Computational Social Sciences

Daniele Miorandi
Vincenzo Maltese
Michael Rovatsos
Anton Nijholt
James Stewart *Editors*

Social Collective Intelligence

Combining the Powers of Humans and Machines to Build a Smarter Society

Springer

Computational Social Sciences

A series of authored and edited monographs that utilize quantitative and computational methods to model, analyze, and interpret large-scale social phenomena. Titles within the series contain methods and practices that test and develop theories of complex social processes through bottom-up modeling of social interactions. Of particular interest is the study of the co-evolution of modern communication technology and social behavior and norms, in connection with emerging issues such as trust, risk, security, and privacy in novel socio-technical environments.

Computational Social Sciences is explicitly transdisciplinary: quantitative methods from fields such as dynamical systems, artificial intelligence, network theory, agent-based modeling, and statistical mechanics are invoked and combined with state-of-the-art mining and analysis of large data sets to help us understand social agents, their interactions on and offline, and the effect of these interactions at the macro level. Topics include, but are not limited to social networks and media, dynamics of opinions, cultures and conflicts, socio-technical co-evolution, and social psychology. Computational Social Sciences will also publish monographs and selected edited contributions from specialized conferences and workshops specifically aimed at communicating new findings to a large transdisciplinary audience. A fundamental goal of the series is to provide a single forum within which commonalities and differences in the workings of this field may be discerned, hence leading to deeper insight and understanding.

More information about this series at http://www.springer.com/series/11784

Daniele Miorandi • Vincenzo Maltese
Michael Rovatsos • Anton Nijholt • James Stewart
Editors

Social Collective Intelligence

Combining the Powers of Humans and Machines to Build a Smarter Society

Springer

Editors

Daniele Miorandi
CREATE-NET
Trento, Italy

Vincenzo Maltese
University of Trento
Trento, Italy

Michael Rovatsos
School of Informatics
The University of Edinburgh
Edinburgh, UK

Anton Nijholt
Faculty EEMCS
University of Twente
Enschede, The Netherlands

James Stewart
The University of Edinburgh
Edinburgh, UK

ISBN 978-3-319-08680-4 ISBN 978-3-319-08681-1 (eBook)
DOI 10.1007/978-3-319-08681-1
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014947873

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Social collective intelligence is an emerging area at the intersection of collective intelligence and social informatics, where social processes between humans are being leveraged and enhanced, by means of advanced Information and Communication Technologies (ICT), to solve challenging problems using the contributions of human collectives. Rather than being a well-defined area, it presents itself—at least for the time being—as a mix of various methods and technologies, such as social media and social computing, human-based computation, social networks and complex systems theory, crowdsourcing, and many other areas which all somehow aim at developing or understanding collectively intelligent systems by combining advanced ICT with the powers of individual and collective human intelligence.

Within this broader area, while novel applications—from mobile social networking services to socially augmented reality systems—are appearing (and disappearing) at an ever-increasing rate, the ability to engineer these systems to concrete design objectives remains, until now, essentially a "black art". Although research in the different areas involved has produced many significant contributions, we are still far from a principled approach for designing and operating these kinds of systems.

This book serves two purposes: On the one hand, while we are not yet in a position to develop textbook-like material for the field of Social Collective Intelligence, we aim to consolidate the fragmented research landscape, gathering contributions that capture the state of the art in all relevant areas, thus providing an up-to-date survey of existing research. In this respect, we put particular emphasis on giving technological and socio-technical aspects equal weight, as we believe that human factors and new technologies need to go hand in hand in developing successful future social collective intelligence systems, maybe more so than in any other area of digital technologies. On the other hand, we focus on the engineering aspect of such systems, thereby taking a distinctly different approach from much of the work done in the complex systems and related social science literature, which primarily focuses on analysis and prediction. While these aspects are also dealt with in several chapters of this book, our objective is to give an overview of appropriate techniques that both scientists and practitioners can use in order to build purposeful and effective social collective intelligence systems.

Based on this overall approach, we expect that this book will be of interest to different audiences: Social scientists who want to understand the computational machinery that drives such applications, and how it interacts with human-centric and societal concerns. Researchers and practitioners in information and communication technologies, who need to acquire an understanding of the socio-technical dimension of these systems, as well as a comprehensive overview of relevant computational techniques. Various stakeholders from businesses, public organisations, and the general public, who want to go beyond a naïve understanding of novel technologies emerging in this area and require adequate knowledge of theoretical foundations and technological potential to make informed decisions, whether this be for commissioning novel systems, regulating their use, or even actively participating in them as a contributor. And, finally, graduate students from various disciplines who are looking for a comprehensive treatment of all aspects of this new type of systems.

This book is divided into three parts: Part I comprises of several chapters covering the foundations and theory behind Social Collective Intelligence. These provide an overview of the area, discuss opportunities and challenges, and investigate fundamental issues and problems. In Part II, we cover the some of the key technologies that are needed to develop social collective intelligence systems. This part addresses core techniques and approaches that can be useful for systems development and analysis, but also more peripheral concerns relevant to the "ecosystem" of social collective intelligence applications. Part III concludes the volume with descriptions of key application domains and several case studies from which insights and lessons can be learnt.

Trento, Italy	Daniele Miorandi
Trento, Italy	Vincenzo Maltese
Edinburgh, UK	Michael Rovatsos
Enschede, The Netherlands	Anton Nijholt
Edinburgh, UK	James Stewart
April 2014	

Contents

Part I Foundations

Part II Technologies

Part III Applications and Case Studies

Part I
Foundations

Towards the Ethical Governance of Smart Society

Mark Hartswood, Barbara Grimpe, Marina Jirotka, and Stuart Anderson

1 Introduction

Smart Society[1] is a term coined by an EU funded Integrating Project (IP) of the same name that aims to capture how contemporary techno-social trends can be harnessed towards solving challenges facing modern society. The "Smart" alludes to the enabling capabilities of innovative, social, mobile and sensor based technologies that in various way are envisaged to create more productive alignments between (growing) demand and (constrained) resources across a number of sectors and application domains.[2] A key example of this is how to meet growing care needs with diminishing resources as the number of elderly people grows as a proportion of the overall population [9]. While the challenges of urban life form the test bed for the Smart Society project, it is likely to become increasingly relevant in other domains such as finance [6].

[1] Smart Society (FP7/2007–2013) Grant agreement n. 600854. http://www.smart-society-project.eu/

[2] Although this might be seen as a tale of improved matchmaking, there are also important nuances in the ways that social computing transforms the sorts of demands, goods and services that are in play. Thus, for example, social platforms can make visible the "hidden" care demands of elders, and also the "hidden" skills of neighbors, and create a market place in which these may be traded (e.g. http://ce.livingitup.org.uk/hidden-talents-scotland/).

M. Hartswood (✉) • B. Grimpe • M. Jirotka
Department of Computer Science, Oxford University, Oxford, UK
e-mail: mark.hartswood@cs.ox.ac.uk; barbara.grimpe@cs.ox.ac.uk; marina.jirotka@cs.ox.ac.uk

S. Anderson
School of Informatics, University of Edinburgh, Edinburgh, UK
e-mail: stuart.anderson@ed.ac.uk

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_1

Smart Society is partly inspired by the idea of the "Smart City", a multifaceted concept [19] that recognizes the benefits of urban living but also the strains that are developing on existing infrastructures and resources due to urban growth. The vision relates how cities made "smart" will be more productive, more sustainable, and pleasanter places to live. One aspect of Smart Cities concerns augmenting service infrastructures (such as transport, energy, health and so on) with sensor-based digital technologies able to visualize patterns of service delivery and use stretching across space and time and with a high degree of fidelity [28]. The idea is that service operators can utilize this information to make efficiency savings by tailoring provision to match demand, and by shaping demand through use of incentives or other motivating feedback mechanisms. At the same time shared resources can be used more effectively if users are aware of the global state of the resource and able to coordinate between themselves about how the resource might be utilised. For example, road users can chose an alternative route if they are made aware of patterns of congestion, and if given the right tools they can offer each other advice based upon their local perspective and knowledge.

Smart Society extends Smart City thinking in a number of ways, for example, by including the ideas of:

Hybrid computing How people and machines working together create new sorts of problem solving capability, for example, as in the "wisdom of the crowds"—but also stemming from peoples' everyday use of their mobile connection to data, algorithms and social networks to solve problems.

Adaptivity Bringing to the appropriate sub-collective to bear to solve a given problem; and

Learning Accreting knowledge of how the system responds to different circumstances and using that to drive subsequent rounds of adaptation.

Smart Society is founded on the idea of "collectives"—groups of people linked by a common identity yet having diverse skills, needs and values. On this definition an example of a collective may be "road users", incorporating several sub-collectives of pedestrians, cyclists, motorists, and bus users whose common identity is established by their dependence on a shared roads infrastructure, but whose needs, values and skills will vary considerably between these different "categories" of use. In Smart Society, collectives are seen as a source of expertise that may be accessed and exploited. At the same time they are consumers of resources whose patterns of consumption can be shaped by appropriate interventions. Diversity within collectives on the one hand provides a resource pool to enable a collective develop a range of responses to a situation, but on the other it is also source of friction and contention. Taking these elements together, the socio-technical entity powering the Smart Society vision is referred to as a Hybrid and Diversity Aware Collective Adaptive System, or HDA-CAS[3] for short.

[3] On the occasions in this chapter when we refer to CASs we are considering Collective Adaptive Systems more generally and not only the Hybrid, Diversity-Aware sort.

Such collectives already exist and are routinely formed on a more or less ad-hoc basis, for instance, through social platforms such as twitter. An example is a collective anchored via the #UKSnow hashtag which collaboratively monitors the impact of winter weather thereby inducing individuals to adapt by adjusting their travel decisions.[4] Smart Society aims to engineer more powerful CASs that behave in more predictable ways, that penetrate further into critical or economically significant city infrastructures and services, and that implicate multiple and diverse user constituencies. These elements of scale and intentional design bring with them a series of risks, including those of naively fixing a narrow range of values and of overlooking the need to create governance structures able to evolve and also to mediate between diverse and conflicting value systems.

Thus a key aspect of Smart Society is how to govern them in ways that permit conflicting and diverse perspectives to co-exist within a large-scale evolving CAS, and this question broadly frames the work we present in this chapter. Smart Society from its inception aimed for a multidisciplinary approach to engineering HDA-CASs that incorporates social science understandings of collectives, and ethical orientations to research and innovation [8]. Triggered by a series of recent EU initiatives and research projects [30, 35] the role of the authors in this project is to bring Responsible Research and Innovation practices [42] to a range of technologies, including CASs, in order to shape their impact upon privacy and other social values. An important aspect of this has been to work towards a framework for the *ethical governance* of HDA-CASs.

Our approach has been to couple a conceptual exploration of governance to a social science enquiry into domains where CASs are envisaged, or where CAS-like systems already exist. This unpacking of governance has lead us to the view that the ways in which *CAS might be regulated* (to operate in socially acceptable ways) are quite intimately tied to the ways in which *CASs themselves aim to regulate collectives* (for example, through targeted incentives). Another way of saying this is that CASs gain their effect by instantiating particular forms of social regulation, and moderating how this is achieved is key to producing CASs that are sensitive to important social values.

Exerting influence on a CAS's participants is central to Smart Society, as the Description of Work (DOW) for Smart Society makes clear:

> "[The aim of the project is] to develop novel incentives, mechanisms and decision-making algorithms able to drive the emergence of desired system-level behaviours in HDA-CASs taking into account the wider information environment and non-incentivised motivations ... To introduce a programming paradigm and an architecture for the management and control of HDA-CASs in a goal-oriented fashion." [10].

In a publication giving an overview of Smart Society these sorts of aims are couched in terms of an everyday example using slightly less technical language:

[4] http://uksnowmap.com/ mashes up #UKSnow tweets and Google Maps to show geographical patterns of reported snowfall, thus providing a sustaining focus for the collective and a mechanism to propagate snow reporting practices through example and a weak obligation to reciprocity.

> "From the analysis of sensor data, machines can "understand" (from low level analysis) that a critical traffic situation has arisen. This initiates a hybrid computation that calculates the best incentives to offer different strata in the driving population in order to align driver behaviour with global policy objectives. ... Incentives will be given to particular target groups depending upon their needs and expectations. People can ignore such suggestions and decide autonomously on what they believe is best for them" [16].

The above description reveals the Smart Society vision to be a complex one that admits diversity and acknowledges conflicting perspectives. The vision allows for autonomy while at the same time seeking to influence with incentives and persuasive technologies. It aims for a degree of self-regulation by giving participants access to information and resources and broadening their capacity to act, whilst at the same time seeking to impart direction and make wider patterns of behaviour align with centrally defined goals. At first inspection, one might doubt the vision's coherence, fearing that it contains inherent and insurmountable contradictions (e.g. autonomy versus control, centralised versus self-regulation, individual versus public interests), and yet as we explore the notion of governance more thoroughly, we discover that it is common, perhaps inevitable or necessary, for multiple governance regimes to coexist simultaneously.

The vision for Smart Society articulated in these quotations raises deeply significant social and ethical issues, including:

- who will set the incentive structures or algorithm parameters?
- who gets to set the ultimate direction or goals of the Smart Society—and if this is the State, what new forms of democratic conventions will be needed to control this new and powerful way of implementing policy?
- will we be aware of the machinations of the "unseen hand" that filters the information we see, targets us with incentives and chooses which resources we can access?
- how are the conflicting interests and perspectives of multiple user constituencies mediated?
- and crucially who will be accountable for the effects of the HDA-CAS should things go wrong?

These questions resonate with the general and long-term problem of governing the global knowledge society in a "smart" way [43].

The chapter is organised as follows: First we present our methodology (Sect. 2) and then explore several potential ethical consequences of the Smart Society vision (Sect. 3). This leads us to unpack the concept of governance to better understand the different forms of regulation and their relationship to one another. We observe that modes of governance are not mutually exclusive, but are rather blended in different proportions to achieve different sorts of regulatory effects (Sect. 4). Understanding how forms of social regulation work provides a foundation for understanding the sorts of regulatory effects that HDA-CASs can achieve. It also provides a basis for applying them mindfully, with care and forethought. Finally, we draw upon a "worked example" to show how governance design can be pursued in a way that is sensitive to social values and emerging ethical concerns (Sect. 5).

2 Methodology

It is impossible to study HDA-CAS "in the wild" as the sorts of HDA-CAS envisaged by Smart Society do not yet exist. Thus to understand the implications of HDA-CAS for ethical governance we need to adopt a series of more indirect approaches. To achieve this we have explored:

1. *Emergent ethical issues of contemporary trends in networked, social and mobile computing*: This has principally involved exploring the extensive existing literatures on this topic.
2. *Existing systems or programmes that have some properties in common with HDA-CASs, or that are driven by a similar vision*: Here we have conducted a series of "elite" interviews with powerful stakeholders driving the Smart City agenda. "Elite interviews" aim to explore and learn from the experiences of those in positions of power and influence within a particular arena, be it politics, business, academia or the public sector [1]. This approach allows us to access the accumulated learning accrued from implementing real-life Smart City visions and CAS-like systems. To date we have conducted the following interviews: Senior police officers (2); IT consultant developing SmartCare "apps" (1); Smart City academics (2); Smart City consultants and system integrators (4); Manager of a Regional Intelligent Traffic Management system (1); Civil Servant facilitating Smart Cities programme (1).
3. *User perspectives in contexts corresponding to Smart Society scenarios*: Focus groups with tourists (young travellers) (1); Interviews with "Ride Sharing" scheme participants (8).[5]
4. *Reflective discussions within the Smart Society project itself*: Smart Society project members naturally reflect on the ethical potential of the technologies during co-located and virtual meetings across the project and these are valid and valuable forms of insight.

3 Addressing Ethical Issues in Smart Society

Drawing upon the above empirical work we have found it useful to distinguish between *contextual* and *emergent* ethical issues in relation to CASs. Contextual issues refer to pre-existing ethical sensitivities within a given socio-technical system that reflect interactions between cultural values, supportive infrastructures and system goals. Emergent ethical issues are ones that arise, or are amplified or diminished as a consequence both of reengineering an existing system to function more like a HDA-CAS, or by virtue of the CAS's evolution. This distinction is important because it enables us to take seriously pre-existing ethical concerns,

[5] In collaboration with our Smart Society partners at Ben Gurion University.

whilst at the same time keeping an open mind as to which ethical issues will assume importance in the future. This awareness of emerging system properties and corresponding ethical issues builds on practice theories of socio-technical order [34, 36]. In the case at hand, it entails an ongoing process for identifying and managing ethical concerns that should function continuously as the HDA-CAS is implemented, as it evolves and as it interacts with wider social and socio-technical systems. Given that ethical concerns are often debatable, conflicting or present as dilemmas then we need to avoid the idea that we can, for the most part, solve ethical problems (cf. [22]). Rather we wish to provide a space for them to be surfaced, negotiated and to enable working compromises to be achieved. We take these processes of identifying and managing ethical concerns to constitute the "ethical governance" of HDA-CASs. We develop some preliminary ideas as to what this governance process should look like and how it intersects with other aspects of CAS governance later on in this document. Here we prime that discussion by drawing upon empirical data to focus on categories of ethical issues that appear relevant to HDA-CASs and Smart City application domains. The intention is to create a sensibility towards relevant ethical concerns, including particular sorts of contemporary or domain specific ethical issues, but also to point to categories of issue attached to wider techno-social trends and anticipated HDA-CAS properties.

3.1 Contextual Issues

A preliminary analysis of interviews and focus groups conducted as to inform this chapter has revealed a variety of pre-existing ethical sensitivities in domains such as social care, tourism and transport. One such example revolves around the safety concerns of those participating in schemes that support "Couch Surfing"[6] as a source of cheap accommodation, and "Ride Sharing"[7] as a means towards inexpensive travel. Some female travellers in particular were concerned they may be exposed to risk using these services for instance if they accepted a lift at night alone with a man they did not know. Interviews and focus groups revealed variation in degrees of concern and an array of ad-hoc strategies used to reduce risk. These included: avoiding use of the service altogether; preferring a telephone conversation to arrange the ride to help gain an impression of the driver's character; keeping a personal record of driver "reputation"; becoming less cautious with experience; and by choosing "safer" rides (e.g. a daytime ride with other passengers). Thus a concerted investigation into a setting can provide valuable insight into important social values that need to be accommodated within the design of a CAS, and yet the

[6] "Couch Surfing"—taking advantage of casual services provided by locals such as offers of accommodation in private homes.

[7] Schemes that allows drivers and commuters to offer and accept lifts and share costs by utilising spare capacity in the cars of those already intending to travel.

process of accommodating social values is often not straightforward. One reason for this is because different social values often compete with each other. In the Ride Sharing schemes, for example, it is hard to balance the need to enhance privacy on the one hand with the need to reveal personal details about drivers and passengers to enhance safety on the other. Layering on properties envisioned for HDA-CASs adds further intricacies to these already complicated situations. An example of this is that users of existing Ride Sharing schemes can choose freely from offers of lifts, whereas Smart Society would use incentives to steer that selection, perhaps to encourage optimal journey times or maximum occupancy. This has the effect of shifting some of the responsibility for choosing a ride to the CAS, implying that if someone should come to harm then liability may be attached to the CAS or its designers. With these complexities in mind our approach is not to attempt a fixed design that roughly satisfies constraints of competing social values as they exist at a point in time, but instead to use an enquiry into social values to inform the design of flexible governance structures that can be renegotiated and modified as circumstances change and as the system evolves. We cover this topic in detail in Sect. 4.1 below

One thing to note from this discussion is that contextual and emerging ethical concerns are not entirely separable. Starting from contextual issues, it is quite natural to then consider how a planned implementation may “mangle” those issues into new types of problem [34]. Thus, understanding existing issues forms the basis for anticipating emerging ethical dilemmas.

3.2 Emerging Issues

Emergence is a key feature of HDA-CASs, and new sorts of ethical dilemmas may arise alongside emerging capabilities and impacts. Forecasting future ethical concerns for evolving, complex, open-ended systems seems a hard task. However, practical methods have been developed towards envisioning a range of alternative possible futures to provide traction for design choices made in the present [17]. These fall under the rubric of “anticipatory governance”, defined as the coupling of foresight and policy to achieve earlier responses to the “unexpected” or emergent consequences of non-linear systems [15]. In this context, foresight is not taken as prediction, but rather as a resource for negotiating possible futures that is informed by combining several sources of knowledge, including: hindsight (i.e. awareness of prior “surprises”), awareness of techno-social trends and dynamics, expertise and perspectives from a range of stakeholders and disciplines, domain overviews, and model based forecasts [15, 17]. Our approach throughout this section has in a modest way been to utilise some of the above anticipation and foresight approaches to understand the implications of governing CASs. For instance, in the remainder of this section we draw out lessons for social values from the accumulated experience of large-scale socio-technical systems with properties similar to CASs that is available from the literature and from our own empirical work. In later sections we seek to understand how CASs may regulate collectives, and anticipate the different propensities attached to alternate governance regimes. Finally we draw these together with an empirically founded "worked example" (i.e. one drawing upon domain expertise) that considers the governance requirements of a HDA-CAS in a care setting detailed in Sect. 5.

Here we return to possible emergent ethical issues for CASs based upon the sorts of social transformations already wrought by existing complex socio-technical systems:

Social Sorting: CAS that are diversity-aware aim to be sensitive to the mix of capabilities and values present within collectives, and able to stratify populations to target incentives and recruit expertise. However, such an approach is open to undesirable forms of social sorting, identified as the ways that surveillance technologies sift populations and thereby regulate entitlement or access to resources [25].

Representation and transparency: Who decides the global goals a CAS should pursue, and to what extent will participants understand that their behaviour is being directed through the use of incentives and persuasive technologies? Although CASs are envisaged as creating societal benefits, various forms of accountability are needed to ensure such ends are not subverted. It may be suspected that CASs really aim to make life more convenient or lucrative for well-off sponsors, thus certain forms of transparency become needed to preserve confidence and trust.

Direction versus autonomy: The metaphor of "herding sheep" has been used to explore how the behaviour of collectives can be directed[8], raising the question as to who gets to set the system's direction—or train and influence the sheepdogs? Similarly the "God of the Smart Society" has been proposed[9] as an evocative metaphor for the unseen hand guiding the collective's behaviour, raising the question as to whether a Smart Society should be more paternalistic or more democratic in its constitution?

Incentives and their effects: Attempts to influence human behaviour can result in "perverse outcomes" on those occasions when incentives drive unanticipated and undesired behaviours [37]. This raises issues about monitoring CASs to ensure that their emergent properties are positive and intended. This becomes harder to achieve as the system scales because of the increasing diversity of outcomes and the increasing diversity of views over what outcomes are actually desirable. So although noble intentions are envisaged for CASs such as, reducing traffic congestion or pollution, or creating community goods where none existed previously, defining such intentions will in practice depend upon negotiating between contested perspectives.

[8] "In a similar fashion to herding sheep, the goal is to steer a group of living individuals to comply with our goals." [2]

[9] By a member of the Smart Society project during a project meeting when the conversation turned to types of ethical concern raised by the project.

CAS boundaries are a further site of ethical concern. Will non-participants be disadvantaged? One can image that business owners who depend on passing trade will be upset by changing commuting patterns as drivers participate in a CAS that aims to reduce congestion. Will these "indirect" stakeholders be given a say in how those CASs are configured?

Hybridity within a HDA-CAS aims to blend the capabilities of humans and machines to solve problems either would struggle to solve alone. Questions arise here whether participation is fairly rewarded, whether professional roles are displaced, and how to guard against malicious forms of participation [24,40].

Flows and mobilities: Attention is needed to the wider impacts of CAS across time and space as they alter flows and mobilities within the proposed Smart City setting. This is because CAS aim to influence the movement of traffic, people, material and immaterial goods, patterns of consumption, transform the knowledge, skills and resources needed to participate in markets, access services and engage in political discourse. With all of these effects there are likely to be winners and losers. As the authors of [18] have argued, increasing the mobility of some stakeholders may entail "immobilities" for other groups.

Automation raises a gamut of issues including the degree of control ceded to algorithms, the redistribution of responsibility and liability (discussed above for the Ride Sharing scenario), the performative shaping of participation (e.g. job applicants aligning their behaviour to the matching algorithm in online job markets such as "Elance"[10]), the opaqueness of algorithms and their adaptations, and the filtering effect they have on human experience of the world [12,20,21,23].

Personal integrity: Finally there are a series of values that relate aspects of personal integrity and autonomy such as trust, safety, security and privacy, some of which are discussed above, and others come to the fore in discussions of privacy elsewhere in this volume.

4 Governing Smart Society

This section sets out a simple example to help illustrate principles of governance, their interrelationship and how they are relevant to Smart Society. Our conceptual analysis of governance has lead us to the view that the ways in which *CASs might be regulated* (to operate in socially acceptable ways) are quite intimately tied to the ways in which *CASs themselves aim to regulate* collectives (through targeted incentives, for example). In other words, a more thorough understanding of how different forms of governance interact to deliver social regulation helps not only with working out how to design CASs effectively to influence how resources are used, but also to see how this can be done in ways that are ethically sensitive to different contexts. The example we present in this section concerns regulation of public

[10] www.elance.com

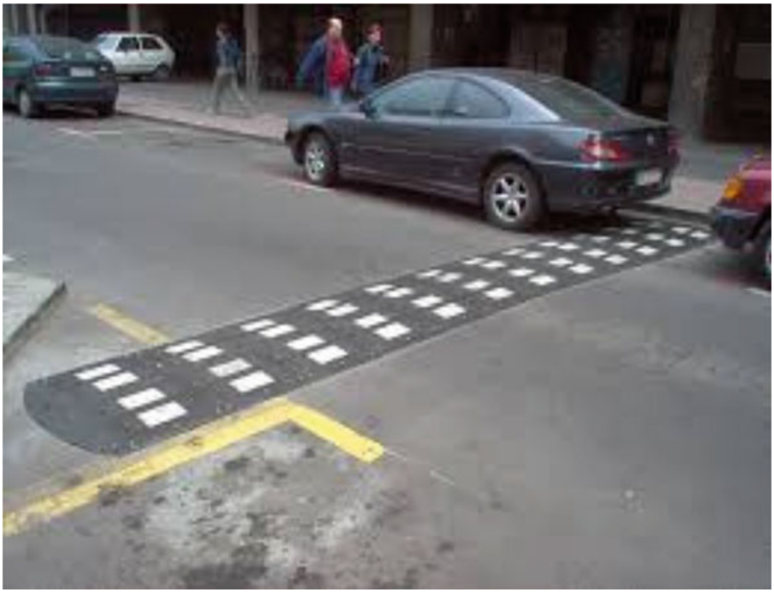


Fig. 1 Example of a "speed bump"

highways to ensure they function effectively as a shared resource. We explore how speed bumps, sometimes known as "sleeping policeman", are employed to regulate traffic speed. We illustrate how "speed bumps" feature simultaneously in several intersecting governance regimes, and discuss how any HDA-CAS must similarly exist at the intersection of several governance regimes. We then make the case that CAS design is shaped by, and shapes, governance design. Finally, we explore what this implies for ethical governance for HDA-CASs.

4.1 Understanding Governance

Speed bumps, like the one shown in Fig. 1, configure the driving environment and help regulate traffic speeds in sensitive areas. They are a small component of a wider system of traffic regulation, which we explore in detail below.

Speed bumps are an example of "**environmentally embedded regulation**" [39] illustrating the approach of configuring the physical environment to constrain driving practices in certain ways, in this case to regulate speed for reasons (perhaps) of pedestrian safety. At a base level the material features of the roads and their organisation create a balance of affordances and constraints that shape the possibilities for road use (e.g. speed and overtaking are possible on straight sections, but not where the road bends). This potential of the built environment to regulate social practices is actively exploited by town planners who configure urban spaces in ways that inhibit crime and anti-social behaviour [41]. Analogously, obtaining desired forms of social computation depends upon carefully structuring virtual user-environments to regulate patterns of social behaviour in specific ways [11]. An example of this is how the moves an ESP game[11] are carefully arranged to produce game play that is generative of useful metadata tags.

Whilst the "rules of the road" might be given physicality in the form of speed-bumps or other traffic calming measures, there are a huge range of regulatory cues (signs, lines, grids, lights etc) that signal conventions of road use but do not by themselves enforce compliance. These are part of a **hierarchical and centralised** mode of regulation deriving from legal or institutional authority and policed by the state. Drivers are socialised to these rules formally via driving lessons and the driving test, and compliance is in part maintained through the threat of state (or professionally or institutionally) authorised sanctions. Centralised or hierarchical forms of regulation, in common with other forms, do not determine behaviour. Policing is imperfect, people are willing to risk sanction for some other benefit and the interpretation of rules is a matter of social convention, as is the degree to which they are enforced. So although shared norms and conventions amongst drivers take account of legally sanctioned regulations, they are not wholly determined by them. An example of this is the difference between the actual speed limit on UK motorways (70 MPH) and the *de facto* speed limit which is closer to 80 MPH.[12] Moreover, circumstances continually arise as part of road use that require improvisation and negotiation that would be impossible if official regulations were adhered to rigidly. In computing, this type of regulation is perhaps analogous to the *terms and conditions* attached to services that typically include expected standards of behaviour, allowable and prohibited ways that the service might be used, and sanctions that might be applied should the code deem to have been broken.

The calming effect of speed bumps depends on drivers noticing them, anticipating the jolt and adjusting their practices accordingly—a process that can become more automatic over time. Much of the moment-by-moment organisation of road use depends upon a mix of prior socialisation and situated decision-making, including an appraisal of environmental cues, what other drivers are doing or are likely to do, what the local conventions are, and expectations of how certain traffic situations are likely to evolve [7]. This in turn depends upon reading the intentions of other road users, signalling one's own intentions, continually adapting one's own approach in response, as well as adjusting to the adaptations of others. This can

[11] A serious game used to generate image metadata such as descriptive tags http://en.wikipedia.org/wiki/ESP_game

[12] A concern voiced about raising the official limit to 80 MPH is that the de-facto limit will then become 90 MPH. The difference arises due to cultural expectations about how regulations are policed. In the UK there is an expectation that the police will not enforce the rule rigidly, but instead allow some leeway, which for all practical purposes leads to raised limit. http://www.independent.co.uk/news/uk/home-news/motorways-not-safe-enough-for-speed-limit-rise-to-80mph-7745678.html

be seen as a form of **polycentric governance** [32],[13] often contrasted to more centralised and hierarchical forms of social regulation, whereby communicating agents *collaboratively self-regulate* their use of a shared resource. This has components of mutual accommodation, sanction and reward, and plays into processes of community norm formation. Polycentric governance is seen to underpin the regulation of knowledge creating communities within Wikipedia, where formation and policing of community norms occurs as part of the communicative practices of community members, rather than being imposed externally. It is also visible in the "discussion fora" of sites like "Zooniverse" where a shared understanding and classificatory practices can emerge for what would otherwise be isolated decision-making tasks of individuals classifying astronomical objects.[14] There are a number of attributes that make polycentric governance a possibility—but a principle among these is "cheap talk" [32]—i.e. easily accessible channels of communication between users of a resource. Design of HDA-CASs should orient to the channels of communication available between participants to take advantage of this type of self-regulation.

Speed bumps are a **motivational** form of governance [31]. They threaten discomfort, the chagrin of passengers and damage to the vehicle should a driver maintain an inappropriate speed. (Of course a thrill seeking teenager might find the bumps a motivation for driving faster.) Many types of social regulation seek to influence human actions through rewards and sanctions built around understandings of how peoples' actions are motivated.[15] Smart Society aims explicitly to regulate the use of resources though motivational mechanisms such as, incentives, persuasive technologies and reputation services. These are also common approaches to Smart City applications and a feature of interviews with Smart City consultants and implementers. Thus programmes towards more effective domestic energy use outlined by interviewees turned upon making energy consumption visible and therefore accountable,[16] either on a household or neighbourhood basis, perhaps with explicit elements of competition and reward. Sometimes motivational aspects were present in stronger or weaker forms. For example, one interviewee in charge of a regional transport information service wanted to encourage network users to use public transport as often as possible and always provided a public transport

[13] Admittedly speed bumps are somewhat peripheral to polycentric modes of governance. But as we argue below, all the forms of governance presented here are interrelated. Thus how the driving environment is organised (including the presence or absence of speed bumps) shapes the sort of polycentric responses that are possible.

[14] https://www.zooniverse.org/

[15] Benkler suggests there are three classes of reward that people are motivated by: Money, Pleasure ("Intrinsic hedonistic rewords") and Social ("Social-psychological rewards") [4].

[16] There are a whole series of ethical issues attached to playing off accountability arrangements, particularly how they can create pressure that vulnerable people may be particularly susceptible to, shape behavior in unwanted ways and encourage "gaming" of the system. The worked example at the end of this chapter shows some of these properties for a technology of accountability operating in a care domain.

option in query results, but stopped short of using explicit incentives, partly so that responsibility for the choice remained with the user.

Speed bumps are an adaptation. They are typically placed in response to neighbourhood concerns or other evidence of incautious driving. The approach of adjusting governance measures in response to changing circumstances is referred to as **adaptive governance** and comprises of iterative cycles of monitoring, policy formulation and implementation [27, 39]. A key element of adaptive governance as applied to socio-environmental systems is to bring together diverse forms of expertise, particularly "native" expertise of people living within the system as to how complex socio-ecological systems might evolve in response to change (ibid). In the context of HDA-CASs, adaptive governance would involve forms of reflection that would bring together the expertise of smart society participants with a range of aggregated data describing how a HDA-CAS is behaving. Adaptive governance processes correspond to the cycle of sensing and adapting envisioned for HDA-CAS that will enable it respond to changing circumstances. However, an evolving CAS will most likely produce unpredictable and non-uniform responses to change—be they as a result of new regulations, counter-adaptations, new ways of making measurements, or environmental changes—in ways that demand the renewal of governance arrangements.

The "speed bump" sits within a nexus of diverse concerns voiced by many interested parties,—road users (of varying stripes), pedestrians, residents, motoring organisations, emergency services, environmental organisations, safety campaigners and so on. In this respect the roads analogy bears a strong resemblance to the ambition of HDA-CAS that aim to support diverse user groups with conflicting interests, since road users often have diametrically opposed interests (e.g. cyclists and motorist) and yet have to be accommodated within the same network. The mechanisms by which these voices are heard, how influence is wielded and how resources are allocated form the system of **political governance** of the highways, usually handled in a multi-tiered way via local and national governments and their agencies, but also via other forms of political expression such as campaigning activity. Political governance is a way of organising power and influence. It can be configured to respond to the diversity of interests and values that have to be brokered to create a functional network that roughly satisfies the requirements of many different users and user constituencies. In order to help satisfy the requirement of diversity within HDA-CAS, thought has to be given as to how those user constituencies can influence HDA-CAS configurations.

4.2 Governance Mechanisms as Layered and Intersecting

It should be clear from the above illustrations that managing a complex shared resource like a roads network involves a constellation of governance mechanisms operating simultaneously that serve a variety of purposes whilst at the same time continually interacting and influencing each other. For example, polycentric and embedded regulation do not preclude one another, but instead tend to occur in mutually supportive (or sometimes disruptive) arrangements. Thus, a junction regulated by traffic lights still depends upon the self-coordinating practices of drivers to achieve its effect. When the lights break down, then traffic will typically continue to flow, but its management shifts towards greater polycentric regulation as the drivers themselves now have to coordinate turn taking [3]. Similar sorts of interdependency relationships can be found with motivational regulation. Coexisting governance arrangements are visible in the way that separate studies of Wikipedia alternately highlight either motivational or polycentric governance mechanisms as accounting for peer production in Wikipedia [4,13]. We argue that these are different perspectives on a composite phenomenon, rather than competing explanations.

Adaptive governance can be seen to intersect with polycentric, motivational, and embedded modes in aiming towards specific regulatory effects by iterative modification of the physical, informational or incentive structures that underpin those regimes. Similarly political regulation operates over a slower time frame (except for some campaigns being enacted as deliberately surprising, quick interventions in public space) and can also appear "layered on" to other mechanisms[17]—although experience of the roads network, communication with other users and access to data about the network are all possible occasions or venues for political discourse or action. Figure 2 shows roughly the relationships between different governance regimes and how they may correspond to Smart Society concepts of evolution and operation. Table 1 shows sample governance mechanism and implementation approaches relevant to computer applications.

Building a CAS can be seen analogously as designing and implementing an ecosystem of governance mechanisms that caters for a diversity of users and fosters the emergence of certain patterns of resource use. This is not the same as designing the behaviour itself. Rather, the relationships between these governance elements need to be carefully thought out in order to allow the system as a whole to emerge in a coherent way.

4.3 Ethical Governance

Now we turn to the role that ethical governance has in relation to these various governance regimes. To maintain the analogy with a roads network, we can consider how the road builders and maintainers may have parallels with the designers,

[17] An article on the history of Speed "Humps" in Berkley on the City Authority's web page (http://www.ci.berkeley.ca.us/ContentDisplay.aspx?id=8238) tells of how speed humps became contentious and how opposition to them led to shaping how humps are used as an adaptive regulatory measure ("speed hump locations chosen must provide clear safety benefits to balance any potential negative impact").

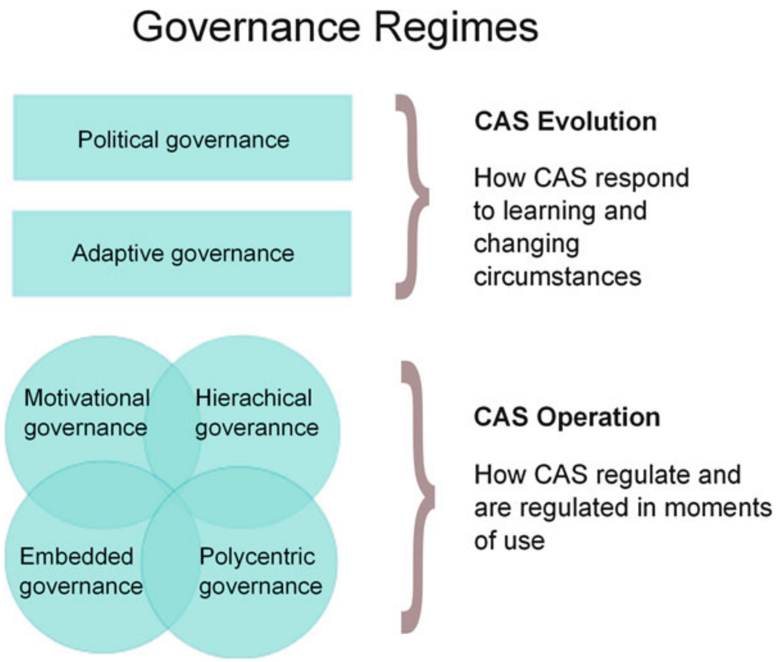


Fig. 2 This figure shows a rough logical arrangement of governance regimes and their relationships to CAS concepts of evolution and operation. This diagram simplifies tremendously the complexity of the relationships between these different aspects of governance

developers and builders of CAS.[18] The road builders wield considerable power over road users in the decisions they make about which roads are built and how the traffic network is regulated—decisions that can affect livelihoods (e.g. where businesses are dependent on passing trade), safety, quality of life (both of drivers and neighbourhoods), the comfort of driving, and impact upon the environment. Designers and implementers of CAS will wield similar powers with respect to a given domain of CAS implementation. Taking care in the production of governance regimes for CAS could include consideration for:

The impact of regulation. Orienting to the practical circumstances in which the activity takes place and considering if the regulation itself poses annoyance, frustration or potential harm to users. The "speed bump" example works well here, because as a mode of regulation it can be potentially very annoying as well as cause damages to vehicles if not noticed. The one in Fig. 1 is painted white to help make speed regulation via bumps less uncomfortable and more palatable.

[18] Assuming the Collective Adaptive System doesn't emerge "spontaneously" as an effect of integrating existing infrastructures and regulatory functions.

Table 1 Sample mechanism and implementation approaches for different forms of governance

Governance regime	Mechanism	Implementation approach
Polycentric	"Cheap talk"—ability to sanction	Discussion boards, chat channels, collaborative filtering, provision of information about the state of the resources and resource users…
Motivational	Seeking of monetary, social-psychological or hedonistic reward. Avoidance of sanction.	Policing, monitoring, logging, reputation services, incentives…
Environmentally embedded	Structuring physical or virtual environment to achieve regulatory effects.	Visibility arrangements, signs, alerts, workflow organisation, ease or difficulty of interactions….
Hierarchical	Laws, regulation, codes of conduct, institutionally backed sanctions and policing.	Terms and conditions, service agreements, codes of conduct, monitoring, penalties, exclusion.
Adaptive	Cycles of monitoring, policy formation and implementation.	Sentiment data, sensor data and provenance data, analytics, engagement with users and other experts, discussion fora, AB testing…
Political	Representation and decision-making processes.	A constitution, stakeholder representation, discussion fora, executive officers, voting, petitioning…

Regulating collectives. Adjusting regulatory mechanisms to achieve some new effect has implications at a collective level where understanding the values and social norms associated with the collective, or with communities within the collective, becomes important. An example here lies with the Ride Sharing scheme where interviews with participants reveal a regime of fixed prices between particular destinations based upon communitarian principles of sharing resources and costs. Attempts to raise the price are typically viewed as being "greedy" and resisted. As part of HDA-CAS we might aim to motivate Ride Sharing participants in new ways (perhaps to improve environmental outcomes), but on the basis of existing norms we can see that achieving this via market based principles might be tricky. This might lead us to select a different approach to motivational regulation that relies less on monetary reward for its effect. The Ride Sharing scheme does not have a central constitution or enforcement mechanisms, but it is evident from the interviews that participants orient to a strong set of community norms and standards of behaviour, indicating a strong polycentric aspect to its regulation. Safety has been identified as of key importance to Ride Sharing, and providing for appropriate social regulation to prevent people coming to harm is an important factor to enable a Ride Sharing CAS to gain acceptance beyond single institutional contexts.

Building on existing regulation. A broader principle building upon the above point is to understand, build upon and build out from existing community norms and regulatory mechanisms.

Anticipating the transformatory power of CAS. When CAS are designed to transform how shared resources are managed over existing practices, perhaps by connecting community members in new ways, then one also has to think through what new sorts of regulation might be required in these transformed circumstances. In the Care House scenario described in Sect. 5, potential of CAS to transform accountability regimes, and the danger of losing a qualitative notion of compassion when care tasks are quantified, calls for specific regulatory mechanisms to safe-guard certain core values.

Balancing Governance Regimes. Fashioning an appropriate balance between regimes is important, as each approach contributes important attributes in a mosaic-like way to the overall system of governance. Thus, a builder of a CAS might ask himself which parts of the regulation need to be freer and community directed, and which need to be more rigid and embedded, and which need to be driven by incentives. Failing to think through provision in a particular area could lead to inequity. For instance, a lack of explicit and appropriate structure for political expression could lead to increasing marginalisation of already vulnerable groups.

Understanding values attached to governance. Governance mechanisms themselves are attached to particular values. They can be more or less democratic or participatory in their implementation, for example. Polycentric governance in particular has an important link to autonomy. In writing about digitally augmented mobilities, Buscher et al. propose that people are "served humanely" when representations of the sensed network are used as a resource for "improvised situated action" rather than centralised control [7]. Thus, a system that minimises polycentrism and drives embedded and motivational governance risks being overly controlling and oppressive.

Designing for adaptive governance. At the point of emergence a CAS might carry a lot of intentional design. Once in operation, however, provision should be made for adaptive governance processes to take over the ongoing redesign of the system. This can be kick-started by making the initial design rounds very much like the adaptive governance cycle, with investigations into the prospective domain, participatory policy formation and trial implementation.

Achieving just the right amount of regulation. Governance design should be proportionate to the scale of the system envisaged and the types of communities implicated. Governance of a nationwide traffic network is immensely complicated and intricate, and has evolved to its current form over the entire history of road use. While it serves as a motivating example for this discussion, one should maintain a sense of proportion when bringing the ideas to any real world example.

Adopting governance structures appropriate to the scale of the CAS. As the scale of a CAS changes, it is likely that governance mechanism may become strained and new patterns of governance will be needed to succeed them. For example, issues that can be handled informally between a pair of collaborating colleagues might need a more formal project management structure to be properly managed within an international research team. An example is how within Wikipedia, governance patterns have changed with changing scale and learning within the wikipedia community [13].

4.4 Guidance for Governance Design

This section considers the sort of design procedure one would follow to realise governance mechanisms with the characteristics outlined above. Treating the design of HDA-CAS as if it were a problem in governance design has the helpful property that social values become first class objects for design, as opposed to being "relegated" to informing categories of non-functional requirements which might be addressed late in the day and/or incompletely. That is to say if one wishes to engineer patterns of social behaviour, then one has to understand and work with sociality. Another way to put this is that if we accept that the speed bump's symbolism is in fact part of its regulatory effect, thinking about how to convey values to influence social orders also becomes an important aspect of design [39].

On this basis, we suggest the following steps for design of HDA-CAS:

1. *Understanding an existing collective, its values and modes of regulation* by characterising the domain in terms of how it functions as a social system—the sort of collective that it corresponds to, the important sub-collectives of which it is composed, how the collective regulates itself, understanding what its core values are and the range of diverse values present.

   There are a number of tools that can help surface social values in a concerted way. Perhaps the most prominent of these are Value-Sensitive Design and Reflective Design approaches [14, 38] that depend upon social science modes of enquiry and "disruptive" design practices to probe existing values. An important research issue is to develop these tools to address dimensions of collectivity since current versions focus more on the values of implicated individual stakeholders rather than of communities. A disclosive computer ethics approach can also be used to surface social values that become silently embedded in computer systems [5]. Anticipatory governance too has an important role to play in helping us see the consequences of alternate design choices by generating insights into possible futures. The Care House scenario in Sect. 5 shows how the altering the balance between different governance regimes can have a significant effect on the overall properties of the system, and illustrates entry points for translating knowledge of social values into governance design.
2. *Draw upon existing knowledge and experience* by bringing together diverse forms of information, expertise and interests, including: the "native" domain expertise of CAS participants, sensor and other quantitative data from existing sources, technical expertise, social science expertise and psychosocial understandings of how human practices are influenced by persuasion and incentives.

This reflects the "enquiry" phase of an adaptive governance cycle and implies strong participatory approaches. It also resonates with Responsible Research and Innovation (RRI) maxims of socially embedded and socially responsive innovation [33]. Participatory design approaches can work at scale [29], and it makes sense to implement these by using the Smart Society platform to engage collectives in design-oriented tasks. Finding ways to balance the influence of designers and different constituencies of native participants will provide clues as to the sorts of political governance mechanisms required.

3. *Designing for governance* by drawing on prior steps, the aim would be to identify key regulatory objectives and implement these through a balance of governance mechanisms. These would aim to produce the desired sorts of social organisation and to regulate the system as a whole to behave in ways that are acceptable to the participating collectives.

   Working out how to translate from information about a domain (from prior steps) into operational governance regimes presents a real challenge to innovate design approaches that can help deliver Smart Society applications. Some starting points include: using our understanding of governance approaches as outlined above as a way of structuring the design challenges (e.g. as a "checklist" of issues that need to be covered); developing a toolkit of governance structures, such as discussion fora, voting mechanisms, chat channels, incentive mechanism, transparency arrangements, constitutional statements etc (see Table 1) that can be composed into a working application; providing mechanisms that set limits or boundaries on the platform that constrain CAS behaviour along particular dimensions to anticipate and contain certain sorts of unwanted adaptation.

5 A Worked Example of Governance Design for HDA-CAS in a Care Setting

This example derives from an interview with a research consultant working on a project to explore how proximity sensors worn by care home staff and residents can be used as an aid to "reflective practice" [26]. The sensors register each time a carer comes within 1.5 m of a resident. The carer can then view analytics that show those residents they were proximal to, when, and for how long, as well as how overall contact time is shared between residents. A sensor is also located on the care home computer to indicate how much time is spent on administrative tasks. The idea is that staff can interpret this data to rethink their own practice, perhaps prompting consideration of who they spend more time with, who less, and why.

This example has a number of advantages for exemplifying Smart Society concepts:

1. It is a simple case that can be easily extended to incorporate features that give it the properties of a HDA-CAS (an elaborated version is described below).

2. There are evident social values and governance issues attached to the system's use.
3. It falls within the application area of social care, which is seen as an important focus for Smart Society as it moves forward, particularly in relation to use of sensors to assist the delivery of care.

The discussion below attempts to illustrate some of the issues and potential solutions in the governance of a HDA-CAS based upon the principles outlined earlier. The idea is to stimulate a certain way of thinking about CAS and their design, particularly to give attention to the issues, tensions and contradictions that emerge when applied to a real world context. The analysis is not meant to be exhaustive and many of the disciplines within Smart Society would have strong suggestions as to the sorts of mechanism or approaches that might be used to address the different issues that are raised, particularly how incentives can be effectively configured; how reputation and provenance can be factored in; and how social orchestration can be designed to help create the "right" sorts of hybridity. Finally, the example does not reflect in any way the actual intentions of the Mirror project[19] which created the original sensor based app for reflective practice. The projection of an extended system exists only within the context of Smart Society.

5.1 Smart Society Extensions

While the computer system is able to ***aggregate*** the pattern and duration of contacts, these aggregated traces are not particularly meaningful by themselves. As the interviewee has it: "[the sensor] doesn't tell you the quality of the interaction, it simply tells you an interaction's occurred". Interpreting the sensor trace depends on the care staff supplying missing contextual detail: where do the residents usually sit? Which residents prefer attention, which prefer to be left alone? Which registrations are likely to be "artefacts", and which correspond to "real" interaction? This interpretation of the pattern of contacts by care staff is already a social computation and demonstrates ***hybridity*** between machine and human capabilities. In particular, it shows how human interpretation can help bridge the ***semantic gap*** between sense data and meaning.

Of course, in developing this as a Smart Society scenario, the contribution of human-factors colleagues would be to improve ***activity recognition*** through better sensors and algorithms, although this is unlikely to eliminate the need for human judgment; but perhaps it would alter the sorts of judgment required, with the human needing less to "repair" sensor readings, and able to concentrate more fully on assessing their significance. While human expertise helps bridge the "semantic gap"

[19] The EU Mirror project aims to create a series of applications to support reflective professional practice. http://www.mirror-project.eu/

between sense data and meaningful interpretations, part of the Smart Society vision is to deliver automated support for sense-making and decision-taking in areas where the computation is easiest for the machine. An extension to the proximity sensor system enabling the discovery of helpful permutations of staff given constraints of duty rotas and shift patterns could be an example of this sort of automation. The work within the project on ***lightweight social orchestration*** would be concerned with how the blend of automation and human control is realised in practice.

The example has elements of ***evolution and adaptation*** built-in, since the aim is for the care staff to adjust their practice on the basis of reflecting on sensor data. Simple extensions to the example provide a means to explore ***diversity and scale***. Diversity could be present in a number of ways, including: perhaps different types of sensor that vary in the way they provide descriptions of proximity, or to incorporate the different preferences, knowledge and skills of carers and residents (this may enable the system to help determine combinations of carers best able to meet a resident's care needs because of shared interests or values). Diversity becomes an increasingly important consideration when the system is ***scaled up*** from a single care home to encompass improving care provision across an administrative region. With scale, ***governance*** issues also come increasingly to the fore, since decision-making and planning would be implicated at multiple levels of organization with each level orienting to different sorts of goals, these are unpicked more fully in a discussion of governance and social values below. Finally, there is scope for building in ***reputation mechanisms*** and ***incentives***, perhaps via resident's rating of the care they receive, through "badges" or other rewards for thoughtful practice.

5.2 Social Values and Governance

The issues presented below represent a value sensitive analysis of the care home example based upon the interview data obtained as part of the empirical component of Smart Society, a conceptual analysis based on our understanding of types of social impact, and an analysis of the technology characteristics. The discussion revolves around design based upon the principles outlined in the governance principles discussed earlier in this chapter.

5.2.1 Embedded Regulation

The following quote is a very good example of how values can be embedded in design, of embedded forms of regulation and how the balance can be struck between different regulatory approaches:

> "the original the developers [developer's name] they came up with a kind of dashboard you know - 100% to 0% - critical and colour coded all the way along - Woo Hoo - I said no, no - take off all values - we are not here to tell them what is good or bad, what's critical or what's adequate (. . . ?) not our job."

Table 2 Different models of the sensor based system depending on how far the sensor data circulates

	Extent of data sharing	Accountability practices
1	Only you see your data	Self reflection
2	The data is shared within the team of carers	Group reflection and oversight
3	The data is available to the care home manager	Managerial oversight
4	The data is shared with residents and or their relatives	Customer oversight

The proposed colour coding pre-configures how “readings” of contact time should be interpreted and as such embeds judgments about what constitutes an appropriate level of contact. These inscribed values imply a regulatory effect similar to that of a thermostat where the aim would be to get the “readings” within an acceptable range. This set-up runs the risk of pushing carers to orient to “getting the reading in the green” as a metric of good care, rather than orienting to quality of interactions and individual need. This points to the more generic danger posed by technologies that *quantify* as framing care in terms of metrics rather than as personal, compassionate, empathic and responsive—characteristics of the quality of interactions. It also shows the power and subtlety of regulatory cues embedded within the user environment and how these should be used mindfully and with sensitivity. In the quote, the IT consultant orients towards a more polycentric mode of regulation that favours greater hybridity by placing a greater emphasis on the discretion and contextual knowledge of the professional carers. We discuss this in further detail in the section on polycentric governance below.

5.2.2 Accountability Regimes

The extent and types of information flows that a technology enables are also implicated in various regulatory effects. An extended version of the sensor system can be configured to create different patterns of disclosure to different audiences and thus, bring different balances of regulatory mechanism into play.

Each of the following patterns of disclosure in Table 2 opens up a different dimension of accountability.

5.2.3 Polycentric Governance

If we think of the care staff as a bounded resource that needs to be allocated effectively to meet the diverse needs of residents then we can also see how, within the context of normal practice, a variety of regulatory structures will play a role in managing the shared resource. One aspect of this will be “centralised” management practices such as the production of a staff rota to ensure that there is appropriate "cover" at all times. These specifications will not, however, detail precisely who does what and when, which will be a matter partly of routine, partly of negotiation and partly of response to contingency—i.e. regulation of care resources at certain levels have a high degree of polycentrism. That is to say it is the staff and residents collaborate in planning and self-organise their moment-by-moment activities around a negotiated and continually evolving shared sense of what needs doing and what division of labour would best achieve those tasks (which will of course be reflected in more static instruments such as the rota).

The sensor system of this example provides an additional source of information that can feed into reflective practices crucial to polycentric forms of self and mutual regulation. As an aid to self-reflection where a staff member only sees data corresponding to their own activities, this perhaps will prompt them to make adjustments to their own work practices. Sharing everyone's data between all team members perhaps has a greater potential for insights, ideas and mutual reweaving of priorities, practices and routines. It will also carry greater risks (in extreme cases maybe associated with work place bullying), and will exert subtle pressures toward conforming to the metric of the system.

5.2.4 Motivational Regulation

One way of viewing the sensor system may be like a rather neutral source of information that can be incorporated into reflective practice to optimise use of a constrained care resource. Another is to acknowledge that at the same time, sensor reading can carry very strong moral overtones as to, for example, whether staff are performing as they should, and whether residents are receiving equal and appropriate care, and so on. Hence the high degree of sensitivity that can be attached to how far the sensor traces circulate how easily subjects can be identified. Thus, while in the original example the system is intended as an aid to *reflective practice*, this ostensive purpose is not fixed, and the tool's strong *evaluative potential* in particular, is something that people can seek to exploit:

> "one of the reviewers he clearly cottoned on to it very quickly and said you are really on to something here - you could sell this, it says, as a quality assess- assurance for relatives - so it's not the carers that get the data it's the relatives that get the data and you think 'oh my god' you know - but that's exactly your issue now - how far down that road - whose data is it?"

It is a very common experience that people are motivated to adjust their practices if they feel they are being observed or assessed, and it would be easy to behave in a way that gave a "positive" account of resident contact time without actually increasing positive interactions with residents. Thus adaptations motivated by these new types of accountability (from managers or relatives) may be quite negative, and may devalue the sensor systems' use as an aid to reflection (because the sensor reading can no longer be trusted).

5.2.5 Adaptive Governance

In the section above on embedded regulation, we saw how the IT consultant argued against the use of "colour coding" precisely to remove evaluative connotations. We can see this as a very simple instance of adaptive governance, where the technology is reconfigured to deliver a different regulatory effect by reflecting upon and anticipating its likely or actual effect.

In the above sections we have formulated a problem. The sensor tool threatens to connect residents, managers, relatives and carers in new ways creating new means of surveillance and accountability that contain the possibility for unwanted and unhelpful adaptations, as well as positive ones. In expanding the system to help beyond personal reflective practices, we have to think of the forms of adaptation that might enable these different functions to more happily co-exist.

One strategy might be to use techniques of anonymisation or aggregation, so that data can be examined at a management level or beyond without implicating individuals or individual care homes. This data would still likely be useful, although not ideal, but provide less strong motivations to "game" the system.

Another might be in finding ways of keeping the carers honest such as, enabling residents to annotate data to give some indication of the quality of the interaction in contexts where this may be possible.

There are many further possibilities and combinations of possibilities that have the potential to shape different patterns of practice. These occur at different levels within the system with different implications for the quality of the data that emerges and whether the "real" goals of the system are being met. The point of adaptive governance is that these types of solution should be investigated, trialled and re-evaluated in an ongoing loop of information gathering, discussion and experimentation.

There may be a number of adaptive cycles at different "levels" within the system. Thus the care staff themselves might experiment with different ways of displaying, sharing and interpreting the data locally that helps maintain an emphasis on the "human" elements of care. While at the same time similar processes could be occurring for how data across the region is used to inform care policy, staffing levels and so on.

5.2.6 Political Governance

> "But you could imagine - or you could very easily imagine - care home managers deciding that they would want to find these things out and the carers will wear these sensors whether they like it or not and there could be problems without a doubt because - we did come across a couple of carers that didn't want to wear them. And obviously, you know, we didn't force them although- . . . I mean it was a small group because we I think there was nine carers in this group and one of them I remember in this test just felt comfortable but peer pressure carried the day and so she says "ok I will do"."

This quote points us towards the politics of the workplace, and by extension, wider spheres of political involvement that would come to encompass unions, professional bodies, governments, resident and relative care pressure groups, particularly as the scale and scope of the system expands.

One issue that is likely to have political ramifications is how such an expanding system would change the nature and character of care work as a profession. A system that more closely matches need with care expertise across a geographic region could lead to changing shift patterns and demand increasing flexibility or mobility of carers. Such a framework might also enable care increasingly to be delivered remotely or virtually or via robots. It could also alter the sorts of qualifications needed to participate into care and entry into the profession, and how care professionals are remunerated. In the end, it could change or challenge broader social attitudes to care. These issues all raise questions as to who should be setting or shaping and monitoring the overall goals of the system, and the sorts of social and political participation needed to review the values underpinning those goals.

6 Conclusions

This chapter discussed a range of intended and possible empirical features of CAS associated with the vision of a Smart Society, and provided some conceptual elements and empirical illustrations for the ethical governance of such systems. The overall point to be made is that any attempt to construct a framework for ethical governance necessarily remains incomplete and contestable, hence our metaphor to sketch a path towards ethical governance rather than provide a full account. The next, more concrete points to be made relate to this general one, they are derived from our analysis and synthesis of existing forms of governance and CAS features.

First, many CAS are built on, and into, existing forms of governance, that is, different ways of steering society and maintaining social order, and more precisely, different approaches to inherent contradictions like autonomy versus control, centralised versus self-regulation, individual versus public interests. We have distinguished five such different governance regimes (and there may be more): polycentric, motivational, environmentally embedded, adaptive, and political. We have argued that in practice, such different governance modes actually interact with one another and influence each other, thus forming a composite phenomenon rather than competing "juxtaposed" alternatives. This composite phenomenon gets more complicated, and may have ever more emergent properties, as the scale of CAS increases (e.g. national road networks), and as human and technical system "components" change over time, often in numerous feedback-loops (and we strongly advocate such a time-sensitive, historical view of CAS for a more realistic understanding).

Second, CAS may be designed to be as diversity-aware as possible, but many real-life settings will include so many diverging stakeholder needs and interests that ethical problems may not be solved completely, for everybody, and once and for all.

We assume that such disharmony and residual conflict are more the rule than the exception. So we make the case that persisting ethical tensions should be perceived as a structural feature of CAS. The latter may be designed to deal with such tensions, but it may not be realistic to assume that any design can reconcile all possible, and emerging, ethical problems “in the wild”.

However, this precaution does not mean that the whole project needs to be abandoned—quite the contrary. We propose that, third, a basic understanding of ethical governance as being mindful in the production of new, or the reproduction of existing governance regimes; and as continuous processes of identifying and managing recurrent ethical concerns. With this double emphasis on facilitating processes of problem identification, and on a second-order awareness of existing governance regimes, we suggest a procedural understanding of “ethics” here, not (only) a substantial one. It remains to be discussed how such a procedural ethical governance is to be realized in practice.

Finally, given the abstractness of a procedural interpretation of ethical governance and its potential downsides, i.e. second-order ethical problems (e.g. who or what decides about the right procedures; who or what actually takes care of (re)producing governance regimes), more substantial, domain-specific values may need to be considered and designed into a CAS (with the possibility of redesign and recalibration of vision). For instance, a substantial value to be accounted for may be a qualitative notion of care that, in this chapter, has been spelled out in greater detail through a worked example of governance design for HDA-CAS in a care setting. The point to be made here is that there is no abstract or theoretical short cut to the development of such substantial domain specific values, nor are they universally and absolutely true. One has to work through the details of different empirical instances of a given domain to develop a careful, ready-to-be-revised preliminary understanding of important substantial ethical values. This seems also to require methodological innovation to build anticipation more strongly into design processes. We believe that this chapter provides a modest example of how this can be achieved, in this instance, by weaving together the ethics attached to contemporary socio-cultural trends, an understanding of governance design, and elaboration of an empirical case. Although there are still many problems to solve, not least achieving a wider representation of stakeholder engagement when envisioning possible futures, we believe we have the basis of a framework for the ethical governance of CAS that we intend to build upon in Smart Society.

Acknowledgements The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 321480 (GREAT) and no. 600854 (Smart Society). We would also like to thank the Engineering and Physical Sciences Research Council (EPSRC) for funding the Framework for Responsible Research and Innovation project Grant No. EP/J000019/1.

References

1. Aberbach, J.D., Rockman, B.A.: Conducting and coding elite interviews. Polit. Sci. Polit. **35**(4), 673–676 (2002)
2. Anderson, S., Bredeche, N., Eiben, A., Kampis, G., van Steen, M.: Adaptive collective systems: herding black sheep. In: BookSprints for ICT Research (2013)
3. Baker, L.: Removing roads and traffic lights speeds urban travel. Scientific American. Online. http://www.scientificamerican.com/article.cfm?id=removing-roads-and-traffic-lights (2009)
4. Benkler, Y.: Coase's penguin, or, linux and the nature of the firm. The Yale Law J. **112**, 369–446 (2003)
5. Brey, P.: Disclosive computer ethics: The exposure and evaluation of embedded normativity in computer technology. Comput. Soc. **30**(4), 10–16 (2000)
6. Bullock, S.: Prospects for large-scale financial system simulation. The future of computer trading in financial markets. Foresight driver review DR14 (2011). http://eprints.soton.ac.uk/272759/ (2011). Accessed 9 Aug 2013
7. Büscher, M., Coulton, P., Efstratiou, C., Gellersen, H., Hemment, D.: Connected,computed, collective: Smart mobilities. In: Grieco, M., Urry, J. (eds.) Mobilities: New Perspectives on Transport and Society, pp. 135–158. Burlington: Ashgate (2011)
8. Collective adaptive systems, expert consultation workshop report. ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/fet-proactive/shapefetip-wp2011-12-02_en.pdf (2009). Accessed 9 Aug 2013
9. Demographic change and an ageing population. report of the finance committee. http://www.scottish.parliament.uk/parliamentarybusiness/CurrentCommittees/59613.aspx (2013)
10. Description of work. Grant agreement for the smartsociety project, annex i. Internal project documentation (2013)
11. Erickson, T., Kellogg, A.: Social translucence: An approach to designing systems that support social processes. ACM Trans. Comput. Hum. Interact. **7**(1), 59–83 (2000)
12. Fleischmann, K., Wallace, W.: Ensuring transparency in computational modeling. Commun. ACM **52**(3), 131–134 (2009). http://dl.acm.org/citation.cfm?id=1467278
13. Forte, A., Larco, Y.V., Bruckman, A.: Decentralisation in wikipedia governance. J. Manag. Inf. Syst. **26**(1), 49–72 (2009)
14. Friedman, B., Kahn, P., Borning, A.: Value sensitive design: theory and methods. Technical report, UW CSE (2002). http://www.urbansim.org/pub/Research/ResearchPapers/vsd-theory-methods-tr.pdf
15. Fuerth, L.S.: Foresight and anticipatory governance. Foresight **11**(4), 4–32 (2009)
16. Giunchiglia, F.: Hybrid and diversity-aware collective adaptive systems in human computer confluence. In: Ferscha, A. (ed.) The Next Generation Humans and Computers Research Agenda, pp. 12–15 (2014). http://hcsquared.eu/hc2-visions-book
17. Guston, H.D.: The anticipatory governance of emerging technologies. J. Kor. Vac. Soc. **19**(6), 432–441 (2010)
18. Hannam, K., Sheller, M., Urry, J.: Editorial: Mobilities, immobilities and moorings. Mobilities **1**(1), 1–22 (2006)
19. Hollands, R.G.: Will the real smart city please stand up? City Anal. Urban Trends Cult. Theory Policy Action **12**(3), 303–320 (2008)
20. Introna, L.D., Nissembaum, H.: Shaping the web: Why the politics of search engines matters. Inf. Soc. **16**(3), 169–185 (2000)
21. Johnson, D.G., Mulvey, J.M.: Accountability and computer decision systems. Commun. ACM **38**(12), 58–64 (1995)
22. Kjolberg, K.: The notion of "responsible development" in new approaches to governance of nanosciences and nanotechnologies. Ph.D. thesis, University of Bergen (2010)
23. Knobel, C., Bowker, C.G.: Values in design. Commun. ACM **54**(7), 26–28 (2011)
24. Lanier, J.: Who Owns The Future? Allen Lane, London (2013)
25. Lyon, D.: Surveillance as social sorting. In: Lyon, D. (ed.) Privacy, Risk and Digital Discrimination. Routledge, London (2003)

26. Maiden, N., D'Souza, S., Jones, S., Müller, L., Pannese, L., Pitts, K., Prilla, M., Pudney, K., Rose, M., Turner, I., Zachos, K.: Computing technologies for reflective, creative care of people with dementia. Commun. ACM **56**(11), 60–67 (2013)
27. McNutt, K., Rayner, J.: Valuing metaphor: A constructivist account of reflexive governance in policy networks. In: 5th Conference on Interpretive Policy Analysis, pp. 23–25, Grenoble (2010)
28. MIT: senseable city laboratory. http://senseable.mit.edu/
29. Neuhauser, L., Rothschild, B., Graham, C., Ivey, S.L., Konishi, S.: Participatory design of mass health communication in three languages for seniors and people with disabilities on medicaid. Am. J. Public Health **99**(12), 2188–2195 (2009)
30. Options for strengthening responsible research and innovations: Report of the expert group on the state of art in europe on responsible research and innovation. http://ec.europa.eu/research/science-society/document_library/pdf_06/options-for-strengthening_en.pdf Accessed 9 Aug 2013
31. Osterloh, M., Frey, B.S., Frost, J.: Managing motivation, organization and governance. J. Manag. Govern. **5**(3), 231–239 (2001)
32. Ostrom, E.: Beyond markets and states: Polycentric governance of complex economic systems. Am. Econ. Rev. **100**(3), 641–672 (2010)
33. Owen, R., Macnaghten, P., Stilgoe, J.: Responsible research and innovation: From science in society to science for society, with society. Sci. Public Policy **39**, 751–760 (2012)
34. Pickering, A.: The mangle of practice: Agency and emergence in the sociology of science. Am. J. Sociol. **99**(3), 559–89 (1993)
35. Progress project. http://www.progressproject.eu/more-rri-resources/
36. Schatzki, T., Cetina, K., von Savigny, K.: The Practice Turn in Contemporary Theory. Routledge, New York (2001)
37. Seddon, J.: Systems Thinking in the Public Sector. Triarchy Press, Axminster (2008)
38. Sengers, P., Boehner, K., David, S., Kaye, J.J.: Reflective design. In: In Proc. Critical Computing, pp. 49–58. ACM Press, New York (2005)
39. Shah, R.C., Kesan, J.P.: How architecture regulates. J. Archit. Plann. Res. **24**(4), 350–359 (2007)
40. Silberman, M.S., Irani, L., Ross, J.: Ethics and tactics of professional crowd work. XRDS: Crossroads, The ACM Magazine for Students **17**(2), 39–43 (2010)
41. Slater, J.: Reducing crime through design, supplementary planning document. Tech. rep., Portsmouth City Council. http://www.portsmouth.gov.uk/media/Reducing_Crime_Through_Design_SPD.pdf (2006). Accessed on 24 Mar 2014
42. Stahl, B.C., Eden, G., Jirotka, M.: Responsible research and innovation in information and communication technology - identifying and engaging with the ethical implications of icts. In: Owen, R., Heintz, M., Bessant, J. (eds.) Responsible Innovation, pp. 199–218. Wiley, London (2013)
43. Willke, H.: Smart Governance: Governing the Global Knowledge Society. Campus, New York (2007)

Collective Intelligence and Algorithmic Governance of Socio-Technical Systems

Jeremy Pitt, Dídac Busquets, Aikaterini Bourazeri, and Patricio Petruzzi

1 Introduction

The methodology of sociologically-inspired computing [10] endeavours to support systems engineering by developing formal and algorithmic models of social processes. The general idea, on encountering an application problem, is to introspect on how people solve such problems, and use that as inspiration for a technical solution. We note, *en passant*, that the paradigm of biologically-inspired computing operates in much the same vein (e.g. [1]), taking instead natural (biological) systems as its source of inspiration.

The methodology, itself a generalisation of Steels' synthetic method [27], is illustrated in Fig. 1. The steps involved are: given a problem, identifying a theory from the social sciences of how people solve that (or an analogous) problem (*theory construction*); developing a formal model of that theory in an appropriate calculus (*formal characterisation*), where by calculus we mean any formal language enabling symbolic representation and manipulation; implementing that formal model (*principled operationalisation*); and then testing the implementation to determine if it provides a solution to the original problem (*controlled experimentation*). Implicitly or explicitly, the methodology has been applied to Dennet's Intentional Stance [7] to produce the BDI agent architecture [25]; cognitive, psychological or physiological models to provide decision-support systems based on trust [15], forgiveness [29] and emotions [18]; legal and organisational models to provide a framework for agent societies [2], and learning by imitation for human-robot interaction [6].

J. Pitt (✉) • D. Busquets • A. Bourazeri • P. Petruzzi
Department of Electrical & Electronic Engineering, Imperial College London,
London SW7 2BT, UK,
e-mail: j.pitt@imperial.ac.uk; d.busquets@imperial.ac.uk;
a.bourazeri11@imperial.ac.uk; p.petruzzi12@imperial.ac.uk

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_2

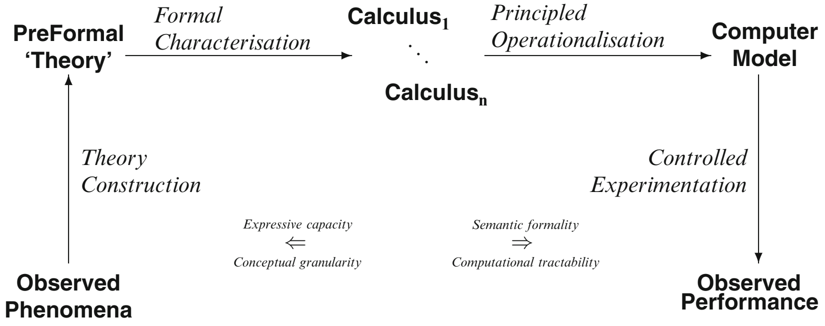

PreFormal 'Theory'
Formal Characterisation
Calculus$_1$
Calculus$_n$
Principled Operationalisation
Computer Model
Theory Construction
Controlled Experimentation
Expressive capacity ⇐ Conceptual granularity
Semantic formality ⇒ Computational tractability
Observed Phenomena
Observed Performance

Fig. 1 Methodology of sociologically-inspired computing [10]

In describing the methodology, Jones et al. identify a number of adequacy criteria for the transition, at each step, between the conceptual theory, formal representation(s), and the implemented model. This is because the final model is *not* a precise testable model of the original social system with predictive and explanatory capacity; and nor is it intended to be. It is designed only to provide an algorithmic solution to an application-specific problem, and in applying the methodology there might have been 'theory loss' (simplification of the theory or the formal representation because the concepts are too complex to formalise or are computationally intractable) and 'application gain' (enrichment of the formal representation or implementation due to domain-specific aspects of the application, not conceptualised by the theory).

On the other hand, it is an intriguing question: *what happens when the algorithmic solution to the engineering problem is offered to the people who have to solve the same problem, i.e., the one that inspired the solution?*

This is the question that is addressed in this chapter. In applying the methodology of sociologically-inspired computing to the idea of self-governing institutions for common-pool resource management, we have established an algorithmic basis for self-organising resource allocation in open computer systems and networks [20, 23] based on *computational justice* [22]. This chapter investigates what happens when this algorithmic basis of 'justice' is made manifest to users in socio-technical systems, and when the technical components have to represent and reason with qualitative values of primary concern to the users.

The issue is investigated from the theoretical concept of 'justice' and the formal representation of different aspects of 'justice' in computational form (Sect. 2), and from the application perspective of decentralised Community Energy Systems (Sect. 3). Then we consider the injection of the algorithmic basis for these concepts of justice being manifested into a socio-technical system for 'fair' demand-side self-organisation in a decentralised Community Energy System. Two systems are presented, in Sect. 4 a system based on collective awareness in a 'serious game', and in Sect. 5 based on representation and reasoning with an electronic form of social capital. We summarise and conclude in Sect. 6.

2 Computational Justice

2.1 Open Systems: Some Issues

Open decentralised computer systems and networks often require the system components to share resources (e.g. bandwidth, memory, energy) in order to achieve their individual goals through the coordinated actions of a group. In the absence of a centralised controller and given the autonomy of the components (i.e. hereafter called agents), let us suppose, in the first instance, there is a system specification defining a set of rules giving the resource allocation method to be used in computing the actual resource allocation.

In fact, the resource allocation problem itself is compounded by a number of other requirements and complicating factors. This includes:

Self-determination. In a system of completely autonomous agents, which may vary over time, and the wide range of possible resource allocation methods available and different outcomes they can produce, the resource allocation method should be determined by the agents themselves. In particular, each agent is entitled to assess the subjective 'quality' of the resource allocation by whatever criteria it considers appropriate, e.g. fairness, equity, utility, etc.

Uncertain resource variation. The system may vary from times in which there is an abundance of resources, to periods where it must operate in an *economy of scarcity* (cf. [26]) in which there are sufficient resources to keep the appropriators 'satisfied' in the long-term, but insufficient resources to meet everyone's demands at any a particular time-point, to times of crisis where the system faces complete failure.

Expectation of error. In the presence of competition from autonomous agents and conflicting goals, sub-ideal behaviour (everything from non-compliance to the specification to 'selfish' behaviour which diminishes the global collective welfare, such as free riding) is to be expected. However, errors may be a result of accident or necessity (e.g. as a consequence of resource variation), as well as malice: in such competitive or transient situations, there is an incentive to maximise individual utility by not contributing to the collective while still benefiting from the contributions of others, i.e. free riding.

Enforcement. Open systems might as well use random allocation and operate under the principle of *caveat emptor*, if agents are not monitored so can transgress at will, or can repudiate agreed rules and sanctions for non-compliance by refusing to abide by their outcomes.

Endogenous resources. In a system where all the resources are provided by the appropriators themselves, as in a sensor network or a micro-grid, all tasks such as determining the resource allocation method, computing the resource allocation itself, and monitoring the resource appropriation, must be 'paid for' from the very same resources. If so much resources are expended on these activities it might leave nothing for 'real' jobs (both [19] and [3] report how the costs of needless and/or excessive monitoring deplete resources in this way).

No full disclosure: the appropriators are autonomous and internal states cannot be checked for compliance (with conventional rules), so incoming agents do not have all the information required for necessarily reliable investment decisions (e.g. contributing to a common pool).

However, all these features are routinely encountered in social situations, and in fact, addressing each of these factors seems to involve some concept of 'justice'.

2.2 Computational Justice: The Programme

'Justice' is a concept that has been of concern in philosophy and jurisprudence (*inter alia*) since antiquity, and we do not intend to review this history or provide a formal definition. However, in the research programme of 'computational justice' we are, intuitively, trying to capture some notion of 'correctness' in the outcomes of algorithmic decision-making (specifically concerned with outcomes of resource allocation processes), thereby trying to accommodate some elements of fairness, utility, equity, proportionality and tractability in the process.

On this understanding, we observe that different 'qualifiers' of justice, that have been used in the social sciences, can be identified to address the key features of open self-organising systems previously specified:

- self-determination requires a concept of *natural* justice in dealing with a shared or common-pool resource (cf. [16]), specifically recognising both membership rights and the right of those affected by rules to participate in the selection of the rules, usually by voting;
- uncertain resource variation not only requires some self-determination in the selection of the rules congruent with the circumstances (abundance, scarcity and crisis), but some familiar fairness and efficiency criteria, like Pareto efficiency and envy/freeness, may be ineffective for all conditions, and a more flexible concept of *distributive* justice [26] is required, including a subjective agreement on fairness norms is required [9];
- expectation of error and enforcement of rules requires monitoring and assessing behaviour, and the enforcement of sanctions for identified non-compliant behaviour, requires a concept of *retributive* justice: this includes distinguishing between different types of error, ensuring that punishments are proportional to the extent of the 'wrong-doing', and offering the chance of redemption and allowing for appeals are essential aspect to consider;
- dealing with endogenous resources requires a concept of *procedural* justice: if the administration of the rules has to be 'paid for' from the same resources that are otherwise allocated for 'useful' jobs, then it is necessary to ensure that they are, in some sense, 'fit-for-purpose' [21]; and
- dealing with lack of full disclosure requires an element of interactional justice, namely *informational* justice, to force disclosure of relevant information.

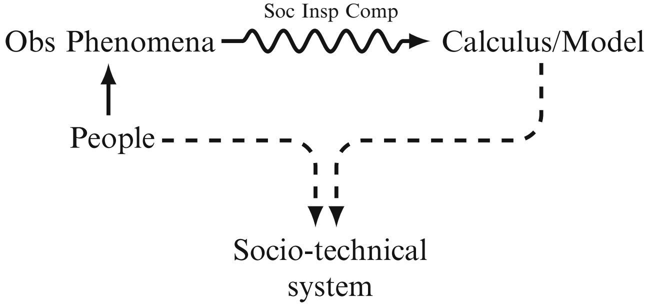


Fig. 2 Computational Justice: from technical systems to socio-technical systems

Therefore, our application of the sociologically-inspired computing methodology has focused on analysing theories of different aspects of justice, formalising them in a calculus—we have used the Event Calculus [11]—and then implementing them as computer models, either directly in Prolog or using the multi-agent system simulator and animator PreSage2 [12]. Amongst others, two significant results to highlight are:

- Showing that Elinor Ostrom's institutional design principles for enduring self-governing institutions [16], which essentially embody many principles of natural and retributive justice, can be axiomatised in computational logic and then used for specifying and implementing self-organising electronic institutions with corresponding properties of endurance and sustained membership [20];
- Showing that Nicholas Rescher's theory of distributive justice [26] based on the canon of legitimate claims can also be axiomatised in computational logic and as complement to Ostrom's principles, used to ensure fairness in resource distribution over time (according to a chosen fairness measure, the Gini index) [23].

The question we now address, see Fig. 2, is what happens when these systems of computational justice are made manifest to users in socio-technical systems. The specific socio-technical systems we use as an exemplar to explore this manifestation are decentralised Community Energy Systems, as described in the next section.

3 Decentralised Community Energy Systems

There are various aspects of power systems presenting situations which need to be solved by an aggregated body comprising a portfolio of smaller resources forming a kind of 'collective'. For example, the concept of *zoning* for self-managed network operation and control could be considered from this perspective as a partitioning/aggregation problem.

Similarly, for energy generation, the idea of the Virtual Power Plant has been studied and implemented [28], where many small(er) generation units are aggregated in an equivalent (virtual) big(ger) power plant. The advantages of these aggregated or collective power plants is threefold. Firstly, they can participate in the markets with higher quantities of energy or of related services, in order to have better prices. Secondly, there are markets where small quantities are not accepted in today's IT support platforms. In addition, some small un-synchronized efforts may not bring at all a visible effect in the network, so small contributors may not participate at all if they think that they are alone.

Usually, however, such aggregations are pre-arranged and usually are backed-up by legal contracts. When the focus is switched from the supply-side to the demand-side, it can be argued that there is a requirement for run-time self-organisation rather than pre-arrangement, and for social contracts rather than legal contracts.

Therefore, we propose that demand-side management of energy distribution and consumption can be addressed by applying a user-centric, self-organizing approach to the various partitioning, aggregation and provision/appropriation problems entailed. In the context of the UK EPSRC Grand Challenge 'Autonomic Power System' [14], we have been studying demand-side self-organisation in decentralised Community Energy Systems (dCES). In a 'traditional' community energy system, there is a central generator serving a set of consumers (e.g. households); in a decentralised community energy decision, both the generation and the decision-making is pushed to the edges, i.e. the households themselves.

An example of a decentralised 'community energy system' is the energy grid of Schönau, Germany [8]. The vision for this grid was a decentralised form of green-energy production, in terms of both increasing the efficiency of energy transmission and empowering citizens to take charge of their energy consumption and production. The idea was to turn energy consumers into prosumers (both producers and consumers), by motivating individuals to produce and save energy, and to sell the surplus back to the grid. This way of thinking initiated the process of equipping the inhabitants of Schönau with resources to produce energy and manage it through a citizen-owned social business, the Power Supplier of Schönau. Most households in this community produced energy by diverse means, and managed the process of its distribution.

In our conception of a decentralized Community Energy System, a group of geographically co-located residences is occupied by prosumers. The residence may have installed photovoltaic cells, small wind turbines or other renewable energy source; and the occupants have the usual requirements to operate their appliances. Note we also consider the issue of storage, and (looking forward) propose to consider the use of electric vehicles as a 'distributed battery' (see Fig. 3).

Therefore, in fact we have two concurrent and co-dependent provision and appropriation systems, one for generation and one for storage, and actions in one system have effects in the other. Furthermore, instead of each residence generating, storing and using its own energy, and each suffering the consequences of over- or under-production, the vagaries of variable supply and demand should be evened out by providing energy to a common-pool and computing a distribution of energy

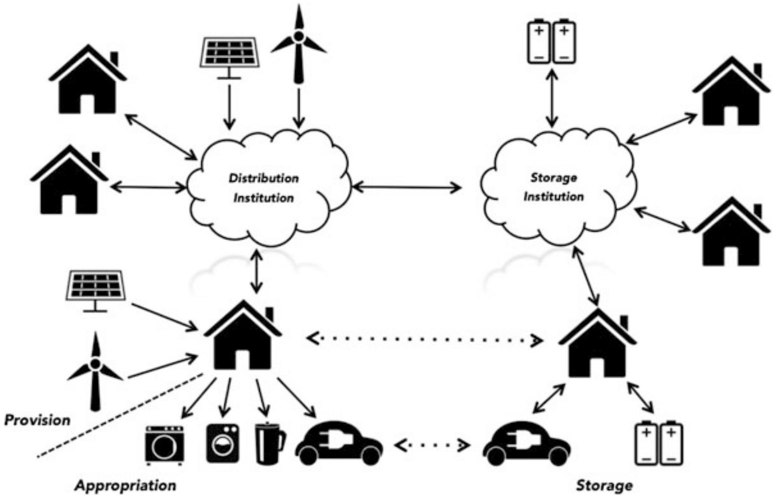


Fig. 3 Decentralised community energy system (dCES)

using algorithmic self-governance specified by institutions. These institutions would operate firstly, Ostrom's design principles for enduring and sustainable common-pool resource management, in which excessive demand, which would otherwise lead to a power outage, could be pre-empted by synchronized action based on collective awareness; and secondly, a social capital framework for successful collective action.

In the next two sections, we present progress in developing frameworks for what are effectively decision-support mechanisms for decentralised community energy systems. The first one is based on collective awareness within a Serious Game (Sect. 4), while the second one is based on social capital for concurrent and co-dependent provision and appropriation systems (Sect. 5). Both are critically dependent on interleaving social and computational intelligence and reasoning with respect to some notion of justice

4 Collective Awareness

Demand-side self-organisation of energy systems depends upon user engagement and active consumer participation. This self-organisation for common pool resource allocation should observe and address different principles, encapsulated by the user-infrastructure interface. This user interface extended to a 'serious game' should support the users in a decentralised community energy system and emphasise on collective awareness, securing at the same time the active participation of users who can be both individual consumers or group of prosumers.

The drive towards demand-side self-organisation of the electricity distribution and supply network is particularly motivated by Elinor Ostom's principles for enduring self-organising institutions. These principles characterise who is a member of the institution, how the resources are managed and allocated, who is affected by the rules of the institution and who can participate in their selection and finally, that no external interference is accepted. These principles are the foundation for user engagement and active consumer participation inside an energy system.

The key issue is how Ostrom's principles can be encapsulated and supported by a user-infrastructure interface, ensuring at the same time that users can actively participate in a decentralised energy system. Serious Games could be a plausible solution; digital games in which Ostrom's principles are supported by both the interface and the rules of the game. Adding ICT to the user-infrastructure interface enables the users to become active participants and make choices which ensure the endurance and fair distribution of the resources in the electricity network.

4.1 Visualisation of Ostrom's principles

Table 1 presents how Ostrom's principles and user participation can be encapsulated in a Serious Game for a Decentralised Community Energy System. Serious Games are digital games, simulations and virtual environments whose purpose is not only to entertain and have fun, but also to assist learning and help users to develop skills such as decision-making, long-term engagement and collaboration. They are experiential environments, where features such as though-provoking, informative or stimulating are as important as fun and entertainment [13]. They can also be used for modelling and simulating new and complex systems, empowering at the same time different groups and communities to exploit the most of the system's possibilities and characteristics.

Principle 1 states that there should be clearly defined boundaries in the institution. This is represented by the player's access to the game. The institution is visualised and represented by a virtual community, where the members of the community need a membership for getting access and having an avatar in the game. Principle 2 refers to the congruence between the rules for appropriation and provision of resources and the state of the local environment. This can be achieved through the collective awareness. Collective awareness among the members of a community enhances the sense of collective responsibility, whereas if it is missing, the members of the community cannot understand the present situation or occurred changes to their local environment. The third principle concerns collective-choice arrangement, stating that those affected by the operational rules should participate in the selection and modification of these rules. This can be represented by a participatory deliberative assembly where all the players can gather and make common choices and decisions concerning the electricity distribution. Principle 4 refers to monitoring behaviours and current state. Smart Meters are assigned this monitoring agency role.

Table 1 Ostrom's principles encapsulated by a serious game

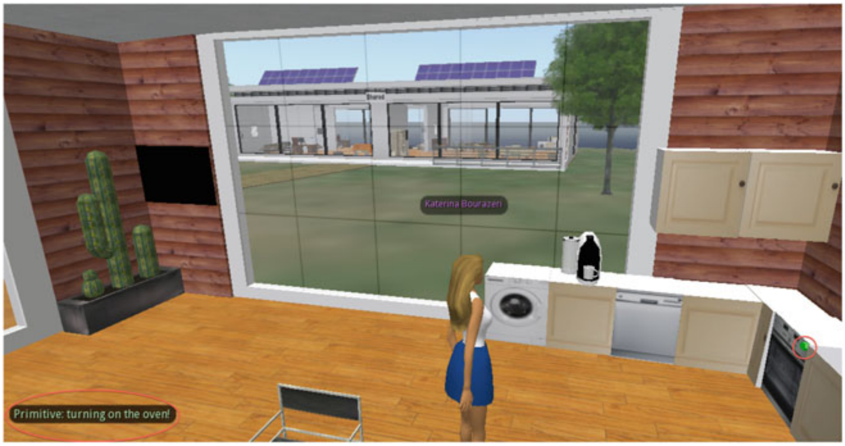
Ostrom's principles	Visualisation in serious games
(1) Clearly defined boundaries	Game access
(2) Congruence between rules and local environment	Collective awareness
(3) Collective choice arrangements	Participatory deliberative assembly
(4) Monitoring	Smart meters
(5) Graduated incentives	Sanctions and rewards
(6) Conflict resolution	Conflict resolution mechanisms

Principle 5 states that there should be graduated sanctions for those agents violating rules, as well as incentives for those complying. This is visualised through a rewarding/sanctioning scheme that it is introduced in the game. This scheme is endorsed to reward the successful game players, whereas it imposes penalties in case of inappropriate behaviours. Finally, Principle 6 is concerned with access to fast, cheap conflict-resolution mechanisms. The game provides different mechanisms such as jury, negotiation or mediation that are used to resolve occurred disputes. Ostrom defines two more principles: no interference from external authorities to ensure that the game cannot be controlled or monitored from the external environment (Principle 7) and systems-of-systems (Principle 8) to allow for nested institutions. However, these two last principles are not represented in the game [5].

4.2 *Visualisation of a Decentralised Community Energy System*

Collective awareness is an important component of a community, as it strengthens the sense of collective responsibility and enables the members of this community to adapt better and easier to their environment. A system based on collective awareness in a serious game can support the demand-side self-organisation of a decentralised Community Energy System. Collective awareness combined with gamification techniques observed in a virtual world, could promote the user engagement and active consumer participation. Gamification is basically the use of game design techniques and mechanics to non-game applications in order to teach, motivate and engage users in a different way.

Drawing attention to these two aspects could enable and support the users of the virtual world to feel part of an online game-based community, where sustainability and adaptability are promoted. The concepts of serious games and gamification can be extended and include social rules and norms, empowering in this way the users who are now enabled to control their avatars, take part in an everyday scenario and being incentivised and driven by social capital rather than money (see Sect. 5).



Primitive: turning on the oven!

Fig. 4 A 3D serious game virtual energy community

A 3D serious game virtual community can provide the necessary requirements for human inclusion and active participation in a decentralised energy system. Five different activities can be defined, enabled and supported through this online virtual community: *(i) Decentralised Energy System Representation*, the virtual community (three different houses one for each type of player—single, couple and shared—with electrical appliances connected to Smart Meters) where the user can control and observe in real time the energy consumption, *(ii) Private & Public Messages*, messages (energy feedback) concerning the energy consumption that users receive in real time and they can be provided both on an individual and common basis, *(iii) Assembly*, another house in the virtual community where all the players can gather and self-organise in a way so that the grid sustainability can be achieved, *(iv) Smart Meters*, an ICT-enabled device that allows both monitoring and reporting of electricity consumption and *(v) Rewards & Sanctions*, where the good players can be rewarded and get prizes, whereas the inappropriate behaviour is sanctioned and the bad players receive penalties (Fig. 4).

When the player gets access to the implemented virtual community, he has to select among three different profiles/houses—single, couple, shared—based on the profile that reflects his everyday life. This real-based choice will enable us to better understand how users are going to consume electricity when they will exit the virtual world and return to their everyday house routine. All the players have to self-organise and coordinate their actions in a way that their electricity consumption in each time slot does not exceed the maximum available energy capacity of the whole community. The different 'installed' blackboards in the houses and in the *Assembly* room display the individual and common energy consumption, whereas the residual capacity is known as well among the players. The demand-side self-organisation is based on collective and coordinated actions among the players and comparative feedback. Collective awareness is particularly important as it supports the collective action and the social networking.

This virtual community can provide decision-support mechanisms for enduring, self-organising institutions and in coordination with gamification mechanisms and techniques a better grid management and resource allocation could be achieved. Demand-side self-organisation based on a common-pool resource management for decentralised community energy systems comes as a user empowerment which highlights collective awareness and choices, whereas consumer behaviour is regulated and organised so as the use participation in the grid is increased.

Users will now have more discerning options and choices inside the virtual energy community system. They are entitled to organise and control their energy consumption and production and as a result they better comprehend concepts which concern grid sustainability, resource allocation and investment decisions. The consumers' inclusion and participation in an energy system require not only a better understanding of the energy consumers' behaviour, but also getting energy consumers to better understand the effects of their behaviour and actions on the electricity network.

4.3 Smart Meters and Systems of dCES

Smart Meters are an ICT-enabled device installed 'at the edge' of a decentralised community energy system, that allow both monitoring and reporting of the energy consumption and production. On top of these services, the two-way communication between the Smart Meters and the central electricity network is enabled and supported. Even though the Smart Meters are not just passive devices which display the energy consumption but they can also serve as agent-based assistants and non player characters, they are received as a "can't-opt-out" technology both centrally imposed and controlled. This obstructs generativity and raises concerns for trust, privacy and security. The end users do not own this technology, although it is their behaviour that is being monitored and controlled.

The introduction of Smart Meters in domestic residences as the basic interface for displaying information needs to be received as an innovative technology for enabling users and making their everyday lives easier [4]. As the energy users cannot spend all their time in front of a screen to monitor and control their energy consumption, intelligence needs to be added to the Smart Meters which will empower them to be adaptive to users' needs and preferences. The user-centric orientation of the Smart Meters will provide awareness to the consumers and visualisation of the different forms of information concerning the energy consumption and production. With this generative, opt-in and at-the-edge technology, the energy users will be able to program their electrical appliances in a more sustainable and efficient way for a community energy system.

Smart Meters being a fundamental element of a decentralised community energy system could provide the computational intelligence, a key aspect which is missing from those systems. The 'intelligence', such as it is now, is definitely not 'at the edge', nor it is operating on behalf of the end-user, i.e. the electricity consumer. Smart Meters should be perceived as assistive-enabled devices which promote and maximise the capabilities and choices of the consumers or prosumers. If the computational intelligence interleaves with the social intelligence coming from the collaboration among the different decentralised community energy systems, then issues such as resource allocation and distribution, investment decisions and energy system's sustainability could be better forwarded and advocated.

5 Social Capital

It has been noted that people's 'attention' is limited, so that users won't spend all their time monitoring their energy consumption. Instead, in the previous section we were relying on social networks and reporting of exceptions to provide the collective awareness to support synchronised, coordinated action. However, to manage the quotidian operation of the system, people need to know how to delegate to the Smart Meters, which in turn need to reason about qualitative values of concern to people. To do this, we propose to use social capital as a way of optimising demand-side self-organisation in provision and appropriation situation; moreover social capital also has significant potential when dealing with multiple concurrent and co-dependent provision and appropriation systems.

5.1 Self Organising Flexible Demand

In decentralised community energy systems, peak consumption times can force them to consume electricity from energy providers. When a community invest in photovoltaic cells, small wind turbines or other renewable energy source, consuming more energy from this source (instead from the energy provider) will be translated into lower electricity prices for them. One method of lowering the consumption peaks is flattening the demand. It implies reducing the difference between the peaks and troughs in electricity usage by creating a levelled usage pattern that lessens the deviation from the average usage.

We propose self organising flexible demand, where consumers can demand an amount of electricity for a certain period of time. Once it is allocated, they can exchange these allocations among them to better satisfy their time preferences. Since consumers might not be available to perform this actions, or not interested in, they can choose their time preferences and delegate the task of exchanging the allocations to their Smart Meters. Furthermore, by introducing the use of Social Capital, consumers cooperate and help each other in order to obtain the allocations they need.

5.2 Electricity Exchange Arena

To enable consumers to self organise we set up an exchange arena in which each day is divided in 24 time slots of 1 h. Consumers can demand amounts of electricity for each time slot based on their needs. Initially, a predefined allocation method performs the first allotment of the consumers' demands. Depending on the method chosen, consumers can receive allocations that are not in their preference; however, the amount of electricity assigned to them is always as much as demanded. Once all the allocations are received, consumers can start to exchange them.

In order to exchange an allocation, consumers can publish which of their allocations they are willing to exchange. All such allocations are publicly visible in a kind of classified advertising board. Consumers can check the ads board and send offers for those allocations they are interested in. The exchange is only between two consumers and they trade only allocations; there are no payments involved. Consumers will accept or deny an offer depending on their preferences or needs of electricity consumption.

5.3 Social Capital in Decentralised Community Energy Systems

Social capital is defined as "the features of social organization, such as networks, norms and trust, that facilitate coordination and cooperation for mutual benefit" [24] and furthermore as "an attribute of individuals and of their relationships that enhance their ability to solve collective-action problems" [17]

The creation of social capital among the consumers not only benefits them individually, but also as a whole. And, since consumers must perform exchanges to obtain the allocations they need, the more they all collaborate the higher the chances they will have to get what they want.

In this work, we implemented a simple form of social capital. At every exchange, consumers check if the received allocation is in their interest. If so, they count it as a "favour received" from the other consumer. In the opposite situation, they count it as a "favour done" to this other consumer. Since the favours calculation is internal for each consumer, an exchange where both consumers get an allocation they want is perceived as a favour received by both of them. These win-win situations help to create social capital among the system.

Our research is focused in developing a framework for representation of and reasoning with social capital. The self-organising processes that social capital facilitates generate outcomes that are visible, tangible, and measurable. The processes themselves are much harder to see, understand and measure.

In the next section we present the experiments done using favours as initial form of social capital.

5.4 Experiments

We have used PreSage2 [12] to develop a simulation of the electricity exchange arena and analyse the self-organisation of flexible demand. The arena was populated with 96 consumers who demanded 4 time slots with 1 kWh of electricity for each. Consumers chose randomly these 4 slots over the 24 available. Two allocation methods were tested; a Random Allocator and an Optimum Allocator. The first, assigns the demands randomly to the available slots. The second, performs the allocations maximising the average consumer satisfaction, which is defined as the proportion of electricity received in their preferred time slots. Both methods allocate up to the daily average for each time slot, i.e. 16 kWh for each slot.

Two type of consumers were added to the system:

- Selfish Consumers: They only accept to exchange if the offered allocation is in their interest, i.e. a time slot that is in its preferences, but was not received at the initial allocation.
- Social Consumers: They check at every exchange if the received allocation is in their interest, and keep the count of favours done and received. They will accept an offer if it benefits them, as the Selfish consumers, but also if they owe a favour to the consumer sending the offer. Through this behaviour, Social consumers will start acting selfishly, but after some exchanges they will start accepting offers in which they are not interested. These exchanges will not decrease their satisfaction, since they are not interested in any of both allocations (the sent and the received), but it will improve their Social Capital.

With this set up, consumers demand, get the allocations and perform the exchanges for a day. The simulations were run for 200 sequential days and the results were averaged over 100 runs.

Figure 5 shows a snapshot of the experiment graphical interface. Each circle represents a consumer and the colour, from red to green, their own satisfaction. The average consumer satisfactions is also showed as a bar on the right. Through the experiment, at each round, an arrow between two consumers will graphically show an exchange of consumption slots among them. When an exchange occur, at least one of the consumers will increase his satisfaction and his colour will change getting greener.

In Fig. 6 we compare the average consumer satisfaction at the end of the exchanges round for each day. The Optimum allocator achieved the highest consumer satisfaction average, and since there is no better allocation distribution, no exchanges were performed. Using this allocation method an average consumer satisfaction of 90 % was achieved. All the values have been normalised to this allocation method. The random method, without any exchange, achieved the lowest consumer satisfaction. By allowing exchanges, the Selfish Consumers considerably improve the results exchanging allocations between them, although their average satisfaction does not vary over time. With the inclusion of Social Capital, Social Consumers start

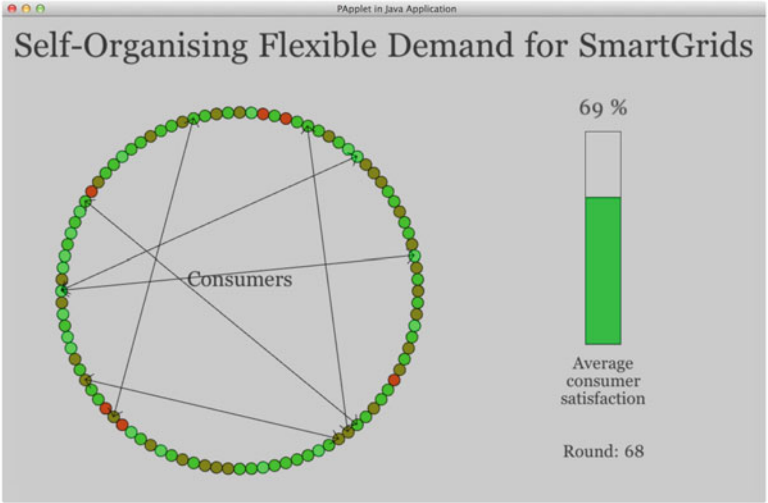

PApplet in Java Application
Self-Organising Flexible Demand for SmartGrids
Consumers
69 %
Average consumer satisfaction
Round: 68

Fig. 5 Snapshot of the experiment graphical interface

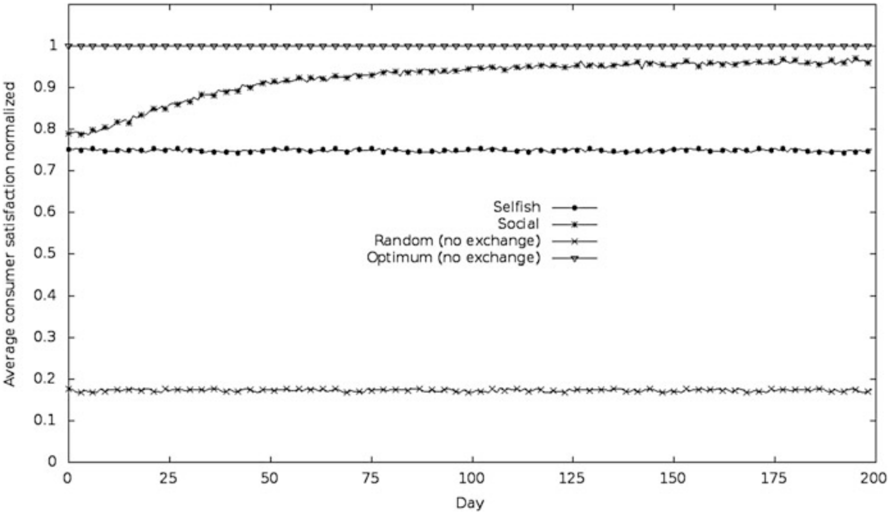


Fig. 6 Average consumer satisfaction normalised at the end of each day

performing as the Selfish Consumers, but their satisfaction increases as they perform exchanges with other consumers. They help other consumers to get the allocations they need as a return of the favours received.

Despite the fact that using a centralised allocation method shows better results, our approach slightly under-performs and frees the systems from the scalability issues. The Optimum Allocation method does not take into account the consumer

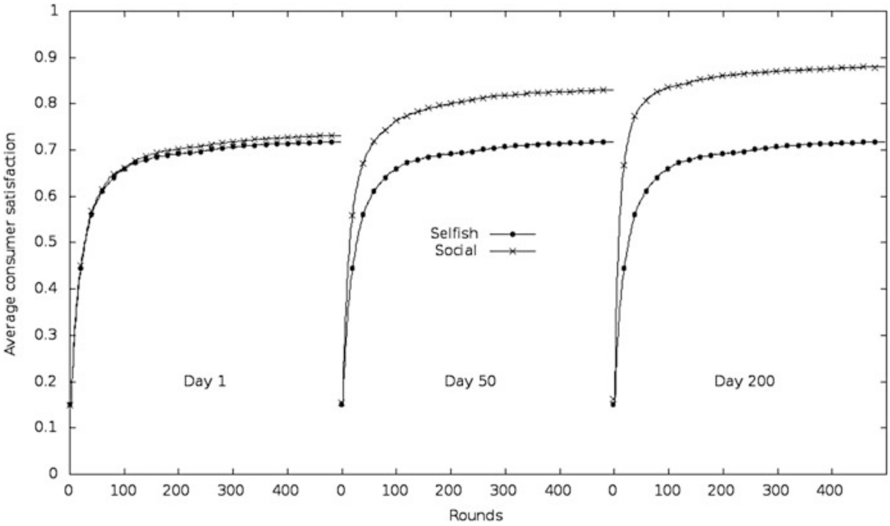


Fig. 7 Average consumer satisfaction during the first, fiftieth and two-hundredth day

flexibility, and including it will require a more complex algorithm. On the other hand, with exchanges, the more flexible a consumer is, the more Social Capital it will be able to generate. Eventually, consumers can also add more constraints or more flexibility to their demands without altering the operation of the whole system, which is not possible in a centralised allocation.

Figure 7 shows the average satisfaction during the exchange period for the first, the fiftieth and the two-hundred day for Selfish and Social Consumers. During the first day both perform equally since very few favours take place. After 50 days, Social Consumers have got a high satisfaction average, because they pay back favours received from previous days. At last, on day two-hundred, the consumers' satisfaction is higher because more exchanges occurred.

6 Summary and Conclusions

In this paper, we have considered decentralized Community Energy Systems, wherein the objective is to create a self-sustaining community of prosumers who provision and appropriate the generation, storage and distribution of energy amongst themselves, independent of a fixed grid infrastructure.

We have considered such systems from the perspective of common-pool resource management; in which case, questions about the 'robustness' of the community and the 'fairness' of the allocation can be addressed using formal theories of natural (or social) justice due to Elinor Ostrom. Furthermore, the co-dependence of concurrent provision and appropriation systems, whereby decisions and actions in one system are leveraged (as social capital) to support and sustain the other, and vice versa, can be addressed using the formal theory of distributive justice due to Nicholas Rescher.

In fact, formal representations of different qualifiers of justice have contributed to a research programme called Computational Justice, providing algorithms for self-regulation of open computer systems and communication networks. The question was then posed, what happens when these computational theories of justice are injected into, and made manifest to the users, in socio-technical systems, i.e. providing an algorithmic basis for self-governance.

Based on this, we discussed how the research programme of computational justice can inform the application Ostrom's theories to the self-organisation and visualization of 'fair' demand-side energy management. We described two approaches, firstly the use of collective awareness within a Serious Games, and secondly the formalisation of social capital mechanisms underlying successful collective action in concurrent, inter-dependent provision and appropriation systems. Two demonstrator systems for 'fair' demand-side self-organisation have been developed, and prospects for combining social and computational intelligence(s) in decentralised community energy systems have been presented.

In both systems, there remains much further work to do: for example, as we move from Serious Games to *gamification* (the use of game-like mechanisms to manage real-life situations), we need to find the correct balance between constant intervention and monitoring by the prosumers and the delegation of their attention to a SmartMeter operating on behalf of (and perhaps programmed by) the users themselves. Having delegated to SmartMeters, the simulation results have shown that a simple form of Social Capital, which creates win-win situations, improves the performance of the demand-side system. We will continue this line of work by developing a Framework for representing and reasoning based on Ostrom's [17] forms of Social Capital. However, we argue that it is *justice* (in its computational form) rather than trust which is the glue between different forms of social capital and successful collective action in socio-technical systems of the kind we have been discussing here. Further specific links between the two self-organizing socio-technical systems and the different qualifiers of justice is illustrated in Fig. 8.

Furthermore, we observe that a decentralised community energy systems can emerge in multiple scales of time and geography. We could have a dCES that could operate as a socio-technical system on a local geographical scale and operate on individual prosumer decision-making. Therefore, we could have a dCES that operates as a 'socio-technical system' composed of individual consumer, but was itself operating as an individual 'technical system' across national boundaries, enabling a community of 'twinned towns' to trade energy. Finally, there could be a dCES which uses concepts of trust, self-organisation and social capital to form a generating body (i.e. we return full circle to the Virtual Power Plant). In particular, we propose to undertake a comparative evaluation of optimisation based on market-based vs. (or with) institution-based approaches to community energy systems, from both business case and operational bases (e.g. computational cost, efficiency, fitness for purpose, compliance, social justice, etc.).

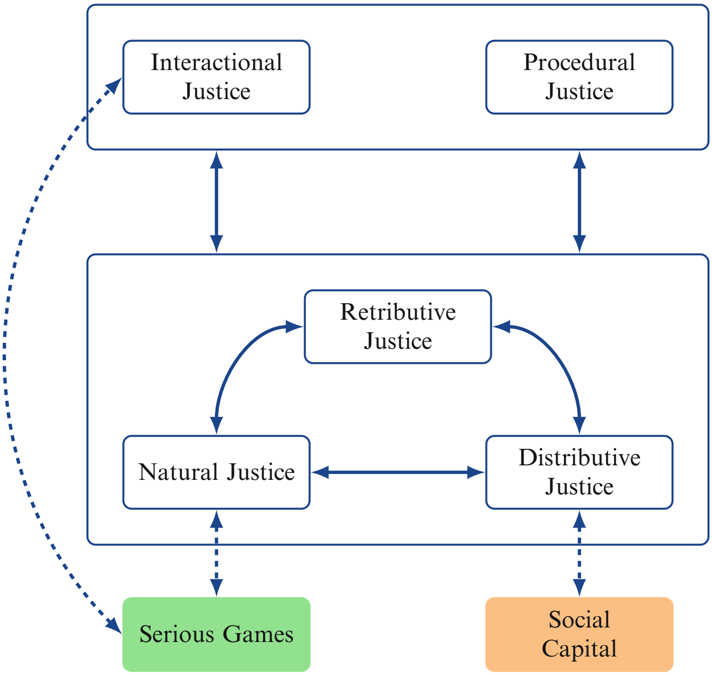


Fig. 8 Link of the presented self-organised systems and computational justice

In conclusion, this chapter has illustrated the potential for using computational justice in open socio-technical systems, such as decentralized Community Energy Systems, and how they can help deliver social justice to the prosumers so involved. However realising the full potential of computational justice in such domains is critically dependent on successfully inter-leaving social and computational intelligence across multiple scales: this is the critical challenge that lies ahead.

Acknowledgements This research has been partially supported by the UK EPSRC Grand Challenge project No. EP/I031650/1.

References

1. Andrews, P., Polack, F., Sampson, A., Stepney, S., Timmis, J.: The cosmos process, version 0.1: a process for the modelling and simulation of complex systems. Technical Report YCS-2010-453, University of York (2010)
2. Artikis, A.: Dynamic specification of open agent systems. J. Log. Comput. **22**(6), 1301–1334 (2012)
3. Balke, T., de Vos, M., Padget, J.: I-ABM: combining institutional frameworks and agent-based modelling for the design of enforcement policies. Artif. Intell. Law **21**(4), 371–398 (2013)

4. Bourazeri, A., Pitt, J.: Serious game design for inclusivity and empowerment in smartgrids. In: First International Workshop on Intelligent Digital Games for Empowerment and Inclusion (2013)
5. Bourazeri, A., Pitt, J., Almajano, P., Rodriguez, I., López-Sánchez, M.: Meet the meter: visualising smartgrids using self-organising electronic institutions and serious games. In: Proceedings of 2nd AWARE Workshop on Challenges for Achieving Self-Awareness in Autonomic Systems, SASO 2012, Lyon (2012)
6. Demiris, Y.: Prediction of intent in robotics and multi-agent systems. Cogn. Process. **8**(3), 151–158 (2007)
7. Dennett, D.: The Intentional Stance. MIT Press, Cambridge (1987)
8. Elektrizitätswerke Schönau: Introducing the elektrizitätswerke schönau (2012)
9. Elster, J.: Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens. Russell Sage Foundation, New York (1992)
10. Jones, A., Artikis, A., Pitt, J.: The design of intelligent socio-technical systems. Artif. Intell. Rev. **39**(1), 5–20 (2013)
11. Kowalski, R., Sergot, M.: A logic-based calculus of events. New Gener. Comput. **4**, 67–95 (1986)
12. Macbeth, S., Pitt, J., Busquets, D.: System modeling: principled operationalisation of social systems using presage2. In: Gianni, D., D'Ambrogio, A., Tolk, A. (eds.) Modeling and Simulation-Based Systems Engineering Handbook. Taylor and Francis, London (2014)
13. Marsh, T.: Serious games continuum: between games for purpose and experiential environments for purpose. Entertain. Comput. **2**(2), 61–68 (2011). doi: 10.1016/j.entcom.2010.12.004
14. McArthur, S.D.J., Taylor, P.C., Ault, G.W., King, J.E., Athanasiadis, D., Alimisis, V.D., Czaplewski, M.: The autonomic power system - network operation and control beyond smart grids. In: 3rd IEEE PES Innovative Smart Grid Technologies (ISGT) Europe, pp. 1–7 (2012)
15. Neville, B., Pitt, J.: A computational framework for social agents in agent mediated e-commerce. In: Falcone, R. (ed.) AAMAS Trust Workshop, pp. 83–91 (2004)
16. Ostrom, E.: Governing The Commons: The Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge (1990)
17. Ostrom, E., Ahn, T.: Foundations of Social Capital. An Elgar Reference Collection. Edward Elgar, Northampton (2003). http://books.google.co.uk/books?id=DZ_YAAAAIAAJ
18. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
19. Pitt, J., Schaumeier, J.: Provision and appropriation of common-pool resources without full disclosure. In: PRIMA, pp. 199–213 (2012)
20. Pitt, J., Schaumeier, J., Artikis, A.: Axiomatisation of socio-economic principles for self-organising institutions: concepts, experiments and challenges. ACM Trans. Auton. Adapt. Syst. **7**(4), 39:1–39:39 (2012). doi: 10.1145/2382570.2382575
21. Pitt, J., Busquets, D., Riveret, R.: Procedural justice and 'fitness for purpose' of self-organising electronic institutions. In: PRIMA, pp. 260–275 (2013)
22. Pitt, J., Busquets, D., Riveret, R.: The pursuit of computational justice in open systems. AI Soc. (2014). doi: 10.1007/s00146-013-0531-6
23. Pitt, J., Busquets, D., Macbeth, S.: Distributive justice for self-organised common-pool resource management. ACM Trans. Auton. Adapt. Syst. (under review) (to appear)
24. Putnam, R.D.: The prosperous community: social capital and public life. Am. Prospect **13**, 35–42 (1993)
25. Rao, A., Georgeff, M.: BDI agents: from theory to practice. In: Proceedings First International Conference on Multi-Agents Systems (ICMAS) (1995)
26. Rescher, N.: Distributive Justice. Bobbs-Merrill, Indianapolis (1966)
27. Steels, L., Brooks, R.: The Artificial Life Route to Artificial Intelligence: Building Situated Embodied Agents. Lawrence Erlbaum Ass, New Haven (1994)

28. Steghöfer, J.P., Anders, G., Siefert, F., Reif, W.: A system of systems approach to the evolutionary transformation of power management systems. In: Proceedings of INFORMATIK 2013 – Workshop on "Smart Grids". Lecture Notes in Informatics, vol. P-220, pp. 1–16. Bonner Köllen Verlag, Bonn (2013)
29. Vasalou, A., Hopfensitz, A., Pitt, J.: In praise of forgiveness: ways for repairing trust breakdowns in one-off online interactions. Int. J. Hum.-Comput. Stud. **66**(6), 466–480 (2008)

A Taxonomic Framework for Social Machines

Paul Smart, Elena Simperl, and Nigel Shadbolt

1 Introduction

Within the context of the World Wide Web, we have witnessed the emergence of a rich range of technologies that support both collaboration and distributed processing. Applications such as Wikipedia, for instance, have demonstrated the power and potential of the Web to facilitate the pooling of geographically dispersed knowledge assets. The result has been the creation of the world's largest online encyclopedia, available for free in more than 200 languages for everyone to access and use. Similar success stories can be found in various other domains. Projects such as Galaxy Zoo,[1] for example, have shown how collective intelligence can be used to inform scientific inquiry, while initiatives such as Ushahidi[2] have played a crucial role in emergency management situations worldwide. These three Web-based systems are representative of a general trend which is characterized by the use of Web-based technologies to enable a wide range of activities that rely on a combination of decentralized human activity and computational processing. This trend has been reflected in research efforts across a variety of areas, including social computing [32], human computation [34], crowdsourcing [11], and collective intelligence [5]. It has also given rise to a variety of new concepts, such as the 'social computer' [38], the 'social operating system' [35] and 'social machines' [18]. This chapter focuses on the last term in this list: the concept of social machines. The term 'social machine' was first used in a Web context by Berners-Lee and

[1] http://www.galaxyzoo.org/.

[2] http://www.ushahidi.com/.

P. Smart (✉) • E. Simperl • N. Shadbolt
Electronics & Computer Science, University of Southampton, Southampton, SO17 1BJ, UK
e-mail: ps02v@ecs.soton.ac.uk; e.simperl@soton.ac.uk; nrs@ecs.soton.ac.uk

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_3

Fischetti [3], and it has since grown in popularity to the point where it is now the focus of large-scale research programs, such as the EPSRC's SOCIAM initiative,[3] the subject of a multitude of academic papers (e.g., [18, 26, 27, 30, 41, 43, 50]) and the basis for a workshop series at the World Wide Web conference.

In spite of the growing interest in social machines, however, there is little consensus, at the present time, as to what the term 'social machine' actually means. In addition, the scientific community seems to have only a very narrow understanding as to what kinds of social machines actually exist. In order to make progress in these areas, we attempt to provide a working definition of the social machine concept that builds on the ideas put forward by Berners-Lee and Fischetti [3]. We also introduce a taxonomic framework for social machines that features a set of dimensions along which all social machines are deemed to vary. This work extends the results of an earlier study, reported by Shadbolt et al. [43], which used knowledge elicitation techniques to generate an initial set of dimensions. The work reported in the current chapter differs from this earlier body of work in two ways. Firstly, the dimensions from the earlier study have been refined and enriched following discussions with members of both the computer science and social science communities. Secondly, the current framework features a complete set of characteristics for each dimension. These characteristics specify the 'values' that each social machine takes with respect to each of the dimensions in the framework (see Sect. 4).

Together, the effort to provide a definition for social machines and the effort to develop a taxonomic framework mark an important step in terms of our attempt to understand the emerging, interdisciplinary research field of social machines. The effort to provide a definition of social machines is crucial because in the absence of an ability to say what social machines are it becomes difficult to know where to focus research and engineering efforts. The lack of a definition also complicates the effort to distinguish social machines from ostensibly similar systems, such as social computing, human computation, crowdsourcing and collective intelligence systems. The development of a taxonomic framework also marks an important step in our attempt to understand how social machines emerge and develop. Most importantly, the taxonomic framework establishes the dimensions according to which different instances of social machines can be compared to derive commonalities, as well as design and behavior patterns. This enables us to identify specific categories of social machines (taxa) that serve as the basis for classification efforts. It also enables us to analyze the overall space of design possibilities and identify areas that have been under-explored by research and development efforts. Finally, a taxonomic framework establishes the basis for future scientific efforts of both an analytic and synthetic nature: analytic efforts are driven by a need to understand why some parts of the design space are more populated than others, and synthetic efforts are driven by the need to explore parts of the design space that may afford opportunities for the creation of entirely novel kinds of social machines.

[3] See http://sociam.org/.

2 Social Machines: A Working Definition

Although there are a variety of views in the literature as to what actually constitutes a social machine, perhaps the most popular characterization is provided by Berners-Lee and Fischetti [3] in their book 'Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web':

> Real life is and must be full of all kinds of social constraint—the very processes from which society arises. Computers can help if we use them to create abstract social machines on the Web: *processes in which the people do the creative work and the machine does the administration.* [our emphasis] (p. 172)

This characterization emphasizes the joint involvement of people and technology with respect to particular processes, and it also makes a distinction between the respective roles that people and machines play with regard to the process being undertaken; in particular, the contributions of the human participants should consist in some form of creative work, while the contributions of the machine components should consist in some form of administrative activity. Assuming that the notion of 'creative work' should be interpreted in terms of the generation of online content (e.g., uploading a photo or writing some text), then it seems that Berners-Lee and Fischetti's understanding of social machines can be applied to many different kinds of Web-based systems. These include, for example, Wikipedia, Twitter, Facebook, YouTube,[4] and Flickr.[5] As is evidenced by the Web traffic system, Alexa,[6] the sites that host these systems are among the most popular on the Web today.

Although Berners-Lee and Fischetti's characterization can be used to support the identification of at least some social machines, it is far from clear that it applies to *all* social machines. One problem is that it is sometimes hard to discern what counts as a form of administrative and creative activity. Wikipedia bots,[7] for example, engage in automated processes that are essential to the ways in which Wikipedia content is managed. In some instances, they use advanced machine learning techniques to perform tasks that not so long ago were exclusively tackled through manual work and human insight, for instance, to detect and remedy deliberate attempts to vandalize Wikipedia articles [33]. Such bot-related (i.e., machine-based) activities could be easily classified as 'creative'. In other cases, we encounter human contributions that could be cast as somewhat administrative in nature. For example, the process of adding tags to Flickr photos plays an important role in terms of organizing the contents of the site, thereby making it easier for certain kinds of content to be accessed by the user community.

[4] http://www.youtube.com/.

[5] http://www.flickr.com/.

[6] http://www.alexa.com/.

[7] For an overview, see http://en.wikipedia.org/wiki/Wikipedia:Bots/Status, retrieved in December 2013.

Another problem with Berners-Lee and Fischetti's characterization is that it seems to overlook cases where the machine elements play an important role in the generation of online content or in enabling activities that are essential to it. PicBreeder,[8] for example, is a system that supports the collaborative and interactive production of images using a mixture of evolutionary computation techniques and human agents [39, 40]. The role of the human agents here is to select the machine-generated images based on (e.g.) aesthetic criteria. These images are then published on the PicBreeder site and are accessible to other users who can use them as the starting point for their own interactive image generation activities. PicBreeder is thus a system in which the machine components arguably play an important role in terms of what appears online (it is, after all, the machine components that are generating the images). If we were to embrace the notion of social machines as systems in which it is the humans that are solely responsible for the creative work, then PicBreeder would seem to be a poor candidate as a social machine. And yet PicBreeder does seem to have many of the features that make it the legitimate target of attention for the social machine community: there is community engagement, issues of human–machine collaboration, the socially-distributed nature of particular tasks, and so on.

In view of these problems, we suspect that a definition that seeks to impose constraints on the precise nature of the contributions made by human and machine components with respect to the performance of a task (administrative, creative, or otherwise) is likely to be overly restrictive in terms of the identification of important and interesting social machine exemplars. More importantly, if we carry such notions forward into the design and development of social machines, we risk delivering systems in which the virtues of human–machine interaction with regard to creative (and administrative) processes are ignored. When it comes to creative processes, for example, we should recognize that some of our best creative accomplishments often come about as a result of our ability to engage and interact with our technological artifacts, and we should seek to exploit this in the context of our design and engineering efforts. A perspective that seeks to limit the kinds of roles that can be performed by human and machine elements, and which additionally seeks to impose a strict (and rather artificial) boundary on where particular processes need to take place, risks blinding us to many of the opportunities that the Web provides in terms of the transformation of traditional processes and the enhancement of both human and machine capabilities.

In response to the problems associated with Berners-Lee and Fischetti's characterization, Smart and Shadbolt [45] have proposed the following working definition of a social machine:

> Social machines are Web-based socio-technical systems in which the human and technological elements play the role of participant machinery with respect to the mechanistic realization of system-level processes.

[8] http://picbreeder.org/.

This definition relaxes the constraint associated with the nature of the functional contribution made by human and machine elements, although it preserves the emphasis on social machines as bio-technologically hybrid systems (i.e., as systems that feature the incorporation of both people and machines). In particular, humans and machines are deemed to be *jointly* involved in the physical realization of processes: they are deemed to constitute part of the social machinery by which such processes are physically realized. This notion of human and machine elements serving as forms of participant machinery [8, p. 207] takes its inspiration from an approach to mind and cognition that sees extra-organismic resources as (on occasion) participating in the material realization of human mental states and processes—such resources are deemed to "form part of the very machinery by means of which mind and cognition are physically realized and hence form part of the local supervenience base for various mental states and processes" [8, p. 207]. A social machine is thus similar to what has been dubbed a 'Web-extended mind' [44] in the context of the Web Science literature.[9] Essentially, we suggest that a social machine is an extended functional organization in which the explanation of certain system-level processes requires an account that adverts to the details of mechanisms that are distributed across both the biological (human) and the technological (conventional computing systems) realms.[10] Such forms of 'explanatory spread' (see [58]) are sufficient for us to approach a social machine as a functionally-integrated system in spite of the heterogeneous nature of its material constitution. One of the crucial differences between the notion of a Web-extended mind and the notion of a social machine concerns the social aspect of the latter: the fact that it is multiple individuals (rather than a single individual) that participate in the realization of processes associated with the larger systemic organization. In addition, the kinds of processes enabled by the two scenarios are not co-extensive: Web-extended minds are concerned with cognitive processes; social machines, in contrast, are more general, referring to processes that may or may not be cognitive in nature.

Based on the above definition, a number of features of social machine systems are worth highlighting. One of these features concerns the fact that social machines are socio-technical systems—that is they involve the participation of human individuals and technological components. In many cases, we can expect the respective contributions of human and machine elements to draw on their distinctive capabilities and to complement one another with respect to the process that is being realized. It is the nature of this complementarity that underlies the interest in social machines as systems capable of a variety of advanced problem-solving capabilities

[9] The notion of a Web-extended mind draws its inspiration from work that goes under a variety of headings, such as 'extended cognition', 'cognitive extension' or 'the extended mind' [8, 9, 28]. Smart [44] defines a Web-extended mind as a system in which some of the informational and technological elements of the Web can be seen to constitute part of the material supervenience base for (at least some of) a human individual's mental states and processes.

[10] The use of the term 'mechanistic realization' in the definition is intended to highlight the importance of this mechanistically-oriented explanatory account [59].

(see [18, 22]).[11] By virtue of their ability to factor in human and machine contributions, social machines are often able to extend the reach of both human and machine intelligence, supporting capabilities that less integrated systems might find difficult to accomplish. In the taxonomic framework introduced in Sect. 4 we will elaborate on this symbiosis with respect to the ways in which this integration is achieved in terms of task assignment mechanisms and the roles that each type of component plays in the overall system.

A second point that is worth emphasizing is that, for our purposes, social machines are cast as Web-based systems. Although we do not rule out the possibility of social machines that are independent of the Web,[12] we suggest that the properties of the Web make Web-based social machines a particularly important focus of social and scientific attention. One virtue of the Web, in this respect, is that it enables us to tap into the capabilities of human agents in a manner (and on a scale) that has never been seen before. The Web is a social technology that interfaces with a large proportion of humanity. By firmly embedding itself within a human social environment, the Web opens up a range of opportunities to incorporate human agents into episodes of machine-based processing. This makes Web-based social machines capable of supporting processes that would be difficult or impossible to realize in other kinds of social (or indeed socio-technical) context.

Thirdly, social machines are systems that consist of multiple (human) individuals. This aspect is crucial for understanding the capabilities of social machines and designing successful systems. By drawing on a large number of individuals, social

[11] Similarly, it is the complementary nature of biological and non-biological resources (in terms of their contrasting representational and computational capabilities) that is often seen as lying at the root of the advanced forms of intelligence exhibited by extended cognitive systems. Sutton [49], for example, writes that "in extended cognitive systems, external states and processes need not mimic or replicate the formats, dynamics, or functions of inner states and processes. Rather, different components of the overall (enduring or temporary) system can play quite different roles and have different properties while coupling in collective and complementary contributions to flexible thinking and acting" (p. 194).

[12] Clocks may provide one example of a social machine that is independent of the Web. In their book, 'Anti-Oedipus', Gilles and Guattari [16] suggest that clocks are a form of 'social machine': "The same machine can be both technical and social, but only when viewed from different perspectives: for example, the clock as a technical machine for measuring uniform time, and as a social machine for reproducing canonic hours and for assuring order in the city" (p. 155). Interestingly, clocks have been seen as providing the technological impetus for the transformation of society. A number of theorists have emphasized the way in which clocks enable the large-scale scheduling and coordination of both individual and collective action, and the way in which the transition from fixed, centralized clock towers to portable wristwatches paved the way for new forms of social interaction and engagement [23]. The invention of portable time-keeping devices, argues Landes [23], made it possible to organize and synchronize activities in a way that had never been possible before, and on the back of this new capability there emerged a new social and economic era. The clock, in this case, can be seen as the technological element of a social machine in the sense that it is influencing social interaction via the delivery of machine-generated temporal representations. These representations serve to structure, sculpt and scaffold forms of social interaction and engagement that progressively shape the contours of the social, economic and cultural landscapes in which we live.

machines are able to accomplish tasks that require significant amounts of effort, for example, the decentralized analysis of large and complex bodies of scientific data (in Sect. 4 we will discuss the types of workflows that support this analysis at scale). In addition, social machines are able to exploit differences between individuals with respect to abilities, skills, insights, perspectives, knowledge, geographical location, experiences, group membership, social position, and so on. This may serve to improve the diversity and quality of the contributions that are made by the human community. Finally, social machines are also able to exploit the performance improvements that are often associated with collective inputs, for example, those associated with the Wisdom of Crowds phenomenon [48].

Fourthly, it follows from the above definition that processes are central to our understanding of what makes something a social machine: we discern a social machine when we encounter a process that demands a (mechanistically-oriented) explanatory account formulated in terms of the joint contributions of multiple individuals and Web-based technological components. It is important to note that we are not saying that social machines *are* processes, as would seem to be implied by the definition of social machines offered by Berners-Lee and Fischetti [3]. Rather, we are saying that social machines are the physical systems that perform, implement or realize such processes. This is an important distinction because the original definition (proposed by Berners-Lee and Fischetti) can result in a certain amount of confusion and conceptual indiscipline when it comes to discussions about social machines. Tinati and Carr [50] thus write that “any task that requires the co-constitutional involvement of humans and technologies is a form of social machine”. This characterization places appropriate emphasis on the importance of socio-technical engagement in the context of particular tasks, but it is a mistake to progress from this to the conclusion that the task itself is a form of social machine. Such conclusions, in our view, reflect a category error concerning the ontological status of social machines.

The centrality of processes to our understanding of social machines throws up a range of interesting issues and questions, some of which are out of the scope of the current chapter. One issue concerns the temporal nature of processes and the implication this has for the lifetime of a social machine. Processes may clearly be of relatively short-lived duration or they may be somewhat more enduring. Inasmuch as social machines exist for the duration of the processes with which they are associated, it would seem likely that social machines have a fair amount of variability with respect to their longevity. It should be possible to encounter social machines that persist for relatively long periods of time (as in the case of temporally-sustained, ongoing processes), as well as social machines whose existence is somewhat more fleeting and evanescent (as in the case of a social machine that supports social coordination in respect of a specific event—the organization of a birthday party, let’s say). Temporality plays a crucial role for several other properties of social machines captured by our taxonomic framework. For instance, the types of contributions made by human participants may change depending on their role in the system; also, the range of activities that are performed automatically might be expanded by the availability of new algorithms (as was the case with the Wikipedia bots discussed earlier). Such temporal variability has implications for efforts that seek to observe and monitor social machines, such as the efforts associated with the Web Observatory initiative [10,51]. In particular, if we assume that persistent social machines are both easier to monitor and also generate the most data (on account of their temporally-enduring nature), then it becomes clear that we face the potential hazard of a sampling bias as part of our monitoring efforts. Equally important is how changes along one or several dimensions of our taxonomic framework (see Sect. 4) affect the frequency of the monitoring exercise and our ability to manage and derive insight from observational data. If our future scientific understanding of social machines is grounded on a limited subset of social machine exemplars (i.e., the long-lived ones), then it is unclear whether our understanding will ever be complete: the properties and dynamics of an entire class of perhaps socially- and cognitively-crucial systems will go unrecorded.

A second issue thrown up by the process-oriented nature of social machines concerns the nature and visibility of the goal that is being realized by the process. In some cases, the goal of the process that is being realized by the social machine will not be visible to the human participants in the system. In other cases, the goal may be visible to one or more of the human participants, perhaps because they are the ones responsible for assembling the social and technological elements into a functionally-integrated information processing ensemble. Importantly, it seems possible to discern some cases where a social machine may be created or emerge from a technological system that was originally designed or configured to perform a different function. A social machine that emerges in the context of a large-scale social networking service, for example, may be concerned with the modification of people's voting behavior (see [31]) or product consumption patterns and actually have very little to do with the formation and maintenance of social bonds. A second class of examples which exemplifies the varying degree of goal transparency/awareness can be found in the area of human computation. An important category of social machines are thus systems referred to as 'games with a purpose' (or GWAP) [55]. In such systems, human agents participate in a game, often interacting with each other, sharing scores and competing against friends from their social network. The inputs collected from the players, in this case, are used to improve the accuracy of computing algorithms; players are not necessarily aware of the actual goal of the game as it was conceived by the game designer, but their social interactions and game play result in useful training data that assists with the development of automated processes. This example also makes clear the ambivalent nature of such goals; one could distinguish among (sometimes overlapping) component-level and system-level goals, each equipped in some cases with a temporal element.

Finally, it is worth noting that social machines, in virtue of involving multiple individuals, are often concerned with processes that are relevant to the social interactions and relationships between individuals. Many of the processes in which human and machine elements participate may thus be glossed as 'social processes': they concern the structure and dynamics of a group of people. Such processes may be many and varied. They include (but are not necessarily limited to) the coordination of collective action (e.g., implementations based on the Ushahidi platform); the pooling and distribution of resources (e.g., YouTube); the influencing of individual thoughts and actions (e.g., Twitter); the formation, maintenance, and dissolution of social relationships (e.g., Facebook); the collaborative creation of socially-shared assets (e.g., Wikipedia); and the social distribution of problem-solving processes (e.g., Galaxy Zoo). In general, the role of the machine or technological elements with respect to these processes is to constrain, control, coordinate or otherwise influence the social interactions between people (e.g., LinkedIn),[13] or, alternatively, to govern the way in which individual human contributions are collectively factored into some other process (e.g., reCAPTCHA[14]). Typically, the influence exerted by the technological elements, in these cases, is mediated by some form of manipulation and processing of the informational inputs that are provided by human agents (this distinguishes social machines from systems which merely act as conduits for the communication of information between individuals). Alternatively, it is possible that the influence may be exerted through the provision of machine-generated representations; for example, system-generated cues play a role in governing the dynamics of person perception processes in the context of systems such as Facebook [52] and Twitter [57].

3 Examples of Social Machines

A broad range of Web-based systems have been considered as candidate social machines within the Web Science community. These include Facebook [18], mySpace [18],[15] Twitter [17], YouTube [43], Ushahidi [41], Galaxy Zoo [17, 41], reCAPTCHA [30], Reddit [43],[16] Wikipedia [17, 18, 41], Amazon's Mechanical Turk [43],[17] and the Web itself [17].[18] As should be clear from this list, social machines are a pretty heterogeneous bunch of systems. For one thing, they seem to occupy a variety of functionally-diverse niches within the ecology of the Social Web. Extant social machines thus include social networking systems (Facebook, mySpace, Twitter), microblogging services (Twitter), video/photo sharing systems (YouTube), citizen science projects (Galaxy Zoo), social news sites (Reddit), collaborative content editing sites (Wikipedia), frameworks for the creation of collaborative systems (Ushahidi) and systems that enable human contributions to

[13] https://www.linkedin.com/.

[14] http://www.google.com/recaptcha.

[15] https://myspace.com/.

[16] http://www.reddit.com/.

[17] https://www.mturk.com/mturk/.

[18] The Web site of the SOCIAM research project lists a large number of additional examples of social machines—see http://www.sociam.org/social-machines.

be productively exploited in the context of automated processes (reCAPTCHA) or more traditional production processes (Mechanical Turk). This diversity has implications for the kind of features that we rely on to discriminate between social machines (see Sect. 4), and it also has implications for the types of social machines that we are able to recognize. The aforementioned list of social machine exemplars also highlights a number of areas of confusion when it comes to an understanding of the social machine concept. Armed with the working definition from Sect. 2, we are now in a position to address these areas of confusion (see also Fig. 1).

The first thing to note is that it is very common for people to refer to specific technologies when they talk about social machines. In many cases, therefore, when people identify a given social machine instance they point to a platform such as Facebook, Twitter or Ushahidi. Figure 1 refers to these as 'frameworks' and 'services', thus emphasizing the key role these socially-active environments play in the development and emergence of a wide range of special-purpose social machines targeting less general audiences. It is important to be clear that when we talk about social machines we are talking about a *socio-technical* system (as opposed to a purely technological system) that is actively engaged in the realization of a particular process [43]. Thus, when we say that Facebook is a social machine, what we mean is that it is the social networking platform (that we typically identify as Facebook) plus the human participants (the social environment) that constitutes the social machine. Any reference to a social machine as being constituted *solely* by the technological system (or subsystem) is, in our view, incorrect. It is for this reason that it is probably a mistake to refer to the human components of a social machine as the 'users' or as forming part of the 'user base' of the social machine.[19] Such terms imply that the social machine is something separate from the human participants: it conjures up an image of social machines as things that are independent of the human communities with which they are associated, and it encourages us to place undue emphasis on the technological aspects of the system. As should be clear from the definition presented in Sect. 2, social machines should be properly conceptualized as socio-technologically integrated systems in which the human 'users' are an intrinsic part of the larger, biotechnologically-hybrid system. This does not, of course, undermine the importance of the technological aspects as a source of scientific interest and a focus of engineering attention. Even in cases where all forms of human participation are absent, we can still recognize a technological system as something that is apt to participate in the formation of a social machine (or a multiplicity of such machines), and treat it as a legitimate target of scientific enquiry. The fact that an aircraft carrier is not, by itself, a socio-technical system does not mean that such vessels are not of considerable interest to naval engineers, even in situations where it is clearly obvious that the processes that the vessel is designed to support could only be realized once the human crew is onboard and certain forms of socio-technical entanglement are established.

[19] We are grateful to Ségolène Tarte (University of Oxford) for pointing this out.

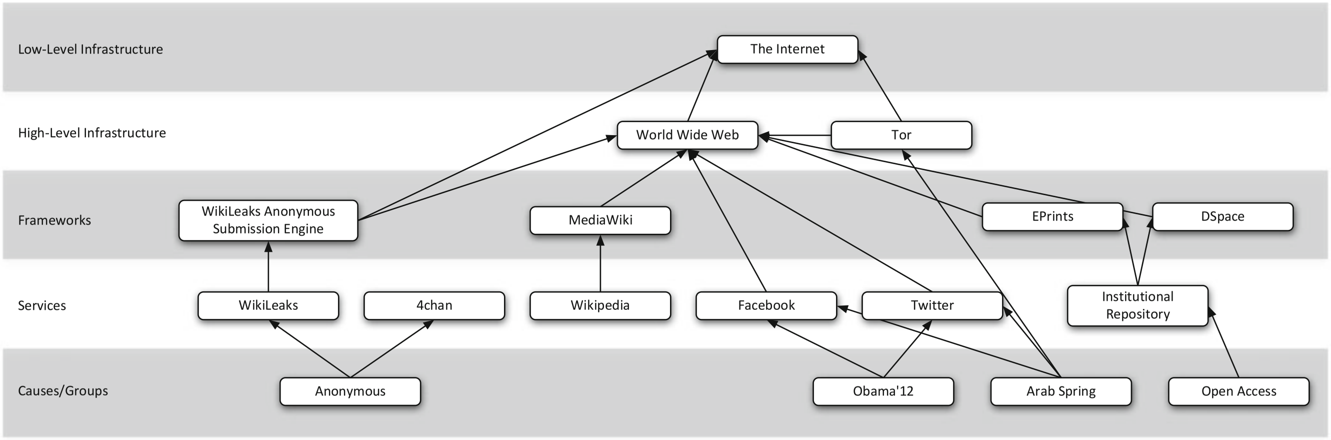


Fig. 1 Distinction between technology and social machines as part of a broader ecosystem [43]

A second issue arising from the aforementioned list of social machine exemplars concerns the distinction between systems that serve mainly as frameworks for the creation of social machines and systems that actually function as social machines (see the framework tier of Fig. 1). Ushahidi, for example, has been used to develop a number of systems that support information collection, visualization and interactive mapping, as in the Ushahidi-based system that supported humanitarian relief efforts in the aftermath of the 2010 Haitian earthquake [60]. An analogy here is with the MediaWiki system, which has supported a wide range of wiki-based collaborative content creation projects (e.g., Wikipedia, Wiktionary, Wikidata and Wikispecies). Wikipedia and Ushahidi are not, therefore, instances of the same class of objects, as might be implied by the above list of social machines. Instead, Ushahidi and MediaWiki should be seen as frameworks that support the creation of specific systems, such as Wikipedia and the Haitian implementation of Ushahidi. Another example of a framework that can be used to support the creation of social machines is Diaspora.[20] It can be used to create social networking services for specific communities of users. Although such frameworks should be distinguished from actual instances of social machines, they are clearly relevant to the project of characterizing and understanding social machines. For one thing, such systems serve as the template for a range of social machines that may possess similar or identical characteristics, and their design greatly influences the way in which a social machine functions and evolves.

A third point of interest concerning the aforementioned list of social machines concerns the way in which some social machines emerge in the context of other systems, which themselves may or may not be regarded as social machines.[21] O'Hara [30], for instance, talks about the use of Facebook to organize a birthday party. In this case, it is the specific use of Facebook to accomplish a particular task (i.e., organize a birthday party) that counts as a social machine rather than (perhaps) a more liberal perspective that sees Facebook itself as a social machine: Facebook is, in this case, merely serving as a form of technological scaffolding that supports the creation of a multiplicity of (probably) short-lived social machines. A similar claim could be made with respect to the relationship between the Web and social machines. Thus, although the Web has been regarded as a social machine [17], perhaps it is more appropriate to see the Web as the technological matrix that gives rise to a variety of social machine systems and in which all such systems are ultimately embedded. Contrary to this interpretation, however, we might argue that nothing in the definition of a social machine—either the original characterization [3] or the more recent definition [45]—would seem to rule out the possibility of

[20] https://diasporafoundation.org/.

[21] This corresponds to the tier termed 'Causes/Groups' in Fig. 1, which builds on a selection of Web-based systems that, through their large-scale user bases and general character, have reached a level of popularity that turns them into frameworks for the development of more special purpose social machines.

either the Web or Facebook counting as social machines. In addition, the possibility of a social machine emerging from the material matrix associated with some other system does not rule out the possibility that the other system is in fact a social machine: it may just be that the material elements of one social machine (i.e., its human and technological components) are simply recruited to form a social machine that is involved in a different process.[22] We suggest that we tend to discern a social machine when we can identify a socio-technical system that is involved in the realization of processes associated with the performance of a particular task. With this in mind, we might feel inclined to see a distinct social machine (one that draws on the technological fabric of Facebook, let's say) whenever we see particular tasks being performed (e.g., organizing a birthday party). However, in many cases, the larger system is also involved in the performance of particular tasks. Thus, in the case of Facebook, we might say that the system is (broadly) engaged in the realization of (the more temporally-protracted) process of social relationship management (i.e., the creation, maintenance, and dissolution of social networks). Inasmuch as we see this process as one in which the technological elements of the Facebook system are playing an explanatorily significant role, then we see no problem with a perspective that views Facebook as part of a functionally-integrated system (i.e., a social machine). Obviously, this does not rule out the possibility that the material elements associated with this system could be involved in a multiplicity of other, perhaps more short-lived, processes.

From an engineering point of view, the realization of such ecosystems depends on technologies, services and generic platforms that not only provide specific functionality—depending on the kind of social process supported by the social machine, this could be anything from communication and coordination of joint efforts to collaborative content generation, knowledge sharing, and decision making—but also promote principles, values, and ideas that match the expectations and motivation of the human participants. In particular, due to the very nature of a social machine and its ecosystem, it is essential that the technologies used to realize it are equipped with the means to tackle scale, decentralization, and concurrent access and processing. As content is created and shared in a distributed fashion, the social machine must be able to establish and associate trust or at least accountability in the ways every component of the system, biological or technological, operates and interacts with the rest of the ecosystem. We will follow up on these aspects in Sect. 4 where we discuss the social machine taxonomic framework.

[22] It is also possible to imagine one or more social machines being 'incorporated' into a larger social machine. In the same way, perhaps, as the neurological subsystems associated with memory, attention and perception merge to form part of the integrated mechanistic substrate that realizes more 'macrocognitive' functions such as sensemaking (see [21]).

4 Characterization of Social Machines

As part of the attempt to understand social machines, it is useful to develop a taxonomic framework that can be used to describe and classify social machine instances. Following Nickerson et al. [29], we define a taxonomy T as a set of n dimensions D_i $(i = 1, \ldots, n)$ each consisting of k_i $(k_i \geq 2)$ mutually exclusive and collectively exhaustive characteristics C_{ij} $(j = 1, \ldots, k_i)$ such that each object (i.e., social machine) under consideration has one and only one C_{ij} for each D_i. We have adopted this definition for our own taxonomic framework. Our approach to taxonomy development is also based on the approach advocated by Nickerson et al. [29], which has its roots in the social sciences (see [2]). The approach consists of three stages (see Fig. 2):

1. **Empirical-to-Deductive Stage:** This stage involves the initial examination of a subset of objects (social machine instances in our case) and the identification of their distinguishing features. As will be clear from the subsequent discussion, we rely on specific techniques in order to support this process. The output of this stage is an initial set of dimensions.
2. **Deductive-to-Empirical Stage:** This stage entails the conceptualization of new characteristics and dimensions. The dimensions elicited in the empirical-to-deductive stage are progressively refined and enriched during this stage.
3. **Taxonomy Application Stage:** This stage involves the use of the taxonomy to identify and characterize new objects. The taxonomy may also be used to inform the design of new objects.

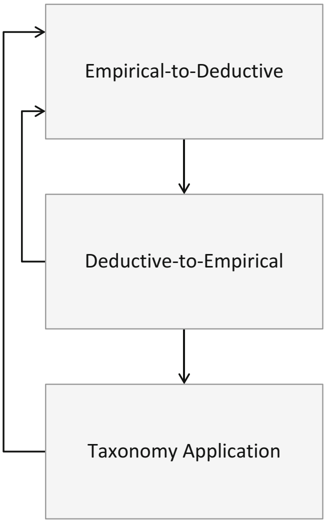


Fig. 2 The three stage approach to taxonomy development that was adopted in the current study (see Nickerson et al. [29] for more details)

As part of the empirical-to-deductive stage, we constructed an initial social machine taxonomy that included a set of dimensions but lacked any characteristics. This work is summarized below and reported in more detail in Shadbolt et al. [43]. In the current chapter, we focus on the deductive-to-empirical stage and present a more complete taxonomy featuring a revised set of dimensions and a complete set of characteristics. Although this taxonomy is subject to further refinement, it is suitable for use within the final taxonomy application stage of the taxonomy development process outlined above. In particular, we have compared the results of the taxonomy development effort with similar efforts that have been made in related areas (see Sect. 5). As part of our future work, we will test the completeness, correctness, and comprehensibility [56] of the taxonomy in experiments in which a new set of social machines will be classified by framework users. We will ask the participants to assess the quality of the framework along these general dimensions, and measure inter-annotator agreement to learn about the usefulness of the classifications produced.

We now turn to a description of the current version of the taxonomy. We will first present the approach taken to elicit information about social machine dimensions from human subjects and then give a summary of the empirical and conceptual work undertaken to define the taxonomy dimensions and their associated characteristics.

4.1 Eliciting Social Machine Dimensions

As illustrated by the examples surveyed in Sect. 3, social machines come in a variety of shapes and sizes. A system such as Facebook, which is concerned with (among other things) the formation and maintenance of social relationships, is clearly different from a system such as Mechanical Turk, which offers a crowdsourced labor market for simple data collection and processing tasks, and both of these are different from Zooniverse, which supports a form of socially-distributed problem solving in the natural and life sciences. Based on the working definition introduced in Sect. 2, we can anticipate a number of ways in which social machines built around these kinds of technological systems might differ. These include differences in the nature of the processes being realized; the kinds of contributions or computations made by human and machine components; the relative balance of processing effort among these components; and the ways that individual contributions are combined in the course of process execution. These, however, are just a few examples of the dimensions that could be used to differentiate between social machines. Other dimensions might be less obvious based on a cursory analysis of a limited subset of what are arguably the most well-known social machines that currently exist. Furthermore, even when larger samples of social machines are surveyed, the task of eliciting a more-or-less complete set of dimensions is not straightforward. People may find it difficult to discern differences between social machines, or find it difficult to communicate their conceptual understanding of these systems in a structured and coherent way, even when they are using these systems on a regular basis.

One way of dealing with the difficulty of eliciting dimensions is to rely on a range of techniques known as knowledge elicitation techniques [42]. These techniques are used as part of knowledge engineering initiatives in order to create the conditions under which domain experts are best able to communicate the knowledge associated with their expertise. Although a broad range of techniques are available, the ones that tend to be most suited for the elicitation of information about the dimensions along which a set of common objects vary are sorting and rating techniques. These include the repertory grid technique, which has its roots in the psychology of personality [13,19,20]. The repertory grid technique is useful because it can be used to support the elicitation of knowledge that is largely tacit in nature, i.e., difficult for an individual to verbalize. In addition, the data that is obtained as part of the technique can be subjected to various forms of statistical analysis in order to obtain insight into the structural organization of domain-relevant concepts.

In a repertory grid exercise subjects are presented with a range of objects, referred to as 'elements', and asked to choose three, such that two are similar and different from the third. This is known as the method of triadic elicitation (e.g., [6]). In order to demonstrate the technique, imagine a subject is presented with the following set of social machines[23]: Facebook, Twitter, YouTube, Wikipedia, reCAPTCHA, Galazy Zoo, Flickr, mySpace, LinkedIn,[24] and Planet Hunters.[25] The participant might choose Twitter and Facebook as the two similar elements and Galaxy Zoo as different from the other two. The subject is then asked to provide the reason for differentiating these elements, and this dimension is known as a 'construct'. Each construct is assigned a name as are the two poles that represent the opposite ends of the construct. In our example, 'size of the user community' might be a suitable construct that differentiates between the selected elements, with 'small' and 'large' serving as the two poles of the construct. Once a construct has been elicited, all the elements can be rated with respect to the construct, with the ratings reflecting the extent to which the subject sees an element as falling to one or other of the construct poles. The process of triadic elicitation is continued with different triads of elements until the subject can think of no further discriminating constructs. At this point we have a matrix of similarity ratings that can be analyzed using techniques such as cluster analysis. This provides us with a dendogram that can reveal conceptually-significant categories of social machines, and it can also shed light on the relationships that exist between the constructs.

In order to test the applicability of the repertory grid technique to the social machine taxonomy development effort, we first undertook a knowledge elicitation session with a computer science researcher from our laboratory. We presented them with the ten social machines listed above and engaged them in a process of triadic elicitation, in each case asking them to select and discriminate between

[23] In fact, as we mentioned in Sect. 3, the technological subsystem is only considered to be one part of the social machine; the human participants are also deemed to be part of the social machine.

[24] https://www.linkedin.com/.

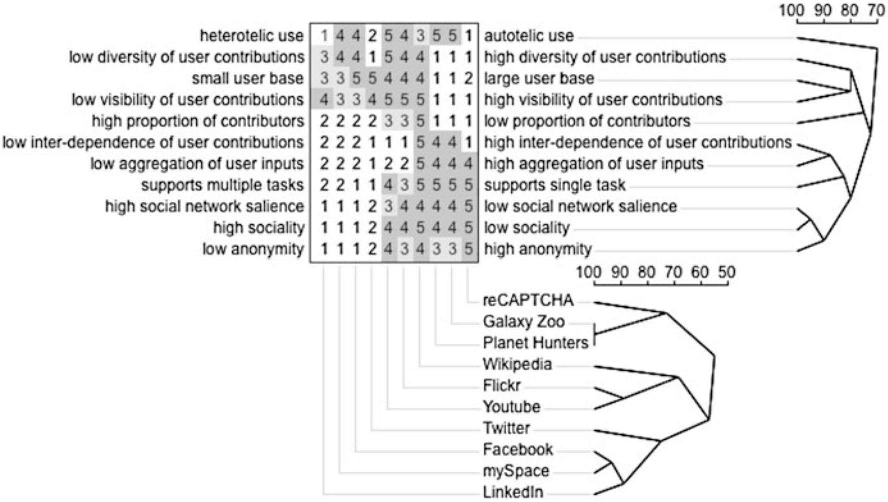
[25] http://www.planethunters.org/.

elements. The following set of constructs were elicited as a result of the application of this technique (elements that are representative of the poles of each construct are presented in square brackets):

1. **Size of the User Community:** the number of users that participate in the system, either as contributors or consumers [small: Galaxy Zoo; large: Facebook].
2. **Extent of User Contribution:** the proportion of users that actually contribute content as opposed to users that merely consume content [small: Wikipedia; large: Galaxy Zoo].
3. **Sociality:** the extent to which the system supports social interaction [low: reCAPTCHA; large: Facebook].
4. **Visibility of Individual User Contributions:** the extent to which individual user contributions are visible to the entire user base of the system [low: reCAPTCHA; high: YouTube].
5. **Inter-dependence of User Contributions:** the extent to which the user contributions are independent of one another with respect to the task being performed by the system [low: Twitter; high: Wikipedia].
6. **Focused on a Single Task:** the extent to which the social machine is focused on a single task as opposed to supporting multiple kinds of tasks [single task: Galaxy Zoo; multiple tasks: Facebook].
7. **Anonymity of Human Users:** the extent to which the system requires users to provide personal information about themselves to other users [low anonymity: Facebook; high anonymity: reCAPTCHA].
8. **Heterotelic vs. Autotelic Usage:** the extent to which the use of the system is motivated by instrumental or professional (heterotelic) concerns as opposed to enjoyment and pleasure (autotelic) [heterotelic: LinkedIn; autotelic: YouTube].
9. **Requires the Aggregation of User Contributions:** the extent to which user contributions need to be aggregated in order for the social machine to perform its primary function [low: Twitter; high: Wikipedia].
10. **Diversity of User Contributions:** the degree of differentiation with respect to user contributions. For example, users may be engaged in a single task (e.g., galaxy classification) or multiple tasks (e.g., uploading, rating, and tagging content) [low: reCAPTCHA; high: YouTube].
11. **Salience of Social Network:** the relative salience or visibility of the social network within the system [low: reCAPTCHA; high: Facebook].

The rating matrix and results of the cluster analysis are presented in Fig. 3 (these results were obtained using the WebGrid 5 system).[26] The ten social machines that were the focus of the repertory grid (e.g., reCAPTCHA, Galaxy Zoo, Planet Hunters, etc.) are listed at the base of the rating matrix, and the labels used to describe the poles of each construct are listed on either side of the matrix (e.g., 'heterotelic use' vs. 'autotelic use'). The numbers that make up the body of the matrix are the ratings made by the user for each of the social machines, with lower

[26] See http://gigi.cpsc.ucalgary.ca/.



heterotelic use	1	4	4	2	5	4	3	5	5	1	autotelic use
low diversity of user contributions	3	4	4	1	5	4	4	1	1	1	high diversity of user contributions
small user base	3	3	5	5	4	4	4	1	1	2	large user base
low visibility of user contributions	4	3	3	4	5	5	5	1	1	1	high visibility of user contributions
high proportion of contributors	2	2	2	2	3	3	5	1	1	1	low proportion of contributors
low inter-dependence of user contributions	2	2	2	1	1	1	5	4	4	1	high inter-dependence of user contributions
low aggregation of user inputs	2	2	2	1	2	2	5	4	4	4	high aggregation of user inputs
supports multiple tasks	2	2	1	1	4	3	5	5	5	5	supports single task
high social network salience	1	1	1	2	3	4	4	4	4	5	low social network salience
high sociality	1	1	1	2	4	4	5	4	4	5	low sociality
low anonymity	1	1	1	2	4	3	4	3	3	5	high anonymity

Fig. 3 The results of the repertory grid technique applied to the domain of social machines [see text for a description of the rating matrix and dendograms for both the constructs (*top dendogram*) and the social machine elements (*bottom dendogram*)]

ratings reflecting a bias towards the pole on the left hand side of the matrix. A score of '1' in the case of the 'Heterotelic vs. Autotelic Usage' construct thus indicates that the social machine is used for heterotelic purposes, whereas a score of '5' indicates that the social machine is used for autotelic purposes. With respect to Fig. 3, we can therefore see that LinkedIn and reCAPTCHA are both examples of social machines that are used predominantly for heterotelic purposes (they both have a low rating), whereas Galaxy Zoo and YouTube are both used predominantly for autotelic purposes (they both have a high rating).

The first thing to note from the dendogram associated with the social machines is that the Planet Hunters and Galaxy Zoo systems emerge as identical systems in this analysis—there are no constructs that differentiate between these two elements. This presumably stems from the fact that both systems form part of the Zoouniverse[27] collection of citizen science projects and both are concerned with the analysis of astronomical data. This result could be used to elicit additional differentiating constructs in situations where the subject did in fact believe there to be differences between the two systems. Another feature to note from the dendogram is that Facebook, mySpace and LinkedIn all seem to form a distinct cluster. We can ask our subject to attempt to say something about this clustering, perhaps supplying a label to identify a class or category of systems. The response, in this case, could be that the systems are all examples of 'social networking systems'. Another category of social machines seems to emerge based on the similarity of YouTube and Flickr.

[27] https://www.zooniverse.org/.

In this case, we might say that these systems are both examples of 'media sharing systems'.

In addition to the dendogram associated with social machines, Fig. 3 also shows the dendogram associated with the constructs. Here we can detect a number of correlations between the similarity scores, and these may reflect interesting contingencies between the features of social machines. For example, systems that exhibit low sociality also tend to be systems in which the social network has low salience. In addition, such systems are also ones that feature high levels of anonymity with respect to user contributions. As one might expect, systems that aim to support social interaction tend to require the disclosure of personal information—such disclosures are, in fact, likely to be a prerequisite for the development of relational intimacy. Another correlation emerges between the inter-dependence of user contributions and the tendency to aggregate user inputs. Again, not surprisingly, systems that feature high levels of interdependence between tasks also tend to be systems that engage in some form of aggregation of the user inputs. As part of our future work, we plan to collect a much larger collection of classifications in order to support the quantitative analysis of these sorts of correlations.

As should be clear from this example, the repertory grid technique can serve as an effective means of eliciting information about the features of social machines. It can also provide insight into the structure of the conceptual landscape associated with social machine systems. In particular, as more and more objects are surveyed, one can use cluster analysis to reveal interesting groupings that may serve as the basis for hierarchically-organized conceptual categories (i.e., taxa within the taxonomic framework). The results of the analysis can also serve as the basis for more focused knowledge elicitation sessions. For example, with respect to the above analysis, we could attempt to differentiate between the Planet Hunter and Galaxy Zoo systems, or we could exploit the ability to identify conceptual categories as a means of identifying additional social machines (e.g., systems that are members of the categories 'social networking system' and 'media sharing system').

4.2 A Social Machine Taxonomy

The analysis of the repertory grid described in the previous section provides some insight into the dimensions associated with social machines.[28] However, in order to expand the range of constructs elicited, it is necessary to draw on the perspectives of multiple individuals with respect to different subsets of social machines. For this reason, we completed an extended study involving ten computer science researchers

[28] The constructs identified in the context of the repertory grid exercise ultimately drive the generation of dimensions associated with the taxonomic framework. A construct such as 'Heterotelic vs. Autotelic Usage' (see Sect. 4.1), for example, is ultimately used as the basis for the 'Motivation Type' and 'Form of Motivation' dimensions listed in Table 2.

from our laboratory [43]. The motivation for using computer science researchers, in this case, relates to the requirements of the repertory grid technique. In particular, the repertory grid technique requires subjects to be familiar enough with the elements being investigated in order for them to make meaningful comparisons and identify distinguishing features. Given that the computer science researchers in our laboratory are currently involved in the analysis of a broad array of social machines, it made sense to draw on their experience in the context of this particular exercise.

After each subject had completed the repertory grid analysis with their self-selected elements, the result was a set of ten repertory grids containing a combined total of 117 different constructs and 56 unique social machine instances. This marked the completion of the empirical-to-deductive phase of taxonomy development. We subsequently reviewed these constructs to identify closely related ones, grouped the resulting list into broader categories, and refined the taxonomy based on insights gained from a review of the relevant Social Web literature.

The results of this second, deductive-to-empirical stage of taxonomy development are presented in Tables 2, 3, 4, 5, and 6 (see appendix). We identified a total of 33 dimensions, which were organized into five categories. The categories relate to the tasks that are being performed by the social machine (or the processes being realized by the social machine), the (human–human, human–machine, and machine–machine) interaction mechanisms by which the social machine operates, the ways in which quality and performance are assessed, the motivational factors and incentive mechanisms that govern user participation in the system, and the technologies used to implement the technical grounding of the system. Across the 33 dimensions, we identified a total of 106 distinct characteristics.

4.3 The Social Machine Morphospace

The dimensions revealed by our analysis constitute the set of dimensions along which all social machines (extant or otherwise) can be deemed to vary. These dimensions can be used to define the axes of a multi-dimensional design space for social machines. This design space constitutes the universe within which all theoretically possible social machines are located, with the location of each social machine dictated by the particular combination of characteristics it possesses. Given the similarity of this design space to the notion of a 'morphospace' in the biological literature [36, 53], we refer to the design space (or universe of social machine possibilities) as the 'social machine morphospace'. As with its biological counterpart, the social machine morphospace aims to chart the space of social machine possibilities with respect to a set of common features (dimensions) along which all social machines vary.

One advantage of the taxonomic framework is that it allows us to assess how much of the design space for social machines has been explored by current development efforts (obviously, given the size of the morphospace, it is likely that this space will be sparsely populated). Regions within the space that are devoid of social machines may represent unexplored regions that provide fertile ground for the creation of novel systems. Alternatively, it may be that such regions are barren for a good reason: perhaps the design candidates that occupy this region are impractical or impossible to implement. In summary, the value of the social machine morphospace is that it provides a view as to the total space of design possibilities for social machines, and it indicates the regions of this space that have been unexplored by current development efforts. Not only does this shed light on the possible nature of future social machines, it may also help us to identify the specific combination of characteristics that determine whether a particular social machine fails or flourishes within the (current) socio-technical ecology of the Web.

5 Related Work

Given the value of taxonomies in advancing our understanding of the conceptual landscape associated with a domain, it is no surprise to discover that taxonomies have been developed for a range of systems appearing in the context of the Social Web. This includes, most notably, crowdsourcing [11, 14] and human computation systems [34], although similar attempts at characterization have been made in respect of social computing [1] and collective intelligence systems [25]. While none of the concepts associated with these systems are synonymous with the notion of social machines (see [43]), there are clear relationships between these various concepts. Instances of at least the technological components of social machines (e.g., Facebook, Wikipedia, Galaxy Zoo, etc.) are sometimes presented as instances of other kinds of systems, and this suggests that some of the dimensions associated with the social machine taxonomy may also surface in the context of other taxonomies. In order to evaluate this, we systematically compared the dimensions listed in Tables 2, 3, 4, 5, and 6 (see appendix) with those appearing in other studies [1,11,14,25,34]. The results of this analysis are presented in Table 1 (see appendix). As can be seen from this table, a number of social machine dimensions have at least some partial mapping to the dimensions identified in other studies.[29] This is particularly noticeable in the case of human computation and crowdsourcing systems (although this may simply reflect the greater attention that has been afforded to these systems in the context of previous taxonomy development efforts) [11, 14, 34].

[29] Note that although two dimensions may be similar, they are only regarded as identical if the set of characteristics associated with the dimensions is the same in each case. In the absence of shared characteristics, a dimension mapping is regarded as ‘partial’.

6 Future Work

The definition of social machines presented in Sect. 2 and the taxonomic framework presented in Sect. 4 form part of an integrated attempt to develop a conceptual foundation for social machine research. It should be clear, however, that much more work needs to be done to make progress in this area. In terms of our conceptual understanding of social machines, for example, a range of perspectives exist concerning the nature of social machines. The definition of social machines that we have adopted here (and also in [45]) emphasizes the role of human and technological elements in the joint realization of processes. We might refer to this as the 'socio-technical perspective' of social machines. Such a perspective is, however, only one among many alternative perspectives that could be countenanced. While our definition is largely consistent with the views expressed by others in the Web Science community,[30] there are a number of competing perspectives available, and these need to be given closer scrutiny. An alternative concept, for example, tends to see social machines as socio-computational systems. According to this view, social machines are socially-extended computational systems in which some aspects of the computational process are delegated to multiple human individuals. Perhaps unsurprisingly, this kind of view tends to emerge in discussions of what has been dubbed 'the social computer' [38]. While there is clearly a certain amount of common ground between the 'socio-computational perspective' and the 'socio-technical perspective' (e.g., both regard social machines as systems that implement certain types of processes), there is a significant difference in terms of the scope of the conceptualizations entertained by each perspective. In particular, the socio-computational perspective seems committed to the view that social machines exist as a specialized form of human computation system [24]. We suggest that this contributes to an unproductive narrowing of the scope of social machine research efforts: it limits our scientific remit to a subset of Web-based systems whose constituent processes can be properly described as 'computational' in nature. In addition, by casting social machines as a specialized form of human computation system, we allow the scientific effort associated with the study of social machines to be too easily subsumed within an existing, and well established, field of scientific enquiry. In our view, the term 'social machine' is best reserved for a class of systems whose most important distinguishing feature is the manner in which system-level processes are realized. This is preferable to a perspective that focuses on issues of whether the process in question is or is not computational in nature. The crucial difference between the two perspectives is highlighted by the emphasis the

[30] Such consistency is evidenced by the way social machines are described in a number of papers. We thus encounter descriptions of social machines as "purposefully designed sociotechnical system[s] comprising machines and people" [10], as systems in which "the human and digital parts. . . [form] a machine in which the two aspects are seamlessly interwoven" [43], and as systems that involve "the co-constitutional involvement of humans and technologies" [50].

socio-technical perspective places on the *way* in which a process is realized (i.e., the details of its mechanistic realization); the issue of whether or not the process in question can be characterized in computational terms is largely irrelevant.

A further focus area for conceptual analytic efforts is to distinguish between the notion of a social machine from a variety of ostensibly similar notions. These include crowdsourcing [7, 11], human computation [34], collective intelligence [25], social computing [32], the global brain [4], the social computer [38] and the social operating system [35]. It has been suggested that the social machine concept is similar to but not synonymous with (at least some of) these other concepts [43]. Additional work is required, however, to elucidate the exact nature of the relationships between the concepts. Furthermore, it will be important to ascertain the degree of overlap in the extensional projections of the concepts expressed by these terms.

As a means of furthering the effort to improve our conceptual understanding of social machines, we may be able to extend the methodological approach that was adopted in the case of the taxonomic framework; i.e., we may be able to make use of a range of knowledge elicitation techniques. Aside from the repertory grid technique (described above), a number of other knowledge elicitation techniques are available, and these could be useful in terms of exploring the social machine conceptual landscape. These include laddering techniques (useful for eliciting hierarchically-organized classes of social machines), concept sorting techniques (useful for identifying the features of social machines) and concept mapping techniques (useful for identifying the relationships between social machines) (see [42]). As with other applications of knowledge elicitation techniques, the results of these studies could serve as the basis for ontology development efforts. Such ontologies could then be used to provide machine-readable characterizations of specific social machine instances.

Regarding the effort to develop a taxonomic framework for social machines, a number of further steps need to be undertaken. Following the methodology advocated by Nickerson et al. [29], our work to date has focused on the empirical-to-deductive and deductive-to-empirical stages. The aim of the third stage—taxonomy application—is to use the taxonomic framework to identify and characterize additional instances of social machines. By situating these instances within the social machine morphospace, we will be able to chart the location of unexplored or under-explored regions of the design space. Of course, given the number of dimensions and the number of potential social machines that may emerge in the context of the current and future Web (recall that the study by Shadbolt et al. [43] yielded an initial sample of 56 social machines), the task of taxonomy application is likely to be something of a laborious undertaking (at least when seen from the perspective of a single individual). Clearly, one strategy for dealing with these sorts of tasks is to draw on the (socially-situated) processing resources made available by the technological infrastructure of the World Wide Web. This is precisely the strategy taken by social machines and other kinds of systems within the context of the Social Web. An interesting possibility, therefore, is to engineer a social machine to expedite the process of taxonomy application.[31] One specific idea that is currently under development is to build a microtask environment, including specific game elements, in which participants are asked to provide answers to atomic challenges that rate and compare a pair of social machine instances according to a dimension in our framework. Such systems may serve as a useful adjunct to ongoing initiatives, such as the Web Observatory initiative [51], which seek to observe the behavior of social machines within the ecological environment of the Web [10].

The use of the taxonomic framework to characterize new social machine instances is also a useful way of validating the framework. In particular, the attempt to characterize novel social machines enables us to answer questions concerning the generality (e.g., can we specify characteristics for all social machines?), accuracy (e.g., are the characteristics associated with a particular dimension mutually exclusive for any given social machine, or can a social machine have multiple characteristics on the same dimension?) and reliability (e.g., is the same system characterized in the same way by multiple users?) of the framework.[32] We may, of course, discover at this stage that some putative social machines cannot be accommodated within the taxonomic framework. This may point to an inadequacy of the framework, or (more positively) it may indicate that the system in question is not, in fact, a social machine. In other words, the taxonomic framework could (ultimately) serve as a useful means of identifying bona fide members of the class of social machines. There are a number of different methodologies in the knowledge engineering literature which describe the steps to be followed in order to carry out the validation, and means to measure and analyze different validation criteria (see, for instance, [56]).

The use of the taxonomic framework to identify and characterize social machines yields a range of benefits. Firstly, by situating social machines within the social machine morphospace, we are able to determine the degree of clustering within the design space. We are able to answer questions concerning the extent to which existing systems are clustered together (like stars within a galaxy) or whether they are more-or-less randomly distributed across the void. This helps to determine whether current design and engineering efforts are focused on particular regions of the design space. Secondly, the population of the morphospace enables us to imagine as yet unrealized forms of social machines. By supporting our ability to focus on previously unexplored regions of the morphospace, the taxonomic framework is functioning as a 'cognitive scaffold' for our imaginative efforts. Such efforts

[31] Note that in the light of our definition, the 'engineering' of a social machine entails more than just software development and deployment; it also includes the assembly of mechanisms that enable and encourage user engagement.

[32] The reliability of the framework is indicated by inter-rater reliability metrics. Poor measures of inter-rater reliability may indicate that some dimensions are more difficult to interpret, understand or discern than others. This may call for the dimension to be refined or removed from the framework.

may feed into the design and development of new kinds of social machines. Thirdly, we can use the body of data associated with the characterization of social machines in order to support efforts aimed at identifying categories or classes of social machines (using quantitative methods). We have alluded to a number of these categories earlier in the paper. For example, as a result of the repertory grid analysis described in Sect. 4.1, we made reference to 'social networking systems' and 'media sharing systems'. Other classes of social machine focus on certain vertical sectors, for instance 'crime social machines' [12] and 'health social machines' [54]. Clearly, the effort to develop a hierarchically-organized set of social machine classes is an important focus area for future work,[33] and it could feed into the aforementioned effort to develop a social machine ontology. Finally, the application of the taxonomic framework yields a body of data that can be used to assess the relationship between particular combinations of characteristics and a range of interesting properties relating to (e.g.) the performance profile of the system and the size of its user community. These kinds of properties tend to be ones that determine how 'successful' a social machine is (e.g., whether it is able to achieve the goals its designers originally intended it to achieve), and thus the collection of correlational data is potentially useful in terms of guiding the design and development of new machines, as well as configuring existing ones. It should also be noted that the dimensions associated with the taxonomic framework can serve as independent variables in the context of experimental efforts intended to elucidate the relationship between particular characteristics and performance outcomes. Such variables are particularly useful in the case of cognitive social simulation studies where large numbers of cognitively-sophisticated agents can be used to shed light on the complex interactions between factors spread across the technological, informational, social and cognitive domains [47]. They can also offer a useful empirical grounding for system designers and inform the engineering and evolution of existing systems. One element of our future work in this area aims to investigate the dynamics of social machines using a combination of multi-agent simulation and cognitive modeling techniques (see [37]). From an engineering and HCI perspective, we will analyze the data collected through the application of the framework to derive best practices and guidelines for system design, which might also prove useful for ongoing initiatives such as the Web Observatory initiative.

[33] The process of identifying categories or classes of social machines is supported by the use of statistical methods that are applied to the social machine morphospace. Cluster analytic techniques are typically used to support these analyses (see Geiger et al. [14] for an example of such techniques applied to crowdsourcing systems).

7 Conclusion

The recent growth and influence of the Social Web has led to an intensification of research efforts to understand the nature and dynamics of Web-based socio-technical systems. As part of these efforts, the term 'social machine' has emerged to help focus attention on a specific class of systems and to help delineate a range of theoretical, empirical and engineering issues. Although there is still no consensus regarding the precise semantics of the term 'social machine', we suggest that the notion of a social machine can best be understood in terms of particular processes (i.e., ones in which our explanatory accounts need to advert to the details of social participation and bio-technological coupling). We thus endorse the following definition of social machines:

> Social machines are Web-based socio-technical systems in which the human and technological elements play the role of participant machinery with respect to the mechanistic realization of system-level processes.

As part of the effort to improve our conceptual understanding of social machines, we have attempted to construct a taxonomic framework. This framework draws on previous work that relied on the use of knowledge elicitation techniques to capture information about the various dimensions along which extant social machines can be deemed to vary [43]. We have extended this initial work by refining the set of elicited dimensions and also identified discrete characteristics for each of the dimensions. The result is a taxonomic framework consisting of a total of 33 dimensions and 106 characteristics. This framework defines a multi-dimensional design space—the social machine morphospace—within which, it is suggested, all social machines (extant or otherwise) can be accommodated.

The effort to develop a taxonomic framework is important for a number of reasons. Aside from the rather obvious sense in which a taxonomic framework improves our understanding of the similarities and differences between social machines, a taxonomic framework can help us to identify unexplored or under-explored regions of the design space. It can also help to identify clusters of social machines that denote conceptually-important classes or categories of social machines. Finally, the taxonomic framework provides a set of variables that can be exploited in the context of more empirical efforts. For example, some of the dimensions may serve as the independent variables for experimental simulations undertaken as part of cognitive social simulation [47] and computational social science [15] studies.

The research that is reported here forms part of a larger effort to establish a conceptual foundation for social machine research. Given that the recent growth and expansion of the Internet, particularly the Web, has been driven by the emergence of systems such as Facebook, Twitter, YouTube and so on, all of which have been regarded as social machines, the study of social machines is of crucial importance to members of the Web and Internet Science community. In addition, the next generation of social machine systems have been implicated in a range of advanced capabilities, including curing diseases, solving world hunger, and deriving strategies to mitigate the effects of climate change [18]. This makes the study of social machines of interest to those concerned with our future individual and collective problem-solving capabilities. Finally, social machines are of critical interest in terms of understanding the relationship between the Web and wider society. By supporting the emergence of new forms of social interaction, organization and coordination, social machines are progressively altering the way a broad array of social activities are performed, ranging from the way we communicate and transmit knowledge, establish romantic partnerships, generate ideas, produce goods and maintain friendships. This establishes the basis for more profound forms of social change in which social machines progressively alter the organization and dynamics of our future society. This potential to effect various forms of social change makes the topic of social machines an important focus of research attention for those working across a variety of social science and engineering disciplines.

Acknowledgements This work is supported under SOCIAM: The Theory and Practice of Social Machines. The SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/1 and comprises the Universities of Southampton, Oxford and Edinburgh.

Appendix

Tables 1, 2, 3, 4, 5, and 6 present the dimensions of the taxonomic framework for social machines.

Table 1 Mapping of social machine dimensions to the dimensions associated with four other systems (i.e., social computing, collective intelligence, human computation and crowdsourcing systems)

Type[a]	Social machine dimension	Target system type	Target dimension	Source
P	Input validation mechanism	Human computation	Quality control	[34]
P	Input validation mechanism	Crowdsourcing	How to evaluate inputs	[11]
P	Human ability	Human computation	Human skill	[34]
P	Human ability	Crowdsourcing	What users can do	[11]
P	Combinatorial strategy	Social computing	From conveyance to convergence content generation	[1]
P	Combinatorial strategy	Human computation	Aggregation	[34]
P	Combinatorial strategy	Crowdsourcing	Aggregation of contributions	[14]
P	Combinatorial strategy	Collective intelligence	How (structure/process)	[25]
P	Combinatorial strategy	Crowdsourcing	How to combine inputs	[11]
P	Task type	Collective intelligence	How (structure/process)	[25]
P	Task type	Crowdsourcing	Type of target problem	[11]
U	Control flow	Human computation	Process order	[34]
U	Task user cardinality	Human computation	Task request cardinality	[34]
U	Sociality	Social computing	From information to people connections	[1]
U	Community specification	Crowdsourcing	Preselection of contributors	[14]
U	Community specification	Collective intelligence	Who (staffing)	[25]
U	Visibility of user contributions	Crowdsourcing	Accessibility of peer contributions	[14]
U	Visibility of user contributions	Crowdsourcing	Nature of collaboration	[11]

U	Response to user contributions	Crowdsourcing	Accessibility of peer contributions	[14]
U	Task assignment policy	Crowdsourcing	Preselection of contributors	[14]
U	Task assignment policy	Collective intelligence	Who (staffing)	[25]
U	Group/individual assignment	Collective intelligence	Who (staffing)	[25]
Q	Quality assessment mechanism	Human computation	Quality control	[34]
Q	Quality assessment mechanism	Crowdsourcing	How to evaluate inputs	[11]
M	Form of motivation	Human computation	Motivation	[34]
M	Form of motivation	Social computing	From utilitarian to hedonic use	[1]
M	Form of motivation	Collective intelligence	Why (incentives)	[25]
M	Reward type	Human computation	Motivation	[34]
M	Reward type	Collective intelligence	Why (incentives)	[25]
M	Reward type	Crowdsourcing	How to recruit and retain users	[11]
M	Reward variability	Crowdsourcing	Remuneration for contributions	[14]

[a] Indicates the type of the dimensions from the social machine taxonomic framework: P = ‘goal, task and process dimensions’, U = ‘user participation and interaction dimensions’, Q = ‘quality assessment dimensions’, and M = ‘motivational factors and incentive mechanism dimensions’

Table 2 Dimensions and characteristics for the category ‘motivational factors and incentive mechanisms’

Dimension	Description	Characteristics
Motivation type	Specifies the type of motivation associated with user participation	Intrinsic/extrinsic
Form of motivation	Specifies the form of motivation associated with user participation	Economic/altruistic/hedonic/reputational/instrumental/other
Reward type	Specifies the type of reward made for user contributions	None/monetary payment/prize/other
Reward variability	Specifies whether reward quantities are fixed or variable. Variable rewards are encountered when rewards are related to individual or collective performance	Fixed/variable/none

Table 3 Dimensions and characteristics for the category ‘technology and engineering’

Dimension	Description	Characteristics
Open source status	Specifies whether the technological elements of the social machine are open source	Open source/not open source
Social machine framework status	Specifies whether the social machine is derived from a generic framework, such as MediaWiki, Diaspora or Ushahidi	Based on framework/not based on framework

Table 4 Dimensions and characteristics for the category ‘goal, task and process’

Dimension	Description	Characteristics
Goal variability	Indicates whether the goal of the social machine is stable across the lifetime of the social machine, or whether it is likely to change	Fixed/variable
Goal visibility	Indicates whether the goal is visible to the human users of the system	Visible/hidden
Output type	Specifies the kind of output that results from the processes performed by the social machine	Physical/social/cognitive/informational
Output ownership	Indicates who owns the results of process execution	System designer/larger community
Task type	Specifies the kind of task that is performed by the system	Evaluating/organizing/sharing/networking/creating/other
Human ability	Specifies the nature of the primary human ability that is required as part of the process	Aesthetic/emotional/epistemic/perceptual/behavioural/social/moral/cognitive/linguistic
Combinatorial strategy	Specifies how the contributions of individual participants are combined during the course of process execution[a]	Additive/compensatory/disjunctive/conjunctive/discretionary
Input validation mechanism	Indicates how individual user contributions are checked or validated	Automatic/manual/none

[a] These characteristics are derived from Steiner’s [46] categories of task independence

Table 5 Dimensions and characteristics for the category ‘quality assessment’

Dimension	Description	Characteristics
Quality assessment mechanism	Indicates how the quality assessment process is undertaken	Automatic/manual/mixed/none
Explicit/implicit nature of quality criteria	Indicates whether quality assessment criteria are explicitly or implicitly specified	Explicit/implicit
User involvement in quality evaluation	Indicates whether users are involved in the evaluation of process outcomes	User involvement/no user involvement
Quality criteria variability	Indicates whether quality assessment criteria are fixed or variable over the lifetime of the social machine	Fixed/variable

Table 6 Dimensions and characteristics for the category 'participation and interaction'

Dimension	Description	Characteristics
Social role differentiation	Indicates whether or not users have different roles within the system	Social role differentiation/ no social role differentiation
Functional role variability	Indicates whether or not users are engaged in different processes or the same process as part of their participation in the machine	Functional role variability/ no functional role variability
User autonomy	Indicates the extent to which users decide what they work on and when they work on it	User autonomy/no user autonomy
Community specification	Indicates whether the user community of the system is a subset of the total population. A subset of users may be based on a variety of characteristics, such as demographic factors or the possession of particular skills and abilities	Specified/unspecified
Task atomicity	Indicates whether the user engages in atomic tasks, multiple tasks of the same kind or a combination of tasks	Atomic/multiple instance/combined
Control flow	Indicates the order in which the tasks performed by multiple agencies are executed	Sequence/parallel/split/ synchronization/asynchronous merge/exclusive choice/ iteration
Visibility of user contributions	Specifies the visibility of user contributions to other users of the system	Restricted/unrestricted/variable
User anonymity	Indicates the extent to which participating users are required to provide personal information about themselves to other users	High anonymity/low anonymity
Response to user contributions	Specifies the kinds of ways in which users respond to the contributions made by other users. User contributions may be enriched (e.g., via tagging) or modified. In addition, one user may respond to the contribution of another user by posting related content	None/enrich/modify/respond
User process awareness	Indicates the extent to which users have full knowledge of what is going on in the system	Local awareness/global awareness
Task assignment policy	Specifies how tasks are assigned to users of the system	Random/role-based/skill-based/contribution-based
Task-user cardinality	Specifies the relationship between specific tasks and user assignments	One-to-one/one-to-many/ many-to-many/many-to-one

(continued)

Table 6 (continued)

Dimension	Description	Characteristics
Group/individual assignment	Specifies whether tasks are assigned to individuals or groups	Individual/group
Proportion of active participants	Specifies the proportion of participants that are actively involved in a process as opposed to those who merely consume the contributions of others	High/low/balanced
Sociality	Indicates the extent to which the system supports social interaction with other members	High sociality/low sociality

References

1. Ali-Hassan, H., Nevo, D.: Identifying social computing dimensions: A multidimensional scaling study. In: Intelligent Conference on Information Systems. Phoenix, Arizona (2009)
2. Bailey, K.D.: Typologies and Taxonomies: An Introduction to Classification Techniques. Sage, Thousand Oaks (1994)
3. Berners-Lee, T., Fischetti, M.: Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. Harper Collins, New York (1999)
4. Bernstein, A., Klein, M., Malone, T.W.: Programming the global brain. Commun. ACM **55**(5), 41–43 (2012)
5. Bonabeau, E.: Decisions 2.0: The power of collective intelligence. MIT Sloan Manag. Rev. **50**(2), 45–52 (2009)
6. Caputi, P., Reddy, P.: A comparison of triadic and dyadic methods of personal construct elicitation. J. Constr. Psychol. **12**(3), 253–264 (1999)
7. Chi, E.H., Bernstein, M.S.: Leveraging online populations for crowdsourcing. IEEE Internet Comput. **16**(5), 10–12 (2012)
8. Clark, A.: Supersizing the Mind: Embodiment, Action, and Cognitive Extension. Oxford University Press, New York (2008)
9. Clark, A., Chalmers, D.: The extended mind. Analysis **58**(1), 7–19 (1998)
10. De Roure, D., Hooper, C., Meredith-Lobay, M., Page, K., Tarte, S., Cruickshank, D., De Roure,C.: Observing social machines Part 1: What to observe? In: WWW2013 Workshop: The Theory & Practice of Social Machines. Rio de Janeiro, Brazil (2013)
11. Doan, A., Ramakrishnan, R., Halevy, A.Y.: Crowdsourcing systems on the World Wide Web. Commun. ACM **54**(4), 86–96 (2011)
12. Evans, M.B., O'Hara, K., Tiropanis, T., Webber, C.: Crime applications and social machines: Crowdsourcing sensitive data. In: WWW2013 Workshop: The Theory & Practice of Social Machines. Rio de Janeiro, Brazil (2013)
13. Fransella, F., Bell, R., Bannister, D.: A Manual for Repertory Grid Technique, 2nd edn. Wiley, Chichester (2003)
14. Geiger, D., Seedorf, S., Schulze, T., Nickerson, R.C., Schader, M.: Managing the crowd: Towards a taxonomy of crowdsourcing processes. In: Americas Conference on Information Systems. Detroit, Michigan (2011)
15. Gilbert, N., Troitzsch, K.G.: Simulation for the Social Scientist, 2nd edn. Open University Press, Maidenhead (2005)

16. Gilles, D., Guattari, F.: Anti-Oedipus. Continuum, London (2004)
17. Hall, W., Tiropanis, T.: Web evolution and web science. Comput. Netw. **56**, 3859–3865 (2012)
18. Hendler, J., Berners-Lee, T.: From the Semantic Web to social machines: A research challenge for AI on the World Wide Web. Artif. Intell. **174**, 156–161 (2010)
19. Jankowicz, D.: The Easy Guide to Repertory Grids. Wiley, Chichester (2003)
20. Kelly, G.A.: The Psychology of Personal Constructs. W.W. Norton and Company, New York (1955)
21. Klein, G., Moon, B., Hoffman, R.R.: Making sense of sensemaking 2: A macrocognitive model. Intell. Syst. **21**(5), 88–92 (2006)
22. Kraut, R., Maher, M.L., Olson, J., Malone, T.W., Pirolli, P., Thomas, J.C.: Scientific foundations: A case for technology-mediated social-participation theory. Computer **43**(11), 22–28 (2010)
23. Landes, D.: Revolution in Time: Clocks and the Making of the Modern World. Viking Press, London (2000)
24. Law, E., von Ahn, L.: Human computation. Synth. Lect. Artif. Intell. Mach. Learn. **5**(3), 1–121 (2011)
25. Malone, T.W., Laubacher, R., Dellarocas, C.: The collective intelligence genome. MIT Sloan Manag. Rev. **51**(3), 21–31 (2010)
26. McBride, N.: From social machine to social commodity: Redefining the concept of social machine as a precursor to new Web development approaches. In: 3rd International Conference on Web Science, Koblenz (2011)
27. Meira, S.R.L., Burégio, V.A.A., Nascimento, L.M., Figueiredo, E., Neto, M., Encarnação, B., Garcia, V.C.: The emerging web of social machines. In: 35th Annual Computer Software and Applications Conference (COMPSAC), pp. 26–27. IEEE, Munich (2011)
28. Menary, R.: The Extended Mind. MIT Press, Cambridge (2010)
29. Nickerson, R., Muntermann, J., Varshney, U., Isaac, H.: Taxonomy development in information systems: Developing a taxonomy of mobile applications. In: European Conference in Information Systems, Verona (2009)
30. O'Hara, K.: Trust in social machines: The challenges. In: AISB/IACAP World Congress 2012: Social Computing, Social Cognition, Social Networks and Multiagent Systems, Birmingham (2012)
31. O'Hara, K.: Social machine politics are here to stay. IEEE Internet Comput. **17**(2), 87–90 (2013)
32. Parameswaran, M., Whinston, A.B.: Research issues in social computing. J. Assoc. Inf. Syst. **8**(6), 336–350 (2007)
33. Potthast, M., Stein, B., Gerling, R.: Automatic vandalism detection in Wikipedia. In: 30th European Conference on Information Retrieval Research. Glasgow, Scotland (2008)
34. Quinn, A., Bederson, B.: Human computation: A survey and taxonomy of a growing field. In: Annual Conference on Human Factors in Computing Systems (CHI'11). Vancouver, British Columbia (2011)
35. Rainie, L., Wellman, B.: Networked: The New Social Operating System. MIT Press, Cambridge (2012)
36. Raup, D.M.: Geometric analysis of shell coiling: General problems. J. Paleontol. **40**, 1178–1190 (1966)
37. Richardson, D., Smart, P.R., Sycara, K., Stone, P., Giammanco, C., Powell, G.: Using ACT-R to model collective sensemaking in military coalition environments. In: Annual Fall Meeting of the International Technology Alliance. Palisades, New York (2013)
38. Robertson, D., Giunchiglia, F.: Programming the social computer. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **371**(1987), 20120379 (2013)
39. Secretan, J.: Stigmergic dimensions of online creative interaction. Cogn. Syst. Res. **21**, 65–74 (2013)
40. Secretan, J., Beato, N., D'Ambrosio, D., Rodriguez, A., Campbell, A., Folsom-Kovarik, J., Stanley, K.: Picbreeder: A case study in collaborative evolutionary exploration of design space. Evol. Comput. **19**(3), 373–403 (2011)

41. Shadbolt, N.R.: Knowledge acquisition and the rise of social machines. Int. J. Hum. Comput. Stud. **71**(2), 200–205 (2013)
42. Shadbolt, N.R., Smart, P.R.: Knowledge elicitation: Methods, tools and techniques. In: Wilson, J.R., Sharples, S. (eds.) Evaluation of Human Work, 4th edn. CRC Press, Boca Raton (2015) (in press)
43. Shadbolt, N., Smith, D.A., Simperl, E., Van Kleek, M., Yang, Y., Hall, W.: Towards a classification framework for social machines. In: WWW2013 Workshop: The Theory & Practice of Social Machines, Rio de Janeiro (2013)
44. Smart, P.R.: The web-extended mind. Metaphilosophy **43**(4), 426–445 (2012)
45. Smart, P.R., Shadbolt, N.R.: Social machines. In: Khosrow-Pour, M. (ed.) Encyclopedia of Information Science and Technology. IGI Global, Hershey, 3rd edn. (2014)
46. Steiner, I.D.: Group Processes and Productivity. Academic, New York (1972)
47. Sun, R.: Cognitive social simulation incorporating cognitive architectures. Intell. Syst. **22**(5), 33–39 (2007)
48. Surowiecki, J.: The Wisdom of Crowds: Why the Many are Smarter than the Few. Random House, New York (2005)
49. Sutton, J.: Exograms and interdisciplinarity: History, the extended mind, and the civilizing process. In: Menary, R. (ed.) The Extended Mind. MIT Press, Cambridge (2010)
50. Tinati, R., Carr, L.: Understanding social machines. In: ASE/IEEE International Conference on Social Computing and International Conference on Privacy, Security, Risk and Trust, Amsterdam (2012)
51. Tiropanis, T., Hall, W., Shadbolt, N., De Roure, D., Contractor, N., Hendler, J.: The web science observatory. IEEE Intell. Syst. **28**(2), 100–104 (2013)
52. Tong, S., Van Der Heide, B., Langwell, L., Walther, J.: Too much of a good thing? The relationship between number of friends and interpersonal impressions on Facebook. J. Comput. Mediat. Commun. **13**(3), 531–549 (2008)
53. Tyszka, J.: Morphospace of foraminiferal shells: Results from the moving reference model. Lethaia **39**(1), 1–12 (2006)
54. Van Kleek, M., Smith, D.A., Hall, W., Shadbolt, N.R.: "The crowd keeps me in shape": Social psychology and the present and future of health social machines. In: WWW2013 Workshop: The Theory & Practice of Social Machines, Rio de Janeiro, Brazil (2013)
55. von Ahn, L., Dabbish, L.: Designing games with a purpose. Comm. ACM **51**(8), 58–67 (2008)
56. Vrandečić, D.: Ontology evaluation. In: Staab, S., Studer, R. (eds.) Handbook of Ontologies. International Handbook on Information Systems, 2nd edn., pp. 293–314. Springer, Berlin (2009)
57. Westerman, D., Spence, P., Van Der Heide, B.: A social network as information: The effect of system generated reports of connectedness on credibility on Twitter. Comput. Hum. Behav. **28**, 199–206 (2012)
58. Wheeler, M., Clark, A.: Genic representation: Reconciling content and causal complexity. Br. J. Philos. Sci. **50**(1), 103–135 (1999)
59. Wilson, R.A., Craver, C.F.: Realization: Metaphysical and scientific perspectives. In: Thagard, P. (ed.) Philosophy of Psychology and Cognitive Science. North-Holland, Oxford (2007)
60. Zook, M., Graham, M., Shelton, T., Gorman, S.: Volunteered geographic information and crowdsourcing disaster relief: A case study of the haitian earthquake. World Med. Health Policy **2**(2), 7–33 (2010)

The Mathematician and the Social Computer: A Look into the Future

Martin Charles Golumbic

1 The Mathematician

Most mathematicians spend their time thinking about tough, ivory tower, theoretical math problems. In a recent article [2], Prof. Gunnar Carlsson from Stanford University is quoted as saying: "*Mathematicians want to work on the deepest, hardest problems and get interesting intellectual results*".

Professor Carlsson is a co-founder of Ayasdi,[1] a Palo Alto tech startup that applies topology to analyze large volumes of data. Big Data has become Big Business. From biotech to cyber-security and social networking, fresh insights are pulled out of huge databases in record time. Today's powerful data analysis algorithms gather vast troves of information—breaking it down to illuminate patterns and relations to better serve the challenges of society. Yet what is being done today, is just a hint of what will be the state-of-art 10 years from now.

This is but one example of combining the strengths of human mathematical reasoning and machine capabilities for problem solving. Research and development in all areas of mathematics, be it topology, combinatorics, or graph theory, holds hidden promise for many applications involving what might be called artificially intelligent agents.

I am a Mathematician. That is both a true statement about me, and also the title of a 1956 book by the legendary mathematician Norbert Wiener [4]. Reading this book as a student, there was one quote that stood out in my mind:

[1] http://www.ayasdi.com/.

M.C. Golumbic (✉)
The Caesarea Rothschild Institute for Computer Science, University of Haifa,
Mt. Carmel, Haifa, 31905, Israel
e-mail: golumbic@cs.haifa.ac.il

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_4

> If the general public ever thinks [positively] of mathematics, it sees it at best as a tool for the physicist and the statistician and at worst as something closely akin to the work of an accountant. Hardly any non-mathematician will admit that mathematics has a cultural and aesthetic appeal, that it has anything to do with beauty or power or emotion.

Wiener categorically denies such a cold and rigid concept of mathematics. He would say that "*the task of the mathematician is to use a rigid and demanding medium to express a new and significant vision of some aspect of the universe; to … reveal something new and something exciting.*" He asks, "*Is a poet less free because his language has a grammar or his sonnets a form?*"

He goes on to partially answer the question. "*What differentiates the appeal of the artist-mathematician from the artist-sculptor and the artist-musician is not the unemotionality of his public but the strict discipline necessary to become a connoisseur of mathematics.*"

Now, many grey hairs later, I think this axiom still applies to mathematical research, and to Social Collective Intelligence as well. During one demonstration of his topology-based system, Prof. Carlsson harvests through genetic data on thousands of breast cancer patients to show which groups of women will respond best to chemotherapy and what their DNA has in common. In Palo Alto, the company Ayasdi and others like it are taking a leap forward to create an Intelligent Agent, and making a significant impact on pharmaceutical, energy, medical and defense organizations through their technology.

2 A Leap Forward: The Intelligent Agent

Welcome to Jeopardy!

During the Fall of 2011, while I was on sabbatical in New York, I had the pleasure of watching the Jeopardy! television quiz show competition between Watson (the IBM program) and the two human world champions Brad Rutter and Ken Jennings. For those not familiar with the game, a phase or sentence is shown and read to the contestants as the "answer" to a question, and the first contestant to press the buzzer must supply the correct "question".

Ladies and gentlemen welcome back to Double Jeopardy! The category is:

Mathematicians of the 20th century.

This 20th century French mathematician wrote the zeroth book on graph theory in 1926.

Any takers?

Who was Deénes König? No, he was Hungarian and wrote the first book on graph theory.

Who was Claude Berge? No, he was only born in 1926!

The correct question is,

"Who was André Sainte-Laguë?", and the title of his book, "Les réseaux (ou graphes)".

I do not know how Watson would have performed on this "answer", but with its 200 million pages of structured and unstructured content (four terabytes of disk space including many databases, dictionaries and the full text of Wikipedia,[2] Watson consistently outperformed its human opponents. By combining and integrating the best that Artificial Intelligence can offer today in the fields of Natural Language Processing, Machine Learning, Search and Hypothesis Generation, the best IBM and university scientists demonstrated that a computer system can now be competitive with humans in ways not possible previously.

The difference between Watson/Jeopardy! as a major challenge project for IBM and its previous Deep Blue grandmaster chess program project 15 years earlier, is the potential to exploit the technology more broadly and to provide significant benefits to areas such as health care. To quote Jon Brodkin of IBM, "*The goal is to have computers start to interact in natural human terms across a range of applications and processes, understanding the questions that humans ask and providing answers that humans can understand and justify.*"

The Jeopardy! competition was only the beginning. Watson is now reading and analyzing vast amounts of industry data, and answering even bigger questions. It and its future cousins will change the way we live and work.

I personally see this as an opportunity to have, within 10 years or less, productive brainstorming sessions between humans and intelligent-appearing software agents. Speech recognition and query understanding may lead us to sitting around a conference table, as illustrated below in Fig. 1, sooner that we think.

[2] http://en.wikipedia.org/wiki/Watson_(computer).

Fig. 1 Human and intelligent agent interaction

Our mathematician just might be able to help an intelligent agent with some of its major challenges. For example, how can it deal with concepts of reasoning about time? What models would be helpful if it needs to monitor a real-time system for a nuclear power plant? What is the difference between regarding time as points or intervals? Should it collect and use temporal data at a very fine granularity like nanoseconds? If its cousin is monitoring elderly patients in a hospital, would a coarse granularity like second-by-second be sufficient, and are there computational implications? Might the agent be interested in a higher level set of activities, or have to deal with partial information and synchronizing time lines? How does it resolve contradictions? It seems to me that a robot has a lot to learn from an experienced mathematician [1].

3 Concluding Remarks

Norbert Wiener wrote in 1950, “The world of the future will be an even more demanding struggle against the limitations of our intelligence, not a comfortable hammock in which we can lie down to be waited upon by our robot slaves.” [3] But I believe it will be, in many ways, a better world for human thought. To quote Hacham Menachem ben Yona, “The machine frees the human mind, and challenges it to new horizons.”

References

1. Golumbic, M.C.: Perspectives on reasoning about time. In: Krüger, A., Kuflik, T. (eds.) Ubiquitous Display Environments, pp. 53–70. Springer, New York (2012)
2. Vance, A.: Ayasdi: Stanford Math Begets a Data Company (2013). http://www.businessweek.com/articles/2013-01-24/ayasdi-stanford-math-begets-a-data-company
3. Wiener, N.: The Human Use of Human Beings: Cybernetics and Society. Houghton Mifflin Co., Boston (1950)
4. Wiener, N.: I Am a Mathematician. Doubleday and Co., Garden City (1956)

Twelve Big Questions for Research on Social Collective Intelligence

Stuart Anderson, Daniele Miorandi, Iacopo Carreras, and Dave Robertson

1 Introduction

In this chapter we provide three top-down challenge areas that we believe could play an important role in shaping the development of the field of social collective intelligence. These challenge areas will help shape bottom-up developments emerging in the relevant R&D&I communities and will help frame how they contribute to the development of deployed social computation systems. Each section provides some examples of the challenge and finishes up with a small number of high-level questions:

1. The first raises the issue of the existence of useful sub-classes of social computation that arise naturally and, on the face of it, appear to be less complex than the full notion.
2. The second stems from the observation that there are many "naturally occurring" social computation systems already in operation and we can challenge our conceptions of social computation using empirical means by identifying characteristic mechanisms in use in some of these social computational systems.
3. The third challenge identifies some key attributes of social computational systems and attempts to identify work in the social science literature that bears on these attributes.

S. Anderson (✉) • D. Robertson
School of Informatics, University of Edinburgh, Informatics Forum, 10 Crichton Street,
Edinburgh EH8 9AB, UK
e-mail: soa@staffmail.ed.ac.uk,dr@inf.ed.ac.uk

D. Miorandi • I. Carreras
CREATE-NET, v. alla Cascata 56/D, 38123, Trento, Italy
e-mail: daniele.miorandi@create-net.org; iacopo.carreras@create-net.org

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social
Sciences, DOI 10.1007/978-3-319-08681-1_5

We then identify twelve “big questions” in social collective intelligence. These represent key challenges that the research community at large should tackle in order to fully exploit the potential embedded in social collective intelligence systems.

The remainder of the chapter is organised as follows. In Sect. 2 we describe some candidate sub-classes of social computation. In Sect. 3 we discuss existing social collective intelligence mechanisms, their relevance and limitations. In Sect. 4 we identify relevant ideas from social sciences. Section 5 includes the description of the twelve big questions we identified for research in social collective intelligence. Section 6 concludes the chapter.

2 Candidate Sub-classes of Social Computation

One important challenge is to identify sub-classes of social computations that are in some sense “simpler” than the general notion. These should be linked to real applications but should be applicable across a range of situations. We have three (or four) sub-classes at the moment but we envisage this list could be extended both by considering particular domains of application but also using features such as architecture or scale to help control complexity:

- Computer-mediated social sense-making of socially generated data: this involves developing social computational systems that allow people to contextualise, correct and interpret data gathered by sensors in the environment. This class is inspired by the problem of contextualising telehealth data gathered by patient in their own home. At the moment most systems generate many false alarms because users, their families and surrounding social context have no way of adding to the raw data and those interpreting the data are both remote and cautious. More generally social interpretation of data seems like a key component in many systems and an important simplified sub-class of social computation.
- Organisational routines [10–12] are notoriously difficult to capture because they involve very large numbers of exceptional cases and necessitate drawing on experience of past situations to help decide the best course of action in new situations but at the same time the situation is constrained since we are working with an identifiable data set and what we aim to deliver is fairly well understood. This provides a good context to consider how social computation can share experience and promote social learning in a relatively constrained and small-scale environment.
- Markets have already been very heavily studied and benefit from a single, financial, notion of value. This is an important sub-class with many well-developed social mechanisms in place. For example, hedge funds, are a good example of Social Collective Intelligence where the social group includes a range of skills, including the capacity to develop new trading algorithms and means of monitoring and visualising market dynamics. Hedge funds typically exist in a complex human/machine symbiosis with market-relevant information shaping

the strategy and execution carried out using high frequency trading. Studies of hedge funds argue that decision taking in a hedge fund is markedly different from decision taking by individuals or groups that do not live in an information-rich environment [5]. Overall, markets comprise many of these new composite social/machine actors competing around maximising profit in a highly monitored risky environment. This context provides a good environment for studying the limits of a single notion of value and in exploring approaches to regulation and its circumvention.

- One could also consider a hybrid of the first two sub-classes above to consider transparent organisations where data about the behaviour of the organisation is gathered continuously and made available to people working in the organisation as a means of helping develop novel lines of communication and action. There has been earlier work on transparent finance in organisations that we could use as a starting point [14].

These categories suggest the following questions:

- What features of social computational systems does each sub-class highlight or eliminate from consideration?
- Are there other such sub-classes that are interesting and might help us investigate social computation?
- Are there refinements of the sub-classes that might be more useful in studying social computation?

3 Existing Mechanisms

This challenge proposes a systematic study of some of the existing social coordination mechanisms such as: recommendation, trading, dating, gaming, friending etc. We should explore how they are deployed at the moment and what social structures they are capable of supporting. One particularly interesting area is to study how aggregated data is imbued with social meaning. For example:

- Friending or linking: how does this process differ between say LinkedIn[1] and Facebook?[2] In Facebook, what is the significance of your number of friends and how was that constructed? Does this construction imbue unfriending with increased significance? How do these mechanisms fit within our framework of social computation and can we account for the social element in the way these mechanisms operate?
- Recommendation: this works well in some contexts and less well in others. As we become more sophisticated users of recommendation systems they are becoming

[1] https://www.linkedin.com/.

[2] https://www.facebook.com/.

more complex and include features such as stratified populations and multi-dimensional recommendation measures. Can we characterise situations in which recommendation is likely to be effective? Can we also identify components and means of composition that can help in the design process of recommendation systems intended to support particular work practices?
- H-index: In academic circles h-index is increasing seen as an important measure of academic influence or significance. How has this been constructed? How have issues around the interpretation of h-index been dealt with? How do different communities interpret h-index? In the longer term will there be an attempt to create other competing impact measures and how will any discrepancies be reconciled or magnified, by which social groupings? This is a good example of performative notions where the social milieu "performs" the definition and it fills its intended social role (and may also have some unintended side-effects).

This is a deliberately short list of mechanisms intended to illustrate the idea. The ideas are apparently simple but they have a character we see in many social computation settings. Human mechanisms need to be conceptually graspable and apparently very simple mechanisms have the capacity to support complex behaviour. One important area will be the development of mechanisms that are easily socially graspable and are useful in a range of settings. This idea of humanly comprehensible or graspable mechanisms is interesting because some social mechanisms that do exist are very useful but remain difficult for people to comprehend. One example in this area is e-voting systems where the mechanisms have good properties but remain underused because people find them opaque and difficult to trust. In developing this area further we will consider the following questions:

- Can we extend this list of mechanisms that are deployed in some social computational system?
- Can we characterise contexts in which the deployment of these mechanisms is likely to be effective?

4 Ideas from the Social Sciences

We should also explore some notions from the social sciences that may give us some grip on barriers of promoters of the social computation concept. A short initial list would include the following:

- Governance: There is a substantial body of work in the social sciences on governance but of particular interest in Social Computation is the work of Elinor Ostrom [9] who has considered approaches to the governance of common pool resources or "commons".
- Decision Making: There is a substantial body of work in social science on decision taking that takes account of limited resources and access to information. This will include game theory from an economic perspective, work like

Giegerenzer's [4, 13] on the use of heuristic decision taking. In addition the work of Tversky and Kahneman's [7] on choice and risk is potentially interesting.
- Trust: Trust is a key factor in the operation of social computations but there is a persistent tendency to attempt to eliminate trust in favour of some evidence-based measure of risk. The social science literature, in particular Mollering [8], avoids this shift and attempt an analysis of trust. This is important because it is a key contributor to smoothly operating human systems.
- Risk: Ullrich Beck's [1] work on Risk Society is a key text in contributing to an understanding of the ethical foundations of large-scale systems and the role of science and technology in social transformation.
- Coordination: Star and Bowker [2] have identified boundary objects as elements in working contexts that allow groups to cooperate without requiring consensus. Boundary objects are ubiquitous in complex organisational settings and taking account of Star and Bowker's work will be critical to the success of social computation.
- Evolution: Bowker, Edwards [3] and others have developed a framework to consider complex human/machine infrastructures that support complex social cooperation. For example, organised and citizen science is one area of study. Paul Edwards study of the climate science infrastructure ("A Vast Machine") points out many complex operations that are required for long-lived infrastructures that may not initially seem to be a requirement of an information infrastructure.

In considering the social in Social Collective Intelligence we will need to take full account of work from the social sciences. This suggests we need to answer two related questions:

- What other social aspects of Social Collective Intelligence do we need to take into account in developing SCI further?
- Can we refine our interest in particular topics (e.g. those listed above) to ask sharper questions of the social science literature?

5 "Big Questions" in Social Collective Intelligence

Social Collective Intelligence has the potential to transform most aspects of human life if it is widely adopted. In particular it offers the potential for a radical transformation in the way information is gathered, integrated and used to support human activity. This chapter outlines some of the key dimensions of Social Collective Intelligence from an interdisciplinary perspective. In assessing the suitability of Social Collective Intelligence as a focus for further work we can reframe this work as a series of "big questions" that such a programme might take as important issues to resolve. Of course there is a balance to be struck between what we anticipate can be achieved by such a programme and setting ambitious "stretch goals". Thus our big questions are not at the level of generality of a so-called "Hilbert question" (e.g. Is mathematics decidable?) but they are set to provide considerable challenge both at the level of individual disciplines and in interdisciplinarity where there is a need for insight that require multiple disciplinary perspectives.

1. *How do we specify and verify Social Collective Intelligence systems?* Borrowing from conventional verification perspectives we will be interested in both safety properties and liveness properties of SCI systems. Safety properties demonstrate the absence of particular behaviour, for example we might be interested to demonstrate an SCI system is incapable of identifying information about individual participants or to demonstrate that particular kinds of instability are absent (e.g. that some individual or group are persistently stressed beyond their capacity). Liveness properties might include demonstrating that an SCI system has the capacity to respond resiliently to some class of events or that the system admits the capacity for particular types of evolution depending on the context of use. Characterising such properties effectively is challenging and developing a verification framework where we can illustrate properties of a system to some level of confidence is also challenging and novel.
2. *What is the data model for Social Collective Intelligence?* There is considerable range of possibilities for data models. For example, we might want to have data that explicitly represents some measure over a population and particular sub populations based on characteristics of that measure. We might also want this to be represented as a time series to capture the dynamics of the measure over time. The general issue is how best to represent captured behavioural data and how to interpret that captured data to provide information on social action.
3. *What is a programming model for Social Collective Intelligence?* We know we need to specify human and computational resources combined by a structure of governance and incentives (or disincentives) associated with particular forms of action. The issue is the design of appropriate structuring mechanisms for the description of these resources and rules and how to test a system under development to observe the likely behaviour of the system in use. Since we envisage that at least some aspects of SCI system will develop after the system has been deployed we need to consider a programming model that includes how to describe the initial state of the system, how it evolves during its lifetime and how to bring things to an orderly close. We might also consider what a highly distributed "socialized" development environment might look like. The programming model would need to develop alongside the data model. An essential aspect of such models is the need to account for governance via policy-like elements that take account of human rule-following characteristics.
4. *What is the process of creating Social Collective Intelligence systems that are "fit for purpose"?* This is the "engineering" process of coordinating a range of resources to bring about a change in the way social groups interact. This will involve identifying the needs of a complex, possibly highly structured, stakeholders group and working to meet needs effectively. This process will involve considering how to structure the human resources to gain the required

effect. This may involve detailed simulations of social action and the negotiation of changes in governance and regulatory structure to enable changes in social action.

5. *What is the network model for SCI?* We will need to consider softer, more flexible ways of modelling networks that include a considerable element of entirely social networking. This will introduce different types of interaction into the graph that take account of human aspects of rule following, transgressions and the open-world structure of human experience. Any networking model will need to take account of locality, polycentrism and the need for the capacity to take coordinated action without requiring agreement between cooperating parties.
6. *How do we manage key dimensions of Scale, Space and Time in Social Collective Systems?* The stability of many governance systems depends on locality and limitation of scale arising from limits on locality combined with loss of information over time. Modern SCI breaks all of these limiting mechanisms. This has many immediate consequences that we can see arising today. For example, we are beginning to see the rise of global-scale online labour markets like elance[3] that have the potential dramatically to transform the nature of work [6]. The emergence of such structures will require the development of stabilizing mechanisms that balance power relationships and permit more reliable prediction of the dynamics of these systems in new conditions. This understanding will pave the way for more effective governance of SCI systems that offer agility and flexibility in the presence of rapidly changing circumstances.
7. *What is the model that supports the access and interpretation of data?* Our goal will be to build on the developing approaches to Web data access and management mechanisms but to include much more of the social dimension in access and interpretation. We envisage the social curation of datasets and the development of a range of mechanisms that allow social sensemaking to play an important role in interpreting data. We will also consider the identification and development of expertise amongst users and the pivotal role played by locality in determining variation in the interpretation of data and understanding of context.
8. *What is the impact of SCI on the nature of work?* We have already touched on the capacity of SCI to transform the way labour and tasks are connected in the emerging new e-labour markets. More generally we envisage that the emergence of SCI as a key element in modern economies will see the development of a whole range of new employment that engages in supporting SCI operation and development.
9. *What are good models to initiate SCI systems?* Often SCI systems will only work effectively if a significant proportion of the population participate in

[3] https://www.elance.com.

using the SCI system. We need to consider generic mechanisms to launch SCI systems in a way that is likely to achieve levels of adoption that allow the system to function as intended.

10. *SCI may transform what we see as intellectual work, how do we provide good governance of IPR in the context of new forms of intellectual production?* Since SCI will transform intellectual work it is highly likely to undermine traditional views of Intellectual Property rights. Understanding better what we want to know about ownership of ideas will allow us to design systems that are adequately instrumented to allow the allocation of property rights in an appropriate manner.
11. *Is it possible to regulate via SCI mechanisms?* One possible role for SCI systems is the development of architectures that provide more agile and responsive regulatory regimes than is currently the case. Such a regulation controller would allow us to consider more dynamic approaches to regulation with very rapid decision-making.
12. *We will want to place monitoring and transparency requirements on SCI, how is this to be achieved?* We will need to provide the means to monitor SCI systems effectively to see that the system is behaving as we anticipate and to learn the dynamics of the system when it encounters radically different environments.

6 Conclusion

This chapter points to the fascinating interdisciplinary challenge in developing the field of Social Collective Intelligence. The main contribution of the chapter lies in the identification of a number of research challenges ("big questions") for moving the field of social collective intelligence forward. While such questions do not constitute a structured research agenda, we do hope that they serve as a stimulus for the research community at large in prioritizing scientific issues to tackle.

References

1. Beck, U.: Risk Society: Towards a New Modernity, vol. 17. Sage, London (1992)
2. Bowker, G.C., Star, S.L.: Sorting Things Out: Classification and Its Consequences. MIT Press, Cambridge (1999)
3. Edwards, P.N.: A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming. MIT Press, Cambridge (2010)
4. Gigerenzer, G., Hertwig, R., Pachur, T.: Heuristics: The Foundations of Adaptive Behavior. Oxford University Press, Oxford (2011)
5. Hardie, I., MacKenzie, D.: Assembling an economic actor: The agencement of a hedge fund. Sociol. Rev. **55**(1), 57–80 (2007)
6. Horton, J.J.: Online Labor Markets. Springer, Berlin (2010)
7. Kahneman, D., Tversky, A.: Choices, values, and frames. Am. Psychol. **39**(4), 341 (1984)
8. Möllering, G.: Trust: Reason, Routine, Reflexivity. Elsevier, Oxford (2006)

9. Ostrom, E.: Polycentric systems as one approach for solving collective-action problems. Available at SSRN 1304697 (2008)
10. Pentland, B.T., Feldman, M.S.: Organizational routines as a unit of analysis. Ind. Corp. Change **14**(5), 793–815 (2005)
11. Pentland, B.T., Feldman, M.S.: Narrative networks: Patterns of technology and organization. Organ. Sci. **18**, 781–795 (2007)
12. Pentland, B.T., Feldman, M.S.: Designing routines: On the folly of designing artifacts, while hoping for patterns of action. Inf. Organ. **18**, 235–250 (2008)
13. Todd, P.M., Gigerenzer, G.: Ecological rationality: Intelligence in the world. Oxford University Press, Oxford (2012)
14. Wehmeier, S., Raaz, O.: Transparency matters: The concept of organizational transparency in the academic discourse. Public Relat. Inquiry **1**(3), 337–366 (2012)

Part II
Technologies

Privacy in Social Collective Intelligence Systems

Simone Fischer-Hübner and Leonardo A. Martucci

1 Introduction

In this chapter we discuss privacy, a fundamental human right, in Social Collective Intelligence Systems (SCIS). Privacy is a key non-functional requirement related to the right of individuals to control information related to them. The fundamentals of SCIS are based on basic concepts such as profiling, provenance, evolution, reputation and incentives. This chapter discusses the impact of such concepts on the individual right to privacy. It also discusses that while SCIS have some inherent characteristics that can be utilized to promote privacy, still several technical challenges remain. Both privacy laws as well as privacy-enhancing technologies are needed to effectively enforce privacy.

This chapter is organized as follows. The concept of privacy is introduced in Sect. 2 and relevant basic privacy principles of the European Data Protection Legal Framework and the Organization for Economic Co-operation and Development (OECD) Privacy Guidelines are presented in Sect. 3. The risks to privacy, mainly in terms of profiling, provenance, trust and reputation in SCIS are listed in Sect. 4. Then, the opportunities provided by the design of SCIS that can help to promote privacy as well as related technical challenges are discussed in Sect. 5. Section 6 outlines legal privacy rules provided by the European Data Protection Legal Framework in regard to profiling and Sect. 7 presents a selection of privacy-enhancing technologies that can technically enforce the basic privacy principles in SCIS. Finally, Sect. 8 briefly summarizes the main findings and open research challenges.

S. Fischer-Hübner • L.A. Martucci (✉)
Karlstad University, 651-88 Karlstad, Sweden
e-mail: simone.fischer-huebner@kau.se; leonardo.martucci@kau.se

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_6

2 Concept of Privacy

Privacy is a core value and is recognized either explicitly or implicitly as a fundamental human right by most constitutions of democratic societies. In the end of the nineteenth century, the American lawyers Warren and Brandeis defined privacy as the "right to be let alone" [45]. Another definition from the early years of computing is by Alan Westin, who defined privacy as the "the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others" [47].

In general, the concept of personal privacy has several dimensions, including the dimensions of informational privacy (by controlling whether and how personal data can be processed or disseminated—see also Westin's definition), territorial privacy (by protecting the close physical area surrounding a person) and privacy of a person (by protecting a person against undue interferences) [20]. In the context of Social Collective Intelligence, the aspect of informational privacy will be the most relevant one and will thus also be the focus of our discussion.

Privacy, however, is not an absolute right, as it can be in conflict with rights of others or other legal values, and because individuals cannot participate fully in society without revealing personal data. Nevertheless, in cases where privacy has to be restricted, the very core of privacy still needs to be protected. Therefore, privacy and data protection laws, as those ones implementing the EU Data Protection Directive 95/46/EC [16], have the objective to define fundamental privacy principles that need to be enforced if personal data is collected, stored or processed. Such fundamental privacy principles will be discussed in the next section.

The EU Data Protection Directive 95/46/EC and most other privacy and data protection laws and guidelines only apply if *personal data* are processed, which are defined by Art. 2 of the Directive as "any information related to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, psychological, mental, economic, cultural or social identity;".

The Opinion 4/2007 of the Article 29 Working Party[1] contains an analysis of the concept of personal data described in the EU Data Protection Directive 95/46/EC. Among its conclusions and clarifications, the Working Party noted that data relates to an individual if it refers to the identity, characteristics or behavior of an individual or if such information is used to determine or influence the way in which that person is treated or evaluated. For instance, data that is related to the individuals behavior profiled under RFID tag identifiers associated to them or the MAC addresses of their smartphone wireless interfaces is personal data, even though these individuals may not be known or identified by their names. The Opinion 4/2007 of the Article 29

[1] The Article 29 Working Party consists of a representative from the data protection authority of each EU Member State, the European Data Protection Supervisor and the European Commission.

Working Party also states that natural persons are 'identified' when, assuming that they are part of a group of persons, they are distinguished from all other members of the group.[2]

3 Basic Privacy Principles

In this section, we provide an overview to internationally accepted, basic legal privacy principles, which are part of the general EU Data Protection Directive 95/46/EC [16] that need to be addressed by SCIS. The Data Protection Directive has been an important legal instrument for privacy protection in Europe, as it codifies general privacy principles that have been implemented in the national privacy laws of all EU member states and of many other states. The principles also correspond to principles of the OECD Privacy Guidelines [36] to which we will also refer to.

1. **Legitimacy:** Personal data processing has to be legitimate, which is according to Art. 7 EU Directive 95/46/EC usually the case if the data subject has given his unambiguous (and informed) consent, if there is a legal obligation, or contractual agreement (cf. the Collection Limitation Principle of the OECD Guidelines).
2. **Purpose specification and purpose binding:** Personal data must be collected for specified, explicit and legitimate purposes and may not be further processed in a way incompatible with these purposes (Art. 6 I b EU Data Protection Directive 95/46/EC—cf. Purpose Specification and Use Limitation Principles of the OECD Guidelines).
3. **Data minimization:** The processing to personal data must be limited to data that are adequate, relevant and not excessive (Art. 6 I (c) EU Data Protection Directive 95/46/EC). Besides, data should not be kept in a personally identifiable form any longer than necessary (Art. 6 I (e) EU Data Protection Directive 95/46/EC—cf. Data Quality Principle of the OECD Guidelines, which requires that data should be relevant to the purposes for which they are to be used). In other words, the collection of personal data and extend to what personal data are used should be minimized, allowing for instance users to act anonymously or pseudonymously. Obviously privacy is best protected if no personal data at all (or at least as little data as possible) are collected or processed.
4. **Restriction for the processing of sensitive data:** According to Art. 8 EU Data Protection Directive 95/46/EC, the processing of so-called special categories of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, or aspects of health or sex life are generally prohibited, subject to exceptions (such as explicit consent).

[2] The Article 29 Working Party statement is close to the definition of anonymity from Pfitzmann and Hansen [37]: "anonymity of a subject from an attackers perspective means that the attacker cannot sufficiently identify the subject within a set of subjects, the anonymity set", which is commonly used in the computer security and privacy area.

5. **Transparency and Rights of the Data Subjects:** Transparency of data processing means informing a data subject at least about the data processing purposes as the identity of the data controller[3] as well as further information, such as information about the possible recipients of the data and the rights and controls of the data subject.[4] The EU Data Protection Directive 95/46/EC provides data subjects with respective information rights according to its Art. 10. Further rights of the data subjects include the right of access to data (Art. 12 (a) EU Directive 95/46/EC), the right to object to the processing of personal data (Art. 14 EU Directive 95/46/EC), and the right to correction, erasure or blocking of incorrect or illegally stored data (Art. 12 (b) EU Directive 95/46/EC, cf. Openness and Individual Participation Principle of the OECD Guidelines).

   Of special interest for SCIS are data subject rights in the context of automated decisions that are, for instance, made based on profiling. According to Art. 12 (a) EU Directive 95/46/EC, the right to access data includes the right to obtain from the data controller "knowledge of the logic involved in any automatic processing of data concerning the data subject at least in the case of the automated decisions". Pursuant to Art. 15 (1) EU Directive 95/46/EC, individuals have in principle "the right not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."
6. **Security of data processing:** The data controller needs to implement appropriate technical and organizational security mechanisms to guarantee the confidentiality, integrity, and availability of personal data (Art. 17 EU Directive 95/46/EC—cf. Security Safeguards Principle of the OECD Guidelines);

In January 2012, the EU Commission published a proposal for a new EU General Data Protection Regulation (GDPR) [17], which defines a single set of modernized privacy rules, and which will (once the regulation will be in force) be directly valid across the EU. On October 12, 2013, the LIBE Committee (Committee on Civil Liberties, Justice and Home Affairs) of the European Parliament voted on compromise amendments to the GDPR [18]. In particular, it includes the principle of data protection by design and by default (Art. 23), requiring building privacy enhancing technologies (PETs) already into the initial system design. Besides, the requirements of transparency of data handling by *concise, transparent, clear and easily accessible policies* (Art. 11) is explicitly stressed. Moreover, the right to erasure is newly introduced in Art. 17 (which was initially branded as the right to be forgotten in the GDPR from January 2012).

[3] According to EU Data Protection Directive 95/46/EC, a data controller is defined as the entity that alone or jointly with others determines the purposes and means of personal data processing.

[4] According to EU Data Protection Directive 95/46/EC, a data subject is a natural person about whom personal data are processed has in regard to his personal data.

Important in the context of Social Collective Intelligence are also newly introduced rules on profiling (Art. 20), including the data subject's right to object to profiling as well as prohibition of profiling that has a discriminatory effect on the grounds of race, ethnic origin, political opinions, religion, philosophical beliefs, trade union membership, sexual orientation or gender identity. "The controller shall implement effective protection against possible discrimination resulting from profiling" (see further discussion below).

Even though the GDPR and its amendment are not enacted yet, it contains legal principles that have been broadly accepted as being important for the protection of privacy in the future.

4 Risks to Privacy in Social Collective Intelligence Systems

SCIS are based on technical concepts, such as profiling, reputation and incentives systems, and data provenance, which all may require the collection and processing of personal data and thus pose privacy risks. Some specific privacy risks related to these technical concepts will be discussed in this section.

4.1 Profiling

Profiles are sets of data that portray significant features of a subject. It aims to represent the extent to which an individual exhibits traits or abilities as determined by tests or ratings [34]. Data used to build profiles are mainly taken from individual's input, which is either explicitly or implicitly revealed or implicitly derived. The explicitly revealed data relate to information and statements that individuals directly disclose about themselves. The implicitly revealed data relate to information that is (automatically) gathered from supervisory systems or sensors that track the activities of individuals. Implicitly derived data are additional data that can be inferred from the data set and it is not produced or collected from individuals. It usually relates to results from statistical analysis on the data set. For instance, social networks contain explicitly revealed data posted by their users; loyalty programs collect data from customers' shopping or traveling activities, i.e. implicitly revealed data, and both social networks and companies running loyalty programs implicitly derive data about the customer habits.

Profiling affects privacy in different respects. As the Council of Europe has discussed in its recommendation CM/REC(2010)13 on profiling [10], the collection, linking, calculation, comparison and statistical correction of data with the objective to create profiles may have significant privacy impacts, as profiling enables a person's personality, behavior, interests and habits to be determined, analyzed and/or predicted. Often such profiling is even happening without the knowledge of the individuals concerned. While profiling may offer benefits for users and society at large, e.g. by providing users with targeted and better services addressing personal and societal interests or by permitting an analysis of risks and fraud. Profiling techniques can also have a negative impact on the individuals concerned by placing them in predetermined categories that may unjustifiably deprive them from accessing certain services and by this discriminate individuals [10].

Moreover, as mentioned above, profiling techniques do not only allow to analyze data that are actually recorded, but also allow to statistically predict or implicitly derive personal information from such records. For instance, it has been shown that sensitive data including political opinions, religious beliefs, intelligence or sexual orientation can be automatically predicted from Facebook Likes (see e.g., [28]).

For these reasons, it is important to protect privacy rights of individuals subject to profiling both by law and technology. Legal rules and privacy enhancing technologies for protecting the user's privacy will be discussed in the subsequent sections.

4.2 Provenance and Reputation

Reputation is a result of past interactions within a given context [12]. Reputation systems help users to select providers offering competing services. Obtaining a good reputation is a powerful incentive for service providers because the better their reputation is, the more services can be delivered or a higher premium can be gained. Hence, both service consumers and providers benefit from reputation systems.

In reputation systems, sequences of past interactions are linked to a subject and the aggregated quality of such interactions is used to determine the reputation of the subject. Provenance is therefore needed to correctly associate an interaction to a subject consuming a service and the service to a service provider. Thus, the correct identification of subjects and services is fundamental for provenance. The process of identification naturally requires some sort of identifier and, in the case of reputation systems and provenance in general, these identifiers are needed to be long-term identifiers because a history of past actions is going to be associated to them.

However, having numerous transactions linked to a single long-term identifier potentially reveals customs and habits of data subjects, i.e., personal data. In addition, decision-making based on reputation systems can be based on direct and indirect interactions, i.e., opinions from other users. Expressing one or multiple opinions about a service can potentially reveal personal information about the users' habits and lead to profiling. Users could then refrain by providing feedback but that would reduce the usefulness of the reputation system. Therefore, privacy in reputation systems has to be considered from the perspective of users providing services and of users consuming services.

From the data protection perspective, short-term identifiers, such as transactions pseudonyms [37], i.e., pseudonyms that are used only once, are able to better protect the privacy of data subjects because their multiple transactions are not easily linked but that would weaken the security of the reputation system, as it could be easily abused. There is a clear conflict between a key requirement of reputation services, i.e., keeping histories of interactions, and general privacy goals, i.e., keeping transaction records unlinkable. Reputation, which is an intrinsic type of incentive, and privacy are core aspects of SCIS that are required to co-exist. Therefore, this notional dissonance needs to be addressed and it is further discussed in Sect. 7.5.

5 Technical Opportunities and Challenges for Protecting Privacy

While SCIS pose different types of privacy risks as we have discussed above, they also have inherent characteristics, such as distribution, hybridity, and the focus on collectives instead of individuals only, which can be utilized for a privacy-enhanced system design. This section discusses opportunities and challenges for designing a privacy-preserving system that takes into account the inherent characteristics and technical concepts of SCIS.

5.1 Formation of Collectives and Privacy

Social collective intelligence is based on hybrid systems, where humans and machines compose and closely cooperate as a collective to solve challenging tasks. A key feature of SCIS is the utilization of group intelligence by composing the "right" collective (or set) of humans and machines that is suitable for solving a given task.

The formation of collectives is related with privacy from two main directions. First, from the anonymity perspective, we evaluate how peers (humans), which are part of collectives, can remain anonymous. Second, from the identity management perspective, we present how collectives can be used for audience segregation and for handling multiple partial identities.

5.1.1 Collectives and Anonymity

The peer profile of a larger collective may not classify as personal data, if the collective is formed in such a way that it does not relate to any identified or identifiable person, i.e., if the individuals of the collective are anonymous and devices that are part of the collectives do not provide personal data. In this case, privacy of individuals is not affected and privacy laws do not apply.

As privacy will be best protected if no personal data are processed at all or if personal data cannot be directly attributed to the data subjects, research challenges to be addressed also include the question how peer profiles can be anonymized or pseudonymized, and/or how peer profiles of collectives can be formed in an anonymous manner.

One leading principle for the formation of collectives in SCIS is diversity [4]. For instance, diversity in opinions helps to eliminate decision bias in collectives and promote different viewpoints. The notion of diversity is also a key component in anonymity metrics, i.e., standards of measurements that aim at quantifying the level of privacy of a subject. Anonymity means that a subject is not identifiable within a set of subjects (the anonymity set) who might have caused a given action [37] or associated to a given piece of information [43]. The cardinality of the anonymity set can be used as a simple privacy metric.

Diversity has a strong impact, either positive or negative, on the privacy of subjects. First of all, diversity decreases the homogeneity of the set of subjects and, thus, may also reduce the cardinality of anonymity sets and the level of privacy for the subjects (persons) that are elements of these sets. The anonymity set size is related to another metric, the k-anonymity.

K-anonymity [43] is a formal privacy protection model that aims at preventing the re-identification of individuals in a given person-specific field-structured data (structured database) while maintaining the utility (usefulness) of the data. The idea behind k-anonymity is that a record from a database is released only if there are at least $(k - 1)$ other similar records, i.e., whose values of quasi-identifiers are indistinguishable from the each other. Thus, there are at least k subjects that can be linked to a given release of data. In addition, k-anonymity can be used to quantify anonymity in location-based services, as shown in [22, 25].

L-diversity [30] is a model that extends k-anonymity. It proposes a solution for the blindness of k-anonymity regarding diversity in sensitive information that can be exploited using attacks that use public (non-sensitive) information to obtain sensitive information. The idea behind l-diversity is that the diversity of sensitive attributes has to be at least l (where $l > 1$). Therefore, lack of diversity of sensitive attributes can also negatively affect privacy.

T-closeness [29] extends l-diversity by proposing restrictions to the disclosed sensitive data, which should follow the distribution of the overall table. Differential privacy [15] is a formal model that ensures that addition or removal of single items of a database does not significantly affect the outcome of an analysis. Differential privacy shows that any statistical database that releases data with a non-trivial utility also leaks personal information. Differential privacy also offers means to quantify the level of loss of personal information against the utility of the data retrieved from the database. Data mining with formal privacy guarantees based on differential privacy is described in [21].

While anonymity is hard to guarantee and hard to measure, still the approaches and metrics mentioned above could help to compose collectives that also form suitable anonymity sets.

5.1.2 Collectives and Privacy-Enhancing Identity Management

Peers can take part in multiple collectives and provide different contributions in terms of f and skills to each collective. In principle, this also allows one human to be represented by different (partial) identities in different collectives or to be represented in one collective with different agents, which represent different (partial) identities of the user in dependence on the current context.

The sociologist Erving Goffman described the concept of audience segregation, meaning that people usually play different roles in different situations and perform differently for different audiences [24]. Privacy-enhancing identity management systems [8] technically enforces audience segregation by allowing users to selectively disclose subsets of their personal data, so-called partial identities, under different pseudonyms to different communication partners dependent on their current context.

While establishing multiple identities prevents users and their agents from being completely profiled under one identity and thus promotes privacy, it also enables compromises by so-called Sibyl attacks. A Sybil attack is an identification attack that occurs when a malicious user influences the network by controlling multiple logical identifiers from a single physical device. In a Sybil attack, malicious users assume multiple identifiers, preventing the usage of security mechanisms based on filters, reputation or trust assumptions [14]. In [32], the concept of self-certified Sybil-free pseudonyms is presented, which allows protecting against Sybil attacks on distributed systems in a privacy-friendly manner.

5.2 Distribution for Promoting Privacy

While centralized systems and collections of data pose privacy risks due to data mining and potential data leakages, Decentralized Systems and Services for Privacy Preservation, such as online social networks, private data storage and backup, or anonymous content dissemination and communication systems, have been developed and researched in the recent years that are removing the need for a powerful centralized provider with its knowledge (see [6]).

Examples are peer-to-peer anonymous communication mechanisms, such as Crowds [41] and Chameleon [31], which are run by the collective of users and based on the compositionality of individual interactions [23]. Tor [13], the most relevant anonymous communication system, is also supported and run by collective that voluntarily offers networking and computing resources to provide anonymity to Internet users.

Online social networks can aggregate collectives and are potential important means for providing compositionality between collectives and machines, as the social networks provide an invaluable source for machines to learn from people. Safebook [11], Peerson [5], and Diaspora are distributed peer-to-peer privacy-friendly online social networks that were proposed and lately implemented.

The distributed nature of SCIS can potentially also be utilized for distributing knowledge and power and thus promoting privacy.

6 Legal Privacy Protection for Profiles

This section discusses how legal privacy rules that are enacted by the EU Data Protection Directive 95/46/EC or proposed as part of the GDPR and its compromise amendment can help to enforce privacy. As reputation scores and personalized incentives schemes [23] can also be viewed as profiles, this section focuses on legal means for protecting personal data of profiles in the form of peer profiles, reputation and incentives schemes.

If a profile contains personal data, then restrictions apply to the propagation or exchange of profiles according to the European data protection legislation. However, if a profile is anonymized and does not contain any personal data, the Directive 95/46/EC does not apply, as its Recital 26 states that "the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable." In practice, the question whether data is anonymous or not is very difficult to answer. This particularly applies to statistical data, "where despite the fact that the information may be presented as aggregated data, the original sample is not sufficiently large and other pieces of information may enable the identification of individuals." For instance, data sets published by AOL, a media company, and by Netflix, a provider of on-demand streaming media, in 2006 that were claimed to be anonymized were later proven not to be since a number of individuals could be re-identified from the data set (see [35]). The Recital 26 demands that for deciding whether data is anonymous "all the means likely reasonably to be used either by the controller or by any other person" should be taken into account.

Basic legal privacy principles, especially those enacted by the EU Data Protection Directive 95/46/EC and the proposed GDPR (cf. Sect. 3), need to be enforced when profiles including personal data are created and processed:

- The collection and processing of personal data in profiles needs to be *legitimate*, which usually implies that the data subjects have given their informed consent (Art. 7—Legitimacy and informed consent).
- Personal data used in the context of profiling must be collected for *specified and legitimate purposes* and may later *only be used for those purposes* (Art. 6 Ib—Purpose specification and binding).
- Furthermore, the amount of personal data and the extent to which they are collected and processed in profiles should be *minimized* (Art. 6 Ic—Data minimization), which implies that if possible data in profiles should be *anonymized* or *pseudonymized*.

- The collection and processing of *so-called special categories of data* in the context of profiling should in principle be *prohibited* (Art. 8 I—No sensitive data), unless the exceptions of Art. 8 II apply.
- Data controllers have to provide the data subjects with sufficient *privacy policy information* pursuant to Art. 10 when personal data are collected in the context of profiling. Data subjects that are being profiled have the right to access (i.e. to obtain information about) their personal data as well as the right to be informed by the data controller about the logic underpinning the processing of their profile data. Furthermore, data subjects have *rights to correction, deletion and blocking of their data*, as well as the *right not to be subject to a "decision which produces legal effects concerning him or significantly affects him and which that is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."* (Transparency and data subject rights).
- The data controller has to implement proper *technical and organizational security measures* for the protection of personal profile data (Art. 17—Security).

The Council of Europe has in an appendix to its recommendation CM/REC(2010) 13 proposed more specific privacy principles that should further strengthen the data subject's protection.

In the context of the EU Data protection reform, the newly proposed General EU Data Protection Regulation (GDPR) [17] introduced "Measures based on Profiling" with its Art. 20. This was however criticized by the Article 29 Data Protection Working Party on focusing merely on the outcome of profiling rather than on the profiling as such [3]. The Article 29 Data Protection Working Party therefore demands a comprehensive approach that also includes legal requirements for the purpose of profiling and the creation of profiles as such, referring to the principles of the appendix to Council of Europe recommendation.

The compromise amendment to the proposed EU Data Protection Regulation [18], which was passed by the LIBE Committee of the European Parliament on October 21, 2013, has taken up this proposal by providing greater transparency and control for data subjects. According to the amended Art. 14 (ga), data controllers should provide "information about the existence of profiling, of measures based on profiling, and the envisaged effects of profiling on the data subject". Besides, the amended proposal includes the right for data subjects to object to profiling (Art. 20 I). Furthermore, pursuant to Art. 20 III, "profiling that has the effect of discriminating against individuals on the basis of race or ethnic origin, political opinions, religion or beliefs, trade union membership, sexual orientation or gender identity, or that results in measures which have such effect, shall be prohibited". Pursuant to Art. 20 V, "Profiling which leads to measures producing legal effects concerning the data subject or does similarly significantly affect the interests, rights or freedoms of the concerned data subject shall not be based solely or predominantly on automated processing and shall include human assessment, including an explanation of the decision reached after such an assessment."

The GDPR defines 'profiling' as "any form of automated processing of personal data intended to evaluate certain personal aspects relating to *a natural person* or to analyze or predict in particular *that natural person's* performance at work, economic situation, location, health, personal preferences, reliability or behavior". Thus, the GDPR and its amendment text refer to profiles comprising data about one individual. However, if profiles contain personal data of several individuals or data that relate to several individuals, the enforcement of the legal rules in regard to consent, transparency and data subject rights discussed above may be practically more difficult to enforce—especially if the data subjects concerned have conflicting interests in regard the transparency, confidentiality or retention of their data.

The amendment text to the GDPR also introduced in Art. 4 (2a) the concept of "pseudonymous data", defined as "personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organizational measures to ensure non-attribution." Recital 58a of the amendment, further states that profiling based solely on the processing of pseudonymous data that cannot be attributed to a specific person should be presumed not to significantly affect the interests, rights or freedoms of the data subject.

The Article 29 Working Party is also pointing out the need for more responsibility of the data controllers. In particular, a data protection impact assessment as foreseen in Art. 33 of the GDPR needs to be conducted as a basis for suitable safeguards for profiles comprising privacy enhancing technologies and privacy friendly default settings (cf. Art. 23 of the GDPR on Data Protection by Design and by Default).

In conclusion, peer profiles relating to individuals may raise privacy concerns. Therefore, suitable legal and technical measures to protect the data subject's rights to information self-determination are needed. While the basic privacy principles of the EU Directive 95/46/EC are implemented by the national laws of the EU member states, the more advanced principles of the proposed EU regulation and its amendment are not finally enacted yet. Still, they reflect important requirements set up by privacy experts and decision makers and are expected to pass as part of the Regulation in this or similar form at least.

The privacy principles discussed above can be enforced more effectively by SCIS and applications by following a Privacy by Design approach. In Sect. 7, we will discuss privacy enhancing technologies for technically enforcing basic privacy principles.

7 Privacy-Enhancing Technologies

In this section, we present a selection of privacy-enhancing technologies (PETs) and show how they can be used to technically enforce basic privacy principles.

7.1 Anonymous Credentials

Anonymous credential protocols are key technologies for enforcing data minimization for applications.

A traditional credential (often also called certificate or attribute certificate) is a set of personal attributes, such as birth date, name or address, signed (and thereby certified) by the certifying party (the so-called issuer) and bound to its owner by cryptographic means. Traditional credentials require, however, that all attributes are disclosed together if the user wants to prove certain properties, so that the verifier can check the issuers signature. This makes different uses of the same credential linkable to each other. Besides, the verifier and issuer can link the different uses of the users credential to the issuing of the credential. Anonymous credentials allow the user to essentially "transform" the certificate into a new one that contains only a subset of attributes of the original certificate, i.e. they allow proving only a subset of its attributes to a verifier (selective disclosure property). Instead of revealing the exact value of an attribute (e.g., the exact birth date or address), anonymous credential systems also enable users to apply any mathematical function to the (original) attribute value, allowing them to prove only attribute properties without revealing the attributes themselves (e.g., one may only reveal the fact that she or he is over 18 and/or lives in Trento—which may be sufficient for authorizing a service). In addition, the Idemix protocol by Camenisch et al. [7], which is an implementation of anonymous credentials, allows the issuer's signature to be transformed in such a way that the signature in the new certificate cannot be linked to the original signature of the issuer. Hence, different credential uses cannot be linked by the verifier and/or issuer (unlinkability property).

7.2 Transparency Enhancing Tools

As mentioned in Sect. 3, transparency of personal data processing for data subjects is a basic privacy principle, and consequently the Legal European Data Protection Framework grants data subjects rights to information for making the processing of their data transparent.

Transparency-enhancing tools (TETs) provide technical means for enforcing these data subject rights. According to [27], TETs can be divided into ex ante TETs which enable the anticipation of consequences before data is actually disclosed, and ex post TETs which inform about consequences if data already has been revealed. Examples for ex ante TETs are privacy policy languages, such as P3P [44] or PPL [38], which could also be used in the context of SCSI for informing users more transparently about privacy policies, e.g. when they have to provide their informed consent to disclose personal data for peer profiling or other purposes.

Ex post TETs comprise tools that provide data subjects with online access to their data at the service provider's side [46] or access to logs documenting how their data were processed. As logs that are recording who has accessed data and how the data has been processed in turn also include personal data (e.g., the fact that a medical record of a patient has been accessed by a psychiatrist reveals sensitive personal information), they have to be designed in a privacy-friendly manner. Privacy preserving transparency logging schemes are for instance introduced in [26, 39]. They propose methods for the encryption of log records in such a way that the records are only accessible by the data subjects to which the records relate.

7.3 *PrimeLife Policy Language (*PPL*)*

Machine-readable privacy policy languages have the objective to make privacy policies of services sides more transparent, negotiable and enforceable. Compared to hard-coded fixed policies, they provide more flexibility, as they allow to easily express, change and extend privacy policies without the need to reimplement the system that enforces the policy. Besides, if the language is agreed-upon or standardised, privacy policies can easily be communicated across interacting entities in different domains [38].

The PrimeLife Policy Language (PPL) for privacy-enhanced access control and data handling was developed in the EU FP7 project PrimeLife [38], and is based on two widespread industry standards, XACML (eXtensible Access Control Markup Language) and SAML (Security Assertions Markup Language). PPL is a language that allows to specify privacy policies of data controllers as well as privacy preferences of users (who are the data subjects in this case), which can be matched to check whether a data controller's policy complies with a user's preferences.

Let's consider the scenario depicted in Fig. 1 involving a data controller requesting personal data from a user. The data controller may later want to forward the personal data to a third party, a so-called downstream controller.

The data controller sends the data request to the user together with a privacy policy, which consists of an access control policy, specifying what information he needs from the user, and a data handling policy specifying how he will treat the revealed data. PPL allows specifying both uncertified data requests as well as certified data requests based on proofs of the possession of (anonymous Idemix [7] or traditional X.509) credentials that fulfill certain properties. The data handling policy is expressed in terms of authorizations, e.g., for what purposes the data will be used, and obligations that the data controller is willing to fulfill for collected data items (e.g., to delete the data after a certain time period or to log all accesses to the data).

Similarly, the data subject's privacy preferences specify to which data controller and downstream data controllers each data item can be released and how users expect their data to be treated.

The PPL engine conducts an automated matching of the data controller's policy and the user's preferences, which can result in a mutual agreement concerning the

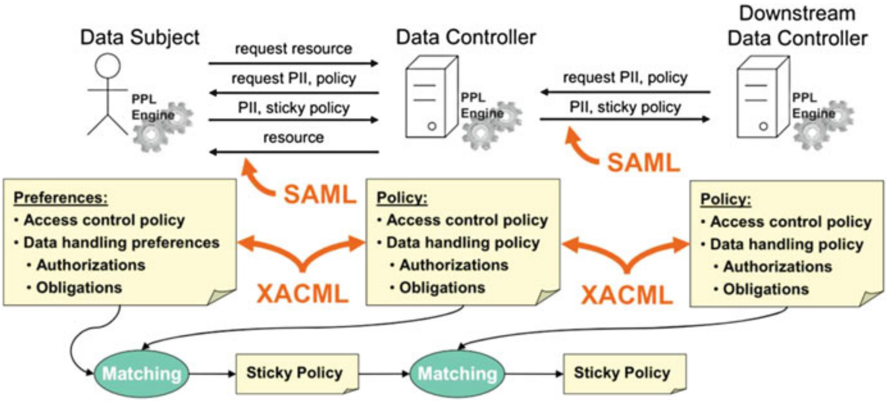


Fig. 1 Matching the data subject's privacy preferences and the data controller's privacy policy [38]

usage of data in form of a so-called "sticky policy", which should "stick to the data" and be stored and enforced at the data controller side (by the XACML access control engine). If data are to be forwarded to third parties (downstream data controllers), the data's sticky policy is first matched with the downstream controller's policy, which may result in a new sticky policy traveling with the data for enforcement at the downstream controller's side.

PPL extends XACML, so that the language can express both data subject's preferences and the (downstream) data controller's policies. Besides, concept of (anonymous Idemix) credential-based access control was integrated in PPL. For the communication between the different parties, SAML was extended to communicate credential-based attribute proofs and to attach sticky policies to the revealed attributes. Privacy policy languages such as PPL can be used to protect the access to personal data contained in peer profiles according to the user's preferences.

7.4 PETs for Protecting Peer Profiles

PPL and other PETs as presented above can help to technically enforce the main legal privacy principles derived from the European Legal Data Protection Framework in regard to peer profiling:

- **Informed consent:** With PPL, a consent form including a short privacy notice can be displayed to the user before the user discloses personal data from his peer profile, informing him about the main aspects of the data controller's privacy policy and about how far it matches with his privacy preferences. Only if the user provides his consent, there will be a valid agreement (in form of a sticky policy) and data will be disclosed. (For proposals of usable PPL user interfaces for obtaining informed consent, please refer to [2]);

- **Purpose specification and binding:** With PPL, the data controller's privacy policy can clearly state the purposes for which requested personal data items will be used. The XACML access control engine will enforce that personal data can only be accessed for the agreed upon sticky policy. This means that the use of data will be restricted to the purposes stated in the sticky policy;
- **Data minimization:** PPL allows the user to disclose certified data in form of anonymous credential proofs, which can via their selective disclosure and unlinkability properties enforce data minimization on application level. Furthermore, obligations, which a user and data controller have agreed upon, in regard to the data retention period, can help to minimize the life time of personal data;
- **Transparency:** As discussed above, privacy policy languages can make it more transparent to users how far a data controller's policy matches their privacy preferences before they disclose personal data (ex ante transparency). Ex post transparency of how data are processed (once they have been disclosed to a data controller) can be enforced by agreeing on obligations that a service provider needs to fulfill. For instance, those obligations may include notifying the data subjects in case that their personal data are accessed or transferred to third parties, and obligations related to the creation of transparency logs, e.g., [40].
- **Technical security:** The XACML access control engine can enforce that the personal data that is disclosed can only be accessed according to the agreed-upon sticky policy. As PPL is based on XACML, the sticky policies can be enforced together with other access control policies by the XACML access control engine.

7.5 Provenance, Reputation and PETs

The notional dissonance between privacy requirements and reputation systems can be partially addressed with PETs. PETs that are designed for reputation systems aim at preserving the privacy of data subjects and/or service providers. In the case of data subjects, PETs aim to prevent third parties to link multiple feedback reports to a data subject—the goal of the third party is to profile which services a data subject uses. In the case of service providers, assuming that a data subject offers a service to other peers, e.g., a carpooling or participatory sensing application, PETs preserve the data subjects privacy by preventing third parties to link reputation values to individuals. In addition, it is fundamental that PETs are designed to thwart attacks against reputation systems, such as *white-washing* [19] and Sybil attacks [14].

To prevent profiling based on recommendation reports, a privacy-enhancing reputation system using role pseudonyms is presented in [33].[5] The proposal is based on self-certified pseudonyms that are valid for a given context or service

[5] Role pseudonyms are pseudonyms that are limited to a specific role or context [37].

and it limits users to have at most one pseudonym per service [1, 32], which prevents Sybil attacks and *white-washing*. In addition, pseudonyms issued for different services are cryptographically unlinkable. Reputation can be transferred between different pseudonyms belonging to a same user using different cloaking techniques, as shown in [9]. Another proposal with the same objective, but based on the homomorphic encryption of the recommendation reports, is described in [42]. It preserves the privacy of the users providing feedback by exchanging and aggregating recommendations under encryption. However, this proposal requires all participants to strictly follow its protocols and it is not robust against misbehaving users.

Privacy-preserving logging schemes can help to determine data provenance and protect users' privacy, such as the ex post TET described in Sect. 7.2.

8 Summary and Open Challenges

In this chapter we discussed privacy in SCIS. SCIS is based on technical concepts such as profiling, provenance and reputation systems, which pose privacy risks, as these techniques allow to track and analyze the users' habits and lifestyle. On the other hand, we also discussed the inherent characteristics of SCIS that can be utilized for a privacy-enhanced system design. While we have pointed out how legal means and PETs can help to protect privacy, still several challenges remain.

Technical challenges to be addressed for promoting privacy-enhanced SCIS in future include: composing peer profiles in a privacy-preserving manner and enforcing privacy-enhancing identity management for audience segregation of peers, utilization of the distributed nature of SCIS for building-in privacy, and combing privacy-preserving logging schemes with data provenance schemes.

References

1. Andersson, C., Kohlweiss, M., Martucci, L.A., Panchenko, A.: A self-certified and Sybil-free framework for secure digital identity domain buildup. In: Information Security Theory and Practices: Smart Devices, Convergence and Next Generation Networks. Proceedings of the 2nd IFIP WG 11.2 International Workshop (WISTP 2008). Lecture Notes in Computer Science (LNCS), vol. 5019, pp. 64–77. Springer, Berlin (2008)
2. Angulo, J., Fischer-Hübner, S., Pulls, T., Wästlund, E.: Towards usable privacy policy display & management for primeLife. Inf. Manag. Comput. Secur. **20**(1), 4–17 (2012)
3. Art. 29 Data Protection Working Party: Advise paper on essential elements of a definition and a provision on profiling within the EU General Data Protection Regulation. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2013/20130513_advice-paper-on-profiling_en.pdf (2013). Accessed 13 May 2013
4. Bonabeau, E.: Decisions 2.0: The power of collective intelligence. MIT Sloan Manag. Rev. **50**(2), 45–52 (2009)

5. Buchegger, S., Schiöberg, D., Vu, L.H., Datta, A.: PeerSoN: P2P social networking - early experiences and insights. In: Proceedings of the 2nd ACM Workshop on Social Network Systems Social Network Systems 2009, co-located with Eurosys 2009, pp. 46–52. Nürnberg, Germany (2009)
6. Buchegger, S., Crowcroft, J., Krishnamurthy, B., Strufe, T.: Decentralized systems for privacy preservation (Dagstuhl Seminar 13062). Dagstuhl Rep. **3**(2), 22–44 (2013). doi:http://dx.doi.org/10.4230/DagRep.3.2.22. http://drops.dagstuhl.de/opus/volltexte/2013/4017
7. Camenisch, J., van Herreweghen, E.: Design and implementation of the idemix anonymous credential system. In: Proceedings of the 9th ACM Conference on Computer and Communications Security, pp. 21–30 (2002)
8. Camenisch, J., Fischer-Hübner, S., Rannenberg, K. (eds.): Privacy and Identity Management for Life. Springer, Berlin (2011)
9. Christin, D., Roßkopf, C., Hollick, M., Martucci, L.A., Kanhere, S.S.: Incognisense: An anonymity-preserving reputation framework for participatory sensing applications. Pervasive Mob. Comput. **9**(3), 353–371 (2013)
10. Council of Europe: Recommendation cm/rec(2010)13 of the committee of ministers to member states on the protection of individuals with regard to automatic processing of personal data in the context of profiling. Available at https://wcd.coe.int/ViewDoc.jsp?id=1710949 (2010)
11. Cutillo, L.A., Molva, R., Strufe, T.: Safebook: Feasibility of transitive cooperation for privacy on a decentralized social network. In: WOWMOM, pp. 1–6. IEEE, New York (2009)
12. Dellarocas, C.: Online reputation systems: How to design one that does what you need. Sloan Manag. Rev. **51**(3), 33–38 (2010)
13. Dingledine, R., Mathewson, N., Syverson, P.F.: Tor: The second-generation onion router. In: USENIX Security Symposium, pp. 303–320. USENIX (2004)
14. Douceur, J.R.: The Sybil Attack. In: Druschel, P., Kaashoek, F., Rowstron, A. (eds.) Peer-to-Peer Systems: Proceedings of the 1st International Peer-to-Peer Systems Workshop (IPTPS), vol. 2429, pp. 251–260. Springer, Berlin (2002)
15. Dwork, C.: Differential privacy: A survey of results. In: Agrawal, M., Du, D.Z., Duan, Z., Li, A. (eds.) TAMC. Lecture Notes in Computer Science, vol. 4978, pp. 1–19. Springer, Berlin (2008)
16. European Commission: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML (1995). Accessed 23 Nov 1995
17. European Commission: Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final 2012/0011 (COD). Available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf (2012). Accessed 25 Jan 2012
18. European Commission: Proposal for a regulation of the European Parliament and of the Council on the protection of individual with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 C7 0025/2012 2012/0011(COD)) Compromise amendments on Articles 1–29. Available at http://www.europarl.europa.eu/meetdocs/2009_2014/documents/libe/dv/comp_am_art_01-29/comp_am_art_01-29en.pdf (2013). Accessed 21 Oct 2013
19. Feldman, M., Chuang, J.: Overcoming free-riding behavior in peer-to-peer systems. SIGecom Exch. **5**(4), 41–50 (2005)
20. Fischer-Hübner, S.: IT-Security and Privacy – Design and Use of Privacy-Enhancing Security Mechanisms. Lecture Notes in Computer Science, vol. 1958. Springer, Berlin (2001)
21. Friedman, A., Schuster, A.: Data mining with differential privacy. In: Rao, B., Krishnapuram, B., Tomkins, A., Yang, Q. (eds.) KDD, pp. 493–502. ACM, New York (2010)
22. Gedik, B., Liu, L.: Protecting location privacy with personalized k-anonymity: Architecture and algorithms. IEEE Trans. Mob. Comput. **7**(1), 1–18 (2008)

23. Giunchiglia, F., Maltese, V., Anderson, S., Miorandi, D.: Towards hybrid and diversity-aware collective adaptive systems. In: Proceedings of FOCAS Workshop on Fundamentals of Collective Systems @ECAL 2013 (2013)
24. Goffman, E.: The Presentation of Self in Everyday Life. Doubleday Anchor Books, Doubleday (1959)
25. Gruteser, M., Grunwald, D.: Anonymous usage of location-based services through spatial and temporal cloaking. In: Proceedings of the 1st International Conference on Mobile Systems, Applications, and Services (MobiSys 2003), pp. 31–42. USENIX (2003)
26. Hedbom, H., Pulls, T., Hjärtquist, P., Lavén, A.: Adding secure transparency logging to the prime core. In: Bezzi, M., Duquenoy, P., Fischer-Hübner, S., Hansen, M., Zhang, G. (eds.) The Future of Identity in the Information Society. Proceedings of the 5th IFIP WG 9.2, 9.6/11.4, 11.6, 11.7/PrimeLife International Summer School, vol. 320, pp. 299–314. Springer, Berlin (2009)
27. Hildebrandt, M.: FIDIS Deliverable D7.12: Behavioural biometric profiling and transparency enhancing tools. Available at http://www.fidis.net/resources/fidis-deliverables/profiling/#c2369 (2009)
28. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proc. Natl. Acad. Sci. **110**(15), 5802–5805 (2013)
29. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity. In: Chirkova, R., Dogac, A., Özsu, M.T., Sellis, T.K. (eds.) ICDE, pp. 106–115. IEEE, New York (2007)
30. Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M.: *L*-diversity: Privacy beyond *k*-anonymity. TKDD **1**(1). Available at http://dl.acm.org/citation.cfm?id=1217302 (2007)
31. Martucci, L.A., Andersson, C., Fischer-Hübner, S.: Chameleon and the identity-anonymity paradox: Anonymity in mobile ad hoc networks. In: Proceedings of the 1st International Workshop on Security (IWSEC 2006), pp. 123–134. Information Processing Society of Japan (IPSJ) (2006)
32. Martucci, L.A., Kohlweiss, M., Andersson, C., Panchenko, A.: Self-certified Sybil-free pseudonyms. In: Proceedings of the 1st ACM Conference on Wireless Network Security (WiSec'08), pp. 154–159. ACM Press, New York (2008)
33. Martucci, L.A., Ries, S., Mühlhäuser, M.: Sybil-free pseudonyms, privacy and trust: Identity management in the internet of services. J. Inf. Process. **19**, 317–331 (2011)
34. Merriam-Webster.com: Profile. Available at http://www.m-w.com/dictionary/profile (2013)
35. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: 30th IEEE Symposium on Security and Privacy, 2009, pp. 173–187. IEEE, New York (2009)
36. Organisation for Economic Cooperation and Development (OECD): Recommendation of the Council concerning Guidelines governing the Protection of Privacy and Transborder Flows of Personal Data (2013) [C(80)58/FINAL, as amended on 11 July 2013 by C(2013)79]
37. Pfitzmann, A., Hansen, M.: A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. Available at http://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.34.pdf (2010). V0.34
38. PrimeLife: PrimeLife – Privacy and Identity Management in Europe for Life: Policy Languages. Available at http://primelife.ercim.eu/images/stories/primer/policylanguage-plb.pdf (2011)
39. Pulls, T.: Privacy-Preserving Transparency-Enhancing Tools. Licentiate Thesis, Karlstad University, p. 57 (2012)
40. Pulls, T., Peeters, R., Wouters, K.: Distributed privacy-preserving transparency logging. In: Proceedings of the 12th Annual ACM Workshop on Privacy in the Electronic Society, WPES 2013, Berlin. ACM, New York (2013)
41. Reiter, M.K., Rubin, A.D.: Crowds: Anonymity for web transactions. ACM Trans. Inf. Syst. Secur. (TISSEC) **1**(1), 66–92 (1998). doi:http://doi.acm.org/10.1145/290163.290168

42. Ries, S., Fischlin, M., Martucci, L.A., Mühlhäuser, M.: Learning whom to trust in a privacy-friendly way. In: Proceedings of the 10th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom 2011), pp. 214–225. IEEE Computer Society, Silver Spring (2011). doi:10.1109/TrustCom.2011.30
43. Sweeney, L.: *k*-anonymity: A model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **10**(5), 557–570 (2002)
44. W3C: Platform for privacy preferences (P3P) project. Available at http://www.w3.org/P3P/ (2006)
45. Warren, S., Brandeis, L.: The right to privacy. Harv. Law Rev. **4**(5) (1890)
46. Wästlund, E., Fischer-Hübner, S.: PrimeLife Deliverable D4.2.2: End user transparency tools: UI prototypes. Available at http://primelife.ercim.eu/ (2010)
47. Westin, A.F.: Privacy and Freedom. Atheneum, New York (1967)

The Future of Social Is Personal: The Potential of the Personal Data Store

Max Van Kleek and Kieron OHara

1 Introduction

A key characteristic common to the various kinds of "social intelligence" described in this volume is one of enhanced autonomy through technological support. Such autonomy allows constituents of a society to form new connections with others dynamically as needed, promoting a more adaptive, flexible and robust social fabric than those of traditional structures, in which efficiency leads a majority to rely on a handful of central, fixed intermediaries. This observation immediately prompts the question of whose interests that "efficiency" is designed to benefit—the intermediaries' or the users'.

While we see technology being applied in many contexts to generalise the benefits and enhance the autonomy thus described, the storage of personal information is one area where it has, thus far, been used to power a perverse reversal towards more centralisation. Currently, a handful of dominant platform vendors and application service providers are grappling for control over individuals' personal information, trying to accumulate as many users as possible in order to maximise understanding of every nook and corner of social interaction—a relentless process satirised in Dave Eggers 2013 novel *The Circle*, about a company with the totalising slogan "All That Happens Must Be Known. This centralising trend, backed by a surveillance-and-analytics business model, began with the rise of so-called "Web 2.0", in which sites became sophisticated apps and content-management platforms designed to facilitate the creation and sharing of user-generated data and content. That content began as a few social network profiles and blog posts, but gradually grew to encompass the entirety of personal data people keep *or generate*, from files and documents to film and music archives. Thus began a migration of personal digital artefacts from

M.V. Kleek (✉) • K. OHara
University of Southampton, University Road, Southampton, SO17 1BJ, UK
e-mail: emax@ecs.soton.ac.uk; kmo@ecs.soton.ac.uk

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_7

individually-administered personal computers into various information spaces of the web. The assimilation of personal data from personal digital devices has accelerated as Web application and service providers have started to create deep integrations with personal computing devices such as Facebook Home,[1] Windows Skydrive[2] and Apple's iCloud.[3] Such services have extended the reach of Web services into the intimate digital spaces of one's personal devices, offering backup and management services for these private data collections as well.

What are the implications of this centralisation? Although the ultimate, long-term implications of this shift are not yet fully understood, several immediate consequences are apparent. Fundamentally, the delegation of responsibility for management of one's personal information to third party service providers necessitates relinquishing control over various aspects of how these data are handled and processed, ranging from how they are stored and represented, to how (and when) they can be accessed, as well as to whom access is granted. When third party delegation accidentally-on-purpose serves the increasingly pervasive business model of deriving revenue directly from these data themselves (through targeted advertising or licensing to third parties), platforms are essentially incentivised to collect from as many individuals as possible, and to create an experience or mechanism that further retains them as long as possible to do as wide a range of things as possible. They are also incentivised to disguise the extent of this delegation, for example by embedding control protocols into complex and legalistic privacy policies whose acceptance is virtually costless (clicking the 'accept' button), binary (yes/no forever) and unconditional, and which are subject to arbitrary change without notice. Platforms get users to disclose as much of their information as possible (to the platforms' benefits) by artificially forcing a trade-off between participation and privacy; in order to enjoy the most basic features of the Web, users have to *give their data away*, thereby sacrificing control over their data and potentially their privacy.

This misalignment of incentives between *what users want to do with their data* and *what platform providers want to do with their data* has the potential to interfere destructively with the development of context-sensitive applications that promise more effective, personalised, behaviourally-adaptive interactions that rely on richer and more sensitive data models, due to either actual or perceived privacy risks entailed. Moreover, the dependency relationships that result from this process place unprecedented power in the hands of these companies, leaving individuals effectively locked in, and unable to switch to alternative providers without greater effort than it is reasonable to expect a privacy-aware non-technical consumer to devote to the problem; the result of this is an overall reduction of autonomy and mobility, potentially ultimately leading to increased fragility, fragmented data spaces and lost or forgotten data [47, 74].

[1] Facebook Home—www.facebook.com/home.

[2] Skydrive—www.microsoft.com/skydrive.

[3] iCloud—www.apple.com/icloud.

While this business model has thus far been hugely successful at creating extremely profitable services from the likes of Facebook, Twitter and Google, the result has been an increasingly fragile ecosystem in which a majority of Web users have come to rely on an oligarchy of service platforms which are in turn amassing a disproportionate quantity of users' personal information. This centralisation, and accompanying power asymmetry, has occurred not just for Web users from the United States, where most of these services are based, but internationally as well, raising concerns pertaining to each country's sovereign rights of access to data of its own versus other nations' citizens, which have been magnified by the information-gathering practices of the US National Security Agency and others revealed by Edward Snowden in 2013. Indeed, thorny issues pertaining to compliance and enforcement of data protection laws across international boundaries [5,12] represent a serious potential risk for this business model, even as the European Union debates a revision to its pre-Web Data Protection Directive. The EUs weak and unsatisfactory 'safe harbor' rule, which allows data sharing with the United States, conveniently diverting attention away from the unsolved problem of differing approaches to privacy and data, looks especially vulnerable—yet where would the cloud be without safe harbor?

However, a basic assumption that powers these dependence relations and underpins the oligarchy is the disparity between the data management capabilities held by the end-users of the Web from those that provide the hosting and storage. In this chapter, we question this "thin client" model of Web computing by examining an alternative approach that places the responsibility of data management back with the users who own it, but in a way that is natural and manageable, while supporting the same social, dynamic interaction flows they are used to on the Web. This set of capabilities we refer to as *personal data stores* (PDSs), the technical goal of which is to augment user computing devices with secure data storage, hosting, and sharing capabilities which can be used to archive and manage valuable information longitudinally, as they interact with one another and third parties respectively.

Our aim in this chapter is to derive the requirements for personal needs for such a platform through insights from the field of Personal Information Management (PIM). To begin with, it is worth reviewing in more detail the dilemmas and asymmetries that current management of "big data" has created, across the public and private sectors, and why the individual is understandably at a loss. Although PDSs cannot conceivably solve or even address all these issues, we should keep them in mind in order to understand the extent to which it makes sense to include PDSs as part of a more equitable longer-term settlement. Second, we present a brief summary of existing platforms being used to manage personal information and their characteristics. The chapter concludes with a discussion of how these platforms may change the socio-economic landscape of the Web, and the ways personal data is shared, collected and handled.

2 The Dilemmas of the Data Economy

Although we would not hazard a guess as to who originated the phrase, we do know that data has been called "the new oil" on many occasions. Of course, the image is intended less as an indication of the deep issues at the core of the data economy, and more as a neat way of conveying excitement in a Powerpoint bullet. However, it *is* indicative, because it can be taken in various, not necessarily exclusive, ways. Oil is a source of great wealth. It is a key factor in many other production and transport processes. It is an essential lubricant. It needs to be mined (well, drilled to be precise) to produce value. It brings great riches to the small number of corporations big enough to exploit it. It raises exchange rates and therefore prices to the detriment of other industries. It has been known to impoverish those whose property is drilled, as elites cream off the main wealth with the help of rapacious corporations and corrupt government. It has, on occasion, led to revolution and the overthrow of *anciens regimes*.

Presumably not all these phenomena associated with the old variety are intended to be predicated of the "new oil." Yet we, as data subjects, presumably want to be sure that we get the good things and not the bad. It is anyway a misleading comparison, because data has properties that oil does not. Data is about people, and can be compromising. Data is generated by people, not by aeons-old trees and animals which have no issues of privacy or dignity. It is not a dwindling resource—we are far from approaching the time of "peak data. It is not a rival good—if I enjoy its use, that does not preclude your exploiting it at the same time. Data becomes valuable when aggregated across communities. Data is a covert way of financing content and services; if the service you receive is free, then you must be the product.

Data is connected to us by an umbilical relation—we generate it in all sorts of ways, and it is about us. We create and provide it; we leave trails of it; it is inferred about us. Yet the flip side of this is that it expresses things we find important (indeed, Luciano Floridi [25] argues that our data are an inalienable part of our identity). By providing a route for others to understand what we are, or what we have done, or where we are situated, it can threaten our privacy, or our dignity, or our autonomy, by diluting the privileged first-person access to our own experience. It creates the possibility of our being counted, measured, judged, steered or influenced without our knowledge by mysterious forces or organisations who may or may not have our best interests at heart. And if the data about us don't exist, we can be profiled, and treated as a standard member of a small demographic, whether this is accurate or not. Maybe this means we get more interesting advertisements—or maybe we will be treated as a potential terrorist and denied access to an aeroplane without adequate explanation.

Because of this, data is regulated. Data protection law is intended to strike a balance between the public good (which may include commercial benefits) of data use, reuse and sharing, and the private good of privacy and individual control. Data subjects have certain rights over their personal data—but not the rights of ownership. If I browse an online bookstore, then I have thereby created a load of data which is of value to someone else. They have constructed a website, and therefore claim ownership rights of the trail—the data results from their investment. In case of dispute, they will cite my consent to their use of my data via some privacy policy that I probably never noticed. It may be argued that I benefit from the collection of this data, because it gives the bookstore sufficient evidence to suggest other books to me that may interest me (and we know, from Amazon's early experience, that a good recommendation algorithm will easily outperform a human recommender). If I were given ownership of my browsing data, then there would be no incentive for the online bookstore to collect it, so it wouldn't be collected, and no-one would benefit from it. If data subjects owned their personal data, then third parties wouldn't bother to collect it, and the data economy would remain a glint in Google's eye.

Data protection is not there to protect privacy; that is at best its secondary purpose. But worse, data protection was a concept developed for the world of standalone databases, not the connected, networked Web with which we are familiar. A tangled skein of legislation struggles to cope with the realities of the personal data economy. Trading personal data goes on at scales previously unimaginable. A user goes online, and literally dozens of organisations will be tracking his or her behaviour. There is talk that this will benefit the data subject, via better devices, better websites and better recommendations; the main 'benefit', arguably, is to become a better target for marketing. I sacrifice my privacy and aspects of my intimate identity for a better class of spam.

Some economists [55] have argued that the release of personal data is a good thing for wider society, as it reduces information asymmetries and enables capital and currency to be allocated in a more informed way. Hmmm, maybe in an ideal world. But arguably the data economy is functioning by ramping *up* the asymmetries—data-using organisations not only know much more about the use they are making of their data than I do, they now in many cases *know more about me* than I do. I, the poor data subject, am sitting at the bottom of so many data asymmetries that the idea that I am too informed for the public good is surely laughable.

Furthermore the concepts of data protection, so valid and timely when they were first introduced, seem at best quaint in 2014. Data minimisation is a great principle, but is it realistic in a world where five billion Google searches and half a billion tweets are generated every day, not to mention the colossal number of mobile phone locations that are logged? Can the use limitation principle be of value in a world where serendipitous reuse is the order of the day? In the data economy, after all, the primary use of data *is* its secondary use. Do notice and informed consent have any meaning in such a world? Do we want to be notified of everything, when our lives are becoming increasingly complex and the choices we already have to make are multiplying? What, when I click the 'yes button, am I consenting *to*? Doesnt virtually everybody treat the 'yes button as an opening to a new and exciting online experience, rather than a notification of the commencement of a complex business relationship which entails a certain amount of risk for the insouciant clicker? Can this truly be glossed as *informed*? One might as well say that the fox is giving informed consent to the hunt when he tunnels under its boundary fence in search of prey.

But how to react to this? We must surely admit the many benefits that data can bring to the subject. Understanding oneself is an important part of managing one’s health or consumption. The benefits accrue not only to the individual, but also wider society. Effective public health, transport and crime management are facilitated by giant quantities of accurate micro-level data. So data sharing with government and businesses cannot be made too difficult to do.

One potential way forward is to move from the current model of data protection, based on regulating the collection of data, for a defined purpose, centred on the data controller, and governed by the consent of the individual. The ‘footprint of the data now stretches far beyond the immediate context and purpose, and regulating for the moment of collection looks anachronistic. A number of commentators, including advocates of ‘big data [39], argue that the time is ripe for a move from subject consent to user accountability. Such a model would regulate the uses of data, and would be centred on the subject who would be given a greater, and less binary, measure of control. For example, Novotny and Spiekermann argue for a three-tier information market, with key distinctions in terms of responsibilities and liabilities between data subjects, service providers, a second tier service space that provides essential support for the top level service relationship, and a tertiary space in which data from the top level relationship is reused on an open but restricted market [46].

What kind of control should the subject be granted? Ownership brings responsibilities as well as rights, and as we noted may mean that potentially valuable data would never be collected at all. Furthermore, many thinkers are nervous that the concept suggests that people’s identities are basically property and commodities, with all the dehumanisation that implies. On the other hand, a human rights approach, for example based on Article 8 of the European Convention, is something of a blunt instrument, and the article is frustratingly vague as to what we actually should do in a particular context. Furthermore, Article 8 is in place and agreed across Europe and many other countries, yet our data is still plundered by the data barons.

In the remainder of this chapter, we will attempt to answer some of these very difficult questions, by tracing a particular idea through conceptual beginnings to a concrete architecture. The next section will consider the notion of personal data or personal information in more detail.

3 Doing Things with Personal Data

Despite the clear importance of the concept, “personal data means different things to different disciplines and communities. In this section, we will consider these different views of data with a view to understanding what capabilities it could afford for subjects, given sufficient access and control. We begin by looking at some of the different definitions, then from the perspective of Personal Information Management look at some of the activities around data and data management, and complete the section by considering the technological support for such activities.

3.1 What Constitutes "Personal Data"? Legal and Operational Definitions

The standard way to conceive personal data is via its legal definition, based on data protection law. This conception has two advantages: first of all it is widely accepted and understood, and secondly it matches the legal liabilities that any PDS management system will need to confront and accept. Personal data, on this definition, is data relating to an identifiable individual. There are a number of issues and indeterminacies here—identifiable by whom? using what methods? in what context?—but these need not detain us here, except to note that they do not make things any easier. The legal definition has not really kept up with technical developments, and it is clear that the ability to identify a data subject is highly context-dependent [43]. 'Personal data is the usual European term, but in the US it can be known as 'personal information or personally identifiable information.

There are strong sanctions against the misuse of personal data without the data subjects consent, but data sharing can still take place if the data controller de-identifies the data by removing identifiers from it or aggregating it (whereupon the new dataset is no longer personal data). There are many techniques for this [77], and there is also a major and unresolved debate [14, 43, 50] about whether de-identified data can be made re-identifiable by cross-referencing it with other datasets, using so-called 'jigsaw identification methods. For instance, the information that a girl in a dataset is pregnant is not identifying, and therefore not personal data, but combined with the information that Mary Jones is the only girl in the dataset, clearly a possibly unwelcome inference can be drawn about the all-too-identifiable Mary Jones. In this chapter, we will not consider the issues raised by such technicalities in detail, except to note that (a) they impinge on data sharing practices and may impose complex liabilities that will be hard to predict, and (b) they can be side-stepped in many cases if the data is exposed by the data subject, who can therefore be assumed to have given consent for use of that personal data by others, given that he or she made the decision to share it in the first place. In the context of giving data subjects greater control over the data that is about them, this is clearly a vital factor to consider.

If we now move from the legal definition, and consider this latter context, an alternative understanding emerges of personal information as the information over which a person has some interest or control, in order to negotiate their environment or order their lives (so, distinct from data in which a person has a commercial interest only). This type of personal information or data is much more in tune with an intuitive understanding of what data means to *me*. And as one would expect, it would include a great deal of crossover with the legal definition of the data from which I am identifiable, but it is likely to include data of which I am the owner, but from which I could not be identified at all (e.g. photographs I have taken, from which it may even be possible that other people might be identifiable, and hence which might be personal data with respect to those people).

The uses to which such data may be put might be social or entertainment, or could be work-related, consumption-related, or administrative; it might also have

Table 1 *Categories of Personal Information*—Jones's proposed taxonomy of personal information [33]

Category	Examples
1. Owned/controlled by me	E.g., Email, files on our computers
2. About me	E.g., my credit/medical history, web history
3. Directed towards me	E.g., phone calls, drop ins, adverts, popups
4. Sent (provided) by me	E.g., Emails, tweets, published reports
5. Experienced by me	E.g., Pages, papers, articles Ive read
6. Relevant (useful) to me	E.g., Somewhere "out there" is the perfect vacation, house, job, lifelong mate

no obvious immediate use, but be stored in case it should have value later on. The data may come from several sources: it could be self-generated, deliberately created, a by-product of other kinds of activity, shared with friends or colleagues, open data from the Web, or have been officially bought or licensed from the (legal) owner. Therefore the data in which a person has an interest will almost certainly be of various types of legal status. Personal information in this sense has been investigated by researchers in Personal Information Management (PIM), and we can draw on some of their insights.

The task of identifying all of the kinds of data a person might need to keep, manage and use is a complex and not easily scoped task. Researchers in PIM have derived various working definitions of *personal information* in order to effectively scope their field of study, and have made progress towards potential functional classifications for kinds of personal information. One such classification by Jones et al. [33] is visible in Table 1.

Jones takes an approach that distinguishes among different kinds of information by how it relates to the individual in question; whether the individual experienced it, kept it, sent it, or received it, or whether this information refers to the individual or his or her activities. The categories *About me* and *Relevant to me* are controversial because these definitions do not require individuals to be aware of the existence of the information; it thus establishes a sphere that goes beyond the scope of information experienced by the user. We discuss the potential implications of including such information within the scope of PDSes in *attentional challenges*.

3.2 Activities Around Personal Information

Each person can access, use and manage information in many different ways in their everyday activities. Moreover, there is considerable variation among the ways that different people manage their information, as documented in studies of people's office and home information environments predating personal computers altogether [37]. As a result, it has been relatively difficult to come up with a single

Table 2 *Categories of PIM activities*—Table comparing Jones's 3-tiered categorisation of information activities [33] versus that proposed by Whittaker [73]

Jones [33], Jones and Teevan [66]	Whittaker et al. [73]
(Re-)Finding	
Keeping	Keeping
Meta-level activities (managing, maintaining)	Management
	Exploitation

characterisation encompassing all of these activities; several classifications have been proposed. Returning to the PIM literature, Jones et al. propose a categorisation centering about a distinction between finding, keeping, and a set of "M-level activities", which encompasses managing and organising information archives (Table 2) [33]. Whittaker et al's slightly different categorisation, meanwhile, simply identifies 3 classes: keeping, management, and what they call "exploitation", as follows:

Jones's classification introduces *finding* as a primary activity that people perform; his definition spans a set of common behaviours including discovery [16], information foraging [53], orienteering [9, 65], searching [68] among other related behaviours in which people purposefully seek information or serendipitously encounter it in the course of other information activities. Once this information is found, information is either consumed and internalised, or kept in an external archive, or both, and this process of saving information externally is referred to as *keeping*. Beyond this activity of archiving, individuals might return to their archives to organise, update, or trim them; such activities are referred to as the *M-level*, for manifold meta- and management, hence *M-level*, activities. Whittaker then includes a fourth category of behaviours, *exploitation*, referring to the set of ways in which the information is used and applied.

Among such uses, while the foremost might be to *inform* an individual making a decision, many other uses of information also exist. For example, information might be created for the explicit purpose of *reminding* a person of past or future events, activities or details. Other purposes might be to *measure* and keep track of the time-evolution of some phenomenon so that it can be easily understood. When this measurement is about the individual's own activities, the purpose might be for providing *feedback*, which may be vital for behavioural modification domains such as cognitive behavioural therapy (CBT)-like programmes. This feedback may, in turn, along with other information, collectively serve to *motivate* further activity or behaviour. Finally, information may serve the purpose of *external cognition*, in which information is created or manipulated for the purpose of facilitating *understanding* or *problem solving*. This set of activities is often referred to as *sensemaking* [54].

3.3 Supporting Information Activities

Technological support for each of these information activities has demonstrated the potential to change not only how they are conducted, but the contexts in which they are applied. One salient example is that of Web search engines, originally created for Web page information retrieval, but which have become a nearly ubiquitous tool for accomplishing tasks across a much broader variety of activities, spanning both desktop and mobile. Another area is in supporting longitudinal curation; tools that automatically perform off-site, incremental, and continuous backup such as Apple's *Time Machine*[4] have become commonplace, allowing end-users to make their stored data more resilient to accidental deletion or data loss.

Yet technological support for most of the other aforementioned personal information activities, including reminding, sensemaking, discovery and orienteering, has remained rudimentary. Reminding in PIM tools, for example, has until only recently been limited to clock/calendar-based alarms that need to be explicitly set for a specific date and time, despite the rich variety of "off-line" strategies people have naturally adopted for their own uses [8]. While the basic calendar alarm remains heavily used, its precision, brittleness and intrusiveness have been documented to undermine effectiveness, sometimes through extended "snooze wars", in which users repeatedly dismiss alarms, resulting in their piling up over time. The alarm can end up a burdensome annoyance, instead of providing the intended assistance.

The mismatch between peoples data management requirements and the technology to support it is not, of course, restricted to PIM. As another example where the promise has not been borne out, Privacy Enhancing Technologies (PETs) [72] have yet to make a mark either. They too have failed to transcend the perennial problem of demanding an investment of time and resources that few want to make, or want to have to make. They also put a relatively inflexible barrier between individuals and organisations, while the individual may in fact have very context-dependent requirements (it is handy, for example, for an online fashion company to know my size, even if I do not want this parameter value bruited abroad). Take-up has been predictably anaemic.

4 Personal Data Stores

Yet surely technology must be part of the solution to a technologically-driven problem. Technology creates data, with the connivance of the data subject, and tools have emerged for large-scale players to exploit their vast datastores. The concept we wish to explore in this paper, in response to the foregoing discussion of the challenges and context, is that of the Personal Data Store (PDS). This is a locus

[4] Time Machine—www.apple.com/uk/support/timemachine/.

of control which leaves open a number of the key questions about ownership and property, while giving power to data subjects. Our aim in this chapter is to set out some of the possibilities of PDSs, and to try to show that at least some of the above dilemmas can be addressed with them. Clearly PDSs will not be the full story—but they should be part of the solution. We hope to suggest some ways this could happen, and how indeed it *has* happened, and to illuminate the potential by refining our account to produce a specific example of a PDS architecture.

The aim of PDSes is to start to narrow the aforementioned data inequality by bolstering the capabilities of individuals for managing, curating, sharing and using data themselves and for their own benefit. The idea is not for such capabilities to replace services, nor for individuals to take their data out of the rich ecosystems that exist today (a feat which would be practically impossible, not to mention potentially destructive), but instead to enable people to collect, maintain and effectively derive value from their own data collections directly on the device(s) under their control. The combination of such capabilities and derived value provides an incentive for individuals to take responsibility for, and invest effort in, the preservation and curation of their data collections, turning to external third parties for specialised services only where needed. The aim of such development would be to try to restore some balance by providing a locus for subject-centric management of data, to complement (and in some cases replace) the current paradigm of organisation-centric data management.

Arriving at an operational definition, we define PDSes as follows:

> A personal data store is a set of capabilities built into a software platform or service that allows an individual to manage and maintain his or her digital information, artefacts and assets, longitudinally and self-sufficiently, so it may be used practically when and where it can for the individuals benefit as perceived by the individual, and shared with others directly, without relying on external third parties.

This description leaves undefined the kinds of activities that might constitute “managing”, “maintaining”, “controlling fully” or “using” this information, nor even what kind(s) of information, owned by whom, that we are talking about. Fortunately, significant insight pertaining to many ways individuals readily use information (in both on-line and off-line contexts) has been gained through studies conducted at the intersection of psychology and computer science, particularly the Human-Computer Interaction (HCI) research community. Beyond insights about existing information practices, various ideas have been proposed dating back nearly a century about how technology might change human-information and human-human relationship, modulated by new emerging information technology.

4.1 Historical Reflections from Memex ...

The genesis of an individual-centric personal data archive pre-dates digital computers entirely, to Vannevar Bush’s Memex vision of 1945 [11], which proposed a mechanical framework for supporting the collection, archiving, and organisation of information to facilitate later cross-reference and retrieval. Among the important contributions of this article was the significant emphasis on reducing the effort needed to capture and retrieve information, due effort being the primary impediment towards effective and frequent information use. To this end, Memex proposed that individuals could wear capture devices on their bodies (a camera strapped to the forehead), store such information compactly, conveniently and indefinitely, and retrieve it later through an associative mechanism modelled upon the human memory, queried naturally via gesture.

Two additional early projects that explored how such information archives might be realised were Ted Nelsons Xanadu [44] and Douglas Engelbart's NLS [21]. Both proposed that information environments could be interlinked through a global network of knowledge sharing, demonstrating many ideas in the 1960s that would not be realised in commercial systems for decades. While the former focused on hypertext and distributed collaboration, the latter focused on structured data collections, including data navigation, creation and management. Engelbart demonstrated an actual prototype of NLS in 1969, capable of synchronous collaboration, complete through a graphical user interface, that incorporated dynamic hierarchies, hyperlinks, and multi-view representations

The introduction of the personal computer (PC) in 1984 provoked the development of the first generation of digital personal information management tools, consisting of a variety of application software products designed to help individuals create and maintain collections of digital data, ranging from flexible, schema-agnostic personal database systems like Filemaker,[5] to specific data types, such as digital calendaring tools, and "digital Rolodex" address books. Seeking to appeal to the first generation of personal computer users, many of these applications borrowed metaphors from paper-based information collection tools, from the notion of "documents", to that of files and folders, and even notebook ledgers and personal diaries. Along with this deliberate shaping of digital information into forms designed to be familiar with paper information organiser came interaction metaphors and organisation methods for them; from deletion of information by "throwing in the rubbish bin" to "desktop" and "filing cabinet"-based information organisation and arrangement.

Meanwhile, research in personal information management continued to pursue the vision put forth by Memex, towards methods of automatically building archives of personal life activities and experiences, so that these might be used as external memory prostheses. The pursuit of this vision was partially responsible for the development of handheld and early wearable computing technology, such as the Xerox PARC Tab [60], arguably the first hand-held computer, which ran arguably the first automatic location-based personal lifelog, PEPYS [45]. Many systems that captured other aspects of context and activities soon followed, such as the Remembrance Agent by Rhodes et al., and the life archive by Clarkson et al., both at the MIT Media Lab's "Cyborg" Wearable Computing group. Since the breadth

[5] Filemaker—www.filemaker.com.

of kinds of activities and experiences that such systems captured transcended paper documents, such research required re-thinking the shape of data away from paper-metaphors to other kinds of collections, including *information streams* (e.g., Lifestreams [24]) and chronological *lifelogs*, such as MyLifeBits [27].

The third, and potentially most profound, transformation of digital information tools occurred with Web 2.0, the rise of a "social Web" replete with dedicated apps and services for managing and sharing nearly any kind of previously imagined personal information, ranging from the sensitive and intimate to the public.

Meanwhile, the data proliferated too. Seeking to monetise the flood of information people were putting online, markets for personal information quickly began to emerge, prompting concerns over privacy, security, and rights of access, which in turn have driven government and regulators interest towards giving citizens more protection over various aspects of how data about them could be collected and handled. This led to international efforts to craft data protection legislation, as discussed above. In terms of the provision of data to individuals, such legislation so far has focused on allowing data subjects to inspect the data an organisation holds about them; on receiving a subject access request, the organisation is obliged to correct inaccuracies, and to respect requirements that the data is not used in any way which may cause damage or distress, and that the data is not used for direct marketing purposes.

However, this is a fairly minimal power which is hardly congruent with the increasing clamour concerning rights to data, including the spread of enforced transparency of data from the private sector [26] and the vogue for freedom of public sector information [49], and technology (and technology policy) together with new attitudes to transparency bring more possibilities. In the UK, a government initiative called *midata* [62] is working to bring about the logical next step of customers getting direct and unfettered access to data kept about them by companies (other similar initiatives include the US Blue Button initiative[6] and the French Mesinfos group[7]). The ultimate success of *midata* will be contingent on several important steps in both technology and regulation, most particularly including realising effective tools such as personal data stores for letting individual users easily consume, consolidate and make use of this data once it is made available.

4.2 . . . To Mydex: Birth of the PDS Concept

Independent of such legislative approaches, both academic and industry-led efforts also began to commit resources to research towards identifying ways that end-user citizens might, in the face of the vast growing repositories of data being held about them, enjoy more control and privacy. An academic consortium known as *Vendor*

[6] www4.va.gov/bluebutton/.

[7] mesinfos.fing.org/.

Relationship Management (VRM) at Harvard's Berkman Center was realised to conduct multifaceted research into socio-legal-econo-technical approaches that might be employed. Among the products of this research was a vision that users might stand as their own information brokers, and start to act as peers with service providers, capable of negotiating fair and equitable mutual terms of data use during interactions with them [1]. Out of this work emerged the earliest mentions of Personal Data Stores for realising such capabilities in the context of online e-commerce, inspiring more than a dozen different Personal Data Store offerings, platforms and services backed by commercial start-ups since 2001 [67].

As an example, consider Mydex, whose proof-of-concept offering dates back to 2009 [31]. Mydex designers worked with data-handling organisations to develop systems to support data transfer and sharing governed by consent and identity verification. Design principles included putting the individual PDS owner in sole charge of consent giving and revocation with a simple 'on/off switch; giving the individual sole access to the private encryption key; verification of all organisations wishing access to data; and comprehensive data sharing agreements going beyond Data Protection Act protections. The business model for Mydex is still experimental, but currently the idea is to fund the stores by charging organisations for access to data; if the charge is set low enough, then they should save by side-stepping other access costs (e.g. the costs of writing a letter to the data subject). The Mydex services are currently free of charge to the individual. Mydex exploits cloud infrastructure with open source software, but its PDSs are discrete collections of files encrypted and controlled by the individual, including—and this seems prescient after the Snowden revelations[8]—the ability to choose the location of the data centre in which the PDS is stored. Similar open source personal data storage containers include The Locker Project,[9] data.fm,[10] Owncloud,[11] and OpenStack,[12] each of which provides various degrees of easy-to-set-up 'personal cloud software that can be used to store and host content on the user's own server on the Web.

A consistent theme of commentary in this area has seen Personal Data Stores (PDS) as important, if not essential, capability for end-users towards growing a healthier global "personal data ecosystem". For example, an independent study commissioned by The World Economic Forum documented ways that the value of personal data might be further "unlocked", citing Personal Data Stores as a core enabling mechanism to turn end-users from consumers into more autonomous data brokers [10]. A separate comprehensive analysis by *Ctrl-Shift* on emerging commercial PDS platforms and offerings projected an enormous economic opportunity for PDS services in the next 5 years [67]. In their view, PDSs are the key to making

[8] www.theguardian.com/world/the-nsa-files, www.ub.uio.no/fag/informatikk-matematikk/informatikk/faglig/bibliografier/no21984.html.

[9] lockerproject.org.

[10] data.fm.

[11] owncloud.org.

[12] www.openstack.org.

sense of the myriad data sources that now surround us, from data we volunteer, to the data that commemorates observations of our behaviour, to the data inferred about us, combined with the data we generate via management of our personal affairs (e.g. in health or finance), and also bringing in data about our activities as customers or consumers, including our contributions to loyalty card schemes.

4.3 Failure to Launch: Barriers to PDS Adoption

Yet despite the extensive needs analysis and market potential identified, early personal data store offerings have thus far failed to attract substantial attention from users. While a number of factors are likely responsible, so the lack of interest among users has been attributed to the fact that many of initial PDS platforms have sought to simply re-create existing end-user experiences offered by popular apps and Web platforms, rather than creating new functionality. Despite the benefit that these PDS offerings provide in terms of data security, users are often less compelled to try something new if the tangible experience nothing new, while data security remains an abstract, inestimable threat which does not necessarily easily compel behaviour change [4]. Finally, since the very purpose of PDS offerings is to protect user data from third party access, these platforms cannot derive revenue from user data and must resort to subscription models—always less attractive to new users than offerings that are completely free to use.

On top of these suppressors of the positive impulse to manage data, we must also remember that the markets work pretty well for some (the most powerful) operators, and so there is a great deal of inertia around. A dogmatic view of revealed preferences of course suggests that individuals lack of interest in the technology shows they have no desire to curate their own data. They happily click on privacy policies they have never read, and they buy the goods that are marketed to them, at least in sufficient quantities to justify the marketers costs. ‘Push models seem to be in the ascendant, because the data oligarchs are the only agents with access to the bigger picture of what data is held about you, what can be inferred from that data, what services are available, and how you relate to the general data context. ‘Pull models struggle, because individuals cannot see the opportunities that are around. In short, the argument is often made that the technological direction of travel is more or less set, that it serves the public good, that the public is uninterested in any alternative, and so, to coin a phrase, “get over it. This deterministic model has been called Zuckerbollocks [48], and it is important to challenge and resist it.

Heath et al. write [31] that “there is market evidence that [the person-centric model of control over personal data] is starting to establish itself, but even they see a challenge to getting the model to work. Three conditions need to obtain simultaneously, on the account of Heath et al: PDSs must (a) make life simpler/better for the individual, (b) appeal to data consumers by solving some of their problems (e.g. costs, or legal liability), and (c) solve some pressing challenge that is holding back developers and entrepreneurs in this space. To these three, we can add a fourth, which is to rejig current data protection thinking. At the moment (2014), there are three key roles in the standard model of data protection: the data subject, the data controller and the data processor. The owner of a PDS is none of these (or none exclusively—he or she is likely to be all three at various times), and it is hard to see how individuals can exercise autonomous control over the data that affects them without some recognition of them as active agents in a different kind of role. Furthermore, data protection legislation is intended to cover cases of personal data being misused by others; it does not cover cases where individuals accidentally (or deliberately) identify themselves. Of course, this is a reasonable starting point for protection, but if it is the only principle, it means that if an individual 'takes charge of his or her data, he or she *loses* the cover of Data Protection Acts.

5 Six Not So Easy Pieces: Challenges Towards Realising the PDS Vision

The goal of providing individuals with the capacity to maintain their own information longitudinally imposes a number of challenges to supporting the kinds of information activities we have described. In particular, we see six broad categories of challenge to be met; the first, most fundamental of which pertains to effective *longitudinal keeping*. Enabling individuals to keep their data safely for a long time, while ensuring its continued accessibility and usefulness impacts both the data formats and methods used to store them. For example, since a person's physical computational hardware is likely to fail with age, methods need to be in place for ensuring robustness to such failures, such as multi-device replication and easy migration from older to new devices over time. Moreover, as evidenced by Moore's law [58], since the technical capabilities and properties of such data storage devices and platforms are likely to change fundamentally, PDSes must be designed to accommodate (and take advantage of) such changes as they arise. The devices and technologies that have made the PDS vision possible date back only a couple of decades, whereas a safe haven for data such as we are envisaging might well have to last a working lifetime (before we even consider the issues surrounding inheritance of data after a death).

A second challenge is allowing individuals who might have little or no experience in the intricacies of data management to cope with the burden of data security and longitudinal maintenance. Using current tools and services, for example, managing your data yourself still means taking pains to ensure that one's personal data is not lost to hardware and software failure, malicious attacks, or safely migrated to new platforms and devices; such efforts require vast investments of time, effort and expertise. A general lack of expertise or willingness to do this means that people currently rarely know how, or bother, to back up or consolidate their data. Thus it is no surprise that individuals have been motivated to outsource maintenance of their data to third parties, such as cloud providers. In order to facilitate autonomy from such services, therefore, PDSes must seek to support directly, and automate where possible, tedious data maintenance tasks that have plagued PC users for decades. Such automation could both ensure compliance for promoting data security and integrity, such as continuous backup regimes, thereby countering recent studies of the extremely low compliance of personal data backup and security maintenance practices [15, 28].

A separate set of challenges arises from the shift back from service-provider controlled data storage to a user-centered model of data management. Although this will re-empower users to control the organisation of their data spaces, and eliminate the pervasive problem of data fragmentation [30, 34], the challenge with the increased flexibility that this approach affords is that it requires re-consideration of how third-party applications and services can interact with such data, which have traditionally been pre-defined to operate on a fixed, typically application-provider established, set of data representation(s) and manipulations. In a consolidated, user-centric data model, on the other hand, such representations may be specified or modified by the individual, or by some other third-party application(s) on behalf of them, and thus applications themselves must be designed to accommodate such variability among representations.

The need to comply with local, national and international data handling requirements pose a fourth set of challenges. In particular, if PDSes are to support the storage of identifiable information, or more critically, regulated sensitive information such as individuals medical records, then PDSes must implement a variety of security standards (e.g. [40]) to ensure secured storage. Perhaps more difficult might be achieving compliance with the additional data handling requirements imposed by these regulations beyond how it is stored and encrypted; in particular, key handling requirements and guaranteeing aspects of physical access to the machine(s). The integrity of data must also be secured—for instance, although a patient should have the right to challenge and correct inaccurate medical data, if the PDS is to store a version of medical data that is likely to be used (for example, in support of medical treatment in a foreign country), the data would need not only to be accurate, but also of appropriate provenance in order to be properly adapted to the standard workflows of medical treatment.

Even if PDSes were to achieve all of the aforementioned goals, individuals would still face the fact that service providers would inevitably continue to profile and amass information about them, as long as it aligned with their incentives to do so (and it is hard to imagine that it will not—for instance, a service provider may need to gather a large amount of personal data in order to ensure correct and appropriate billing for its services). Thus, if PDSes are to give users the degree of autonomy and independence from profiling, they would need to include privacy-enhancing technologies, such as IP anonymisers, user-agent randomisation and cookie blocking. This may be difficult or impossible to do on "closed" platforms such as iOS that prevent these techniques because they are perceived as "hacking.

Perhaps the ultimate set of challenges, however, pertain to accommodating change as it affects both the information itself and the practices and activities surrounding it, over the years that a PDSes is intended to operate. Technologies that bring in new ways that data is used and generated seem to be introduced every quarter, placing new demands how this information needs to be accessed, created and used. The most recent examples include wearable computing and "always on" wearable sensor technology, from simple devices such as Fitbits[13] and Fuelbands[14] that unobtrusively but nearly constantly measure simple aspects of an individual's activity, to complex computational devices that can both deliver and capture information in high fidelity and quantity anywhere, such as Google Glass.[15] Such devices, as well as innovative new apps, can, in some cases, bring about changes in norms pertaining to people's activities, including the ways people think about technologies themselves.

Looking forward at some of the ways such technologies might impact information activities, some have looked at the possible consequences and implications that ever-increasing information capture and access might have on the kinds of activities mentioned above. While Bell and Gemmel have argued [7] that such increased capture and access could create near-perfect records of our daily lives, allowing people to examine with unprecedented scrutiny their everyday activities, others such as Mayer-Schonberger have argued that such a utopian views overlooks a great number of potential unintended consequences [39].

The difficulties that this community has encountered have led us to reconsider, from the ground up, the need(s) these platforms are meant to address, so that they can be used to design a platform that will fulfill needs beyond secure data storage, towards new applications that promote the more effective use of data in both personal and social contexts.

6 Survey of Online Data Platforms and Services

Given this characterisation of the various kinds of *personal data* and activities around it, we can identify the ways that current online services fulfil the needs towards people's information types and activities.

Table 3 lists the top five personal data cloud platforms by number of users. While Facebook may not be considered an end-user personal data storage provider of the likes of Dropbox, it remains one of the world's largest brokers of personal information. Of particular interest is the introduction of its Timeline feature in December 2011, which took the format of a visual chronological lifelog starting at the individuals birth. In order to compel users to backfill information about their lives into their Timelines from before they joined Facebook, the platform introduced prompting questions, asking for information such as all of the places one has lived, ones family members and favourite activities. Somewhat surprisingly, the negative

[13] Fitbits—www.fitbit.com.

[14] Nike+ Fuelband—www.nike.com/fuelband.

[15] Google Glass—www.google.com/glass.

Table 3 *Most popular commercial cloud data storage providers*—Most popular service-centric data storage providers in 2014, listed with descriptions of kinds of user data managed

Facebook	Profile incl. Timeline; Friends; Events; Group memberships; Biographical history; States favourites; Preferences; Message archives; Liked pages, images, products; Places visited
Google Drive	Any files; Google Docs; calendar; G+ profile; identify and profiles of friends; search history; page access history; bookmarks; locations visited
iCloud	iWork Documents, Photos, Calendars, Passwords (Keychain)
Dropbox	Any files
Skydrive	Office Documents; Any files

"backlash surrounding Timeline has been predominantly surrounding its aesthetic and usability [51], rather than its privacy intrusiveness, except among a small but vocal minority [18]. This aggressive strategy, however, has successfully driven millions of individuals to divulge rich histories of their lives with unprecedented fidelity.

Facebook only supports the storage of very specific information forms, spanning status updates, likes, photos, messages to individuals and so forth. Among the remaining services, Google Drive, Dropbox and Skydrive support general file storage, with the former two providing full versioning history support, while Skydrive providing versioning only for MS Office documents with a full paid Skydrive Pro membership.[16] iCloud, meanwhile, in a move congruent to their push for their mobile devices to render user-visible filesystems obsolete, does not support the general storage of files.

A survey of why people used the file-oriented storage services revealed that while backup had previously been the main reason for using online cloud services, multi-device access and sharing/collaboration have quickly eclipsed backup for reasons people use such services online [70]. The primary use of Facebook, meanwhile is to stay connected with others, as well as several emotional reasons, spanning reasons of self-actualisation and to fulfill the need to belong [42].

However, these services primarily pertain to the management of a fraction of the personal data encompassed by Jones's definition above, specifically "data owned/controlled by me". If we also extend consideration to online services that host and collect "data about me" as well, there are now an increasing number of sensor-driven apps and services that facilitate the tracking of various, routine aspects of everyday life activities, spanning purchases, movements, wellbeing vital statistics; we list such life tracking sites in Table 4.

While both categories of services broker significant amounts of data, these do not generally meet the requirements for personal data stores, as service providers ultimately remain in control how this data is stored, secured, and have full access

[16] Skydrive Pro- http://office.microsoft.com/en-001/office365-sharepoint-online-enterprise-help/manage-document-versions-in-skydrive-pro-HA103158256.aspx.

Table 4 *Activity and life trackers*—Popular web “lifelogging services that facilitate the capture and logging of everyday life experiences

Service	Description	Logging Method
Foursquare; FB Places	Visits made to points of interest	Manual check-ins
Moves	Complete history of a person’s movements throughout the day as recorded from smartphone app	Sensed via mobile
Mint; BUDGT	Access to personal banking records (tracking spending)	Sensed
Withings	Weight, blood pressure,	Sensed
Garmin HRM; Polar HRM; Cardiio App	Heart rate over time	Sensed
Zensorium Tinke	Blood volume pulse, stress markers	Manual measurement
Fitbit; Fuelband; Jawbone	Daily activity levels	Sensed
Wattvision; Stepgreen	Energy consumption	Sensed
Moodpanda; Mappiness; Gotafeeling	Mood/emotion/stress	Experience Sampled
CalorieCounter; Fooducate	Daily calorie consumption	Manual

Table 5 *Early PDS Offerings*—Personal Data Store offerings which encrypt data to provide a high degree of user data security, e.g., only the user has access

Personal.com	Cloud svc for keeping important structured data of specific schema types (passwords, contact details)
Mydex	Cloud svc centered around specific structured data and identity verification

to its contents. Other services, meanwhile have been launched with their primary offering centered on end-user privacy and control; such services (examples of which are listed in Table 5) are sometimes referred to as the first generation of “personal data store” offerings.

These offerings protect user data through encryption and by adhering to data handling standards; however, user data still physically reside in data centres operated by these service providers, where they ultimately remain under their control. Similarly, these services thus far are both highly restrictive on the kinds of information they are designed to manage, with support for a handful of different information types in specific schemas.

A different approach that embraces a “DIY model [56] for PDSes are software packages that people can install on their hardware devices of choice, in order to provide data management and security capabilities. An example of such software packages are Table 6. While *aerofs* and *bittorrent sync* are proprietary commercial software packages, the remaining are released under libre, open source licenses [71].

The open source model provides a number of advantages in terms of realising PDSes. First, these licenses allow these systems to be appropriated, in whole or in parts, and mixed with other systems, in order to construct bespoke PDS

Table 6 *Self hosted data management software*—A sampling of commercial and FOSS software designed to facilitate management of personal information

aerofs	Commercial solution for self-hosting a centralised dropbox-like service
bittorrent sync	Commercial peer to peer file synchronisation software for personal computers
gitannex	FOSS Distributed file metadata maintenance system for advanced users
cosicloud	FOSS self-hosted cloud platform for plug computers offering mail, photo, contact and metadata hosting and storage
data.fm	FOSS RDF-based Web data store with linked data support

functionalities in any ways seen fit. It also encourages transparency by allowing anyone (and everyone) to consult, verify and improve its code, while at the same time making it difficult to hide malicious code within them, such a malware.

It is worth noting that while these DIY PDS platforms are largely platform and hardware agnostic, there are a number of hardware personal data archiving solutions for personal use, ranging from simple external hard drives, automatic-backup solutions that provide version histories, such as Apples Time Capsule,[17] NAS storage devices (e.g. WD MyClouds[18]), to systems that provide data resilience, access control and some degree of data security such as *Drobo*.[19]

In terms of support for the kinds of aforementioned PIM activities, these examples demonstrate that the majority of personal data services have thus far focused on prioritising data durability and multi-device data access. Beyond data backup, a few provide full data versioning, and some offer data security guarantees as well.

Sharing is another kind of support that is central to all of these services, reflecting their common roots in social Web 2.0 services. This comes in terms of real-time collaboration for some (e.g.,, Google Drive and Skydrive), while nearly all of the cloud and DIY platforms above provide some support for asynchronous collaboration, including disconnected operation (e.g., Dropbox Google Drive, Skydrive, gitannex).

The specific approaches taken to supporting PIM activities also vary considerably; some have bundled application front-ends, or sets of application utilities"baked into them (such as Google Drive and WD MyCloud), while others simply function as generic storage containers for existing applications (e.g., Dropbox and gitannex). Still others stand as their own platforms for future PIM apps and services (e.g. cosicloud). Two trends are clear, however; first, that the commercial centralised cloud offerings currently outpace the self-hosted options in terms of features and lowest immediate visible cost, both in terms of subscription costs (due to the pervasive freemium models) and in terms of time and effort to set up and use. In terms of long-term costs, however, the advantages of provider-hosting are less clear; the DIY approaches promise much greater flexibility by facilitating the creation

[17] Apple Airport Time Capsule—www.apple.com/airport-time-capsule/.

[18] WD MyCloud www.wdc.com/en/products/network/networkstorage/.

[19] Drobo—www.drobo.com.

of re-appropriable, custom-tailorable PIM solutions. Moreover, since self-hosted solutions place the ultimate responsibility on the user for the maintenance of his or her data, they provide much greater potential for long term data durability and security.

Based on this perceived disparity between cloud-hosted and DIY solutions, we organised an open source community effort around identifying and realising advanced PIM support in self-hosted, DIY platforms. The goal of this effort was to identify how to realise a system that would overcome the barriers to using self-hosted platforms while leveraging its benefits; specifically supporting research questions on how to better support users long-term data retention and management needs.

7 INDX: A Research Programme Around Personal Data Stores

The substantial challenges just described towards realising an actual PDS platform that achieves the goal set out in the introduction makes deriving a requirements specification daunting. Such a specification would require a well-defined and limited set of capabilities, provided in sufficient detail to be realised in a software (or software-hardware system). Yet, it is not clear how such a set of capabilities (out of many) should be chosen, nor how to choose a such a set to satisfy the requirement of minimality (to avoid overspecification). Nor, finally, is it entirely clear how to verify whether any such set could reach its intended goal.

Therefore we believe a research-centric, rather than development-driven, approach may be the most suitable for bridging the gap between the high-level challenges discussed and the evaluation of potential solutions. Towards this end, we have begun a research project centred about a set of core questions for investigation, and an open experimental research PDS platform called INDX.[20]

The purpose of INDX and the research efforts around it, are several; from a research coordination perspective, it aims to serve as a common ground where various research communities may identify interrelated issues. This is a particularly critical role, as the kinds of work emerging from usable security, privacy, data durability, decentralised social systems, could both be informed by, and used to inform others about how approaches might fit into an integrated picture of future information management systems.

The second role is to serve as a base platform upon which various PDS technical and interface experiments can be tested in a real world setting. To this end, INDX will provide a basic implementation of what one might consider the most elementary kinds of services that PDSes are likely to need. We outline the specific such functionality in the next section. The reason that a complete, open implementation

[20] INDX source code and distributions—http://indx.es.

of a basic set of components is necessary for evaluation is to provide essential functionality to enable PDS researchers to focus on particular problems one at a time, rather than having to re-implement these basic components per experiment.

The third, and perhaps most critical reason for INDX is that a concrete implementation is necessary to even start to interrogate many of the goals pertaining to how the systems might be used by individuals. A deployable implementation of a PDS architecture opens up the possibility of running field experiments, which can be vital to understanding how individuals might perceive or adopt functionality in actual use. Just as the social mechanisms of the Web could not be effectively studied until years after it was built (and continues to evolve), the various interface and interaction mechanisms of PDSes may set off different usage(s) that would altogether be difficult to anticipate prior to deployment. Such is particularly important for personal information management practices, which have been shown to be highly slippery and idiosyncratic; people appropriate and change the ways they use the tools in their collections in unexpected, creative ways in order to satisfy their particular needs.

7.1 Base Functionality of the INDX PDS Platform

The base architecture of INDX consists of three components; a versioned database for semi-structured data, a distributed identity subsystem, and management logic that glues the components together. Each is described below, along with rationale for its design.

7.1.1 The Data Store

A key question in implementing the core component of a PDS is choosing the right database—what kind of data model should it use? What query language should it support? How should it store the data to ensure longevity?

As databases have evolved over the years, many kinds of database models have been proposed and improved. The INDX design process brought us to consider many popular database types, including “traditional relational databases, document oriented (or “NoSQL) databases, graph based data stores, XML databases, and RDF triple stores, to name a few. Each offers a few distinct advantages over the others, and many open source implementations exist of each type.

Since there are several advantages to using pre-existing databases, the most obvious of which is the fact that using mature, open-source software is likely to be more reliable and require less engineering than creating a bespoke solution from the ground up. Beyond this purely practical development consideration, there is a greater argument for being database-agnostic [17], rather than sticking to a single implementation. In order to realise the PDS vision of longevity, an unavoidable fact is that hardware and software is going to change dramatically, as will the database systems built on top of them; moreover, there may be a need to accommodate a variety of different data demands, with uses and needs continually increasing, as data streams become more numerous, personal data archives become larger, and query and sharing functionality is tasked with increasingly challenging applications. What may make sense to run on a single “conventional PC today might need to be run on a thousand nodes in some virtualised computer architecture in the future in order to accommodate an individuals increased storage and query capacity.

Therefore, using the age-old engineering principle of modularity, we sought to create the INDX PDS as an adapter on top of one or two basic underlying database systems. This decision has enabled us to target multiple databases at the outset, ranging from desktops and servers to mobile devices.

The question of finding an appropriate data and query model for PDSes is a more delicate question because the design choices made at this level are visible to, and thus directly affect, application developers, and to a certain extent, end-users. A variety of considerations need to be made when selecting the data model; first, whatever target model is chosen must be sufficiently flexible to accommodate (with reasonable transformation) the kinds of data that the platform will be managing. A poorly suited data model for the target will likely introduce inefficiencies that will either slow down performance, increase complexity or both.

Fortunately, most of the aforementioned data models are fairly general, each with specific characteristics; for example, relational databases require data to be factored into tables, which assumes a certain degree of data regularity; XML databases represent data as hierarchical structured documents; more general document-oriented stores manage collections of (either structured or unstructured) documents with limited metadata (comprising sets of keys for retrieval), while RDF ultimately represents data their granular components: triples.

Another dimension is that certain types of databases more typically afford guarantees that others do not; for example, many relational databases offer grades of ACID (Atomicity, Consistency, Isolation and Durability) guarantees [41], while few document-oriented or RDF triple stores do, partly due to technicalities arising from realising these guarantees in these settings. An additional advantage to relational databases is that extensive research on them has yielded well-known methods to tune performance, such as ways to factor tables to avoid otherwise computationally expensive query operations, the creation of indexes and so on, whereas such methods and query performance predictability is remains less well established for other database types.

The culmination of these observations, with the availability of an highly respected implementation have led us to target a relational database, Postgres [63], for desktop and server hosted INDX stores.

7.1.2 Datastore Management

However, despite its large feature set, Postgres does not, “out of the box meet all of the capabilities required of a PDS by the definition we arrived upon earlier. Given the need for PDSes to continue to meet changing information needs over an individuals lifetime, it is rather unlikely that any database will ever be devised at any point in time that will be able to fulfill all future information needs itself. Thus, this is where the design of the PDS has to provide incremental functionality extension, again, through encapsulation and modular design.

One of the immediate such functionality that must be added in order to use Postgres as the core data store is support for schema-less storage. Being a relational database, this is not straightforward; typical scenarios of the deployment of Postgres involves having a database programmer specifically create a bespoke set of schemas per data type being stored, consisting of tables and related views. Yet, in terms of PDSes, such needs may not be known at the time of set-up, and may change dramatically over time; moreover, it is practically impossible to know at design time the structure of all the data any user might want to store.

A second example also relevant to long-term data retention was providing the capability of a revisitable history of all data objects kept in the store. There are many uses for such a history, such as letting a user retrieve old versions of their objects, such as their documents, that were subsequently lost or altered, or determining how particular objects were changed over time. Such capabilities have started to become available in commodity software such as Apples Time Machine, platforms such as Dropbox and Skydrive, as well as many collaborative software tools. Thus, we believe that it such a capability will soon become a standard capability assumed by users.

Other capabilities that in the works for INDX include managing replicated copies (for enhanced resilience against datastore failure and corruption), sharing (such as object-level sharing support), and encryption for handling sensitive data. Such platform-level data capability allows PDS platform application writers to take advantage of sophisticated functionality and data security without having to implement them within apps themselves, allowing the unilateral improvement of data handling without adding application-wise complexity.

An important piece of functionality that the PDS management logic also has to assume is to access control, which involves orchestration of at least three separate components: access control policies specified by the user and stored as rules, the databases own gate-keeping mechanisms for granting access to the data kept within, and digital identities of users and applications requesting access, described next.

7.1.3 Distributed Identity Management

The current predominant model of identity management is that service providers perform this management directly for users; for example, service providers allow users to create principals with them, and provide authentication mechanisms as well. This model is inconvenient for a decentralised model of interaction, however, as it requires users to register a new principals with every single individuals PDS prior to interacting with them.

Distributed identity management protocols [35] offer a solution to this problem, by separating the problem of identity establishment and verification from its use. This permits, for example, an individual to grant access to sensitive data in their PDS to a verifiable identity of an entity, for example, their GP, even if their GP has never previously interacted with their PDS. Currently popular distributed identity management implementations include OpenID [57], WebID [32], Mozilla Persona [75].

A related problem that is distinct to identity management is that of allowing third parties to request and securely receive access to data (with the users permission). For this purpose, protocols such as OAuth [29] and SAML [3] have been developed and implemented across a large number of data providers, including Facebook, Instagram, and others. Such mechanisms allow these particular parties to continue to share data on behalf of the user once permission has been granted once, without subsequent user intervention.

INDXs reference implementation uses OAuth in conjunction with OpenID to allow interoperability with current Web services, particularly for the purpose of permitting transparent archiving of content that users distribute across the Web. It currently supports the archiving of content posted to social networking sites and services such as Twitter and Facebook, activity logging sites such as Nike+, Withings, and Moves, financial tracking sites such as Mint, open data sources such as OpenWeatherAPI, with support for other services to follow.

8 Looking Forward: Functionality for Future Information Management

In this final section, we wish to touch upon a few potential ways that PDSes might change the ways individuals will work with information in the future. A key goal will be to achieve consolidated data models from heterogeneous sources, for which we discuss the role of semantic technology and ontology matching and alignment algorithms; and the implications.

8.1 The Challenge of Automatic Consolidation

If one were to make an assumption that Personal Data Stores will eventually be able to draw in information obtained from hundreds to even thousands of third party data sources, for example, ranging from social networking posts to retail sites to Wikipedia to ones electronic medical record providers, so that such data may be safely archived, versioned and conveniently accessed, a question remains—how will this information be organised?

While this information could be kept separate and archived in its original form as provided, there are significant advantages to a user if this heterogeneous data is consolidated. By consolidation, we imply the act of combining complementary information from multiple sources into fewer, coherent and more complete and consistent representations. If this is done, like information items can be displayed in a consistent fashion, making coherent presentation and manipulation of items simpler; such consolidated information can be used by the user (and by the users applications) uniformly, effectively eliminating the aforementioned problems of fragmentation mentioned earlier. The advantages to the user of a single consolidated data model are many, and we discuss a few of the potential ways this may enable applications to do more sophisticated things for users later in this section.

If all information service providers adopted the a single unified schema for all information coming into and out of them, this goal could be achieved relatively simply, since data records from separate sources could be directly compared. However, it is fairly well accepted that achieving such a singular data representation is as unlikely as convincing the entire world to speak exactly one dialect of a single language; the degree of diversity and continued independent evolution of systems practically guarantees that this will never happen [6].

Thus to tackle this challenge, we must perform a kind of information integration, in which data are transformed into a consistent representation. For any pair of fixed sources, bespoke mapping could be specified by a programmer manually. However, if the applications are not known, or if the data came in arbitrary forms unknown in advance (such as if they came directly from a user), other methods must be employed. It is this latter situation that is likely to be quite common for PDSes, particularly considering the wide range of potential data and applications a user might need. We briefly discuss how semantic technology and ontology matching algorithms may be able to help.

8.1.1 Semantic Technology

Research pertaining to the Semantic Web has looked at methods by which automatic inference over heterogeneous information can be made possible by grounding such representations in ontologies related through ontology languages, such as OWL [2]. Such semantics establish a framework by which machine translation of information representations become made possible through the formal stated connections made about such representations. The role of *semantic reasoners* thus are to take information represented in such formats, along with their source ontologies, and to allow relationships among such information items to be deduced.

A requirement for such technology to work, however, is that all information providers provide appropriate mappings for their information representations against common ontologies using languages such as OWL. Thus far, few Web data sources outside of research and a few specific domains have embraced such techniques, making the use of such ontology languages, meaning that other approaches may also have to be employed. One such is the use of automatic ontology matching algorithms.

8.1.2 Ontology Matching: Automatic and Interactive Methods

Two other approaches have been taken to this problem; one is the use of machine-learning techniques for ontology matching (e.g. [20, 23], or *instance matching* [13, 64]). In such approaches, an algorithm is given a collection of examples of ontologies (or instances) and their corresponding semantic relationships, and the algorithm extrapolates properties to new, yet unseen relationships. This remains a rather computationally difficult task, however, and these methods have remained highly imperfect.

One promising approach has been to use such methods in combination with interactive approaches, that is to let users help such matching algorithms out when they get stuck. The *end-user programming* community has sought interfaces that can leverage information from non-expert individuals, who are empowered to assist and orchestrates the process of reconciliation at various levels of specificity. Systems that use this approach include "mash-up makers" (such as Mashmaker, [22], Marmite [36], Vegemite [76]) and interactive data workbenches, such as Data-Palette [69].

8.2 Defragmentation and "Placeless" Data

One of the greatest advantages of the Web is that it has started enable pervasive information access; for an increasing proportion of the worlds population, people can now access any information, anytime, anywhere from their desktops or mobile devices in nearly any setting [52]. Yet, the silos on the web have created artificial "places in themselves; so now it is necessary "go to facebook or "log into my universitys portal or "go to my health care provider, using the dedicated search and navigation facilities of these sites in order to get behind their *walled gardens*—even when the information being sought is the users own information!

Such walls impede individuals abilities to quickly access information needed, and in some cases, entirely preclude the ability for this information to be effectively cross-referenced, by preventing links from being established between these data items and increasing the barrier to accessing them. The result is often that the user experience of the Web has reverted back from the Memex vision of being able to navigate fluid "association lines of investigation aimed to complement the associative mechanisms of human memory and creative thinking, instead getting back to a series of online disparate bulletin board systems.

The vision of the PDS may reverse this at least for ones personal information, by providing consolidated representations of all of the information items distributed across silos that can be arbitrarily cross-referenced and linked. Doing this has its subtleties, however; as argued by, Marshall et al, "simply archiving by harvesting a persons out of all of ones third party services necessarily decontextualises from the context of its original location, application or Web service in which it was created or found [38]. In order to avoid having this loss of context, PDSes could provide “wormholes from the consolidated representation—which is better suited for sensemaking, to the individual services hosting the rich context of content.

8.3 Supporting Information Management for Life: Context-Sensitive Automation and Behaviour Change

Since it can be easily argued that the most valuable features of tools are the ones that are the greatest felt, we briefly touch on a few ways that the capabilities afforded by PDSes might directly impact peoples lives.

The all too familiar feeling of data loss that occurs when we have had a hard drive fail, or the frustration that arises from not being able to find a particular important document or photo demonstrates the potential for technology to save people from distress in many immediate and direct ways. The position of personal information tools, as the most intimate and direct mechanism for satisfying a majority of our information needs, means that small changes in these tools can have substantial long-term effects.

Across many of the biggest information management problems are a host of well-known techniques that are simply not used because they are simply too time consuming, require expertise, or that people simply forget. For example, data loss can be practically avoided in relatively simple ways through the creation of off-site backups and vigilance in continuing to back data up over time. However, the low-compliance rate to backup regimens simply comes from the fact that people are often either too busy, forgetful, or simply do not know how to carry out such backups regularly. Similarly, limited time, attention, effort and expertise serve as the root cause for many other problems concerning long-term data preservation and access, including disorganisation, ensuring data security, and accidental deletion.

One potential solution to all such problems is the judicious application of automation in supporting a broader set of information management and maintenance practices. Just as spam filters transparently and automatically remove unwanted mail to save people from having to delete it themselves, or Apples Time Machine continuously creates generational backups of the information on ones desktop and notebook without the user usually even being aware of it, we can imagine other information management activities being facilitated by more of such “attention free support.

A particular kind of automation that has thus far been technically challenging to realise but well-suited to the capabilities afforded by PDSes is *context-aware* and adaptive automation that is sensitive to a users needs, location, and activities. Since PDSes consolidate multiple information streams about a persons sensed activities (such as through wearable activity sensors or apps), it can consolidate the most complete digital “shadow of the individual. This shadow, can, in turn be used by applications to provide attention-free automation support; for example, by using information about one credit card statement (such as from Mint) with ones current location (sensed via ones smartphone) and purchase history (collected in ones PDS over a long term), a future application might infer automatically that one is at risk of going over credit limit and intervene, either by warning the user, or automatically transfer money on his or her behalf to avoid over-transaction fees. Many such context aware scenarios have been proposed before (e.g. [19, 59, 61]), but their inability to get accurate, high-dimensional data of the users context have impeded progress. PDSes seem an appropriate solution to this, particularly in situations such as the above where the involved in the inference is highly sensitive and personal, such as ones bank account balance, current location, medical conditions, and so on.

When such context-sensitive and adaptive approaches are applied to health and wellbeing, it can be used to play a role delivering better personalised coaching and intervention support. Simple forms of fitness coaching are already becoming available on the market, usually delivered as part of low-cost commercial activity sensors such as Nikes FuelBand, or Withings body scale and blood pressure products. However, few of these applications are able to perform sophisticated tailoring due to the limited information available about the user from these single, simple sensor streams. Therefore, the kinds of multi-stream consolidation of user context may be helpful here towards more effective digital support in wellbeing maintenance, intervention and recovery.

9 Conclusion

In this chapter we have attempted to position the notion of Personal Data Stores as a (partial) response to the pressing problem of the autonomy of the data subject, and the asymmetry of power between the subject and large-scale service providers and data consumers. Given what Novotny and Spiekermann have called the “missing governance of personal data markets [46] threatens to undermine subject trust in data sharing practice, and given that data sharing underlies not only a series of very valuable public services but also a whole economy, PDSs are highly suggestive of a means of putting the data subject at the centre of the data markets institutional structure.

The notion of ‘personal data is, for obvious reasons, in thrall to a legal definition that governs liability and policy, but the narrow legalistic coverage that this subtends should surely be supplemented by a more intuitive notion of the data which is of interest, importance or value to individuals. Such a rethink would help both individuals, many of whom are concerned, if only in the abstract, that their privacy is being undermined by the collection, storage, aggregation and mining of their data, and data consuming organisations, many of which are concerned about a potential backlash. The rules governing ownership of data seem unlikely to change, as this would hamper the development of an equitable data economy, but regulatory and technical models are emerging in which the rights and responsibilities of various stakeholders are redistributed. PDSs are part of that emerging picture. It is also worth pointing out, however, that even with an unchanged regulatory position, PDSs have made some progress (e.g. [31])—and the regulatory position is unlikely to remain unchanged in the charged atmosphere (at the time of writing) caused by the Snowden revelations and the revisions to the EUs Data Protection Directive.

Earlier, we set out six challenges facing PDSs, and described a reference implementation called INDX. The intention for INDX was not to make a claim that it would in its current state (or ever) solve all of the challenges, but to serve as a common artefact around collaborative research discourse for investigating socio-technical issues and user needs. As a functional open platform, our hope is that it might be adopted as an instrument that accelerates research towards more flexible and adaptive information environments that assume dramatically different forms and shapes than our current models of silo-encapsulated hegemonies in the cloud.

Acknowledgements The work described in this chapter is supported through an interdisciplinary ESPRC Research Grant EP/J017728/1, The Theory and Practice of Social Machines. The design and development of the INDX platform has benefitted from contributions from SOCIAM researchers, including Daniel Alexander Smith, Laura Dragăn, Markus Lucsak-Roesch, Ramine Tinati, and Dave Murray-Rust, and with direction and advising from Sir Nigel Shadbolt, Sir Tim Berners-Lee, and Dame Wendy Hall.

References

1. Agustin, J.M., Albritton, W.M.: Vendor relationship management (2001)
2. Antoniou, G., Van Harmelen, F.: Web Ontology Language: OWL. In: Handbook on Ontologies, pp. 67–92. Springer, New York (2004)
3. Armando, A., Carbone, R., Compagna, L., Cuellar, J., Tobarra, L.: Formal analysis of SAML 2.0 web browser single sign-on: breaking the SAML-based single sign-on for Google Apps. In: Proceedings of the 6th ACM Workshop on Formal Methods in Security Engineering, pp. 1–10. ACM, New York (2008)
4. Bandura, A.: Self-efficacy: toward a unifying theory of behavioral change. Psychol. Rev. **84**(2), 191 (1977)
5. Banisar, D., Davies, S.: Global trends in privacy protection: An international survey of privacy, data protection, and surveillance laws and developments. John Marshall J. Comput. Inform. Law **18**(1) (1999)
6. Bannon, L., Bødker, S.: Constructing common information spaces. In: Proceedings of the Fifth European Conference on Computer Supported Cooperative Work, pp. 81–96. Springer, New York (1997)
7. Bell, C.G., Gemmell, J., Rosson, C.: Total recall. Dutton (2010)
8. Bellotti, V., Dalal, B., Good, N., Flynn, P., Bobrow, D.G., Ducheneaut, N.: What a to-do: studies of task management towards the design of a personal task list manager. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 735–742. ACM, New York (2004)
9. Benyon, D., Höök, K.: Navigation in information spaces: supporting the individual. In: Human-Computer Interaction INTERACT97, pp. 39–46. Springer, New York (1997)
10. Boston Consulting Group: Unlocking the value of personal data: From collection to usage (2013)
11. Bush, V.: As We May Think. The Atlantic Monthly, Boston (1945)
12. Bygrave, L.A.: Data Protection Law. Kluwer Law International, New York (2002)

13. Castano, S., Ferrara, A., Montanelli, S.: Matching ontologies in open networked systems: Techniques and applications. J. Data Semant. V 25–63 (2006)
14. Cavoukian, A., El Emam, K.: Dispelling the Myths Surrounding De-identification: Anonymization Remains a Strong Tool for Protecting Privacy. Office of the Privacy and Information Commissioner, Ontario (2011)
15. Chervenak, A., Vellanki, V., Kurmas, Z.: Protecting file systems: A survey of backup techniques. In: Proceedings Joint NASA and IEEE Mass Storage Conference, vol. 3 (1998)
16. Chi, E.H., Pirolli, P., Chen, K., Pitkow, J.: Using information scent to model user information needs and actions and the web. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 490–497. ACM, New York (2001)
17. Chohan, N., Bunch, C., Krintz, C., Nomura, Y.: Database-agnostic transaction support for cloud infrastructures. In: Cloud Computing (CLOUD), 2011 IEEE International Conference on, pp. 692–699. IEEE (2011)
18. Choney, S.: Facebook timeline poll: 'overwhelming negative' reaction. Today (2012). URL http://www.today.com/tech/facebook-timeline-poll-overwhelming-negative-reaction-84717
19. Dey, A.K.: Understanding and using context. Personal Ubiquitous Comput. **5**(1), 4–7 (2001)
20. Doan, A., Madhavan, J., Dhamankar, R., Domingos, P., Halevy, A.: Learning to match ontologies on the Semantic Web. VLDB J. **12**(4), 303–319 (2003)
21. Engelbart, D.C., English, W.K.: A research center for augmenting human intellect. In: Proceedings of the December 9–11, 1968, Fall Joint Computer Conference, Part I, pp. 395–410. ACM, New York (1968)
22. Ennals, R., Brewer, E., Garofalakis, M., Shadle, M., Gandhi, P.: Intel Mash Maker: join the web. SIGMOD Rec. **36**(4), 27–33 (2007)
23. Euzenat, J.: An API for ontology alignment. Proc ISWC '04 pp. 698–712 (2004)
24. Fertig, S., Freeman, E., Gelernter, D.: Lifestreams: an alternative to the desktop metaphor. In: Conference Companion on Human Factors in Computing Systems, pp. 410–411. ACM, New York (1996)
25. Floridi, L.: The Ethics of Information. Oxford University Press, Oxford (2013)
26. Fung, A., Graham, M., Weil, D.: Full Disclosure: The Perils and Promise of Transparency. Cambridge University Press, Cambridge (2007)
27. Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C.: Mylifebits: fulfilling the memex vision. In: Proceedings of the Tenth ACM International Conference on Multimedia, pp. 235–238. ACM, New York (2002)
28. Grasso, M.A., Yen, M.J., Mintz, M.L.: Survey of handheld computing among medical students. Comput. Meth. Programs Biomed. **82**(3), 196–202 (2006)
29. Hardt, D.: The oauth 2.0 authorization framework (2012). IETF RFC 6749
30. Heath, T., Bizer, C.: Linked data: Evolving the web into a global data space. Synth. Lect. Semantic Web Theory Tech. **1**(1), 1–136 (2011)
31. Heath, W., Alexander, D., Booth, P.: Digital enlightenment, mydex, and restoring control over personal data to the individual. In: Hildebrandt, M., OHara, K., Waidner, M. (eds.) Digital Enlightenment Yearbook 2013: The Value of Personal Data, pp. 253–269. IOS Press, Amsterdam (2013)
32. Huang, G., Mak, K.: WeBid: a web-based framework to support early supplier involvement in new product development. Robot. Comput. Integrated Manuf. **16**(2), 169–179 (2000)
33. Jones, W.: Keeping Found Things Found: The Study and Practice of Personal Information Management: The Study and Practice of Personal Information Management. Morgan Kaufmann, Amsterdam (2010)
34. Karger, D.R., Jones, W.: Data unification in personal information management. Comm. ACM **49**(1), 77–82 (2006)
35. Koshutanski, H., Ion, M., Telesca, L.: Distributed identity management model for digital ecosystems. In: Emerging Security Information, Systems, and Technologies, 2007. SecureWare 2007. The International Conference on, pp. 132–138. IEEE, New York (2007)
36. Lin, J., Wong, J., Nichols, J., Cypher, A., Lau, T.A.: End-user programming of mashups with vegemite. In: Proc. IUI '09, pp. 97–106. ACM, New York (2009). DOI 10.1145/1502650.1502667. URL http://doi.acm.org/10.1145/1502650.1502667

37. Malone, T.W.: How do people organize their desks?: Implications for the design of office information systems. ACM Trans. Inform. Syst. (TOIS) **1**(1), 99–112 (1983)
38. Marshall, C.C.: Challenges and opportunities for personal digital archiving. Digital: Personal Collections in the Digital Era pp. 90–114 (2011)
39. Mayer-Schönberger, V., Cukier, K.: Big Data: A Revolution That Will Transform How We Live, Work and Think. John Murray, London (2013)
40. McCallister, E.: Guide to Protecting the Confidentiality of Personally Identifiable Information. DIANE Publishing, Darby, PA, US (2010)
41. Muth, P., Rakow, T.C.: Atomic commitment for integrated database systems. In: Data Engineering, 1991. Proceedings. Seventh International Conference on, pp. 296–304. IEEE, New York (1991)
42. Nadkarni, A., Hofmann, S.G.: Why do people use facebook? Pers. Indiv. Differ. **52**(3), 243–249 (2012)
43. Narayanan, A., Shmatikov, V.: Myths and fallacies of "personally identifying information. Comm. ACM **53(6)**, 24–26 (2010)
44. Nelson, T.H.: Literary Machines: The Report On, and Of, Project Xanadu Concerning Word Processing, Electronic Publishing, Hypertext, Thinkertoys, Tomorrow's Intellectual Revolution, and Certain Other Topics Including Knowledge, Education and Freedom. Nelson Ted; Schooley's Mountain, NJ: distrib. by Distributors (1987)
45. Newman, W.M., Eldridge, M.A., Lamming, M.G.: Pepys: Generating autobiographies by automatic tracking. In: Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW91, pp. 175–188. Springer, New York (1991)
46. Novotny, A., Spiekermann, S.: Personal information markets and privacy: a new model to solve the controversy. In: Hildebrandt, M., OHara, K., Waidner, M. (eds.) Digital Enlightenment Yearbook 2013: The Value of Personal Data, pp. 102–120. IOS Press, Amsterdam (2013)
47. Odom, W., Banks, R., Kirk, D., Harper, R., Lindley, S., Sellen, A.: Technology heirlooms?: considerations for passing down and inheriting digital materials. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 337–346. ACM, New York (2012)
48. OHara, K.: Are we getting privacy the wrong way round? IEEE Internet Comput. **17**(4), 89–92 (2014)
49. OHara, K.: The information spring. IEEE Internet Comput. **18**(2), 79–83 (2014)
50. Ohm, P.: Broken promises of privacy: responding to the surprising failure of anonymization. UCLA Law Rev. **57**, 1701–1777 (2010)
51. Olmstead, T.: Facebook timeline and users: Not quite a love affair. Mashable (2012). URL http://mashable.com/2012/01/31/facebook-timeline-poll-results/
52. Perry, M., O'hara, K., Sellen, A., Brown, B., Harper, R.: Dealing with mobility: understanding access anytime, anywhere. ACM Trans. Comput. Hum. Interact. (TOCHI) **8**(4), 323–347 (2001)
53. Pirolli, P., Card, S.: Information foraging. Psychol. Rev. **106**(4), 643 (1999)
54. Pirolli, P., Card, S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: Proceedings of International Conference on Intelligence Analysis, vol. 5, pp. 2–4 (2005)
55. Posner, R.A.: The economics of privacy. Am. Econ. Rev. **71**, 405–409 (1981)
56. Purdue, D., Dürrschmidt, J., Jowers, P., O'Doherty, R.: Diy culture and extended milieux: Lets, veggie boxes and festivals. Socio. Rev. **45**(4), 645–667 (1997)
57. Recordon, D., Reed, D.: OpenID 2.0: a platform for user-centric identity management. In: Proceedings of the Second ACM Workshop on Digital Identity Management, pp. 11–16. ACM, New York (2006)
58. Schaller, R.R.: Moore's law: past, present and future. IEEE Spectrum **34**(6), 52–59 (1997)
59. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: Mobile Computing Systems and Applications, 1994. WMCSA 1994. First Workshop on, pp. 85–90. IEEE, New York (1994)

60. Schilit, B.N., Adams, N., Gold, R., Tso, M.M., Want, R.: The parctab mobile computing system. In: Workstation Operating Systems, 1993. Proceedings., Fourth Workshop on, pp. 34–39. IEEE, New York (1993)
61. Selker, T., Burleson, W.: Context-aware design and interaction in computer systems. IBM Syst. J. **39**(3.4), 880–891 (2000)
62. Shadbolt, N.: Midata: towards a personal information revolution. In: M. Hildebrandt, K. OHara, M. Waidner (eds.) Digital Enlightenment Yearbook 2013: The Value of Personal Data, pp. 202–224. IOS Press (2013)
63. Stonebraker, M., Rowe, L.A.: The design of postgres. In: Proc. of International Conference on the Management of Data, pp. 340–355. ACM, New York (1986)
64. Suchanek, F., Abiteboul, S., Senellart, P.: PARIS: probabilistic alignment of relations, instances, and schema. Proc. VLDB 11 **5**(3), 157–168 (2011)
65. Teevan, J., Alvarado, C., Ackerman, M.S., Karger, D.R.: The perfect search engine is not enough: a study of orienteering behavior in directed search. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 415–422. ACM, New York (2004)
66. Teevan, J., Jones, W., Bederson, B.B.: Personal information management. Comm. ACM **49**(1), 40–43 (2006)
67. The new personal data landscape (2011). URL http://ctrl-shift.co.uk/wordpress/wp-content/uploads/2011/11/The-new-personal-data-landscape-FINAL.pdf
68. Vakkari, P.: Task-based information searching. Annu. Rev. Inform. Sci. Tech. **37**(1), 413–464 (2003)
69. Van Kleek, M., Smith, D.A., Packer, H.S., Skinner, J., Shadbolt, N.R.: Carpé data: supporting serendipitous data integration in personal information management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2339–2348. ACM, New York (2013)
70. Van Kleek, M.G., Bernstein, M., Panovich, K., Vargas, G.G., Karger, D.R., Schraefel, M.: Note to self: examining personal information keeping in a lightweight note-taking tool. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1477–1480. ACM, New York (2009)
71. Von Hippel, E.: Learning from open-source software. MIT Sloan Manag. Rev. **42**(4), 82–86 (2001)
72. Wang, Y., Kobsa, A.: Privacy-enhancing technologies. Social and Organizational Liabilities in Information Security, pp. 203–227 (2006)
73. Whittaker, S.: Personal information management: from information consumption to curation. Annu. Rev. Inform. Sci. Tech. **45**(1), 1–62 (2011)
74. Whittaker, S., Hirschberg, J.: The character, value, and management of personal paper archives. ACM Trans. Comput. Hum. Interact. (TOCHI) **8**(2), 150–170 (2001)
75. Williams, N., Howard, L.: A SASL and GSS-API mechanism for the BrowserID authentication protocol (2013)
76. Wong, J., Hong, J.I.: Making mashups with marmite: towards end-user programming for the web. In: Proc. CHI '07, pp. 1435–1444. ACM, New York (2007). DOI 10.1145/1240624.1240842. URL http://doi.acm.org/10.1145/1240624.1240842
77. Yang, M., Sassone, V., OHara, K.: Appendix 3: Practical Examples of Some Anonymisation Techniques, pp. 80–103. UK Information Commissioners Office (2012)

An Auditable Reputation Service for Collective Adaptive Systems

Heather S. Packer, Laura Drăgan, and Luc Moreau

1 Introduction

A subject's reputation is a measure of how much a community rates it. Reputation plays a core role in online communities such as eBay and StackExchange since it can influence the communities perception and interactions, and affect computational activities within the system. In eBay, the reputation of a seller can help a buyer decide whether they want to purchase an item from this seller. In StackExchange, reputation is a key incentive for people to contribute as it leads to kudos and potential employment offers.

A subject's reputation can be derived from feedback, which may be of two kinds:

1. User feedback consists of ratings or comments provided by users who participate in the system, and have interacted with the subject.
2. System feedback consists of various metrics directly measurable by the system, including performance, timeliness and responsiveness.

Reputation can be evaluated either manually or automatically, on a set of criteria which differs across domains.

For a subject to achieve and maintain a good reputation, it is important to understand how different factors influence its calculation. Given that reputation varies over time as feedback is submitted, it is desirable for a reputation provider to be accountable. In order to be accountable it is required to explain how reputation measures have changed over time, which feedback reports affected the reputation, and any changes to how it is measured. Hence, auditing is a key mechanism that a reputation provider can offer its clients to inspect reputation measures it derives.

H.S. Packer (✉) • L. Drăgan • L. Moreau
University of Southampton, University Road, Southampton, SO17 1BJ, UK
e-mail: hp3@ecs.soton.ac.uk; lcd@ecs.soton.ac.uk; l.moreau@ecs.soton.ac.uk

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_8

In this context, provenance can be used to provide descriptions of the entities, activities, and agents that may have influenced a reputation measure.

While the use of reputation is frequent in Collective Adaptive Systems (CAS), there is a lack of agreed methods for its use, representation, and auditability. The aim of this chapter is to investigate key facets of an auditable reputation service for CAS, summarised by the following contributions:

1. Use cases for reputation and provenance in CAS, which are categorised into functional, auditable, privacy and security, and administrative.
2. A RESTful Reputation API, which allows users access to subject feedback and to access feedback reports and reputation measures.

In Sect. 2 we outline related work on trust and reputation, and social computation. Following that, in Sect. 3, we describe generic provenance use cases. Then in Sect. 4, we discuss a reputation API. In Sect. 5, we describe in detail a use case for a ride share application. Finally, Sect. 6 concludes the paper.

2 Background and Related Work

The following sections define provenance, trust and reputation, and describe their use in the context of CAS.

2.1 Provenance

Provenance has varied emerging applications: it may be used to make CASs accountable and transparent [35]; provenance can help determine whether data or users can be trusted [12]; and provenance can be used to ensure reproducibility [18] of computations.

In this chapter, we adopt the W3C (World Wide Web Consortium) Provenance Working Group's definition, given in the PROV Data Model specification [20]:

> Provenance is defined as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing.

PROV is a recent recommendation of the W3C for representing provenance on the web. PROV is a conceptual data model (PROV-DM [20]), which can be mapped and serialized to different technologies. There is an OWL2 ontology for PROV (PROV-O [15]), allowing mapping to RDF, an XML schema for provenance [10], and a textual representation for PROV (PROV-N [21]).

PROV can be divided into a core, forming the essence of provenance, and a set of extended constructs, catering for more advanced uses. There are three views according to which the PROV data model can be presented: *data flow* view, *process flow* view, and *responsibility* view, Fig. 1 displays these three views of provenance and the associated core classes and properties. In Sect. 5.2 we show an extension of the PROV ontology, in the description of the Provenance for CAS ontology.

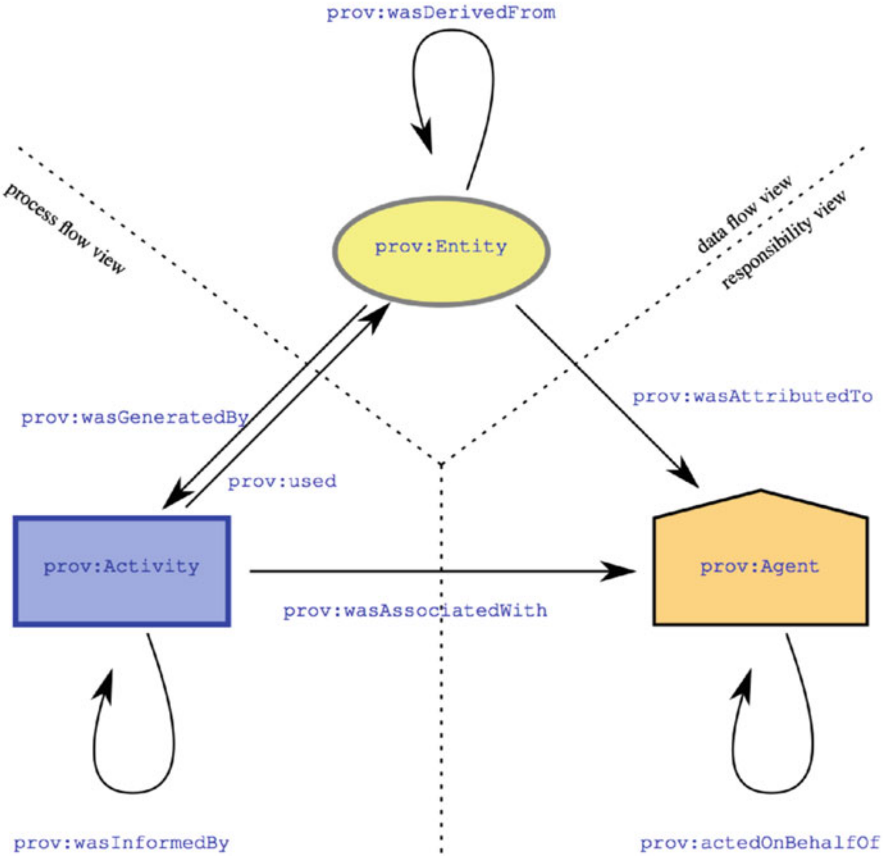


Fig. 1 Three different views of the core of PROV. The figure adopts the PROV layout conventions: an entity is represented by a *yellow ellipsis*, an activity by a *blue rectangle*, and an agent by an orange pentagon. We note here that the diagram is a "class diagram" illustrating the classes that occur as domain and range of properties. Taken from [19]

2.2 Trust and Reputation

The topic of trust and reputation has been extensively reviewed [4, 7, 25, 28], with Artz and Gil's [4] being one of the more comprehensive survey of definitions and existing systems. Sabater and Siera [28] propose a set of criteria for the characterisation of computational models, and then proceed to qualify a selection of existing models according to these criteria. Pinyol and Sabater-Mir [23] more recently surveyed trust and reputation models for open multi-agent systems, a category to which CASs belong. This second survey cites the previous characterisation criteria as just one of several existing classification systems.

Goldbeck's work, which spans several years [6–9, 11] tackles the way trust and reputation is defined and computed in online social networks. Golbeck describes trust and reputation as "socially loaded" terms, and reputation as a "more social notion of trust" [8]. The general definition of trust used in her PhD thesis [9] is:

> Alice trusts Bob if she commits to an action based on a belief that Bob's future actions will lead to a good outcome.

Different types of trust are described and studied in the research literature, and trust is often equated with the mechanism for authentication, such as digital signatures and encryption. The survey by Artz and Gil [4] lists the diverse research areas of trust, from security and access control to decision making and game theory, differentiating between policy-based and reputation-based trust—the former focusing on "hard security" and the latter on "social" aspects. The survey focuses on trust representations in the Semantic Web.

In the context of the Semantic Web, Tim Berners-Lee envisioned [2] that provenance is a crucial component in establishing trust. In [1], Berners-Lee introduces the idea of an easy way to access provenance information provided on websites with an "oh yeah?" button:

> At the toolbar (menu, whatever) associated with a document there is a button marked "Oh, yeah?". You press it when you loses that feeling of trust. It says to the Web, "so how do I know I can trust this information?".

The idea is that if we can determine where data and documents come from, we can decide whether it can be trusted. Li et al. [16] outline how trust can be developed, or distrust minimised, through provenance on the Semantic Web, by describing and generalizing several use cases where possible "distrust events" occurred. Prat and Madnick [24] define a provenance-grounded measure of the believability of data on the web, relying on a measure of trustworthiness of an agent as one of three dimensions.

A large amount of research in the area of trust comes from the multi-agent domain [22, 23, 25]. In a multi-agent context [25], trust is defined as:

> Trust is a belief an agent has that the other party will do what it says it will (being honest and reliable) or reciprocate (being reciprocative for the common good of both), given an opportunity to defect to get higher payoffs.

The focus of multi-agent systems trust is on actions performed by agents. It is to be distinguished from trust of information (or content trust [5]) defined as follows:

> Content trust is a trust judgement on a particular piece of information in a given context.

In this chapter we use a definition of trust that is based on a mix of provenance and reputation and content-based trust, not including the hard security.

Trust and reputation models typically introduce a measure, to represent how trustworthy or reputable a system or individual is. This measure is then typically summarised in a value, discrete (e.g. 1 to 9 [6], or 1 to 5 star rating as used by online commerce sites like Amazon and Ebay) or continuous (e.g. [0,1] [27]), or a label (trustworthy/not trustworthy). This value is then used to make trust decisions.

Reputation about a subject is acquired from the experiences of others in a community. According to [8] "reputation is synonymous with the measure of trust", and from a more social perspective [4]:

> Reputation is an assessment based on the history of interactions with or observations of an entity, either directly via personal experience or as reported by others.

Reputation systems can be centralized or distributed [3, 13, 14, 34]. Liu [17] describes criteria for classifying and analysing centralised reputation systems. CASs which are centralised usually use a centralised reputation system. However, some aspects in the reputation system can be decentralised, where participants can set preferences which favour or reject some sources, without taking into consideration the reputation, thus trusting or distrusting them implicitly—an example is TrustMail described by Golbeck and Hendler [8].

2.3 Provenance, Trust and Reputation in CAS

Berners-Lee and Fischetti [2] define a social machine, which are a synonym for CAS in the context of this work,[1] as systems "in which the people do the creative work and the machine does the administration", where both human and machines contribute to the completion of a task which they could not achieve separately. The characterisation of social machines is also described in the chapter [31] entitled "A Taxonomic Framework for Social Machines".

A number of terms and research areas involve the intersection of social behaviour and computational systems: social computing, crowdsourcing, collective intelligence, human computation [29]. CAS are used for a large variety of tasks, too complex for humans or computers to achieve independently, but which can be divided in small simpler tasks for one of the other. These include annotation, disaster relief, mapping, collaborative filtering, online auctions, prediction markets, reputation systems, computational social choice, tagging, and verification. Many existing systems employ strategies which can qualify them as CAS, including Wikipedia, OpenStreetMap, Ushahidi, re-Captcha.

The environment for collaboration in social CASs varies from system to system, it can be loosely mediated as in Twitter, or under stricter control of community policies and guidelines like in Wikipedia. Stein and Hess [32] show that in the German Wikipedia the quality of contributions is connected to the reputation of participants. The ability to uniquely identity and assess participants inputs in a collective based on their past actions or perceived domain knowledge in the system,

[1] In general a CAS may be constituted only by artificial agents, with no humans, and thus with no social elements.

is a factor in measuring the trust and reputation of the collective output.[2] Although some CASs go out of their way to prevent uniquely identifying users, such as 4chan, most will have a way of identifying participants, usually through user accounts.

The adoption of a CAS depends on a combination of many factors, some of them include:

- the purpose set out when the CAS was created,
- the perceived benefits to the participants,
- the amount and type of tasks, or
- the level of participation required.

Human participants have to trust that the machine will deliver the expected outcomes and that any information provided by them will be used for the purpose which was described. Specifically, the users must trust that the machine will not do anything with the information that they provide, which conflicts with the purpose for which this data was captured.

Weitzner et al. [35] argue that, for information, "accountability must become a primary means through which society addresses appropriate use". For them, "information accountability means the use of information should be transparent, so it is possible to determine whether a particular use is appropriate under a given set of rules, and that the system enables individuals and institutions to be held accountable for misuse". Dynamically assembled systems need to be made accountable for users to gain confidence in them, i.e., their past executions must be auditable so that compliance with, or violations of, policies can be asserted and explained. They also note the similarity between accountability and provenance in scientific experiments. Provenance is a key enabler for accountable systems since it consists of an explicit representation of past processes, which allows us to trace the origin of data, actions and decisions (whether automated or human-driven). It therefore provides the necessary logs to reason about compliance or violations. As users delegate important tasks to systems and endow them with private data, it is crucial that they can put their trust in such systems. Accountability is a way by which trust can be built, since action transparency and audit help users gain trust in systems. However, users may not always want (or have the resources) to audit systems; instead, they would like to be given a measure of trust, which they can rely upon to decide whether to use a system or not.

The output of CASs may be influenced by many factors including collective human input and machine processes. Because information is derived from many agents and processes, it can be challenging to understand and evaluate the results. Provenance information can be used to understand better the results, allowing their reliability and quality to be analysed. Therefore, understanding CAS hinges on capturing the provenance of the creation of data by both humans and machines.

[2] Uniquely identifying participants does not require or imply the use of any personally identifiable information, which would connect the participant to the real person.

3 Use Cases for Provenance, Trust and Reputation

In order develop a fully auditable reputation service, it is necessary to capture provenance information from the applications which use it. Hence, we first consider important design issues and decisions for applying provenance, trust and reputation to social machines. We ground our recommendations in a set of generic use cases for social machines. The use cases were outlined by considering generalised user requirements of social machines in the public domain, by investigating a subset of the social machines identified by the SOCIAM and SmartSociety projects.[3] We discuss methods for using provenance in social machines.

We have categorised the use cases into several types: functional, audit, privacy and security, and administration. We note that not all of the use cases are suited to all social machines, and we have described a machine's applicable attributes. We refer to participants as either humans or machines, which take part in the function of the social machine.

3.1 Functional Use Cases

The following use cases describe scenarios where provenance, trust and reputation can be used to support a social machine's functional requirements.

Use Case 1. *A participant creates or edits a piece of information in the system.*

This is the basic use case of such systems, and we require that provenance is captured for the new piece of information created, or the changes to existing information. In social machines like Wikipedia, where the generated information is the actual final output of the system, this kind of information is very important. When an editor changes an article, the information logged comprises of user name or IP address of the editor, the date and time, and the changes made. These information items are made visible to all other participants, passive or active, through the "View history" tab. Depending on how the systems allow access (which based on ownership, access rights, or roles) to the objects they manage, some participants might not be able to edit part of the information, in which situation the next use case is relevant.

Use Case 2. *A participant annotates existing resources in the system.*

An annotation is any meta-information about the main objects used by the social machine. This includes ratings of existing users or products in an online store, feedback on user activities in a listing of service providers, feedback on the quality of data. Information which is the main focus in one social machine, can

[3] A comprehensive list of the systems identified can be found at: http://sociam.org/social-machines.

be considered annotation in another system, for example ratings and comments on products on eBay might be annotations, while comments and ratings on TripAdvisor and Yelp are the main focus of the system. If we consider creating annotations as adding new data in the system, then this use case is a sub-part of the previous one.

Use Case 3. *A participant has to make a decision which requires her to select from a subset of the objects (or users) available in the system.*

This use case is relevant to social machines that rely on reputation and user preferences to enable participants to make decisions—make a selection. For example: On Amazon users select which product best suits their needs and rely on the product information and reviews; on TripAdvisor users select hotels or restaurants based on their location preferences, and reviews; and on Stack Overflow users post programming related problems and questions, and receive solutions, and can then select which is the best solution based on their opinion, a voting system, and user reputations.

3.2 Audit Use Cases

The provenance and reputation information collected in a social machine can support the ability to audit it.

Use Case 4. *A participant wants to know who created or changed a resource.*

This use case requires that the system provides a way to expose the users to provenance data captured as a result of the first or second use cases listed above. This use case is applicable to data that has been edited collectively, like for instance Wikipedia articles, where it is important to be able to see when and who made a change. Wikipedia also provides "Talk pages" where edits to articles can be discussed, and which allow participants to understand the motivations behind the changes, so that future edits take into account considerations of past motivations. The next use case refers to annotations, in a similar manner.

Use Case 5. *A participant wants to know who annotated a resource and when.*

Amazon and eBay are social machines whose participants benefit from being able to see provenance information of annotations. The users can decide on the value of ratings by checking who and when they were posted. This will allow the user to make an informed decision about the vote rating.

Use Case 6. *A participant wants to know how the reputation of a user is computed by the system.*

This use case requires that the reputation scores for participants also have provenance information which makes them auditable as well. The method used in computing the reputation scores should be easy to understand and available to participants, as part of the provenance of the reputation. For most online stores, reputation of sellers is computed in a straightforward manner by averaging the ratings received over time. However, some systems might decide to discard ratings older than a given date, or given by users with a low reputation. Such choices in the formula might improve the overall accuracy of the resulting scores, but they should be known to participants.

Within the scope of this use case is included also the possibility of the user enquiring about their own reputation, to see how they are perceived by other participants in the system and what factors influenced it. This leads to the next use case.

Use Case 7. *A participant has audited their reputation score, but they consider it is incorrect and would like to influence it.*

Some social machines allow users to verify information provided about them by other users, and take under consideration evidence that refutes the incorrect information. An example of this is the eBay Resolution Center, which allows buyers and sellers to resolve conflicts in a controlled environment, before negative ratings are submitted. Yelp on the other hand does not arbitrate reviews, but they do allow businesses to post public responses to reviews, in which to address the issues.

3.3 Privacy and Security Use Cases

The use cases in this section describe scenarios where provenance, trust and reputation can be used to support a user's security and privacy requirements. They are applicable mainly to human users of social machines, especially those systems which request and store personal details, for example Facebook, LinkedIn, Twitter.

Use Case 8. *A participant wants to change her personal information and preferences stored by a social machine.*

This use case includes adding new information, changing existing values, and removing previously set personal data from the system. The users should also be informed what other usage information the social machine captures, and should be able to decide if they agree to this data being stored. For example, Google uses location data from Android phones to map congestion areas in Google Maps,[4] but they allow users to opt out of this crowdsourcing experiment.

Use Case 9. *A participant wants to know who has access to her personal information.*

This use case is as much about auditing the system as it is about privacy and security. It includes situations when the user is concerned about who can see and use her personal information as stored in the system, both among other internal users, and

[4] http://googleblog.blogspot.co.uk/2009/08/bright-side-of-sitting-in-traffic.html.

external entities, like advertising companies for example. LinkedIn in particular employs this information to show its users who of its other users has viewed their profile information, as a possible show of interest.

3.4 Administrative Use Cases

The use cases in this section refer to the use of provenance, trust and reputation for the administration of the social machine.

Use Case 10. *An administrator wants to quantify how much of a goal of the system has been achieved.*

This use case is relevant to social machines which have quantifiable goals. It is useful if the overall goal is divided in sub-goals, which must be achieved before the next state or activity can occur. For example, in the CollabMap[5] [26] social machine for map building for emergency response, an evacuation route from a building can only be created after the building outline was created, and validated by a given number of independent users. Some social machines have a running goal, which can never be completed, like for example Wikipedia's aim of capturing world knowledge. Some social machines can have dual goals, one or all of which can be quantifiable, like for instance reCAPTCHA [33] which on one side is used to validate that the user is human by being able to decipher images of words, and at the same time the result is used for digitization of books.

Use Case 11. *An administrator wants to analyse statistics about the users' behaviour and achievements.*

User statistics can be used for tracking the adoption of a social machine, but also for feeding back information into the social machine, and influencing its further development. For example, this use case is applicable to social machines which use gamification elements like star ratings, badges and leader boards. In this case, usage analysis can help to identify behaviours to reward, or how much certain actions should contribute to a user's score, based on provenance. For example, the protein folding game Foldit could analyse provenance data to find who performed complex moves and reward them with a higher score or ranking.

Use Case 12. *An administrator wants to check that the resources are used according to the agreed upon rules.*

This use case is applicable to social machines which require resources to be created in accordance to specific policies. Auditing the provenance data allows the user to validate the sequence of activities performed over entities by an agent.

[5] http://collabmap.org/.

4 Reputation Service API

In this section, we present a REST API for a reputation service, which stores and retrieves user feedback, and retrieves of reputation information. We use the API to support the capture of provenance for the ride share application, discussed in the following section. It will be used to aid in the development of reusable provenance patterns for REST services.

CASs, such as LinkedIn, Stack Overflow, and eBay, use reputations to allow users to make trust judgements, and also to instil trust in the system. A reputation service will compute and publish reputation scores for a set of subjects (such as users, goods, service providers and services) within a CAS, based on the opinions of other users about a subject. The opinions are typically ratings and are sent to the reputation service, which uses a specific reputation algorithm to dynamically compute the reputation scores based on the received ratings.

Users of a CAS use the reputation scores for decision making: a subject with a high reputation score will normally attract more business than a subject with a low reputation score. It is therefore in the interest of users to: have a high reputation score; know what factors influenced the reputation score; and understand how to improve their score. Also, a transparent reputation system, which is clear in the way it computes scores, appears more trustworthy to its users.

In order to allow CASs to provide reputation data and access the reputations, we contribute a RESTful API because it helps organise a complex application into simple resources which it makes it easy for new clients to use the application, even if it is not specifically designed for them. The following REST API described in Table 1, has four resources: subjects, feedback reports, reputation reports, and events. A subject is the subject about which feedback or reputation describes, and is derived from feedback reports provided by an author. A feedback report can be associated with an event in which a subject took part. In more detail, an event is identified with a time and date range to which the feedback is pertaining to.

5 The Ride Share Application

This section describes a ride sharing application that allows drivers and commuters to offer and make request rides. These offers and ride requests include details about required travels, timing, locations, capacity, prices, and other details relevant for ride sharing. It performs automatic matching of commuters to available cars, by considering departure and destination locations, routes, capacity and reputation. The interactions of a driver and commuters differ and result in different outcomes. The following list describes the flow of interactions when a driver offers a ride.

1. Drivers and Commuters post *ride requests* to the server.
2. Matching is performed and some potential ride plans are generated based on the previously submitted ride requests from commuters that are matching the

Table 1 The reputation service's URIs

Action	Description
GET /subjects/	Get the URIs of subjects which have reputations
GET /subjects/:subject/feedback-reports/	Get the URIs of the feedback reports about the subject
GET /subjects/:subject/reputation-reports/	Get the URIs of the reputation reports about the subject
GET /subjects/:subject/feedback-reports/:report/	Get a feedback with a report identifier about the subject
GET /subjects/:subject/reputation-reports/:report/	Get a reputation report with a report identifier about the subject
POST /subjects/:subject/feedback-reports/	Post a new reputation report about the subject
GET /subjects/:subject/events/	Gets the URIs of events a subject is associated with (for example, in the ride share application users are associated with ride request id)
GET /subjects/:subject/feedback-reports/?event=:event	Gets the URIs of the feedback reports about the subject from an event
GET /subjects/:subject/feedback-reports/?author=:user	Get the feedback reports that is authored by a user about a subject
GET /subjects/:subject/reputation-reports/summary-latest	Get latest reputation report, which is the latest generated summary from the reputation service about a user

constraints of the ride request posted by the driver. This gives rise to *ride plans* that appear as potential ride plans both for the driver as well as for the commuters who have already submitted ride requests in the past. These ride requests have not been finalised (i.e. a ride record does not exist for them) and are matched to the ride request of the driver.

3. When at least one driver or commuter indicates their willingness to follow a specific ride plan the specific potential ride plan becomes a *potentially agreed ride plan*.
4. When all participants have expressed an interest in a potentially agreed ride plan, the driver selects one and attempts to finalise negotiation.
5. When all the commuters who appear in the driver agreed ride plan also agree that this is their selection among their driver agreed ride plans, the agreement has been reached and this gives rise to an *agreed ride plan*.
6. Together with the agreed ride and all the other ride plans that exist, both for the driver as well as the commuters, automatically become invalid ride plans for the specific requests that generated them.

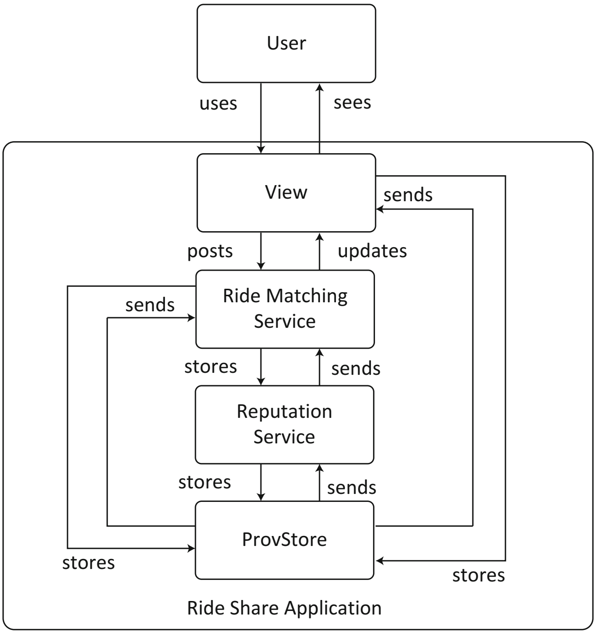


Fig. 2 Components of the ride share application

5.1 The Ride Share Architecture

The ride share application has five core components: a view, ride matching service, reputation service, and ProvStore[6] (see Fig. 2). The view provides the user with the graphical components with which to enter their ride requests, and to view and select potential rides. The matching service provides matches containing drivers and commuters, which the users can select. The reputation service is designed to store feedback reports and, generate and store reputation reports. The ProvStore is a specialised service for storing provenance using W3C PROV standard, the view, matching service and reputation service all send provenance data to it.

The components communicate using REST APIs. In more detail, we show the interactions between the ride share service, reputation service and the ProvStore in Fig. 3. The interactions use a REST API for the reputation and ProvStore services. The figure describes five interactions:

1. The ride share application requests the latest reputation of a user. It sends a GET request for the latest generated reputation of the user, and receives a JSON object containing the reputation. The reputation service also generates the provenance data recording this request and the outcome, and posts it to the

[6] ProvStore: https://provenance.ecs.soton.ac.uk/store/.

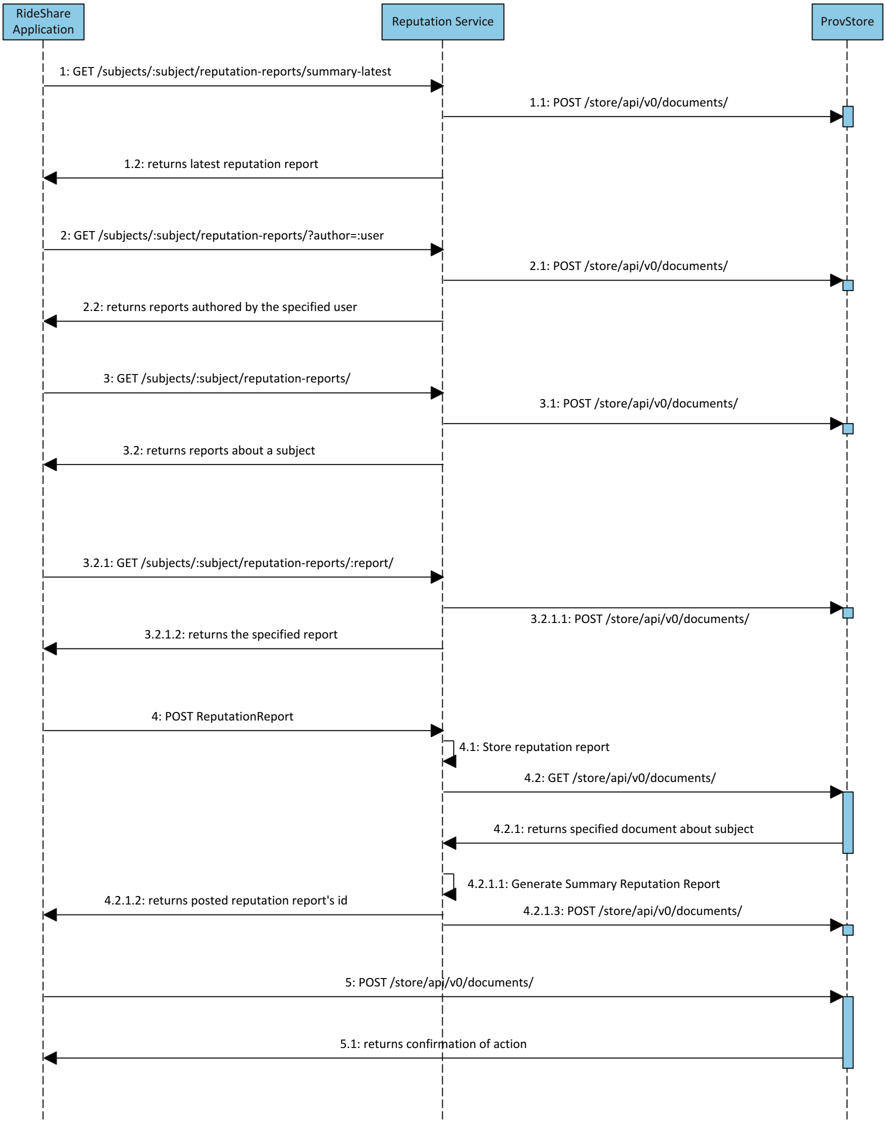

RideShare Application
Reputation Service
ProvStore
1: GET /subjects/:subject/reputation-reports/summary-latest
1.1: POST /store/api/v0/documents/
1.2: returns latest reputation report
2: GET /subjects/:subject/reputation-reports/?author=:user
2.1: POST /store/api/v0/documents/
2.2: returns reports authored by the specified user
3: GET /subjects/:subject/reputation-reports/
3.1: POST /store/api/v0/documents/
3.2: returns reports about a subject
3.2.1: GET /subjects/:subject/reputation-reports/:report/
3.2.1.1: POST /store/api/v0/documents/
3.2.1.2: returns the specified report
4: POST ReputationReport
4.1: Store reputation report
4.2: GET /store/api/v0/documents/
4.2.1: returns specified document about subject
4.2.1.1: Generate Summary Reputation Report
4.2.1.2: returns posted reputation report's id
4.2.1.3: POST /store/api/v0/documents/
5: POST /store/api/v0/documents/
5.1: returns confirmation of action

Fig. 3 Interactions between the ride share application, reputation service and ProvStore

ProvStore. This interaction may occur when a user requests to view the reputation of another, and the ride matching algorithm filters the potential rides.

2. The ride share application requests all feedback reports about a subject authored by a given author. This interaction occurs when the ride manager is matching drivers and users for rides, if an author rates the other participant (the subject) highly then this is more likely to result in a match. The reputation service

returns a JSON object containing the requested feedback reports, and sends the provenance data recording this request to the ProvStore.

3. The ride share application requests the feedback reports about a user, the ride share application sends two requests. The first GET returns the dictionary of reports describing a user, and the second GET returns a particular report. The reputation service sends the provenance data recording this request to the ProvStore.
4. The ride share application submits a feedback report. The reputation service stores the feedback report, and then generates a reputation report of the user. In order to generate the reputation of a user, it requests details from the ProvStore about the user. The reputation service posts the provenance data recording this request to the ProvStore.
5. The ride share application submit provenance data to the ProvStore. The ride share application posts the provenance data contained in a bundle to the ProvStore. This interaction occurs when the ride share application creates entities and performs activities on entities, such as receiving ride requests from users and generating ride plans.

5.2 Provenance for CAS

Provenance is a record of the entities, activities, and agents, that describes the flow of data across a single or multiple systems. In order for provenance to be traceable through heterogeneous systems, which may have their own ways of representing information, it is important to use a shared vocabulary. In order to support heterogeneous CASs, we present an ad-hoc ontology provenance for CAS, which defines a vocabulary for the classification of agents, entities and activities. Specifically, these three concepts are reused from PROV-O,[7] where:

1. A *prov:Entity* is a physical, digital, conceptual, or other kind of thing with some fixed aspects; entities may be real or imaginary.
2. A *prov:Activity* is something that occurs over a period of time and acts upon or with entities; it may include consuming, processing, transforming, modifying, relocating, using, or generating entities.
3. A *prov:Agent* is something that bears some form of responsibility for an activity taking place, for the existence of an entity, or for another agent's activity.

The ontology is expressed using the OWL2 Web Ontology Language. It provides a set of classes, properties, and restrictions that can be used to represent and exchange provenance information generated in different systems and under different contexts. It can also be specialised to create new classes and properties to model provenance information for CASs.

[7] PROV-O: http://www.w3.org/TR/prov-o/.

Table 2 prov:Agent classes and their descriptions

Class	Description
Machine	Identifies software components which run software. Subclasses of Machine include: **WebServer**, **WebApplication** or **Database**
User	Identifies agents which have a user role within a CASs. A user can be classified as: an **AdminUser** who has administrative privileges; a **GuestUser** where the user is not registered with a social machine; and a **LoggedInUser** who has registered with a user name and password. A GuestUser and LoggedInUser may be the same person at different times and applications, however we chose to differentiate between them because typically social machines allow logged in users different privileges to guest users. While it is important to understand the provenance of an agent, it is also important to be able to describe collectives of agents
Collective	Identifies a group of agents that are motivated by at least one common issue or interest, or work together to achieve a common objective. The Collective class denotes a group of agents, which can be composed of just one type, like Users or Machines, or a mix of Users and Machines. The notions of the dimensions or characteristics used to define collectives are largely undefined, with respect to social machines. Therefore, the collective subsumption is most likely to evolve with new research efforts. However, the notion of a collective is used when describing or comparing the outcome of a social computation. For example, in GalaxyZoo people classify with collective performance as good as professional astronomers [30]

- → prov:Agent
- → Collective
  - → Machines
  - → Users
  - → MachinesAndUsers
- → prov:Person
  - → User
    - → AdminUser
    - → GuestUser
    - → LoggedInUser
- → prov:SoftwareAgent
  - → Machine
    - → WebApplication
    - → WebServer
    - → Database

The PROV:Agent class is a parent to three subclasses: prov:Person, prov:SoftwareAgent, and Collective (see Table 2 for a description of key concepts).

The PROV:Entity class is parent to seven subclasses DataStore, prov:Location, MachineOutput, NegotiationOutcome, Plan, UserInput, and Utility, as shown in Table 3.

- → prov:Entity
- → DataStore
- → prov:Location
- → MachineOutput
- → Plan

Table 3 prov:Entity classes and their descriptions

Class	Description
DataStore	Identifies data repository entities of a set of integrated objects
prov:Location	Identifies the geographical location of an entity
MachineOutput	Defines entities that were created by a machine process, this process may transform the inputs to this activity
NegotiationOutcome	Defines entities that were produced from a negotiation, where a negation may result in an agreement, rejection or counter offer
Plan	Defines entities that are a detailed proposal for doing or achieving a goal
UserInput	Classifies entities that are inputted by users, these may include personal details, user preferences, feedback information such as votes or ratings, or user requests
Utility	Describes entities that have utility for social computation, such as price

Table 4 prov:Activity classes and their descriptions

Class	Description
PerformNegotiation	Describes negotiation activities
ProvideInformation	Describes activities that provide a prov:Entity
CompleteTask	Describes activities that are performed by a user or machine and may result in Plan, Utility, UserInput, and or Negotiation entities
PublishData	Describes activities which can be performed by a user or machine who publishes data, a user may publish their contribution, and a machine may publish the result of a computation
RunPlan	Describes activities that use a plan
StoreData	Describes activities which store data

  - → TransformativeInformation
- → NegotiationOutcome
  - → NegotiationAgreement
  - → NegotiationCounterOffer
- → UserInput
  - → PersonalDetails
  - → UserPreference
  - → FeedbackInformation
    - → Feedback
    - → Vote
  - → NegotiationInput
  - → UserRequest
- → Utility
  - → Price

The PROV:Activity class is parent to six subclasses (see Table 4):

- → prov:Activity
- → PerformNegotiation
  - → SubmitApproval
  - → SubmitCounterOffer

    - → SubmitDisagreement

- → ProvideInformation
    - → ProvideFeedback
        - → ProvideVote
        - → ProvideStarRating
    - → CompleteTask
        - → CompleteMirotask
        - → PlayGame
    - → CreateOriginalContent
    - → ProvideDeviceCollectedData
- → PublishData
- → RunPlan
    - → PerformDataManipuation
        - → Additive
        - → Subtactive
        - → Transformative
    - → PerformHouseKeeping
- → StoreData
- → RetrieveData

This ontology is used to describe the agents, activities and entities in the ride share application. For example, the ride manager is an instance of a *WebApplication*, a ride plan is an instance of a *MachineOutput*, and a matching activity is an instance of *RunPlan*. Defining the types of the items in a provenance graph supports social computation through querying and the reuse of provenance.

5.3 Provenance Example for the Ride Share Application

In order to describe the provenance generated by the ride share application, we run through an example where two users post ride requests and describe the provenance generated in each step.

1. Alice who is a driver wants to car pool because it saves her money on petrol. Therefore, she logs into the ride share application and creates a ride request with the details of the ride, including the departure and destination locations and times, how many seats are available and her preferences.

   The provenance recorded from posting a ride request is shown in Fig. 4. Alice who has the URI *rs:users/0*, posted the ride request *rs:rideRequest/#0*[8] to the *ride_manager*, who stored the ride request at *rs:rideRequest/0*.
2. Bob would like a ride, and he logs in to the ride share application with the username agent2. He adds his request to the application.

[8] The hash indicates that this entity isn't stored in memory.

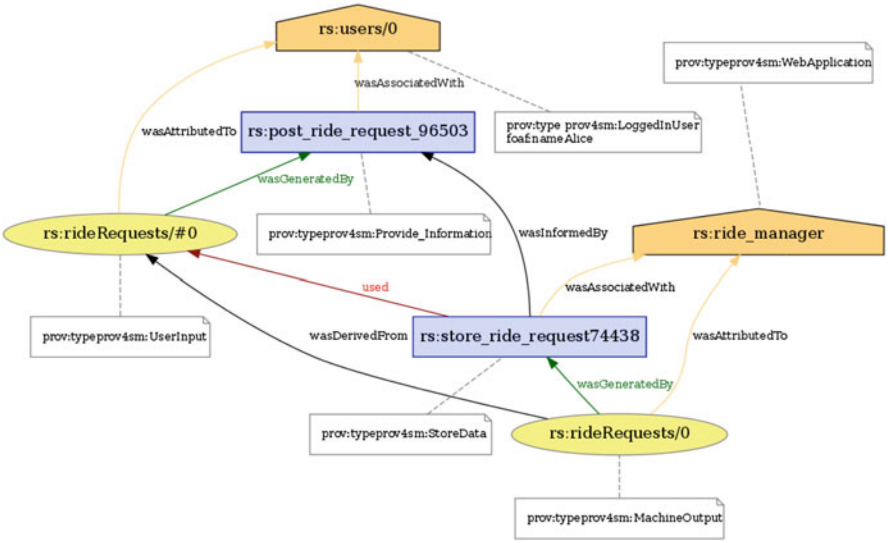

rs:users/0
wasAssociatedWith
prov:typeprov4sm:WebApplication
wasAttributedTo
rs:post_ride_request_96503
prov:type prov4sm:LoggedInUser
foaf:name Alice
wasGeneratedBy
rs:rideRequests/#0
prov:typeprov4sm:Provide_Information
wasInformedBy
rs:ride_manager
used
wasAssociatedWith
prov:typeprov4sm:UserInput
wasDerivedFrom
rs:store_ride_request74438
wasAttributedTo
wasGeneratedBy
prov:typeprov4sm:StoreData
rs:rideRequests/0
prov:typeprov4sm:MachineOutput

Fig. 4 A graph showing the provenance generated by Alice's ride request

The provenance recorded Bob, who has the URI *rs:users/1* posting the ride request *rs:rideRequest/#1* to the *ride_manager*, who stored the ride request at *rs:rideRequest/1*.

3. After each ride request is submitted, the ride share application runs a ride matching algorithm. It uses the information in the two ride requests Alice and Bob submitted because their departure and destination locations and time were compatible, it also used their reputation which was stored by the ride manager, to generate a match.

   The provenance recorded from this step shows that the matching algorithm used Alice and Bob's ride requests, and their reputations which retrieved from the *reputation manager*, to generate the ride plan, *rs:ridePlans/0*, which is shown in Fig. 5.
4. Alice and Bob can then view that there is a match. Alice views Bob's reputation and then accepts the ride.

   This provenance generated by Alice's acceptance is shown in Fig. 6. It shows that Alice viewed Bob's reputation *repser:subjects/28/reputation-reports/v/11* and accepted the ride plan *rs:ridePlans/0*, which changed the state of the ride plan into a potentially agreed ride plan *rs:ridePlan/0/v/parp*.
5. Bob then views Alice's reputation, and accepts the ride.

   This provenance which was generated by this acceptance, shows that Bob viewed Alice's reputation *repser:subjects/27/reputation-reports/v/2* and accepted the ride plan *rs:ridePlan/0/v/parp*, which changed the state of the ride plan into an agreed ride plan *rs:ridePlan/0/v/arp*, as shown in Fig. 7.

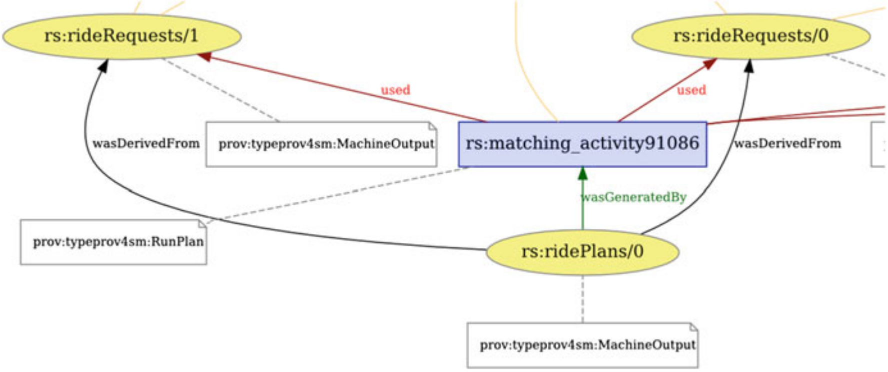


Fig. 5 The matching algorithm used Alice and Bob's ride requests

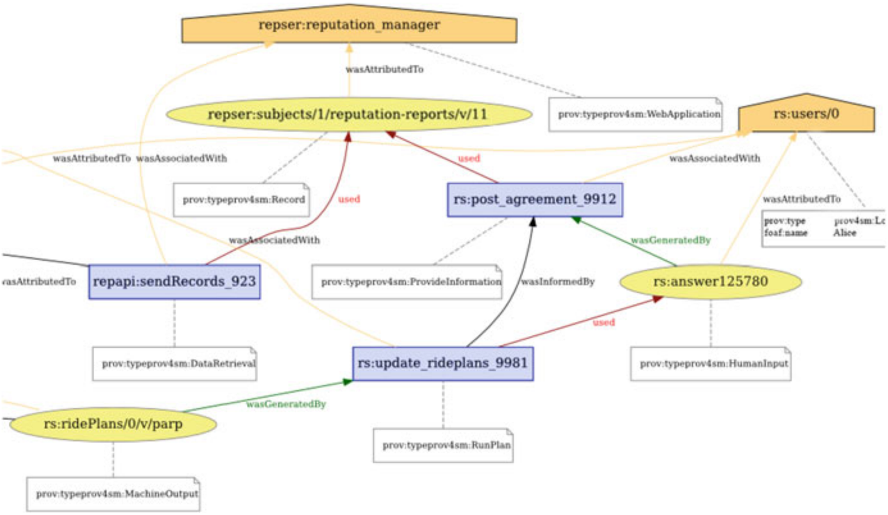


Fig. 6 A detail of the provenance generated by Alice's acceptance of a ride plan

5.4 Feedback and Reputation Reports

Once an agreed ride has taken place, the ride share application and users can submit feedback reports about the users involved in the agreed ride. These reputation reports are used to determine a user's reputation, thus once a feedback report is submitted the reputation manager generates a new reputation report.

In our ride share example, Bob fills in a feedback form rating Alice as a Driver, which in turn triggers the reputation manager to recalculate Alice's reputation.

The provenance generated by the submission of Bob's feedback, shows that the *rs:ride_manager* posts a reputation report *repser:subjects/27/reputation-*

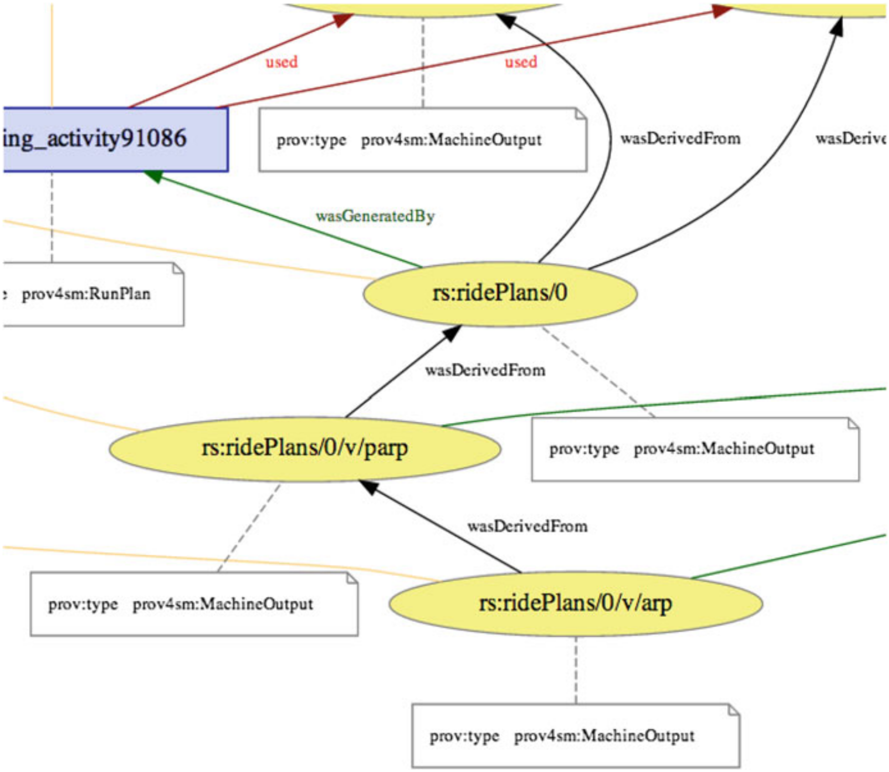


Fig. 7 The agree plan was generated

reports/#490 and is used to generate a reputation report with the activity *repapi:generateRepuationReports_1244*. This activity generates the reputation report *repser:subjects/27/reputation-report/491*, which is derived from the properties (such as the *#total_completed_rides*) in the report *repser:subjects/27/reputation-reports/490*.

The reputation report *repser:subjects/27/reputation-report/491* contains Alice's reputation in JSON format shown in Table 5, where the blue, pink, and light blue properties were derived from Bob's feedback report, a system feedback report, and the provenance recorded from the generation of this reputation report, respectively. The reputation report includes a *hasProvenance* property, which is a URI that links to the provenance of the reputation report. The properties in italics denote that they were totals derived from properties in feedback reports about Alice, and the properties in bold denote that they were averages derived from properties in feedback reports.

The provenance recorded in the steps in Sects. 5.3 and 5.4 allows users to perform audits, including:

Table 5 Alice’s reputation report *repser:subjects/27/reputation-report/491*

{ “report_type” : “reputationReport”,	Indicates the type of the report
“user_id” : 0”,	Is the subject’s user id
“total-stars”: 321,	Is the total number of stars a user has been reward with
“total_completed_rides”: 201,	Is the number of rides a user has participated in
“number_of_repeat_riders” 82,	Is the number of repeat riders the user has travelled with
“average_overallStarRating”: 4.5,	Is the average number of stars awarded as an overall rating
“average_ride_Price”: 3,	Is the average number of stars awarded for the ride’s price
“average_ride_Route”: 4,	Is the average number of stars awarded for the ride’s route
“average_ride_Car/Environment”: 2,	Is the average number of stars awarded for the environment of the ride
“average_ride_OnTime”: 5,	Is the average number of stars awarded for the ride being on time
“average_individual_Reliability”: 5,	Is the average number of stars awarded to a subject based on their reliability
“average_individual_Communication”: 4,	Is the average number of stars awarded to a subject based on their communication
“average_individual_DrivingSkill”: 3,	Is the average number stars awarded to a subject based on their driving skill
“average_individual_Friendliness”: 4,	Is the average number of stars awarded to a subject based on their friendless
“average_outsideFactors_Traffic”: 2,	Is the average number of stars awarded to a subject based on the traffic experienced during the ride
“average_outsideFactors_Weather”: 2,	Is the average number of stars awarded to a subject based on the weather experienced during the ride
“freeTextComments” :	Is a list of all the free text comments given by other users
‘[“Problems with traffic but otherwise fine.],	
“average_number_of_negotiations”: 6,	Is the average number of negotiations used to achieve a ride plan
“average_number_of_price_negotiations”: 12,	Is the average number of price negotiations used to agree upon a ride
“average_difference_in_price_negotiations”: 12,	Is the average difference in price used in negotiations used to agree upon a ride
“average_distance_deviated_from_original_route”: 5,	Is the average distance the subject agrees to deviate from their original route
“average_number_of_potential_rides_plans”: 5,	Is the average number of potential ride plans their ride offer was associated with
“average_number_of_potential_rides_selected”: 1,	Is the average number of potential ride plans were selected by other users
“hasProvenance” :	Is a link to the provenance recorded from generating the reputation report
“https://provenance.ecs.soton.ac.uk/store/documents/1665/”,	
“provenanceTimestamp” : “2014-01-30T21:00:00.250” }	Is a timestamp of the time the provenance was generated

1. Who created ride requests;
2. Which ride requests were used to generate which ride plans;
3. Who accepted and rejected ride plans;
4. How users' rides are generated and what influenced their generation;
5. How users' reputations were generated.

Moreover, it allows the users to overview the provenance of more than one service, the ride and reputation manager, which can be difficult in heterogeneous systems. It also gives the users the awareness that their actions are accountable, and it allows users to alter their actions so that they might improve their chances of ride matches and being selected by other users. Specifically, it supports the following use cases, which are derived from the general use cases presented in Sect. 3:

Ride Share Use Case 1. *The user wants to be able to make choices between available rides based on the participants and their preferences, reputation, and their opinion of them. This use case is a specialisation of use case 3.*

Ride Share Use Case 2. *The user wants to be able to analyse quickly the possible participants, and if the choice is too large then they should be filtered to include only rides that fulfil their preferences with users that have good reputations. This use case is a specialisation of use case 3.*

Ride Share Use Case 3. *The user wants to be able to understand why they were recommended particular ride matches, so that they can see which factors affected the recommendation, such as their preferences or reputation. This use case is a specialisation of use case 3.*

Ride Share Use Case 4. *The user wants to be able to understand how they are viewed by others in the ride share application, and which factors influenced this. This use case is a specialisation of use case 6.*

Ride Share Use Case 5. *The user wants to understand how their personal details are used by the ride share application system. This use case is a specialisation of use case 9.*

6 Summary

Provenance describes the flow of data across multiple systems, as we have demonstrated in the previous section with the ride and reputation manager. Moreover, provenance is independent of the technologies used in those systems executions. This is crucial because heterogeneous systems implemented by different developers or companies, and may each have their own way of representing information. In order to support heterogeneous CASs, we present the an ad-hoc ontology called provenance for social computations, which defines a vocabulary for the classification of agents, entities and activities for the ride share application.

Auditing private data processing is crucial so that authorities and system administrators can check its compliance with regulations. Provenance is a record of data entities, and it details how entities are created, modified and used, and this record can be used to audit these processes. In order to consume the provenance data generated by a CAS it must be retrievable.

Exposing provenance data increases public awareness and promotes accountability. Organisations are required to manage personal information in accordance with regulatory frameworks. However, there have been several cases where personal information has been leaked and exposed to unauthorised recipients. Future work will explore how to expose provenance without revealing the identity of users.

Reputation plays a core role in CASs, directly by interactions between users of social machines, and indirectly by influencing social computations. Therefore, in this chapter we provide a generic REST API for exposing reputation information.

Acknowledgements We would like to acknowledge the SmartSociety and SOCIAM projects. The SmartSociety Project is funded by the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n 600854. The SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/1 and comprises the Universities of Southampton, Oxford and Edinburgh. The authors would also like to acknowledge Avi Segal, Dimitrios I. Diochnos, Kevin Page, Kobi Gal and Michael Rovatsos, for their work on the ride share application.

References

1. Berners-Lee, T.: Design Issues: Cleaning up the User Interface. W3C Note, World Wide Web Consortium (1997). URL http://www.w3.org/DesignIssues/UI.html
2. Berners-Lee, T., Fischetti, M., Foreword By-Dertouzos, M.L.: Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor. HarperInformation (2000)
3. Chirita, P.A., Nejdl, W., Schlosser, M.T., Scurtu, O.: Personalized reputation management in p2p networks. In: ISWC Workshop on Trust, Security, and Reputation on the Semantic Web (2004)
4. Gal, K., et al.: Rideshare: A smart society application. Web Semantics: Science, Services and Agents on the World Wide Web (2014). URL https://docs.google.com/document/d/1cNhX-sW8pVyG5XYzrDNXzrn8R-NeKrzGszkDYJ3KwKQ/edit?invite=CPnQk7sG&pli=1
5. Gil, Y., Artz, D.: Towards content trust of web resources. Web Semant. Sci. Serv. Agents World Wide Web **5**(4), 227–239 (2007). DOI http://dx.doi.org/10.1016/j.websem.2007.09.005. URL http://www.sciencedirect.com/science/article/pii/S1570826807000376. <ce:title>World Wide Web Conference 2006Semantic Web Track</ce:title>
6. Golbeck, J.: Trust on the world wide web: A survey. Found. Trends Web Sci. **1**(2), 131–197 (2006). DOI 10.1561/1800000006. URL http://dx.doi.org/10.1561/1800000006
7. Golbeck, J.: Introduction to computing with social trust. In: Computing with Social Trust, pp. 1–5. Springer, London (2009)
8. Golbeck, J., Hendler, J.: Accuracy of Metrics for Inferring Trust and Reputation in Semantic Web-Based Social Networks. In: Motta, E., Shadbolt, N.R., Stutt, A., Gibbins, N. (eds.) Engineering Knowledge in the Age of the Semantic Web, Proceedings of the 14th International Conference, EKAW 2004, Lecture Notes in Computer Science, vol. 3257, pp. 116–131. Springer, Berlin/Heidelberg (2004). DOI 10.1007/978-3-540-30202-5_8. URL http://dx.doi.org/10.1007/978-3-540-30202-5_8

9. Golbeck, J.A.: Computing and applying trust in web-based social networks. Ph.D. thesis, College Park, MD, USA (2005). AAI3178583
10. Hua, H., Tilmes, C., Zednik (eds.), S., Moreau, L.: PROV-XML: The PROV XML Schema. W3C Working Group Note NOTE-prov-xml-20130430, World Wide Web Consortium (2013). URL http://www.w3.org/TR/2013/NOTE-prov-xml-20130430/
11. Huang, B., Kimmig, A., Getoor, L., Golbeck, J.: A flexible framework for probabilistic models of social trust. In: Social Computing, Behavioral-Cultural Modeling and Prediction, pp. 265–273. Springer, Berlin/Heidelberg (2013)
12. Huynh, T.D.: Trust and reputation in open multi-agent systems (2006). URL http://eprints.soton.ac.uk/262759/
13. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decis. Support Syst. **43**(2), 618–644 (2007). DOI 10.1016/j.dss.2005.05.019. URL http://dx.doi.org/10.1016/j.dss.2005.05.019
14. Koutrouli, E., Tsalgatidou, A.: Reputation-based trust systems for p2p applications: Design issues and comparison framework. In: Proceedings of the Third International Conference on Trust, Privacy, and Security in Digital Business, TrustBus'06, pp. 152–161. Springer, Berlin, Heidelberg (2006). DOI 10.1007/11824633_16. URL http://dx.doi.org/10.1007/11824633_16
15. Lebo, T., Sahoo, S., McGuinness (eds.), D., Behajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S., Zhao, J.: PROV-O: The PROV Ontology. W3C Recommendation REC-prov-o-20130430, World Wide Web Consortium (2013). URL http://www.w3.org/TR/2013/REC-prov-o-20130430/
16. Li, X., Lebo, T., McGuinness, D.L.: Provenance-based strategies to develop trust in semantic web applications. In: McGuinness, D.L., Michaelis, J., Moreau, L. (eds.) Provenance and Annotation of Data and Processes, Lecture Notes in Computer Science, vol. 6378, pp. 182–197. Springer, Berlin/Heidelberg (2010). DOI 10.1007/978-3-642-17819-1_21. URL http://dx.doi.org/10.1007/978-3-642-17819-1_21
17. Liu, L., Munro, M.: Systematic analysis of centralized online reputation systems. Decis. Support Syst. **52**(2), 438–449 (2012). DOI 10.1016/j.dss.2011.10.003. URL http://dx.doi.org/10.1016/j.dss.2011.10.003
18. Moreau, L.: Provenance-based reproducibility in the semantic web. Web Semant. Sci. Serv. Agents World Wide Web **9**(2), 202–221 (2011). URL http://eprints.soton.ac.uk/271992/
19. Moreau, L., Groth, P.: Provenance: An Introduction to PROV. Morgan and Claypool, San Rafael, CA, US (2013). URL http://dx.doi.org/10.2200/S00528ED1V01Y201308WBE007
20. Moreau, L., Missier (eds.), P., Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., Miles, S., Myers, J., Sahoo, S., Tilmes, C.: PROV-DM: The PROV Data Model. W3C Recommendation REC-prov-dm-20130430, World Wide Web Consortium (2013). URL http://www.w3.org/TR/2013/REC-prov-dm-20130430/
21. Moreau, L., Missier (eds.), P., Cheney, J., Soiland-Reyes, S.: PROV-N: The Provenance Notation. W3C Recommendation REC-prov-n-20130430, World Wide Web Consortium (2013). URL http://www.w3.org/TR/2013/REC-prov-n-20130430/
22. Pinyol, I., Sabater-Mir, J.: Arguing about social evaluations: From theory to experimentation. Int. J. Approximate Reason. **54**(5), 667–689 (2013). DOI http://dx.doi.org/10.1016/j.ijar.2012.11.006. URL http://www.sciencedirect.com/science/article/pii/S0888613X1200196X
23. Pinyol, I., Sabater-Mir, J.: Computational trust and reputation models for open multi-agent systems: a review. Artif. Intell. Rev. **40**(1), 1–25 (2013). DOI 10.1007/s10462-011-9277-z. URL http://dx.doi.org/10.1007/s10462-011-9277-z
24. Prat, N., Madnick, S.: Measuring data believability: A provenance approach. In: Proceedings of the Proceedings of the 41st Annual Hawaii International Conference on System Sciences, HICSS '08, pp. 393–. IEEE Computer Society, Washington, DC, USA (2008). DOI 10.1109/HICSS.2008.243. URL http://dx.doi.org/10.1109/HICSS.2008.243
25. Ramchurn, S.D., Huynh, D., Jennings, N.R., et al.: Trust in multi-agent systems. Knowl. Eng. Rev. **19**(1), 1–25 (2004)
26. Ramchurn, S.D., Huynh, T.D., Venanzi, M., Shi, B.: Collabmap: crowdsourcing maps for emergency planning. In: The 5th Annual ACM Web Science Conference, pp. 326–335 (2013). URL http://eprints.soton.ac.uk/350677/

27. Richardson, M., Agrawal, R., Domingos, P.: Trust management for the semantic web. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) The Semantic Web—ISWC 2003, Lecture Notes in Computer Science, vol. 2870, pp. 351–368. Springer, Berlin/Heidelberg (2003). DOI 10.1007/978-3-540-39718-2_23. URL http://dx.doi.org/10.1007/978-3-540-39718-2_23
28. Sabater, J., Sierra, C.: Review on computational trust and reputation models. Artif. Intell. Rev. **24**(1), 33–60 (2005). DOI 10.1007/s10462-004-0041-5. URL http://dx.doi.org/10.1007/s10462-004-0041-5
29. Shadbolt, N.R., Smith, D.A., Simperl, E., Kleek, M.V., Yang, Y., Hall, W.: Towards a classification framework for social machines. In: SOCM2013: The Theory and Practice of Social Machines (2013). URL http://eprints.soton.ac.uk/350513/
30. Shadbolt, N.R., Smith, D.A., Simperl, E., Van Kleek, M., Yang, Y., Hall, W.: Towards a classification framework for social machines. In: Proceedings of the 22nd International Conference on World Wide Web Companion, pp. 905–912. International World Wide Web Conferences Steering Committee (2013)
31. Smart, P., Simperl, E., Shadbolt, N.: A Taxonomic Framework for Social Machines. In: Miorandi, D., Maltese, E., Rovatsos, M., Nijholt, A., Stewart, J. (eds.) Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society. Springer, New York (2014)
32. Stein, K., Hess, C.: Does it matter who contributes: A study on featured articles in the german wikipedia. In: Proceedings of the Eighteenth Conference on Hypertext and Hypermedia, HT '07, pp. 171–174. ACM, New York, NY, USA (2007). DOI 10.1145/1286240.1286290. URL http://doi.acm.org/10.1145/1286240.1286290
33. von Ahn, L., Maurer, B., Mcmillen, C., Abraham, D., Blum, M.: reCAPTCHA: Human-Based Character Recognition via Web Security Measures. Science **321**(5895), 1465–1468 (2008). URL http://dx.doi.org/10.1126/science.1160379
34. Vu, L.H., Aberer, K.: Effective usage of computational trust models in rational environments. ACM Trans. Auton. Adapt. Syst. **6**(4), 24:1–24:25 (2011). DOI 10.1145/2019591.2019593. URL http://doi.acm.org/10.1145/2019591.2019593
35. Weitzner, D.J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., Sussman, G.J.: Information Accountability, pp. 82–87. ACM, New York (2008)

Part III
Applications and Case Studies

Surfacing Collective Intelligence with Implications for Interface Design in Massive Open Online Courses

Anna Zawilska, Marina Jirotka, and Mark Hartswood

1 Introduction

SCI is heralded as having a transformational impact on many domains, one of which is education [28]. While formal educational systems are expected to be slow to change [21], new opportunities for informing learning online are emerging and attracting considerable popularity. One opportunity is the platform called the massive open online course (MOOC) which attracts high course enrolment rates in the order of tens of thousands of teenagers and adults per course [16]. The capacity of MOOC infrastructure to draw and support such large user groups give the MOOC the potential to be developed into a significant SCI platform.

Currently, MOOCs are not purposefully designed [2] to be an SCI platform, and with SCI research in its nascent stages, foundational work is required. This foundational work should enhance our understanding of MOOC users, and ultimately lead to providing a critical lens for evaluating and prioritising ideas, as well as informing tools and practices.

The study described in this chapter contributes toward this foundational work by engaging with the learner community of a live MOOC to investigate whether or not an untapped collective intelligence exists, and therefore whether MOOCs might become an instantiation of SCI. Further, we hope that the elicited data will surface emergent important issues and requirements for interface design practice for these platforms to harness SCI.

A. Zawilska (✉) • M. Jirotka • M. Hartswood
University of Oxford, Oxford, UK
e-mail: anna.zawilska@cs.ox.ac.uk; marina.jirotka@cs.ox.ac.uk; mark.hartswood@cs.ox.ac.uk

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_9

2 Background

In this section, we provide a review of important prior literature and concepts which formalise the idea of untapped intelligence of online platform users, describe MOOCs in more detail, and argue for the pedagogical plausibility of developing the MOOC into an SCI platform.

2.1 The Cognitive Surplus

In education there are several different kinds of emergent socio-technical systems where SCI may play a role. ICTs to support learning are allowing new learning models to emerge, such as MOOCs and, as another example, the 'flipped classroom' [19] scenario where students learn factual information independently, and collaborate on performing tasks in the classroom. Each system has the potential to benefit in a number of ways from different forms of SCI. In this study, we concentrate on exploring the potential for the popularity of MOOCs to be harnessed for SCI, and therefore focus on SCI as part of a system of a large group of geographically dispersed learners connected online.

A useful concept when considering online platform user 'intelligence' is 'cognitive surplus'. It refers to those aspects of human reasoning which may remain untapped when people use online platforms such as, when playing repetitive online games or during the static viewing of video content [27]. This cognitive surplus may be associated with reasoning which humans perform better than machines such as, semantic reasoning [35].

For the MOOC, our starting point in this study is the hypothesis that a cognitive surplus exists in the form of the ways students relate to course content. By 'relate', we mean how the students make sense of the course content in terms of their own experiences and the body of knowledge they are attempting to acquire. They may (re)structure the course content as part of their learning, in terms of their personal goals, formal examination requirements, existing interests, and understandings. For example, content containing a narrative on how Internet technologies were developed may be viewed as affording an understanding of the technicalities of hardware and software development over time; or it may be viewed from an alternative, sociological perspective where it affords an understanding of the changes in society which may have affected their development. The significance of this relevancing work as a component of learning is discussed in Sect. 2.3.

Furthermore, the relations drawn by one student to content may be a resource for the relation creation of another. Given what we understand about the wisdom of crowds [33], we may also consider it likely that a collective of people with different attitudes and perspectives will perform better at subject-scoping than even a group of experts; one of the conditions of this being that the group is sufficiently diverse [33]. Diversity is important because it is suggested that groups which are too much alike tend towards an asymptote of collective 'learning' as less new information and fewer new perspectives are brought to bear on the base issue. Homogeneous groups tend to excel at one approach but tend to perform poorly at exploring alternatives. Therefore, in order to harness the SCI the platform may require affordances to not only collect but also make visible the relations (through appropriate visualisations) and then consolidate them.

2.2 Interactions in Current Massive Open Online Courses

Generally, there are two different kinds of MOOCs. The kind on which we focus in this chapter are those which have gained significant popularity in the last few years, so-called xMOOCs [23]. The second kind are termed cMOOCs [8], and were first offered many years earlier than xMOOCs, in 2008. They are based on a different pedagogical paradigm to xMOOCs: connectivism and networking. cMOOC enrolment rates are usually only a fraction of that of xMOOCs, therefore we focus in this study on xMOOCs (hereinafter simply referred to as 'MOOCs').

Currently MOOC platforms focus on broadcasting content and assessments and do not attempt to accommodate the relation-making process [2]. The MOOC interface is argued to have been developed without a clear goal for participation of students, primarily driven by the ideas of recreating an offline classroom setting online [24]. The focus of the courses remains on providing factual content and administering standard tests with correct and incorrect answers based on this content.

Coursera is arguably the leader in MOOC platforms, having hosted the first substantially popular MOOC in 2012 [26]. The platform provides the framework of functionality for the course, with the course content provided by academics at prominent universities. One instance of a MOOC is the case study chosen for this paper. This is the Coursera course titled 'Internet History, Technology, and Security' (referred to as 'IHTS' in this chapter for brevity [6]). The course ran over the period May-June 2013. The enrolment number for this course was 20,665.

Anyone who has an Internet connection with which to regularly access the platform may search and enrol in as many courses as they wish. There are no requirements on the student to participate in the course in any particular way to continue to be enrolled, although students are required to pass assessments if they wish to get a certificate of accomplishment for the course. MOOCs may sometimes use incentives and rewards to encourage the user to participate in a standard, pre-determined way. Incentives, combined with the use of assessments, reflect behaviorist theories of learning which are based on the idea that knowledge is acquired through controlled stimulus/response conditioning [36].

Research into participation in MOOCs has mainly focused on the collection of quantitative platform server log data that uses machine learning algorithms to create categories of student behavior [18]. These statistics are helpful in understanding some forms of collective behavior but are limited to describing activities and behaviors already captured by the system. Therefore, these statistics may not be of primary value when seeking insight into the cognitive surplus which is not harnessed by the platform. In this study, the primary concern is to unpick in-depth the relationships that are created between the content and a learner's experiences and thus, how a user relates to content. This lends itself more to qualitative analysis. Therefore, in this study we focus on collecting qualitative data.

2.3 Constructivism and Pedagogical Plausibility

In this section, we consider the pedagogical plausibility and value of exploring the development of MOOCs into an SCI platform. The goals of interface design for MOOCs depend strongly on the designer's definition of 'learning', and the kinds of activities and metrics which are defined to provide evidence for the learning process. Therefore, the goals of interface design are related to the kinds of learner-platform interactions which are considered 'correct'. The questions around the 'right' and 'wrong' forms of interaction are both a foundational and controversial issue amongst the members of the online educational community [1, 10, 20, 29]. We suggested in the previous Section that Coursera-type MOOCs engender a behaviourist pedagogical paradigm. We suggest now that, as an alternative to behaviourism, SCI may be more closely aligned with the constructivist educational paradigm. Constructivism is a more adaptive view of knowledge acquisition which focuses on discovery, reflection, and use of personal experiences when viewing new content. In this case, content is not characterised only by its facts, but provision is made for multiple representations of reality, for complexity, reflecting on experiences, context-focus, and collaborative knowledge construction rather than competition [9].

Proponents of constructivism and SCI researchers share the common goal of accommodating and leveraging individual relation-making processes. In the first case, proponents of constructivism view relation-making as an important part of learning. In the second case, SCI researchers view the individual ways a platform user views content as an important resource to leveraging a group's inherent collective intelligence [33].

The practice of relation-making implies that the scope of a learner's sense making and reasoning is not determined mechanically, but is actively constituted by the learner and influenced by their personal history. This understanding of perception is echoed in theories such as Goffman's Frame Analysis [14]. We draw upon the concept of 'frame' below to order the sorts of relation-making activity discovered in the data. Relation-making as described here may be a core activity within the educational paradigms: lifelong learning [11], constructivism [38], and deep learning [22], providing theoretical grounds for the pedagogical advantage of using MOOCs for SCI based on relation-making. In this paper, we aim to investigate in greater depth the idea that a significant untapped cognitive surplus exists in MOOCs via the empirical study explained in the next section. We use frame analysis in this study in a general sense to identify this research as one of a larger group which considers perception as a constitutive act, rather than perception as the ability to "address the true properties of the world, classify its structure, and evolve our sense to this end" [15] which reflects a more traditional and alternative understanding.

3 Surfacing Untapped Forms of Intelligence in a Live MOOC

In this section, we discuss some methodological issues regarding our data elicitation, as well as provide some preliminary observations from our analysis.

3.1 Online Survey and Nonprobabilistic Research

The chosen elicitation method was an online survey. While this method permits the collection of qualitative data from platform users, there can be issues around sampling and statistical generalisability for online surveys, including difficulty in obtaining a sampling frame [37] and self-selection bias because whether or not a participant replies to a survey depends on their initiative [30, 34]. These issues limit online surveys to be used for nonprobabilistic research only.

However, the statistical generalisability is not the primary concern of this study; rather it is investigating and revealing phenomena around participation in MOOCs, substantively under-explored, which drives future research. Nonprobabilistic research is useful for identifying new phenomena, causal processes, counter examples or additional examples about existing theories, conceptual frameworks, and phenomena. In so doing, this kind of research may illuminate new territories for thinking about issues. In addition, it is the first-step in identifying empirical data which extend current theory and from which other abstractions, models, frameworks, and theory may ultimately emerge. An example of previous non-probabilistic work which had a major impact on computer system design is Suchman's 'Plans and Situated Actions' [32] which drew upon fieldwork and ethnomethodologically-informed analysis [13].

3.2 Grounded Theory Sensitised by Frame Analysis

In our study, qualitative data was collected through open-ended questions to students with free text responses. The purpose of the open-ended questions was to reduce as far as possible the influence of questions on answers and to allow students to use their own vocabulary to illustrate their own understandings. To attempt to compensate for any ambiguity of the survey questions, the survey was piloted with 3 anonymous users.

The qualitative data was analysed using grounded theory [31] and thematic analysis [5], resulting in the emergence of themes and issues. Grounded theory was applied adaptively with the concept of frame analysis being a ‘sensitising concept’ [4] providing a starting point for further exploration.

3.3 Emergent Learner Frames

In this section, we discuss preliminary observations from our analysis, providing the reader with illustrative fragments of data.

While several open-ended survey questions were administered, in this study we focus on the particular survey question: *Where in your everyday life do you hope to apply what you learn in this course? How do you hope to apply what you have learnt?* The purpose of this question was to get students to reflect on the targets and goals of their participation in the MOOC. This question probes at the body of knowledge the learner hope to acquire, goals, interests, and understandings and therefore, we propose, characteristics of their ‘relation making’ practices. The response rate (rr) to this question was as follows:

$$rr = \frac{no.\ unique\ submissions}{no.\ unique\ students\ who\ viewed\ survey\ invitation} = \frac{670}{3596} = 19\,\% \qquad (1)$$

This is reasonable when compared to response rates documented from other online surveys. Email-only surveys have been reported to have a rate around 20 % [17]. Web surveys have had reported responses of as low as 2 % [25].

The themes and issues which emerged are explained next, along with illustrative quotes and a discussion on the interface design issues highlighted by each. We found that the emergent themes, characterised by a particular relation to content and an external context which informed this relation, lent themselves to being called a ‘frame’ [14]. The responses of many students fit into a frame, with some responses fitting into more than one frame. Some particularly illustrative quotes of responses categorised under each frame are provided below. Each quote, unless otherwise stated, is from a different student. The data in this paper provide a ‘snapshot’ and therefore will not attempt to infer how learners’ analytic frames change with time.

1. **The general interest frame**—The participant as the curious explorer, looking to expand current knowledge with non-professional interests in mind.

> “At 77 years of age, I’m just expanding my knowledge of the electronic world and my fascination with it.”
>
> “I’m glad to have an opportunity to learn more about this history, but I don’t see any particular application. Maybe some of the interviews will inspire me to pursue bright ideas. I think that, in general, the study of history is apt to be a useful pursuit, even when the specific applications to current affairs cannot be specified.”

> "Mainly I will use it like a general knowledge. It can be useful during interactions with my friends and [colleagues], especially with those working on computer sciences as I have a lot of them."
>
> "I hope to use it with my sons, who have computer degrees, in order to contribute to conversations with them."
>
> "So much of my life involves Internet technology that I am sure it will be consistently useful in work, home and education. I want to be able to explain this to my children before they get their first devices, and I want to be able to help explain security and other user issues to my not-so-tech-savvy parents."

Learners within this frame participate in the MOOC without a professional goal or sometimes without any readily definable focus at all beyond 'general interest'. These students appear to be motivated by how their learning can contribute broadly to their general knowledge, or by a non-focally motivated interest in the subject area. This does not mean such learning is necessarily purposeless, for instance, some students took the course so they could communicate knowledge about the Internet to family members and friends. Here students can be seen to be converting course content into social capital by acquiring forms of knowledge valued in their local social networks. These modes of engagement are a potential resource towards SCI if the platform were able to capture and share the ways learners engage with course content with an eye to its further transmission. MOOCs are thus revealed to be part of a broader learning network that goes beyond a dyadic relationship with the enrolled learner, illustrating the rich ways external social situations may come to bear on how a person relates to content.

The students' reasoning in relating the content to their real-world conversations is a resource for the collective: First, the student may be a channel through which other perspectives from conversation members are brought to bear onto the content, potentially further incorporating more diversity into the relation-making process. Second, the student may bring to bear on the content perspectives regarding the communication of IHTS course content in different social circumstances and in interactions with people of different backgrounds. The interface, in attempting to harness this reasoning, may provide mechanisms for the user to document and consider not only how the content connects to the interests of others with whom they are connected, but also the method of communication of the content.

The student who brings this frame to bear on the content may also benefit from an interface mechanism which provides searchable annotations and labels to content that particularly suit one kind of social interaction. For example, parents taking the course may wish to easily view those parts of the course annotated as useful for explanations about IHTS content to children beginning to use technology, and to contribute to this annotation by providing their own insights on this process to other students.

2. **The current profession frame**—The participant as seeker of knowledge for current professional work.

> "In my work with developing countries—I want to help them understand the origins of the technology and how they can use it to their advantage"

> "Some of my staff are responsible for our company's website, but I have no IT background at all. I am hoping this course will give me a better understanding of today's Internet and some of the issues our firm needs to prepare[d] to deal with."

> "I am doing research on hacktivist culture from a sociological, psychological perspective in my master degree at the moment. Furthermore I will use it for my PhD."

These students wish to apply the course content to their current professional contexts. One profession which is of particular interest is that of teaching. The data show two primary ways in which teachers responded that they may benefit from taking the course, each resulting in different interface design issues. Two illustrative quotes are shown next, each from a different participant.

> "I teach a course about cyberpsychology at the university I work in, and [this] is precious information for contextualising that material."

> "I teach Vocational English lessons at an ICT high school. I am an English Language Teacher who is interested in technology. So, I think, by learning a lot in this course, I can impress my pupils and set an example for them to dig deeper for their learning."

In the first quote above, we find evidence for teachers wishing to use the MOOC knowledge to enhance the courses they teach. In the second quote, the teacher is hoping to use their enrolment and participation in the MOOC as evidence to their students of the value of education. In this case, the platform user may want to bring something back to their real-world classroom which serves as motivation such as a qualification or new knowledge. To support this the enhanced MOOC could provide evidence of progress by revealing artefacts of learning, such as visualisations which track and display the progress of the student through the MOOC, which could possibly be made visible to others if the user so wishes. This frame suggests that the interface should not just facilitate learning but also provide information on and evidence of the process of learning enacted through the course content.

Taking the teaching profession as an example, the data illustrate some richer dynamics in the different ways the course content may be of value to the students, showing a more detailed reflection of how a user may come to view particular aspects of the course content as salient. The potential capacity for the collective to scope the content in terms of many of its possible needs in a particular profession may be an important part of its intelligence towards a particular subject matter.

3. **The future profession frame**—The participant as seeker of knowledge for future professional path.

> "I hope to apply [what I learn] from this course to my work with my own Internet startup"

> "I plan to [self-learn] AP Computer Science and I hope to gain insight on this rather interesting aspect of a subject I plan to delve deeper into next year. I hope to keep this course as general knowledge and to maybe, one day, be able to contribute or start a conversation in my academic future with the unique material in this course."

> "I hope to use this information in giving me a new direction in my career in the military. So far the lectures have been helpful in my understanding of the INTERNET and how it has become what it is now and how it will grow."

> "I plan to get back into the programming field after taking 20 years off to raise my family (and a few other things). What better place to start than with the history of what happened since I've been out of touch."

The responses grouped under this frame illustrate that the course content may be contrasted by students to practices in particular industries with attention paid to how the content may illuminate areas in these industries for future professional opportunities. Because these opportunities are not conceived separately from each student's own personal requirements, what counts as an opportunity will depend on the personal history and interests of the student, and what the student predicts will hold value for them in the future. Although there is a strong sense of individuality present here, sharing may broaden others' appreciation for what opportunities there are or pique their interest in a previously unconsidered career path.

Foresight and reasoning about opportunities in different industries may be better achieved by humans than machines, and therefore this data again suggest that by harnessing this frame SCI may contribute towards matching the benefits of human reasoning with the capabilities of computer systems.

4. **The privacy/security frame**—The participant as seeker of information for online privacy/security.

> "I hope to use what I have learnt in my classes about computing Internet and privacy as [an] introduction. Most of the time I stress about privacy and security of Internet"

> "I hope to apply what I learn from this course in my family's business, which involves home security and networking."

> "When I check into a hotel, I want to no longer worry about the security of the wifi because I will know exactly how it works."

> "Better understand the current privacy/terrorism prevention/wikileaks and snowden turmoils"

The responses here are characterised by a focus by the student on the part of the course concerned with online security and privacy. The data provide evidence for the student reasoning through these topics with regard to personal security practices and available software. To a lesser degree, responses also indicate an interest in understanding aspects of privacy/security online in order to understand related news stories popular at the time of the survey such as Wikileaks [12] and Edward Snowden [3]. The course content is framed in wider discourse and controversy which may be seen as involvement in processes of norm formation and the linking of the academic and the public realms by the student. Sharing this may help others understand and value personal security practices as well as assist the collective in scoping and collecting perspectives on current issues important to the public.

Many of the responses referred to privacy/security software. Because an important feature of software is its usability, which one may argue is best assessed through human rather than machine reasoning, the focus on software for this frame underscores the value of coordinating human agents to compile information about this collectively. Additionally, human agents may be able to discern the relation between context and preference, and a sense of what privacy/security risks exist in different circumstances; for example, Internet use in a home context may be regulated in a different way and with different implications for personal privacy than in a work context. This human reasoning regarding software adds a dimension to content on security and is therefore a resource for SCI. The focus on software also suggests that SCI may be harnessed by the interface by allowing students to annotate course content on software with links to software downloads.

5. **The everyday usage frame**—The participant as seeker of knowledge relating to everyday use.

> "Given that computers are and will continue to be a part of our everyday life, this course is giving me the historical background to understand the people and events that effect how I use computers today. Much like taking an American History course to understand how our country came to be, this course helps me understand how my computer usage came to be. Now, when I hear terms, I am not as bewildered as I was before taking this course. It is essential to understanding why I exist in this computer tech world."

> "I hope I can learn more about who is behind things like the browser I use, the programming of applications I games I use."

The responses in this frame are characterised by students expressing that the course content has inherent value because of the everyday necessity of Internet technologies in modern society and the widespread use of particular technologies such as the Internet browser. In these responses learners link the content to their everyday practices. Sometimes, they express the wish to link the software they use to the wider narrative behind the particular piece of technology to assist the learner to make sense of their participation in it. Shared perspectives of the everyday value of technology may inform others about the value of particular software, and may spark counter debates or discussions, and it may spark conversation about the value of software, its primary purpose, and shortfalls.

In these responses, the students appear to link the practical everyday software with wider patterns of meaning and an understanding of the forces that shape the technological landscape. It is the narrative behind a particular piece of technology which may assist the learner to make sense of their participation in it. This may be linked to broader spiritual or political discourse.

This frame suggests that content on Internet technologies may be supplemented using references to particular pieces of software which enact the principles/theory of Internet technology. Again as in the privacy/security frame, the usability, personal reviews and understandings of this software may be very important, underscoring the value of SCI expanding the course content by harnessing a collective of human participants.

6. **The historical trends frame**—The participant as the curious seeker of knowledge relating to historical trends.

   > "I believe that history goes in circles and sometimes in order to answer complicated questions of future we may need to look back into past. There is a chance that such a problem or similar was already solved."

   > "I hope to use the knowledge of the course to the analysis of technology and its impact in society. In history you can learn tendencies. I'm part of the Free Software Movement, and we're trying to construct the concept of popular technology in my city."

   > "One benefit of the course is sharpening ability to teach others about the power of independent, innovative thinking. Another benefit, based in particular on Week Four lectures, is practical and emphatic reinforcement of personal values and principles. Noble civic motivations do not always spring to mind in analysing the march of technology. Yet, a frequent theme embedded in much of the course dialogue so far is just that. How to make society better[?] The applications of this concept in daily life are endless."

   This frame captures the perspectives of students who are interested in what course content may indicate about the historical narratives of the development of Internet technology and what it may indicate about future trends both in technology but also more broadly in society. For example, the third quote illustrates a learner using the course content to reinforce narratives with a moral/ethical dimension ('personal values and principles') in technological development. Students sharing these perspectives may assist others in contextualising current events as the effect of technology on society increases, and create semantic arcs which may speak to discussion about the future of technology and society.

   The idea that students may prefer a different semantic ordering and arc of the course content than established by the course lecturer was mentioned in the discussion of the everyday usage frame, but is perhaps more clear for this frame since the data speak more directly to the issue of narratives. In this case, the student may be interested in particular changes over time and wish to focus on the characteristics of the transformation and change for example, the circumstances around a particular change; the drivers, and the results. This data illustrate again the value of SCI using human agents because creation and consideration of a different semantic arc through content is better performed by human agents rather than machines [35].

These data provide a strong sense that a richer negotiation of meaning by the students is underway as opposed to a mechanic progression through the course content in order to pass assessment and earn a qualification. Our analysis provides a 'snapshot' of some of the types of intelligence (for example perception, problem solving, and judgement) of the learners towards course content which is currently not harnessed by the design of the MOOC platform. A summary of the emergent frames is given in Table 1.

Overall, for each frame seen, the data reveal contrasting kinds of contexts of application which impart particular ideas of salience on the content. In beginning

Table 1 Summary of emergent frames

Classes	Action mechanism
The general interest frame	The participant as the curious explorer, looking to expand current knowledge with non-professional interests in mind
The current profession frame	The participant as seeker of knowledge for current professional work
The future profession frame	The participant as seeker of knowledge for future professional path
The privacy/security frame	The participant as seeker of information for online privacy/security
The everyday usage frame	The participant as seeker of knowledge relating to everyday use
The historical trends frame	The participant as the curious seeker of knowledge relating to historical trends

to accommodate and leverage these diverse relations to content, the interface may need to create an intuitive way for users to impart their categories on the content. In so doing, it may facilitate the creation by students of new semantic connections both within the course content (new semantic arcs/narratives through course content) and externally to other resources such as: descriptions of experiences, online resources like current news stories, and links to software downloads.

The revealing of detailed and rich modes of participation by the students in the course has shown the value of using qualitative research in computer system design, particularly for exploratory work which aims to reveal new territories for exploration. Analysis of the qualitative responses has informed some preliminary new features of the working models of the system user, providing potential for the development of an SCI infrastructure. The next section considers the kinds of implications which may be drawn from this data analysis for interface design of MOOCs.

4 Interaction Affordances for Learners Impacting Back on Course Content

Having discussed the results of data analysis, we now consider the implications of these on interaction design of MOOCs. The first implication is that the data seem to support the existence of relation-making and therefore a cognitive surplus. The second insight concerns the nature of relation-making which would be very difficult to replicate with an algorithmic computation because it depends upon people's ability to reason semantically, to be creative, and to function socially [35]. This leads to a distinct opportunity to create an SCI by blending human relation-making practices with the affordances of the learning platform. The third main implication concerns how the MOOC interface may be redesigned to afford SCI.

In particular, the data motivate the MOOC designer to orient away from solely focusing on standard tasks, ergonomics, and the broadcasting of fixed content to the learner. Additional affordances are needed to harness the cognitive surplus of learners in their relation-making for SCI. There may be many different mechanisms underpinning the affordances of SCI, such as discussion forums. The data suggest we may wish to develop mechanisms which focus on capturing and visualising the different ways learners relate to content. One approach, we suggest here, would be to provide an annotation or tagging feature in the MOOC interface. Other interfaces such as discussion forums may not suffice because they do not intuitively allow relations to be drawn between a variety of content in an easily visualisable and consolidatable way.

An example of how the annotation system may operate, mentioned previously, is that of the interface perhaps allowing students who are parents to tag and view parts of the course annotated as useful for explaining Internet technology to children, and to contribute to this by providing their own insights and experiences. As another example, course content may be annotated according to its perceived value in particular industries allowing inter- and intra-industry perspectives to be collected and consolidated as part of the collective‘s problem scoping and solving. A further approach would be to allow the learners to tag to parts of course content descriptions of experience, online resources like current news stories, and links to software downloads. In so doing, the students may remake the ways the pedagogical narrative is woven by the perspectives they bring to bear, impacting back on the way the course is presented. If the various annotations, and their possibly diverse features, can be consolidated, the SCI of the learners may be harnessed.

A future question is how this re-making of the course content may become a collaborative endeavour between enrolled learners and how different paths through the material can be made sharable and discoverable through search and browsing. This collaboration could include most of the enrolled learners, or smaller clusters of learners tied together by stronger links as suggested by the emergence of groups of students with similar frames in the data.

A number of issues arise in attempting to exploit cognitive surplus to take advantage of SCI. These include:

1. The ‘cold start’ problem

   There will likely be a less favorable balance between contribution and benefit for early adopters, and the critical mass of content needed to animate the SCI may never be achieved. This may be dealt with using incentive structures or gamification [7].
2. Navigating content

   In a functioning SCI there would be voluminous descriptions of relations that need to be navigated. Possible solutions for this would need to include carefully considered visualizations of relations. Intuitive navigation of relations along with incentive structures mentioned above may also motivate contributions by learners.

3. Managing diversity
   It is possible that the learner community may stratify into groups that have sometimes overlapping and sometimes disjoint interests, creating similar sorts of navigation problems to the above point. The issue of managing diversity would need to be addressed through MOOC design in such a way that useful alignments may be found without sacrificing serendipitous discovery. There may be computational approaches to managing diversity such as finding and making visible patterns in the annotations or relations.

In following on from this study, we would like to collect data of the relation-making process as close to *in situ* as possible through, for example, quasi-naturalistic experiments which capture aspects of learner-MOOC interactions including relation-making. If a satisfactory understanding of the practices of relation-making is achieved, it may be possible to develop interfaces and design interactions which support and harness relation-making.

5 Conclusions

MOOCs have emerged as a promising online socio-technical system which may be developed into an SCI platform. This study takes some of the first steps towards exploring this development by attempting to surface some form of untapped intelligence of learners in a live MOOC. Our focus is on the students' 'relation-making' which concerns the links they create between the course content and the real world as governed by their prior experiences and the body of knowledge they are attempting to acquire. Through the analysis of 670 free-text responses to an open-ended survey question, we suggest that relation-making is underway; a more complex negotiation of the course content by students as opposed to simply a mechanistic progression through course content which the platform is currently designed to accommodate.

The contributions of this study include providing one of the first collections of empirical evidence supporting the potential for MOOCs to be developed into an SCI platform. Furthermore, the data analysis motivate that the MOOC interface and interactions designer orient more towards designing new interactions through interface features which allow learners to actively shape the course content, such as an annotation system. By allowing students to create relations, we suggest the platform design would align with many of the practices associated with educational paradigms like constructivism, adding pedagogical plausibility to enhancing MOOCs in this way. There are many challenges to developing MOOCs into an SCI platform, including encouraging initial contributions by students and managing contributions should they become voluminous.

We end by suggesting that future work look more closely at the situated practice of relation-making and how an understanding of this might inform requirements for an annotation system which facilitates relation-making.

References

1. Balch, T.: About MOOC Completion Rates: The Importance of Student Investment (2013). URL http://augmentedtrader.wordpress.com/2013/01/06/about-mooc-completion-rates-the-importance-of-investment/?goback=.gde_3724233_member_203726319
2. Bates, T.: What's right and what's wrong about Coursera-style MOOCs (2012). URL http://www.tonybates.ca/2012/08/05/whats-right-and-whats-wrong-about-coursera-style-moocs/
3. BBC: Profile: Edward Snowden (2013). URL http://www.bbc.co.uk/news/world-us-canada-22837100
4. Blumer, H.: What is wrong with social theory? Am. Socio. Rev. **18**, 3–10 (1954)
5. Boyatzis, R.E.: Transforming Qualitative Information: Thematic Analysis and Code Development. SAGE, London (1998)
6. Coursera. Internet History, Technology, and Security. URL http://www.coursera.org/course/insidetheinternet
7. Deterding, S., Dixon, D., Khaled, R., Nacke, L.: From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp. 9–15. ACM, New York (2001)
8. Downes, S.: MOOC and Mookies: The Connectivism & Connective Knowledge Online Course (2008). URL http://www.slideshare.net/Downes/mooc-and-mookiesthe-connectivism-connective-knowledge-online-course-presentation
9. Duffy, T., Jonassen, D.: Constructivism and the Technology of Instruction: A Conversation. Lawrence Erlbaum Associates, Hillsdale (1992)
10. ECampus: Top 10 reasons for low MOOC completion rates (2013). URL http://www.ecampusnews.com/top-news/top-10-reasons-for-low-mooc-completion-rates/?
11. Field, J.: Lifelong Learning and the New Educational Order. Trentham Books, Stoke on Trent (2006)
12. Fildes, J.: What is Wikileaks? (2010). URL http://www.bbc.co.uk/news/technology-10757263
13. Garfinkel, H.: Studies in Ethnomethodology. Prentice Hall, Englewood Cliffs (1967)
14. Goffman, E.: Frame Analysis. Harvard University Press, Cambridge (1974)
15. Hoffman, D.: The Interface Theory of Perception. In: Object Categorization: Computer and Human Vision Perspectives by Ven Dickenson, Michael Tarr, Ales Leonardis, and Bernt Schiele. Cambridge University Press, Cambridge and New York (1989)
16. Jordan, K.: MOOC Completion Rates: The Data (2013). URL http://www.katyjordan.com/MOOCproject.html
17. Kaplowitz, M.D., Hadlock, T.D., Levine, R.: A comparison of web and mail survey response rates. Publ. Opin. Q. **68.1**, 94–101 (2004). URL http://www.uwyo.edu/studentaff/_files/docs/survey_calendar/kaplovitz_hadlock_levine_a_comparison_of_web_and_mail_survey_reponse_rates.pdf
18. Kizilcec, R.F., Piech, C., Schneider, E.: Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Proceedings of 3rd Conference on Learning Analytics and Knowledge. Leuven, Belgium (2013)
19. Knewton: Flipped Classroom—The Flipped Classroom Infographic (2014). URL http://www.knewton.com/flipped-classroom/
20. Kolowich, S.: Who Takes MOOCs? (2012). URL http://www.insidehighered.com/news/2012/06/05/early-demographic-data-hints-what-type-student-takes-mooc
21. Kolowich, S.: Why Some Colleges Are Saying No to MOOC Deals, at Least for Now (2013). URL http://chronicle.com/article/article-content/138863/
22. Lublin, J.: Deep, surface and strategic approaches to learning. Tech. rep., UCD Dublin (2003)
23. Moe, R.: MOOC or xMOOC? URL https://allmoocs.wordpress.com/2013/06/09/mooc-or-xmooc/#more-443
24. Morris, S.M.: Making Composition Massive: a #digped Discussion (2013). URL http://www.hybridpedagogy.com/Journal/files/Making_Composition_Massive.html

25. Petchenik, J., Watermolen, D.J.: A cautionary note on using the Internet to survey recent hunter education graduates. Hum. Dimens. Wildl. **16**(3), 216–218 (2011)
26. Round, C.: The Best MOOC Provider: A Review of Coursera, Udacity and Edx (2013). URL http://www.skilledup.com/blog/the-best-mooc-provider-a-review-of-coursera-udacity-and-edx/
27. Shirky, C.: Cognitive Surplus: Creativity and Generosity in a Connected Age. Penguin Press, London (2010)
28. Social-IST: High Impact Application Areas in Social Collective Intelligence (2013). URL http://social-ist.eu/high-impact-application-areas/
29. Spaniel, W.: Can We Stop Caring about MOOC Completion Rates? (2013). URL http://wjspaniel.wordpress.com/2013/05/27/can-we-stop-caring-about-mooc-completion-rates/
30. Stanton, J.M.: An empirical assessment of data collection using the Internet. Person. Psychol. **51**(3), 709–725 (1998)
31. Strauss, A.L., Corbin, J.M.: Basics of Qualitative Research: Grounded Theory Procedures and Techniques. SAGE Publications, Newbury Park (1990)
32. Suchman, L.: Plans and Situated Action: The Problem of Human-Machine Communication. Cambridge University Press, Cambridge (1987)
33. Surowiecki, J.: The Wisdom of Crowds. Random House Digital, New York (2005)
34. Thompson, L.F., Surface, E.A., Martin, D.L., Sanders, M.G.: From paper to pixels: Moving personnel surveys to the Web. Person. Psychol. **56**(1), 197–227 (2003)
35. Wang, Y.: On abstract intelligence: Toward a unifying theory of natural, artificial, machinable, and computational intelligence. Int. J. Software Sci. Comput. Intell. **1**(1), 1–17 (2009)
36. Watson, J.B.: Behaviorism. W W Norton & Co., New York (1930)
37. Wright, K.B.: Researching internet-based populations: advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. J. Comput. Mediat. Comm. **10**(3) (2005). URL http://jcmc.indiana.edu/vol10/issue3/wright.html
38. Yager, R.E.: The constructivist learning model. Sci. Teach. **58**(6), 52–57 (1991)

Who Were Where When? On the Use of Social Collective Intelligence in Computational Epidemiology

Magnus Boman

1 Introduction

1.1 Motivation

A burgeoning application area of social collective intelligence is computational epidemiology: the field concerned with all aspects of communicable disease save the purely medical ones. The reason is two-fold. Firstly, digital traces of human social activities abound, and such traces are purportedly useful for tracking disease in space and time. Second, the so-called race-to-trace employs new kinds of analytics of data notoriously difficult to make sense of, and social collective intelligence is purportedly useful for such analytics. Since the possible gains are enormous, measured quantitatively and qualitatively in diminishing the effects of disease and in reducing human suffering, many new technologies and tools have been implemented and deployed to assist the modern epidemiologist.

1.2 Methodology

A triangular approach will be used to investigate if and how social collective intelligence is useful to computational epidemiology. The first method employed is empirical, resting on own work and observations in both areas, and takes the form of a case study. The second method is theoretical, resting on inductive conclusions

M. Boman (✉)
SICS, Box 1263, SE-16429 Kista, Sweden, and KTH/ICT/SCS,
Forum 120, SE-16440 Kista, Sweden
e-mail: mab@sics.se

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_10

in turn based on the advent of new algorithms and theories for inference in massive data analytics. The third method is a largely deductive literature study, only relevant parts of which will be used here.

Arguments will be put forth below for the fact that modern epidemiology may benefit from social collective intelligence. The hypothesis is that social collective intelligence can be employed for assisting in converting data into useful information through intelligent analyses by deploying new methods from data analytics that render previously unintelligible data intelligible. The key observation is that new methods from data analytics allow for massive data analytics to stay in micro, providing the individual with tailored advice and relevant policies, without resorting to macro-analyses of the kind traditionally used in population studies and health economics. In economics, market approaches have been forced from treating customers as a flock of sheep into studies of distributions of customer behaviours, and recognising outliers [48]. There is a corresponding trend in healthcare that is offering more tailored medicine and other forms of customised medical assistance [43]. Personalised medicine is in this trend a way of empowering patients, offering them a say in medical decision situations, and recognising that their personal health data has a value.

In Sect. 2, the role of the computational epidemiologist is explained. This section rests heavily on the literature study. Section 3 shows how social collective intelligence can be linked to computational epidemiology by building a conceptual bridge between the concepts of *crowd signals* and *syndromic surveillance*. Because other chapters in this volume cover the ontological and the epistemological aspects of social collective intelligence, Sect. 3 instead turns towards relevant recent developments in complex systems, again resting on the literature study. Section 4 covers health data, and the analytics that today goes with it, with a special eye on syndromic surveillance. Section 5 presents a case study. While the earlier sections did benefit from empirical work, this section deals with it proper, by delving deep into means to fighting the spread of methicillin-resistant bacteria. This section revisits work previously only reported in short form [6] in 2006, and now with the focus on our hypothesis. The latter is investigated in the section that follows, which concludes this chapter.

2 Computational Epidemiology

Computational epidemiology (see [37] for a recent survey) is a promising and potentially very important area of research and development. It promises to save billions for tax payers as well as human lives, mostly just through deduction: making use of what is already there, making the invisible connections visible.

2.1 Models of Epidemics

While the traditional SIR (for Susceptible/Infected/Removed, cf., e.g., [2]) family of models—compartmental models based on differential equations—has proved immensely useful for the last century of fighting disease, strong arguments for complementing it by various other families of models have been put forth [21, 22, 26, 27, 35, 42]. The advent of new algorithms and tools within computational epidemiology has been a strong driver of this development (see, e.g., [20]).

Social collective intelligence and computational epidemiology are connected in that the latter is a form of spatiotemporal reasoning in which social link structures are employed. The social links can be part of macro-structures like demographics, population statistics, or organisational structures. These structures can in turn depict people mobility, political structures, power networks, etc. The smallest part of these macro-structures, the indivisibles of any model, are the individuals. In computational epidemiology, not just any social mechanism [28] is modelled, only those that have bearing on the spread of infectious disease. As a consequence, the relatively simple models of sociology are often replaced by more elaborate ones; in the last decade even by sophisticated multi-agent systems, allowing for considerable heterogeneity and fairly advanced studies of local (micro-)effects [7]. Alas, the computational complexity of executable micro-models also becomes forbidding, making simulations and sensitivity analyses a challenging engineering problem. One way of addressing this problem is to consider hybrid models: models taking into account micro-data (such as the geographical position and the personal health record of each individual), meso-data (such as the family structure of an individual or the properties of the neighbourhood of the dwelling of an individual), and macro-data (i.e., the structures mentioned above).

The micro/macro-distinction is sometimes used also in population biology of parasites to distinguish between parasites with direct reproduction within the host (microparasites) and those without (macroparasites) [1]. In the former category, most viral and bacterial parasites can be found. In the epidemiology literature, it is often argued that SIR models are appropriate for microparasites (see, e.g., [2], p.13). There is also the possibility of considering the different stages of an outbreak or a pathology in a multi-scale model, modelling the different phases in ways that have been deemed adequate; for instance, that the early stages of a disease outbreak is best captured by one kind of model, whereas the final stages, when the epidemic is panning out, are best captured by another [19, 54].

2.2 Shortcomings of the SIR Family of Models

A well-known shortcoming of SIR models is their proviso of homogeneous mixing: there is no heterogeneity within the population. To the computer scientist, this is bit of a mystery, not least because of all the efforts spent on multi-agent systems in the last 30 years. Protocols, languages, and ontologies for engineering cooperation and competition in populations where heterogeneity is key to successful teamwork have been laboriously worked out (see, e.g., [24, 29, 49, 52]). Likewise puzzling is the proviso of total susceptibility within the population. For most pathogens, there is some form of residual immunity in any real population, but because this immunity is all but impossible to estimate empirically, it is set to zero in SIR models. This can be seen as a form of caution among epidemiologists, by reverting to the worst case. A third shortcoming is the normalisation of outcomes of SIR model analyses, pivoting around the concept known as R_0: the number of people the disease is passed on to from an infected person (again, assuming full susceptibility). If analyses produce an R_0 inconsistent with earlier empirical observations, infectiousness or other factors are routinely adjusted. To the computer scientist, this is not validation, but reverse engineering. It is also a case of questionable reductionism, since R_0 is not a primitive concept but a higher order measure determined by a number of primitive concepts. The latter is a parameter space covering prevalence, infectiousness, morbidity, mortality, and more. To the theoretical scientist, this parameter space could (in theory) be understood in full through complex systems analysis [46]. To the engineer, such an analysis is not much different from analysing the parameter space of, say, the elements of a supply-chain in a car factory.

2.3 The Computational Epidemiologist

The computational epidemiologist—arguably the closest instrumentalisation of a Sherlock Holmes in contemporary society—faces the task of unifying the methodology of the epidemiologist, the scientist, and the engineer. Ideally, the epidemiologist is assisted in the race-to-trace by the wisdom provided by computational analyses. The computational epidemiologist can also have another role, addressing policy makers. Because the latter govern epidemiologists in some sense, and must understand at some level the results of epidemiological analyses in order for governance to be effective, the former can assist with such understanding [14].

3 Social Collective Intelligence

Rather than surveying the entire area, only some cross-sections of social collective intelligence methods will be brought forth here. Of the methods relevant to computational epidemiology, a partitioning into passive and active methods is the most straightforward to carry out. Passive methods include surveillance, for example. Here, crowd signals will serve as the example of passive collective intelligence, and its computational epidemiology counterpart will be syndromic surveillance, as will be explained in some detail below. An example of active methods would be health status reporting. This comes in several forms. Many individuals today use apps and mobile services to keep track of their own health and well-being. Some of the data gathered, and the information resulting from making sense of that data, is fed back to health authorities and healthcare providers [45]. In most countries, general practitioners and laboratories follow policies, practices, rules, laws, and clinical guidelines for reporting diseases that are listed on national or regional registers over diseases mandatory to report. Internationally, there are also dozens of systems in use for analysing health status reports and lab reports, turning them into intelligence then delivered to epidemiologists and policy makers via user-friendly services like automated email digests [15]. Finally, some countries allow for extensive simulation studies built either on registry data [12, 13] or on census data and synthetic populations [40].

Another development worth mentioning is the employment of artificial intelligence in health. Starting in expert systems and moving on to knowledge-based systems, engineered decision support is now being pushed into care-giving institutions with sales claims like 90 % of nurses following machine-generated advice [53]. In this case (IBM's Watson), the prospect of having many instances of the system to assist the computational epidemiologist is thrilling, not least in view of the data analysis required for passive micro-data. Moving some of the analysis from the human analyst closer to the data source, as in pre-processing or intelligent stream analysis, would arguably (see, e.g., [23] for a contrasting view) buy the human analyst more time in the race-to-trace. This development requires at least a thousand-fold increase in the number of instances of Watson's or similar implementations, as well as considerable development in the methodology and use of human-machine analysis interplay. This would need to include investigations into ethics, IPR, and professional conduct and responsibilities. These obstacles are not necessarily delaying the prospect of a many-Watson system assisting the human epidemiologist, however, and the collective producing social intelligence may in the future include artificial members.

3.1 Crowd Signals

A *crowd signal* is an indicator with its value derived as a side-effect from large groups of people performing tasks. For example, if one seeks to monitor the spread of some infectious disease, one may rest on reports from hospitals and primary care units. This is known to medical professionals as *disease surveillance*. But in recent years, *syndromic surveillance*—the collection and integration of other relevant data—has also proved useful; such as harvesting twitter posts [17, 18] or Web search queries with the appropriate keywords [31]. In the case of monitoring influenza, one would perhaps use "cough", "fever", and "influenza" as keywords for this purpose. A crowd signal may also be an indirect trace of peoples' activities and not just a monitoring of some activity like search. How people have moved geographically, and how people have profiled themselves on various kinds of social media platforms can also be used to the same end.

3.2 Complex Systems

Thanks to recent advances in the area of complex systems, crowd signals can be modelled and understood with enough efficiency for them to be adopted for use in practice. In fact, the dynamics of complex systems nearing a critical point have generic properties that unify and transcend all areas of application, albeit near impossible to observe or measure. In recent years, however, new means to understanding such transitions have emerged. The concept of a detected outbreak in epidemiology, or a systemic collapse point in population studies or ecology can now be grouped together in various classes of bifurcations. By studying the resilience of a system, recovery rates around the time of a bifurcation can provide important means to measurement of severity. At times before a collapse, such recovery rates may act as early signals. Crowd signals here constitute a class of early signals, of particular importance. At times after a collapse, recovery rates can measure resilience and shock severity, e.g., through autocorrelation. This theoretical advance makes possible new kinds of systemic studies.

In practice, many systems handling information and communication that we normally rely on break down at the time of a serious crisis. These systems include Internet connections, telephone networks, and the services provided by government and industry. The information available is also more uncertain than we are accustomed to. What is ultimately threatened in such situations is the resilience of society.

In theory, the resilience of a system can be measured as size of a basin of attraction [30]. Two points F_1 and F_2 on an equilibrium curve are always dependent on system parameters, which means that there is no guarantee for a system to stay in a state in which catastrophic events cannot occur, i.e. thresholds are never fixed values [46]. So-called fold bifurcations, in which F_1 and F_2 are points on a folded curve, constitute a class of bifurcations that may be used to describe systems that are in some sense vulnerable to perturbations. While a strong resilience means that the basins of attraction are wide or deep, this also entails that a return to a stable state is difficult. If, on the contrary, resilience is weak, even very small perturbations can cause critical state changes [47]. For most interesting systems, there are many stable states, and the return to any of them is obtained through a positive feedback loop. An example widely used in the social sciences is when positive feedback equals wealth, and the system has two stable states of Rich and Poor. Parental income and wealth have been proved to be very strong indicators of at least an American individual's wealth [9], something that traditional economic theory has large problems with explaining (cf. [46]). Some systems are so complex that they are never stable. Instead, they repeatedly suffer critical shifts, sometimes converging asymptotically towards a cycle (forming another class of bifurcations, viz. the *Hopf bifurcations*), or even towards a chaotic state [44]. To the computational epidemiologist, one of the most important kinds of system is of the cycle kind with periodic oscillation of the environment. Seasonal influenza (in the temperate zones) constitutes the best example of so-called *periodic forcing*. The interplay between, on the one hand, a Hopf bifurcation system with periodic forcing, and on the other hand the perturbations and indicators of pandemic influenza provides a great research challenge. Crowd signals and other forms of social collective intelligence can play a role in meeting that challenge.

4 Health Data

The recently almost all-encompassing interest in data analytics has prompted investigations also into how to assist public health and well-being by analysing health data. One must begin by understanding what sort of data is available, how to analyse it, and how to deal with its sensitivity. The scope will here be narrowed down to questions pertaining to the hypothesis.

4.1 Availability

Big data can be pragmatically defined as data impossible to analyse directly, due to its size. The vast majority of health data is structured, and while electronic health records can be long and numerous, the digital handling and maintenance of the documents is not a big data issue. Medical imaging data may be unstructured or semi-structured, and thus hard to use for analysis if databases are massive, but this is more of an exception than the rule. By contrast, much of the data about individuals not directly but indirectly relevant to their health and well-being is unstructured and definitely big data. An incomplete list would include data on an individual's use of transport, purchases, active memberships, phone calls, tweets, blog posts, and forum comments. Health data, by contrast, only comes in the following forms [4]:

- Image data
  - Medical imaging
  - Microscopy
- Sequence data
  - Transcriptomics (RNA)
  - Epigenomics (DNA)
- Text data
  - Electronic health records
  - Scientific publications

To the computational epidemiologist, the engineering challenge when it comes to data thus lies chiefly outside health data. Since most of the big data is on individuals (micro data), just as in epidemiology, the promise of using micro data for analyses has a number of appealing properties when compared to using population (macro) data.

4.2 Analytics

Disease surveillance is performed for both communicable and non-communicable disease. Case reports (and lab reports) are essential parts of traditional disease surveillance. Syndromic surveillance adds data originally collected for other purposes. Recent advances in ICT have created new opportunities for syndromic surveillance, and many systems have been developed to take advantage of the new sources of data. Efficiently interpreting the combined output collected by these systems, however, remains an open problem. In many cases, the populations surveyed by the systems differ significantly, complicating the application of traditional statistical (macro) methods to analyse the collected data on individuals (micro).

Conceptually, syndromic surveillance systems can be divided into three parts: data collection, analysis, and reporting. *Data collection* includes lists of available data sources, collection methods for each data source, additional formatting of the collected data, and storage solutions. *Analysis* contains all the methods used to extract additional signals from the collected data. These methods include temporal, spatial, or spatio-temporal methods, as well as machine learning applications for larger data sets. Analysis of syndromic surveillance data typically aims at detecting outbreaks, shifts in long-term trends, or everyday monitoring of ongoing infections in the population. The final component, *reporting*, includes all the ways in which the analysis results are communicated to interested parties. The results can be presented in many forms: numerical output from statistical analyses, incident plots displaying exceeded thresholds, "heat maps" coloured to indicate different levels of observed activity, or even simple notices instructing the experts to check a data source for further information. The medium of reporting is also varied: messages can be transmitted via email, SMS, regular updates on a Web page, or dedicated display units placed at institutions tasked with monitoring.

Data sources (or indicators) are the most important part of any syndromic surveillance system. In the first stage of development, the availability of the sources must be determined. There are a wide variety of sources, and their availability depends on the local context of the project, with regards to existing laws regulating data access and privacy access, as well as practical concerns such as ensuring sustained connectivity to the data sources. Data most often used for syndromic surveillance include [16]:

- emergency department visit chief complaints;
- ambulatory visit records;
- hospital admissions;
- over-the-counter (OTC) drug sales;
- triage nurse telephone calls;
- emergency hotline (112) calls;
- work or school absenteeism data;
- veterinary health records;
- laboratory test orders;
- laboratory test results;
- infectious disease case databases.

The analysis methods in infectious disease informatics generally detect anomalies, and gradual shifts in trends. Temporal methods try to explain how the data evolve over time. Spatial methods aim to do the same over geography, often corresponding to the jurisdiction borders of local health authorities. Spatio-temporal methods combine the features of both approaches. If a spatial method is used, the same variable can also be displayed on a map to illustrate the geographical results of the analysis. The challenge remains, regardless of the kind of analysis: to extract, integrate, and visualise huge amounts of information to first responders, in a timely manner, without compromising the quality of the information.

4.3 Sensitivity

Composing mobility models is not straightforward, since data sources are heterogeneous and usually noisy. This not only in technical terms, i.e., with respect to spatial and temporal scales, accuracy, and statistical skewness, but also in terms of degree of potential privacy intrusion [32]. Such intrusions may be legally or otherwise perceived by stakeholders (or public opinion) as controversial [45]. Just a few years ago, social collective intelligence via location-based services or individual service-usage patters existed almost exclusively in research labs; today such services are used by millions, and monitored by thousands of stakeholders. This gives researchers an enormous opportunity to explore how people perceive and construct space, and to understand the connection between digital and physical places. The potential of using user-generated content and its metadata is clear [39], but analysis of local social media is still in its very early stages. In particular, network operator data (such as call data) by contrast consists of traces of location data points generated without explicit user involvement. Using this data for indicators or for gathering intelligence, requires a micro-level understanding of the user actions that have resulted in the data points, as well as the reasons for noise, bias, and skewness.

People mobility data gathered from social- and access networks are typically massive, but also highly structured [50, 51], since most people move in a conservative manner; e.g., with respect to travel distances and to locations [25]. This behaviour has been captured by diffusion-based models such as Lévy flights that have been used to describe indicators for human mobility, e.g., bank note dispersal [11].

5 Case Study

A case study from Stockholm will serve as a basis for empirical observations relevant to the hypothesis. Some of these empirical observations were made also in other work contexts. All observations will be scrutinised in the final section of this chapter.

5.1 Visualising Contact Networks for MRSA

A tool for visualising contact networks will be presented. It generates interactive 3D network visualisations. Its general purpose visualisation engine can support multiple applications and varying pathogens. The main purpose is to trace, in the case of an outbreak, contacts among individuals known to have been at the same place.

Several software tools for contact tracing were available to the epidemiologists constituting the user group, but these tools used rudimentary and flat structure displaying methods, such as lists and reports. This fact meant that epidemiologists resorted to lots of browsing in physical paper binders to search for information that could lead them to useful constructions of contact networks. The new tool provided, named *asimplot*, generates interactive 3D network visualisations. Its general purpose visualisation engine can support multiple applications and varying pathogens, and is used in this example for a contagious infection called methicillin-resistant *Staphylococcus aureus* (MRSA) [38]. *S. aureus* is a bacterial species commonly found on the skin. If it gains entrance through breaches in the skin it sometimes causes skin and soft tissue infections. MRSA is a variant of *S. aureus* carrying resistance towards penicillin antibiotics and all antibiotics chemically related to penicillin (so-called *beta-lactam antibiotics*). MRSA has been called the world's biggest problem regarding nosocomial (hospital-acquired) infections. Epidemic spread within a healthcare system is a major issue: not only does it defy the purpose of the system; it also demands a costly effort for remedy when the spread crosses a certain threshold [33].

5.2 Data and Models of MRSA

In healthcare systems, registers and databases are maintained for administrative and economic purposes, which contain information on when individuals are in contact with the healthcare system as an in- or outpatient, and for how long as an inpatient. This gives a unique opportunity to map when individuals have made contact by visiting the same outpatient clinic, or by being admitted to the same hospital ward simultaneously. This type of detailed data reflecting contacts between individuals and relevant to the transmission of infectious agents does generally not exist for other parts of society.

Observation A: In the race-to-trace, epidemiologists need information on "who were where when?". While healthcare systems provide some data on this, the granularity is notably larger than desired. Passive surveillance of individuals can provide epidemiologists with more accurate information. It is possible that patients will endorse or encourage (even active) surveillance, if the added value of improved health and safety in the hospital is demonstrated and explained to them.

At the time of the Stockholm study, there were a few surveillance tools for monitoring nosocomial infections, but they were all for use in single-hospital environments [3, 8]. Thus, they failed to diagnose an infection as nosocomial if it originated from another hospital. In countries such as Sweden, which maintains country-wide patient data in high-quality registers, it is possible to do surveillance at a higher level. A reasonable ambition is at least city-wide contact tracing.

Observation B: Even for a relatively small local outbreak, the area to survey and model must be extensive if prediction, or any kind of prescriptive advice, is to be of high quality. Since systems developed for healthcare often have a limited scope (hospital, ward, geographical region, political region,. . .), mash-ups and amalgamations with systems developed for other purposes can be employed in order to get better coverage.

Since the aims of epidemiology is to provide early warnings, policy advice, and other kinds of information to mitigate spread, a related study on the cooperation between policy makers and computational epidemiologists is also of relevance here:

Observation C ([14]:220, references omitted): All population data sets are regional. To have access to data on the entire population on the planet is not a realistic goal. Hence, most studies are limited to one geographic region, such as a city, a state, or a country. This means that the universe of discourse includes not only the individuals in this geographic region, but also that a certain proportion of the individuals must be allowed to leave the region. Moreover, visitors and immigrants from other regions should be included in the population data. Some computational epidemiology projects employing micro data use census data, others extrapolate from samples, and yet others use synthetic data. In the rare cases where registry data is available for a large population—as is the case for the Swedish population—hard methodological questions must still be answered regarding the generalisability of results: which parts of a scenario execution in Sweden are likely to be analogous to ones in Norway, Iceland, or the state of Oregon?

Scientific visualisation libraries usually have limited functionality and scope regarding network visualisations. Being a specialised field [10], network visualisation have enjoyed mature and stable software packages, such as Pajek [5], for some time. Pajek is recommended for large network analyses, but was developed for producing static outputs (cf. Fig. 1). Pajek also has a relatively slow learning curve.

In the event of an MRSA spread within a hospital, where a patient possibly infects another while they are admitted to the same ward, either the first patient was tested positive before he or she was admitted to the ward, or not. The latter possibility is due to the fact that infected individuals can spread the disease before their diagnosis. Both individuals who have a disease caused by MRSA (such as a skin infection) and individuals who are just carriers, can spread it to others by direct and indirect contact. Therefore, the term infected here refers to the transmission of MRSA regardless of if the individuals involved in the transmission are diseased by

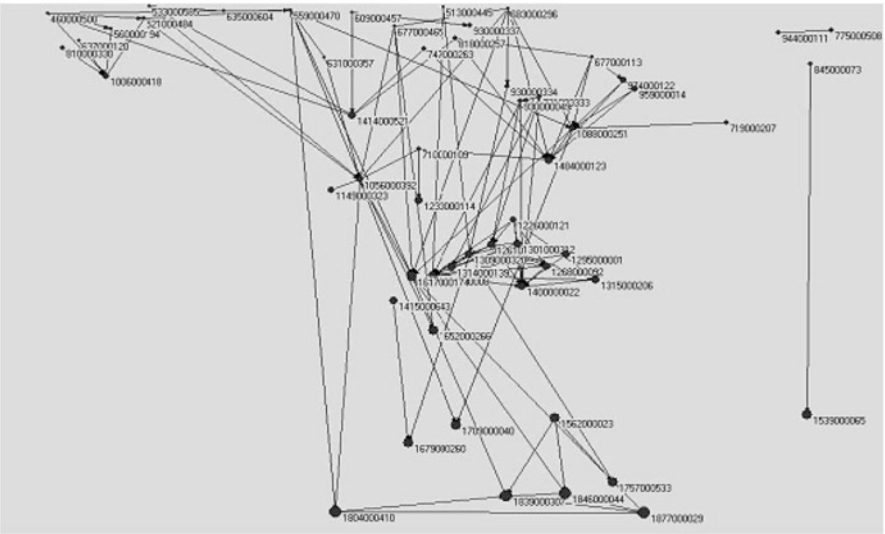


Fig. 1 Contact network of MRSA patients drawn by epidemiologists at Karolinska Institute, Stockholm, using Pajek, and adjusted manually

MRSA or just carry the bacterium. Patient log data is available for the healthcare system in Stockholm county. This healthcare system consists of several hospitals and outpatient clinics. Each hospital consists of many clinical departments, and each clinical department can have several wards. The log is maintained at the ward level. A new log entry is made if a patient moves from one ward to another, within the same clinic or otherwise. The data is complete in the sense that within the system a log entry is always made when a patient moves between wards. However, it must be recognised that a patient can infect other patients outside the healthcare system.

5.3 Design, Implementation, and Evaluation

The users evaluated the tool on real data. Due to the sensitivity of this data, real data could not be used for testing and developing asimplot. Therefore, several test data sets were constructed using the statistical parameters calculated from the real data. The usage of statistically fabricated data assisted in avoiding any over-fitting to the real data set, but the extent of success can only be validated once the tool is utilised with several real data sets, and no such validation attempts were made. In MRSA, one real data set is not representative of any other real data set, since the regional variations are large.

The stakeholder in this case study was an epidemiologist, and in short, the user requirements specification contained the following constraints on asimplot.

1. Visualise contacts between agents as a network visualisation.
2. Discriminate type and level of contact.
3. Discriminate strength of contact.
4. Make nodes (and edges) moveable and selectable.
5. Discriminate (agent) disease type, if so required.
6. Map axes dynamically on (agent) properties, and expose all data fields in the MRSA data set to this mapping, along with additional derived fields.
7. Give a multiple-level view based on the three types of MRSA used in the standard classification.
8. Provide an easy way of inspecting key properties of selected nodes and edges.
9. The tool should be executable with acceptable response times, on a standard personal computer system.
10. There should be no unreasonably slow or jerky movements for medium-sized data sets of around 500 agents.

Through analysis of the requirements, a design decision was taken to develop a tool with two layers. General purpose 3D-network visualisation functionality should reside in the base layer, and the MRSA-specific features should reside in the top layer. The former would hopefully constitute a generic visualisation engine. One should be able to develop several applications on a given engine and not just for MRSA. In the case of MRSA, asimplot can also be adapted to any hospital standard regarding the typology of the bacterium.

The application creates and sets properties of all nodes and edges and then passes them on to the engine, which plots the graph. The engine is thus left with the minimal information needed for drawing each node, such as an agent's representation in 3D, and colour information.

After finalising the design, a prototype was developed in seven iterations. Each iteration included a meeting with users, who then proceeded to test the prototype. The tool has two logical parts: a general-purpose network visualisation engine (NVE) and an MRSA-specific application (MRSAApp). The team developed NVE with DirectX-9, using C/C^{++}. The problem domain is completely represented using C^{++} classes, but no effort was made to encapsulate all parts of the tool behind classes; hence, a large part was written in C. The first complete version of asimplot had just over 2,500 lines of code.

Users can move the camera along all three axes, with z-axis movement providing zoom in/out functionality. Alternatively, users can rotate the generated model along X-, Y-, and Z-axes while keeping the camera fixed. Strafing is also available, which essentially means moving the camera perpendicular to its line of vision. The NVE can combine its knowledge about the agents that it is displaying with its ability to strafe in all three axes, to make the centre of the visible agents the new centre of the screen. Colouring the agents allows for the users to discriminate between the agents based on some property chosen at the application level. Edges can have different widths and colours. These two properties can be used to show various information associated with the edges, such as strength, category, etc. In the NVE output, an edge is the line between any two nodes and represents a contact in the MRSAApp.

Six colours were used to discriminate between the edges and they represent the following for any two connected agents N1 and N2:

- Red: If N1 and N2 have a same disease Type 1 and (Case1:) one of them was already tested positive when they entered a ward.
- Yellow: If N1 and N2 have a same disease Type 1 and (Case2:) none of them was tested positive when they entered a ward.
- Blue: If N1 and N2 have the same disease classification Type 2, and Case1.
- Purple: If N1 and N2 have the same disease classification Type 2, and Case2.
- Green: If N1 and N2 have the same disease classification Type 3, and Case1.
- Cyan: If N1 and N2 have the same disease classification Type 3, and Case2.

A disease-type here means an edge-level and is very different from an edge-type which is defined by the two cases, so there are two types of edges. Type-1, Type-2, and Type-3 are category information about the node's MRSA strain. Therefore, an edge formed because of a Type1 strain is a level one edge; in MRSAApp, there are three edge-levels. Another usable visual property of any edge is its width. For each edge, the width is calculated using $1 - (1 - r)^n$, where r is a constant with a value between 0 and 1 and n is an exponent parameter. The result will also be in this range. Value 0 then means the edge will be the thinnest possible one, and 1 the thickest edge. For the MRSAApp, n will be some property of the edge, such as the duration for which the two agents connected by the edge were in contact with each other (e.g., days spent together in the same hospital ward, or number of visits in the same day to the same outpatient clinic).

5.4 Results

In order to describe the features of asimplot, we present in this section some representative output for a real data set. Because 3D-visualisations do not print up well, the graphs presented here are chiefly in 2D, but one 3D-example is included, for the purpose of illustration. For the default graph, the x-axis was mapped onto Day Tested Positive and the y-axis onto First Day In. Many rows had missing values for Day Test Positive, due to missing data. In sharp contact tracing, such data is sought to be completed. Here, it was assumed that these individuals could be tested positive in the future so a high constant value was assigned to their Day Test Positive. Hence, there are many nodes towards the right side of the graph. Similarly, if there is no date for when the first positive culture was taken then the earliest documented date after that date should be used (such as the day the culture result was answered). After 1,000 days, the graph of Fig. 2 results, and after 2,000 days the MRSAApp had calculated and visualised the graph of Fig. 3. There is a significant number of red and yellow links. The graph shows that a large number (about 50) of untested individuals have made links with individuals who have tested positive.

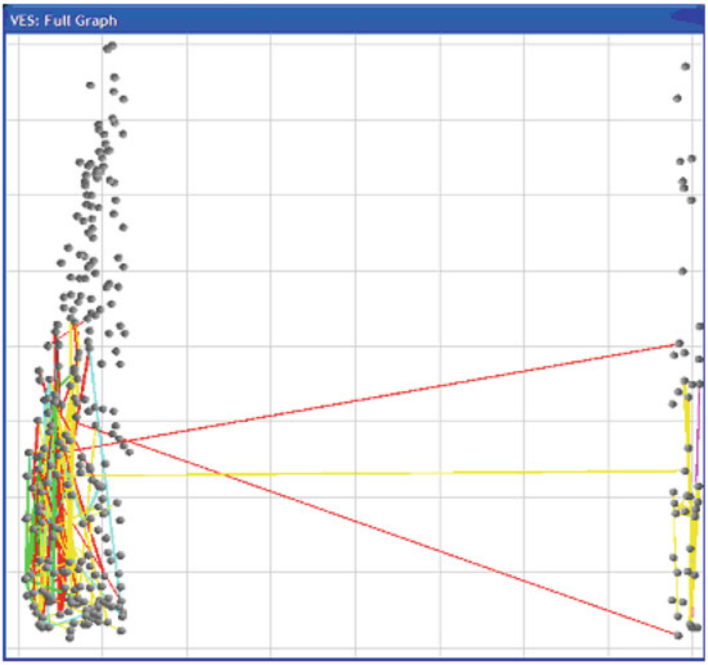


Fig. 2 MRSAApp plot for MRSA Dataset Stockholm 2004, with the x-axis mapped onto Day Test Positive and the y-axis mapped onto First Day In, at t = 1,000 days

The implications of this are two-fold. First, doctors need to test these people for MRSA to make sure that they are not infected, and secondly, these individuals should be classified as high-risk due to their potential acting as bridges.

Observation D: Missing values from hospital logs are often impossible to complete. Instead of *ad hoc* guesses, social collective intelligence can be used to (automatically) complete logs based, e.g., on passive surveillance or voluntary reporting.

Observation E: The power of visualisation is well-known in pedagogy, and properly visualised epidemiological data can be used to communicate with patients and care-givers, for instance when motivating or explaining the reasons for questions about a person's actions or whereabouts. In particular, the epidemiologist can provide the doctor with hands-on prescriptive advice, in this case, e.g., who to test for MRSA.

Figure 4 shows a plot in which the x-axis is mapped to in-degree and y-axis is mapped to out-degree. These two values are useful in contact tracing when it is important to guess who infected whom in a given relation. As there is no accurate way of knowing this, for the MRSAApp, in- and out-degrees are calculated by comparing Day Test Positive of the two nodes in a relation. At the top of Fig. 4, there are nodes representing individuals who have infected at least ten others. We call

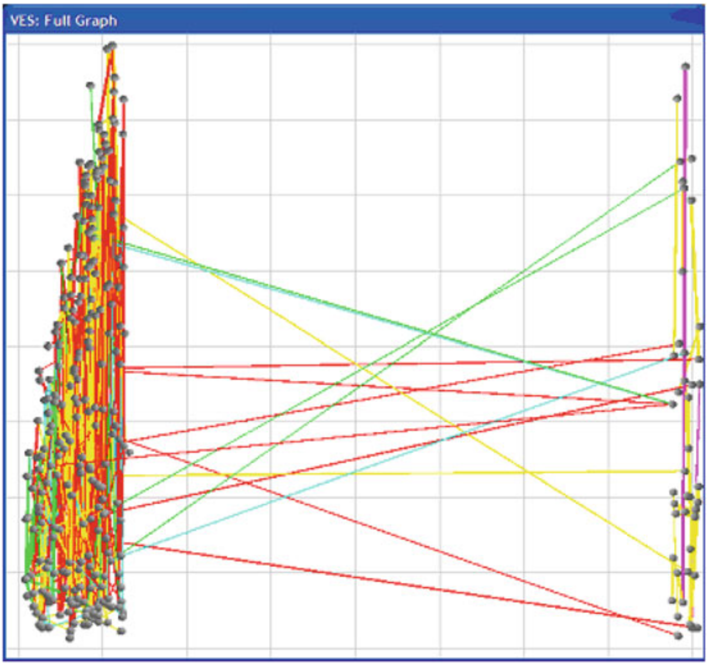


Fig. 3 MRSAApp plot for MRSA Dataset Stockholm 2004, with the x-axis mapped onto Day Test Positive and the y-axis mapped onto First Day In, at t = 2,000 days

such individuals super-spreaders and this figure reveals some of their interesting properties (see [41] for basic network theory terminology). For example, the one with the maximum out-degree has only four red links; this means that the individual was diagnosed after he or she possibly had infected most of his or her infected contacts. One other super-spreader only has red edges, indicating that the individual was already diagnosed when he or she infected others. There is also one super-sink: the node on the extreme right. Moreover, there are several bridge nodes, representing individuals that get infected within the healthcare system, and then infect others there. One of these has been labelled, and that node serves as a sink for a half-dozen nodes, and then as a possible source to another half-dozen.

In Fig. 5, we have remapped the x-axis to Day First In. This graph very clearly shows that all super-spreaders entered the system earlier then most other nodes. There might be more super-spreaders but the MRSAApp can only identify them when new data, representing more recent events, becomes available.

Observation F: When the epidemiologist can see an individual identified as a bridge, or another important vertex in the contact network, the abstract world of graph theory becomes linked to the real world in which this individual resides. It has often been assumed in traditional epidemiology that the classification of individuals into various behaviour- or morbidity-patterns, as in the SIR

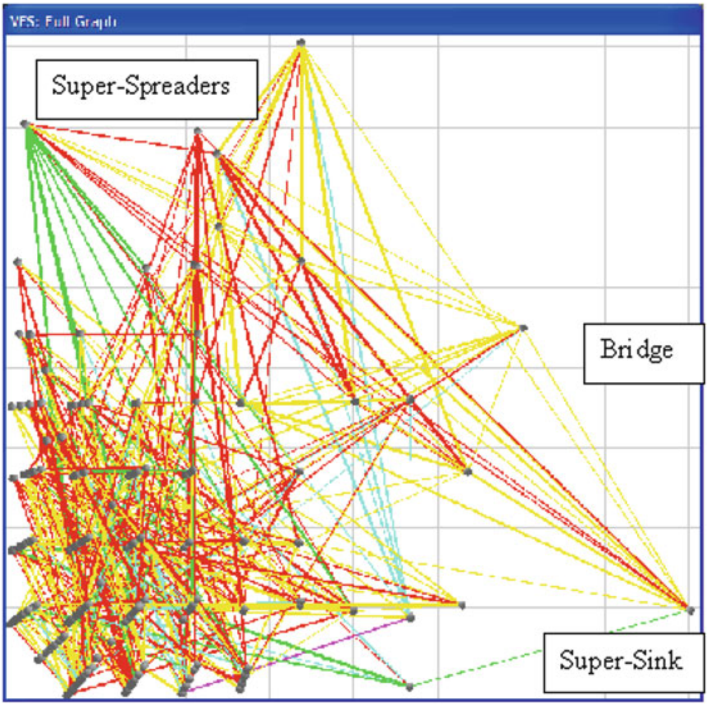


Fig. 4 MRSAApp plot for MRSA Dataset Stockholm 2004, with the x-axis mapped onto in-degree and the y-axis mapped onto out-degree, at t = 2,000 days

family of models, is enough to conduct studies into contact networks. To some extent, this is true, but the advent of new methods and of using social collective intelligence to assist in contact tracing can move analyses away from compartments of individuals, into analyses of the individuals themselves. Since so much health data and non-health data both are linked to the individual, the use of computers for computation and visualisation allow for staying in micro, rather than going through the micro-macro-micro chain: observation (micro) — statistics (macro) — action (back to micro). This is significant since it allows for studying not just “patients” but people, and their entire social life, as necessary and available for study. It also allows for studying co-morbidity: people rarely suffer from one disease or condition at a time, and two conditions suffered simultaneously are rarely independent from each other.

In Fig. 6, we illustrate the 3D-visualisation features of asimplot, although the power of the tool can only be appreciated in full by interacting with it.

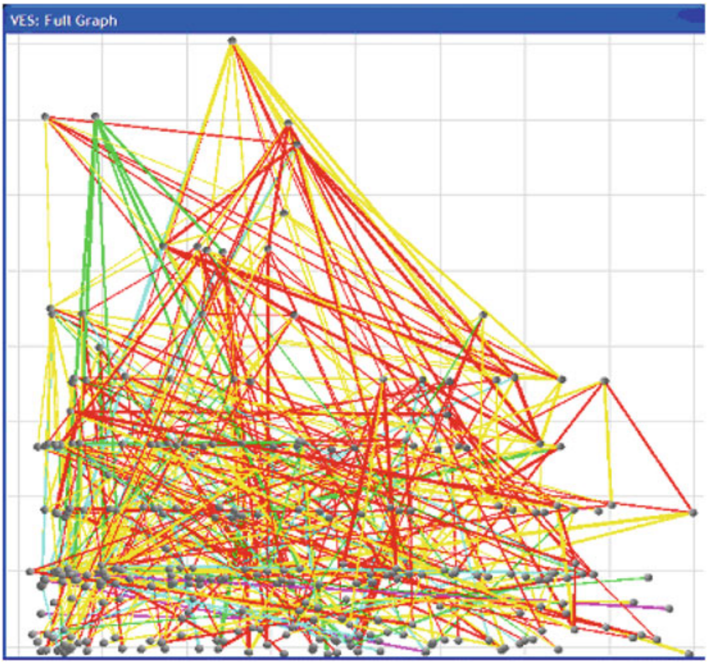


Fig. 5 MRSAApp plot for MRSA Dataset Stockholm 2004, with the x-axis mapped onto First Day In and the y-axis mapped onto out-degree, at t = 2,000 days

6 Conclusion

In the case study, a number of empirical observations were made, each of which provides some support to the hypothesis that *social collective intelligence can be employed for assisting in converting data into useful information through intelligent analyses by deploying new methods from data analytics that render previously unintelligible data intelligible*. In short, these were:

Observation A: Epidemiologists need as much information as they can get on "Who were where when?", and at a fine level of granularity, some of which can be delivered by social collective intelligence.

Observations B and C: Even for a relatively small local outbreak, epidemiologists do not know the bound for their studies (i.e., the universe of discourse of their models), and systems developed for non-health purposes can be employed in order to get better coverage than their regional data sets.

Observation D: Noisy or incomplete hospital logs can benefit from social collective intelligence for their completion, e.g., via passive surveillance or voluntary reporting.

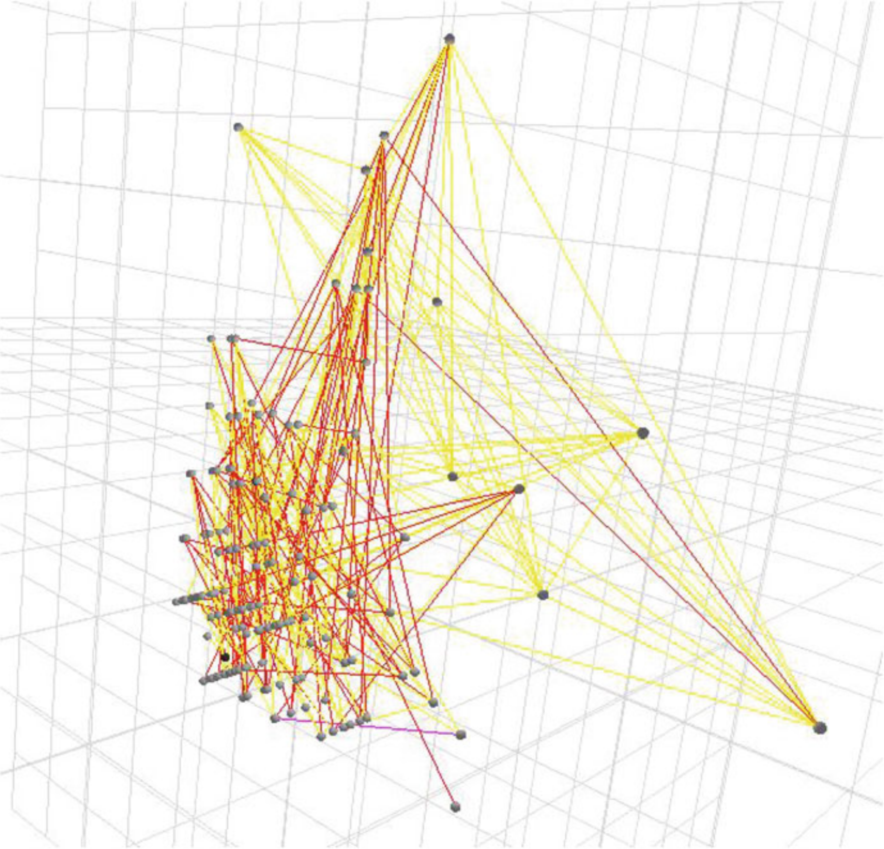


Fig. 6 A sample 3D screen dump from asimplot

Observation E: Properly visualised epidemiological data can be used to communicate with patients and care-givers, and to select who to test for disease, for instance.

Observation F: New methods based on social collective intelligence can move epidemiological analyses away from compartments of individuals, into analyses of the individuals themselves, merging health data with non-health data for personalised care.

The last observation is arguably the heaviest weighing argument for the correctness of the hypothesis. New methods from data analytics allows for massive data analytics to stay in micro (or meso), without resorting to macro analyses of the kind used in traditional epidemiological modelling, population studies, and health economics. The theoretical study also points to advances in complex systems as relevant to new forms of micro-studies of ([55]:241)

> ... systems in which the microscopic properties and processes can be immensely complex and seemingly noisy, yet on larger scales they exhibit certain classes of simple behaviour that seem insensitive to the mechanistic details.

Any individual facing disease is doing so in micro. To be treated in silos; for one condition at a time by one specialist at a time and being prescribed one medicinal treatment at a time, is something of a nightmare to such an individual. The macro-properties (i.e., the statistics) of conditions, specialists, treatments, etc. should not be discarded entirely but they are not necessarily needed for personalised medication, understanding, and advice. Dependencies, such as co-morbidity, are best understood (in the language of complex systems) as a customised parameter space in a bifurcation-type model. This is in keeping with the trend of informing the patient, and practising medicine in a transparent way without compromising privacy or security at any point in the care process.

The last three observations above, and Observation F in particular, assume that the individual assesses enough value to personalised care to take the time to participate (cf. [34]). Moreover, the privacy of the individual might be at risk in sharing such data [23]. Exploitation of non-health data should likewise be privacy-sensitive [45], and surveillance in itself poses many problems, even with the best of intentions [36].

Most of the empirical work was carried out in Sweden, a country with a superb health data reputation and some of the positive results cited herein do not hold for every country. Fighting communicable disease is a global issue, however, and the possibility of an amalgamation between the two areas studied in this chapter provides for serious engineering challenges as well as for considerable health and well-being rewards should they succeed.

Acknowledgements This chapter emerged from experiences gathered over a period of almost 10 years of working as a computational epidemiologist. The author spent time in 2011 and 2012 as a research fellow at the Etisalat BT Innovation Center (EBTIC) in Abu Dhabi, serving the local health authorities (HAAD), and some of the work reported on in this chapter was carried out at EBTIC, where Benjamin Hirsch led the work. Two keynotes at specialist conferences and a short presentation at a Social-IST meeting in 2013 generated a lot of comments and questions, some of which have been included here. The section on MRSA benefited considerably from illustrations and text from the co-authors of an earlier short paper: Asim Ghaffar, Fredrik Liljeros, and Mikael Stenhem. Some of the ideas in this chapter have also been used for research applications over the years, and some of the formulations here were in connection with this improved by SICS colleagues Anders Holst, Björn Bjurling, Markus Bylund, Pedro Sanches, Baki Cakici, and Daniel Gillblad. Baki Cakici provided the author with important comments on earlier sketches of this chapter. Last but not least, the author wishes to express his sincere thanks to Daniele Miorandi for generously sharing his insights on social collective intelligence.

References

1. Anderson, R.M., May, R.M.: Population biology of infectious diseases: Part 1. Nature **280**(5721), 361–367 (1979)
2. Anderson, R.M., May, R.M.: Infectious Diseases of Humans—Dynamics and Control. Oxford Univ Press, Oxford (1991)
3. Arantes, A., Carvalho, E.S., Medeiros, E.A., Farhat, C.K., Mantese, O.C.: Use of statistical process control charts in the epidemiological surveillance of nosocomial infections. Rev. Saúde Pública **37**(6), 768–774 (1993)
4. Aurell, E., Kirkpatrick, S., Koski, T., Skoglund, M., Öktem, O.: KTH-Aalto initiative on big data to small information. ICT platform White Paper (2013). KTH
5. Batagelj, V., Mrvar, A.: Pajek—program for large network analysis. Connections **21**(2), 47–57 (1998)
6. Boman, M., Ghaffar, A., Liljeros, F., Stenhem, M.: Social network visualization as a contact tracing tool. In: Jennings, N.e. (ed.) Proc AAMAS Workshop on Agent Technology for Disaster Management, pp. 131–133. Future University, Hakodate, Japan (2006)
7. Boman, M., Holm, E.: Multi-agent systems, time geography, and microsimulations. In: Olsson, M.O., Sjöstedt, G. (eds.) Systems Approaches and their Application, chap. 4, pp. 95–118. Springer, Netherlands (2004)
8. Bouam, S., Girou, E., Brun-Buisson, C., Lepage, E.: Development of a web-based clinical information system for surveillance of multiresistant organisms and nosocomial infections. In: Proc AMIA Symp, pp. 696–700 (1999)
9. Bowles, S., Gintis, H.: The inheritance of inequality. J. Econ. Perspect. **16**(3), 3–30 (2002)
10. Brandes, U., Kenis, P., Raab, J., Schneider, V., Wagner, D.: Explorations into the visualization of policy networks. Theor. Polit. **11**, 75–106 (1999)
11. Brockmann, D., Hufnagel, L., Geisel, T.: The scaling laws of human travel. Nature **439**(7075), 462–465 (2006)
12. Brouwers, L., Boman, M., Camitz, M., Mäkilä, K., Tegnell, A.: Micro-simulation of a smallpox outbreak using official register data. Eurosurveillance **15**(35) (2010)
13. Brouwers, L., Cakici, B., Camitz, M., Tegnell, A., Boman, M.: Economic consequences to society of pandemic H1N1 influenza 2009: Preliminary results for Sweden. Eurosurveillance **14**(37) (2009)
14. Cakici, B., Boman, M.: A workflow for software development within computational epidemiology. J. Comput. Sci. **2**(3), 216–222 (2011)
15. Cakici, B., Hebing, K., Grünewald, M., Saretok, P., Hulth, A.: CASE: a framework for computer supported outbreak detection. BMC Med. Inform. Decis. Making **10**(14) (2010)
16. Chen, H., Zeng, D., Yan, P.: Infectious Disease Informatics: Syndromic Surveillance for Public Health and Bio-Defense, 1 edn. Springer, New York (2009)
17. Corley, C.D., Cook, D.J., Mikler, A.R., Singh, K.P.: Text and structural data mining of influenza mentions in web and social media. Environ. Res. Publ. Health **7**(2), 596–615 (2010)
18. Culotta, A.: Detecting influenza outbreaks by analyzing Twitter messages. arXiv:1007.4748v1 [cs.IR] (2010)
19. Eagle, N., Pentland, A.: Eigenbehaviors: identifying structure in routine. Behav. Ecol. Sociobiol. **63**(7), 1057–1066 (2009)
20. Espino, J.U., et al.: Removing a barrier to computer-based outbreak and disease surveillance–The RODS Open Source Project. MMWR Morb. Mortal Wkly. Rep. **53**(Supplement), 32–39 (2004)
21. Eubank, A., et al.: Modelling disease outbreaks in realistic urban social networks. Nature **429**, 180–184 (2004)
22. Ferguson, N.M., et al.: Strategies for mitigating an influenza pandemic. Nature **442**, 448–452 (2006)

23. French, M.A.: Picturing public health surveillance: Tracing the material dimensions of information in ontario's public health system. Ph.D. thesis, Queen's University, Kingston, Ontario, Canada (2009). Dept of Sociology
24. Genesereth, M.R., Ketchpel, S.: Software agents. Comm. ACM **37**(7), 48–ff. (1994). DOI 10.1145/176789.176794. URL http://doi.acm.org/10.1145/176789.176794
25. González, M.C., Hidalgo, C.A., Barabási, A.L.: Understanding individual human mobility patterns. Nature **453**(7196), 779–782 (2008)
26. Hall, M., Gani, R., Hughes, H.E., Leach, S.: Real-time epidemic forecasting for pandemic influenza. Epid Inf. **135**(3), 372–385 (2007)
27. Halloran, M.E., et al.: Modeling targeted layered containment of an influenza pandemic in the united states. PNAS **105**(12), 4639–4644 (2008)
28. Hedström, P., Swedberg, R. (eds.): Social Mechanisms: An Analytical Approach to Social Theory. Cambridge University Press, Cambridge (1998)
29. Hewitt, C.: Offices are open systems. ACM Trans. Inf. Syst. **4**(3), 271–287 (1986). DOI 10.1145/214427.214432. URL http://doi.acm.org/10.1145/214427.214432
30. Holling, C.S.: Resilience and stability of ecological systems. Ann. Rev. Ecol. Stat. **4**, 1–23 (1973)
31. Hulth, A., Rydevik, G., Linde, A.: Web queries as a source for syndromic surveillance. PLoS ONE **4**(2), e4378 (2009)
32. Kirkpatrick, M.: Meet the firehose seven thousand times bigger than Twitter's. ReadWriteWeb (2010)
33. Liljeros, F., Giesecke, J., Holme, P.: The contact network of inpatients in a regional healthcare system. a longitudinal case study. Math. Popul. Stud. **14**(4), 269–284 (2007). DOI 10.1080/08898480701612899
34. Lipsitch, M., et al.: Managing and reducing uncertainty in an emerging influenza pandemic. NEJM **361**(2), 112–115 (2009)
35. Longini, I.M., et al.: Containing pandemic influenza at the source. Science **309**(5737), 1083–1087 (2005). DOI 10.1126/science.1115717. URL http://www.ncbi.nlm.nih.gov/pubmed/16079251
36. Lyon, D.: Surveillance Studies: An Overview. Polity Press, Cambridge (2007)
37. Marathe, M.V., Vullikanti, A.K.S.: Computational epidemiology. Comm. ACM **56**(7), 88–96 (2013)
38. Mulligan, M.E., et al.: Methicillin-resistant staphylococcus aureus: A consensus review of the microbiology, pathogenesis, and epidemiology with implications for prevention and management. Am. J. Med. **94**(3), 313–328 (1993)
39. Naaman, M.: Social multimedia: highlighting opportunities for search and mining of multimedia data in social media applications. Multimed. Tools Appl., 1–26 (2010)
40. Nagel, K., Beckman, R.J., Barrett, C.L.: TRANSIMS for regional planning. Int. J. Complex Syst. (1998). Manuscript 244
41. Newman, M.E.J.: The structure and function of complex networks. SIAM Rev. **45**, 167–256 (2003)
42. Ottino, J.M.: Engineering complex systems. Nature **427**(6973), 399 (2004)
43. Personalised medicine. European Commission, Futurium, Digital Agenda for Europe (2013). Http://ec.europa.eu/digital-agenda/futurium/en/content/personalised-medicine
44. Rosenzweig, M.L.: Paradox of enrichment: Destabilization of exploitation ecosystems in ecological time. Science **171**(3969), 385–387 (1971)
45. Sanches, P., Svee, E., Bylund, M., Hirsch, B., Boman, M.: Knowing your population: Privacy-sensitive mining of massive data. Netw. Comm. Tech. **2**(1), 34–51 (2013)
46. Scheffer, M.: Critical Transitions in Nature and Society. Princeton University Press, Princeton (2009)
47. Scheffer, M. et al.: Early-warning signals for critical transitions. Nature **461**, 53–58 (2009)
48. Shiller, R.J.: From efficient markets theory to behavioral finance. J. Econ. Perspect. **17**(1), 83–104 (2003)

49. Smith, R.G., Mitchell, T.M., Chestek, R.A., Buchanan, B.G.: A model for learning systems. In: Proc IJCAI, pp. 338–343. Cambridge, MA (1977)
50. Song, C., Koren, T., Wang, P., Barabási, A.L.: Modelling the scaling properties of human mobility. Nat. Phys. **6**(10), 818–823 (2010)
51. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of predictability in human mobility. Science **327**(5968), 1018–1021 (2010)
52. Steels, L.: Cooperation between distributed agents through self-organisation. In: Decentralized A.I: Proc Modelling Autonomous Agents in a Multi-Agent World (MAAMAW), pp. 175–196. North-Holland (1990)
53. Upbin, B.: IBM's Watson gets its first piece of business in healthcare. Forbes (2013). TECH 2/08/13
54. Vespignani, A.: Predicting the behavior of Techno-Social systems. Science **325**(5939), 425–428 (2009)
55. Zlemells, K.: Complex systems. Nature **410**(6825), 241 (2001)

Social Collective Awareness in Socio-Technical Urban Superorganisms

Nicola Bicocchi, Alket Cecaj, Damiano Fontana, Marco Mamei, Andrea Sassi, and Franco Zambonelli

1 Introduction

The widespread adoption of sensor networks, actuators and computational resources capable of interacting with people is transforming urban environments [6]. Citizens will have the possibility of being continuously connected in a situation-aware and socially-aware way, both with each other and with the entities around, e.g., typically via some situation-aware social networking infrastructure [28]. This will eventually contribute to define a dense ecosystem whose individual components will enable collaboration between ICT devices and humans, towards the realization of advanced urban services. Such services can contribute to the smart city vision along several dimensions from intelligent transportation systems, to environmental sustainability and participatory governance [15, 24]. Even more, it has been envisioned how they could radically transform urban environments into socio-technical urban superorganisms [12, 28].

Future pervasive urban services, rather than being limited to sensing what happens in the city (as most current approaches do) will leverage on the complementary sensing, computing, and actuating capabilities of humans and ICT devices.

The urban system *as a whole* will be able to: (a) combine a wide range of information sources (e.g., environmental data from sensor networks, mobility data and social network posts) and sense the current state of the city. (b) perform

N. Bicocchi (✉)
Dipartimento di Ingegneria dell'Informazione, Universitá di Modena e Reggio Emilia,
Corso Canal Grande, 64, Modena, Italy
e-mail: nicola.bicocchi@unimore.it

A. Cecaj • D. Fontana • M. Mamei • A. Sassi • F. Zambonelli
Dipartimento di Scienze e Metodi dell'Ingegneria, Universitá di Modena e Reggio Emilia,
Corso Canal Grande, 64, Modena, Italy

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_11

advanced reasoning on the data to extract patterns and routines taking place. (c) engage in large-scale coordinated tasks to achieve specific goals (e.g., optimize traffic flow in the city, make it more environmentally sustainable, etc.).

That is, the overall urban environment will act as a single superorganism made up of individual organisms (humans and ICT devices) and capable of directing its behaviour towards the achievement of specific urban-level goals.

Engineering collaborative and coordinated services capable of harnessing human and ICT capabilities at large scales challenges current engineering practices and middleware architectures. In this context, the contribution of this work is twofold:

- It sketches the key concepts of urban superorganisms and of the collective awareness mechanisms at the basis.
- It outlines the key research challenges to be faced to realize such kind of systems
- It proposes a self-aware middleware architecture conceived around the goal of tackling the identified challenges.

The paper is organized according to the above points. Section 2 presents our vision on urban superorganism and on the associated collective awareness. Section 3 discusses key research challenges. Section 4 proposes a general-purpose middleware architecture addressing some of those challenges. Section 5 provides a technical overview of the same architecture presenting also some initial experiments we conducted to analyze the feasibility of the architecture. Finally, Sect. 6 provides some concluding remarks.

2 The Urban Superorganism

In future ubiquitous computing scenarios, the very large number of inter-connected entities that can be found in urban environments, whether humans or ICT devices, can potentially be exploited to create what has been defined as a superorganism [12]. In particular, closing the sensing, computing, and actuating capabilities in a loop (see Fig. 1), and making such activities collaborative ones, it is possible to realize coherent collective behaviours, as it is observed in many natural situations, e.g., in ant colonies [4].

A single ant has very limited, local sensing and actuating capabilities, and little or no cognitive abilities. Yet, ants can indirectly coordinate their movements and activities, via spreading and sensing of pheromones in the environment, so as to exhibit, as a colony, very powerful capabilities of sensing (finding food in the environment), computing (finding the shortest path from food back to nest), and action (carrying large amounts of food in the nest). These capabilities make the whole colony seemingly intelligent and certainly adaptive in its foraging activities.

Research in self-adaptive and self-organizing systems have focused on defining a catalogue of bio-inspired mechanisms, with the intent to overcome the limit of ad-hoc implementations that prevent their systemic reuse [7]. Thus, the basic idea is that of providing the bio-inspired self-organizing pattern modules with a

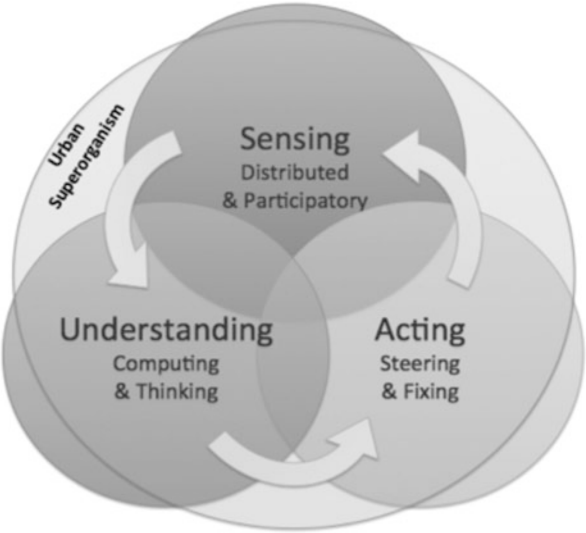
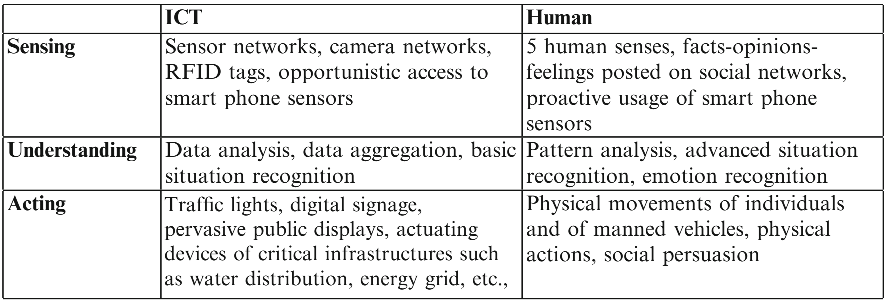


Fig. 1 Collaborative sensing, awareness and action among humans and ICT systems can be put at work in urban superorganisms in the form of a closed feedback loop

set of reusable patterns that could be used to ease engineering of artificial and collective behaviors of urban superorganisms. It is probable that the complementary capabilities of humans and ICT devices and the ability to coordinate and organize them could overcome current approaches and promote collective awareness and complex behaviors.

More in detail, Fig. 1 illustrates the feedback cycle creating collective awareness in the super organism. Advanced finalized and coordinated activities are the result of: *Sensing* activities in which users supported by ICT devices and services get information about the current state of the environment (e.g., people location data). *Understanding* activities in which advanced forms of context information are derived from the sensed data (e.g., citizens mobility patterns are identified from the collected location data). *Acting*: goal-directed coordinated tasks supported by the extracted information (e.g., traffic management on the basis of the identified mobility patterns, car sharing on the basis of people mobility routines, etc.). The results of the activities being performed are then sensed again closing the feedback cycle.

Previous works in opportunistic and participatory sensing have tried to involve users by making use of their devices as sensors [14, 21]. On the opposite side, other works try to detect events or situations by observing users activities on online social networks [23]. However, these works lack a general and unified vision and do not completely tackle the complexity of the global scenario, i.e. they do not explore all the possible convergence of humans and ICT devices. Moreover, they do not



	ICT	Human
Sensing	Sensor networks, camera networks, RFID tags, opportunistic access to smart phone sensors	5 human senses, facts-opinions-feelings posted on social networks, proactive usage of smart phone sensors
Understanding	Data analysis, data aggregation, basic situation recognition	Pattern analysis, advanced situation recognition, emotion recognition
Acting	Traffic lights, digital signage, pervasive public displays, actuating devices of critical infrastructures such as water distribution, energy grid, etc.,	Physical movements of individuals and of manned vehicles, physical actions, social persuasion

Fig. 2 The table summarises sensing, computing, and actuating capabilities of both humans and ICT devices that could mutually interact within an urban superorganism

fully make use of the very large number of inter-connected individuals and their complementary capabilities to realize a collective awareness.

Recent works are addressing the "smart city" scenario [5]. Current approaches to smart cities (e.g., as those that are carried on by IBM [15]) are mostly related to the "sensing and understanding" facets of the urban superorganism scenario (cfr. Fig. 1), i.e., collecting (typically in a centralized way) data about various aspects of a city life and get a meaning out of it, for the sake of driving decision makers in planning future urban infrastructures. The acting aspect, i.e., the possibility of dynamically involving citizens and ICT actuators to dynamically influence the city dynamics is mostly disregarded. Only a few studies in this direction have been performed, and mostly oriented to steer crowd via mobile phones (e.g., the Tag my Lagoon Project in Venice[1]), or at directing traffic towards zones with available parking space (e.g., the Santander Smart City Project[2]).

2.1 The Role of Humans and ICT Systems

In this section we provide more details on the sensing, computing and actuating activities to be integrated for the sake of enabling the above described collective awareness feedback cycle. People are increasingly equipped with smart phones with powerful capacities in terms of battery life, sensing, computational power and connectivity. At the same time, autonomous ICT infrastructures (sensor networks, security cameras, robots, etc.) are likely to pervade cities in the near future. Accordingly, the future urban environment is becoming a sort of very dense digital ecosystem (Fig. 2).

[1] www.tagmylagoon.com.

[2] www.smartsantander.eu.

The components that are going to pervade urban environments are characterized by heterogeneous and complementary sensing, computing, and actuating capabilities, that can cooperate in a goal-directed way. In particular:

Sensing

- ICT Side. The capabilities in sensing from the ICT side are provided by (a) mobile phones equipped with GPS, accelerometers and cameras; (b) sensors networks and smart objects that follow the Internet of Things paradigm [1]; (c) tags that exploit the wireless short-range communication technologies (NFC, RFID and Bluetooth).
- Human Side. From the human side, the five senses of humans, which in many situations can supply and be more accurate than ICT sensors, can be put at work for the community, due to the possibility of continuous access to social networks. In addition, users can make available via social networks any other information, thus acting as sorts of social sensors [22].

Understanding

- ICT Side. The capabilities in computing from the ICT side makes it possible to collect and digest very large amounts of urban data in a short time, and to perform some limited pattern analysis on such data.
- Human Side. From the human side, on the other hand, one can exploit the capability of recognizing complex situations and patterns (so called human computation [27]), which machines can hardly tackle.

Acting

- ICT Side. The capabilities in actuating from ICT side can be provided by (a) traffic controllers supporting pervasive solutions in the mobility dimension; (b) public displays that will be exploited to promote adaptable citizens behavioural steering; (c) all kinds of actuators related to critical infrastructures (e.g., energy grid).
- Human Side. From the human side the key actuating element involved is the user himself, which can perform a variety of actions related to moving or moving items around or changing the properties of some physical entities. In other words, citizens could accomplish actions, by realizing an impact on the environment.

The goal-directed integration of the above capabilities and activities, will allow to close the collective-awareness feedback loop enabling large-scale coordinated behavior among humans and ICT devices and services.

2.2 Application Scenarios

In this section we present some exemplary application scenarios that could be enabled by the vision of the super organism and by the above defined collective awareness cycle.

Mobility. Among many capabilities that future urban superorganisms will exhibit, the first that we expect to be in place, and for which we already observe embryonic examples around, will relate to urban mobility [10, 13, 25]. Specifically, it will relate to the capability of affecting (i.e., steering) the movement of vehicles or pedestrians, and thus improving the overall efficiency of urban mobility while reducing the stress of users, due the improvement of traffic flows and the avoidance of traffic congestion. A variety of sensors already exist to detect the conditions of traffic or crowd in urban environments. In addition, users are increasingly given the possibility to contribute to such sensing activities by posting information on social networks or by opening access to their navigators and smart phone sensors. All this information can be used to understand how to improve traffic flows or how to avoid congestion. To this end: actuators such as traffic lights and digital traffic signs can be put at work for vehicles; public (wall mounted) and private (smart phone) displays can be exploited to suggest directions to pedestrian. However, one could push the capabilities of superorganisms much further. For instance, one can think of dynamically matching the similarity of planned vehicle routes and of merchandise to be delivered to dynamically self-organize a very flexible ride sharing and shipment services. In general, urban superorganisms induce a change in the dominant paradigm for the provisioning of mobility services: from sensing mobility patterns and adapt existing services to them, to dynamically collect mobility needs and self-organize the role and mobility patterns of vehicles accordingly.

Sustainability. As an additional example of how the capabilities of future urban superorganisms can impact urban life, just imagine sensing in real-time information related to energy consumption, to compute sorts of instantaneous urban carbon footprints for specific areas of the city or for specific groups of citizens, other than for the city as a whole. Public displays can then be exploited to share this information and possibly some analysis of the factors contributing to it, and personal displays can be possible exploited to let individuals and groups to become aware of their own contributions to the urban carbon footprint. On these bases, one could think of steering the behaviour of individual citizens towards more energy efficient behaviours, or at engaging groups of citizens in self-organized collaborative actions aimed at solving/improving specific energy problems in specific urban areas to supply the lack of actuators suitable to the purpose (e.g., detecting open windows and closing them).

Taking Care. Via similar means, it could be possible to dynamically involve citizens in proactively helping to taking care of the city, e.g., to help keeping it city cleaner or making it a safer place for everyone. For instance, one can think of dynamically engaging people to temporarily take care of (or simply take a look at for some minutes) children on their way to school, whenever the current activity and known habits of some persons suggests. Ideally, in the presence of enough matching persons willing to be involved, and possibly complementing sensors (e.g., cameras) and actuators (e.g., robots) already in place for that purpose, one can make sure that the whole path from home to school of every children in a city is properly covered and taken care of.

Feeling Part of It. Beside thinking at measurably useful objectives and services for which urban superorganisms can be put at work, their advantages could also be in the (not easily measurable) way by which they will improve our experience of living in urban environments. In particular, acting and moving around in a city by being given feedbacks on the effect of our own existence in it, and by making possible to observe ourselves in relation with our environment and with the other citizens, can make most of our everyday actions inherently more pleasant and rewarding, and can promote a renewed and stronger sense of citizenship.

In addition to these exemplary applications, we think that in the next future innovative collaborative and collective behaviours, expressing various forms of urban awareness and intelligence, will take place. These will dramatically change the way we move, live, and work, in our urban environments.

3 Challenges for Superorganism Architectures

The vision of the urban superorganism presents a number of challenges that can be hardly dealt with by present middleware architectures. In this section we present a number of such challenges, while in the next one, we present a middleware proposal addressing them.

3.1 Heterogeneity and Interoperability

The software architecture has to provide an abstraction layer on top of different individuals, both humans and ICT devices, by adapting their heterogeneous and complementary sensing, actuating and computing capabilities [9, 19]. The complementary use of sensing and understanding capabilities of humans and ICT devices has to lead the coordinated learning process towards reconstructing a collective awareness of the state of the city. A common example that fits this situation is expressed by the capturing of pictures of a traffic jam both from users cameras and Closed Circuit TeleVision (CCTV) cameras, to make sure that all the images are properly tagged with information automatically generated by devices (e.g., amount of vehicles involved) and further enriched by humans (e.g., the reason they are stuck in traffic), due to their higher classification capabilities. The challenge is to realize an abstraction layer able to continuously observe the superorganism status and plan strategies to reach specific goals by making use of heterogeneous individuals with evolving sensing, actuating and computing capabilities.

3.2 Dynamic Re-configurability

As emerged by the heterogeneity challenge, the software architecture should show a certain degree of flexibility. Thus, the ability to dynamically execute heterogeneous code with heterogeneous sensing, actuating and computing capabilities is a key challenge to be tackled. For instance, in the case of an accident report executed by a citizen, the architecture has to support a service that requires the user interaction and thus has a user interface; while, in the case of the same task executed by a remote camera, the latter has to support services exploiting the sensing, actuating and computing capabilities of devices. Furthermore, dynamic service composition and re-configuration is needed. The design of an architecture dealing with these challenges will make applications interoperable and able to self-reconfigure and self-optimize depending on the execution context.

3.3 Interconnection

As emerged by the case study, individuals have to be connected and able to exchange messages for both (a) supporting collective behaviors (e.g. the coordination effort required to steer individuals to take photos of a road intersection), and (b) gathering individual awareness to infer complex situations that involve the superorganism rather than specific individuals. It is worth noticing how this requirement calls for innovative data fusion techniques. In fact, at both the individual and superorganism levels multiple data streams of information sources have to be processed and put together to build up a coherent picture of operating conditions.

3.4 Behavioural Steering

The effectiveness of the infrastructure supporting the urban superorganism is determined by the amount of people actually involved in the collective sensing, computing and actuating phases. So, from a social perspective, it depends mainly on how deep individuals are steered by collaborative behaviors. Academic literature present several studies on modeling and evaluating behavioral changes driven by ICT devices [16, 17, 29]. As precisely described by Klein et al. [16], behavioral changes can be supported by taking a closer look at underlying determinants of behavior change, focusing on how users can be persuaded to establish a desired behavior. This practice results in the identification of behavioral patterns composed by a sequence of behavior determinants, which are solicited through proper HCI techniques (e.g., provide the user with automated reminders, valuable suggestions, and tailored feedback on her activity), and then evaluated with a related computational model based on theoretical frameworks of behavior change, to classify the degree of individuals motivation, awareness, and commitment to adopt new desired habits. However, a satisfying evaluation of behavioral change models effectiveness needs to be protracted for years, in order to distinguish short-lived changes from permanent ones.

3.5 Dynamic Selection

One should evaluate which individuals are more suitable to be involved by taking in account different constraints. For instance, in the case study, many CCTV cameras and humans could be available at the same time to capture traffic information to feed any Intelligent Transportation System. Strategies taking into account different constraints (e.g. geographical areas, individuals status and sensing, actuating and computing capabilities) are needed to pick up the most affine individuals for the desired behavior. These strategies could be based both on explicit interactions (e.g., sending a message to an individual) or implicit interactions (e.g., sub-sampling the whole population of individuals satisfying specific constraints).

3.6 Context Awareness

As technologies evolve, new types of sensors become available. Chemical, electric, optical, proximity and position sensors can provide data about environment, weather, presence or movements of and between different entities part of the city life. Furthermore humans can also act as a type of social sensor through social networks or their mobile phone signals [9, 19]. These data sources, made of humans and ICT devices, will produce, continuous streams that will generate a very big data set, specially, if we consider the temporal dimension of data. From a computation concern, the issue is about finding suitable pattern analysis algorithms to extract high-level knowledge from sensed data [2]. As the number of available data sources and algorithms to process them is constantly increasing, the perception is that the algorithms to extract relevant information from data are already there. The important challenge is to find ways of combining them together so that results coming from one data can validate and further describe results from other data [3, 11, 18].

3.7 Mixing Bottom-Up and Top-Down Design

Designing with a top-down approach means that all the requirements of a software architecture have to be taken into account first; systems engineered in this way have a predictable and measurable behavior but are not capable of coping with dynamic execution-context; while systems designed with a bottom-up approach are more robust and suitable for a pervasive environment but predicting their behavior and controlling them by design is not an easy task. In the design of architectures for urban superorganisms, both of the two approaches are needed and finding and tuning the optimal trade-off between them is a key challenge to be tackled [6, 26].

4 Architecture

As a first step to create the basis for the collective awareness schema and to address some of the challenges presented in the previous section, we designed and developed a middleware architecture supporting collective sensing, understanding and actuating capabilities. Such an architecture constitutes a general-purpose awareness framework that could be used as a starting point for many superorganism services.

4.1 Conceptual Viewpoint

The architecture (see Fig. 3-right) is structured around four layers, namely *sensor*, *classifier*, *awareness* and *actuator* layer. Each layer can host multiple modules connected to each other via application-definable topologies. The data flow from

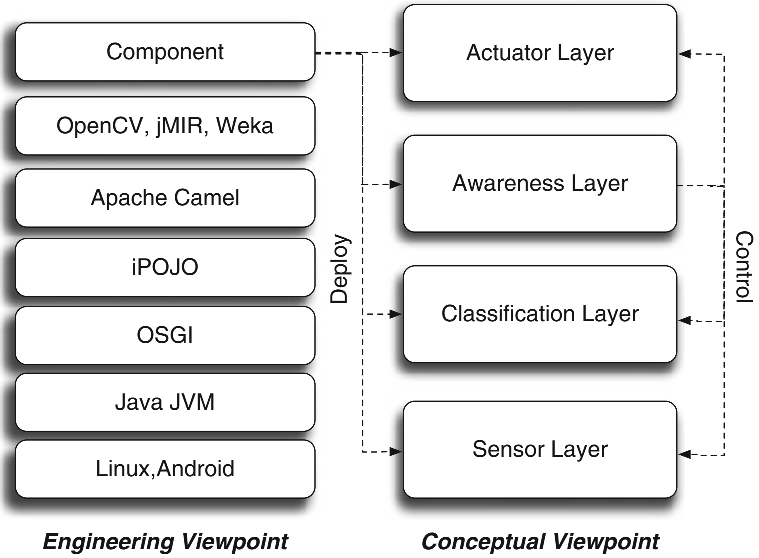


Fig. 3 The architecture seen from both conceptual and engineering viewpoints. From the conceptual viewpoint, it is structured around three layers, namely *sensor*, *classifier* and *awareness* layer. From an engineering viewpoint, the architecture is implemented on the top of industrial-level Java technologies. Each module is actually an OSGi component enriched with iPOJO and Apache Camel functionalities (*right*). The skeleton of a simple component is also reported (*left*)

sensors (i.e., both hardware and software) trough the whole architecture by means of in-memory queues enabling modules decoupling and many-to-many asynchronous communications. Each layer can host multiple modules (i.e, sensors, classifiers, awareness modules, actuators and queues).

The *sensor layer* hosts modules that are in charge of retrieving raw data from physical sensors and of preprocessing them. An example could be a module acquiring images from a camera and cropping and resizing them. Other examples could be modules acquiring facts from social networks, such as Twitter, Facebook or Foursquare. At the time of writing, we have already implemented modules for reading data from Android devices. In general, this layer addresses the heterogeneity and interoperability challenge (Sect. 3.1) in that it provides a uniform access to the various sensors and devices.

The *classification layer* hosts modules that consume data coming from the sensor layer and classify (i.e., generate semantically richer information) them. An example could be a module able to classify the activity performed by a user by processing accelerometer data. At the time of writing, we have implemented modules for classifying user activity, location, speed, vehicle used on the basis on common smartphone sensors. It is worth noting that our goal is to build a general-purpose awareness framework that could be used as common basis for both research and application development, not to solve every possible classification problem. Specific applications will need their own modules to be developed. This layer tackles most of the context awareness (Sect. 3.6) challenges in that it provides a wide range of context-modules to be integrated to fulfil the application needs.

The *awareness layer* hosts modules consuming labels produced in the classification layer and feeding external applications with situational information. These modules might have different goals depending on the application. However, they could be divided into two main classes. The former comprises modules delegated to sensor fusion processes. These modules receive labels, eventually conflicting, coming from multiple classification modules and apply algorithms to achieve higher semantic levels. The latter, instead, is related with the capability of the framework of monitoring and controlling itself. In a sense, the awareness layer could be the key for building an awareness module that is aware of itself. For example, it would be possible to integrate within this level modules observing the internal status of the framework and activating different classifiers and sensors depending on the operating conditions. This capability could be used to achieve both improved classification accuracies and reduced power consumption levels by continuously selecting the most suitable classifiers and sensors. The strategies used to select sensors and classifiers might change depending on the application. This layer tackles both the context awareness (Sect. 3.6), and dynamic selection (Sect. 3.5) challenges in that it allows to fuse information together and to choose the most valuable information providers.

The *actuator layer* hosts modules to enact specific activities and to steer the behavior of users toward specific goals. On the one hand, the actuator layers contains modules to drive the actuators of physical sensors (e.g., to control smart household devices). On the other hand, it contains visualization and user-interaction modules to ask/steer the users to specific actions. At the time of writing, we are implementing modules for user interaction for Android devices. In the long term, the idea is to incorporate in this layer incentivization modules [20] and mechanisms taken from persuasion theory [8]. The use of these latter techniques addresses the behavioral steering (Sect. 3.4) challenges.

4.2 Engineering Viewpoint

From an engineering viewpoint, the architecture is implemented on top of industrial-level Java technologies (see Fig. 3-left). Each module is actually an OSGi component. Because of this, modules (i.e, components) within this architecture can be plugged, removed and reconfigured at runtime. These capabilities, related with the adoption of a Service Oriented Component approach are crucial to address the dynamic reconfigurability (Sect. 3.2) and interconnection (Sect. 3.3) challenges.

On top of OSGi, we have an iPOJO layer. iPOJO is a container-based framework handling the lifecycle of *Plain Old Java Objects (POJOs)* and supporting management facilities like dynamic dependency handling, component reconfiguration, component factory, and introspection. Moreover, the iPOJO container is easily extensible and allows pluggable handlers, typically for the management of non-functional aspects.

On top of the iPOJO framework we build the support for the staged and layered architecture by making use of Apache Camel. This framework provides components with the capability of asynchronously processing data streams and communicate through in-memory queues. These queues allow modules belonging to different layers to continuously communicate each other with minimum hardware requirements. Considering that pattern classification and analysis has a central role in situation awareness, we wrapped well-know data manipulation libraries within the framework. For instance, experiments presented in this paper made use of Weka and jMIR.

Overall, the proposed architecture allows developers to select the required modules, define the topology of data flows and specify their reconfiguration strategies. The middleware takes care of connecting all the modules and to reconfigure them whenever needed. Such high-level support will allow developers to focus on software engineering-level tasks such as those emphasized in Sect. 3.7 about mixing top-down and bottom-up approaches.

5 Experimental Results

To quantitatively assess and validate the feasibility, in terms of performance, of exploiting our awareness module in a superorganism architecture, we have deployed and tested it on servers with Core Duo 2 CPUs operating at 2.2GHz and 5GB

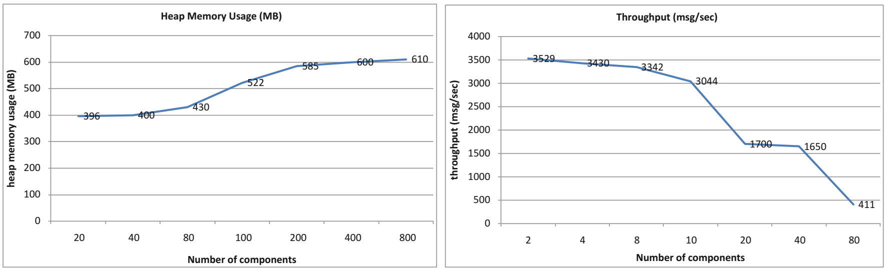


Fig. 4 We have evaluated memory consumption and throughput of the awareness module by dynamically increasing the number of parallel components in the classifier layer that process a data stream without further computations

of RAM. In the tests we used MacOSX 10.6 with JVM v1.6. The framework is running on Apache Karaf 2.3.0 OSGi container. To validate our implementation, we have stressed the awareness module by realizing a sensor that produces a high traffic load of 10,000 messages (with small payload). As experimental evaluation, we have estimated the performance by dynamically increasing the number of parallel components in the classifier layer that process this data stream without further computations. By dynamically increasing the number of components, we increase the number of messages that have to be processed in the entire module, because the classifiers subscribe the same data stream. The first metric used to evaluate the performance is the memory usage. The corresponding experimental results are reported in Fig. 4. In particular, the figure show how the heap memory usage increases linearly with the number of messages processed.

The second metric used to evaluate the performance of the awareness module is the throughput, in terms of average number of messages per second processed by components in the classifier layer. The corresponding experimental results are reported in Fig. 4. In particular, the graph shows how the average throughput of the system decreases linearly with the increase of parallel components that consume the data stream. However, in the worst case represented by eighty parallel components in execution in the classifier layer, the average throughput remain acceptable, by considering that all the components are able to process 411 messages per second. Considering that this test has been run a single machine, we believe that this would be enough to handle the large majority of circumstances.

These results demonstrate how (a) the bottleneck in the awareness module is the CPU that limits the scalability in terms of number of messages processed per second; (b) the overhead introduced by a self-reconfigurable and highly adaptable awareness module is negligible, because the performance decreases linearly with an increasing load and number of components.

6 Conclusion and Future Works

As we have discussed in this paper, we believe that it will be possible to exploit socio-technical superorganisms to deliver complex collective urban-level services. In our opinion, innovative collaborative collective behaviours expressing various forms of urban awareness and intelligence will take place, and dramatically change the way we move, live, and work, in our urban environments. However, to reach this goal, many research challenges need to be addressed, and suitable middleware infrastructures have to be developed. At the time of writing, we are in the process of completing a first prototype implementation of the proposed architecture. In addition, our future work includes testing the infrastructure in controlled (campus-level) environment and, later on, start experiencing it in real-world situations with simple urban awareness services.

Acknowledgements Work partially supported by the ASCENS project (EU FP7-FET, Contract No. 257414) and by the SPINNER project MUCCA.

References

1. Atzori, L., Iera, A., Morabito, G.: The internet of things: A survey. Comput. Network **54**(15), 2787–2805 (2010). DOI 10.1016/j.comnet.2010.05.010. URL http://dx.doi.org/10.1016/j.comnet.2010.05.010
2. Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., Riboni, D.: A survey of context modelling and reasoning techniques. Pervasive Mobile Comput. **6**(2), 161–180 (2010)
3. Bicocchi, N., Castelli, G., Lasagni, M., Mamei, M., Zambonelli, F.: Experiences on sensor fusion with commonsense reasoning. In: IEEE Workshop on Context Modeling and Reasoning. Lugano (CH) (2012)
4. Bonabeau, E., Dorigo, M., Theraulaz, G.: Swarm Intelligence: From Natural to Artificial Systems. Oxford University Press, London (1998)
5. Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J.R., Mellouli, S., Nahon, K., Pardo, T., Scholl, H.J.: Understanding smart cities: An integrative framework. In: IEEE Hawaii International Conference on System Sciences. Maui (HI), USA (2012)
6. Conti, M., Das, S., Bisdikian, C., Kumar, M., Ni, L., Passarella, A., Roussos, G., Trster, G., Tsudik, G., Zambonelli, F.: Looking ahead in pervasive computing: Challenges and opportunities in the era of cyberphysical convergence. Pervasive Mobile Comput. **8**(1), 2–21 (2012)
7. Fernandez-Marquez, J., Serugendo, G.D.M., Montagna, S., Viroli, M., Arcos, J.: Description and composition of bio-inspired design patterns: a complete overview. Nat. Comput. **12**(1), 43–67 (2013)
8. Fogg, B.: Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann, Amsterdam (2002)
9. Girardin, F., Blat, J., Calabrese, F., Fiore, F.D., Ratti, C.: Digital footprinting: Uncovering tourists with user-generated content. IEEE Pervasive Comput. **7**(4), 36–43 (2008)
10. Gong, X., Liu, X.: A data mining based algorithm for traffic network flow forecasting. In: International Conference on Integration of Knowledge Intensive Multi-Agents Systems. Boston (MA), USA (2003)

11. Helaoui, R., Riboni, D., Stuckenschmidt, H.: A probabilistic ontological framework for the recognition of multilevel human activities. In: ACM International Joint Conference on Pervasive and Ubiquitous Computing. Zurich (CH) (2013)
12. Holldobler, B., Wilson, O.: The Superorganism: the Beauty, Elegance, and Strangeness, of Insect Societies. W. W. Norton and C, New York (2009)
13. Hu, X., Wang, W., Leung, V.: Vssa: A service-oriented vehicular social-networking platform for transportation efficiency. In: International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications. New York (2012)
14. Kanhere, S.: Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. In: IEEE International Conference on Mobile Data Management. Bengaluru, India (2012)
15. Kehoe, M.: Understanding IBM Smart Cities. Redbook Series. IBM Corporation, New York (2011)
16. Klein, M.C.A., Mogles, N.M., van Wissen, A.: Why won't you do what's good for you? using intelligent support for behavior change. In: Salah, A.A., Lepri, B. (eds.) HBU, Lecture Notes in Computer Science, vol. 7065, pp. 104–115. Springer, New York (2011). URL http://dblp.uni-trier.de/db/conf/hbu/hbu2011.html#KleinMW11
17. Lathia, N., Pejovic, V., Rachuri, K.K., Mascolo, C., Musolesi, M., Rentfrow, P.J.: Smartphones for large-scale behavior change interventions. IEEE Pervasive Comput. **12**(3), 66–73 (2013)
18. Mamei, M.: Applying commonsense reasoning to place identification. Int. J. Handheld Comput. Res. **1**(2), 36–53 (2010)
19. Mitchell, M.: Self-awareness and control in decentralized systems. In: AAAI Spring Symposium: Meta-cognition in Computation. Palo Alto (CA), USA (2005)
20. Rahwan, I., Dsouza, S., Rutherford, A., Naroditskiy, V., McInerney, J., Venanzi, M., Jennings, N., Cebrian, M.: Global manhunt pushes the limits of social mobilization. IEEE Comput. **46**(4), 68–75 (2010)
21. Rana, R., C., Chou, Kanhere, S., Bulusu, N., Hu, W.: Ear-phone: An end-to-end participatory urban noise mapping system. In: International Conference on Information Processing in Sensor Network. Stockholm, Sweden (2010)
22. Rosi, A., Mamei, M., Zambonelli, F., Dobson, S., Stevenson, G., Ye, J.: Social sensors and pervasive services: Approaches and perspectives. In: IEEE Workshop on Pervasive Collaboration and Social Networking. Seattle (WA), USA (2011)
23. Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: Real-time event detection by social sensors. In: World Wide Web Conference. Raleigh (NC), USA (2010)
24. Smart cities ranking of european medium-sized cities. In: http://tinyurl.com/bqh83np. Vienna, Austria (2007)
25. Tubaishat, M., Zhuang, P., Qi, Q., Shang, Y.: Wireless sensor networks in intelligent transportation systems. Wireless Comm. Mobile Comput. **9**(3), 287–302 (2009). DOI 10.1002/wcm.v9:3. URL http://dx.doi.org/10.1002/wcm.v9:3
26. Wurtz, R.: Organic Computing. Springer, Berlin (2008)
27. Yuen, M., Chen, L., King, I.: A survey of human computation systems. In: International Conference on Computational Science and Engineering. Vancouver, Canada (2009)
28. Zambonelli, F.: Toward sociotechnical urban superorganisms. IEEE Comput. **45**(8), 76–78 (2012)
29. Zapico, J.L., Turpeinen, M., Brandt, N.: Climate persuasive services: changing behavior towards low-carbon lifestyles. In: Chatterjee, S., Dev, P. (eds.) PERSUASIVE, ACM International Conference Proceeding Series, vol. 350, p. 14. ACM, New York (2009). URL http://dblp.uni-trier.de/db/conf/persuasive/persuasive2009.html#ZapicoTB09

Collective Intelligence in Crises

Monika Büscher, Michael Liegl, and Vanessa Thomas

1 Introduction

Collective intelligence is part of disruptive innovation in disaster response, that is, innovation that transforms the social, economic, political, and organizational practices that shape this domain [12,27,41,47]. One of the earliest examples of collective intelligence in this context arose during the Virginia Tech shootings, where students who had been told to stay in their dorm rooms connected online to work out who had been hurt or shot. Converging on a Facebook group called 'I'm OK at VT', the students exchanged information, verified reports and constructed accurate lists of who had been killed, several hours before the authorities released the same information. Under the pressures of the unfolding tragedy, they spontaneously developed social conventions and practical measures to ensure that information was accurate [63]. Since then, collective intelligence has been an integral part of wider transformations in crisis response.

'Crisis informatics' is a field of research that studies these transformations through interdisciplinary investigations of how members of the public use information technology and social media during crises [45]. A key insight derived from these studies is that local communities can be connected through complex communicative networks and in crises extend links to national and global communities, including diasporas, globally distributed 'crowds' of digital volunteers and emergent 'digital humanitarian organizations' who perform increasingly important

M. Büscher (✉) • M. Liegl
Centre for Mobilities Research, Department of Sociology, Lancaster University, Lancaster, UK
e-mail: m.buscher@lancaster.ac.uk; m.liegl@lancaster.ac.uk

V. Thomas
HighWire Centre for Doctoral Training, Lancaster University, Lancaster, UK
e-mail: v.thomas1@lancaster.ac.uk

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_12

responsibilities of gathering, verifying, geo-locating and mapping information from afar (such as CrisisMappers, Standby Task Force (SBTF), Humanity Road, and Virtual Operations Support Teams (VOST)) [54]. This can support faster and more detailed awareness of the needs of affected communities and the nature and extent of damage, which makes the public use of social media interesting as an informational service for official emergency responders. Collective intelligence is an integral part of this in two ways. Firstly, digitally connected crowds, networks and communities literally produce 'intelligence' about an incident taking pictures and posting situation reports online 'from the ground'. Secondly, volunteers enter into complex collaborative engagements to crowdsource, verify, map, list, aggregate and analyze information and make it available to others. The concept of 'social' collective intelligence is in some sense tautological (how could a collective activity not be social?). However, the concept draws attention to two important dimensions of the sociality of collective intelligence:

1. The social practices involved in producing intelligence and in collectively reasoning about it.
2. The societally transformative momentum that such practices exert.

In this chapter, we explore both dimensions. A richer understanding of the sociality of collective intelligence may help find answers to the question of how bridging between collective intelligence community-based and official emergency response efforts might be enhanced and how IT innovation can support this. We proceed through a selective review of related research to provide a background for a set of three recent examples where social media have been used to mobilise and organise different forms of collective intelligence. In the discussion that follows, we explore positive and negative frictions and avenues for innovation. The chapter concludes with a brief summary.

2 Background

Ulrich Beck's 1999 landmark diagnosis of a 'World Risk Society' [2] has led into a twenty-first century that has been labeled a 'Century of disasters', following a Royal Society report [18]. Humans are deeply implicated in both the effects and causes of disasters, whether they be due to storms, droughts, flooding, accidents or conflict and violence. Indeed, recent discussions suggest that we have reached a new era in the history of the earth—the 'anthropocene'—where 'the human imprint on the global environment has now become so large and active that it rivals some of the great forces of Nature in its impact on the functioning of the Earth system' [58]. Beck and other scholars in the sociology of risk argue that disastrous technological accidents (such as Chernobyl or Bhopal) and environmental threats (such as climate change) have engendered growing public awareness of this human responsibility [25]. The resulting changes in forms of public engagement in debates about risk and science and technology have, we would argue, prepared the ground for the emergence of collective intelligence as an important contemporary phenomenon. In risk societies, modern science has lost its monopoly on the production of knowledge and truth. Knowledge is no longer solely bound to professional expertise, and diverse new publics such as environmental movements or patients' rights movements are demanding a voice in science and technology decision making [38]. Society's relationship with science has become ambivalent, oscillating between blaming science and technology for ecological, technological and health crises and at the same time seeing it as potentially the only solution.

The crisis of science has re-assembled the relationship between experts, the media and the public and led to the emergence of new kinds of publics, making their own claims to legitimate knowledge, and demanding a place at the table of fact production. These new enactments of citizenship introduced new interactive and collaborative practices where concerned and affected citizens become participants in scientific research and media debates [25, 38]. The social web has amplified this reconfiguration of public engagement by combining the affordances of mass media and social networks. The hashtag function in twitter is of particular importance allowing instant formation of 'ad hoc issue publics' around certain topics as well as their equally speedy dispersal [10].

Yet, many analysts are skeptical about the practical and political leverage of such publics. Clay Shirky, for example, warns that participation in online communities does not translate into organizing groups for change, 'because participation in online communities often provides a sense of satisfaction that actually dampens a willingness to interact with the real world' [53]. Jodi Dean talks of 'communicative capitalism', where a concern with expression and circulation of messages replaces a commitment to listen, respond and engage in debate [16]; and Jaron Lanier argues that 'collectives can be just as stupid as any individual, and in important cases, stupider' [29]. Social media may support the performance of the democratic entitlement to an opinion, but ultimately be inconsequential for practical and politically democratic action, merely fueling the proliferation of messages addressed at others without genuine support for listening to and debating with these others, deliberating and taking considered collective action (amongst the members of ad hoc issue publics, the twitter convention of addressing others through their @name may be inadvertently symbolic of a practice of speaking at, and not with, each other).

In the field of disaster response, there is some evidence that a rise in digital volunteering is accompanied by a decline in real world volunteering, especially in urban areas, where anonymous neighbourhoods, fear over liability for damage or misconduct and high expectations of public services combine to prevent members of the public to take responsibility in emergencies [48]. However, at the same time, there is extremely vigorous social and technical innovation in and around 'digital humanitarianism' in major crisis events. The development of activities that connect online and offline, such as geo-tagging, location based social networking, and micro-blogging, makes it hard to accept wholesale criticism of social media publics as practically and politically ineffective or even corrosive. Many online activities now maintain a close connection to activities in the 'real' world, and this is especially true for citizen science and crisis response. Proponents of citizen science celebrate the potential of engaging members of the public, frequently using qualitative and economic arguments: 'We can employ citizens to gather data that we cannot get any other way . . . we can't afford to hire enough research assistants . . . to gather data on a larger geographical scale' [13]. Similar motivations apply in disaster response where attempts to leverage collective intelligence to enhance and augment 'situation awareness' have become an area of intensive social and technological innovation.

The concept of 'collective intelligence' commonly describes two activities: (1) data collected by collectives and (2) self-organising, synergistic collective reasoning [3, 31]. Discussions of these activities abound in both popular and academic literature, ranging from descriptions of crowdsourcing and micro-tasking through platforms such as Amazon's 'Mechanical Turk' [3], to concepts that posit the emergence of new forms of collective cognition or 'we-think' around examples such as Wikipedia and Alternate Reality Games [30] or Rheingold's 'smartmobs', who use digital technologies to coordinate protests or campaigns [49], and concepts that focus on peer production (e.g. of open source software) in new digital economy 'commons' [3].

In disaster response, crowdsourcing 'actionable' information is one of the key tasks pursued by pioneers in digital volunteering. At the time of writing one example includes the search for Malaysian Airlines Flight 370, which disappeared on 8th March 2014 over the Indian Ocean. More than 3 million globally distributed digital volunteers are participating in efforts to find debris across a vast area of land and ocean by poring over satellite photography, often from their homes. They have examined 'over a quarter-of-a-billion micro-maps and have tagged almost 3 million features in these satellite maps' [35]. The company that provided the satellite photography, Tomnod, is coordinating the search, triangulating between the tagged images to identify areas of greatest consensus amongst the crowd. Activities like these are described as 'collective intelligence', because they break a complex task down into 'micro-tasks', such as identifying objects that could be debris from a plane crash, in a way that can become part of larger efforts of complex reasoning, the solving of a larger 'puzzle'. Coordination often involves centralised control over the goal and work process, a relatively narrow set of motivations and incentives (in this case altruistic and ludic, in other contexts there may also be micro-payments), and it takes place within organizational contexts. Mainstream organisations that orchestrate crowdsourcing collective intelligence also include *Innocentive*, a commercial broker organization that liaises between clients with complex problems ('seekers') and the crowd [3]. In disaster response, crowds as well as digital humanitarian organisations and Virtual Operations Support Teams and their crowds often form around members of technical, humanitarian or gaming communities, but they also mobilise large numbers of ordinary everyday social media users.

The Virginia Tech crisis informatics study we mentioned in the introduction shows that such crowdsourcing is part of collective intelligence in crises, but it is not the whole story. Complex social practices of interpretation, coordination, information verification and aggregation are necessary, and they are a key element of more complex self-organised forms of collective intelligence. Pierre Lévy describes synergy of collective reasoning as the hallmark of collective intelligence 'a form of universally distributed intelligence, constantly enhanced, coordinated in real time, and resulting in the effective mobilization of skills ... [where] no one knows everything, everyone knows something ...' [31]. In a recent ethnographic study of one of the major digital humanitarian organisations, *Humanity Road*, Starbird and Palen examine emergent social and cultural practices of such coordination in working and sustaining a virtual 'Disaster Desk' [56]. They detail sophisticated communicative practices and mobilization of a plethora of digital tools (from Skype to Google docs) that allow episodic participation from large groups of individuals with highly diverse skills and knowledge situated across widely different contexts, enabling them to come together to support disaster response through information that can enhance understanding of unfolding events ('situation awareness') and support those 'sheltering in place'. An important element of these practices are forms of 'curating' contributions, which involves a number of different roles, reaching from 'trainers', who show the crowd what kind of information is valuable and how it should be produced, to 'archivists', who find, collect and aggregate information, to 'librarians', who organise, classify and categorise it, to 'storytellers' and 'editors', who filter, prioritise and contextualise it, and 'docents', who can facilitate best use of the archive [32].

The self-organising collective reasoning outlined by Lévy and detailed by peer production and crisis informatics scholars can complement but also contrast with the popular concept of 'Wisdom of Crowds' put forward by James Surowiecki [59], which underpins many crowdsourcing-focused definitions of collective intelligence. Unlike socially organised collective reasoning, wisdom of crowds is produced through the aggregation of mass produced data such as estimates of weight or value. In a famous example, Surowiecki describes how, in the immediate aftermath of the Challenger disaster, stock owners began to sell their shares in the four major companies involved in the building of the space shuttle. Morton Thiokol, which had manufactured the O-ring seals whose failure caused the explosion of the shuttle, lost nearly 12 % in one day. Surowiecki implies that the crowd of share owners correctly deducted that Morton Thiokol bore responsibility and that this would affect their market value. He suggests that when it comes to complex problems, there seem to be mechanisms that have similarity to the market principle of the invisible hand. 'Wisdom' is, according to Surowiecki, produced when a large number of people each enter their own calculations *without influencing* each other's findings. He suggests that independent individual reasoning is key to accuracy: 'ask a hundred people to answer a question or solve a problem, and the average answer will often be at least as good as the answer of the smartest member' [59]. In this model, intelligence is conceptualised as a purely individual capacity and reasoning as an individual practice. The 'added value' of the collective is seen as providing a critical mass of contributions to calculate averages.

Lévy's model, in contrast, focuses attention on reasoning as a social collaborative practice, and collective intelligence as involving social, collaborative deliberative processes that emerge in online communities as participants listen, share information, correct and orient towards each other, and coordinate their activities. At the heart of Lévy's synergistic collective reasoning are social mechanisms for participation, recognition given by peers, and mechanisms for effective self-governance [3,30]. Studies in crisis informatics detail such practices and add insight into practices of recipient design of contributions, their sequential organization and practices of listening. The study of the students' response to the VT shooting, for example, identifies practices of subtly documenting access to privileged information (such as information about boyfriend/girlfriend relationships) amongst the students and of demonstrating that a 'best attempt' at providing correct information has been made, e.g. by providing contextually authoritative sources. Early studies like this emphasised that collective intelligence is 'best understood as being emergent and collective rather than orchestrated' [63], which suggests that collaboration and coordination are at their best when they are self-organised. More recent investigations into strategies of 'stewarding the commons' [56], information 'curation' [32], the formalization of collective intelligence for disaster response [54] and attempts to design IT support for such strategies [23, 55, 57] shed doubt on the analytical and practical utility of contrasting self-organisation with orchestration. More recent insights into the detail of collective intelligence practices suggest that emergence and orchestration may actually complement each other. By drawing on insights from the field where collective intelligence practices have been established for the longest time, namely studies of online and alternate reality gaming, we can perhaps sidestep these contradictions.

Reporting on her role in one of the most celebrated examples of collective intelligence, the alternate reality game 'We love bees', Jane McGonigal states:

> *I was . . . one of four puppet masters designing the live missions . . . The gamers exercise of free will has long been assumed to be a core aspect of gaming. But the rise of the puppet master . . . suggests that in the new ubiquitous computing landscape, many gamers want to experience precisely the opposite . . . [34]*

McGonigal suggests that what participants in collective intelligence efforts seem to need above all is careful orchestration by people who 'move with' the participants, able to spot and encourage positive emergent behaviour and discourage behaviour that does not suit the overall aims. In the quote above she contrasts 'free will' which would be central for the emergence of self-organization and 'puppetmastering' a strong form of orchestration, but later on in the same article, she qualifies the relationship:

> *The first time I told this story at a lecture, an audience member challenged me: "You puppet masters must really get a kick out of manipulating these players to do whatever you want. That must be such a power trip." But in fact, the exact opposite was true. We didn't get a rush of power . . . We actually felt completely out of control. We had worked so carefully to craft just the right text for our mission scripts, and yet from the very first moment of gameplay, our actual, effective authority was stripped away. Yes, we could give the players a set of instructions but clearly we could not predict or dictate how they would read and embody those instructions. We were absolutely not in control of our players' creative instincts. [34]*

McGonigal's reflective analysis of her and her colleagues' actual experience of 'puppetmastering' shows that the term is misleading. 'Puppetmasters' do not have complete control and power over players. On the contrary, the game designers and puppetmasters actually needed to orchestrate the game in a way that was extremely responsive to the creative interpretation of instructions by the players. These analyses suggest that both orchestration and emergence based on individual creativity must be supported for collective intelligence endeavours to self-organise successfully.

The link between online and 'real' world activities in crisis situations adds another dimension to this discussion. Crisis informatics can service practical self-organised mobilisation and coordination of local resources, knowledge, and efforts *in situ*. During the floods in Germany in 2013, for example, 29 % of Twitter-messages focused on coordinating help and resources locally [67]. Reports from sandbag filling stations appeared alongside calls for help and a crowdsourced map of the current need for volunteers in different places [39]. Lüge [33] suggests that these examples index a shift in the use of social media for emergency management. It seems that the informational service function for official response that can be addressed by crowdsourcing and coordinating digital volunteers is increasingly complemented by a practical service function where local community help and resources are crowdsourced and coordinated by both local and potentially globally distributed digital volunteers, where, for some, work in online and real world spaces can be combined. Yet, research in crisis informatics is still mainly focused on understanding and developing means for *extracting* more valuable, reliable, 'actionable' information from social media for enhanced situation awareness, especially for professional responders. There are only a small number of studies that explore how *self-organizing* might be supported through collective intelligence. These studies are beginning to highlight positive and negative friction between self-organisation and orchestration amongst members of connected local and digital publics and the professional response (which is also both improvised and orchestrated [36]). They include studies of how connected communities coordinated the mobilization of resources during the 2010 Haiti earthquake [55], the 2011 Norway Attacks [46], and the 2011 'Super Outbreak' tornado in Alabama [48].

This review of existing literature documents a multiplicity of social practices involved in collective intelligence in crises which can be broken down into four main activity types:

1. **Gathering** the activity of crowd-sourcing 'intelligence' about disasters through mass public participation in sensing, documenting, defining, and collecting relevant data;
2. **Reasoning** making sense of information and needs, analyzing data and making information useful or 'actionable' for affected populations and professional emergency responders, by leveraging individual and collective capacities of information processing, mobilizing different knowledge and skills;
3. **Curating, stewarding and orchestrating** defining strategies to identify information needs of affected populations and emergency responders, monitoring and guiding information production, providing incentives and coordinating training,

collection, archiving, categorization, aggregating, assembling, analyzing, filtering of information, visualizing it in maps and reports, and facilitating their use;

4. **Acting** coordinating resource mobilization *in situ* through pulling or feeding aggregated information to official responders or local volunteer communities.

These activities have made social collective intelligence an important force in connecting people within and beyond local communities in disaster situations, which has begun 'to fundamentally alter the very nature and arc of emergencies' [47].

3 Examples

A discussion of examples will now illustrate some core dimensions of these transformations and related collective intelligence phenomena in the context of emergency response practices. The focus is on the interface between self-organised community efforts and official efforts during the response phase of crisis management.

3.1 The Haiti Earthquake

January 2014 was the 4th anniversary of the Haiti earthquake, where over 220,000 people were killed and over 300,000 were injured. The earthquake made more than 1.5 million people homeless, and resulted in an 'immense humanitarian crisis, highlighting long-lasting development challenges' [44]. With the Haiti earthquake two important and related things changed in disaster response: self-organised mass-reporting with digital media took place in unprecedented numbers and at the same time 'online communication enabled a kind of [global] collective intelligence to emerge' [42]. Thousands of volunteers from all over the world

> *aggregated, analyzed, and mapped the flow of messages coming from Haiti. Using Internet collaboration tools and modern practices, they wrote software, processed satellite imagery, built maps, and translated reports between the three languages of the operation [...] [42]*

Volunteers coordinated some of these efforts via formalised crowdsourcing tools, including OpenStreetMap and Ushahidi. It was their use of the latter tool, Ushahidi, that marked a milestone in the development of crisis informatics for humanitarian emergency response. Ushahidi is a free, open-source crowdmapping tool that was initially developed in the aftermath of the 2008 elections in Kenya [1]. Ushahidi relies on the power of the crowd. Anyone can contribute to an Ushahidi map by using social media, text messages and the Ushahidi website to share geographically tagged information, news stories, videos and pictures. By mapping this information, the software helps people make sense of complex situations. When a team of international volunteers decided to deploy Ushahidi in Haiti following the earthquake, its novel ways of crowdsourcing and mapping information were applied to a complex crisis. Over 4,000 volunteers contributed to the Ushahidi Haiti Project (UHP) map, and their work provided valuable support to a number of in-the-field organisations, including the US Marines and the United Nations Disaster Assessment Search and Rescue teams [40]. The UHP map even supported the task of deploying resources to people in need. Morrow, Mock, Papendieck and Kocmich, for example, describe how the Department of State Analysts for the US government inter-agency task force and US marines used UHP information to enhance situation awareness and identify 'centers of gravity' for the deployment of field teams [40].

However, the Ushahidi Haiti Project was not the only example of social collective intelligence following the Haiti Earthquake. Innovations like Project Epic's 'Tweak the Tweet' (TtT), a standard which suggests a uniform format for reports through hashtagging needs, locations and contact details, promoted a shared 'grammar' that facilitated computational parsing and mapping of tweeted information [55]. Starbird and Palen observe how volunteer translators or 'voluntweeters' translated reports from different sources, such as text messages or tweets, using the TtT syntax in response to the Haiti crisis, and worked as 'remote operators' to facilitate assistance, resource coordination and collaboration from a distance. Amongst other things, they promoted the international transfer of small funds via Paypal to many Haitians' pay-as-you-go mobile phones, and even coordinated the provision of trucks to specific locations and local volunteers, with messages sent back and forth, including confirmation of resolution of resource coordination challenges.

However, despite the successes Morrow et al. and Starbird and Palen note, they also found significant barriers to the use of microblogging by official responder agencies. They quote one of their most experienced emergency responder interviewees as describing UHP as 'a shadow operation that was not part of the emergency response plan' [40]. Further to this, Starbird and Palen [55] describe how voluntweeters felt frustrated and 'obstructed when the "formal" response moved into place'. One of the most challenging issues for integrating local, online and official response was the reliability of information. Despite the fact that many experts and government organisations like the US Federal Emergency Management Agency, the Department of State as well as international organisations like the United Nations Office for the Coordination of Humanitarian Affairs agree that integration of digital volunteers and humanitarian organisations with formal emergency response efforts is invaluable, and while they are establishing interfaces to grassroots networks, there are serious obstacles:

> *Federal agencies are legally obligated to provide data that are accurate, reliable and useful. They must take steps to ensure the integrity of information . . . prevent the release of data that breach the privacy or security of citizens or organizations, violate nondisclosure agreements, or endanger national security [15].*

These constraints are hard to overcome. We will explore them further in our discussion, but before we turn to this, two more examples will draw out important political and problematic aspects of collective intelligence in crises.

3.2 Flooding in Alberta, Canada

In June 2013 the Canadian province of Alberta was hit by sudden and unprecedented flooding. During a 48-h period, the floods left four dead, forced over 100,000 people to evacuate their homes, and caused over $5 billion CAD in damage [50, 64]. The floods forced Alberta to declare its first-ever State of Provincial Emergency, with 29 communities classified as in a state of emergency [20]. One of those communities was the City of Calgary, the third largest city in Canada, which experienced severe damage to its hospitals, roads, bridges, schools and water treatment facilities. As the disaster unfolded, the city also experienced the unifying potential of social collective intelligence.

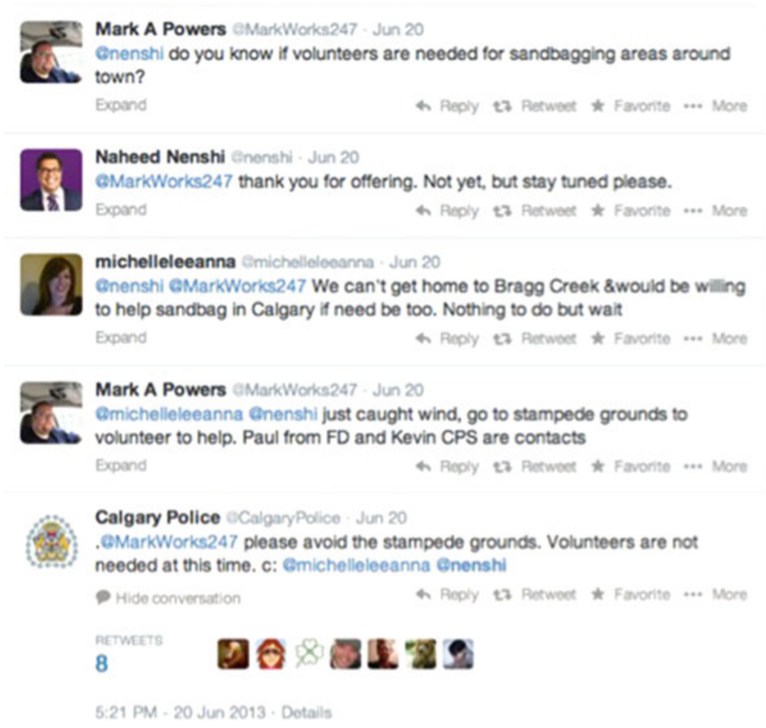
At a very early stage in the flooding, Naheed Nenshi, the Mayor of Calgary, committed to actively and regularly communicating with residents about the municipality's disaster response and recovery efforts [11]. Although Nenshi hosted regular television conferences, his Twitter and Facebook accounts became two of the primary sources for news updates and resource coordination. Calgarians took notice. Nenshi effectively became one of the 'puppetmasters' in the coordination of the emergency response, contributing to the orchestration of a combined community-based and professional effort, which in turn also supported his success as a politician. Between 19 and 30 June, Nenshi's Twitter followers jumped from 94,000 to 122,000 [62]. During that same period, his followers tweeted at him over 89,000 times [65], often with logistical questions and concerns. He would respond quickly, efficiently and often using creative and popular hashtags, such as #yyc, #abflood and #yycflood.[1] But his followers also used creative hashtags to communicate directly with Nenshi, highlighting personal and affective aspects of engagement in collective intelligence in crises when they used a hashtag (#nap4nenshi) to plead with him to take a nap after working for 43 h [8].

When Toronto faced a (much less severe) flooding crisis several weeks after Calgary, things turned out differently. The already discredited Toronto mayor Rob Ford (who had been in the news for drug abuse allegations) attempted to follow in Nenshi's footsteps and use twitter to address the crisis. The first round of criticisms for Ford came when the Toronto Mayor Ford @TOMayorFord account tweeted that the worst was over, hours before the rainfall peaked, using the wrong measurements for rain, and deleting the tweet soon after, which was detected and highlighted by Toronto Star reporter Daniel Dale [61]. Things got worse, when 'Toronto Sun reporter Don Peat described that the mayor was with his kids and in his SUV 'rather than coordinating disaster relief, informing the public or whatever it is big city mayors do in times of crises' [22].

In Calgary, Nenshi was not alone in his efforts during the floods. He worked directly with the City of Calgary's Emergency Management Agency (CEMA) and

[1] A Canadian twitter convention to tag places is to use airport codes for referring to cities. For example, 'YYC' is the airport code of Calgary, so Calgary is #yyc, Edmonton is #yeg, Toronto is #yyz, Vancouver is #yvr.

the Calgary Police, which used their Twitter and Facebook accounts to support Nenshi's efforts and also to share service-specific information, often responding directly to requests from members of the public (Fig. 1).



Mark A Powers @MarkWorks247 · Jun 20
@nenshi do you know if volunteers are needed for sandbagging areas around town?

Naheed Nenshi @nenshi · Jun 20
@MarkWorks247 thank you for offering. Not yet, but stay tuned please.

michelleleeanna @michelleleeanna · Jun 20
@nenshi @MarkWorks247 We can't get home to Bragg Creek &would be willing to help sandbag in Calgary if need be too. Nothing to do but wait

Mark A Powers @MarkWorks247 · Jun 20
@michelleleeanna @nenshi just caught wind, go to stampede grounds to volunteer to help. Paul from FD and Kevin CPS are contacts

Calgary Police @CalgaryPolice · Jun 20
.@MarkWorks247 please avoid the stampede grounds. Volunteers are not needed at this time. c: @michelleleeanna @nenshi

RETWEETS
8

5:21 PM - 20 Jun 2013 · Details

Fig. 1 Twitter Conversations around the Calgary Floods. From https://twitter.com/search?q=from%3Acalgarypolice%20%40nenshi%20since%3A2013-06-19%20until%3A2013-06-27&src=typd&f=realtime [Accessed 9th April 2014]

Calgary Police and other established emergency response agencies also used Twitter extensively. For example, on the same day, they tweeted the following message:

> **@CalgaryPolice** (20 June 2013 10:39 PM): Due to #yycflood we are unable to take any non-emergency calls. Please save your calls until the state of emergency has been lifted. #yyc

Although the impact of that tweet on 911 calls was not tracked, it was retweeted 136 times and likely reached thousands of people. In a similarly untracked but clearly effective tweet, CEMA used the City of Calgary's twitter account to issue the following call for volunteers:

> **@cityofcalgary** (24 June 2013 6:24 AM): Ready to volunteer? If you're 18 or older, meet up at McMahon Stadium at 10 a.m. Info is here: http://ow.ly/mkdW8 #yychelps #yycflood

With only three and a half hours between the time of the tweet and the launch of the volunteer event, the City hoped that 600 volunteers would arrive at McMahon Stadium. However, after the tweet was shared on Twitter and Facebook, over 3,000 people arrived to offer their help [7]. The unexpected reach of and overwhelming response to the call for help was one of the first clear indications to the official responders that the residents of Calgary were organizing their efforts by using the #yychelps hashtag.

In the early days of the flood, Calgarians who were asking for and offering help also used #yychelps to connect with one another. People used the hashtag to share resources, including heavy-duty equipment and food, as well as to publicise examples of illegal price gouging, which occurred when stores sold goods at a higher price than usual to take advantage of the crisis. To make IT-enabled citizen coordination efforts easier, a small group of Calgarians eventually created a website, Twitter account and Facebook page that shared the same name as the hashtag, YYCHelps. It became one of the central community hubs for coordinating resources, for listing volunteer opportunities, links to municipal resources (e.g. the City of Calgary's road closures map), and information about existing community initiatives, such as citizen-coordinated food kitchens, offers of temporary housing and fundraising events [66]. They put out calls via the #yychelps Twitter hashtag for volunteers who were willing to donate time, skilled trades and heavy duty equipment, and every call was met by hundreds of volunteers [7]. Through this work, they transitioned into a self-organizing connected community, and one that crossed geographies and social boundaries. Just outside of Calgary, severe flooding also hit the Siksika First Nations reserve; however, official responders and the media largely ignored the disaster here until a call for help was posted on Facebook.[2] A link to the Facebook post was shared on Twitter using the #yychelps hashtag, and the situation quickly changed. The #yychelps community coordinated food, clothing and temporary shelter for displaced residents, and then demanded increased media coverage of the crisis there.

3.3 The Boston Marathon Bombing

Our final example brings out some more challenging issues. The annual Boston Marathon came to a sudden end on April 15, 2013 when two bombs exploded close to the finishing line, killing three people and injuring an estimated 264 others [28]. Within hours, the FBI called upon bystanders to submit their photographs and videos from the event, triggering a massive 'crowdsourced intelligence gathering' [28]. Two days later the police released a photograph of one of the suspects and asked the public for help in identifying him. But within these two days, the 'digital bystanders' had not waited patiently. They had already turned to 'crowdsourced crime solving' [57], analyzing image content, collecting clues and listening to

[2] https://www.facebook.com/SiksikaAbFlood2013Info/posts/143125142550831?stream_ref=10.

and posting recordings from the police scanner. This was largely organised on social news and activism websites, 'Reddit' and '4chan'. When a tweet noted a resemblance between the suspect on the police photo and a tweeter's former classmate, his name was posted on Reddit along with another name from the police scanner. This resulted in this widely retweeted tweet:

> **@ghughesca** (April 19, 2:43pm): BPD has identified the names: Suspect 1: Mike Mulugeta. Suspect 2: Sunil Tripathi. [cited in [57]]

For a short time the crowd detectives celebrated this as a victory: 'Reddit solved the bombing. Before the Feds' [60]. But soon the FBI and news outlets released completely different names for the real suspects: the Tsarnev brothers; exposing the crowd as 'digital vigilantes' who had spread rumours slandering two innocent men [57].

The crowdsourced manhunt after the Boston bombing highlights some of the risks officials take when collaborating with volunteers. It also showed that there is good reason for the media to be more cautious of using crowdsourced intelligence as a source. Speculation by digital volunteers led reputable media organizations and news agencies to effectively disseminate misinformation. This, too, was initially celebrated as a victory by some members of the crowd. Greg Hughes (@ghughesca), who was one of the first to spread the wrong names, for example, said: 'Journalism students take note: tonight, the best reporting was crowdsourced, digital and done by bystanders' [60]. The effect of this reporting and its spread into even highbrow mainstream media was highly problematic. The family of Sunil Tripathi especially suffered severe anguish as a result of his being implicated. The 22-year-old had committed suicide and was missing when he was named as a suspect. As his family desperately searched for him and his name was associated with the Boston Marathon bombing on twitter, doors began to close. One homeless shelter the family enquired at is reported to have told them 'we do not aid terrorists' [17]. Reflections amongst the media in the aftermath of this confusion call for higher 'benchmarks for reliability and truth-telling through a revival of journalism based upon ethics and humanity' [17]. Such calls echo calls from digital humanitarian organizations and practitioners, who have begun to formulate ethical codes of conduct. There are calls for a 'code of ethics' for social media use in crises [47] and some early formulations of 'Twitter Commandments' for 'voluntweeters', providing 'guidance about sorting accurate from inaccurate 'rumour', and for "tweeting responsibly" during disasters' [56], as well as guidelines for crowdsourcing information from populations affected by conflict [24].

4 Discussion

The use of social media for self-organised mobilization of knowledge, resources and self-help in crises by nested digital and local communities raises opportunities for positively disruptive innovation in emergency response as well as challenges. The turn to collective intelligence to augment local communities' capacity for self-help can help address needs more swiftly and effectively. This is extremely useful as economic pressures, increased frequency and severity of disasters, heightened vulnerability through ageing infrastructures and populations, coupled with a generation change in the emergency services are creating a 'new reality' for these services [27, 41]. This is characterised by a need to increase efficiency, meeting higher demands with fewer resources and a less experienced but more technology-savvy workforce. In this new reality, enhanced community resilience presents new economic, social, political, legal and ethical openings. Some see the future of emergency response in spreading the burden of responsibility by engaging communities more closely. The US Federal Emergency Management Agency (FEMA), for example, argues that natural or man-made crises (floods, storms, violent attacks) can be addressed better with a 'Whole Community' approach, where 'officials can collectively understand and assess the needs of their respective communities' and communities can play an active part in emergency planning and management [19]. In some sense, this acknowledges communities as an agency in multi-agency crisis management. However, for established emergency response organizations it is practically and politically difficult to switch from approaches focused on protecting and managing the public to engaging with communities. This is exacerbated by the fact that their notion of a clearly defined community whose needs can be assessed by 'their' respective officials is outdated, for communities are dynamic, their commitment to volunteering seems to be waning [48], and it is misleading to think of communities as purely local when they are potentially globally connected and capable of mobilizing global collective intelligence, especially in disasters.

There is significant research regarding the 'curation' and 'orchestration' of crowdsourced forms of collective intelligence for situation awareness. Practitioners and researchers already analyse and address social, political, economic, ethical and legal issues, ranging from approaches that identify misinformation through to analyses that show that information can undermine 'information superiority' and endanger operations [37, 43], lead to vigilantism [57], tort liability for civil wrongs for volunteers and various challenges for professional responders [51]. However, current research focuses on practices of information extraction and processing, and neglects practices of self-organised mobilization of resources by nested digital and local communities.

The examples above exhibit the momentum of social and technical innovation in relation to these practices of self-organised mobilization of knowledge and resources, and they highlight different dimensions of how new technologies emerged along with new practices of collective intelligence and emergency response, introducing new forms of agency and actors and provoking negotiation and contestation of competences and responsibilities. In this emerging new reality of emergency response we see six types of entities/agencies negotiating their relationships and roles:

- **Established response organizations**, whose roles are being renegotiated and who are under pressure from budget cuts, technological innovation, a generation change with large numbers of experienced senior personnel retiring, and rising expectations from the public, as well as heightened media scrutiny.

- **Elected officials**, who have always played a role in crisis communications, but who are now being placed under new demands of swift, decisive and visible interventions through social media.
- **Established media organizations**, who use social media as a source and a channel for disaster reporting and analysis
- **Digital Volunteers** acting as individuals ‘in the wild’, members of diasporas or otherwise connected to affected populations, or simply seeking to contribute something to the disaster response.
- **Digital humanitarian organizations**, emergent organizations, where individuals can come together, receive training and instruction, and act as part of collectives organised in networks, and communities, gathering, curating, orchestrating and processing crowdsourced information with a view to supporting official crisis response and community efforts.
- **Self-organizing connected communities**, who combine local with sometimes globally networked communications for improvised micro-coordinated mobilization of help, knowledge, resources and community efforts.

The central question in the negotiation of capacities and responsibilities is how all these agencies could coordinate their activities more productively and easily. This includes questions about when it is appropriate and when it may *not* be appropriate to work together, and questions about the ethics of collaboration. The examples can help us explore specific socio-technical aspects around these questions and opportunities for innovation.

In the first large scale mobilization of connected communities in the aftermath of the Haiti earthquake, distributed crowd communities were able to map affected areas and thereby provide a baseline for translating and mapping needs. A lesson from this effort was however, that there are limitations in terms of collaboration between digital volunteers and official responder agencies. Once the mapping was done, the officials more or less took over, which led to some degree of alienation on the volunteers’ side. At the same time, there are many open questions about the mapping. How can the reliability of the information provided by volunteer services such as the UHP be ensured? How to ensure that legal responsibilities can be met? Who can address all the needs that connected communities identify and how? Whose responsibility are the needs that are made visible? How does greater visibility of needs affect expectations from affected populations, local publics and global media publics? Who analyses the needs and defines what is to be addressed (first)? What counts as damaged and in need of (urgent) rebuilding? Are all people affected connected or are some left out of the loop technologically or otherwise? What efforts can be made to bridge digital divides? Locating and making more needs visible may seriously exceed the capacity of formal response organizations to address them and make self-help a necessity, as well as opening up more long term and political questions over the resourcing of crisis management and emergency response.

Moreover, the fact that many digital volunteers are located in the global (urban) North, volunteering for incidents in developing countries of the South, raises challenges. In her analysis of the aftermath of the Haiti earthquake, Mimi Sheller shows that disaster response logistics amplified North/South inequalities through measures 'in which the outsider has the power to move, to bring in supplies, to access information, or to come and go at will, while the local victim experiences . . . decreasing access to mobility, and high levels of random and turbulent serial displacement' [52]. She describes the physical and digital influx of highly mobile international responders with their ability of aerial surveys of damage and GPS-enabled satellite data collection systems coinciding with a local population which at large had neither the means nor the right to move outside the danger zone or to leave their country. Part of this unequal mobility manifested itself in the ability of foreigners such as the World Bank, but also the digital volunteers and crisis mappers, to make aerial images and access satellite data to assess the damage and ultimately (help) decide what needed rebuilding. They based this on 'an aerial view that few Haitians had due to lack of Internet access and (because they usually are not in a position to fly) will ever have of their own city', translating 'visual power through the aerial gaze' into material socio-economic and political decisions on the ground [52]. Sheller cautions, that

> *applications of virtual mobility via informational mobility are not innocent, but are directly related to the operationalization of mobility regimes that enable foreign travel into Haiti and foreign control of logistics, while largely preventing Haitians from leaving their country . . . [and] interfering with their self determination of rebuilding processes [52]*

The Calgary floods demonstrate how politicians are involving themselves directly and can, in a more positive manner, support collective intelligence to help organise a combined community-based and professional effort during a crisis. Official responders targeted their efforts based on information that citizens and the Mayor of Calgary shared via social media, and they, in turn, coordinated self-help initiatives with these official efforts. Communities, who were not being served by official responders or the media, such as the Siksika First Nations reserve community, could make themselves visible and connect with residents of Calgary. They then used social media to self-organise, coordinate and mobilise resources.

The Calgary floods also highlight challenges and opportunities for leadership in crisis. When diverse publics use social media effectively to produce an overview of the disaster, organise emergency relief, and often know about needs before formal responders do, stewardship of these efforts by local leaders, like in this case Mayor Nenshi can function as a catalyst. Tapping into the social media emergency response infrastructure may allow politicians to satisfy expectations and provide authentic hope and confidence, which as research has shown is 'how mass communication in crises is best done':

> *It should explain the crisis, its consequences and what is being done to minimize the consequences. It should also offer 'actionable advice,' explaining what should be done, by whom, and why. [5]*

Being able to 'play' the media can be crucial for a politician's reputation when crisis hits. Analysis of leadership styles often compare Rudy Giuliani's successful response to 9/11 with George W. Bush's widely considered failure in handling the flooding of New Orleans after Hurricane Katrina. It showed that the media and the public in such times are looking for compassionate, hands-on leadership on the ground, which Bush with his ‘principled’, rigid and managerial leadership style did not try and arguably could not have delivered authentically [4]. Social media offer opportunities for new ad hoc and often very active issue publics [10] that in some sense take out the middle-man of mainstream media for established emergency response agencies and politicians, allowing them to speak directly with members of the public. Calgary’s mayor Nenshi and Calgary’s established emergency organizations engaged with these publics effectively, but such engagement can also go wrong. In a crisis-induced ‘information storm’ officials are exposed to critical scrutiny. Toronto’s mayor Rob Ford’s efforts were compared unfavourably (mostly on social media) with Nenshi’s virtuous handling of the media. A large factor in Ford’s tweet-fail might have been pre-existing troubles, with the public seizing this opportunity to ridicule a politician who had already fallen out of favour. But this also showed that part of a politician’s successful performance is the ability to effectively link online and ’real world’ activities, to (micro-)publicise this in social media, to be in the trenches, tweet about it, and even tweet about tweeting about it.

The political fallout of information mobilised during collective intelligence endeavours in disaster response is closely linked to elected officials’ capacity to be aware of, motivate, orchestrate and integrate diverse efforts. They can become highly effective ‘puppetmasters’ who can nurture and channel collective intelligence by coordinating with established emergency response organizations and entering into a dialog with members of the public. This is not indicative of a spread of ‘communicative capitalism’.

However, highly problematic communicative practices do arise in the context of collective intelligence in crises. The Boston example highlights a need to distinguish more carefully between the different types of emergencies where collective intelligence can be employed and the different agencies involved. Unlike established emergency response organizations and news agencies, the social media crowd of digital volunteers is currently unorganised, untrained, unregulated, uncertified and largely anonymous. There are some effective social informational practices of self-regulation in collective intelligence in such groups, but these do not seem to function in ‘manhunt’ circumstances [57]. Equipped with images sourced from the ephemeral local community of visitors and participants in the marathon, the crowd launched into crowdsourced crime solving and falsely accused two innocent men. Collectives clearly do not necessarily produce intelligent behaviour or moral integrity, indeed they can be ‘stupider’ than individuals [29]. Debates over the intelligence and morality of crowds have a long tradition, for example, in the psychology and sociology of Le Bon and Simmel [6], and Lanier’s verdict regarding digital crowds echoes their debates. In crises, the dynamic of ‘clicktivism’ is powerful, feeding on and feeding into sensationalist media reporting and even vigilantism [57] in a way that raises questions about responsible social media use on behalf of all parties involved. The Boston example highlights the entanglement of social and technical innovation at the frontiers of crisis informatics and its transformational implications for the relationships between established responder organizations, the media, digital volunteers and self-organizing connected communities. It opens up questions over how collective intelligence may be leveraged in ways that are more ethically circumspect.

These questions have contributed to the emergence of digital volunteer organisations (like Humanity Road) and Virtual Operations Support Teams, who acknowledge that in order to assure reliability, build trust, detect and prevent misuse and manage phenomena of collective intelligence in crises more effectively, some professionalization is necessary [54, 56]. This mirrors the support for 'real world' volunteers in organisations such as the German Civil Protection Organisation (THW), which regularly trains thousands of volunteers. In relation to digital volunteers, it involves building identifiable and accountable organizational frameworks, upholding an ethic of care and information security, where members adhere to codes of conduct such as 'verify twice, tweet once' [56], and institutionalizing some organizations as non-governmental organizations. Such professionalization also supports forming more formal relationships with emergency managers and acting as a 'steward of the commons' (Hess & Ostrom in [56]). Early indications of such professionalization suggest that collective intelligence in crises can be enhanced through curation and quality control done by more formalised collective intelligence 'orchestration agencies'. At the same time, efforts are being made to explore how indigenous mechanisms of identifying and correcting misinformation may be computationally supported, for example through automatically detecting corrections as indicators of rumours or misinformation being spread [21, 37, 57], or through artificial intelligence solutions to making social media analysis more efficient and reliable [23].

5 Conclusion

In this chapter we have used the concept of social collective intelligence to highlight two dimensions of the sociality of collective intelligence: (1) the social practices involved and (2) the societally transformative momentum of these social collective intelligence practices, in the hope that a richer understanding of the sociality of collective intelligence can inform more socially and ethically circumspect social and technical innovation. We have argued that the World Risk Society has given rise to new practices and new technologies for public engagement in science and technology and, more recently, disaster response. As the twenty-first Century unfolds as a 'Century of disasters', digital humanitarianism is part and parcel of a transformation of social, economic, and political practices of disaster response. Two related forms of collective intelligence are taking shape in this context.

Firstly, crowdsourcing-focused collective intelligence describes the way in which local and globally distributed but connected communities can generate information that can be highly valuable for understanding the impact of disasters and the needs of affected populations, especially targeted at professional emergency response organisations. Secondly, connected communities can use collective intelligence to self-organise the mobilization of resources and self-help activities. Both forms have been studied within the field of crisis informatics, but the emphasis has so far often been on crowdsourcing forms of collective intelligence. Opportunities and challenges, such as the ability to micro-task large numbers of volunteers to gather, verify, analyze and aggregate information, along with threats of misinformation, rumours, and vigilantism have been discussed. In crisis informatics studies the focus is on the social practices involved, detailing how people subtly recipient design and tailor their contributions to indicate their relevance, authority and accuracy. Studies also highlight the difficulties arising in amongst a distributed, uncertified, unregulated and anonymous crowd. The professionalization of digital volunteering, the development of codes of conduct and self-regulation measures, and the development of computational support for the practices involved as well as their integration into official and community efforts are beginning to leverage the potential of crowdsourced forms of collective intelligence in crises. However, it is not clear how successfully such traditional forms of professionalization can be adapted to and enforced across globally distributed communities of episodic digital volunteers. Moreover, these innovation efforts in crisis informatics focus almost exclusively on leveraging crowdsourcing and wisdom of the crowd forms of collective intelligence. Only a few studies are beginning to explore how these can be dovetailed with self-organised improvisation and micro-coordination in connected communities. In these mixed online/'real world' efforts, self-organization and orchestration can complement each other and help multiple agencies established emergency response agencies, elected representatives, the media, digital volunteers, digital humanitarian organisations and connected communities to come together productively.

The most important insight arising from the examples and analysis in this chapter is that there is a need to engage and support local communities more deeply and seriously, and to produce technologies that can help with this. There is scope to build on experiences from the co-production of public services in other domains [9] to emergency response. Cole et al. cite Furedi to argue that

> *[A] highly centralized professional response cannot deal with every contingency. In the end, encouraging people to take responsibility for their own well-being is essential for an effective response to an emergency situation. [Furedi, cited in [14]]*

And they proceed to show that engaging communities and crowds might allow current thinking and practice to be extended beyond professional first response. However, this requires a transformation of crisis services to facilitate their opening up to citizen activities on the ground, as well as those that are digitally mediated and enabled. Connecting affected local populations more richly with digital volunteers and the other agencies involved in disaster response also provides an opportunity to counteract the perpetuation of neo-colonial unequal (im)mobility regimes and exploitation (with extremes documented in Naomi Kleins analysis of 'disaster capitalism' [26]), which can be an unintended consequence of the efforts of digital volunteers in cases like the 2010 Haiti earthquake. Integrating local communities would not only make the emergency response more intelligent, since locals possess context information necessary for sensibly interpreting aerial images of damage, it could also make the response fairer and more democratic. Integration will be easier where affected local communities and digital volunteers have access to the same kind of technology, and the same economic means and rights. From this perspective, for a situation such as that in Haiti, closing the gap between official response and digital volunteers seems a less pressing issue than the fact that affected populations are being excluded from shaping the response and decisions about rebuilding. Clearly the perceived responsibilities of digital volunteers and many digital humanitarian organizations stop well short of such questions. Four years after the earthquake, the United Nations find that 817,000 Haitians still need humanitarian assistance [44], yet most digital humanitarians have moved on to the next crisis. By supporting connected communities in self-organising and orchestrating self-help in the context of professional and volunteer efforts in a more targeted fashion, they gain more opportunity to put themselves on the map.

Acknowledgements The research presented here is part of the BRIDGE and SecInCoRe projects, funded by the European Union 7th Framework Programme under grant number 261817 and the Topic SEC-2012.5.1-1 Analysis and identification of security systems and data set used by first responders and police authorities.

References

1. Banks, K., Hersman, E.: FrontlineSMS and Ushahidi—a demo. In: Proc. 2009 Information and Communication Technologies and Development (ICTD) International Conference, p. 484, Doha, 2009
2. Beck, U.: Risk Society: Towards a New Modernity. Sage, London, Newbury Park and New Delhi (1992)
3. Benkler, Y., Shaw, A., Hill, B.M.: Peer production: a modality of collective intelligence. In: Bernstein, M., Malone, T. (eds.), Collective Intelligence (2013). Retrieved from http://mako.cc/academic/benkler_shaw_hill-peer_production_ci.pdf [Accessed 4th April 2014]
4. Boin, A., Hart, P.T., McConnell, A., Preston, T.: Leadership style, crisis response and blame management: The case of Hurricane Katrina. Public Admin. **88**(3), 706723 (2010)
5. Boin, A., Kuipers, S., Overdijk, W.: Leadership in times of crisis: a framework for assessment. Int. Rev. Public Admin. **18**(1), 79–91 (2013)
6. Borch, C.: Between Destructiveness and Vitalism: Simmels Sociology of Crowds. Conserveries Morielles, (#8) (2010). Retrieved from http://cm.revues.org/744
7. Bowman, J.: Calgary volunteers create YYCHelps.ca to organize cleanup (2013a). http://www.cbc.ca/newsblogs/yourcommunity/2013/06/calgary-volunteers-create-yychelpsca-to-organize-cleanup.html [accessed 6 March 2014]
8. Bowman, J.: Calgary Mayor Naheed Nenshi gets online hero status (2013b). http://www.cbc.ca/newsblogs/yourcommunity/2013/06/calgary-mayor-naheed-nenshi-gets-online-hero-status.html [accessed 6 March 2014]
9. Brandsen, T., Pestoff, V.: Co-production, the third sector and the delivery of public services. Public Manag. Rev. **8**(4), 493501 (2006)

10. Bruns, A., Burgess, J.E.: The use of Twitter hashtags in the formation of ad hoc publics. In: 6th European Consortium for Political Research General Conference, 25–27 August 2011. University of Iceland, Reykjavik (2011)
11. CBC News: Alberta floods: a look back at the first 48 hours (2013). [Videorecording]. http://www.cbc.ca/news/canada/calgary/alberta-floods-a-look-back-at-the-first-48-hours-1.1871826 [accessed 6 March 2014]
12. Chesbrough, H.W.: Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business Press, Boston, MA, US (2003)
13. Cohn, J.P.: Citizen science: Can volunteers do real research? BioScience **58**(3), 192–197 (2008)
14. Cole, J., Walters, M., Lynch, M.: Part of the solution, not the problem: the crowd's role in emergency response. Contemp. Soc. Sci. **6**(3), 361375 (2011)
15. Crowley, J.: Connecting Grassroots and Government for Disaster Response. Commons Lab, Wilson Center (2013). http://www.wilsoncenter.org/sites/default/files/crowleyupdated2.pdf [Accessed 17 February 2014]
16. Dean, J.: Communicative capitalism: circulation and the foreclosure of plitics. Cult. Polit. **1**(1), 5174 (2005)
17. EJN: How Tragedy Strikes When Journalism and Social Media Lack Ethics and Humanity. Ethical Journalism Network (2013). http://ethicaljournalismnetwork.org/en/2013/how-tragedy-strikes-when-journalism-and-social-media-lack-ethics-and-humanity [Accessed 7 April 2014]
18. eScience: Earth Faces a Century of Disasters, Report Warns (2012). http://esciencenews.com/sources/the.guardian.science/2012/04/26/earth.faces.a.century.disasters.report.warns [Accessed 15 August 2013]
19. FEMA: Whole Community (2012) | FEMA.gov. http://www.fema.gov/whole-community [accessed 24.3.2013]
20. Government of Alberta Flood Recovery Task Force: Southern Alberta 2013 Floods: The Provincial Recovery Framework (2013). http://alberta.ca/albertacode/images/Flood-Recovery-Framework.pdf [accessed 6 March 2014]
21. Gupta, A., Lamba, H., Kumaraguru, P.: $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing Fake Content on Twitter. IEEE APWG eCrime Researchers Summit (eCRS), 17 Sep–18 Sep 2013, San Francisco, USA (2013)
22. Huffington Post: Rob Ford's Toronto Flood Response Criticized On Twitter. /textitHuffington Post 7th September (2013). http://www.huffingtonpost.ca/2013/07/09/toronto-flood-rob-ford-twitter_n_3568245.html [Accessed 9th April 2014]
23. Imran, M., Castillo, C., Lucas, J., Meier, P., Vieweg, S.: AIDR: Artificial Intelligence for Disaster Response. In: 23rd International Conference on the World Wide Web WWW14 Companion, April 7–11, 2014, Seoul, Korea. ACM, New York, 159–162 (2014)
24. International Committee of the Red Cross: Professional standards for Protection Work, 1115 (2013). http://www.icrc.org/eng/assets/files/other/icrc-002-0999.pdf [accessed 24.2.2014]
25. Irwin, A.: Citizen Science. A Study of People, Expertise and Sustainable Development. Routledge, London, New York (1995)
26. Klein, N.: The Shock Doctrine: The Rise of Disaster Capitalism. Penguin, London (2008)
27. Knight, K.: Facing the future. Findings from the review of efficiencies and operations in fire and rescue authorities in England. Her Majestys Stationery Office (2013). https://www.gov.uk/government/publications/facing-the-future [accessed 3rd March 2014]
28. Kotz, D.: Injury Toll from Marathon Bombs Reduced to 264. The Boston Globe, April 24 (2013). http://www.bostonglobe.com/lifestyle/health-wellness/2013/04/23/number-injured-marathon-bombing-revised-downward/NRpaz5mmvGquP7KMA6XsIK/story.html [accessed February 22, 2014]
29. Lanier, J.: On 'Digital Maoism': The Hazards of the New Online Collectivism. Edge Third Culture (2006). http://edge.org/3rd_culture/lanier06/lanier06_index.html [accessed 20 June 2010].
30. Leadbeater, C.: We-Think. Mass Innovation, not Mass Production. Profile Books, London (2008)

31. Lévy, P.: Collective Intelligence. Mankind's Emerging World in Cyberspace. Translated by R. Bononno. Perseus Books, Cambridge (1997)
32. Liu, S.B.: Trends in distributed curatorial technology to manage data deluge in a networked world. Upgrade **11**(4), 1824 (2010)
33. Lüge, T.: Social Media und Crowdsourcing in Katastropheneinsätzen—internationale Perspektiven. Heidelberg: Fachtagung: Web 2.0 und Social Media in Katastrophenschutz und Hochwassermanagement (2013). http://kats20.leiner-wolff.de/vortraege-3/
34. McGonigal, J.: The Puppetmaster Problem: Design for real world, mission based gaming. In: Harrigan, P., Wardrip-Fruin, N. (eds.) Second Person, pp. 251–264. MIT Press, Cambridge (2006)
35. Meier, P.: Results of the Crowdsourced Search for Malaysia Flight 370 (Updated). Irevolution (2014). http://irevolution.net/2014/03/15/results-of-the-crowdsourced-flight-370-search/ [Accessed 5the April 2014]
36. Mendona, D., Jefferson, T., Harrald, J.: Emergent interoperability: collaborative adhocracies and mix and match technologies in emergency management. Comm. ACM **50**(3), 45–49 (2007)
37. Mendoza, M., Poblete, B., Castillo, C.: Twitter under crisis: can we trust what we RT? In: Proc. of 1st Workshop on Social Media Analytics (SOMA 2010), pp. 71–79. ACM, New York (2010)
38. Michael, M.: Publics performing publics: of PiGs, PiPs and politics. Public Understand. Sci. **18**(5), 617–631 (2009)
39. Mildner, S.: Brgerbeteiligung beim Hochwasserkampf—Chancen und Risiken einer kollaborativen Internetplatform zur Koordination der Gefahrenabwehr (2013). Fachtagung, Heidelberg: Web 2.0 und Social Media in Katastrophenschutz und Hochwassermanagement. Retrieved from http://kats20.leiner-wolff.de/vortraege-3/
40. Morrow, N., Mock, A., Papendieck, A., Kocmich, N.: Independent Evaluation of the Ushahidi Haiti Project, Development Information systems International (2011). (http://www.alnap.org/pool/files/1282.pdf). [accessed 20 February 2014]
41. New York State: Reimagining New York for a New Reality. 2014–15 New York State Executive Budget, New York (2014). http://publications.budget.ny.gov/eBudget1415/fy1415littlebook/ReimaginingNewYork.pdf [accessed 3rd March 2014]
42. OCHA: Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies (2013). http://www.unocha.org/top-stories/all-stories/disaster-relief-20-future-information-sharing-humanitarian-emergencies [Accessed 10 April 2014]
43. Oh, O., Agrawal, M., Rao, H.R.: Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. Inform. Syst. Front. **13**(1), 111 (2010)
44. Oxfam: Haiti earthquake: 4 years later | Oxfam International. Retrieved March 02 (2014). From http://www.oxfam.org/en/haitiquake [accessed 1st March 2014]
45. Palen, L., Vieweg, S., Sutton, J., Liu, S.B.: Crisis informatics: studying crisis in a networked world. Soc. Sci. Comput. Rev. **27**(4), 467480 (2009)
46. Perng, S.-Y., Büscher, M., Wood, L., Halvorsrud, R., Stiso, M., Ramirez, L., Al-Akkad, A.: Peripheral response: microblogging during the 22/7/2011 Norway attacks. Int. J. Inform. Syst. Crisis Response Manag. **5**(1), 41–57 (2013)
47. Raymond, N., Howarth, C., Hutson, J.: Crisis Mapping Needs an Ethical Compass. Global Brief (2014). http://globalbrief.ca/blog/2012/02/06/crisis-mapping-needs-an-ethical-compass/ [accessed 22.2.2–14]
48. Reuter, C., Heger, O., Pipek, V.: Combining Real and Virtual Volunteers through Social Media. In: Comes, T., Fiedrich, F., Fortier, S., Geldermann, J., Mller, T. (eds.) Proceedings of the Conference on Information Systems for Crisis Response and Management, pp. 780–790. Baden-Baden, Germany (2013)
49. Rheingold, H.: Smart Mobs: The Next Social Revolution. Perseus, Cambridge (2003)
50. Schnebele, E., Cervone, G., Kumar, S., Waters, N.: Real time estimation of the Calgary floods using limited remote sensing data. Water **6**, 381–398 (2014). Doi:10.3390/w6020381
51. Shanley, L., Burns, R., Bastian, Z., Robson, E.: Tweeting up a Storm. The Promise and Perils of Crisis Mapping (2013). http://www.wilsoncenter.org/sites/default/files/October_Highlight_865-879.pdf

52. Sheller, M.: The islanding effect: post-disaster mobility systems and humanitarian logistics in Haiti. Cult. Geogr. **20**(2), 185–204 (2013)
53. Shirky, C.:"Is Social Software Bad for the Dean Campaign?" Many-2-Many, posted on January 26 (2004). http://www.corante.com/many/archives/2004/01/26/is_social_software_bad_for_the_dean_campaign.php [Accessed 12 April 2014]
54. St. Denis, L.A., Hughes, A.L., Palen, L.: Trial by Fire: The Deployment of Trusted Digital Volunteers in the 2011 Shadow Lake Fire. In: Proceedings of the 9th International ISCRAM Conference, p. 110, April 2012, Vancouver, Canada (2012)
55. Starbird, K., Palen, L.: "Voluntweeters": Self-Organizing by Digital Volunteers in Times of Crisis. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, Vancouver, Canada (2011)
56. Starbird, K., Palen, L.: Working and sustaining the virtual "Disaster Desk." In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work, CSCW 13, pp. 491–502 (2013)
57. Starbird, K., Maddock, J., Orand, M., Achterman, P.: Rumors, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing. In: iConference 2014 (2014)
58. Steffen, W., Grinevald, J., Crutzen, P., McNeill, J.: The Anthropocene: conceptual and historical perspectives. Phil. Trans. Roy. Soc. A Math. Phys. Eng. Sci. **369**(1938), 842–867 (2011)
59. Surowiecki, J.: The Wisdom of Crowds. Random House, New York (2005)
60. Tapia, A.H., LaLone, N., Hyun-Woo, K.: Run amok: group crowd participation in identifying the bomb and bomber from the boston marathon bombing. In: Hiltz, S.R., Pfaff, M.S., Plotnick, L., Shih, P. (eds.) Proceedings of the 11th International ISCRAM Conference, University Park, Pennsylvania, USA (2014)
61. Tucker, E.: How Toronto Mayor Ford handled the crisis: social media reacts (2013) http://globalnews.ca/news/704726/how-torontos-mayor-ford-flood-response-compared-to-calgarys-mayor-nenshi/ [Accessed 2nd April 2014]
62. Turner, K.: Mayor Nenshi's Twitter popularity swells during Alberta flood (2013). http://metronews.ca/news/calgary/736562/mayor-nenshis-twitter-popularity-swells-during-alberta-flood/ [accessed 6 March 2013]
63. Vieweg, S., Palen, L., Liu, S., Hughes, A., Sutton, J.: Collective Intelligence in Disaster: An examination of the phenomenon in the aftermath of the 2007 Virginia Tech shooting. In: Proceedings of the 5th International ISCRAM Conference, Washington DC, May (2008)
64. Wood, J.: Province boosts cost of Alberta floods to $6 billion (24 September 2013). http://www.calgaryherald.com/news/Province+boosts+cost+Alberta+floods+billion/8952392/story.html [accessed 6 March 2014]
65. Yablonksi, C.: The impact of social media on the Calgary Flood (2013). http://www.inboundinteractive.ca/the-impact-of-social-media-on-the-calgary-flood/ [accessed 6 March 2014]
66. YYCHelps: Internet: www.yychelps.ca (2014). [accessed 6 March 2014]
67. Zipf, A.: Nutzergenerierte Geodaten im Crisis Mapping. Stand der Forschung & Perspectiven. Fachtagung, Heidelberg (2013). Web 2.0 und Social Media in Katastrophenschutz und Hochwassermanagement. Retrieved from http://kats20.leiner-wolff.de/vortraege-3/

The Lean Research: How to Design and Execute Social Collective Intelligence Research and Innovation Projects

Daniele Miorandi, Iacopo Carreras, and Imrich Chlamtac

1 Introduction

Research can be, broadly speaking, defined as a "systematized approach to gain new knowledge" [11]. Various research methodologies have been proposed for usage in different fields of investigations, based on a large number of approaches [13, 15]. As a matter of fact, various scientific disciplines developed their own de facto standard research methodology, which build on a number of common building blocks, most of which can be traced back to Greek philosophers and systematized by Galileo in the seventeenth century [8].

In this chapter, we focus on *research and innovation (R&I) projects*, defined as systematic approaches to gain new knowledge in a well-scoped area and in a goal-oriented fashion. The question we ask is the following: "Is there a method for architecting and running R&I projects on Social Collective Intelligence?". This is motivated by the fact that Social Collective Intelligence (SCI) [23] refers to a class of socio-technical systems that combine, in a coordinated way, the strengths of humans and groups in terms of competences, knowledge and problem solving capabilities with the communication, computing and storage capabilities of advanced information and communication technologies. As such, projects in SCI present a set of distinctive features, compared to other scientific disciplines. In particular, we refer here to the 'hybrid' nature of SCI systems, which include elements of human and machine-based computation which cannot be considered in isolation. This is strongly linked to the inherently multidisciplinary character of any initiative in the field. Besides this, SCI is about innovating in complex socio-technical systems, which are based on a set of interwoven feedback loops

D. Miorandi (✉) • I. Carreras • I. Chlamtac
CREATE-NET, v. alla Cascata 56/D, 38123—Povo, Trento, Italy
e-mail: daniele.miorandi@create-net.org; iacopo.carreras@create-net.org; imrich.chlamtac@create-net.org

D. Miorandi et al. (eds.), *Social Collective Intelligence*, Computational Social Sciences, DOI 10.1007/978-3-319-08681-1_13

involving individuals, collectives and technologies. And any type of innovation or change in such context can effectively be understood as a ‘perturbation’ (to use a physics terminology) of the current state of the system, whose effect cannot be fully predicted a priori and needs to be tested in vivo, through an experimentally-driven approach [29].

Based on the outcomes of a number of consultation workshops and events held within the scope of the EU-funded Social-IST project,[1] and inspired by a parallel with the ‘lean startup’ approach [21], we introduce in this chapter a set of methodological guidelines for maximising success chances and potential impacts of R&I projects in the field of Social Collective Intelligence. Such guidelines aim at representing the seeds of a blueprint for planning and implementing research and innovation initiatives in the field of SCI.

The remainder of this chapter is organised as follows. In Sect. 2 the key elements of the ‘lean research’ methodology are introduced. Such elements are then discussed in details in Sects. 3–8. Finally, Sect. 9 concludes the chapter highlighting the key challenges to be faced.

2 Towards a “Lean Research” Approach

Taking inspiration from the startup movement, and in particular from the concept of the “lean startup” [21], we do believe that a research programme on SCI should adopt an innovative “lean research” methodology in order to achieve truly transformational impacts.

The lean start-up approach is a set of methodological steps, first introduced by E. Ries [21], that were observed to represent a recurrent pattern in many successful entrepreneurial ventures. Quoting from [1],

> It’s a methodology called the “lean start-up,” and it favors experimentation over elaborate planning, customer feedback over intuition and iterative design over traditional “big design up front” development.

Conceptually, the lean startup approach builds upon an iterative three-steps process, represented in Fig. 1. Starting from ideas, companies following such an approach go through a (short) building phase, in which some prototype of the product/service is created and delivered to end users. The outcome of this initial phase is called a Minimum viable Product (MVP), as it refers to the minimum version of the product that is able to deliver a “value” for the customers. In this phase, the company measures the acceptance of the prototype/MVP by customers, collecting data on user feedback and utilization. Data is then analysed (the ‘validated learning’ phase) in order to understand how to tune/change the original design. The loop is continuously iterated upon, in order to keep on improving the fit between the product and the market.

[1] http://social-ist.eu/.

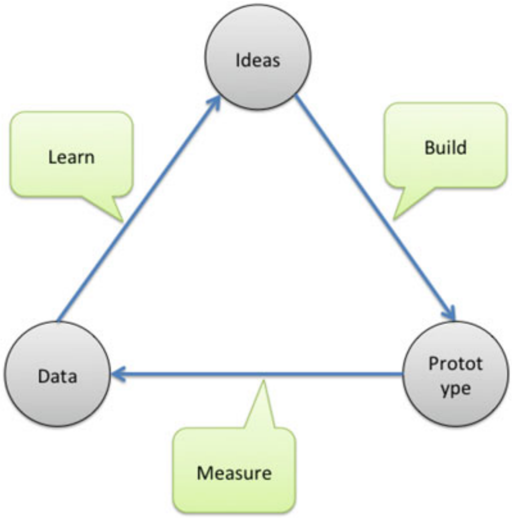


Fig. 1 Conceptual representation of the lean startup approach: key phases

Of course, research and innovation projects in SCI are very different from startup companies. Yet, they share one key commonality, that is, the fact that they operate in a setting characterised by very high uncertainty and for which an iterative feedback loop is fundamental in order to ensure a proper progress. If for startups this relates mostly to understand the market and the customers needs [21], for an R&I project in SCI this refers to how individuals and collectives would react to a new technological infrastructure or service introduced. That is why we do believe that some of the key concepts underpinning the lean startup approach could well be applied (after a proper 'translation' process) to SCI research and innovation projects.

We now introduce a set of guidelines that constitute a 'lean research' approach for SCI research projects; details are provided in the following sections.

- **Experiments from day zero.** SCI systems are naturally people-centric. This requires adopting and extending user-centric/co-design approaches [17, 25] well beyond what currently done in most ICT projects. Empirical activities with individuals and collectives should represent the starting point around which new technological enablers will be introduced iteratively and incrementally. This experimental set-up should combine lab activities in controlled environments as well as tests in real-world settings where SCI is already emerging, for example in e-labour markets, digital science, e-health and care, or in e-mediated creative and cultural industries [2]. Using lab and live settings in parallel will allow the development of new syntheses of scientific and social methodologies that will be based on empirical evidence through the use of detailed observations and big data.
- **Fast incremental cycles.** SCI projects should be based upon an agile, incremental approach, whereby technologies/solutions/systems developed in the lab are

concurrently exposed to test and evaluation 'in the wild' (i.e., with real users and applications in selected, rapidly evolving real-world contexts), and where carefully monitored exposure in the real world is calibrated to the experimental setting in order to adjust/adapt/evolve the solution iteratively in response to experience. This process will also consider the extent to which the technologies transform the social context in an attempt to evaluate potential risk and to scope the potential reach and effects of the deployment of SCI technologies in different settings [19].

- **Stakeholders engagement.** To achieve a real impact SCI projects must directly involve stakeholders in empirical activities. This requires substantial efforts and the development of skills and competences, which are out of the background of most scientists. This is particularly relevant for young researchers, who should be educated to work with stakeholders as part of their working methodology. This may involve significant transformations in the way researchers work. For example, it may be necessary to embed a researcher with SCI knowledge and development skills into a live context to "co-realize" solutions for stakeholders. This will require careful monitoring and control to understand the effects and transform them into transferable knowledge.
- **Sustainability.** SCI has sustainability "built in"—any true SCI system is "self-propelling" using the effort and resources of participants to power the enterprise via individual and collective motivation and incentives. If this is not the case, the SCI will never "power up" when it is switched on (of course achieving critical mass or operating levels of activity is an interesting issue in itself for some SCI systems since they may need special measures or incentives to achieve eventual sustainability). Thus SCI projects cannot afford the luxury of not considering the development of structures and operating models that support sustainable business models (understood in the widest sense). The design of SCI systems and applications should always include the development of appropriate measures for becoming self-sustainable. Business models are not restricted to the monetary aspects, but will provide a clear description of the different value classes utilized by the SCI together with the value created by the system/application and of the measures needed for its deployment (including the design of appropriate incentives for individuals and collectives).
- **Public sector engagement.** The public sector is the locus for many systems that could benefit from SCI approaches. The public sector is also a key regulator for such systems and is often responsible for the development and deployment of governance regimes. These components are critical to SCI and without the active engagement of the public sector it will be impossible to fully understand the development, operation, evolution and oversight of SCI.
- **Sharing and open access.** Open access policies should be adopted to ensure re-use of the knowledge generated within the projects. This will bring along a twofold advantage. On the one hand, it will reduce the risk of duplication of efforts among different projects, which may well encounter similar problems in their execution. On the other hand, it will foster the take-up by third parties of the knowledge developed inside the projects, maximizing impacts beyond the

boundary of consortia. Openness is a key feature of the development of SCI because part of the work will result in the creation of an infrastructure built out of existing and evolving structures combined with new tools, components and services developed by the SCI programme.

3 Experiments from Day Zero

Typically most research projects in Information and Communication Technology push experiments towards late in the work plan. The idea is to first design the technology, then develop and integrate it and last test it and validate it, when possible with potential (early) adopters. While many approaches exist for involving the users in the early stages of the design process [4, 6, 24], typically empirical activities are still postponed until a prototypical/stable version of the new technology developed is available.

We challenge this approach and claim that projects on SCI should start experiments from the very beginning. The idea is that experiments are not a mere tool for validating the ability of a new technology to respect some requirements or achieve certain performance. Rather, experiments should be considered a fundamental tool for both understanding the problem and devising a solution. As such, we believe that SCI projects should make use of existing socio-technical facilities for running experiments with real users 'in the wild', without requiring new infrastructure to be developed. In this context, any project working in this field, should carefully plan the system to be designed, as well as how to experiment it with potential real users (or early adopters) from day zero. In particular, this latter part should take into account the many aspects involved in the identification, as well as the engagement of a users community. This includes the channel to be used for interacting with them (e.g., social media, mobile, web, etc.), the most appropriate engagement strategy (e.g., incentives) as well as the required critical mass to validate research results.[2]

4 Fast Incremental Cycles

In line with what described in Sect. 3 in terms of experimental/empirical activities, the work plan of SCI projects should be built around fast and incremental cycles, where new elements get integrated and tested on a fast pace, allowing the projects to undergo various trial-and-error phases. Drawing inspiration from agile methodologies nowadays widely adopted in software development processes [7] we call for a similar approach to R&I projects as well. This implies a fundamental shift also

[2] An interesting and open research question is to understand the minimum sample size for validating a given research hypothesis in SCI.

in the way research projects are planned and managed, requiring a more flexible adaptation of activities and tasks schedules, in order to cope with the lessons learned during the previous cycles.

Each cycle should include research, design and experimental components. Each cycle should allow the project team to *learn* sufficiently about the problem to enhance the solution devised. In this respect, the "fear of failure" so widely found in research circles should be replaced by the "fail often, fail quickly" motto adopted in the high tech startup environment. Failures should be actually seen as a key constituency of the creative innovation process, in line with Schumpeter's theories [27].

Initial examples of experimentally driven SCI projects are taking advantage of the Mechanical Turk[3] facilities for delivering tasks to real world users [12, 16, 20]. In this case, an experiment consists of a task to be performed by one of the many users registered to the system. To the completion of each task, corresponds a monetary reward that is specified by the experimenter. While this represents an initial facility for performing SCI experiments, it is not the most appropriate for progressing SCI research, as it not able to account for the many contextual aspects who can play a significant role in the dynamics of the system [18]. Furthermore, it is complicated to target specific communities of users, over which to perform iterative build/measure/learn cycles.

5 Stakeholders Involvement

In many sectors where ICT R&D projects are carried out, the involvement of stakeholders is known to be a non-trivial issue. When dealing, e.g., with eHealth solutions, projects aiming at achieving high impact should involve health agencies and providers, policy and decision makers, patient associations and, in many cases, even professional associations of doctors, nurses etc. Similarly, projects on "smart cities" should involve municipalities, citizens' associations, mobility providers and agencies (when transportation is considered) etc.

In many cases stakeholders will end up being end-users of the SCI project's results, since typically such projects are rooted in a given application domain. They are then in the best position to validate many of the assumptions on the systems design. Furthermore, they often have the necessary knowledge to interpret the results of the experimental activities performed through collectives of users. Taking for example projects on smart urban mobility, transportation authorities typically have knowledge of mobility demands (origin/destination matrix), mobility offers (e.g., public transportation), events which may alter mobility patterns (e.g., concerts, public transportation strikes etc.) as well as regulatory constraints and policy directives. Their involvement is then fundamental for the design of any SCI

[3] https://www.mturk.com/mturk/welcome.

project in the field of smart mobility, and to maximally exploit the intelligence emerging from a collectives of citizens.

Typically, scientists tend to see this as a burden, as it implies dealing with a number of issues which do not contribute to strengthening the quality of the research outcomes of the project. Furthermore, some of such stakeholders may have very cumbersome and lengthy internal decision processes, which may conflict with the limited time-life of cooperative R&I projects.

Yet, stakeholders involvement in design choices and empirical activities (the two being tightly linked as discussed above) represents an instrumental element in achieving long-term success for a SCI project. This requires both substantial investment (budget-wise, but also time-wise) in ensuring a proper involvement of stakeholders as well as a change in researchers' mindset. The latter aspect call for an element of novelty in the design of educational curricula (at least at the PhD level), enabling the future generation of SCI researchers to be able to work consistently and effectively with stakeholders.

6 Sustainability

From a research policy perspective, one of the main pitfalls of public spending in research and development projects relate to the low impact on economy and businesses. One of the reasons relates to the fact that too often R&I projects focus on developing solutions but do not take in proper account sustainability aspects. In many cases exploitation plans and business models are worked out a posteriori, once the technology/innovation has already been developed and it is too late to make any change to its fundamental assumptions. While this is rather bad per se, it becomes even worse in SCI. Social collective intelligence systems are, indeed, self-propelling by design. SCI systems should embed in their inner fabric incentives able to motivate the participation of users, whose contribution extends and augments the 'value' of the system, attracting more users and so on so forth. This 'sustainability' aspect is a key one, which should be considered from the very beginning of the system inception. The term 'business model' for SCI systems should therefore be understood at large, including not just financial aspects, but focusing on the ability of the system to create value for all relevant actors. A fundamental aspect in this sense is represented by incentives design for SCI system. While this is still an open research field, early works on incentives and rewarding schemes for social computing provide certainly significant insight [26].

7 Public Sector Involvement

The public sector plays an instrumental twofold role in the inception of Social Collective Intelligence systems. First of all, as SCI systems are deeply embedded in society the public sector should actively participate to their governance [10]. Second, the public sector could play a very important role in fostering the take-up and adoption of SCI systems in their early-stage, in particular if such systems have the potential to tackle some relevant societal challenges. In this sense, pre-commercial procurement (PCP) [5, 22] appears a very appealing tool for the public sector to foster the adoption of SCI approaches in a setting where it can play a key role in the governance of the system and directly benefit from it.

8 Sharing and Open Access

Sharing and open access are becoming key pillars of modern science, as they are recognized as being able to fully unleash the potential of R&I activities beyond the boundaries of organizations and projects. This trend, which aligns with the concept of 'open innovation' being popularized by Chesbrough and others [3], should be considered a requirement for R&I endeavours in SCI. In particular we do see a stringent need to promote sharing and reuse of:

- **Software tools and infrastructures.** The sharing of software tools and infrastructures is fundamental to lower the entry barrier for scientists and innovators interested in experimenting with SCI concepts and systems. Examples of such software frameworks include platforms to easily recruit and crowd-source end-users or for automating tasking on top of Mechanical Turk [14, 16] .
- **Open Data.** Open data [28] is increasingly becoming a reference paradigm to stimulate innovation by making freely available large data sets to anyone interested in using them. There are many examples of open data, including everything from demography and population data over geographic, economic, education and health data to transport, travel and mobility data. The idea is that there is great value in this data, but this value can only be unlocked by making it available to communities of users who can make the best use of it. The availability of Open Data will allow scientists and organizations to gauge better insight into how novel ICT can benefit and change social processes and societal structures. While policy-makers are already taking steps in this direction [9], further work is needed to harmonize and extend such approaches. In particular, nowadays Open Data is mostly providing access to static information (e.g., datasets), while when it comes to SCI projects the availability of real-time information, accessible through open APIs is what is really needed.
- **Incentive Schemes.** Incentive strategies are very specific to a given application domain and users community. The availability of empirical results on the effectiveness of different incentive strategies for engaging community of users is of paramount importance for accelerating the design and implementation of SCI R&I. Indeed, the validation of incentive schemes typically requires long and expensive trials over large community of users. Having access to a repository

of experimental results, together with a description of the setting and of the incentives use, would greatly accelerate the iterative cycle described in Sect. 4.

- **Communities of users.** One of the key issues in empirical SCI activities is the access to communities of users. Creating and maintaining a community of users willing to experiment with SCI prototypes is a complex task. It would be desirable to have means for sharing, across R&I projects, groups of users, in much the same way Living Labs [14] are starting to offer access following an 'as a service' model.

9 Conclusions

In this chapter we have identified a set of guidelines which can help R&I projects on Social Collective Intelligence to maximise their impacts. The guidelines bear a striking resemblance with the lean startup approach advocated by E. Ries in [21], hence the use of the term 'lean research' approach.

These guidelines can be useful for both scientists as well as research policy makers, in order to identify guidelines for the construction of high-impact SCI projects. Such guidelines should become part of the design of any project in this field, and will allow the SCI research community to rapidly grow, sharing results and best practices in SCI projects. Challenges include the need for a major shift in the researchers' mindset and in educational curricula, in order to enable the arising of a novel generation of SCI researchers and practitioners, able to embrace the distinctive features of Social Collective Intelligence systems and to leverage on them to effectively tackle major societal challenges.

Acknowledgements The work reported here was partially supported by the European Community's Seventh Framework Program (FP7/2007–2013) under grant agreement n. 317681 Social-IST: Social Collective Intelligence (http://social-ist.eu/).

References

1. Blank, S.: Why the lean start-up changes everything. Harv. Bus. Rev. **91**(5), 63–72 (2013)
2. Carreras, I., Anderson, S., Robertson, D., Miorandi, D.: Roadmap for FET initiatives in social collective intelligence (2013). URL http://social-ist.eu/files/2013/12/D3.1.pdf. Social-IST FP7 Project Deliverable D3.1
3. Chesbrough, H.W.: Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business Press, Boston (2003)
4. DeBellis, M., Haapala, C.: User-centric software engineering. IEEE Expert **10**(1), 34–41 (1995)
5. Edler, J., Georghiou, L.: Public procurement and innovation-resurrecting the demand side. Res. Pol. **36**(7), 949–963 (2007)

6. Eriksson, M., Niitamo, V.P., Kulkki, S.: State-of-the-Art in Utilizing Living Labs Approach to User-Centric Ict Innovation-a European Approach. Lulea: Center for Distance-spanning Technology. Lulea University of Technology Sweden, Lulea. Online under: http://www.cdt.ltu.se/main.php/SOA_LivingLabs.pdf (2005)
7. Fowler, M., Highsmith, J.: The agile manifesto. Software Dev. **9**(8), 28–35 (2001)
8. Galilei, G.: Discorsi e dimostrazioni matematiche, intorno à due nuove scienze (discourses and mathematical demonstrations relating to two new sciences). Leiden (1638)
9. Guidelines on open access to scientific publications and research data in Horizon 2020 (2013). URL http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020- hi-oa-pilot-guide_en.pdf
10. Hartswood, M., Grimpe, B., Jirotka, M., Anderson, S.: Towards the ethical governance of smart society. In: Miorandi, D., Maltese, V., Rovatsos, M., Nijholt, A., Stewart, J. (eds.) Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society. Springer, New York (2014)
11. Hufford, M.E.: The romance of research. School Sci. Math. **35**(3), 273–284 (1935)
12. Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with Mechanical Turk. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 453–456. ACM, New York (2008)
13. Kothari, C.: Research Methodology: Methods and Techniques. New Age International, New Delhi (2004)
14. Kulkarni, A.P., Can, M., Hartmann, B.: Turkomatic: automatic recursive task and workflow design for mechanical turk. In: CHI'11 Extended Abstracts on Human Factors in Computing Systems, pp. 2053–2058. ACM, New York (2011)
15. Kumar, S., Phrommathed, P.: Research Methodology. Springer, New York (2005)
16. Little, G., Chilton, L.B., Goldman, M., Miller, R.C.: Turkit: tools for iterative tasks on mechanical turk. In: Proceedings of the ACM SIGKDD Workshop on Human Computation, pp. 29–30. ACM, New York (2009)
17. Maguire, M.: Methods to support human-centred design. Int. J. Hum. Comput. Stud. **55**(4), 587–634 (2001)
18. Mason, W., Suri, S.: Conducting behavioral research on Amazon's mechanical turk. Behav. Res. Meth. **44**(1), 1–23 (2012)
19. Owen, R., Macnaghten, P., Stilgoe, J.: Responsible research and innovation: From science in society to science for society, with society. Sci. Publ. Pol. **39**(6), 751–760 (2012)
20. Paolacci, G., Chandler, J., Ipeirotis, P.G.: Running experiments on Amazon Mechanical Turk. Judgment Decis. Making **5**(5), 411–419 (2010)
21. Ries, E.: The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. Random House LLC, New York, NY (US) (2011)
22. Rigby, J.: Review of pre-commercial procurement approaches and effects on innovation. Tech. Rep. Working Paper 13/14, NESTA (2013). URL http://www.nesta.org.uk/sites/default/files/review_of_pre-commercial_procurement_approaches_and_effects_on_innovation_revised_ajr-14-11-2013_final.pdf
23. Robertson, D., Anderson, S., Carreras, I., Miorandi, D.: White paper on research challenges in social collective intelligence (2013). URL http://social-ist.eu/files/2013/12/D2.1.pdf. Social-IST FP7 Project Deliverable D2.1
24. Roy, R., Goatman, M., Khangura, K.: User-centric design and kansei engineering. CIRP J. Manuf. Sci. Tech. **1**(3), 172–178 (2009)
25. Sanders, E.B.N., Stappers, P.J.: Co-creation and the new landscapes of design. Co-design **4**(1), 5–18 (2008)
26. Scekic, O., Truong, H.L., Dustdar, S.: Incentives and rewarding in social computing. Comm. ACM **56**(6), 72–82 (2013)
27. Schumpeter, J.A.: Business Cycles, vol. 1. Cambridge Univ Press, Cambridge (1939)
28. Uhlir, P., Schröder, P.: Open data for global science. Data Sci. J. **6**(0) (2007)
29. Vespignani, A.: Predicting the behavior of techno-social systems. Science **325**(5939), 425 (2009)