Increasing Transparency and Privacy for Online Social Network Users – USEMP Value Model, Scoring Framework and Legal

A. Popescu[1], M. Hildebrandt[2], J. Breuer[3], L. Claeys[3], S. Papadopoulos[4], G. Petkos[4], T. Michalareas[6(✉)], D. Lund[5], R. Heyman[3], S. van der Graaf[3], E. Gadeski[1], H. Le Borgne[1], K. deVries[2], T. Kastrinogiannis[6], A. Kousaridas[6], and A. Padyab[7]

[1] CEA, Paris, France
[2] Radboud University Nijmegen, Nijmegen, The Netherlands
[3] iMinds-SMIT, Etterbeek, Belgium
[4] CERTH-ITI, Thessaloniki, Greece
[5] HW Communication Ltd, Lancaster, UK
[6] VELTI SA, Marousi, Greece
tmichalareas@velti.com
[7] Luleå Tekniska Universitet, Luleå, Sweden

Abstract. In this paper we present research results from the multi-disciplinary EU research project USEMP (USEMP is a project funded from EU research framework, additional information about project scope and deliverables are available at project's public website at: http://www.usemp-project.eu/). In particular, we look at the legal aspects of personal data licensing and profile transparency, the development of a personal data value model in Online Social Networks (OSNs) and the development of disclosure scoring and personal data value frameworks. In the first part of the paper we show how personal data usage licensing and profile transparency for OSN activities provides for Data Protection by Design (DPbD). We also present an overview of the existing personal data monetization ecosystem in OSNs and its possible evolutions for increasing privacy and transparency for consumers about their OSN presence. In the last part of the paper, we describe the USEMP scoring framework for personal information disclosure and data value that can assist users to better perceive how their privacy is affected by their OSN presence and what the value of their OSN activities is.

1 Introduction

USEMP is a multi-disciplinary research project, integrating the perspectives of lawyers, engineers, computer scientists, marketing experts and social scientists, aiming at developing a framework that will empower Online Social Network (OSN) users by enhancing their control over the data they distribute or interact with, with an eye on what can be inferred from the personal data shared in OSNs. At the core of this objective lies the idea of reducing the existing asymmetries of processing and control between OSNs organizations and citizens.

B. Berendt et al. (Eds.): APF 2015, LNCS 9484, pp. 38–59, 2016.
DOI: 10.1007/978-3-319-31456-3_3

For that purpose, we will briefly indicate how transparency tools operate in the legal framework of Data Protection by Design (DPbD), notably for profile transparency. Additionally, we will present how an ecosystem - as envisaged in USEMP - can evolve in the future, identifying business opportunities and challenges that could arise when the user has the means to assert more meaningful control in a sustainable manner. In addition to the proposed considerations for a personal data value model in OSNs, the USEMP research team has also developed a scoring framework that can be used to collect and compute indicators for the information disclosure and value of shared personal data, which is also described in the last part of the paper.

2 The Legal Angle: Data Licensing and Profile Transparency

At the global level privacy and data protection is in turmoil, creating unprecedented uncertainty over the business ethics that drive the new economy. We believe that it is pivotal that the upcoming EU GDPR stabilizes the expectations within the internal market of the EU, making sure that citizens have a legitimate expectation of the consequences of sharing data. USEMP has developed two types of legal tools to support the data-driven ecosystem of OSNs, which underpins the previous– technical and economic– account.

2.1 Data Licensing Agreement (DLA)

EU data protection law requires that any processing of personal data is based on one of six legal grounds. Current business models are often based on consent and on the legitimate interest of the data controller (i.e. the party that determines the purpose of processing). These are not very reliable grounds: Consent can be withdrawn at any time and the legitimate interest of the data controller needs to be balanced against the fundamental rights and interests of the data subject. In both cases the legal relationship between user and service provider remains opaque and distant, usually determined by privacy policies and terms of service that are oriented to assumptions of US law. In the context of USEMP we have opted for a Data Licensing Agreement that requires the active participation of the user, by means of mutually obligatory agreement (quid quo pro) that determines what each party commits to deliver in the context of the economic value chain. The DLA thus licenses the use– and if applicable– the re-use of personal data for explicitly defined and specified purposes by specific data controllers under the conditions set forth in the DLA. The core obligation on the side of the user is to download and install the USEMP tools on her device and to license her data for the purpose of computing a set of scores related to disclosure dimensions that can help the user understand what can be inferred based on shared data. The core obligation on the side of the provider is to provide transparency tools that deliver user-friendly visualizations of the (perceived) sensitive information that can be inferred from the user's data points. The DLA contains explicit consent for the processing of (legally) sensitive data (for the specified purpose only) and explicit consent for downloading the USEMP tools (or any other tracking mechanism).

The next version of the DLA will be modular, enabling a more granular type of licensing, comparable with the various types of Creative Commons Licenses for copyright protection. This should enable users to specify e.g. for which purposes their data may and may not be used, and/or re-used, by which parties and for what time.

2.2 Profile Transparency Tools

The current and the upcoming EU legislative framework of data protection requires that any automated decisions that have a significant impact on the user must comply with a number of conditions. Crucially, under the upcoming framework, three transparency requirements must be met for decisions based predominantly on automated profiling: the existence of such profiling, the logic of processing and the envisaged consequences (Hildebrandt 2012). The upcoming framework will also require the implementation of Data Protection by Design mechanisms that ensure that data controllers by default comply with their legal obligations (see [5]). This means that once state of the art mechanisms have been developed, their employment will become mandatory. USEMP is developing tools to provide profile transparency. By inferring disclosure scores based on the disclosure dimensions, USEMP is capable of indicating what kind of profiles people match, thus also indicating how they may be targeted by third parties (advertising, insurance companies, employers etc.). We hope that once USEMP-type platforms emerge as viable playgrounds for the testing of inferred user profiles, OSN providers will be forced to collaborate with them to increase the trust of their users and to improve their reputation. This may also contribute to compliance with their legal obligation to provide profile transparency. In the next section of the paper we discuss the results of research with respect to the effect of transparency in the current advertising value chain and we briefly discuss end-consumers perception of privacy.

3 Increasing Transparency of Use for Personal Data in OSNs

The research regarding transparency undertaken in the USEMP project has been based on conducted expert interviews with nine (9) diverse actors from the advertising ecosystem to get first hand insights; interviews covered roles from ad.networks, marketing companies, organizations related to the practices of advertising and social networks, all of which have a business interest in the utilization of users' data (the interviews method and details are described in detail in [6], we have also added a summary in Table 1).

We start with a description of the advertising and marketing ecosystem in terms of its use of personal data as it is today. The problems that users face in the value network underpinning personal data are central to the USEMP project: non-transparent re-use of personal data, often through unaccountable and untrustworthy third parties. This causes information asymmetries and power imbalances, disfavouring the user. Our research has shown that many actors on the industry side also see shortcomings in the status-quo, and we note that their interests are not per se contrary to those of users.

Table 1. Overview of USEMP conducted interviews with advertising/marketing industry stakeholders

Interviewee	Role
Lien Brusselmans	L. Brusselmans Marketing Communication Manager at Engagor
Roland Siebelink	R. Siebelink is head of quality, productivity and best practice of Rocket Fuel
Theodoros Michalareas	T. Michalareas is head of product development in Velti, a provider of mobile marketing and advertising services
Joelle Frijters	J. Frijters is co-founder and CEO of ImproveDigital, a European provider of independent publisher monetisation technology
Chris Payne	C. Payne is Public Affairs Manager at World Federation of Advertisers. The federation is a global organization representing marketers and advertisers. (http://www.wfanet.org/en)
Niels Baarsma	N. Baarsma is co-founder and CTO of Yieldr, a demand side platform provider.
Kimon Zorbas & Ionel Naftanaila	K. Zorbas is Director at the Digital Business Consultancy Group (www.dbcg.eu). Before that, he was CEO and Vice President for IAB Europe. Dr. I. Naftanaila is EDAA's Programme Development Director and brings with him a wealth of industry knowledge and experience
Mario van Lommel	M. van Lommel is Technical Sales Engineer at Be-Mobile, a leading provider of traffic and mobility content and services for the automotive industry, mobile, media and government road operators
Joost Roelandts	J. Roelandts is COO at the social network Twoo

A concrete and objective value of personal data cannot be easily defined. All actors involved in the OSNs value model depend to some degree on some kind of personal data, but each of them on different types and for different ends and purposes. Thus the value of personal data from OSN users differs for each actor. This also relates to the fact that it makes no sense to reduce 'value' to either monetary value or ethical value, as both are simultaneously at stake [11]. We propose a mechanism that can increase transparency of personal data usage in the ecosystem [1]. This would not only contribute to a business ecosystem that is more respectful to the content creators such as social network users, while enabling innovation that complies with EU Data Protection law. It would also increase the value of (personal) data for each actor and thus for the ecosystem on the whole.

The low level of trust in the industry is a main barrier in reaching such a goal (see also [2]). Trust is not only from the OSN end-user towards the industry (rightfully so in many regards), but also between commercial actors in the value network. The reason for this is among others the non-transparency of data-related operations. Linked to non-transparency, the low quality of utilized data is a major issue, impeding the industry's functioning, reducing efficiency and thus also profits. For the user, non-transparency creates not only information asymmetries, but also diminishes the user experience (for example, irrelevant advertising and longer page loading times from the effect of third party tracking).

Additionally, the USEMP team has performed a number of focus group interviews and related analysis to understand what the social requirements are for privacy enhancing tools. This social requirement analysis has highlighted the need for a personal data management platform that currently the USEMP team is building (see [16] for details of requirements and method of analysis).

3.1 Personal Data Business Model Evolution and Related Legal Aspects

Firstly, we will assess how the ecosystem will evolve, if the market is left to operate alone, without appropriate regulation or mechanisms as the ones suggested by USEMP.

The low quality and reliability of data is a real problem for the industry. In order to avoid being responsible and accountable for personal data issues, actors often outsource the collection and pseudo-anonymization[1] of data to data brokers or similar companies. Subsequently OSN providers purchase end-user data they need for their operations from such third parties (advertising networks, data brokers or similar companies), in some cases pseudo-anonymized. This can be seen in a similar light as rights clearance in the field of intellectual property. After buying data in this fashion, the actor may reason that it is no longer personal data, or if it is, at least lay the blame on the third party. This is– obviously– incorrect, and partly caused by the fact that some of the big players believe they can afford to base their operations on US law, even when processing personal data of EU citizens.

As the interviews with industry experts have illustrated, the functioning of the market is strongly affected by major OSNs, mostly Google, Twitter, Facebook and their related competition (for example see [8]). These actors might be able to offer a remedy regarding data quality. They have full access to personal data collected through user profiles in their respective OSNs. In addition, they have access to additional behavioural data spanning over additional data points (like web sites that include OSN tracking code).

End-User Demand for Data Protection and the Effect to Smaller Companies. In such an environment end-users might increasingly demand stronger data protection and prefer those actors and services that they trust and offer tools that respect their data protection rights. If the market is left alone to deal with that, the already strong position of the prominent actors today will allow them to fortify their position. In contrast to smaller companies, these actors have the necessary means at their disposal to invest and develop new tools and mechanisms. Additionally data protection and privacy tools would be

[1] On the legal effect of pseudo-anonymization see: Art. 29 WP Opinion 05/2014 on Anonymisation Techniques, and the upcoming General Data Protection Regulation (GDPR) that mitigates some of the obligations of data controllers if they process personal data that have been pseudo-anonymized. The legal definition of pseudo-anonymization (art. 4. 2(a) of the upcoming GDPR reads: 'pseudonjymous data' means 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. A data controller is whoever determines the purpose of processing, i.e. the business model. The liability for compliance with EU Data Protection law rests solely with the data controller.

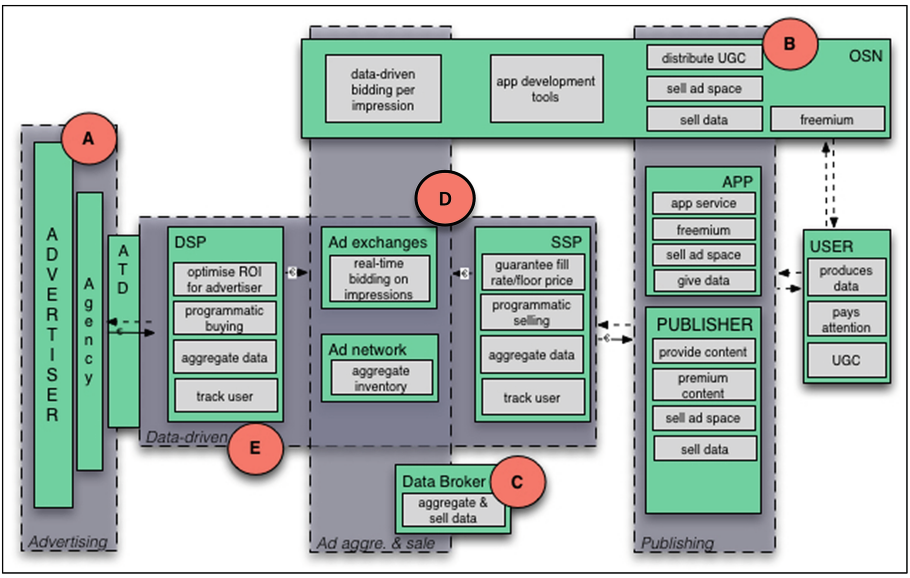

Fig. 1. The Value Network of Online Advertising and options to introduce transparency tools

controlled by those companies, which already misuse their access to user data today. This could be regarded in the light of feigned vendor relationship management, pretending that users have control (for examples see [3]). The advantage for OSN organizations increases even further because they can leverage user lock-in to ask for more personal data or further reaching users without losing end-users to competitors.

Our research and interviews with ecosystem actors indicates that the currently weak enforcement of data protection and privacy leads to adverse effects; it harms the European industry, weakening innovation, hurting SMEs and start-ups, while favouring big companies, especially from non-European countries that feel they can ignore the EU Data Protection legislation[2]. Moreover evidence of this harm is now documented in industry reports that show the increased use of ad-blockers plugins from end-users and the corresponding revenue losses (see for example [15]).

Need for Personal Data Management Tools that Provide Awareness. The concept of privacy by design, incorporating privacy from the first step of the design project, aims for a sustainable data life cycle management. Prior consent in the form of opt-in can be seen as a prime example of the failure of these types of solutions. Though consent supposedly enables the user to take all decisions over his/her data, the user basically has very little or no clue as to the consequences. One of the problems is that privacy is reduced to the ability to hide one's data, whereas the real issues reside in the inferences that may be drawn from the data. Data Protection, moreover, requires that

[2] Cf. expert interviews reported in [4] indicate a strong need on the side of the industry for a level playing field that will enable enterprises to act ethically sound, once they are sure that their competitors are forced to abide by the same rules.

users become aware of the purpose of processing, while forcing providers to process only those data necessary (data minimisation). We therefore believe that consent is not the best way to engage the users. A more sustainable and meaningful way to ensure user participation is the use of a Data Licensing Agreement (DLA) that involves clearly demarcated mutual obligations based on a fair exchange.

Business Challenges for Personal Data Management Tools. The implementation of current personal data management tools entails several challenges. Most significantly, the costs can be challenging for small companies. Efforts to ensure data protection, as demanded by legislation, are coupled with investments in respective technologies to support such tools. Cookie regulation sets an example in this regard, but also the right to be forgotten: the request to delete all data related to one user might be difficult to comply with for practical reasons rather than not being willing to do so[3].

Another decisive challenge regarding opt-in is that it cannot work as OSN providers, whose operations are mostly data-driven, are located between advertisers and publishers. The latter companies do not offer any kind of value proposition that is relevant directly to the user. As a result, they have no leverage that interests users to care sufficiently about their personal data. The most successful data-driven companies of today, on the other hand, have several advantages. As social networks, they are able to provide opt-ins simple through existing user lock-ins. The value proposition is, at least partly, that the only alternative for the user to being tracked is to stop using the service[4]. Their scale furthermore allows them to create independent data management tools, owned and controlled by them.

Value Proposition for Personal Data Management Tools. Thus, the value proposition of personal data management tools, i.e. the incentive to use them, is a central issue in this regard. First, it depends on the interests of the users not only in knowing what is happening with their data, but also in investing effort and time in actually controlling it. Second, even those groups in society that are concerned about their personal data do often lack the knowledge to assess their own "value". Expectation management would thus be integral to a tool in question, as an individual's data is worth much less to the industry than many would think (the value is in the connections and inferences, not in the individual data points). In addition, it needs to be clarified how much value a user already derives from using free services, as these are sponsored through the personal data market. If these two aspects are taken into account, a value proposition for a personal data management tool may become feasible.

Last but not least, such a tool could make clear that, if a user voluntarily provides certain data, for instance through tracking, a company would make money of it, while the user could benefit from free services and more personalized marketing, provided there is transparency with respect to which personal data are used and for what purpose.

---

[3] Once the legal ground or the purpose for processing has been exhausted personal data should be erased or anonymized, cf. art. 12 and 14 of the current Data Protection Directive D/95/46/EC (DPD).

[4] Note that art. 7.4 of the upcoming GDPR may prohibit this: 'the execution of a contract or the provision of a service shall not be made conditional on the consent to the processing of data that is not necessary for the execution of the contract or the provision of the service pursuant to Article 6(1), point (b)'.

This could provide re-usable data for the benefit of several actors. However, we believe that such an exchange is only sustainable if users are informed about what they inadvertently disclose when interacting with their OSN.

3.2 USEMP: A Centralised Tool for Transparency

In the following paragraph we present how the USEMP tool for personal data management can become effective and interesting enough for users and businesses to utilize. The research conducted in USEMP illustrates (see [6]) that direct monetization of personal data by the data subject/OSN user is unrealistic, due to the low monetary value of an individual's information. Moreover, the value that users derive from free services, such as Google, Facebook, Twitter, news etc., is already immense, yet, indirect. Instead of monetization, a tool that focuses on transparency in the first place (with the potential to be extended) seems to promise most potential, both for the user and the business. USEMP's interdisciplinary use cases and scenarios highlight the importance of transparency and awareness (see [16]). They demonstrate the need to hand over a certain amount of control to users, notably over the definition of 'privacy' and 'sensitiveness of data' (even if this will not overrule mandatory law regarding the treatment of sensitive data as defined by law). Foremost, a tool should create knowledge and awareness about the personal data market, which in itself would already be a major step forward. In privacy literature this has been coined as relating to 'institutional privacy', as opposed to 'social privacy'.

The latter concerns disclosure of personal information to one's peers, the former concerns the capturing of (inferred) personal information by public and private service providers. Whereas users have developed intricate privacy strategies with regard to social privacy, they are hardly aware of their lack of institutional privacy (see for example [9]). Not only would the actual value of personal data become more clear, not only would it clarify the actual value of personal data more clear, it would also illustrate the implications of free online services. The development of comprehensive visualisations is certainly of main importance in this regard, as it improves the user experience and comprehensibility of these complex dynamics.

From the value network perspective, which we have adopted for this work, it is a central question where such a tool:

- needs to be "inserted" in the value network, in order to be most effective,
- which actor is most directly affected by the tool, and
- how can this actor pass obligations and benefits to its partners in the value network.

Due to the strong connectedness of the ecosystem in question, the right location in the value network can affect the whole structure. Depending on where the tool is located, different advantages may also arise for different actors, and for users.

Options for Implementing a Transparency Tool in the Ecosystem. The broad distinction we can make regarding the value network is between 'in the middle', 'at the sides', and at the front-end (user side).

Ad.networks and Technology Providers. So far, much attention has gone to the actors in the middle part of the ecosystem depicted in Fig. 1. These, the publishers and app developers, are data-driven companies, and all user-related data might flow through their systems. However, they are only intermediaries, facilitators of the actions of others. Also these actors are obliged to implement consent mechanisms in their operations, due to the so-called cookie legislation (which has been twisted by the industry to force users' hand, so in point of fact such consent has little meaning [5]. This is probably the least effective spot to realize effective data protection. Although the core of their business is data, and often pseudonymous data, they have no direct contact with the user. They depend also in this regard on the sides of the value network. Focusing on the actors in the middle is unfortunate, as they create value indirectly for the user, backing the free model by increasing efficiency and relevance of advertising and other content. Without the data, they cannot work, as targeting, delivering and evaluating all depend on it.

Publishers and Advertisers. The actors on the sides, i.e. publishers and advertisers (A and B respectively in Fig. 1), are arguably a better location to implement a personal data management tool. Indeed, they are legally obliged to implement data protection, including transparency tools. USEMP tools will thus mitigate their liability for violations by integrating empowering tools for DPbD. This is far more effective because they have direct contact with the user or customer. They also need a good reputation to stay attractive for the user. Furthermore, almost all other actors depend on the sides. Thus, they do not only have leverage over the users, but also over the actors in the value network. Due to the strong connections and dependencies, business-to-business pressure down the value chain should be utilized as a powerful accomplice in strengthening personal data protection rights. The user can only build a trust relationship with those actors he/she has direct contact with, if these actors provide reliable insights about their operations and how these may affect the user.

Benefits of a Centralized Tool. The most significant outcome of all possible scenarios thus seems to be that a centralized mechanism on the side of the OSN providers promises the most desirable outcomes; this would provide a tool with overarching effects on the whole ecosystem. This is not least the case because:

- the economic value of personal data is indirect, and through centralising it in a transparent way, it becomes clearer;
- it can have benefits for both the user and the industry by increasing transparency that is missing most in the ecosystem;
- most importantly, a centralised transparency tool could also promote smaller and European companies, which might otherwise not be able to invest in appropriate tools themselves;
- centralisation that clearly complies with the legal framework of data protection, notably by providing DPbD and therefore backed by competent Data Protection Authorities, would be a good starting point for creating trust on the side of users;
- finally, a tool where all data management related operations are centralised could create legal certainty and ensure accountability through transparency.

The Case for an Independent Platform. We note, however, that it is not obvious that OSN providers are willing to provide the kind of transparency that is required. Below, we will explain that the emphasis is on profile transparency, which is part of mandatory data protection law. OSN providers may wish to keep their inferences behind the walls of trade-secret and IP rights, finding them to be a central part of their competitive advantage. At the same time, users may not be willing to trust OSN providers' information about what they infer. They may prefer an independent platform to secure more impartial inferring tools. The USEMP tools aim to provide precisely such an independent platform.

Challenges for the Adoption of Centralized Platforms and USEMP Specifications. The challenge for a centralized platform such as the one proposed by USEMP, is how this can be applied in free market conditions. There are two types of driving force that can help the adoption of such a platform: a) regulatory policies that require such tools to be implemented, and b) self-regulatory policies that can be enforced by the industry itself (see for example [17]). In addition to any regulation, the proposed solution acts also as a set of specifications that can be implemented by more than one platform by consortia of business actors that can see the benefits of introducing such solutions (for example to increase trustworthiness for the end-users).

The USEMP consortium plans to experiment in such a centralised solution where there will be contractual relationship with the user and the platform that will allow the sharing of user data on the basis of a granular license to use and/or reuse the data. This will, for instance, enable a one-time licence (or prohibition) of the processing of specific types of data, specified reuse, or particular third parties. Such granular licensing can be implemented clearly all over the value network. To render the consent that is involved in concluding a data licensing agreement (DLA) meaningful, it is important to develop indicators of personal data privacy and value to end users. In the following sections of the paper we describe the proposed USEMP framework for the collection of the necessary information and computation of such indicators.

4 USEMP Disclosure and Personal Value Indicators Framework

In order to develop tools that will increase the transparency of usage of personal data by advertisers and other third parties, it is important to develop personal data value and disclosure indicators that can be used to provide end-users meaningful insights. In the following paragraphs, we present a framework that has been developed as part of the USEMP project. Since these indicators are part of the personal data for each end-user of the USEMP tools (referred to also collectively as Databait), their use is also governed by the proposed DLA approach in Sect. 2.

In the early stages of the development of the USEMP disclosure scoring framework we identified a list of personal data attributes that have been qualified as sensitive or

valuable by the users.[5] For users, to better perceive the different aspects of their privacy, it is useful to organize the attributes in a semantic manner. To this end we organize the identified attributes in a number of high-level categories that we refer to as disclosure dimensions[6]. This organization allows for clear and intuitive presentation and handling of the different aspects of a user's personal information.

For instance, one of the disclosure dimensions to be considered is demographics, which includes user attributes such as age, sex, etc., and another is health factors, which includes attributes such as smoking and drinking, etc. Such a grouping has multiple benefits for the end user. First, it enables him/her to form a succinct, easy to grasp mental model of the disclosed personal information and to prioritize its different parts. Second, it enables the use of different compact visualization methods that will further augment the user's awareness with respect to his/her personal information.

From a legal perspective we note that much of this data falls within the ambit of sensitive data, which has stringent legal implications. Providing these disclosure dimensions will enable users, OSN providers and other stakeholders to get a better understanding of the sensitive data that are inferred from user data and will thereby enable a clear attribution of liability for the processing of such inferred data.

On top of this disclosure dimensions framework, we develop the USEMP disclosure scoring model, by enriching it with disclosure and data value scores. Having organized the user attributes in the disclosure dimensions structure, we proceed by enriching it with disclosure and data value scores. Disclosure scores are about quantifying the potential negative impact entailed by the disclosure of different parts of the personal information of a user. The economic (though not monetary) value of a user's shared personal data in OSNs (e.g. posts) is inferred by measuring its impact on the user's social graph, i.e. audience (in terms of reactions, e.g. likes, shares, comments).

4.1 Disclosure Dimensions

We define eight (8) key categories of personal attributes, which we name *disclosure dimensions*. These are: (A) Demographics, (B) Psychological Traits, (C) Sexual Profile, (D) Political Attitudes, (E) Religious Beliefs, (F) Health Factors and Condition, (G) Location, and (H) Consumer Profile.

These dimensions cover a wide variety of personal information, which OSN users in many cases consider of private nature (perceived privacy), and also encompass information that is considered sensitive from a legal perspective (legally sensitive data). In addition, based on current business practices (mainly stemming from the marketing industry),

[5] We recognized that we need to qualify this as 'perceived' sensitivity, since when the law qualifies certain data as sensitive, based on art. 8 Data Protection Directive (DPD), this has major legal effect, which, however, does not depend on how a user 'feels' about the data.

[6] Clearly, these dimensions are not exhaustive and they do not necessarily match with the legal right to privacy as stipulated in art. 8 of the European Convention of Human Rights, or with the fundamental rights to privacy and data protection of the Charter of Fundamental Rights of the European Union. It is pivotal that perceived privacy and the right to privacy are understood on their own merits, taking note that the latter aims to provide the level playing field for users to develop their own privacy preferences.

the identified dimensions are associated with certain value levels, i.e. they carry a certain level of utility for (marketing) companies that are interested in targeting consumers. Table 2 summarizes the eight identified disclosure dimensions, along with the value levels associated with them.

This set of eight disclosure dimensions constitutes the current top-level *schema* of the USEMP privacy model, and although we do not foresee considerable changes at this level, the implementation of the overall framework is generic enough and can accommodate such changes if needed (e.g., addition of a new dimension, splitting of an existing dimension into more).

Table 2. Overview of USEMP disclosure dimensions

#	Name	Description	Potential threats-Sensitivity	Value (for advertisers)
A	Demographics	Personal data, such as Gender, Age, Nationality, Ethnic background, etc.	Discrimination in a variety of settings. The most frequently used type of information.	**High:** advertisers wish to target users of certain demographic criteria
B	Psychological Traits	Defined by psychologists (extraversion, openness, etc.)	Discrimination, e.g. in personnel selection	**Low:** a limited number of advertisers can connect type of personality to their product
C	Sexual Profile	Relationship status, preferences, habits	Discrimination, e.g. in workplace, education, housing	**High:** advertisers wish to target consumers based on their relationship status/lifestyle related to their sexual profile
D	Political Attitudes	Supported politicians, parties and stance	Discrimination, e.g. in workplace or personnel selection	**High:** advertisers wish to target consumers based on the political affiliations since these are related to their general profile
E	Religious Beliefs	Religion (if any) and beliefs	Discrimination, e.g. in the sale or rental of housing, job selection, workplace.	**Moderate:** advertisers wish to target consumers based on their religious and cultural beliefs

(*Continued*)

Table 2. (*Continued*)

#	Name	Description	Potential threats-Sensitivity	Value (for advertisers)
F	Health Factors and Condition	Habits (e.g. smoking, drinking), medical conditions, disabilities, health factors (exercise)	Discrimination, e.g. health insurance denial or discriminatory pricing.	**High:** advertisers wish to target consumers based on their habits
G	Location	Characteristic locations of the individual and history of previous locations	Discrimination, e.g. house insurance, stalking	**High:** advertisers wish to target consumers based on their current location or their home location
H	Consumer Profile	Preferred products and brands	Ad targeting and discrimination in online price-setting	**High:** advertisers wish to target consumers based on their consumer profile attributes like the devices the use to access digital content

The disclosure dimensions framework effectively defines a hierarchy where the top level represents the OSN personal data profile, the next level contains the eight (8) disclosure dimensions, the level below contains the set of attributes of each dimension and the lower level contains a predefined set of possible values for each attribute (please note that the word value here refers to the possible values an attribute may take and should not be confused with the concept of personal data value).

We link OSN data to this framework by considering a variety of detection and analysis mechanisms, e.g. multimedia information extraction techniques and inference techniques, and OSN presence data (typically in the form of observed user activities, e.g. likes [18], posts, user interactions, or volunteered profile information). In short, we use a user's OSN presence data (e.g. posted content, likes, set of friends), in order to predict the values of the user's attributes. This involves both utilizing explicitly provided information and also producing a number of inferences.

4.2 Disclosure Scoring

Initially, the USEMP research team identified the need for a scoring framework that would help the end-user understand better which OSN or web behaviour actions may disclose personal information related to the disclosure dimensions presented above. Such a mechanism is based on perceived privacy and highlights what can be disclosed

about users based on inferences on their data. We make clear to the users that this disclosure is based on algorithmic decision making and need not at all be correct, but emphasize that this is how the current personal data ecosystem operates: the value chain is based on such probabilistic disclosures. Additionally, overall disclosure scores are computed at each level of the proposed hierarchy. In the next subsection we provide more details about the scoring framework.

Structure: A schematic illustration of the disclosure scoring framework is shown in Fig. 2. USEMP disclosure scoring framework builds on top of the disclosure dimensions and assigns a number of scores to the elements of each level. These scores express different aspects of disclosure, e.g. the perceived sensitivity of different types of information, the confidence that some value holds for some user, an overall disclosure score and other aspects that will be described shortly Clearly, the two important characteristics of this framework are the following: (a) it is tailored to the hierarchical structure of the disclosure dimensions, (b) there are multiple scores associated with the elements of each level of the hierarchy. Hence, the framework enables the following two kinds of user awareness: (a) navigation through the levels of the hierarchy and understanding of how the scores for some particular value *affect* or *are affected by* the levels above and below it, and (b) focus on specific aspects of the factors that are related to perceived privacy; e.g., it will be possible to focus on visibility, the overall disclosure score, etc.

Here, we consider an additional level at the root of the framework, which contains any type of data that is generated as a result of a user's behaviour and interaction with the services of an OSN operator. This includes posted content (text, images), explicitly declared profile information, user network data, sets of likes, etc. We call this the *OSN presence data layer* and consider it as the primary source for populating the disclosure scores for the given user. Naturally, between the perceived privacy values level and the online presence data, there is a layer of modules that perform various mining and inference procedures. The overall framework is visualized in Fig. 2.

Computation is carried out in a bottom-up manner. That is, information that comes from the OSN and the inference mechanisms is used to fill the scores at the values level and then the computed scores at the values level are used to fill the scores at the upper levels, one level at a time. Starting from the level of values, the scores that characterize each value are the following:

(a) *Confidence*. This is a continuous value in the range from 0 to 1. It represents how confident we are that the corresponding value is true and is typically computed by the inference algorithm along with the produced inference. It needs to be noted that the confidence values under the same attribute should sum to 1 (except for the case that an attribute can take multiple values simultaneously).
(b) *Sensitivity*. This score represents how sensitive the user feels that this particular piece of information is. It also ranges from 0 to 1.
(c) *Visibility*. This score attempts to quantify the set of people to which the relevant information is accessible and consists of three sub-scores. The first is the overall visibility score and is also a continuous value in the range from 0 to 1 (lower

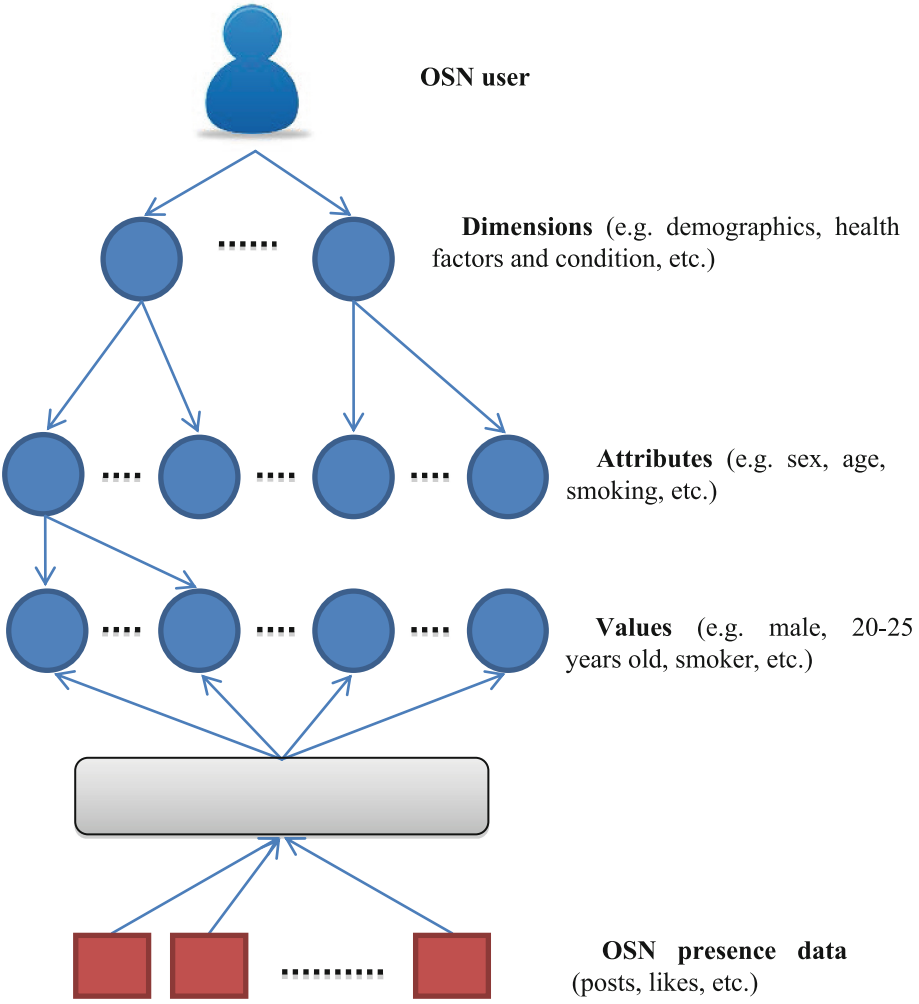

Fig. 2. Overview of USEMP disclosure scoring framework

values denote that the corresponding piece of information is accessible by fewer people, whereas a value of 1 denotes public information). The second visibility sub-score is a qualitative label that is related to the overall disclosure score and expresses the widest possible audience to which this information is accessible. For instance, a value whose overall visibility score is 0 has a visibility label of "Private", a value whose overall visibility score is 1 has a visibility label of "Public" and an intermediate value denotes the widest group of people that have access to the value, e.g. "Friends" or "Friends of friends", etc. This sub-score is called "visibility label". The third visibility sub-score expresses an estimate of the *actual audience* that sees this value and we refer to it as "actual visibility". It is an

integer number representing the actual number of users that are aware of that value and depends on the estimates of the actual audience of the content that has been used to infer that value. It should be noted that the current implementation of the scoring framework does not compute the actual audience, however, different approaches are considered for estimating it.

(d) *Declared/Inferred*. This is a binary value that defines whether our knowledge about the particular value comes from explicitly provided information that the user has provided or has been inferred (derived). It is not an actual score but reflects information that is important for maintaining a complete view of disclosure with respect to some particular value. Additionally, in some cases a value may be both declared and inferred. In such cases, the value will be considered as declared (i.e. declared will override inferred).

(e) *Support*. This field is not really a score, but rather provides a link to the OSN presence data that support the particular value. In the case that the support for the value is associated with an inference mechanism, this field points both to the inference mechanism and the data that the inference mechanism used. This field is particularly important because it allows the user to understand the types of content that are important for his privacy.

(f) *Level of control.* This score represents the ability of a user to control the disclosure of data about him/her. It ranges from 0 to 1; low values will denote a limited ability to control the disclosure of this particular data about the user. The ability of a user to control the disclosure of data about him/her may be limited by the fact that the support of some value may involve also data posted by other users that the user him/herself cannot control. This score is set by evaluating each piece of shared information with respect to (a) ownership of the data from the end-user (or someone else), (b) the permissions framework of the social network that may allow the user to stop this information from being shared or not.

(g) *Disclosure score.* Eventually, the framework includes an overall perceived score that provides a succinct idea about the overall privacy status of an OSN user (see for example [13]). It is a function of other scores: confidence, sensitivity and visibility. Higher values of the score denote a higher exposure of personal information that is perceived to be of private nature. Note that although the disclosure score essentially summarizes the other scores, the model maintains a separate list of the individual scores (confidence, sensitivity, and visibility) in order to support richer visualization and analysis capabilities (e.g. separate visualization of visibility and sensitivity).

The three upper levels of the proposed disclosure scoring structure, namely the user, the dimensions and the attributes, are all associated with the following set of scores: (a) Visibility, (b) Disclosure score and (c) Level of control. These have the same meaning as the corresponding scores at the values level. In addition, the top level (user) is also associated with an overall personal data value score (please see the next subsection). For a full description of proposed estimators see [7, 19].

4.3 Personal Data Value Indicators Framework

In addition to disclosure scores, a set of personal data value indicators are required so that the end-users can be informed about the value of the data they are sharing. The proposed personal value indicators are based on the activities of the end user in the OSN environment and his/her OSN social graph. Two basic indicators are initially proposed:

(a) a measure of influence for a specific person, referred to as Influence score and denoted with I, that is based on the history of the objects that the specific person has created in the OSN;
(b) a measure of the importance of an object (picture/video/post), denoted with M, that is posted to the OSN. M is calculated taking into consideration the type of action on the specific object of the first- and second-hop friends of the object creator.

The *Influence score* of a specific person should be estimated based on the history of the objects that the specific person has created, while taking into consideration the

- number of connections comparing to the total number of users of the network;
- types of actions (share, like, comment) of the first and the second hop friends on the objects that the corresponding person has uploaded/created to the OSN.

For the calculation of user influence *I*, the following parameters are proposed to be collected and used:

- number of objects (i.e., picture/video/post) that a user has created
- number of first- and second-hop friends
- total number of first- and second-hop friends that had an action on each object (i.e., picture/video/post)
- type of action (i.e., share, like, comment) of user j on the object i

The parameters listed above can be collected and combined to different formulas to compute values for variables *I* and *M*.

The proposed personal data value combines these two factors (*I*, *M*) and is calculated as follows:

$$V = I \cdot M$$

This initial set of defined value indicators (data value *V*, user influence score *I*, object importance *M*) is defined so that it can be computed from actual OSN data (like Facebook). As part of future work these parameters will be collected and computed with actual data from pilots on top of Facebook and with simulated data from theoretical models and various formulas for *I*, *M* will be tested (for a full description of proposed estimators see [7]).

4.4 Experimental Tools and Visualization

In order to evaluate the ideas developed in the USEMP project, a set of pilots have been scheduled and a set of tools are developed to provide a testbed for collecting data and end-user feedback, these are referred by USEMP partners as the Databait tool. These include:

- **Databait web browser plugin:** a browser plugin that is used to collect users' browsing data during the pilot and that allows end-users to block tracking behaviour or offer users recommendations with respect to sharing data.
- **Databait webapp:** a web application that allows end-users to view indicators of their online social network sharing behaviour with respect to transparency and data value.
- **Databait backend:** a framework and set of services that collect data about user behaviour and compute a number of indicators with respect to disclosure and data value that are shared with end-users via the Databait webapp visualizations.

In the following Figs. 3, 4 and 5, we present some examples of the UI/UX of the web browser plugin and webapp visualizations for the Databait tools (this is work in progress to be validated after the completion of the USEMP pilots):

- Figure 3: presents the look and feel of the Databait web plugin that is responsible for selecting trackers and blocking (or unblocking them).
- Figure 4: presents the look and feel of the Databait web application that provides the end-user with access to a number of visualizations/tools and information from OSN shared data.
- Figure 5: presents the look and feel of two types of visualization that allow the user to understand if the shared data disclose any location (from Facebook posts analysis) or interests (derived from image analysis) and present them in an intuitive way.

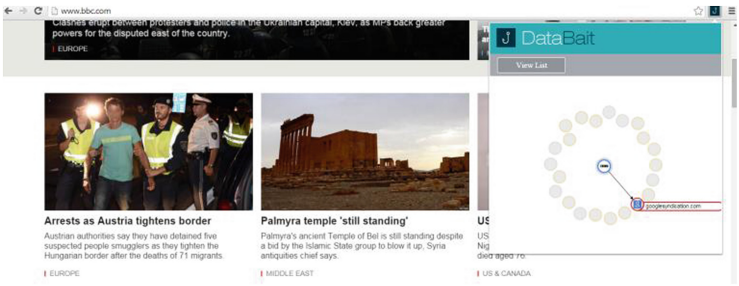

Fig. 3. Databait web plugin – allows end users to view third-party tracking services and block them (or unblock them)

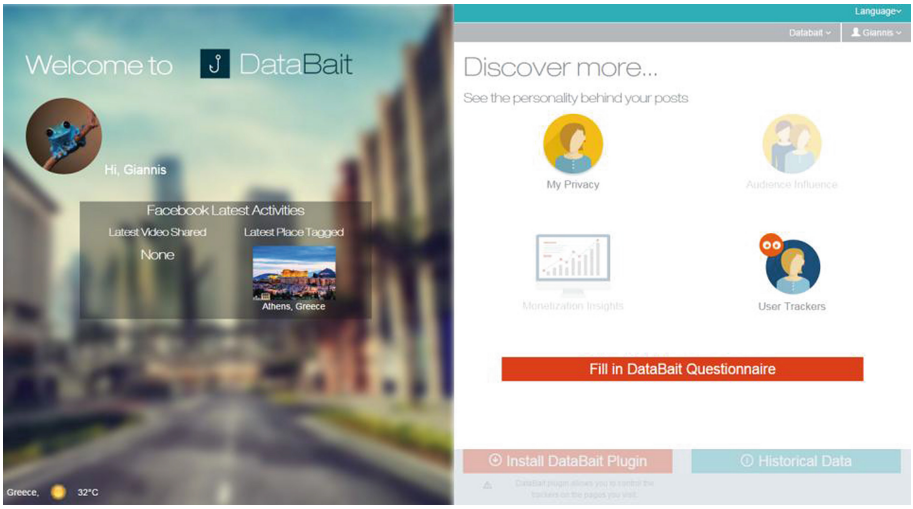

Fig. 4. Databait web application – access to visualizations for transparency and data value

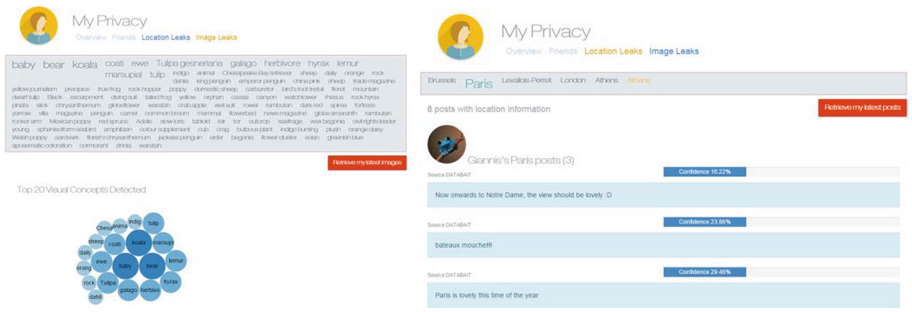

Fig. 5. Databait web application – access to visualizations for transparency and data value: (a) left screen shows the concepts detected in a user's shared images, (b) right screen show the locations detected in a user's shared posts

4.5 Evaluation of USEMP Tools

We have performed a series of user studies within the USEMP project in order to build the Databait tools on solid grounds of user acceptance in the form three focus group sessions with end users. Hence the study acted as a formative evaluation approach by involving end users to the design and evaluation of tools while the project moves forward. Three focus group interviews were conducted in English within March 2015 together with 15 participants from Botnia Living Lab[7]. The design of the focus groups

[7] Botnia Living Lab is an environment in Sweden for human-centric research and the development and innovation of new ICT based solutions.

was based on gathering end user insights on main theme of transparency of personal information through USEMP tools. For this purpose we presented the participants with the mock-ups of the USEMP tools/Databait (as presented in Figs. 3, 4 and 5). Each function demonstration was then followed by questions targeted on insights for values, motivations and barriers to use. We briefly discuss the participant feedbacks here.

We first asked users about the normal social media and internet usages in order to indirectly and directly capture their awareness of the ways they disclose information. During the course of a normal social media usage there are different communicative actions which are established. Users have personal motivations and external influencing factors that force them to use social media. Therefore their usage is not totally optional and for this reason their personal information is inevitably disclosed. Among personal motivational factors that can be enumerated are the willingness to reach a wider audience in order to promote themselves for example with different political activities, keeping in touch with the families, friends, acquaintances, keeping track of their events like their friend's birthdays, to keep themselves updated about what is happening in surroundings and etc. Disclosing various types of information is evident in these types of social media usages. We asked them about the kinds of information they think they are disclosing in their everyday usage of social media to capture the level of their awareness towards privacy issues. Our analysis showed that most users think of privacy as only the basic personal information they disclose voluntarily like name, age, relationship status. The awareness towards the observed and inferred data sources is extremely low among the users.

Next, the Databait tools were presented using the mocks of the future tool and the expected results (e.g. inferences). The participants were allowed to freely discuss about the features and ask questions about the functions. Therefore we created a milieu for the lively discussion and to capture their concerns and how the tool could serve them in different scenarios. For example we could observe that most of the participants were curious about the features and at some points were shocked by the level of the tool's sophistication. One of their pivotal points was related to the unconscious disclosure of personal information that might have an impact on one's public image and how data processors could gain value of their information. The users saw the benefits in this awareness awakening through manipulation of informed disclosure. The benefits were also associated to the disclosed information at various levels determined by each and every user's beliefs, cultures, economical values gained, political outreach and etc. Databait's personality trait function showed to be beneficial in this sense since users can be sensitive to different subjects.

Photo and location leaks functionalities could draw user's attention on various levels of disclosure both those revealed intentionality and those that are unintentional. From intentional point of view users find this helpful with respect to the values of the contents to the Social Media owners. So what made them more aware of their shared content was the ability to see the profits of their contents from the social media owner's perspective; to see what could be gained from the contents and how those could be inferred. Even though they are aware of their shared contents, their perception of the contents' secondary usage was limited so that social media owner's bad intentions could hide in the user's low institutional privacy awareness. Unintentionally revealed sensitive information interpreted by Databait could help the participants learn more

about the adverse effects of their actions and seek to possible solutions e.g. deleting photo/location leak or limiting audience.

To summarize, the result from evaluation of designed concepts, showed that users are curious about the revelation of values that could be drawn from personal information and generated content. We found that users are willing to be more educated through the tool about adverse effects of their sharing habits triggered by a sense of dread that could raise their awareness. Here the idea is that the users are more intrigued when they see dangers more explicitly. This has then led the users to perceive such privacy tools to be more effective. Simplicity showed to have an impact on how the users are willing to adopt a tool as well. Most of the users agreed that the tool needs to have a 'simple to use' settings with self-explanatory features. Our aim in USEMP project is to take this end user's perspective into account for the next versions of the tools.

5 Conclusions and Future Work

In this paper we have presented the results of the multi-disciplinary project USEMP in developing a value model for the use of personal data in advertising and OSNs that empowers the end user and offers more transparency to the use of personal data. The presented research describes:

- what are the legal aspects of users' privacy in OSNs that can be addressed by transparency tools that are based in the principle of Data Protection by Design (DPbD);
- how a DLA model is more appropriate than that of simple prior-consent to improve trust and user control on sharing personal data in OSNs;
- how a centralized tool developed as an independent platform is more appropriate for the business ecosystem to improve trust of the end-consumers to the advertising and marketing industry;
- how a disclosure scoring framework can be developed to support such a transparency tool;
- an overview of visualization methods that can be used as part of such a tool.

The ideas examined in this paper are currently implemented in the form of Databait tools and they are under evaluation from the pilot experiments organized from the USEMP project. Their impact on the end-user perception of privacy and the creation of new innovative business models that can support DPbD in OSNs and online advertising/marketing will be presented in future publications.

References

1. Popescu, A., Hildebrandt, M., Papadopoulos, S., Claeys, L., Lund, D., Michalareas, T., Kastrinogiannis, T., Pierson, J., Padyab, A.M.: User Empowerment for Enhanced Online Presence Management– Use Cases and Tools, APC15, 2015 (forthcoming)

2. van der Graaf, S., Vanobberghen, W., Kanakakis, M., Kalogiros, C.: Usable trust: Grasping trust dynamics for online security as a service. In: Tryfonas, T., Askoxylakis, I. (eds.) HAS 2015. LNCS, vol. 9190, pp. 271–283. Springer, Heidelberg (2015). The valorization of surveillance: Towards a political economy of Facebook. Democratic Communiqué, 22(1), 5–22
3. Searls, D.: The Intention Economy (book) 2012
4. London Economics, Study on the economic benefits of privacy-enhancing technologies (2010). http://londoneconomics.co.uk/blog/publication/study-on-the-economic-benefits-of-privacy-enhancing-technologies-pets/
5. Hildebrandt, M., Tielemans, L.: Data protection by design and technology neutral law. Comput. Law Secur. Rev. **29**(5), 509–521 (2013)
6. USEMP Deliverable D3.5, Socio-economic value of personal information, 2015-04-01, this report presents a socio-economic perspective on the tool for user-centred personal data, management as envisioned by the USEMP project. http://www.usemp-project.eu/documents/deliverables
7. USEMP Deliverable D6.1 (under review), USEMP privacy scoring framework, 2015-0-15, the current deliverable is a technical report accompanying the first version of the USEMP privacy scoring framework, a tool that aims at raising the awareness of users about the disclosure and value of their personal information. http://www.usemp-project.eu/documents/deliverables/
8. Constine, J.: Facebook, Google, and Twitter's war for app install ads (2014, November 30). Retrieved January 27, 2015. http://techcrunch.com/2014/11/30/like-advertising-a-needle-in-a-haystack/
9. Young, A.L., Quan-Haase, A.: Privacy protection strategies on Facebook: The Internet privacy paradox revisited. Inf. Commun. Soc. **16**(4), 479–500 (2013)
10. Van Dijck, J.: The culture of connectivity: a critical history of social media. Oxford University Press, Oxford; New York (2013)
11. Hildebrandt, M., O'Hara, K., Waidner, M. (eds.): The Value of Personal Data. Digital Enlightenment Yearbook 2013. IOS Press, Amsterdam (2013)
12. Hildebrandt, M.: "The Dawn of a Critical Transparency Right for the Profiling Era", in Digital Enlightenment Yearbook 2012, pp. 41–56. IOS Press, Amsterdam (2012)
13. Liu, K., Terzi, E.: A framework for computing the privacy scores of users in online social networks. ACM Trans. Knowl. Discov. Data, **5**(1), Article 6 (2010)
14. Moody, D.L., Walsh, P.A.: Measuring the value of information: An asset valuation approach. In: Morgan, B., Nolan, C. (eds.) Guidelines for Implementing Data Resource Management (4th Edition). DAMA International Press, Seattle, USA (2002)
15. Pagefair 2015 ad-block Report. http://blog.pagefair.com/2015/ad-blocking-report/
16. USEMP Deliverable D4.1, Social Requirement Analysis, 2014-08-18, this document presents the methodology used and the results of the first user research including focus groups that highlight the requirements for a privacy enhancing tool. http://www.usemp-project.eu/documents/deliverables/
17. NAA ad-choices initiative for consumers opt-out for behavioural targeting. http://www.networkadvertising.org/choices/
18. Theodoridis, T., Papadopoulos, S., Kompatsiaris, Y.: Assessing the reliability of facebook user profiling. WWW (Companion Volume) **2015**, 129–130 (2015)
19. Petkos, G., Papadopoulos, S., Kompatsiaris, Y.: PScore: Enhancing privacy awareness in online social networks. In: International Workshop on Multimedia Forensics and Security (MFSec) (2015)