Fostering Citizens' Participation and Transparency with Social Tools and Personalization

Vittorio Scarano, Delfina Malandrino, Michael Baker, Françoise Détienne, Jerry Andriessen, Mirjam Pardijs, Adegboyega Ojo, Michael Hogan, Albert Meijer, and Erna Ruijer

Abstract In this paper we present innovative solutions to the problem of transparency in Public Administrations (PAs) by opening up public data and services so that citizens participation is facilitated and encouraged with a Social Platform and a personalized user-friendly Transparency-Enhancing Toolset.

Introduction

In our research, transparency is a characteristic of an interactive collaborative process between local government and its citizens. Different forms of collaboration are characteristic of how local governments and its citizens interact, specific for

V. Scarano (✉) • D. Malandrino
Dipartimento di Informatica, Università degli Studi di Salerno, Salerno, Italy
e-mail: vitsca@dia.unisa.it; dmalandrino@unisa.it

M. Baker • F. Détienne
CNRS - Telecom ParisTech, Paris, France
e-mail: michael.baker@telecom-paristech.fr; francoise.detienne@telecom-paristech.fr

J. Andriessen • M. Pardijs
Wise & Munro Learning Res, Den Haag, The Netherlands
e-mail: jerry@wisenmunro.eu; mirjam@wisemunro.eu

A. Ojo
Insight Centre for Data Analytics, National University of Ireland Galway,
Lower Dangan, Galway, Ireland
e-mail: Adegboyega.Ojo@insight-centre.org

M. Hogan
National University of Ireland, Galway, Ireland
e-mail: michael.hogan@nuigalway.ie

A. Meijer • E. Ruijer
Utrecht School of Governance, Utrecht, The Netherlands
e-mail: A.J.Meijer@uu.nl; H.J.M.Ruijer@uu.nl

A. Ojo, J. Millard (eds.), *Government 3.0 – Next Generation Government Technology Infrastructure and Services*, Public Administration and Information Technology 32, DOI 10.1007/978-3-319-63743-3_8

each local context. When such interactions become more transparent, this may mean greater accessibility, more sharing of ideas, better understanding on how the other works and thinks, etcetera. Such processes can be studied at different levels of abstraction, from that of individuals to that of democratic systems. It is our contention that social media can support transparency, especially when the interactions are meaningful and deliberately focused on joint issues and backed up with knowledge about ongoing developments and facts related to the issue at stake. Such knowledge comes from the use of Open Data. Wide-spread access to the Internet has greatly reduced the cost of collecting, distributing, and accessing government information. But an important effect of the diffusion of networks in the population is the potential of ICT, by promoting good governance, strengthening reform-oriented initiatives, reducing potential for corrupt behaviours, enhancing relationships between government employees and citizens, allowing for citizen tracking of activities, and by monitoring and controlling behaviours of government employees, is able to effectively reduce corruption (Bertot et al. 2010).

Open data are significantly seen, in general, as the main instrument to improve transparency, at every level. In fact, transparency for local government can be defined in different ways. The traditional view of transparency is that governments provide information on their work, on the other hand, governments are requiring transparency also from their dependents (such as non-profit-organizations and entities that they regulate in the private sector) (Michener and Bersch 2011). Hence, transparency is an interactive concept.

The early attempts and most of the current open data efforts are somehow failing to enable the transparency in its interactive aspects. Some of the main barriers often cited by researchers is the complexity of the information that is provided and the lack or inadequateness of tools that can help citizens in making sense out of the highly specialized datasets that are provided by government. Of course, this is crucial, as noticed by Michener and Bersch (2011), since the quality of transparency does not only depend on how (and how much) information is made visible and accessible, but also on how the information provided can be fruitfully used to accurate inference.

Some researchers (Mishory 2013) have indeed noticed that transparency is not an "object" but it is rather a "relationship", and, in order to yield better outcomes from transparency programs, it is of primary importance to design a relationship toward greater trust between the "supplier" of open-data based transparency (government) to the "recipients" (i.e. the citizens). In this context, it is crucial to ensure that citizens must be facilitated in their work toward comprehension of what data is important for them.

Many have observed (Bonsón et al. 2012) that local governments in Europe are often well behind their citizens in the use of the social dimension of the Internet, since they usually prefer to use ICT technologies (and social networks and the Internet) only as a one-way channel where they can communicate easily and at low cost.

As noticed by Davies (2012), the task of opening up data to citizens is seen as completed once data is correctly published, missing the important dimension of

open data as the catalyser of discussions, conversations and collaborations around data, between citizens and between citizens and the government. The lack of citizens engagement has led Tim Davies to present the "Five stars of Open Data Engagement", that range from the data being demand-driven (1 star), being put in context (2 stars), being supporting conversations (3 stars), being able to build capacity, skills and network (4 stars), to the full collaboration on the data as a common resource (5 stars) (Davies 2012).

Furthermore, as noticed in Colpaert et al. (2013), not only open data needs to be of good quality for others to transform them into knowledge and make them useful, but open data programs must also ensure that the citizens as well as developers can discuss about the datasets and around them, in order to stimulate and favor its re-use.

Summing up, research is clearly pointing out the direction where citizens

(1) must not be left alone in making their comprehension of open data and should be able to collaborate and interact socially around them,
(2) must be provided with tailored and contextualized data,

so that they can relate open data and transparency as a relationship between them (and their daily problems) and the government.

Our research is, then, motivated in providing an interactive solution for improving the engagement of the citizens

- by making them able to socially interact over open data, by forming or joining existing online communities that share common interest and discuss common issues of relevance to local policy, service delivery, and regulation;
- by providing a robust and more holistic understanding of transparency, by underpinning the next generation open-data based transparency initiatives, ensuring that published data are those of value to citizens, with a personalized view in different forms to different segments of the citizens and public based on their profiles for facilitate better understanding.

Our main purpose here is to engage citizens through a "purposeful and personalized relationship" between citizens and open data, not only on a personal basis, but between government and networks of citizens that collectively attribute meanings to this information. The information provided by Open Data is shared, interpreted, personalized, made easier to understand and discussed, to assess its meanings.

Concretely, we describe here our approach instantiated in the design, development and evaluation of

- a Social Platform for Open Data (SPOD) enabling social interactions among open data users and between open data users and government data;
- a Transparency-Enhancing Toolset (TET) as extension for existing major Open Data Platforms, enabling easier access to the relevant dataset, a better understanding of these datasets and integration with social platforms for sharing and discussing datasets

Our research is conducted within a Horizon 2020 European funded innovation project, called ROUTE-TO-PA (www.routetopa.eu) to improve the impact of ICT-based technology platforms for transparency. The ROUTE-TO-PA team is strongly heterogeneous and multidisciplinary. It integrates theory, research, innovation and transformation of local practices, by encompassing research partners, small and large companies, pilot Public Administrations (from four different countries) and one non-profit foundation. Our research partners have expertise in e-government, computer science, learning sciences and economy. In a sense, our team has been designed to tackle the challenges of transparency with a trans-disciplinary approach. To wit, the research described here is strongly embedded into all the areas of reference for the project.

Organization of the Chapter

The cyclical approach involves elicitation of requirements, design and implementation of the technology and evaluation of the result, in our trans-disciplinary effort consist of four phases. The first two phases involve focussing on user requirements from the bottom up (user workshops) and from a top-down perspective (model of the societal context). In out bottom-up approach we adopted a collective intelligence and scenario-based design approach (Warfield, 19..; Hogan, ...). Based on initial scenarios developed for each Public administration site we derived user stories based on general TET and SPOD affordances. On these, we collected user feedback about barriers and possible solutions to develop use case models and descriptions to obtain more detailed functionalities and capabilities of the system.

At the same time, we studied the context around the platform, i.e. the people and their relations, within their professional practices, in order to provide further requirements coming from an abstract representation of the "Societal Activity Model of Open Data Users" described in successive Section**.**

At this point we were able to begin the third phase, i.e., to provide the design and implementation of the ROUTE-TO-PA platform, that encompasses the two tools: the Social Platform for Open Data (SPOD) and the Transparency-Enhancing Toolset (TET), as described in section "The ROUTE-TO-PA Platform". The details on each tool is given in dedicated subsection.

Finally, in the fourth phase, the platform is undergoing an evaluation at the same user sites that were involved already during the user workshop, in Prato (Italy), The Hague and Groningen (The Netherlands), Dublin (Ireland) and Issy-les-Moulineaux (France).

It is worth noticing how the phases of modeling, design and evaluation are occurring iteratively, in repeating cycles, during the project, in order to allow experiments and pilots' feedback to be directly taken into consideration into the design and implementation.

Requirements and User Involvement

A series of carefully designed workshops were conducted, one in each pilot site, for the purpose of developing a comprehensive set of user needs, from the viewpoint of key stakeholders. Each workshop brought together experts, academics, industry specialists, open data practitioners, representatives of governments, open data researchers, and potential users (including citizens, representatives of citizens and social service institutes, and journalists) to brainstorm on open data platform adoption challenges, solutions to the challenges and a set of needs and requirements necessary for consideration in the design of the ROUTE-TO-PA platform. The emphasis on citizen participation and collaborative design in the methodology seeks to address the goals of improved government transparency and accountability for decision-making. Each workshop began with a collective intelligence (CI) analysis of barriers to accessing, understanding and using open data, followed by an analysis of options that may overcome these barriers. Participants then worked to develop scenario-based user needs, which involved profiling user needs in light of the barriers and options and high level scenarios of open data usage.

The methodology used to gather user-level requirements is inspired by a scenario-based design (SBD) approach (Rosson and Carroll 2002), but builds upon this approach by adding a collective intelligence (Warfield 2006) and agile user story development (Cohn 2004) approach. In the current application of CI, workshop participants worked to develop scenario-based user needs, which involved profiling user needs in light of the barriers and options and high level scenarios of open data usage. This included a separate focus on (1) information needs, (2) social and collaborative interaction needs, and (3) understandability, usability and decision-making needs. Idea writing was used for each cluster of needs. High level scenarios including multiple users were used to prompt thinking in relation to user needs. All the short user stories generated by participants were generated in the form:

As user type _____ , I want _____, so that I can _____

The wants (or needs) generated by participants across each pilot site were then analysed and key categories of user needs identified. Reasons for specified user needs were also analysed, and this analysis was used to advance our understanding of the scenarios and prospective use case models. This work in turn has shaped the test and evaluation framework (see Fig. 1).

The scenarios used addressed various contextual issues, relevant to each of the workshop sites, and aligned with the primary case focus in each pilot site. For example, the Dublin workshop focused on community networking and opportunity creation; the Groningen workshop focused on the use of Open Data in overcoming issues associated with population decline; the Den Haag workshop focused on Open Data in relation to employment and opportunity creation; the Prato workshop centered on local policy and budget issues; and finally, the workshop in Paris focused on Open Data in relation to start-up companies and the digital economy.

As such, there was some variety in user needs generated, across all there categories of needs: information needs, social and collaborative needs, and understandability, usability and decision-making needs (Hogan et al., submitted).

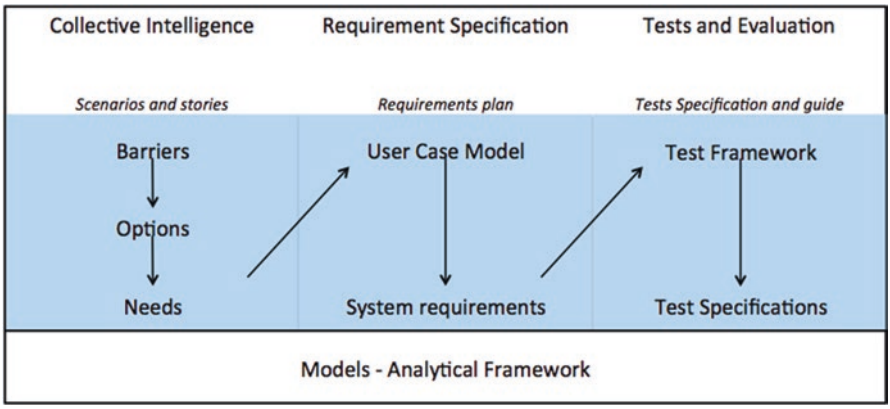

Fig. 1 Workflow for collective intelligence and link between collective intelligence and evaluation framework and test specifications

As workshop participants in each pilot site were working with a variety of scenarios, the user information needs generated were numerous and diverse. The information needs included, for example, demographic information needs; legal information needs; health information needs; social and community information needs; planning information needs; services, amenities and event information needs; business and financial information needs; jobseeker information needs. Essentially, the data and information that different pilot sites need depend on the problems they are working to solve in their scenarios. The ROUTE-TO-PA team are working to collate all available open data to make it available on the platform.

Participants then moved on to identify social and collaborative needs based on the user stories, in order to provide input for the design of SPOD. Social and collaborative needs were commonly specified across pilot sites. Categories of needs here included: dialogue and discussion spaces; moderation and maintenance of these spaces; platform tool capabilities for interaction; varied forms of social media interaction; personalisation of user spaces; and requesting and sharing information. Broadly speaking, participants identified a variety of forms of interaction which could be used over Open Data, and suggested a number of considerations and affordances which would increase the impact and appeal of such social and collaborative platforms.

Participants also used the scenarios provided to design user stories around understandability, usability, and decision-making needs, which will inform the design of the TET. The major categories of understandability, usability and decision-making needs were common across sites. Categories of needs here included: Affordances for the visualisation of complex information; data analysis and reporting tools; decision-making support tools; guidance and usage support tools; affordances for personalising platforms and/or data; and certification tools. Broadly speaking, participants frequently cited the need for data visualisation tools, among others, which would make data more easily understood, whether for personal or professional use.

Subsequent to gathering and integrating scenario-based user needs across all pilot sites, the Route-to-PA design team engaged in an exercise designed to rate the relative impact and feasibility of specified needs. This resulted in the first set of user needs selected for agile software development from M6 to M12 and this process continues iteratively into Year 2 as the design team revisits user needs and ways in which SPOD and TET design features can support those needs.

Societal Activity Model of Open Data User

Next to identifying user requirements for an open data platform from the bottom-up via collective design based workshops, we also identified user requirements from a top down perspective. After all, in order to design and implement a successful ICT platform, "the context" that includes people and their relations (Kuutti 1999) needs to be taken into account as well.

Information technologies should be able to support active users, while dealing with the organizational and societal context (Kuutti 1999). Yet, often this context of broader social forces and structures that influences the interaction between users and information technology, is left unexamined (Engeström 2005).

Therefore, based on democracy, transparency and activity theory, the Societal Activity model of Open Data use (Ruijer et al. 2016) was developed. The model takes three democratic processes as a starting point for the design of open data platforms: monitorial, deliberative and participatory democracy (Meijer 2012).

The Societal Activity Model of Open Data Use (Ruijer et al. 2016) enhances our understanding of user requirements of open data in a societal context. It helps to find the best fit between; on the one hand, the impetus for governmental organizations to provide open data, to increase accountability and transparency, and on the other hand the specific needs of citizen-users in particular domains.

The model was tested in five pilot sites, using interviews, analysis of official documents (where available), and workshops or focus groups where open-data providers and users met and discussed The findings show that different societal processes call for different roles of citizens and government and different user requirements for the design of open data platforms, and, also, provided input for the design of the ROUTE-TO-PA platform.

The ROUTE-TO-PA Platform

Our project will produce software by using open-source licensing model, and the platform will be given to the community of PAs and developers that, after the end of the project, will ensure further development and widespread, sustainable and scalable exploitation of the results achieved. The results of the project (both software and guidelines) will allow PAs to follow the economic and budgetary pressures that

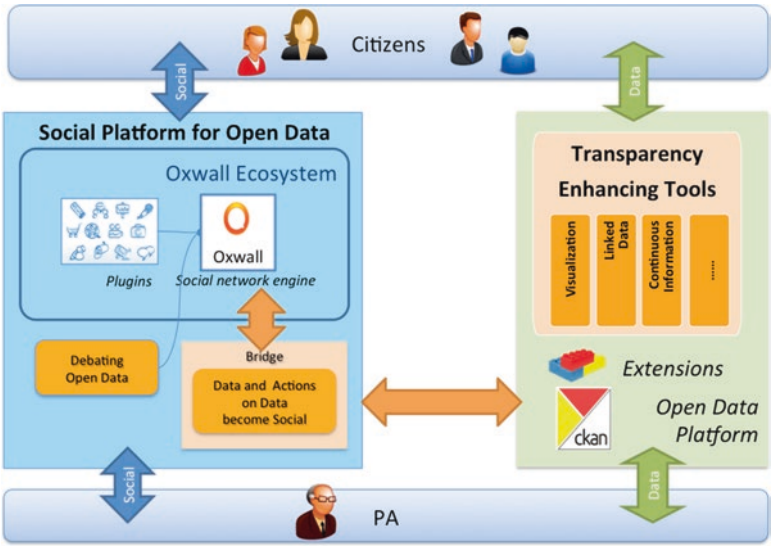

Fig. 2 ROUTE-TO-PA architecture

force administration to be more efficient and to reduce cost (EU DGCONNECT 2013) by adhering to the philosophy of "Doing more with less".

The software and the experiences will be highly reusable, given that they will be piloted in different contexts, in different countries, whose results will be elaborated in guidelines and recommendations at the end of the project. "Reuse rather than reinvent" is one of the guidelines to long-term success of ICTs in Transparency (Bertot et al. 2010) that we adopt as project "mantra" both from technological point of view (i.e., integration of existing popular open source products) and from the PA point of view (plug ROUTE-TO-PA onto existing experiences and needs by involving the Pilots) (Fig. 2).

Social Platform for Open Data (SPOD)

The SPOD platform architecture has multiple decoupled and modular components that communicate together. The architecture is based on mainstream, open source and modular technologies to guarantee interoperability with other external systems. The overall architecture is distributed, as the load of different tasks is taken by different servers (components), both server-side and client-side (e.g., the client-side visualization of data), thereby achieving the important non-functional requirement of Performance Efficiency.

SPOD is a Social Platform for Open data, so its primary requirement is the retrieval of data from Open data Providers. Therefore, SPOD interoperates with

TET, any CKAN based platform, UltraClarity and OpenDataSoft (and additional interoperability with OASIS is planned).

In addition, SPOD can retrieve open data from other existing third party data providers that use restful API. Hence, the interoperability with data provider platforms is based on Web 2.0 mainstream technologies (fulfilling an important non-functional requirement of Interoperability); in this way, SPOD can retrieve the open data to use within the social discussions. For instance, the user can create visualizations from the data available in the open data provider and use them to support its argumentation. SPOD can be configured to allow easy access to associated data providers, so that their datasets can be shown easily used to build visualizations (see Datalets below).

The platform administrator, using the administration pages, can add another data provider and make it available to end-users. In addition, in order to maximize the flexibility, during the creation of a data visualization, any user can copy and paste the data URL from any other external open data Provider as well as post directly the link or content on SPOD (Fig. 3).

The architecture has a ROUTE-TO-PA Authentication Server (RAS), which acts as OpenID Authentication Provider and administration tool to manage users' accounts. In accessing to the ROUTE-TO-PA platforms, users must seamless switch between SPOD and TET features in a user-friendly way. Therefore, a user must access to SPOD and TET, and any other feature federated system, with a unique

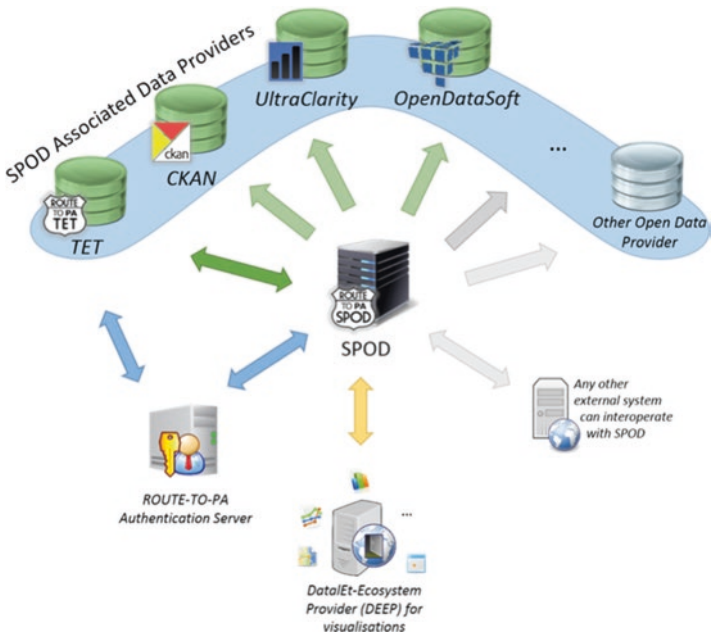

Fig. 3 SPOD architecture

username and password credential. In order to provide this, the architecture has a ROUTE-TO-PA Authentication Server (RAS) based on the OpenID protocol. Any time a user logs in SPOD or TET, his/her browser redirects to the authentication provider log in page. All the authentication server pages have a consistent Graphical User Interface (GUI) with SPOD and TET, indeed they have been specifically designed within the ROUTE-TO-PA project and it is based and compliant with the material design. In this way, TET, SPOD and authentication look and feel is the same, and the switching among their pages is seamless.

The overall architecture has a specific server for the authentication and the entire platform has deployed with a dedicated authentication server. In order to support the authentication through OpenID, SPOD team developer designed a new Oxwall plug-in that supports OpenID. SPOD can be deployed without the activation of the OpenID plug-in so it works without a dedicated authentication server and uses the existing Oxwall registration, log in pages and user accounts management.

The introduction of a ROUTE-TO-PA Authentication Server allows the interoperability of other systems with the ROUTE-TO-PA platforms, following the non-functional requirement of Interoperability. Both SPOD and TET can provide specific services and data to external systems in form of Restful API. Therefore, not only SPOD interoperates with existing data providers, but also itself can provides services to other platforms. Any other external system can authenticate to RAS and interact with SPOD through the restful API services. For instance, based on this architecture, the platform can provide a social widget to embed within any web site to easily share open data, add the content in the own private room or participate in a discussion. In a federated architecture, after the authentication the federated system can invoke a restful API service to perform an action on SPOD (e.g., post of content on SPOD, etc.).

SPOD enables the social collaboration around open data; in particular, it aims to support the collaboration around visualization of open data, allowing their creation, sharing, change and comment. The SPOD architecture provides the visualizations and their services through the DatalEt-Ecosystem Provider (briefly DEEP). DEEP is a repository of visualization web-components to use within SPOD and within any other web site or system that needs to visualise data.

The overall architecture is also useful to envision a “federation” of ROUTE-TO-PA systems. Several federated SPODs and TETs or other institutional systems can access to the ROUTE-TO-PA Authentication Server and share the same ID.

The SPOD architecture is modular and scalable: for instance, the DEEP component can be replicated and distribute around the world to improve performances and serve visualizations to end-users with high availability. Of course, multiple architecture instances can be deployed in different places.

Our distributed architecture, in general, follows the non-functional requirement of Replaceability, as each single component can be substituted with another specific one with the same interface (like a different dataset provider, or another Authentication server with OpenID).

SPOD on Oxwall

Oxwall is a free and Open Source Software (FOSS) social network engine that is able to power customizable online social networks and community-enabled websites. It provides all the basic functionalities of a social network, such as users' friendship, posting text or media comments, handling (private) groups group or event creation/joining. Oxwall architecture is based on plugins, few core plugins handle user/platform/access management. Additional features can be provided by plugins.

SPOD consists of several additional plugins for Oxwall that add to the "standard" social network (friends, status, etc.) the following functionalities:

- It is possible to attach to any social comment, status or answer a *Datalet* (i.e. a re-usable Web Component that provides real-time visualization of datasets, located on any compatible server (see the subsection below). In this way social interactions become enriched with the actual datasets, providing discussions with easy-to-use and easy-to-understand factual evidence. The process to build a datalet is provided by a user-friendly, wizard-like component, that provides the choice of the dataset (among the suggested ones from known providers or from a new one added by providing the link to the RESTful call), the filtering capabilities (choicing columns and rows) and the visualization (choice of the charts, parameters, preview). It is a very important characteristics of Datalets that load and show the actual dataset (1) directly from the source and (2) in real-time, when the user is loading the page, i.e., ensuring authoritative datasets.
- Any user is given the possibility to access a *Personal space*, i.e. a place where he/she can collect and annotate material that can be fruitfully re-used in social discussions later. The user can collect links to webpages, by providing URLs and a datalet is showing the real page in a miniature (that can be also navigated), Datalet for particular visualizations, and plain text notes. All the items can be further annotated, and full text search capabilities allow easy management of the material. The main purpose of the Personal space is to provide a space for *reflection* as the citizen may need time and thoughts to build an argument to be re-used in discussions.
- Discussions occur in *Public Rooms* grouped in an *Agora*. Each Public Room is a traditional threaded chat on the left, with the possibility to add Datalets to the discussion and the possibility to add an opinion (Neutral, Agree, Disagree) to the comment. The right part of the screen is used to provide synthetic information about the discussion, that also makes easier to navigate through long discussions. A graph based representation of the discussions is shown, where nodes are the comments and edges join the answer to a comment, with colors to represent opinions and size of the node representing the number of answers below the comment (see Fig. 6). Navigation is synchronized: clicking on a node on the tree shows the corresponding comment on the threaded chat on the left. Other graphs showing the datalets and the users can be also shown.

Datalets and DEEP

The ROUTE-TO-PA software architecture exploits a modular programming design in order to develop independent software. To achieve this design goal, an architecture completely decoupled from the main project SPOD was designed. The key idea was to realize a repository of components (software services) to be used for different purposes, so that it is possible to enclose some functionalities in a kind of widget and make it available on the Web. This architecture is realized using the Web-Component (WC) standard. We designed a Web service that allows distributing the code of each software component dynamically.

Therefore, each software component is a WC that is, an auto-consistent and independent component that provides some functionalities. A *datalet* is a Web Component that is an output presentation to the user based on the data dynamically loaded from the data source.

The service that allows downloading and using the Datalets is the DatalEt-Ecosystem Provider (DEEP). Within the SPOD software, DEEP architecture is an open, extensible, modular and pluggable service that provides WCs for visualization of open data datasets. DEEP allows sharing, collaboration and creating around customized data visualizations. Further users can create, reuse and share visualizations both in SPOD or in any Web page or other Web-based systems. Its modularity and extensibility fulfils the non-functional requirements of Adaptability and Replaceability.

DEEP is developed as a simple Restful service, providing the list of available datalets (i.e., listing service) and the mapping among the visualization names and their relevant URL within the WC repositories. The system is online and is the base of the architecture of SPOD (http://deep.routetopa.eu/).

Both the DEEP and the WC repository have been designed to be extensible: they can collect all the visualization requests so, as planned future work, they could also provide aggregated statistics on both users preferences and on data and their visualizations. For instance, the most popular datalet visualizations, most used datasets, most popular visualizations for a particular dataset, most visualised fields for a particular dataset, and so on.

The DEEP main task are the listing services which provides a list of available dataset and the mapping between the visualization names and their relevant URL within the “datalets repositories”.

Transparency-Enhancing Toolset (TET)

What Is TET? The TET comprises a set of tools designed to extend available features on popular Open Data Platforms (ODP) to more adequately support transparency related qualities desirable by different categories of ODP end-users. Starting with the Comprehensive Knowledge Archive Network (CKAN) Platform (OKF

2014), the vision for TET is to extend major open ODPs platforms with features enabling easier access to the relevant dataset, a better understanding of these datasets and integration with social platforms for sharing and discussing datasets.

Technical Features TET is implemented as a set of plugins to extend the available features on the well-known CKAN Open Data Platform. The Alpha version of TET described in this report supports the following eight extensions: (1) Support for the use of the WordPress Content Management System as a rich client for CKAN, (2) Enhanced metadata schema to support provenance and alignment with latest W3C guidelines for publishing data on the web, (3) Validation of metadata quality, (4) Linking of related datasets, (5) Enhanced user profiles for personalization and recommendation, (6) Personalised search and dataset recommendation to users, (7) Recommendation of similar dataset to users and (8) Extension of data analytics function on CKAN to support pivot operations on datasets, (9) Interface with the Social Platform for Open Data (SPOD). These features are explained in briefly explained below.

1. Integration of Content Management System for Richer Client Experience – CKAN Integration with content management system enables publishers to publish content related to datasets and publish updates related to portal in an easy way. TET Wordpress plugin seamlessly integrates TET enabled CKAN to provide rich content publishing features to publishers and intuitive interface to end-users.
2. Enhanced Metadata for Improved Context – More metadata fields are added to dataset upload form to enable data publishers to specify richer metadata that will help users in discovery and in getting better understanding of the datasets. The metadata fields are guided by and comply with aligned with the "W3C guidelines for publishing data on the web". Additional fields supported include: Basic details related to the dataset; Target audience of the dataset; Theme/Category; Versioning; Provenance; Geospatial Coverage; and Temporal Coverage.
3. Metadata Quality Check and Validation – Additional validations to dataset entry form are added to prevent data entry errors and to ensure consistency. Quality check indicators guide publishers about the quality of metadata being entered. The features will also help end-users in assessing the metadata quality of dataset.
4. Relating Datasets – Dataset linking feature allow users to specify explicit links between datasets, which can be exploited for recommendations and data integration purposes.
5. Enhanced User Profiles for Personalization and Recommendation – Default CKAN user registration page is modified to allow more details related to the user to be captured, the feature plays essential role in creating a personalized user experience for the end user.
6. Personalization Search and Dataset Recommendation for users – enables users to search for datasets based on their profile or based on the desired category.

Users can select appropriate profile from the list of profiles provided or could select the category they are interested in from the list (see Fig. 7).

7. Recommendation for related datasets – enables recommendation of more datasets based on user group and dataset category selected in the user profile in addition to other contextual information. The feature guides users to find potentially useful and relevant datasets.
8. Extension of Available of Data Analytics Functions – CKAN platform lacks data analysis capabilities essential for working with data. To overcome this limitation as the first step, we added PivotTable feature which allows users to view, summarize and visualize data.
9. Enhanced Interface with the SPOD – builds on the CKAN APIs to enable the SPOD platform access datasets managed on the CKAN with the enhanced features for visualisation, sharing and discussion.

The above features resulted from analysis of the information on the barriers to open data use and needs of end-users gathered from the series of Collective Intelligence Sessions hosted by pilot partners in respective partner countries. In addition, transparency qualities including Accessibility, Usability, Understandability, Informativeness and Auditability described in Cappelli et al. (2013) and other transparency constructs such as those in Fung (2013) underpinned the development of the above features.

TET "Plug-and-Play" Architecture

In implementing the above features, the base CKAN platform were extended with a number of additional components implemented as "plug-ins". The architectural decision to implement TET as plug-ins is to enable easy coupling and removal of TET components and consequently minimal operational impact to the based open data platforms. This architectural style allows us to experiment easily with alternative design and implementations of the feature (Wang and Avrunin 2008). The TET components are grouped into three categories: Frontend, Data Platform and Analytics components as shown in Fig. 4.

In addition, an additional element (or plugin) enables integration with the social platform. The development of the above features was carried out through an Agile Software Development process which enabled the development an early prototype and subsequent short "develop-test-release" cycles to engage potential end-users of TET platform.

Status and Exploitation The above features have been successfully implemented as part of the Alpha release available at http://srvgal100.deri.ie:8081/. Some of the functionalities have already been deployed as part of CKAN instance managed by the Dublin City Council, Republic of Ireland.

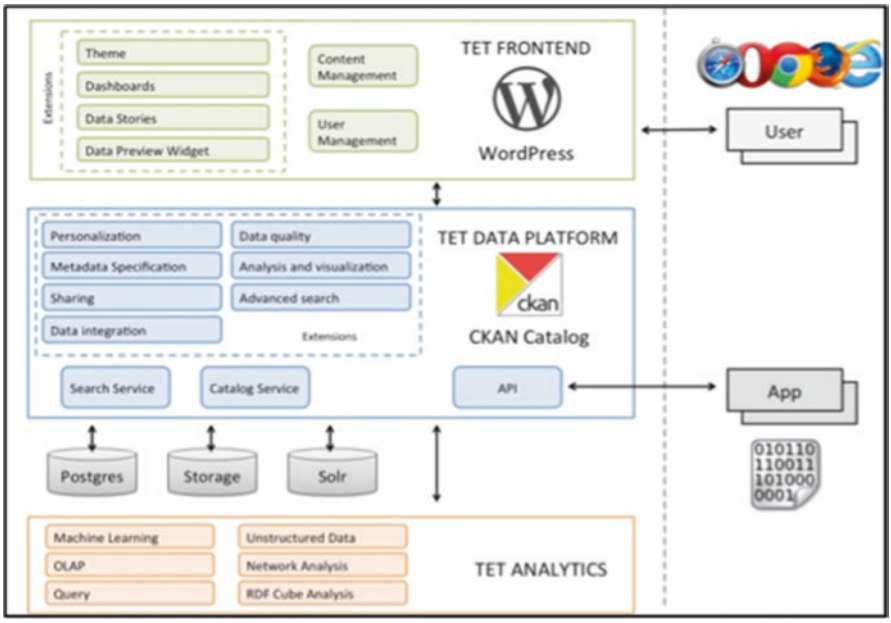

Fig. 4 TET architecture

Evaluation Plan and Pilots Description

Description of the Five Pilot Sites

The City of Prato (Italy) scenario case concerns *city budget management*. The goal is for citizens to monitor the allocation of the city budget by municipality and to propose expenditure priorities and suggestions. The city would like to increase transparency on budget management and possibly collect some of citizens' expenditure suggestions. It is important to know that this process is intended to be an improved version of existing democratic procedures of involving citizens in budget decisions.

The object or policy issue in **Groningen (NL)** focuses on population decline. One of the areas with the highest population decline in the Netherlands is situated in Groningen. The potential community of stakeholders is diverse consisting of citizens, public organizations (schools, health care organizations etc.) and private companies. Open data as an instrument can provide insight in the consequences of and might be able to contribute to innovative and collaborative solutions for population decline.

The **Dublin, Ireland** City Council is one of the founding partners of Dublinked. Dublinked is an ideas and information sharing network which connects the Dublin

region's four local authorities with universities, companies and entrepreneurs. The scenario chosen for ROUTE-TO-PA evolves around capacity building, which focuses on increasing citizen engagement in a deliberative process with their city or more specifically their community. The scenario thus focuses on building community awareness with the ultimate outcome to "make my city great".

Issy- les- Moulineaux (France) is a city located near Paris. The city hosts many IT companies and welcomes start-up companies in the field of new technologies. The scenario focuses on the activity of two central groups: young entrepreneurs in ICT domain who form a major part of the local economic base and public administrators from Paris Region who collaborate with Issy-les-Moulineaux in the global open data policy.

The pilot in **The Hague (NL)** focuses on collaboration between public administrators and employers and can be characterized as a participatory process. Employers and the City of The Hague have a longer history of collaboration and meeting, the relationship between the Department and the local employers is quite good. The specific scenario or policy issue suitable for exploiting Open Data will be jointly developed, whereby the focus is on finding solutions for existing problems together. This is called co-creation.

Research Approach: Four Levels of Evaluation

Concerning evaluation of developments and outcomes at the five user sites, in terms of transparency, we have several sources of criteria. The identification of users' needs (section "Requirements and User Involvement"), are translated into design specifications for the tool, that is, the actions users are supposed to perform with the tool leads to a set of evaluation criteria at the technology and user levels. The modelling activity (section "Societal Activity Model of Open Data User") has provided abstract models at the society and community-levels. On the basis of these models we will be able to characterise and compare all cases with respect to their current and future states at four different levels:

1. The technology functioning according to design specifications, tested by usability studies and user consultation.
2. The individual user carrying out different actions: we shall develop a framework based on the well-known and studied Technology Acceptance Model (TAM) (more details in the next subsection).
3. The community, or small group working together to generate new ideas: the main dimensions of evaluation on this level can be formulated in terms of the characteristics of the participants, the structure of their interactions and their content within the OECoP, studied from a developmental perspective (Engeström 1987). The developmental, or diachronic, analysis draws on synchronic analysis of the OE-CoP (at a given point in time), and identifies relevant differences across time-points, with a view to identifying the overall trajectory (Dreier 1999; Ludvigsen et al. 2011) of the community, i.e. where it is heading.

4. The society, or the organisational context, adapting to the new possibilities for creating transparency. We will investigate the impact of our project within the organisation by observations, interviews and surveys.

Our evaluation activity also relates closely to the Models and Methods developed:

1. The Societal Model (year 1) relates to the above in terms of analysis of the object of activity (including community rules), and the expression of tensions. In these cases, the level of content analysis is relevant, for identifying 'what the participants are trying to achieve' (i.e. object of activity) in given exchanges, and, on the level of communicative functions, given that tensions will be correspond to argumentative functions and, most likely, to the salient expression of strong emotions.
2. The Community Model (year 2) provides the theoretical background and methodological tools for addressing the main research question, concerning evolution of SPOD-TET mediated collective activity towards a veritable epistemic community.
3. The Social Representations Model (year 3) concerns the evolutions of the OE-CoP participants' representations (attitudes, appraisals) of the community, of their perceived self-efficacy, of the degree of transparency and cooperativity of the Public Administration, and of the SPOD-TET tools themselves. This will be studied using interviews and questionnaires, but also on the basis of content analysis (what are the attitudes expressed using SPOD?). This can draw on appraisal theory (White 2002), i.e. the positive and negative attitudes and affects expressed in language.

The User Level

To evaluate the artifacts, we shall develop a framework based on the well-known and studied Technology Acceptance Model (TAM). The model was published by Davis (1989) and is the most widely accepted model for understanding the usage of Information Systems (IS) and its acceptance. It suggests, that external variables (such as system design and rich features) influence the perceived usefulness and perceived ease of use. Over time the model evolved to TAM2 (Venkatesh and Davis 2000) and was extended with additional external variables, relevant to IS utilised in the workplace: the social influence variables (i.e. subjective norm) and the cognitive instrumental variables (i.e. result demonstrability). Original TAM is presented in Fig. 5.

The definitions of the additional variables in TAM2 are defined as follow:

1. Voluntariness – "the extent to which potential adopters perceive the adoption decision to be non-mandatory".
2. Subjective norm – "a person's perception that most people who are important to him think he should or should not perform the behavior in question"

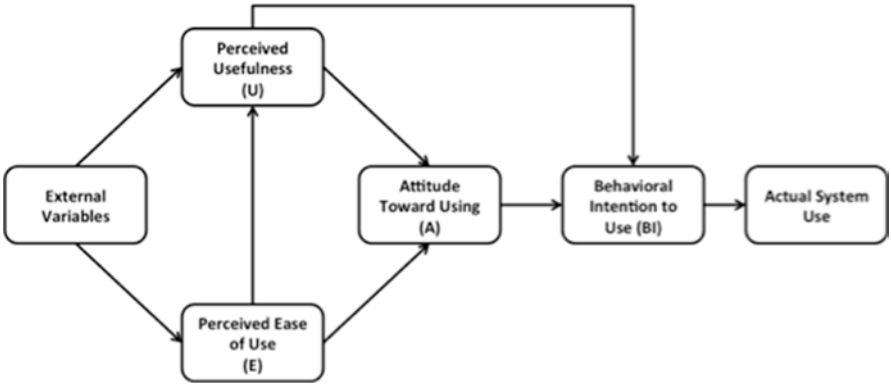

Fig. 5 Technology Acceptance Model (TAM)

3. Image – "the degree to which use of an innovation is perceived to enhance one's status in one's social system"
4. Job relevance – "an individual's perception regarding the degree to which the target system is applicable to his or her job"
5. Output quality – "the tasks a system is capable of performing and the degree to which those tasks match the job goals"
6. Result demonstrability – "the tangibility of the results of using the innovation"

Based on the Technology Acceptance Model in the pilot evaluation, we evaluate the following parameters:

1. Relevance – how relevant is the use of open data and the TET/SPOD to the user's job and daily life,
2. Output quality – what is the quality of datasets available on the platform? What is the perceived quality of TET and SPOD platforms?
3. Result Demonstrability – does the use of the TET and SPOD address the relevant user needs provided by users during the Scenario development workshop?
4. Perceived Ease of Use – how easy is it for non-technical users to use TET and SPOD?
5. Perceived Usefulness – how useful do users perceive TET and SPOD?
6. Intention to use – how wiling are the users to use TET and SPOD to support their information needs and decision making needs?

Conclusions and Future Actions

The team has just released the first Alpha prototypes of SPOD and TET as the project first year just finished (January 2016). SPOD and TET will be tested (in a first round) in the five pilots for the year 2016, starting late September. In 2017 a new version, beta, will be submitted to another round of testing from the pilots, leading to the release of a highly tested, jointly designed and citizen-centered software.

In this paper we have described a holistic, multi-disciplinary approach that starting from collective intelligence and scenario-based design approach, produced first scenarios and then user stories that were used to feed the initial technological design. At the same time, the activities on modeling, with the Societal Activity model of open data user provided contextual information, that further motivated the design. The design produced the first prototypes of a Social Platform for Open Data, for fostering participation, and Transparency-Enhancing Toolset, for improving transparency for citizens. The prototypes are going to be tested in September 2016 on five pilot sites that provide diversity and heterogeneity in our evaluation.

It must be stressed that our holistic multidisciplinary approach employs methodologies that, although coming from different fields, share a common vision of continuous user-centered design, from the collective intelligence approach, to the activity model to the technological employment of agile methodologies. We believe that our approach, that involves stakeholders since the very beginning of the design and development will be one of the key factors to the success of our project.

Of course, we are well aware that our research is strongly based on only five pilots in Europe, that, although heterogeneous in countries, size, state of the open-data programs, and scenarios, do offer a limited view of the overall scenario in the whole world with very diverse needs and contexts. Nevertheless, we are confident that, as the limitations of our work are evident to us as researchers, that methodologies and the technologies that have been designed and developed in order to be of wide impact, could be partly re-used and fruitfully employed to tackle, at least partially, the engagement of citizens through Open Data in diverse contexts.

The project will finish its activities on January 2018, with two rounds of pilot studies (February 2016 and February 2017) of SPOD and TET in the five pilots (see http://www.routetopa.eu for updates on the status of the activities) (Figs. 6 and 7).

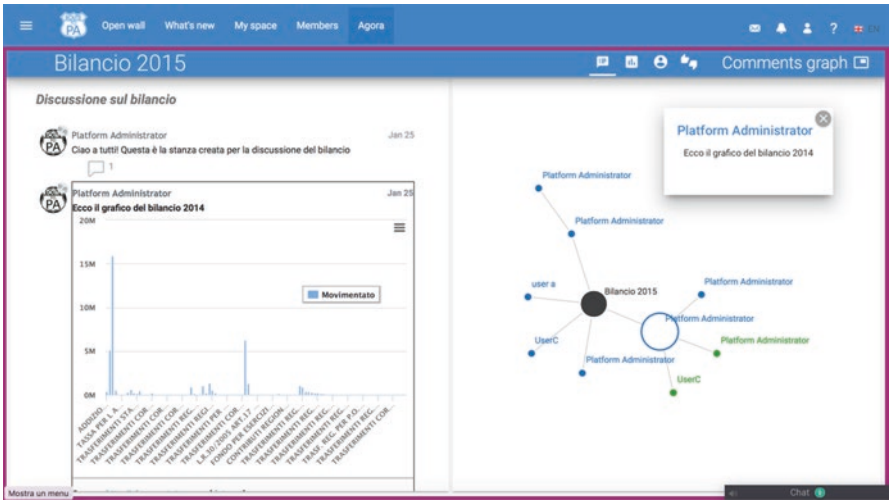

Fig. 6 SPOD example screenshot

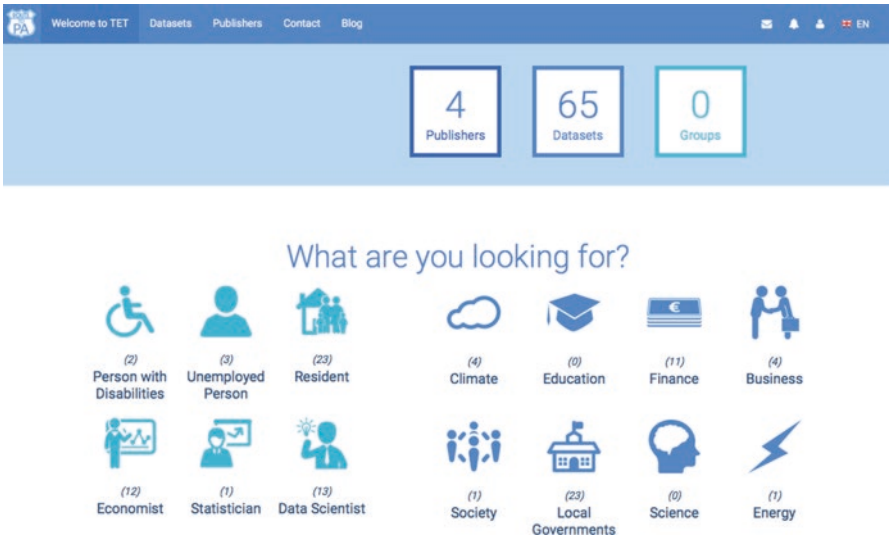

Fig. 7 TET example screenshot

Acknowledgments This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 645860.

References

Bertot J, Jaeger P, Grimes JM (2010) Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. Government Information Quarterly *27*:264–271

Bonsón E, Torres L, Royo S, Flores F (2012) Local e-government 2.0: social media and corporate transparency in municipalities. Government Information Quarterly 29(2):123–132

Cappelli C, Engiel P, Araujo RMD, Cesar J, Leite P (2013) Managing transparency guided by a maturity model. In: 3rd global conference on transparency research, HEC, Paris, 24–26 October 2013, pp 1–17

Cohn M (2004) User stories applied for agile software development. *Addison-Wesley*, Boston

Colpaert P, Sarah J, Peter M, Mannens E, Van de Walle R (2013). The 5 stars of open data portals. In: 7th international conference on methodologies, technologies and tools enabling e-Government (MeTTeG), University of Vigo, Spain, pp 61–67

Davies T (2012) Supporting open data use through active engagement. In: Proceedings of W3C workshop "using open data: policy modeling, citizen empowerment, data journalism", 19–20 June 2012. http://www.w3.org/2012/06/pmod/pmod2012_submission_5.pdf

Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13:313–339

Dreier O (1999) Personal trajectories of participation across contexts of social practice. Outlines: Critical Social Studies 1:5–32

Engeström Y (1987) Learning by expanding: an activity theoretic approach to developmental research. Orienta Konsu, Helsinki

Engeström Y (2005) Developmental work research: expanding activity theory in practice. Lehmanns Media, Berlin
EU DG-CONNECT (2013). A vision for public services. European Commission, "Public Services" unit of Directorate-General for Communications Networks, Content and Technology (DG-CONNECT). 13 June 2013
Fung A (2013) Infotopia: unleashing the democratic power of transparency. Polit Soc 41:183–212
Kuutti, K. (1999) Activity theory, transformation of work, and information systems design. In: Engestrom, Y.; Miettinen, R. & Punamaki, R.-L. (eds) Perspectives on activity theory. Cambridge University Press, Cambridge, 360-376
Ludvigsen S, Rasmussen I, Ingeborg K, Moen A, Middleton D (2011) Learning across sites: new tools, infrastructures and practices, Intersecting trajectories of participation: temporality and learning. Routledge, London, pp 105–121
Meijer AJ (2012) The do it yourself state. Inf Polity., IOS Press *17*:303–314
Michener G, Bersch K (2011) Conceptualizing the quality of transparency. In 1st global conference on transparency. Political concepts, Committee on Concepts and Methods Working Paper No. 49
Mishory E. N (2013). Clarifying transparency: transparency relationships in government procurement. In Government Procurement Seminar, Chris Yukins & David A. Drabkin, 4 November 2013
Open Knowledge Foundation (OKF) (2014). Comprehensive Knowledge Archive Network (CKAN) Developer Documentation
Rosson M, Carroll J (2002) Scenario-based design. In: The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. Lawrence Erlbaum Associates, Mahwah
Ruijer E, Grimmelikhuijsen S, Enzerink S, Meijer AJ (2016) The societal activity model of open data use. Deliverable 3.1
Venkatesh V, Davis FD (2000) A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 46:186–204
Wang S, Avrunin G (2008) Plug-and-play architectural design and verification. In: Architecting dependable systems V. Springer, Berlin/Heidelberg
Warfield J (2006) An introduction to systems science. World Scientific, Hackensack
White PRR (2002) Appraisal—the language of evaluation and stance. In: Verschueren J, Östman J-O, Blommaert J, Bulcaen C (eds) The handbook of pragmatics. John Benjamins, Philadelphia, pp 1–23

Vittorio Scarano is Computer Science Associated Professor at the Dipartimento di Informatica of Università di Salerno. He is coauthor of more than 150 papers in journals and international conferences (with reviews) at IEEE, ACM etc. and was advisor (or co-advisor) of 12 PhD thesis in Computer Science at the University of Salerno. His research is mainly focussed on open data, distributed, parallel and cooperative systems, and to their applications in the fields of collaborative learning and virtual interactive environments (Cultural Heritage).

Delfina Malandrino is an Assistant Professor at the Dipartimento di Informatica of the University of Salerno. Her research activities mainly focus on: privacy, distributed systems, collaborative systems, social networking, privacy, social network analysis, green computing.

Michael Baker is a Research Director (tenured Research Professor) in psychology and language sciences, of the Centre National de la Recherche Scientifique (CNRS). He currently works in the Social and Economics Science Department of Telecom ParisTech, the French National Telecommunications Engineering School. His research aims to analyse and model the processes of knowledge elaboration in dialogues produced in learning and work situations, focussing on argumentation and the role of technological mediation. Michael Baker recently published the following book, with Baruch Schwarz: "Dialogue, Argumentation and Education: History, Theory and Practice" (Cambridge University Press, New York, 2017).

Françoise Détienne is research director at CNRS, Paris. Her research in cognitive ergonomics focusses on collaboration, design, creativity and online communities as well as on the role of socio-technical systems. She is associate editor of International Journal of Human Computer Studies, and member of the editorial board of Interacting With Computers and CoDesign.

Jerry Andriessen is a senior scientist and research director of Wise & Munro learning Research, The Hague, Netherlands. Wise & Munro participates in many innovative European projects, on technology, open data, collaboration, primary education, and cybersecurity. These projects add societal relevance to their scientific implications. Jerry has more than 25 years of experience in uses of technology for support of learning and collaboration in various settings. His current interest is in the qualitative interpretation of discourse and action, including the roles of emotions.

Mirjam Pardijs is an independent educational scientist. She is co-director of Wise & Munro Learning Research, the Hague, Netherlands. She specialises in the design, coaching, and interpretation of collaborative learning in educational and professional contexts. Her main interests are narratives as a tool for learning, appropriation of technology, and the role of pedagogical support for innovation in learning settings.

Dr. Adegboyega Ojo is Senior Research Fellow at the Insight Centre for Data Analytics, National University of Ireland Galway (NUIG). He leads the E-Government Group at Insight Centre, serves as Adjunct Lecturer at the College of Engineering and Informatics. His current research interests include data driven innovations in government, Open data policies and Infrastructures, data analytics and governance of smart cities. He is member of the Editorial Boards of the Government Information Quarterly and International Journal of Public Administration in the Digital Age.

Michael Hogan is a senior lecturer and researcher at NUI, Galway, Ireland. His research foci include individual, social and technology factors contributing to adult learning, motivation, and collaborative performance. He is a co-director of the Structured PhD in Perception, Cognition and Action, co-director of the Structured PhD in Learning Sciences, and a co-leader of the Health and Wellbeing theme at the Whitaker Institute for Innovation and Social Change, NUI, Galway.

Albert Meijer is a Professor of Public Innovation at the Utrecht School of Governance in the Netherlands. His research focusses on e-government, government transparency, co-production in the public sector and public innovation.

Erna Ruijer is a postdoctoral researcher at the Utrecht School of Governance in the Netherlands. Her research focuses on open data, government transparency, open government and government communication.