A taxonomy of open government data research areas and topics

Yannis Charalabidis, Charalampos Alexopoulos & Euripidis Loukis

Taylor & Francis
Taylor & Francis Group

A taxonomy of open government data research areas and topics

Yannis Charalabidis, Charalampos Alexopoulos ⬡, and Euripidis Loukis

Department of Information and Communications Systems Engineering, University of the Aegean, Athens, Greece

ABSTRACT

The opening of government data, in order to have both social and economic value generated from them, has attracted the attention and interest of both researchers and practitioners from various disciplines, such as information systems, management sciences, political and social sciences, and law. Despite the rapid growth of this multidisciplinary research domain, which has led to the emergence and continuous evolution of technologies and management approaches for open government data (OGD), a detailed analysis of the specific areas and topics of this research is still missing. In this article, a detailed taxonomy of research areas and corresponding research topics of the OGD domain is presented: it includes four main research areas (ODG management and policies, infrastructures, interoperability and usage and value), which are further analyzed into 35 research topics. An important advantage of this taxonomy, beyond its high level of detail, is that it has been developed through extraction and a combination of relevant knowledge from three different sources: important relevant government policy documents, research literature, and experts. For each of the 35 research topics we have identified, its research literature is summarized and main research objectives and directions are highlighted. Based on the taxonomy, an extension of the extant OGD lifecycle is advanced; also, under-researched topics that require further research are identified.

1. Introduction

Open government data (OGD) has been attracting the growing attention and interest of both researchers and practitioners from various disciplines, such as information systems, management sciences, political and social sciences, and law, due to its widely recognized potential to generate public value through driving innovation and economic growth as well as also scientific research, and by promoting transparency and substantial evidence-based political dialogue (Stevens 1984; Conradie and Choenni 2012; Janssen 2011b). The concept of open data is strongly associated with innovative capacity and transformative power (Davies, Perini, and Alonso 2013). It is increasingly recognized that proactively opening public data can create considerable benefits for several stakeholders, such as firms and individuals interested in the development of value-added e-services or mobile applications, by combining various types of OGD, and possibly other private data, or scientists, journalists, and active citizens who want to understand better various public problems and policies through advanced data processing and production of analytics (Zuiderwijk et al. 2014a; Janssen 2011a). However, it should be noted that at the same time there are a few articles discussing unintended consequences and negative side effects of opening data (Blakemore and Craglia 2006;

Zuiderwijk and Janssen 2014b). At the same time, as mentioned in the Call for Papers of this Special Issue, "Yet organizations are struggling to generate value from big and open data."

Furthermore, OGD, as a new organizational invention gradually diffusing in government is under a continuous renegotiation over its meanings and practices, and therefore a gradual formulation of its "organizing vision" (using the term proposed by Swanson and Ramiller 1997). According to Tammisto and Lindman (2012), the first level of renegotiation in the context of OGD took place initially in relevant policy discussions, public and professional press, and consultancy. The second level of renegotiation is taking place when organizations gradually understand how to benefit from open data and drive the development of social and economic value from it. This renegotiation and evolution of this new domain can be greatly assisted by establishing a common code of understanding concerning the main areas and topics of research on OGD. The development of a detailed taxonomy of current research areas and topics in the domain of OGD will address the communication gap in this new domain, and facilitate better interaction among researchers as well as with interested practitioners. Also, it can provide a solid base for future research in this domain, and thus contribute to reaching higher levels of maturity in the practices of opening and exploiting government data, and in the generation of social and economic value from them; in general, it can contribute to the development of a body of knowledge in this area, which will enable improving and optimizing the technology, and also the design, operations, and performance of the units of government agencies responsible for opening data. Such a taxonomy is of critical importance for the development of a "science base" (see Charalabidis, Gonçalves, and Popplewell (2011)) in the OGD domain.

Furthermore, it can be useful for information and communication technologies (ICT) firms (assisting them in developing better OGD technological infrastructures), government agencies (for improving their OGD practices), and firms developing innovative value added e-services or mobile applications based on OGD (Statistical Office of the European Communities 2005). However, despite the rapid growth of this multidisciplinary research domain, which has led to the emergence and continuous evolution of technologies and management approaches for open government data (OGD), a detailed analysis of the specific areas and topics of this research is still missing (see Section 2 for more details).

This article contributes to filling the research gaps. In particular, it makes the following contributions:

(1) It develops a detailed taxonomy of research areas and corresponding research topics of the OGD domain has been developed, including four main research areas, which are further analyzed into 35 research topics.
(2) This taxonomy has been developed through extraction and combination of relevant knowledge from three different sources: important relevant government policy documents, research literature, and experts.
(3) For each of these 35 research topics we identified, its research literature has been summarized and the main research objectives and directions have been highlighted; also, under-researched topics that require further research have been identified.
(4) Based on the above taxonomy an extension of the existing in the literature OGD lifecycle has been proposed (including important additional stages).
(5) Our OGD research taxonomy extends and elaborates previous research taxonomies for the "ICT-enabled Governance" and "Policy Making 2.0" domains, which have been developed in the European projects CROSSROAD and CROSSOVER.
(6) Finally, directions have been formulated for future multi-disciplinary research based on OGD aiming to address current societal challenges.

This article is structured as follows. Section 2 describes the methodology we followed for developing the taxonomy. In section 3 the main findings of literature review we have conducted for this purpose are presented and discussed. Then section 4 presents the taxonomy, including descriptions of the identified main research areas, and the particular research subareas/topics for each of them. Finally, a discussion of findings is provided in section 5, while section 6 concludes the article.

2. Methodology

This study is focused on two main research questions, which constitute a first step toward the creation of a "descriptive theory" of the OGD domain that will enable the development of a science base of it: (a) what are the main research areas and topics of the OGD domain, and (b) how they can be categorized? Gregor (2002) proposed five types of theories that need to be developed in the information systems domain; the first and more fundamental of them, which is necessary for the development of the other four more advanced ones, is the "descriptive theories," which "describe or classify specific dimensions or characteristics of individuals, groups, situations, or events." There are two categories of descriptive theories: naming theories and classification theories (Stevens 1984). A naming theory is a description of the main dimensions or characteristics of some phenomenon. A classification theory is more elaborate in that it also includes interrelations between such dimensions or characteristics of given phenomena.

This article contributes to the development of description theory for the OGD domain, both a naming and classification theory, which are of critical importance for the development of more advanced types of theories in this domain (e.g., concerning relationships between various dimensions or characteristics of them), and in general for the development of its scientific base. In particular, we developed an OGD research areas taxonomy (OGDRAT), based on relevant government policy documents, previous relevant research literature, and experts' knowledge. For this purpose we followed the bottom-up approach to taxonomy development proposed by Ramos and Rasmus (2003) and Sujatha and Rao (2011), which includes the four stages shown in Figure 1 (our research has focused on the first three of them). The scope of this study focuses on the first three steps of this approach, as it is illustrated with the dashed box in Figure 1.

In particular, the methodology we followed for the development of the OGDRAT was based on content analysis (Krippendorff 2012) of different documents (government policy documents, previous relevant research literature, and minutes of experts' workshops). It consisted of the following eight steps (shown also in Figure 2):

(1) Initially we identified and analyzed important relevant government policy documents concerning OGD, which define the main terms, issues, and perspectives, and also the main problems and challenges posed in this domain. The most important of them were (a) European Commission Directives and Communications (European Commission 2011, 2012, 2013a, 2013b), (b) US Government documents (Executive Office of the President 2009; Obama 2012), (c) UK Government documents (O'Hara 2011; UK Cabinet Office 2011; HM Government 2012), and (d) Horizon 2020 Information and Communication Technologies Work Programme (European Commission 2014). The outcome of this step was a first set of ODG related terms, which were used for constructing the first version of OGDRAT in step 3.

(2) Then we identified and analyzed previous research papers that propose categorizations of research areas and perspectives of the OGD research domain. Additionally, we identified and analyzed previous research literature concerning barriers to OGD publishing and
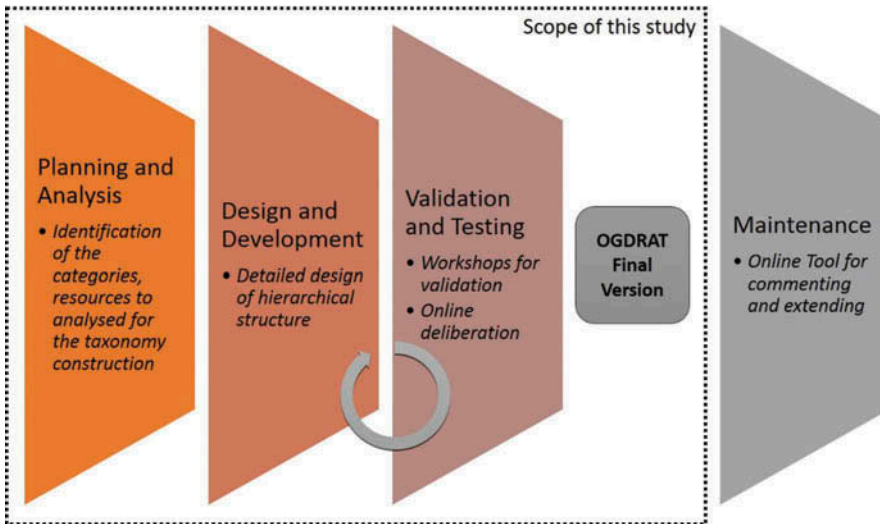
Figure 1. Taxonomy development approach.

exploitation, and also uptake of OGD and value generation from them. A brief review of this literature is presented in section 3 (while we refrain from presenting a review of the relevant government policy documents on OGD identified and analyzed in step 1, in order to keep the paper in acceptable level). The outcome of this step was another set of ODG related terms (having some overlapping with the ones of the set produced in the previous step), which were used as well for the construction of the first version of OGDRAT in step 3.

(3) After realizing the above first two steps, the main research topics in the OGD domain were defined, and then were grouped in higher level research areas; this was a first version of the OGDRAT.

(4) A thorough literature search was then conducted, based on the E-Government Reference Library (EGRL—faculty.washington.edu/jscholl/egrl/), which is a widely recognized and frequently updated electronic library of peer-reviewed papers in the electronic government/governance domain, using as keywords the terms of the first version of the OGDRAT. In particular, the EGRL was searched by paper title and abstract for each of these terms, and the most relevant papers were retained and read in detail. This led to the identification of additional research topics in the OGD domain, which were used for the construction of a second version of the OGDRAT.

(5) The realization of the fourth step resulted in the second version of the OGDRAT.

(6) A workshop was organized for the discussion, evaluation, and validation of the second version of the OGRAT, aiming at the assessment of its main research topics, and the possible proposition of new ones, and also at the assessment of their grouping, and the possible proposition of changes. In this workshop participated 20 OGD experts from 11 different EU countries (NL, UK, DE, GR, BE, IT, AU, RO, ES, BG, LV), from different organizations (public administrations, universities, and firms) and different educational levels (Professors, PhD and MSc holders), in order to validate and further elaborate second version of the OGRAT. All of the participants were selected based on their experience in the area of OGD and they are characterized as very experienced in the OGD domain, having been or currently being involved in OGD related projects (national or European).

(7) Based on feedback collected from this workshop, which included the proposition of new research topics, such as the topics 2.7 ("citizen-generated open data") and 2.8 ("sensor-generated open data") described in section 4, and also of changes in their grouping in
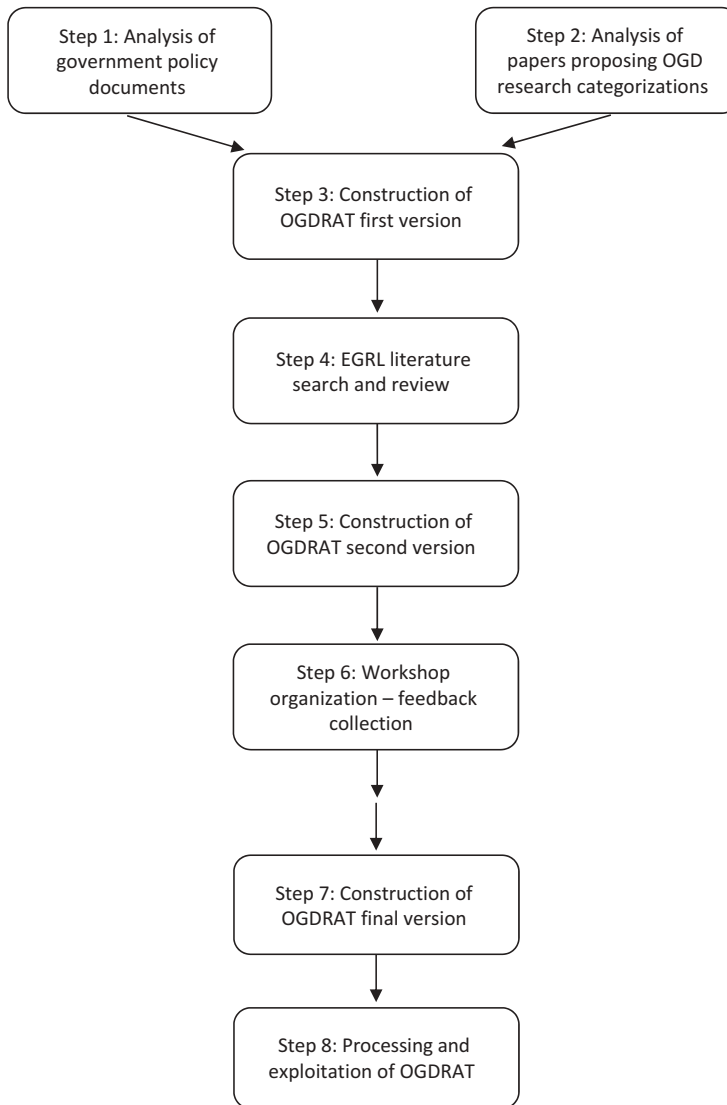
Figure 2. Steps of OGDRAT development methodology.

research areas, the final version of the OGRAT was produced, which is presented in the section 4.

(8) Finally we proceeded to further processing and exploitation of it, and the results are presented in section 5.

3. Literature review

In the step 2 of our methodology (described in the previous section) we identified four previous research papers that propose categorizations of OGD research in areas and themes (Davies, Perini, and Alonso 2013; Zuiderwijk et al. 2014a; Lindman, Rossi, and Tuunainen 2014; Harrison, Pardo, and Cook 2012), which were reviewed because they include elements that can be useful for the development of the OGDRAT.

Davies and colleagues (2013, 11) argued that "over its short history as a field of action a number of distinct fronts of research into open data have developed, responding to different practice, policy and knowledge needs. These can be usefully classified into three broad groups: 1) open data readiness assessments, 2) open data implementation studies and 3) impact studies." Readiness studies aim to assess whether the conditions in public administrations are appropriate for the effective development of open data initiatives. Implementation studies aim to assess whether the conditions for open data itself actually exist in terms of open data availability, extent of publishing government agencies, and importance of published datasets. Finally, impact studies aim to assess to what extent open data initiatives have led to change and public value.

The second study by Zuiderwijk and associates (2014a, 2) identified seven perspectives of OGD research, namely, political, social, economic, institutional, operational, legal, and technical and argued that "combining perspectives may be more effective in dealing with the issues related to open data and stimulating innovation." Further, it also identifies a number of OGD research directions, and categorizes them under three major topics: (1) open data theory and development; (2) open data policies, use, and innovation; and (3) open data infrastructures and technologies.

Another study conducted by Lindman and coauthors (2014, 4) focused on the research challenges concerning open data services, and categorized the relevant issues based on the work systems framework (Alter 2010). The authors argued that "there are two basic approaches for organizing the research issues according to the challenges that emerge when data is made available to the public, and further provided as services. These are: 1) an analysis of the life-cycle of the data and 2) an analysis of the levels of inquiry at which the open data phenomenon is studied." The proposed categories for the organization of open data services research are: technologies, information, processes and activities, products and services, participants, customers, and environment; each of them includes several research questions.

Finally, Harrison and colleagues (2012, 23) examined the Open Government "ecosystem," concluding that OGD emerges as an essential dimension of the open government concept, arguing that "the importance of developing the social and material infrastructures for creating, managing, and sharing data in the short term, along with the governance structures through which innovative architectures, infrastructures, and standards will be negotiated for the future." Then they defined the main themes of the research required in order to realize this vision, along with the workflow of defining data of interest, prioritizing data collection, conducting data collection, publishing the data, and then using them and generating value.

Furthermore, there is another research stream dealing with the barriers to OGD publishing and exploitation (Conradie and Choenni 2012; Janssen 2011a; Janssen et al. 2012; McDermott 2010; Barry and Bannister 2013). We reviewed this research stream, because the main findings of it (e.g., identified barriers) might correspond to important research topics (e.g., concerning new ways of overcoming these barriers), so they can be useful for the development of OGDRAT. Finally, for the same reason we also reviewed another research stream dealing with the uptake and use of OGD, and their exploitation for innovation and value generation (Bason 2010; Borins 2001; Hartley 2005; Kundra 2012; Mohr 1969; Windrum and Koch 2008; Yang and Kankanhalli 2013). The main conclusions of this stream of research indicate that the uptake and use of the OGD, and also the generation of innovation and value in general from them, are not straightforward, being complex, and requiring the collaboration of several actors.

From the literature review it has been concluded that although there are some previous studies that propose categorizations of OGD research of OGD in areas and themes, these are at a too high level, and lack the detail required for directing future research, for facilitating a better interaction among researchers, and also with interested practitioners, and in general for providing the development of a "science base" in this domain. Our research, as mentioned in the introduction, contributes to filling this gap.
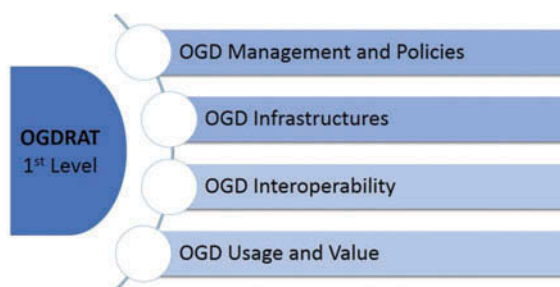
Figure 3. First level of OGDRAT (Research Areas).

4. The OGD research areas' taxonomy

The final version of the OGDRAT we developed (the outcome of step 7 of our methodology described in section 2) consists of four major research areas (in its first level): OGD Management and Policies, OGD Infrastructures, OGD Interoperability, and OGD Usage and Value (shown in Figure 3), which includes 35 research topics (in the second level). These 35 identified research topics were initially divided into two categories: the technological and nontechnological ones; the latter correspond to the OGD Usage and Value research area.

By examining the former we distinguished two clear subgroups of research topics, concerning the interoperability and the management of the OGD, respectively, which led to the definition of the OGD Interoperability and the OGD Management and Policies areas; the remaining technological factors concerned the OGD infrastructures, so they were grouped in a separate research area. This grouping of the identified research topics into the four research areas has been confirmed by the experts who participated in the workshop mentioned in section 2; however, research area changes were proposed for some research topics. The OGDRAT has been constructed in an online tool "mind42.com" and is available for commenting.[1]

4.1. *OGD management and policies*

The first research area of the OGDRAT has been named "Open Government Data Management and Policies." Data and Information Management is an important research topic in the broader information systems domain, from which concepts, theories, and frameworks can be borrowed and elaborated for further analysis and investigation of OGD management challenges.

Policy issues are closely related to the data management, in a broader definition, since policy decisions create the context of OGD management, so it affects data management procedures. Data management is a challenge both for OGD providers (public organizations) and for OGD users (e.g., scientists, analysts, journalists, active citizens). Therefore this research area includes several research topics corresponding to important OGD management challenges (such as methods for OGD anonymization, cleansing, visualization, linking, publishing, mining, and also quality assessment). It is worth mentioning that within the workshop there were comments on whether we should put some of the research topics, such as OGD linking and mining in the category of infrastructures, since they are supported and provided by the developed infrastructures.

Finally, it was agreed that the OGD management capabilities, due to their importance for the use and the generation of value from OGD, should be viewed as a separate research area. In Figure 4 we can see the research topics of the "OGD Management and Policies" research area, while in Table 1 these OGD research topics are described in more detail, supported by some representative relevant literature from the EGRL.

---

[1]http://mind42.com/public/f2a7c2f6-63ec-475f-a848-7ed5abe6c5a4.

Figure 4. Research topics for the OGD Management and Policies research area.

Table 1. Description of the research topics of OGD Management and Policies research area.

Research topic	Description
1.1 Policy and legal issues for OGD	This research topic concerns the investigation of different policies, strategies, and principles for opening data, as well as specific measures and instruments in this direction (Blakemore and Craglia 2006; European Commission 2013a and 2013b; Zuiderwijk and Janssen 2014c). Formulating an OGD policy is a complex multidisciplinary problem, and as such it is associated with many of the following research topics.
1.2 OGD anonymization methods	The current practice in data publishing relies mainly on policies and guidelines as to what types of data can be published and on agreements concerning the use of published data. A major precondition for opening data of government agencies is not to disclose sensitive private data of citizens and firms. Therefore this research area focuses on methods for the anonymization of opened data. Privacy-preserving data publishing (PPDP) provides methods and tools for publishing useful information while preserving data privacy (Benjamin et al. 2010).
1.3 OGD cleaning methods	This research topic deals with data cleaning methods for OGD, which aim to correct errors in quantitative attributes of datasets, or even other types of attributes (Hellerstein 2008). Data cleaning is a process used to determine inaccurate, incomplete or unreasonable data, and then improve their quality through correcting of detected errors and omissions. Generally data cleaning reduces errors and improves the data quality (Natarajan, Li, and Koronios 2010).
1.4 OGD quality assessment frameworks	This research topic deals with data quality, a major issue in information management in general, highly important for OGD in particular. Data quality problems occur anywhere in information systems, and they are solved by data cleaning (see previous research topic). After applying data cleaning, the quality of the data can be assessed in a number of ways, based on the internal consistency of the data and comparison of the corrected intensities with the corrected standard deviations (Chapman 2005).
1.5 OGD visualization methods and tools	Visualization methods and tools is an important research topic, aiming to provide simple mechanisms for understanding and communicating large amounts of data. There is a need for exploratory mechanisms to navigate the data and metadata in these visualizations. It is therefore highly important to develop features and tools for facilitating the creation of visualizations by users on OGD (Graves and Hendler 2013).
1.6 OGD linking	The principles, frameworks, techniques, and tools for OGD linking are the subjects of this research area (Kalampokis, Tambouris, and Tarabanis 2013; Bojārs et al. 2008). The term linked data refers to data published on the web so that they are machine-readable, their meaning is explicitly defined, can be linked to (and from) other external datasets (Bizer, Heath, and Berners-Lee 2009). The advancements on this research topic concentrate on how we can structure our data so that we can find, link, and process them more easily. Knowledge management representation systems have been created and continue evolving in order to link different data.

(*Continued*)

Table 1. (Continued).

Research topic	Description
1.7 OGD publishing	The OGD Publishing research deals with and investigates all the issues of the publishing workflow and its involved actors (Bizer, Heath, and Berners-Lee 2009; Dawes and Helbig, 2010; Helbig et al. 2012). It also examines the interconnection between the OGD publishing processes and their context (main actors and their interests and goals), and also their effects on OGD use and outcomes, and on their dynamics.
1.8 OGD mining	The OGD mining research aims to exploit and elaborate the algorithms and methods developed in the area of data mining, in order to extract useful patterns and knowledge from OGD. Data mining uses a broad family of computationally intensive methods which include decision trees, neural networks, rule induction, machine learning and graphic visualization (Bakirli et al. 2012; Mostafa and El-Masry 2013; Kum, Duncan, and Stewart 2009; Mannila 2002).
1.9 OGD rating and feedback	This research focuses on policies and mechanisms for closing the feedback loop between OGD users and providers, through establishing communication channels between them. Another important objective of this research is to enable OGD providers to manage efficiently comments and requests from OGD users. Thus, tools for supporting the rating of OGD and their infrastructures, providing feedback to the corresponding public organizations are more than essential. The use of OGD users–providers collaboration techniques for the previously mentioned purposes are also investigated in this research area, for example, through Web 2.0 oriented mechanisms (Alexopoulos et al. 2014; Charalabidis, Loukis, and Alexopoulos 2014).

4.2. OGD infrastructures

The second research area of OGDRAT has been named "Open Government Data Infrastructures." It includes research topics concerning various important technological aspects of the ICT infrastructures developed by government agencies in order to make OGD accessible to different groups of actors, such as their architectures, APIs provision, and personalization capabilities; another important research topic is OGD storage and long-term preservation, and also the use of cloud services in this domain. Further, although the main source of OGD is the information systems of government agencies, two more sources are gradually emerging, sensors and citizens; therefore, researching them and their exploitation is an important research challenge. In Figure 5 we can see the research topics of the OGD Infrastructure research area, while in Table 2 these OGD research topics are described in more detail, supported with representative literature from the EGRL.

Figure 5. Research topics for the OGD infrastructures research area.

Table 2. Description of the research topics of the OGD infrastructures research area.

Research topic	Description
2.1 OGD portals architecture	This research aims at defining the architectures of OGD portals, with respect to their scope and provided data and functionalities (Alexopoulos et al. 2014; Charalabidis, Loukis, and Alexopoulos 2014; Helbig et al. 2012). Various types and generations of architectures are proposed and discussed from various perspective. Additionally, some research is conducted concerning the development of architectures of ICT Infrastructures that allow for and support application development utilizing OGD.
2.2 Open web services/APIs	This research aims at facilitating and providing well-designed standards for application programming interfaces (APIs) in OGD platforms, in order to ensure the exploitation and re-usability of published data. It is of high importance to use APIs for machine-to-machine operations for OGD. Unfortunately many of the OGD are not machine readable or the data are provided in a proprietary format (Braunschweig et al. 2012). Open web services in this domain should conform to a set of conventions that define how a client searches for and interacts with a service (Paolucci et al. 2002; Kleijnen and Raju 2003).
2.3 OGD user profiling and service personalization	This research focuses on user profiling, which can offer big opportunities to make OGD related services more personalized, to infer and predict citizens' behavior, and to even influence their behavior (Pieterson, Ebbers, and Van Dijk 2005). Like the private sector, the public sector makes more and more use of user profiling in order to personalize the electronic services that are being offered to citizens (Mostafa and El-Masry 2013).
2.4 OGD long-term preservation	This research topic can be found in every ICT-related research domain, dealing with the ways and methods for the long-term preservation of data, which is particularly important for OGD (Agrawal and Srikant 2000).
2.5 OGD storage	This research topic concerns the optimization of OGD storage, combining knowledge from various domains, such as databases and algorithms.
2.6 Cloud computing for OGD	The use of private and public cloud computing technologies and services (Lewis 2013) for hosting and providing OGD is an important research challenge, taking into account the increasing adoption of cloud in the public sector (Joshi 2012). The Linked Open Data Cloud creation supporting the vision of the Web of Data is also a research challenge classified under this research topic (Sorrentino et al. 2013; Jain et al. 2010b).
2.7 Citizen-generated open data	This research aims to investigate the emerging and continuously growing volunteered user-generated content, which is often used to replace existing commercial or authoritative datasets, for example, Wikipedia[a] as an open encyclopedia, or OpenStreetMap[b] as an open topographic dataset of the world (Richter and Winter 2011). Open data generated by citizens, e.g., through e-participation platforms and social media, and their use for "crowdsourcing" purposes, are an emerging research topic of this research area (Heipke 2010).
2.8 Sensor-generated open data	This emerging research topic involves tools, methods and techniques for OGD generation through sensors, which will be made freely available to the public. Big data are becoming of critical importance for science and commercial applications development (e.g., Elgendy and Elragal 2014), so exploiting the knowledge developed in this domain and elaborating it for the OGD can be quite useful. This research topic also includes the development of methods of processing such data, calculation of analytics, and finally exploitation of them (for scientific and business purposes).

[a]http://en.wikipedia.org/wiki/Main_Page.
[b]http://www.openstreetmap.org/.

4.3. *OGD interoperability*

Interoperability is a highly important feature of all types of information systems, and this gave rise to the development of a well-established research domain, which attracts considerable research interest, motivated by the increasing need of data exchange among organizations (both of the private and the public sector) (Jardim-Goncalves et al. 2013). Interoperability has many aspects, mainly technical, semantic, and organizational. It becomes increasingly important in government, since "The divergent interpretations of data, the lack of common metadata, and the absence of universal reference data hinder governments from seamless data exchange, information systems integration, and the delivery of cross-border public services" (Shukair et al. 2013, 10). Therefore our third research area deals with the interoperability issue in the specific domain of OGD. It includes research topics concerning OGD metadata, semantic annotation, ontologies and controlled vocabularies and codelists, and also on OGD platforms technical interoperability, services interoperability standards, and organizational

OGD Management and Policies	OGD Infrastructures	**OGD Interoperability**	OGD Usage and Value
• Policy and Legal Issues for OGD	• OGD Portals Architecture	3.1 Metadata for OGD	• Skills management for OGD
• OGD Anonymisation Methods	• Open Web Services / API's	3.2 Multilinguality Issues	• Reputation Management
• OGD Cleansing Methods	• OGD User Profiling and Service personalisation	3.3 Services Interoperability Standards	• OGD Uses
• OGD Quality Assessment Frameworks	• OGD Long Term Preservation	3.4 Semantic Annotation	• OGD-based Entrepreneurship
• OGD Visualisation methods and tools	• OGD Storage	3.5 OGD Ontologies	• OGD Value & Impact Assessment
• OGD Linking	• Cloud Computing for OGD	3.6 Platform & Technical Interoperability	• OGD Needs Declaration
• OGD Publishing	• Citizen - generated Open Data	3.7 Organisational Interoperability	• OGD-based Accountability
• OGD Mining	• Sensor - generated Open Data	3.8 Controlled Vocabularies / Codelists Preservation	• OGD Readiness Assessment
• OGD Rating and Feedback			• OGD Portals Evaluation Frameworks
			• OGD Innovation

Figure 6. Research topics for the OGD interoperability research area.

interoperability. In Figure 6 we can see the research topics of the OGD Interoperability research area, which are described in more detail, and also supported with relevant literature from the EGRL, in Table 3.

4.4. OGD usage and value

The fourth research area of OGDRAT is directed toward the measurement and deeper understanding of the use of OGD as well as the impact and value generation from them. It includes research topics concerning OGD needs, readiness, use, skills management, and reputation management as well as OGD related value and impact, innovation, entrepreneurship, and contribution to accountability/transparency. In Figure 7 we can see the research topics of this "OGD Usage and Value" research area, while an elaboration of them and EGRL literature support are provided in Table 4.

5. Discussion

This section presents the outcomes of the further processing and exploitation of the OGDRAT conducted finally as part of step 8 of our research methodology (see section 2): analysis of EGRL publications for each of the identified research topics (5.1); development of an extended OGD lifecycle (5.2); exploitation of OGDRAT for OGD Science Base Creation (5.3); association of OGDRAT with the ICT-enabled Governance research taxonomy developed in the CROSSROAD and the CROSSOVER projects, use of the former in order to extend the latter (5.4); and formulation of direction for multidisciplinary research on important societal challenges using OGD (5.5).

5.1. EGRL publications for research topics

For all the OGD research areas identified and presented in the previous section (of the final version of the OGDRAT produced in step 7 our methodology, section 2) we searched for relevant publications in the EGRL. In Figure 8 we can see the number of publications found for each topic (the topics are sorted in descending order of publications' number); for the few publications that concern more than one of these topics we proceeded to their classification in the one judged as dominant (after discussion and consensus reaching among the authors). We remark that there are significant differences among these research topics as to the number of

Table 3. Description of the research topics of the "OGD Interoperability" research area.

Research topic	Description
3.1 Metadata for OGD	This research topic includes various OGD metadata-related research subtopics: data models, schemata, taxonomies, codelists, and ontology-based extended metadata sets for OGD, and also other types e-Government Resources. The term semantic interoperability asset is widely used to refer to these types of resources (Charalabidis, Lampathaki, and Askounis 2009; Zuiderwijk, Jeffery, and Janssen 2012; Robertson et al. 2001).
3.2 Multilinguality	Multilinguality is a research topic that has been attracting a growing interest by supranational institutions, such as the European Union. It includes research associated with using, extending, combining, and developing semantic assets toward the support of multilinguality in the domain of OGD (Houssos, Jörg, and Matthews 2012).
3.3 Service interoperability standards	This research topic is concerns mainly the identification, composition, and execution of various applications (designed and implemented independently) offered as services. This research investigates standards that can be used for seamless interconnection among OGD related services, in order to serve different OGD uses and user scopes (Jardim-Goncalves et al. 2013). It includes the development of information systems and registries consisting of workflow models and process descriptions in an integrated knowledge base (Sourouni et al. 2008).
3.4 Semantic annotation	This research focuses on methods and tools for the semantic annotation of OGD generated by public organisations and sensors, as well as the semantic annotation of user-generated content (UGC) (Deng 2013). Semantic annotation techniques capture not only the semantics, but also the pragmatics of the resources, such as who, when, where, how, and why the resources are used (Kiryakov et al. 2004; Warner and Chun 2009; Dill et al. 2003). The major objective of this research is the development of algorithms and tools for semantic integration (Bergamaschi, Castano, and Vincini 1999), and also for automated extraction of metadata (self-extracted metadata).
3.5 OGD ontologies	This research topic includes investigation of the proper release of OGD and the use of ontologies behind these sources (Parundekar, Knoblock, and Ambite 2010). Ontologies for the description and use of OGD as well as the sense of ontology alignment are under investigation in this research (Jain et al. 2010a; Alexander et al. 2009). The Linked Open Data (LOD) paradigm is the major outcome of this research area.
3.6 Platform technical interoperability	This research examines various technical issues involved in linking OGD systems and services, such as open interfaces, interconnection services, data integration, middleware, data presentation and exchange, accessibility, and security services (Sarantis, Charalabidis, and Psarras 2008; Jardim-Goncalves et al. 2013).
3.7 Organizational interoperability	The main objective of this research is the investigation of the processes by which different organisations, such as different government agencies, collaborate in order to achieve mutually beneficial agreed e-Government OGDservice-related goals (Sarantis, Charalabidis, and Psarras 2008; Jardim-Goncalves et al. 2013), which concern the publishing and the management of OGD.
3.8 Controlled vocabularies and codelists preservation	This research includes investigation regarding preservation, indexing, and retrieval of semantic assets, such as vocabularies and codelists (Kiryakov et al. 2004).

relevant publications: for some of them we have found more publications, for example, in research topics concerning OGD use, portals evaluation frameworks, publishing, policy, and legal issues, while in some others we found significantly less or even no publications, for example, in research topics concerning sensor-generated OGD, OGD storage, long-term preservation, reputation management, and skills management (for these five research topics there is no relevant literature in the EGRL, and were proposed in the workshop (step 6 of our OGDRAT development methodology (section 2)) by the experts who participated as major issues of OGD). Also, from Figure 8 we can see that there are many under-researched topics with very small numbers of relevant publications. Therefore further research is required on these research topics with very small numbers or even no publications, since they constitute interesting emerging topics, which can be significant for the achievement of higher maturity in OGD practices and value generation from them.

OGD Management and Policies	OGD Infrastructures	OGD Interoperability	OGD Usage and Value
· Policy and Legal Issues for OGD · OGD Anonymisation Methods · OGD Cleansing Methods · OGD Quality Assessment Frameworks · OGD Visualisation methods and tools · OGD Linking · OGD Publishing · OGD Mining · OGD Rating and Feedback	· OGD Portals Architecture · Open Web Services / API's · OGD User Profiling and Service personalisation · OGD Long Term Preservation · OGD Storage · Cloud Computing for OGD · Citizen - generated Open Data · Sensor - generated Open Data	· Metadata for OGD · Multilinguality Issues · Services Interoperability Standards · Semantic Annotation · OGD Ontologies · Platform & Technical Interoperability · Organisational Interoperability · Controlled Vocabularies / Codelists Preservation	4.1 Skills management for OGD 4.2 Reputation Management 4.3 OGD Use 4.4 OGD-based Entrepreneurship 4.5 OGD Value & Impact Assessment 4.6 OGD Needs Analysis 4.7 OGD-based Accountability 4.8 OGD Readiness Assessment 4.9 OGD Portals Evaluation Frameworks 4.10 OGD Innovation

Figure 7. Research topics for the OGD Usage and Value research area.

Table 4. Description of research topics for the "OGD Usage and Value" research area.

Research topics	Description
4.1 Skills management for OGD	This research aims to identify and understand better the necessary skills required for OGD analysis and processing (by OGD users' side), and also for OGD publishing and management (by OGD providers' side). They are usually defined in terms of skills frameworks (also termed as competency frameworks or skills matrices); each of them consists of a list of skills, and a grading system, with a definition of what it means to be at particular level for a given skill.
4.2 Reputation management	This research includes investigation of the use of reputation systems in OGD value chain. It examines various algorithms and methods for the reputation management of various OGD stakeholders (Bani and Paoli 2013; Hansson et al. 2013).
4.3 OGD use	It includes studies that describe and analyze examples, ways and paradigms of OGD use for various purposes, not only by citizens (e.g., scientists, journalists, active citizens, firms active in the development of value-added e-services, and mobile applications), but also by government as well (e.g., for policy making: Kalampokis et al. (2011a) combined social data and ODG for participatory decision-making in government).
4.4 OGD-based entrepreneurship	This research topic concerns mainly business models for exploiting the potential value of OGD and initiating OGD value chains (Ferro and Osella 2012, 2013).
4.5 OGD value and impact assessment	The current OGD research on this topic focuses on analyzing OGD initiatives that have led to the generation of some kind of public value (Davies, Perini, and Alonso 2013; Jetzek, Avital, and Bjorn-Andersen 2012; 2013; Charalabidis, Loukis, and Alexopoulos 2014), analyzing the positive—and sometimes the negative as well—aspects of OGD use and impacts.
4.6 OGD needs analysis	This research includes studies of OGD users' needs, with respect to both government datasets, and also functionalities of OGD infrastructures, aiming to lead to further developments of OGD strategies of public organizations, and also functionalities of ODG infrastructures/portals. For instance, this research led to the identification of needs for collaboration workflows and feedback mechanisms (Alexopoulos et al. 2014), and also needs for better metadata and semantic annotation mechanisms (Zuiderwijk, Jeffery, and Janssen 2012).
4.7 OGD-based accountability	This research investigates the use of OGD as part of anticorruption programs in order to increase public sector accountability and credibility. Many government organizations publish a variety of datasets on the web, in order to promote transparency, accountability, and satisfy relevant legal obligations (Böhm et al. 2012; Alon 2011).
4.8 OGD readiness assessment	The main objective of this research is to develop frameworks and methods for assessing from various viewpoints (both "internal" and "external" ones) the degree of readiness of a national, regional, or municipal government—or even individual agencies—to implement OGD initiatives (Davies, Perini, and Alonso 2013; World Bank 2013).
4.9 OGD portals evaluation frameworks	This research aims at the creation of roadmaps, guidelines, and benchmarking frameworks for the evaluation of OGD portals and infrastructures from various viewpoints (Charalabidis, Loukis, and Alexopoulos 2014; Alexopoulos et al. 2013; Kalampokis et al. 2011b).

(*Continued*)

Table 4. (Continued).

Research topics	Description
4.10 OGD innovation	The main objective of this OGD research is to identify and analyze innovations driven by OGD, both in the private sector (e.g., e-services innovations), and in the public sector (Zuiderwijk et al. 2014a). According to this literature, OGD innovation concerns mainly three domains: (a) research, (b) business, and (c) transparency (Jetzek, Avital, and Bjorn-Andersen 2012; 2013). While US literature and practice focuses mainly on (b), EU tends to focus on (a), but both of them are equally interested the potential of promoting (c) through OGD.

5.2. *An extended OGD life cycle*

Taking into account the OGD research topics identified during the development of our taxonomy, and also the discussions we had in the workshop with the experts who participated in it, which revealed a wide range of tasks to be performed during the life cycle of OGD, we proceeded to the development of an extended OGD life cycle, based on the combination of the Linked OGD Life Cycle[2] (Open Data Support Working Group) and the Curation Life Cycle.[3] It is shown in Table 5: it consists of nine stages (create, preprocess, curate, store/obtain, publish, retrieve/acquire, process, use, and collaborate with users), and for each of the stages, associated tools and methods that can be used in the particular stage are shown.

5.3. *Contribution to OGD science-based creation*

As mentioned in section 2 the research presented in this article contributes to the development of "description theory" for the OGD domain, so it constitutes the first step toward the creation of a science base for it. According to Charalabidis, Gonçalves, and Popplewell (2011) the science base of a domain should include the main concepts, methods, tools, and standards of the domain, and also supportive relevant experiments, surveys, and case studies that have been conducted and the body of knowledge produced in the domain, and also various types of "proofs of concept" aiming to assist practitioners in this domain to solve particular problems and generate value. Our OGDRAT contributes in these directions, as (1) it identifies the main concepts, methods, and tools in OGD; and (2) provides directions for future research in this domain, aiming to increase maturity of these methods and tools, so that finally OGD stakeholders (government, scientific communities, journalists, active citizens, and e-/m-services development firms) can be systematically assisted in their relevant activities, leading to higher value generation from OGD.

5.4. *Extension of ICT-enabled governance taxonomy*

The OGDRAT is associated with and extends/elaborates the ICT-enabled Governance research taxonomy developed in the CROSSROAD[4] and the CROSSOVER[5] European projects. In particular, the CROSSROAD project has developed a research areas taxonomy for the ICT-enabled Governance domain, which consists of five main research themes, 17 research areas, and more than 80 research subareas (Lampathaki et al. 2010). One of the research themes of this taxonomy is "Open Government Information & Intelligence for Transparency," which includes three research areas concerning "Open and Transparent Information Management," "Linked Data," and "Visual

---

[2]https://joinup.ec.europa.eu/sites/default/files/D2.1.1%20Training%20Module%202.1%20The%20Linked%20Open%20Government %20Data%20Lifecycle_v0.11_EN.pdf.
[3]http://www.dcc.ac.uk/resources/curation-lifecycle-model#sthash.FnrCA3Kf.dpuf.
[4]http://www.2020-horizon.com/CROSSROAD-CROSSROAD-A-Participative-Roadmap-for-ICT-Research-in-Electronic-Governance-and-Policy-Modelling(CROSSROAD)-s9412.html.
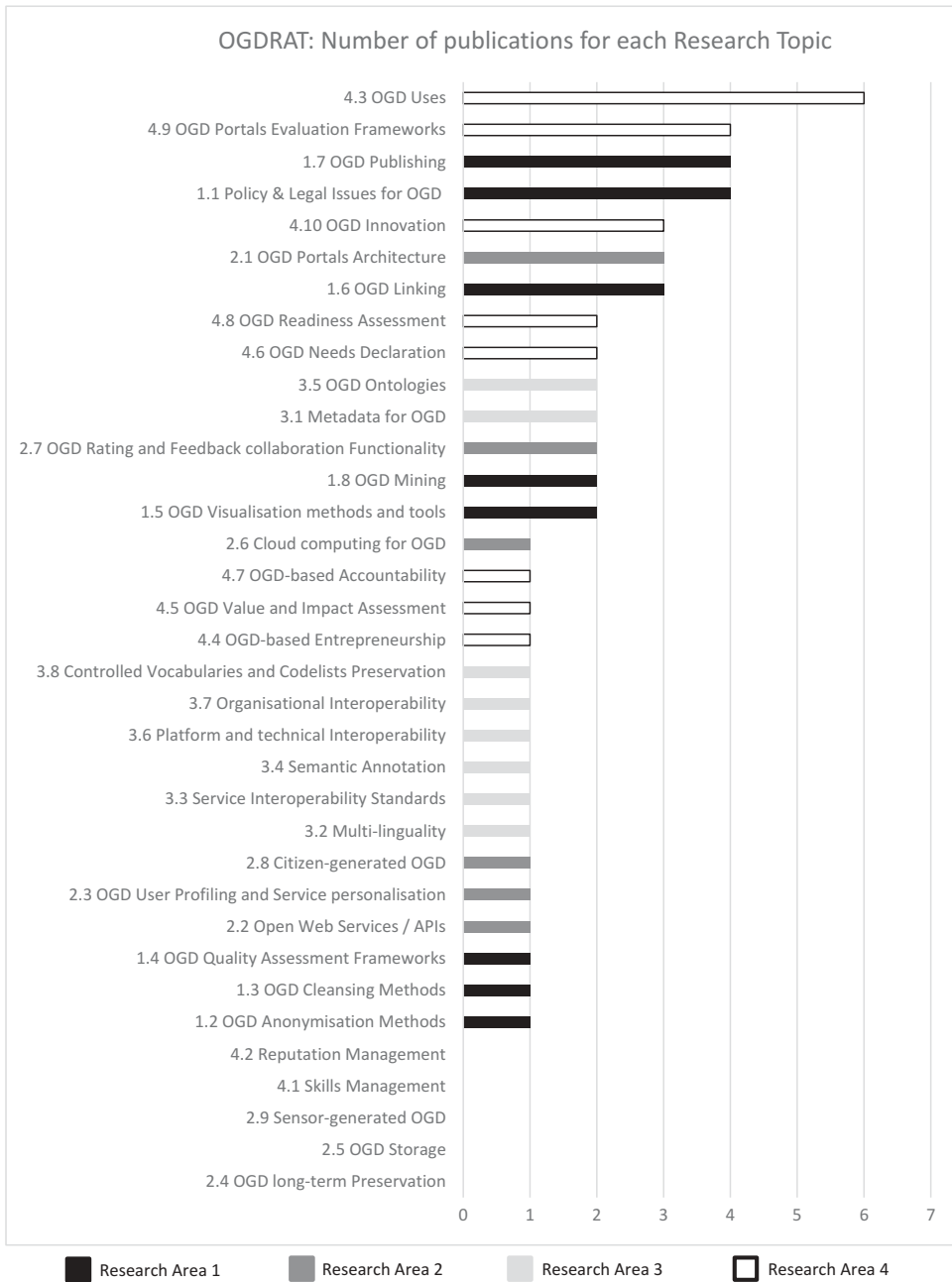[5]http://www.crossover-project.eu/ResearchRoadmap.aspx.

Figure 8. Ranking of OGD research topics based on EGRL relevant literature.

Analytics." The OGDRAT extends and elaborates this research theme, because the main research areas and topics of the former can replace the research areas and subareas of the latter, providing a higher level of detail and adding recently emerged research topics.

Also, the CROSSOVER project developed a taxonomy of research challenges in a related but narrower domain, concerning the next generation of public policy making in the Web 2.0 social media context (policy making 2.0) (CROSSOVER Project Deliverable 2.2.2, 2013), which categorizes these research challenges under two research themes: (1) Data-powered Collaborative Governance

Table 5. Extended OGD life cycle and associated tools and methods.

Life cycle stage	Tools	Methods
Create	• Sensors, RFID, IS, human, connection with already gathered OGD	• Automated data creation • Manual data entry • Linking with OGD Portals
Pre-process	• Detailed metadata standards • Evaluation metrics and models • Maturity matrices	• Conceptualization • Structuring • Evaluation
Curate	• Web tools for LOD (open refine external tool) • Individual/native tools	• Metadata refinement • Change data format • Data cleansing
Store/Obtain	• Repository and data center • Cloud infrastructures	• Versioning • Data linking
Publish	• Data publication mechanisms • Data publication sites	• Licensing • Intellectual property rights
Retrieve/Acquire	• Advanced search techniques (i.e., multilingual) • Download capabilities	• Open Access • 3-layer metadata schema
Process	• External data processing tools • Web tools for LOD (open refine external tool)	• Data enrichment • Create linked open data • Different datasets combination
Use	• Internal visualization tools • External visualization tools • Statistical packages • Linking with external artefacts	• Statistical analysis • Map visualization • Chart visualization • Plot visualization
Collaborate	• Web 2.0 capabilities and tools • Collaboration workflows • Feedback mechanisms and tools	• Exchange notes/emails/ideas • Data quality rating • Create groups of common interests • Needs and requests on OGD

and (2) Policy Modelling, in order to develop a roadmap on policy making 2.0. The OGDRAT extends and elaborates the "Link Open Government Data" research challenge of the "Data-powered Collaborative Governance" theme.

5.5. *Multidisciplinary research on societal challenges based on OGD*

In the workshops it was emphasized by the participating experts that the most important and socially beneficial research OGD research can be conducted by using them as a basis of multidisciplinary research on important societal problems and challenges that modern societies face. These data can be used by multidisciplinary scientific teams, for example, including members from various "neighboring scientific domains," such as economic, political, social, management, and behavioral sciences (and using theoretical foundations from these sciences) in order to perform various sophisticated analyses from various disciplinary perspectives and gain useful synthetic insights into serious problems and challenges of modern societies; these can be quite important for the design of effective solutions and public policies for addressing them. Some directions for such mutlidisciplinary research were mentioned, and are summarized in Table 6.

6. Conclusions

As mentioned in the Introduction, the OGD research domain is still in its early stages, so it is important to develop a taxonomy of its main research areas and topics. This article makes the following contributions in this direction:

- It develops a detailed taxonomy of research areas and corresponding research topics of the OGD domain (which was missing from OGD previous literature, despite its importance for the progress of this domain toward higher levels of maturity), through extraction and combination

Table 6. Directions of multidisciplinary research on societal challenges based on OGD.

Societal challenge	ICT-enabled governance research topic	OGD research topic	Neighboring scientific domain
Language divide and lack of cross-communities communication	• Language and cultural interoperability	• Metadata for OGD	• Information intelligence
		• Multilinguality	• Computer science (translation tools)
		• Controlled vocabularies and codelists preservation	• Behavioral sciences
Anticipating unexpected crises	• Social—Economic simulation models	• Semantic annotation	• Social and economic sciences
	• Policy modelling	• Organizational interoperability	
	• Process optimization for OGD (accurate provision)	• Sensor-generated open data	
Enhanced collective cognitive intelligence (human/ICT-enabled) for better Governance	• Modelling and simulation	• OGD mining	• Economics
	• Policy analysis	• Citizen-generated open data	• Mathematics
	• Identity management	• Visualization	• Sociology
		• Information management	• Computer science

- of relevant knowledge from three kinds of sources: important relevant government policy documents, research literature, and experts. For each of these OGD research topics relevant literature from the EGRL has been identified and analyzed, which enables a better understanding of them and their main research objectives and directions.
- Based on the taxonomy, an extension of the existing OGD lifecycle has been initially developed, which includes important additional stages that can contribute to higher maturity in the OGD-related practices of both providers and users of them, and finally more social and economic value generation from them. Furthermore, our OGD research taxonomy has been connected with two previous research taxonomies for the "ICT-enabled Governance" and "Policy Making 2.0" domains, respectively, which have been developed in the European projects CROSSROAD and CROSSOVER, providing extensions and elaborations of them for the OGD domain. Finally, directions have been formulated for future multidisciplinary research based on OGD for addressing important challenges that modern societies face.

The findings of our study reveal the interesting thematic "richness" of the OGD research domain, which currently includes a wide range of research topics, both technological and nontechnological ones, concerning both the opening and publishing of government datasets, and also their usage (by various actors, such as e-service or mobile apps developers, scientists, analysts, journalists, active citizens, etc.), exploitation, and value generation from them. This reflects the inherent complexity of opening of government data to the society and the economy, and then creating value from them, which the OGD research aims to address. In particular, we identified a multitude of technological research topics in the OGD research domain, with most of them concerning the exploitation of existing or emerging technologies, on one hand in the opened datasets (e.g., anonymization, cleansing, mining, metadata, linking, and semantically enriching technologies), and on the other hand in the OGD infrastructures (e.g., web services, storage, cloud computing, interoperability technologies), in order to enrich their usefulness. Furthermore, we identified a multitude of non-technological OGD research topics, which concern mainly OGD needs, use, impact, value, and entrepreneurship.

However, our study has revealed significant differences among the identified OGD research topics as to the "quantity" of the research conducted on them. For some of these topics there are limited or even no publications at all (e.g., for research topics sensor-generated OGD, OGD storage, long-term

preservation, reputation management, and skills management); so further research is required on these under-researched topics.

Our research taxonomy has interesting implications for research and practice. With respect to research it provides directions and structure for future research in the OGD domain, and also facilitates communication and interaction among researchers (through the "common language" it introduces), and also with interested practitioners. Also, it contributes to the development of a "description theory" of the OGD domain, which can be useful for the development of other more advanced types of theories (as mentioned in Section 2). Finally, it identifies important under-researched topics, on which further research is required. With respect to practice our OGDRAT is useful to government agencies, as it proposes to them possible dimensions of their OGD strategies, practices, and infrastructures, on which they should focus their attention in order to improve the value generated from them. Also, this detailed taxonomy can contribute to the development of new knowledge in this domain, which will enable improving and optimizing the technology, and also the design, operations, and performance of the units of government agencies responsible for opening data. Finally, OGDRAT is useful to ICT firms developing OGD technological infrastructures, as it provides them directions for improving their products and services.

A limitation of our study is that for practical reasons we organized only one workshop (although we had participants from 11 EU countries, and from different types of organizations, such as public administrations, universities, and firms). So it is necessary to organize more workshops in order to further validate the OGDRAT, and probably have proposals for additional research topics, with participants from all major stakeholder groups, such as such as e-service or mobile apps developers, scientists, analysts, journalists, active citizens, and public servants. In this direction the proposed taxonomy is available on the web and can be accessed by the following link http://mind42.com/public/f2a7c2f6-63ec-475f-a848-7ed5abe6c5a4, so that we can collect ratings, comments, and ideas from the OGD community for further elaboration and update. Another limitation is that we have identified and analyzed relevant literature for all the research topics of OGDRAT only from the EGRL; so it would be good to exploit other research libraries as well. In addition, more research is needed to map the multiple OGD research projects that are currently in progress (e.g., supported by European Commission or US research programs) to the first-level research areas and the second-level topics of the OGDRAT, and possibly based on them elaborate and update the taxonomy.

ORCID

Charalampos Alexopoulos 🔟 http://orcid.org/0000-0002-6610-0675

References

Agrawal, R., and R. Srikant. 2000. Privacy-preserving data mining. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data (SIGMOD '00)*, ACM, New York, NY, USA, 439–450. doi:10.1145/342009.335438.

Alexander, K., and M. Hausenblas. 2009. Describing linked datasets-on the design and usage of void, the 'vocabulary of interlinked datasets. In *Linked Data on the Web Workshop (LDOW 09), in conjunction with 18th International World Wide Web Conference (WWW 09)*. Cambridge, MA: W3C.

Alexopoulos, C., E. Loukis, Y. Charalabidis, and A. Zuiderwijk. 2013. An evaluation framework for traditional and advanced open public data e-infrastructures. In *CEG2013-13th European Conference on eGovernment: ECEG 2013*, 102. Academic Conferences Limited.

Alexopoulos, C., A. Zuiderwijk, E. Loukis, and M. Janssen. 2014. Designing a second generation of open data platforms: Integrating open data and social media. In *Electronic Government*, 230–241. Berlin Heidelberg, Germany: Springer.

Alon, P. 2011. When transparency and collaboration collide: The USA open data program. *Journal of the American Society for Information Science and Technology, Wiley Subscription Services, Inc., A Wiley Company*. doi:10.1002/asi.21622.

Alter, S. 2010. Viewing systems as services: A fresh approach in this field. *Communications of the Association for Information Systems* 26(11):2010.

Bakırlı, G., D. Birant, E. Mutlu, A. Kut, L. Denktaş, and D. Çetin. 2012. Data mining solutions for local municipalities. In *European Conference on e-Government*, 74. Academic Conferences International Limited.

Bani, M., and S. D. Paoli. 2013. Ideas for a new civic reputation system for the rising of digital civics: Digital badges and their role in democratic process. In *ECEG2013-13th European conference on eGovernment: ECEG 2013*, 45. Academic Conferences Limited.

Barry, E., and F. Bannister. 2013. *Barriers to open data release: A view from the top*. Edinburgh: European Group for Public Administration.

Bason, C. 2010. *"Leading Public Sector Innovation," co-creating for a better society*. Bristol, United Kingdom: The Policy Press.

Benjamin, C. M., F. K. Wang, R. Chen, and P. S. Yu. 2010. Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv* 42(4):53. doi:10.1145/1749603.1749605.

Bergamaschi, S., S. Castano, and M. Vincini. 1999. Semantic integration of semistructured and structured data sources. *ACM Sigmod Record* 28(1):54–59. doi:10.1145/309844.

Bizer, C., T. Heath, and T. Berners-Lee. 2009. Linked data—The story so far. *International Journal on Semantic Web and Information Systems* 5(3):1–22. doi:10.4018/jswis.2009081901.

Blakemore, M., and M. Craglia. 2006. Access to public-sector information in Europe: Policy, rights and obligations. *The Information Society* 22(1):13–24. doi:10.1080/01972240500388180.

Böhm, C., M. Freitag, A. Heise, C. Lehmann, A. Mascher, F. Naumann, V. Ercegovac, M. Hernandez, P. Haase, and M. Schmidt. 2012. GovWILD: Integrating open government data for transparency. In *Proceedings of the 21st international conference companion on World Wide Web (WWW '12 Companion)*, ACM, New York, NY, USA, 321–324. doi:10.1145/2187980.2188039.

Bojārs, U., J. G. Breslin, A. Finn, and S. Decker. 2008. Using the semantic web for linking and reusing data across Web 2.0 communities. *Web Semantics: Science, Services and Agents on the World Wide Web* 6(1): February 200821–28. ISSN 1570-8268. 10.1016/j.websem.2007.11.010.

Borins, S. 2001. Encouraging innovation in the public sector. *Journal of Intellectual Capital* 2(3):310–19. doi:10.1108/14691930110400128.

Braunschweig, K., J. Eberius, M. Thiele, and W. Lehner. 2012. The state of open data: Limits of current open data platforms. Paper presented at the International World Wide Web Conference, Lyon, France. http://www2012.wwwconference.org/proceedings/nocompanion/wwwwebsci2012_braunschweig.pdf.

Chapman, A. D. 2005. Principles and Methods of data cleaning—Primary Species and Species—Occurrence Data, Version 1.0, Report for the Global Biodiversity Information Facility, Copenhagen.

Charalabidis, Y., R. J. Gonçalves, and K. Popplewell. 2011. Towards a scientific foundation for interoperability. In *Interoperability in digital public services and administration: Bridging E-government and E-business*, ed. Y. Charalabidis, 355–373. New York: Information Science Reference, Hershey.

Charalabidis, Y., F. Lampathaki, and D. Askounis. 2009. Metadata sets for e-government resources: The extended e-government metadata schema (eGMS+). In *Electronic government: 8th international conference (EGOV 2009)*, eds. M. A. Wimmer, H. J. Scholl, M. Janssen, and R. Traunmüller, vol. 5693, 341–52. Berlin: Springer Verlag.

Charalabidis, Y., E. Loukis, and C. Alexopoulos. 2014. Evaluating second generation open government data infrastructures using value models. In *2014 47th Hawaii International Conference on System Sciences (HICSS)*, 2114–2126. IEEE.

Conradie, P., and S. Choenni. 2012. Exploring process barriers to release public sector information in local government. In *Proceedings of the 6th international conference on theory and practice of electronic governance* (ICEGOV), 5–13. ACM.

CROSSOVER Project—Deliverable 2.2.2. 2013. Towards policy-making 2.0: The International Roadmap on ICT for Governance and Policy Modelling http://crossover-project.eu/Portals/0/0205F01_International%20Research%20Roadmap.pdf.

Davies, T., F. Perini, and J. M. Alonso. 2013. Researching the emerging impacts of open data. ODDC conceptual framework. http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf.

Dawes, S., and N. Helbig. 2010. Information strategies for open government: Challenges and prospects for deriving public value from government transparency. In *Electronic government*, 50–60. Berlin Heidelberg, Germany: Springer.

Deng, D., G. Mai, C. Hsu, C. Chang, T. Chuang, and K. Shao. 2013. *Linking open data resources for semantic enhancement of user–generated content*. Berlin Heidelberg, Germany: Springer. doi:10.1007/978-3-642-37996-3_30.

Dill, S., N. Eiron, D. Gibson, D. Gruhl, R. Guha, A. Jhingran, and J. Y. Zien. 2003. SemTag and seeker: Bootstrapping the semantic web via automated semantic annotation. In *Proceedings of the 12th international conference on World Wide Web*, 178–186. ACM.

Elgendy, N., and A. Elragal. 2014. Big data analytics: A literature review paper. *Advances in Data Mining, Applications and Theoretical Aspects Lecture Notes in Computer Science* 8557:214–227.

European Commission. 2011. Digital agenda: Turning government data into gold. *European Commission, Brussels.* P/11/1524, 2011.

European Commission. 2012, December. Directive 2003/98/EC of the European parliament and of the council of 17 November 2003 on the re-use of public sector information. European Commission. http://ec.europa.eu/informationsociety/policy/psi/rules/eu/ index en.htm.

European Commission. 2013a. EU implementation of the G8 open data charter. http://ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=3489.

European Commission. 2013b. Digital agenda: Commission's open data strategy, questions & answers. http://europa.eu/rapid/pressReleasesAction.do?reference=MEMO/11/891&format=HTML&aged=1&language=EN&guiLanguage=en.

European Commission. 2014. Decision C (2014) 4995 of 22 July 2014. HORIZON 2020 LEIT ICT Work Programme. http://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/main/h2020-wp1415-leit-ict_en.pdf.

Executive Office of the President. 2009. Open Government Directive. http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf.

Ferro, E., and M. Osella. 2012. Business models for PSI re-use: A multidimensional framework. In *Using open data: Policy modeling, citizen empowerment, data journalism workshop*. Brussels: European Commission.

Ferro, E., and M. Osella. 2013. Eight business model archetypes for PSI re-use. In *Open data on the web workshop*. Shoreditch, London.

Graves, A., and J. Hendler. 2013. Visualization tools for open government data. In *Proceedings of the 14th Annual International Conference on Digital Government Research (dg.o '13)*, ACM, New York, NY, USA, 136–145. doi:10.1145/2479724.2479746.

Gregor, S. 2002. A theory of theories in information systems. In *Information systems foundations: Building the theoretical base*, eds. S. Gregor, and D. Hart, 1–20. Canberra, Australia: Australian National University.

Hansson, K., H. Verhagen, P. Karlstrom, and A. Larsson. 2013. Reputation and online communication: Visualizing reputational power to promote collaborative discussions. In *2013 46th Hawaii International Conference on System Sciences (HICSS)*, 748–758. IEEE.

Harrison, T. M., T. A. Pardo, and M. Cook. 2012. Creating open government ecosystems: A research and development agenda. *Future Internet* 4(4):900–928. doi:10.3390/fi4040900. 2012.

Hartley, J. 2005. Innovation in governance and public services: Past and present. *Public Money and Management* 25 (1):27–34. 2005.

Heipke, C. 2010. Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing* 65(6):550–557. ISSN 0924-2716. 10.1016/j.isprsjprs.2010.06.005.

Helbig, N., A. M. Cresswell, G. B. Burke, and L. Luna-Reyes. 2012. *The dynamics of opening government data, a white paper*. New York: Center for Technology in Government, University at Albany, State University of New York.

Hellerstein, J. M. 2008. *Quantitative data cleaning for large databases*. United Nations Economic Commission for Europe (UNECE).

HM Government. 2012. Open Data White Paper—Unleashing the Potential. http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf.

Houssos, N., B. Jörg, and B. Matthews. 2012. A multi-level metadata approach for a Public Sector Information data infrastructure. Paper presented at 11th International Conference on Current Research Information Systems, Prague, Czech Republic, June 6–9, 2012.

Jain, P., P. Hitzler, A. P. Sheth, K. Verma, and P. Z. Yeh. 2010a. Ontology alignment for linked open data. In *The semantic web–ISWC 2010*, 402–417. Berlin Heidelberg, Germany: Springer.

Jain, P., P. Hitzler, P. Z. Yeh, K. Verma, and A. P. Sheth. 2010b. Linked data is merely more data. In *Linked data meets artificial intelligence*. Technical Report SS-10-07, eds. D. Brickley, V. K. Chaudhri, H. Halpin, and D. McGuinness, 82–86. Menlo Park, California: AAAI Press.

Janssen, K. 2011a. Legal interoperability—Barriers to the harmonization of licenses. Paper presented at the ICRI - Share PSI workshop, Brussels, Belgium, May 2011.

Janssen, K. 2011b. The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly* 28:446–456. doi:10.1016/j.giq.2011.01.004.

Janssen, M., Y. Charalabidis, and A. Zuiderwijk. 2012. Benefits, adoption barriers and myths of open data and open government. *Information Systems Management* 29:258–268. doi:10.1080/10580530.2012.716740.

Jardim-Goncalves, R., A. Grilo, C. Agostinho, F. Lampathaki, and Y. Charalabidis. 2013. Systematisation of interoperability body of knowledge: The foundation for enterprise interoperability as a science. *Enterprise Information Systems* 7(1):7–32. doi:10.1080/17517575.2012.684401.

Jetzek, T., M. Avital, and N. Bjorn-Andersen. 2012. The value of open government data: A strategic analysis framework. In *Proceedings of SIG eGovernment pre-ICIS Workshop*, Orlando, USA.

Jetzek, T., M. Avital, and N. Bjorn-Andersen. 2013. The generative mechanisms of open government data. In *Proceedings of the 21st European Conference on Information Systems (ECIS)*, 156. Utrecht, The Netherlands.

Joshi, A. 2012. Challenges for adoption of secured effective e-governance through virtualization and cloud computing. In *Towards e-governance in the cloud*, ed. C. Unnithan and B. Fraunholz, 135–142.

Kalampokis, E., M. Hausenblas, and K. Tarabanis. 2011a. Combining social and government open data for participatory decision-making. In *Electronic participation*, eds. E. Tambouris, A. Macintosh, and H. Bruijn, vol. 6847, 36–47. Berlin, Heidelberg: Springer.

Kalampokis, E., E. Tambouris, and K. Tarabanis. 2011b. *Open government data: A stage model*. Berlin Heidelberg: Springer. doi:10.1007/978-3-642-22878-0_20.

Kalampokis, E., E. Tambouris, and K. Tarabanis. 2013. Linked open government data analytics. In *Electronic government*, eds. M. A. Wimmer, M. Janssen, and H. J. Scholl, 99–110. Berlin Heidelberg, Germany: Springer.

Kiryakov, A., B. Popov, I. Terziev, D. Manov, and D. Ognyanoff. 2004. Semantic annotation, indexing and retrieval. *Web Semantics: Science, Services and Agents on the World Wide Web* 2(1):49–79. doi:10.1016/j.websem.2004.07.005.

Kleijnen, S., and S. Raju. 2003. An open web services architecture. *Queue* 1(1):38–46. doi:10.1145/637958.637961.

Krippendorff, K. H. 2012. *Content analysis—An introduction to its methodology*, 3rd ed. Sage Publications.

Kum, H.-C., D. F. Duncan, and C. J. Stewart. 2009. Supporting self-evaluation in local government via knowledge discovery and data mining. *Government Information Quarterly* 26(2):295–304. doi:10.1016/j.giq.2008.12.009.

Kundra, V. 2012. *Digital fuel of the 21st century: Innovation through open data and the network effect*. Joan Shorenstein Center on the Press, Politics and Public Policy.

Lampathaki, F., Y. Charalabidis, S. Passas, D. Osimo, M. Bicking, M. Wimmer, and D. Askounis. 2010. Defining a taxonomy for research areas on ICT for governance and policy modelling. In *Electronic government*, Vol. 6228. eds. M. A. Wimmer, J.-L. Chappelet, M. Janssen, and H. J. Scholl, 61–72. Springer.

Lewis, G. A. 2013. Role of standards in cloud-computing interoperability. *System Sciences (HICSS), 2013 46th Hawaii International Conference on* 1652, 1661. January 7–10. doi:10.1109/HICSS.2013.470.

Lindman, J., M. Rossi, and V. K. Tuunainen. 2014. Open data services: Research agenda. *2013/01/01, 2014 47th Hawaii International Conference on System Sciences*. doi:10.1109/HICSS.2013.430.

Mannila, H. 2002. *Local and global methods in data mining: Basic techniques and open problems*. Berlin Heidelberg: Springer. doi:10.1007/3-540-45465-9_6.

McDermott, P. 2010. Building open government. *Government Information Quarterly* 27:401–413. doi:10.1016/j.giq.2010.07.002.

Mohr, L. B. 1969. Determinants of innovation in organizations. *The American Political Science Review* 63(1):111–126. doi:10.2307/1954288.

Mostafa, M. M., and A. A. El-Masry. 2013. Citizens as consumers: Profiling e-government services' users in Egypt via data mining techniques. *International Journal of Information Management* 33(4):627–641. 10.1016/j.ijinfomgt.2013.03.007.

Natarajan, K., J. Li, and A. Koronios. 2010. *Data mining techniques for data cleaning*. London: Springer.

O'Hara, K. 2011. Transparent government, not transparent citizens: A report on privacy and transparency for the Cabinet office, London, UK, 272–769.

Obama, B. 2012. Digital government. Building a 21st Century Platform to Better Serve the American People. http://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government.html.

Paolucci, M., T. Kawamura, T. R. Payne, and K. Sycara. 2002. *Semantic matching of web services capabilities*. Berlin Heidelberg: Springer. doi:10.1007/3-540-48005-6_26.

Parundekar, R., C. A. Knoblock, and J. L. Ambite. 2010. Linking and building ontologies of linked data. In *The semantic web–ISWC 2010*, 598–614. Berlin Heidelberg: Springer.

Pieterson, W., W. Ebbers, and L. Van Dijk. 2005. *The opportunities and barriers of user profiling in the public sector*. Berlin Heidelberg: Springer. doi:10.1007/11545156_26.

Ramos, L., and D. Rasmus. 2003. Best practices in taxonomy development and management. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.201.4848.

Richter, K. F., and S. Winter. 2011. *Citizens as database: Conscious ubiquity in data collection*. Berlin Heidelberg: Springer. doi:10.1007/978-3-642-22922-0_27.

Robertson, W. D., E. M. Leadem, J. Dube, and J. Greenberg. 2001. Design and implementation of the National Institute of Environmental Health Sciences Dublin core metadata schema. In *International Conference on Dublin Core and Metadata Applications*, 193.

Sarantis, D., Y. Charalabidis, and J. Psarras. 2008. Towards standardising interoperability levels for information systems of public administrations. *The Electronic Journal for E-commerce Tools & Applications (eJETA) Special Issue on Interoperability for Enterprises and Administrations Worldwide* 2.

Shukair, G., N. Loutas, V. Peristeras, and S. Sklarss. 2013. Towards semantically interoperable metadata repositories: The asset description metadata schema. *Computers in Industry* 64(1):10–18. doi:10.1016/j.compind.2012.09.003.

Sorrentino, S., S. Bergamaschi, E. Fusari, and B. Beneventano. 2013. *Semantic annotation and publication of linked open data*. Berlin Heidelberg, Germany: Springer. doi:10.1007/978-3-642-39640-3_34.

Sourouni, A. M., F. Lampathaki, S. Mouzakitis, Y. Charalabidis, and D. Askounis. 2008. Paving the way to eGovernment transformation: Interoperability registry infrastructure development. In *Electronic government*, 340–351. Berlin Heidelberg, Germany: Springer.

Statistical Office of the European Communities. 2005. *Oslo manual: Guidelines for collecting and interpreting innovation data (No. 4)*. Publications de l'OCDE.

Stevens, B. J. 1984. *Nursing theory. Analysis, application, evaluation*, 2nd ed. Boston: Little, Brown.

Sujatha, R., and B. R. K. Rao. 2011. Taxonomy construction techniques—Issues and challenges. *Indian Journal of Computer Science and Engineering* 3(5).

Swanson, E. B., and N. C. Ramiller. 1997. The organizing vision in information systems innovation. *Organization Science* 8:458–74. doi:10.1287/orsc.8.5.458.

Tammisto, Y., and J. Lindman. 2012. Definition of open data services in software business. In *Software business*, 297–303. Berlin Heidelberg, Germany: Springer.

UK Cabinet Office. 2011. Public Data Corporation to free up public data and drive innovation. https://www.gov.uk/government/news/public-data-corporation-to-free-up-public-data-and-drive-innovation.

Warner, J., and S. A. Chun. 2009. Semantic and pragmatic annotation for government information discovery, sharing and collaboration. In *Proceedings of the 10th annual international conference on digital government research: Social networks: Making connections between citizens, data and government*, 199–205. Digital Government Society of North America.

Windrum, P., and P. Koch. 2008. *Innovation in public sector services. Entrepreneurship, creativity and management*. Cheltenham, UK: Edward Elgar.

World Bank. 2013. Open data readiness assessment tool. Open Government Data Working Group. http://data.worldbank.org/sites/default/files/1/.

Yang, Z., and A. Kankanhalli. 2013. Innovation in government services: The case of open data. In *Grand successes and failures in IT. Public and private sectors*, 644–651. Berlin Heidelberg, Germany: Springer.

Zuiderwijk, A., B. Helbig, J. R. Gil-García, and M. Janssen. 2014a. Special issue on innovation through open data—A review of the state-of-the-art and an emerging research agenda: Guest editors' introduction. *Journal of Theoretical and Applied Electronic Commerce* 9(2):Talca May 2014. doi:10.4067/S0718-18762014000200001.

Zuiderwijk, A., and M. Janssen. 2014b. The negative effects of open government data - Investigating the dark side of open data. Proceedings of the 15th Annual International Conference on Digital Government Research, 147–152. doi:10.1145/2612733.2612761.

Zuiderwijk, A., and M. Janssen. 2014c. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly* 31(1):17–29. doi:10.1016/j.giq.2013.04.003.

Zuiderwijk, A. M. G., K. G. Jeffery, and M. F. W. H. A. Janssen. 2012. The necessity of metadata for open linked data and its contribution to policy analyses. In *Proceedings CeDEM2012 Conference*, 281–294.

Notes on contributors

Yannis Charalabidis is an assistant professor in the Department of Information and Communication Systems Engineering at the University of Aegean. In parallel, he serves as director of the Innovation and Entrepreneurship Unit of the University, designing and managing youth entrepreneurship activities, and head of the Information Systems Laboratory, coordinating policymaking, research, and pilot application projects for governments and enterprises worldwide. He has more than 20 years of experience in designing, implementing, managing, and applying complex information systems as a project manager in Greece and Europe. He has published more than 150 papers in international journals and conferences, while actively participating in international standardization committees and scientific bodies.

Charalampos Alexopoulos is a PhD candidate at University of the Aegean and Research Associate in the Information Systems Laboratory (ISL/AEGEAN-RU), working on high-level policymaking, research and pilot application FP7, and national projects (ENGAGE, SHARE-PSI 2.0, EU-COMMUNITY, PADGETS, NOMAD, NET-EUCEN, PLUG-IN). He is a computer science graduate from the University of Peloponnese with an MSc in management information systems from University of Aegean. Mr. Alexopoulos has published in scientific conferences and journals on IS evaluation, open data, semantic interoperability, and e-Government. He has also been a member of the organizing, reviewing, and scientific committees of several conferences (ICTDM, Samos Summit, IEEE SOLI, SSCORE, and ECIS).

Dr. Euripidis Loukis is an associate professor of Information Systems and Decision Support Systems at the University of the Aegean. Formerly, he has been Information Systems Advisor at the Ministry to the Presidency of the Government of Greece, technical director of the Program of Modernization of Greek Public Administration of the

Second Community Support Framework, and national representative of Greece in the programs "Telematics" and Interchange of Data between Administrations of the European Union. He is the author of numerous scientific articles in international journals and conferences. His current research interests include e-government, e-participation, information systems impact and internal/external determinants, business process adaptation and medical decision support systems.