Open government data

Areas and directions for research

Charalabidis, Yannis; Zuiderwijk, Anneke; Alexopoulos, Charalampos; Janssen, Marijn; Lampoltshammer, Thomas; Ferro, Enrico

DOI
10.1007/978-3-319-90850-2_9

Publication date
2018

Document Version
Final published version

Published in
Public Administration and Information Technology

Citation (APA)
Charalabidis, Y., Zuiderwijk, A., Alexopoulos, C., Janssen, M., Lampoltshammer, T., & Ferro, E. (2018). Open government data: Areas and directions for research. In *Public Administration and Information Technology* (pp. 173-194). (Public Administration and Information Technology; Vol. 28). Springer. https://doi.org/10.1007/978-3-319-90850-2_9

Chapter 9
Open Government Data: Areas and Directions for Research

"and still, so much to be done towards unveiling the true potential of open data."

9.1 Introduction

The concept of open data itself is strongly associated with innovative capacity and transformative power (Davies, Perini, & 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 digital services or mobile applications, by combining various types of Open Government Data (OGD), and possibly other private data. On the other hand, OGD also empowers scientists, journalists and active citizens who want to understand various public issues and policies through advanced data processing and production of analytics (Janssen, 2011; Zuiderwijk, Helbig, Gil-García, & Janssen, 2014).

Due to its recognised potential to generate public value through driving innovation and economic growth, the OGD movement has been attracting a growing attention and interest of both researchers and practitioners from various disciplines, such as information systems, management sciences, political and social sciences and law. Research on open data has also been targeting the promoting of transparency and the substantiation of evidence-based decision making in policy formulation (Conradie & Choenni, 2012; Janssen, 2011; Stevens, 1984). At the same time, a few articles discussing unintended consequences and negative side effects of opening data have started to appear (Blakemore & Craglia, 2006; Zuiderwijk & Janssen, 2014a).

OGD, as a rather 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

Y. Charalabidis et al., *The World of Open Data*, Public Administration and Information Technology 28, https://doi.org/10.1007/978-3-319-90850-2_9

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 the 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. 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.

The development of a detailed taxonomy of current research areas and topics in the domain of OGD, presented in this Chapter, as part of the work done in (Charalabidis, Alexopoulos, & Loukis, 2016), will address the communication gap in this new domain, and facilitate better interaction among researchers and interested practitioners. It can also provide a solid base for driving future research in this domain, and thus contribute to reaching higher levels of maturity in the practices of opening and exploiting government data, as well as in the generation of greater social and economic value. The research taxonomy can assist in the development of a body of knowledge in this area, which will enable improving and optimizing the technology, the service design elements, the operations and overall 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' (Charalabidis, Gonçalves, & Popplewell, 2011) in the OGD domain.

Research topics organisation is also extremely useful for Information and Communication Technology firms, assisting them in developing better OGD technological infrastructures, more innovative value added digital services or mobile applications based on OGD. This chapter contributes to filling the above-mentioned research gaps. In particular, it makes the following contributions:

(i) 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 analysed into 35 research topics.
(ii) It comprises multi-sourced knowledge extraction process. The development of this taxonomy includes the extraction and combination of relevant knowledge originated from three different kinds of sources: important relevant government policy documents, research literature and experts from research and practice.
(iii) It ascertains these 35 research topics summarizing relevant research literature for each one of them. The main research objectives and directions have been highlighted and under-researched topics that require further research have been identified.
(iv) 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 FP7 European projects CROSSROAD and CROSSOVER.

(v) Finally, directions have been formulated for future multi-disciplinary research based on OGD aiming to address current societal challenges.

Part of the research presented in this chapter has been conducted within the FP7 ENGAGE project "An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens".

The chapter is structured as follows: Section 9.2 describes the methodology we followed for developing the taxonomy. In Sect. 9.3 the main findings of the literature review we have conducted for this purpose are presented and discussed. Then Sect. 9.4 presents the taxonomy, including descriptions of the identified main research areas, and the particular research sub-areas/topics for each of them. Finally, a discussion of findings is provided in Sect. 9.5, while Sect. 9.6 concludes the chapter.

9.2 Taxonomy Design Methodology

This study is focused on two main research questions, which constitute a first step towards 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) proposes 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 chapter 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, based on relevant government policy documents, previous relevant research literature and also 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 Fig. 9.1 (our research has focused on the first three of them).

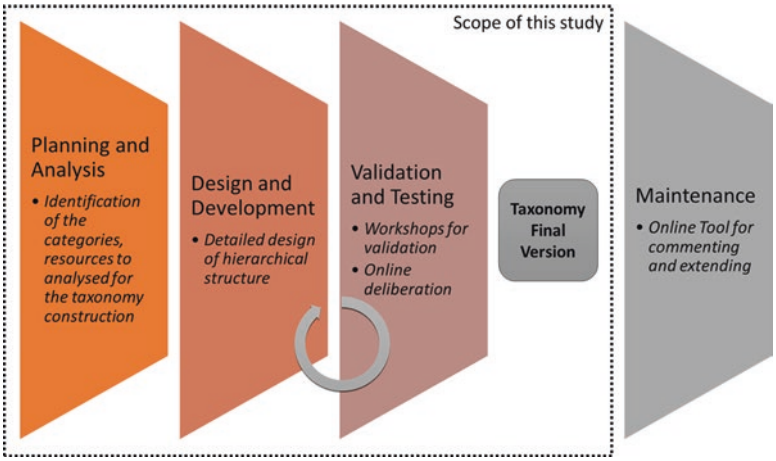

Fig. 9.1 The Open Data Research Taxonomy development approach

In particular, the methodology we followed for the development of the taxonomy was based on content analysis (Krippendorff, 2013) of different kinds of documents (government policy documents, previous relevant research literature and minutes of experts' workshops). It consisted of the following eight steps (shown also in Fig. 9.2):

1. Initially we identified and analysed 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, 2010a, 2011b, 2011d, 2012, 2013b, 2013d),
   (b) US Government documents (Executive Office of the President, 2009; Obama, 2012b),
   (c) UK Government documents (HM Government, 2012; O'Hara, 2011; UK Cabinet Office, 2011), 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 the taxonomy, in step three.

2. Then we identified and analysed previous research papers that propose categorisations of research areas and perspectives of the OGD research domain. Additionally, we identified and analysed previous research literature concerning barriers to OGD publishing and exploitation, and also uptake of OGD and value generation from them. A brief review of this literature is presented in the following Sect. 9.3 The outcome of this step was another set of OGD related terms

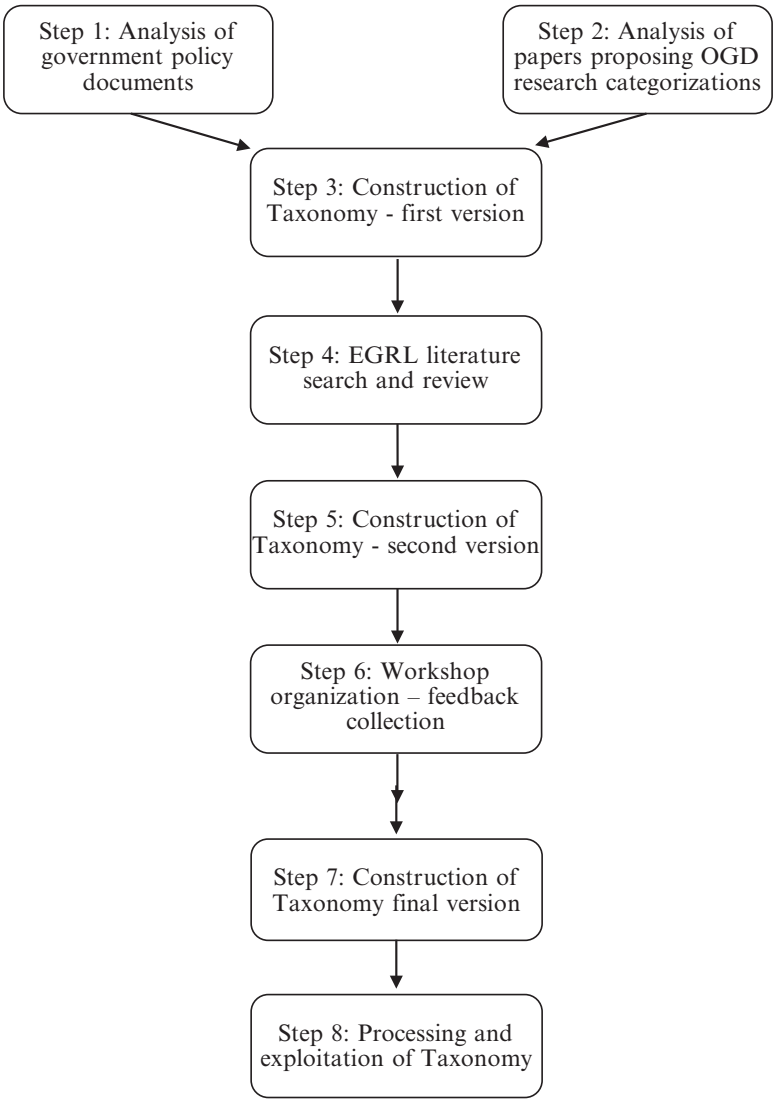

Fig. 9.2 Steps of the development methodology

(having some overlap with the ones of the set produced in the previous step), which were used as well for the construction of the first version of the taxonomy in step three.

3. After realising 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 Open Data Research Taxonomy.

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 above first version of the OGD Research Areas Taxonomy. 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 taxonomy.
5. The realisation of the fourth step resulted in the second version of the taxonomy.
6. A workshop was organised for the discussion, evaluation and validation of the above second version of the taxonomy, 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. Twenty OGD experts participated in this workshop in order to validate and further elaborate the second version of the taxonomy. These experts came from 11 different EU countries (NL, UK, DE, GR, BE, IT, AU, RO, ES, BG, LV), from different kinds of organizations (public administrations, universities and firms) and had different educational levels (Professors, PhD and MSc holders). All of the participants were selected based on their experience in the area of OGD and they are characterised 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 topics 2.7 ('citizen-generated open data') and 2.8 ('sensor-generated open data') described in Sect. 9.4, and also of changes in their grouping in research areas), the final version of the taxonomy was produced, which is presented in Sect. 9.4.
8. Finally we proceeded to further processing and exploitation of it, and the results are presented in Sect. 9.5.

9.3 Background and Literature Review

During step 2 of the methodology as described in the previous section, we have identified four previous research papers that propose categorisations of OGD research in areas and themes (Davies et al., 2013; Harrison, Pardo, & Cook, 2012; Lindman, Rossi, & Tuunainen, 2014; Zuiderwijk et al., 2014), which were reviewed as they include elements that can be useful for the development of the Open Data Research Taxonomy.

The study of Davies et al. (2013, p.11) argues 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 exists 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 et al. (2014, p.2) identifies seven different perspectives of OGD research, namely, (a) political, (b) social, (c) economical, (d) institutional, (e) operational, (f) legal and (g) technical and argues that *"combining perspectives may be more effective in dealing with the issues related to open data and stimulating innovation"*. Furthermore, it also identifies a number of OGD research directions, and categorises them under three major topics: (i) open data theory and development, (ii) open data policies, use, and innovation, and (iii) open data infrastructures and technologies.

Another study conducted by Lindman et al. (2014 p.4) focuses on the research challenges concerning Open Data Services, and categorises the relevant issues based on the work systems framework (Alter, 2010). It argues 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 organisation of open data services research are: (1) Technologies, (2) Information, (3) Processes and Activities, (4) Products and Services, (5) Participants, (6) Customers and (7) Environment; each of them includes several research questions.

Finally, the study of Harrison et al. (2012, p.23) examines 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 define the main themes of the research required in order to realise 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 (Barry & Bannister, 2014; Conradie & Choenni, 2012; Janssen, 2011; Janssen, Charalabidis, & Zuiderwijk, 2012; McDermott, 2010). We reviewed this research stream, as 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 the taxonomy. 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 & Koch, 2008; Yang & Kankanhalli, 2013). The main conclusions of this stream of research indicate that the uptake and use of the OGD, and also the genera-

tion of innovation and value in general from them, are not straightforward, being complex, and requiring the collaboration of several actors.

From the above literature review we conclude that although there are some previous studies that propose categorisations of OGD research in areas and themes, they are at a too high level, and lack the detail required for directing future research. In order to provide the development of a 'science base' in this domain, we have to facilitate a better interaction among researchers and interested practitioners. Our research, as mentioned in the Introduction, contributes to filling this gap.

9.4 The Open Government Data Research Taxonomy

The Open Government Data Research Taxonomy 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 Fig. 9.3), which include 35 research topics (in the second level). These 35 identified research topics were initially divided into two categories: the technological and non-technological ones; the latter correspond to the abovementioned OGD Usage and Value research area.

By examining the former we distinguished two clear sub-groups of research topics, concerning the interoperability and the management of the OGD respectively, which lead 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 above four research areas has been confirmed by the experts who participated in the workshop mentioned in the 'Methodology' Sect. 9.2. Changes were also proposed for some research topics and the research area they were associated with. The full taxonomy is available for reviewing and commenting online at mind42.com mind-mapping service[1].

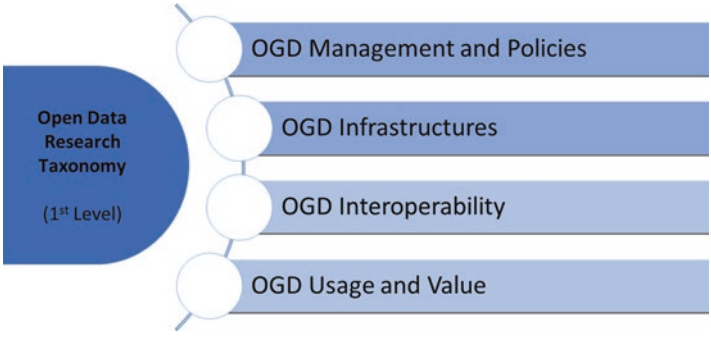

Fig. 9.3 Top-level open data research areas

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

9.4.1 OGD Management and Policies

The first top-level research area of the taxonomy 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 anonymisation, 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 Fig. 9.4 we can see the research topics of the 'OGD Management and Policies' research area, while in Table 9.1 these OGD research topics are described in more detail, supported by some representative relevant literature from the EGRL.

Fig. 9.4 Research topics for the OGD Management & Policies research area

Table 9.1 Description of the research topics of OGD Management & Policies research area

Research topic	Description
1.1 Policy & 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 & Craglia, 2006; European Commission, 2013b, 2013d; Zuiderwijk & Janssen, 2014b). 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 Anonymisation 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 anonymisation of opened data. Privacy-preserving data publishing (PPDP) provides methods and tools for publishing useful information while preserving data privacy (Fung, Wang, Chen, & Yu, 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, & 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 Visualisation 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 & Hendler, 2013).
1.6 OGD Linking	The principles, frameworks, techniques and tools for OGD linking are the subjects of this research area (Bojārs, Breslin, Finn, & Decker, 2008; Kalampokis, Tambouris, & Tarabanis, 2013). 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, & 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 kinds of data.

(continued)

Table 9.1 (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 et al., 2009; Dawes & Helbig, 2010; Helbig, Cresswell, Burke, & Luna-Reyes, 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 (Bakirl et al., 2012; Mostafa & El-Masry, 2013).
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 (Zuiderwijk, 2015a). 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 above purposes are also investigated in this research area, e.g. through web 2.0 oriented mechanisms (Alexopoulos, Zuiderwijk, Loukis, & Janssen, 2014; Charalabidis, Loukis, & Alexopoulos, 2014b).

9.4.2 OGD Infrastructures

The second research area of the Taxonomy 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 personalisation capabilities; another important research topic is OGD storage and long – term preservation, and also the use of cloud services in this domain.

Furthermore, though the main source of OGD is the information systems of government agencies, two more sources are gradually emerging, sensors and citizens, so researching them and their exploitation is an important research challenge. In Fig. 9.5 we can see the research topics of the 'OGD Infrastructure' research area, while in Table 9.2 these OGD research topics are described in more detail, supported also with representative literature from the EGRL.

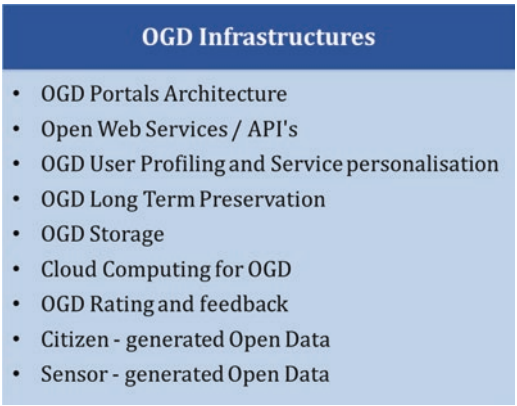

Fig. 9.5 Research topics for the OGD Infrastructures research area

Table 9.2 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, 2016; Charalabidis et al., 2014b; Helbig et al., 2012). Various types and generations of architectures are proposed and discussed from various perspectives. Additionally, some research is conducted concerning the development of architectures of ICT infrastructures that allow for and support application development utilising 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, Eberius, Thiele, & Lehner, 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 (Kleijnen & Raju, 2003; Paolucci, Kawamura, Payne, & Sycara, 2002).
2.3 OGD User Profiling and Service personalisation	This research focuses on user profiling, which can offer big opportunities to make OGD related services more personalised, to infer and predict citizens' behaviour, and to even influence their behaviour (Pieterson, Ebbers, & Dijk, 2005). Like the private sector, the public sector makes more and more use of user profiling in order to personalise the electronic services that are being offered to citizens (Mostafa & 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 & 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.

(continued)

Table 9.2 (continued)

Research topic	Description
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 (Jain, Hitzler, Sheth, Verma, & Yeh, 2010; Jain, Hitzler, Yeh, Verma, & Sheth, 2010; Sorrentino, Bergamaschi, Fusari, & Beneventano, 2013).
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 encyclopaedia, or OpenStreetMap[b] as an open topographic dataset of the world (Richter & Winter, 2011) and Zooniverse[c] platform for people-powered research (many individual volunteers, relying on a version of the 'wisdom of crowds' to produce reliable and accurate data). Open data is 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 is becoming of critical importance for science and commercial applications development (e.g. Elgendy & Elragal, 2014b), 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/
[c] https://www.zooniverse.org/

9.4.3 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, Grilo, Agostinho, Lampathaki, & Charalabidis, 2013). Interoperability has many aspects, mainly technical, semantic and organisational. 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, Loutas, Peristeras, & Sklarss, 2013, p.10).

So our third research area deals with the interoperability issue in the specific domain of OGD. It includes research topics concerning OGD metadata, semantic

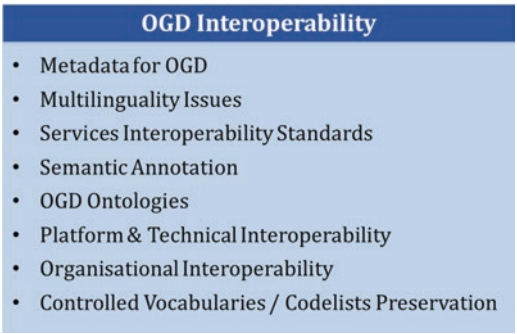

Fig. 9.6 Research topics for the OGD Interoperability research area

annotation, ontologies and controlled vocabularies and codelists, and also on OGD platforms technical interoperability, services interoperability standards and organizational interoperability. In Fig. 9.6 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 9.3.

9.4.4 OGD Usage and Value

The fourth research area of the research areas taxonomy is directed towards 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 on one hand OGD needs, readiness, use, skills management and reputation management, and on the other hand OGD related value and impact, innovation, entrepreneurship and contribution to accountability/transparency. In Fig. 9.7 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 9.4.

9.5 Discussion

In this section the outcomes of the further processing and exploitation of the Research Areas Taxonomy are presented, conducted finally as part of the step eight of our research methodology (see Sect. 9.2): analysis of EGRL publications for each of the identified research topics (Sect. 9.5.1); exploitation of the Taxonomy for OGD Science Base Creation (Sect. 9.5.2); association of OGD Research Areas Taxonomy with the ICT-enabled Governance research taxonomy developed in the CROSSROAD and the CROSSOVER projects, and also use of the former in order to extend the latter (Sect. 9.5.3); and formulation of direction for multi-disciplinary research on important societal challenges using OGD (Sect. 9.5.4).

Table 9.3 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 sub-topics: 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, & Askounis, 2009; Robertson, Leadem, Dube, & Greenberg, 2001; Zuiderwijk, Jeffery, & Janssen, 2012b).
3.2 Multi-linguality	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 towards the support of multi-linguality in the domain of OGD (Houssos, Jörg, & Matthews, 2012).
3.3 Service Interoperability Standards	This research topic 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, Lampathaki, Mouzakitis, Charalabidis, & Askounis, 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 et al., 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 (Dill et al., 2013; Kiryakov, Popov, Terziev, Manov, & Ognyanoff, 2004; Warner & Chun, 2009). The major objective of this research is the development of algorithms and tools for semantic integration (Bergamaschi, Castano, & 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, & Ambite, 2010). Ontologies for the description and use of OGD, as well as the sense of ontology alignment are under investigation in this research (Osterwalder & Pigneur, 2010; Jain, Hitzler, Sheth et al., 2010; Jain, Hitzler, Yeh et al., 2010). 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) (Jardim-Goncalves et al., 2013; Sarantis, Charalabidis, & Psarras, 2008).

(continued)

Table 9.3 (continued)

Research topic	Description
3.7 Organisational 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 OGD service-related goals (Jardim-Goncalves et al., 2013; Sarantis et al., 2008), 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).

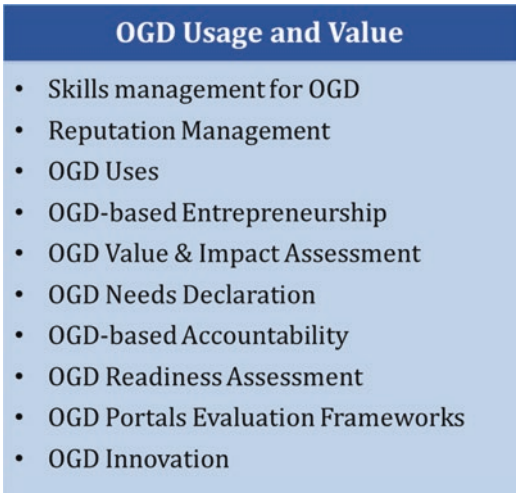

Fig. 9.7 Research topics for the OGD Usage and Value research area

Table 9.4 Description of the research topics of 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 the investigation of the use of reputation systems in the OGD value chain. It examines various algorithms and methods for the reputation management of various OGD stakeholders (Bani & Paoli, 2013; Hansson, Verhagen, Karlstrom, & Larsson, 2013).

(continued)

Table 9.4 (continued)

Research topics	Description
4.3 OGD Use	It includes studies that describe and analyse 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 the government (e.g. for policy making: Kalampokis, Hausenblas, and Tarabanis (2011), Kalampokis, Tambouris, and Tarabanis (2011b) 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 & Osella, 2012, 2013).
4.5 OGD Value and Impact Assessment	The current OGD research on this topic focuses on analysing OGD initiatives that have led to the generation of some kind of public value (Charalabidis et al., 2014b; Davies et al., 2013; Jetzek, Avital, & Bjorn-Andersen, 2012, 2013), analysing the positive – and sometimes also the negative – 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, 2015a).
4.7 OGD-based Accountability	This research investigates the use of OGD as part of anti-corruption programmes, 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 (Alon, 2011; Böhm et al., 2012b).
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 et al., 2013; World Bank, 2013b).
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 (Alexopoulos, 2016; Charalabidis et al., 2014b; Kalampokis et al., 2011).
4.10 OGD Innovation	The main objective of this OGD research is to identify and analyse innovations driven by OGD, both in the private sector (e.g. e-services innovations), and in the public sector (Zuiderwijk et al., 2014). According to this literature, OGD innovation concerns mainly three domains: (a) research, (b) business and (c) transparency (Jetzek et al., 2012, 2013). While the US literature and practice focuses mainly on (b), the EU tends to focus on (a), but both of them are equally interested in (c) OGD promotion towards transparency.

9.5.1 EGRL Publications for Research Topics

For all the OGD research areas identified and presented in the previous section (of the final version of the Taxonomy produced in step seven, see our methodology (Sect. 9.2)) we searched for relevant publications in the EGRL. In Fig. 9.8 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 relevant publications: for some of them we have found more publications, e.g. for research topics concerning OGD use, portals evaluation frameworks, publishing, policy and legal issues. For some others we found significantly less or even no publications, e.g. for 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. These topics were proposed in the workshop (step six of our OGD Research Areas Taxonomy development methodology by the experts who participated as major issues of OGD). Also, from Fig. 9.8 we can conclude 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.

9.5.2 Contribution to OGD Science Base Creation

As mentioned in Sect. 9.2, the research presented in this chapter contributes to the development of 'description theory' for the OGD domain, so it constitutes the first step towards the creation of a Science Base for it. According to Charalabidis, Gonçalves, and Popplewell (2010) 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 produced a body of knowledge in the domain, and also various types of 'proofs of concept', aiming all to assist practitioners in this domain to solve particular problems and generate value.

Our OGD Research Areas Taxonomy contributes in the above-mentioned directions, as (i) it identifies the main concepts, methods and tools in OGD, and (ii) 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.

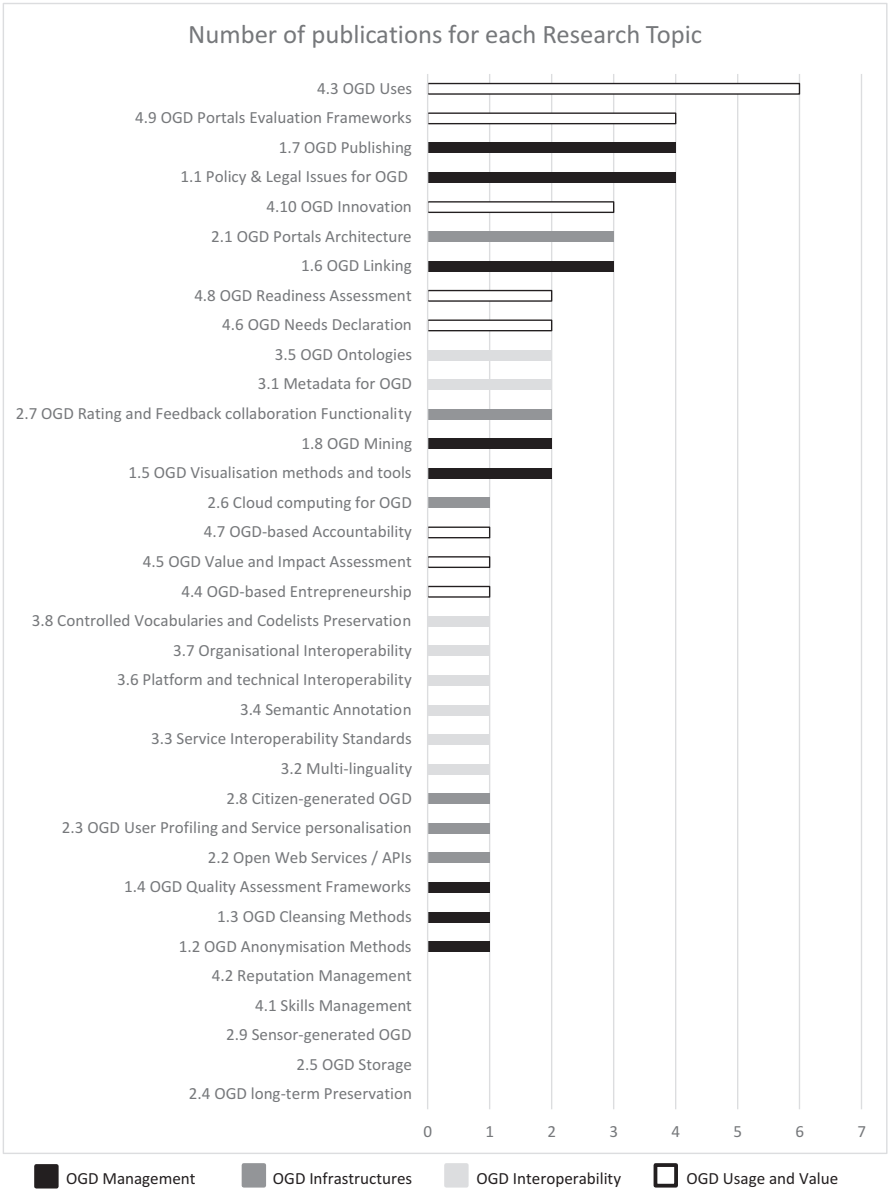

Fig. 9.8 Ranking of OGD research topics based on EGRL relevant literature

9.5.3 Extension of ICT-Enabled Governance Taxonomy

The OGD Research Areas Taxonomy is associated with and extends/elaborates the ICT-enabled Governance research taxonomy developed in the CROSSROAD[2] and the CROSSOVER[3] 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 sub-areas (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 Analytics". The OGD Research Areas Taxonomy extends and elaborates this research theme, as the main research areas and topics of the former can replace the research areas and sub-areas 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 categorises these research challenges under two research themes: (a) Data-powered Collaborative Governance and (b) Policy Modelling, in order to develop a roadmap on policy making 2.0. The OGD Research Areas Taxonomy extends and elaborates the "Linked Open Government Data" research challenge of the "Data-powered Collaborative Governance" theme.

9.5.4 Multi-disciplinary 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 multi-disciplinary research on important societal problems and challenges that modern societies face. These data can be used by multi-disciplinary scientific teams, e.g. including members from various 'neighbouring scientific domains', such as economic, political, social, management and behavioural 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 multi-disciplinary research were mentioned, and are summarized in Table 9.5.

[2] http://www.2020-horizon.com/CROSSROAD-CROSSROAD-A-Participative-Roadmap-for-ICT-Research-in-Electronic-Governance-and-Policy-Modelling(CROSSROAD)-s9412.html

[3] http://www.crossover-project.eu/ResearchRoadmap.aspx

Table 9.5 Directions of Multi-disciplinary Research on Societal Challenges Based on OGD

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

9.6 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. The Open Government Data Research Taxonomy 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 Fig. 9.3), which include 35 research topics (in the second level).

These 35 identified research topics have been validated through their association with relevant literature from the EGRL, as well as their importance to the experts of the workshop. The validation steps enabled a better understanding of them and their main research objectives and directions. Our OGD research taxonomy has been also 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 multi-disciplinary 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 non-technological 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 com-

plexity 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. anonymisation, 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.

Our study has revealed significant differences among the above 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 Sect. 9.2). Finally, it identifies important under-researched topics, on which further research is required. With respect to practice, the OGD Research Areas Taxonomy 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, the OGD Research Areas Taxonomy is useful to ICT firms developing OGD technological infrastructures, as it provides them directions for improving their products and services.

As the domain is evolving, it is necessary to organize more workshops in order to further validate the OGD Research Areas Taxonomy, 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 through 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. Finally, it would be interesting to exploit other research libraries except EGRL and the multiple OGD research projects which are currently in progress (e.g. supported by European Commission or USA research programs) towards a better understanding of the implications in each research topic.