UNIVERSITY OF THE AEGEAN
DEPARTMENT OF INFORMATION AND COMMUNICATION SYSTEMS ENGINEERING

DOCTORAL DISSERTATION

OPEN GOVERNMENT DATA INFRASTRUCTURES: RESEARCH CHALLENGES, ARTEFACTS DESIGN AND EVALUATION

CHARALAMPOS ALEXOPOULOS

SAMOS, NOVEMBER 2016
OPEN GOVERNMENT DATA INFRASTRUCTURES:
RESEARCH CHALLENGES, ARTEFACTS DESIGN AND EVALUATION

SUPERVISION COMMITTEE:
COMMITTEE CHAIR:

YANNIS CHARALABIDIS, ASSOCIATE PROFESSOR,
UNIVERSITY OF THE AEGEAN

COMMITTEE MEMBERS:

EURIPIDIS LOUKIS, ASSOCIATE PROFESSOR,
UNIVERSITY OF THE AEGEAN

MANOLIS MARAGKoudakis, ASSISTANT PROFESSOR,
UNIVERSITY OF THE AEGEAN

EXAMINATION COMMITTEE:
COMMITTEE CHAIR:

YANNIS CHARALABIDIS, ASSOCIATE PROFESSOR,
UNIVERSITY OF THE AEGEAN

COMMITTEE MEMBERS:

EURIPIDIS LOUKIS, ASSOCIATE PROFESSOR, UNIVERSITY
OF THE AEGEAN

DIMITRIOS-FRAGKISkos LEKKAS, ASSOCIATE
PROFESSOR, UNIVERSITY OF THE AEGEAN

CONSTANTINOS KOUTRAS, ASSOCIATE PROFESSOR,
UNIVERSITY OF PELOPONNESE

DIMITRIOS ASKOUNIS, ASSOCIATE PROFESSOR,
NATIONAL TECHNICAL UNIVERSITY OF ATHENS

MANOLIS MARAGKoudakis, ASSISTANT PROFESSOR,
UNIVERSITY OF THE AEGEAN

DIMITRIOS DROSOS, ASSISTANT PROFESSOR, UNIVERSITY
OF THE AEGEAN
ABSTRACT

Numerous government agencies worldwide are making big investments for developing information systems that open data they possess to the society, in order to be used for scientific, commercial and political purposes. It is therefore highly important to develop advanced open government infrastructures, which not only publish government data, but also provide support for the individual and collaborative value generation from them. It is also necessary, for both the ‘traditional’ and the advanced open government data infrastructures, to understand what value they create and how, and at the same time – since this is a relatively new type of information systems – to identify the main improvements they require, as well as, the infrastructure development priorities. Filling this research gap and following the Design Science research methodology five research questions were formulated further evolving this research field. A thorough literature review and a taxonomy creation conclude the main research areas of the Open Government Data (OGD) domain. Continuously, a model for a desk-based research was developed in order to analyse the current landscape of OGD infrastructures. Following the results of the above studies, a scenario-based design was applied in order to identify the requirements of the next generation infrastructures. Moreover, an evaluation framework and a value model have been developed driving to the next versions of the infrastructure. Finally, a new platform was realised and applied to the Greek context maximising the value for Collaborative and Individual use of OGD. Addressing the five basic research questions of this dissertation, different issues accrued and handled that are of major importance for the development of an OGD infrastructure. These issues have been discussed in the conclusions and were assimilated into the greater domain of OGD research articulating the future research.
ΠΕΡΙΛΗΨΗ ΣΤΑ ΕΛΛΗΝΙΚΑ

Πληθώρα κυβερνήσεων και κυβερνητικών οργανισμών σε όλο τον κόσμο πραγματοποιούν μεγάλες επενδύσεις για την ανάπτυξη πληροφοριακών συστημάτων που βοηθούν στη δημοσίευση των δεδομένων που έχουν στην κατοχή τους ώστε να τα καταστήσουν κοινώς προσβάσιμα, προκειμένου να χρησιμοποιούνται για επιστημονικούς, εμπορικούς και πολιτικούς σκοπούς. Επομένως, κρίνεται πολύ σημαντική η ανάπτυξη προηγμένων υποδομών ανοιχτών κυβερνητικών δεδομένων, οι οποίες όχι μόνο θα δημοσιεύουν τα στοιχεία της κυβέρνησης, αλλά και θα παρέχουν τις κατάλληλες δυνατότητες για την παραγωγή ατομικής και συλλογικής αξίας από αυτά. Είναι επίσης απαραίτητο για την ανάπτυξη, τόσο των «παραδοσιακών» όσο και των προηγμένων υποδομών ανοιχτών κυβερνητικών δεδομένων, να κατανοηθεί η αξία που δημιουργείται από τη δημοσίευση τέτοιων δεδομένων και συγχρόνως πώς, δεδομένου ότι αυτό είναι ένα σχετικά νέο είδος των πληροφοριακών συστημάτων, μπορούν να προσδιοριστούν οι κόρες βελτιώσεις και οι προτεραιότητες κατά την ανάπτυξη των υποδομών αυτών. Προσπαθώντας να συμπληρώσουμε αυτό το ερευνητικό κενό και ακολουθώντας τη μεθοδολογία ανάπτυξης και σχεδιασμού επιστημονικών αντικειμένων (Design Science Research Methodology), πέντε ερευνητικά ερωτήματα διατυπώθηκαν με στόχο την περαιτέρω εξέλιξη αυτού του πεδίου έρευνας. Μια λεπτομερής επισκόπηση της βιβλιογραφίας και η δημιουργία μιας ταξινομίας είχε ως αποτέλεσμα την αναγνώριση των βασικών ερευνητικών περιοχών του τομέα (Open Government Data - OGD). Εν συνεχεία, πραγματοποιήθηκε η ανάπτυξη ενός μοντέλου με σκοπό την ανάλυση των σύγχρονων υποδομών ανοιχτών κυβερνητικών δεδομένων καθώς και η ανάλυση των μεγαλύτερων υποδομών παγκοσμίως. Μετά τα αποτελέσματα των παραπάνω μελετών, ακολούθησε ο προσδιορισμός των απαιτήσεων βασισμένος στην μεθοδολογία ανάπτυξης σεναρίων χρήσης και η υλοποίηση της υποδομής νέας γενιάς. Επιπλέον, ένα πλαίσιο αξιολόγησης αναπτύχθηκε και εφαρμοστήκε της με στόχο την εφαρμογή της στο ελληνικό τοπίο, ώστε να ενισχύσει την παραγωγή αξίας και καθοδηγούμενη η μελλοντική έρευνα.
Acknowledgements

During the drafting of this dissertation I had the opportunity to cooperate with many people without their assistance this thesis could not be completed. This great knowledge travel towards the exploitation of open data in various levels evolve my analytical and research skills. At this point I would like to take the opportunity to convey my thanks to the efforts of those people who helped and supported me at various stages in the completion of this work.

First of all, I would like to thank my Supervisor, Associate Professor Yannis Charalabidis, for his valuable guidance and tireless support all these years. Many thanks to the members of my Supervision Committee, Associate Professor Euripidis Loukis and Assistant Professor Manolis Maragkoudakis for their substantial support.

Special thanks to my colleagues of the ENGAGE project for the great collaboration and fruitful insights. This dissertation is partially related to the ENGAGE FP7 Infrastructure Project (An Infrastructure for Open, Linked Governmental Data Provision towards Research Communities and Citizens), that started in June 2011 (www.engage-project.eu and www.engagedata.eu).

Another great “Thank you” to my family – birth and in-law (Dimitris, Voula, George, Helen, Dimitris, Voula, Efthimia, Christos) – for their ubiquitous support, enablement and encouragement. Last but not least, I express my gratitude to my colleagues – friends Aggeliki Androutsopoulou and Vasiliki Diamantopoulou, since without their support, I would not have been able to complete my work.

This dissertation is dedicated to Penny.
Table of Contents

1. Introduction .................................................................................................................. 16
   1.1. The Public Sector Information Directive .............................................................. 17
   1.2. The Digital Agenda 2020 ...................................................................................... 19
   1.3. Research Questions ............................................................................................... 21
2. Overall Methodology ..................................................................................................... 23
3. Background and literature review ................................................................................... 26
   3.1. A taxonomy of open government data research areas ............................................. 26
   3.2. Methodology for creating the OGD Research Areas Taxonomy ......................... 26
   3.3. Literature Review .................................................................................................. 30
   3.4. The OGD Research Areas Taxonomy ..................................................................... 32
      3.4.1. OGD Management and Policies ................................................................. 33
      3.4.2. OGD Infrastructures .................................................................................. 37
      3.4.3. OGD Interoperability ............................................................................... 40
      3.4.4. OGD Usage and Value ............................................................................ 42
   3.5. Answering Research Question 1 ............................................................................. 45
      3.5.1. Limitations .................................................................................................... 47
4. Current State of Play in OGD infrastructures ................................................................ 49
   4.1. Background and key concepts ................................................................................ 51
      4.1.1. Open Government Data recent history .......................................................... 51
      4.1.2. OGD Thematic Categories ........................................................................... 54
      4.1.3. OGD Sources Capabilities/Functionalities .................................................... 55
      4.1.4. Semantic Technologies and OGD ............................................................... 56
   4.2. Research Method .................................................................................................... 57
   4.3. Open Data Sources ................................................................................................ 62
   4.4. Results ..................................................................................................................... 66
      4.4.1. Functional Perspective Analysis .................................................................... 66
      4.4.2. Semantic and Data Perspective Analysis ...................................................... 67
      4.4.3. Technological Perspective Analysis .............................................................. 67
   4.5. Answering Research Question 2 ............................................................................. 67
5. Defining and Designing the Solution ............................................................................. 72
   5.1. Methodological Approach ....................................................................................... 74
5.1.1. Scenario Based Design ................................................................. 74
5.1.2. Description Template ................................................................. 75

5.2. Usage Scenarios .............................................................................. 78
5.2.1. Storing, Linking, Annotating and Visualising a PS data set ................. 79
5.2.2. Getting the metadata specifications ............................................... 84
5.2.3. Getting useful information (through browsing datasets or visualisations) ..... 88
5.2.4. Getting data for my research work ............................................... 93
5.2.5. Linking my system for uploading data ........................................... 96
5.2.6. Linking my system for downloading data ....................................... 102
5.2.7. Storing data in draft form (in the buffer) ....................................... 107
5.2.8. Put data and annotate them according to the platform standard and my needs 109
5.2.9. Put my needs for Public Sector data ............................................. 112
5.2.10. Get information and training on metadata, open data, platform ........... 115

5.3. Validation and Rating of the Results ............................................... 118

5.4. Answering Research Question 3 ..................................................... 120
5.4.1. Social Media Use in Governance .................................................. 120
5.4.2. Closing the feedback loop ......................................................... 121
5.4.3. Users and Final Requirements ..................................................... 123

6. Evaluation of an OGD e-infrastructure ............................................ 127

6.1.1. Theories for evaluation ............................................................. 127
6.1.2. Methodical concepts for performing evaluations ............................ 132
6.1.3. Review on Evaluation Models .................................................... 134
6.1.4. Conclusions on Evaluation Models ............................................. 147

6.2. Designing an Evaluation Framework ............................................. 150
6.2.1. Evaluation procedure ............................................................... 152
6.2.2. Application on the ENGAGE platform (version 1) ......................... 154
6.2.2.1. External Evaluation ............................................................... 154
6.2.2.2. Internal Evaluation ............................................................... 168
6.2.2.3. Web Analytics and Metrics .................................................... 174
6.2.3. Summary of results ................................................................. 184

6.3. A value model for ODG e-Infrastructures ...................................... 186
6.3.1. Value Model Definition ............................................................. 186
6.3.2. Value Model Estimation Algorithm ............................................ 191
6.3.3. Model Validation and Results .................................................... 193
6.3.4. Conclusions from the value model application ........................................... 200

6.4. Answering Research Question 4 .......................................................................... 201

7. 3rd Generation of OGD Platforms ........................................................................... 202

7.1. Literature Review ................................................................................................. 202
  7.1.1. Market Places ................................................................................................. 202
  7.1.2. Open Data Platforms ....................................................................................... 203
  7.1.3. Open Web Services .......................................................................................... 204
  7.1.4. Current status in open data and open services repositories ............................... 206

7.2. Research approach ............................................................................................... 208

7.3. Identifying developments influencing open data marketplaces ......................... 209
  7.3.1. Electronic marketplaces .................................................................................... 209
  7.3.2. Open data platforms and marketplaces ............................................................. 210

7.4. Organizing developments: expert discussions ..................................................... 212

7.5. Answering Research Question 5 .......................................................................... 213
  7.5.1. Identifying elements for open data marketplaces ............................................ 213
    7.5.1.1. Bring stakeholders together (match supply and demand) ....................... 214
    7.5.1.2. Provide rich metadata ............................................................................... 215
    7.5.1.3. Enable data quality assessment ................................................................. 216
    7.5.1.4. Ensure trust, security and critical mass ...................................................... 217
    7.5.1.5. Have an appropriate revenue model ......................................................... 217
    7.5.1.6. Provide use cases, training and support ..................................................... 218
    7.5.1.7. Provide technical support: Open data processing tools .............................. 218
    7.5.1.8. Provide a full API for machine-to-machine operation ............................... 219
    7.5.1.9. Target multiple nationalities ...................................................................... 219

  7.5.2. Realising an Open Data and Open Services Marketplace for Greece ............ 219
    7.5.2.1. The proposed Architecture ....................................................................... 219
    7.5.2.2. Design and Development ......................................................................... 223
    7.5.2.3. Further Development Steps ....................................................................... 226

8. Conclusions ............................................................................................................... 229

  8.1. Which are the main research challenges and opportunities for OGD and how we can map them? ................................................................. 230

  8.2. What is the current state-of-play in the OGD infrastructures development? .... 231

  8.3. What are the requirements and features for a new more advanced generation of OGD platform? .............................................................. 232

  8.4. Which are the existing evaluation models and how to evaluate an OGD Initiative? ........................................................................................................ 236
8.5. How to maximise value for Collaborative and Individual use of OGD and how to apply it to the Greek context? ................................................................. 239

8.6. Discussion and further research ................................................................. 241
8.6.1. An extended OGD Life-Cycle .................................................................. 242
8.6.2. Open Data Stakeholders ........................................................................ 245
8.6.3. A maturity model for OGD e-Infrastructures .......................................... 248
8.6.4. Multi-disciplinary Research Based on OGD ............................................. 251

Appendix A: References ..................................................................................... 252

Appendix B: Author’s Publications ................................................................. 267

B1. Journal papers ......................................................................................... 267
B2. Conference Papers .................................................................................. 267
B3. Workshop papers ..................................................................................... 267
B4. Chapters in Books ................................................................................... 268
B5. Book Authorships .................................................................................... 268

Appendix C: Scenarios based Design Template ............................................ 269

Appendix D: List of European Open Government Data Initiatives ............... 271

Appendix E: List of Greek Open Government Data Sources ..................... 275
1. Introduction

Open Government Data (OGD) 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, due to its widely recognised potential to generate public value through driving innovation and economic growth, and also scientific research, and by promoting transparency and substantial evidence-based political dialogue (Stevens 1984; Conradie et al. 2012; Janssen K. 2011a). The concept of open data is strongly associated with innovative capacity and transformative power (Davies et al. 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 K. 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 et al. 2006; Zuiderwijk et al. 2014b). In parallel, 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 et al. 1997). According to Tammisto et al. (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. 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.

Furthermore, it can be useful for 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.

However, this potential is to this day non-systematically exploited, as there are significant barriers that hinder the effective exploration, management and distribution of the vast amounts of available public sector information towards the research communities (e.g. operational data, financial performance data, process-related information, indices and metrics, key performance indicators, tacit knowledge). Furthermore, there is a shortage of experimental methods and tools that would allow effective knowledge mining, visualisation or further computation, empowering the integration of information and communication technologies into government practices and their adoption by the public. In the context of governance, multiple aspects have to be taken into account (e.g. financial, social, political, administrative, legal), indicating that in order to make administrative transformation a success, a multi-disciplinary approach has to be adopted, as opposed to the sole use of ICT.

1.1. The Public Sector Information Directive

Scientists are not the only ones with an interest in accessing this information. There are several groups of stakeholders that will benefit from better access to Public Sector Information (PSI), and these include citizens, public organisations, business and industry. Open public data improve trust and promote transparent governance, support education, enhance the policy-making process and public dialogue and support business and economic growth. The recent EC communication “ICT Infrastructures for e-Science”\(^1\) highlights three key aspects of a renewed ICT strategy: e-Science, e-Infrastructures and innovation (the three vectors). The third vector focuses on the innovation potential of e-Infrastructures and proposes “the transfer of expertise

to areas beyond science (e.g. e-Health, e-Government, e-Learning)”. Similar reference has been done in the Aho report\(^2\), stating that e-Infrastructures ‘are needed in sectors such as e-Government” making a special note in procurement. The EU Competitiveness Council Conclusions of December 3rd 2009\(^3\) invite the European Commission and Member States to "explore how to extend the benefits of e-Infrastructures to industrial research and innovation, to public services and to SMEs”. E-infrastructures, in the Council’s view, are a significant driver of innovation, socioeconomic progress and the “fifth freedom”: the free movement of knowledge.

Public sector information is actually the single largest source of information in Europe with its value in the EU being estimated at €28 billion, which is four times the EU market for mobile roaming services\(^4\). These facts show the central role of public sector content in the digital age as a key driver of economic activity, while a further increase in the use of this resource shall therefore directly contribute to the EU’s goals of increasing competitiveness. Traditionally, PSI has been published in different ways and formats, from the early paper days to the early Web days in which information was being published online in whatever format was more convenient for the government organisation in charge of publishing it and according to the normative at that time. In this context, organisations were usually busy locking up their data and preventing the use of correspondingly advanced technologies on knowledge outside their influence. Recently, though, the realisation that the power of data comes from the ability to sort, search and transform it for a wide range of applications has started to dominate leading to the so-called open data initiatives.

The promise of open government and the potential for increased awareness by citizens of government activity and participation in the institutions of governance have never been greater than they are today, as a number of important technical and

---


philosophical developments have combined in the past 10–15 years to bring the open data argument into the mainstream\textsuperscript{5}.

1.2. The Digital Agenda 2020

Furthermore, the recent Digital Agenda for Europe (DAE)\textsuperscript{6} communication is conceived as one of the flagships of Europe 2020 strategy for smart, sustainable and inclusive growth; it provides a policy framework aiming at delivering sustainable economic and social benefits from a digital single market based on fast and ultra-fast Internet and interoperable applications. DAE has many references to e-Government and interoperable, secure and effective e-Government services ensuring “interoperability based on standards and open development”. It is explicitly stated that “e-Government services offer a cost-effective route to better service for every citizen and business and participatory open and transparent government. e-Government services can reduce costs and save time for public administrations, citizens and businesses. They can also help mitigate the risks of climate change, natural and human-made hazards by including the sharing of environmental data and environment related information. Today, despite a high level of availability of e-Government services in Europe, differences still exist amongst Member States and the take-up of e-Government services by citizens is low”.

“European governments are committed to making user-centric, personalised, multiplatform e-Government services a widespread reality by 2015\textsuperscript{7}. To that end, governments should take steps to avoid any unnecessary technical requirements, for example applications that only work in specific technical environments or with specific devices”. “Most public online services do not work across borders to the detriment of the mobility of businesses and citizens. Public authorities have so far focused on national needs and have not sufficiently taken into account the single market dimension of eGovernment”. “Therefore Europe needs better administrative

\textsuperscript{7}The Commission is working with Member States on an Action Plan to implement the commitments in the 2009. eGovernment Declaration Malmö.
cooperation to develop and deploy cross-border public online services”. This includes the implementation of seamless e-Procurement as well as practical e-identification and e-authentication cross border services”.

Regarding e-Infrastructures, DAE observes that “Europe should also build its innovative advantage in key areas through reinforced e-Infrastructures” highlighting GÉANT high-capacity electronic communication network and the EGI European Grid Infrastructure. “Developing an EU-strategy on cloud computing” is explicitly featured in DAE. High-Performance Computing and Scientific Data infrastructures are also key e-Infrastructure components with increasing attention. The latter is further highlighted in the recently published EC's High Level Expert Group on Scientific Data report entitled “Riding the Wave – How Europe can Gain from the Rising Tide of Scientific Data”8.

Scientific e-infrastructure is essential to help address the greatest challenges of today’s society: finding alternative sources for energy, and understanding climate change to name just a few. The EU Digital Agenda, its Framework Programmes and the policies encompassed in its European Research Area initiatives are key instruments to achieve such grand challenges.

Given all those facts, an intriguing opportunity emerges: there would be significant added-value from the deployment of an ICT platform for the collection, harmonisation, storage and dissemination of public sector information, coupled with powerful search, visualisation and analytics tools according to the needs of its user-base (primarily for researchers, but also including citizens, industry and government).

Such a platform would require researching and implementing methods for the effective exploration, management and distribution of the vast amounts of available public sector information (e.g. operational data, financial performance data, process-related information, indices and metrics, key performance indicators, tacit knowledge). Standardisation, harmonisation and interoperability concerns would need to be addressed, as well as issues pertaining to knowledge mining, visualisation or further computation. It would be necessary to create a persistent, responsive and distributed infrastructure for managing and archiving large amounts of data. Finally, the project would need to build a community of participating public organisations and data

8 Digital Agenda to unlock the full value of scientific data: High-Level Group presents report, 6 October 2010http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=6204
repositories, domain experts, academic and industry which would foster cooperation in the field of data management and e-infrastructures for public data.

1.3. Research Questions

The research designed and implemented through this thesis is targeting to implement an OGD infrastructure that will support all the above-mentioned capabilities for all the target groups. The proposed research questions are as follows:

1. Which are the main research challenges and opportunities for OGD and how we can map them?

Trying to identify the research challenges and opportunities in the bibliography we created a Research Areas Taxonomy about Open Government Data. 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, and also with interested practitioners. Such a taxonomy is of critical importance for the development of a ‘science base’ in the OGD domain. 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.

2. What is the current state-of-play in the OGD infrastructures development? –

Trying to shed light in the current state of play in the OGD domain we created a tool for the analysis of open data sources in order to perform a review on them. This review concluded the basic characteristics of OGD platforms around the world as they existed back in the beginning of this effort in 2011.

3. What are the requirements and features for a new more advanced generation of OGD platform?

In order to identify the requirements of a new advanced OGD infrastructure, we created a tool for capturing the new needs from the scientific communities and the other target groups. The description template is based on scenario development. After the results of question 2 the research team was in position to formulate advanced scenaria for the functionality of the envisioned advanced OGD infrastructure. These
scenarios were assessed and prioritised by the user target groups and the first version of
the platform was developed.

4. Which are the existing evaluation models and how to evaluate an OGD Initiative? –

A thorough research on IS evaluation models and methods was conducted in order to
answer this question. The first result was an Evaluation Framework combining
qualitative and quantitative methods of evaluation targeted on the OGD infrastructures.
This framework were applied to the already developed infrastructure and its results led
to its next versions. Furthermore, a value model were developed and a development
priorities identification algorithm was proposed in order to lead to the final version. All
the evaluation tools and their results are presented in chapter 6. The tools were
designed to identify the incentives for the users and publishers of the platform; to
identify the factors influencing people’s intention to use an OGD e-infrastructure and to
prioritize the developments of an OGD infrastructure based on users' opinions.

5. How to maximise value for Collaborative and Individual use of OGD and how to
   apply it to the Greek context?

The new developments in the field motivate us to further elaborate on how to create
and maximize value based on collaborative and individual use of OGD. New business
models were created leading us to the new generation of OGD platforms exploiting the
value flows through a marketplace development. The primary goal of the marketplace is
to be the place of exchanging added-value services or selling them for a price. These
services will be based on the exploitation of OGD and would be asked and paid by the
users. Finally, we attempt to apply this new concept in the Greek reality.
2. Overall Methodology

Even though each one of the steps of our research follows its own research methodology, a design science approach was conducted, since “design [...] is concerned with how things ought to be, with devising artefacts to attain goals” (Simon, 1996). In particular, the Design Science Research Methodology (DSRM) of Peffers et al. (2008) was used, consisting of the following six steps:

1. **Problem identification and motivation.** The problem identified was that there was little support for creating value of OGD by users, whereas Web 2.0 social media tools can be used for this. There would be significant added-value from the deployment of an ICT platform for the collection, harmonisation, storage and dissemination of public sector information, coupled with powerful search, visualisation and analytics tools according to the needs of its user-base (primarily for researchers, but also including citizens, industry and government). Additionally, a taxonomy of OGD research areas was developed and presented in chapter 3 in order to identify the current gaps towards the realisation of this vision.

2. **Define objectives of a solution.** A desk-based research was conducted – presented in chapter 4 – in order to identify the elements for the development and the current landscape of OGD infrastructures. Additionally, various sources were used to define the particular objectives of a solution to the above problem, including six semi-structured interviews, a questionnaire and four workshops. The interviews were conducted with open data experts between December 2011 and January 2012. The questionnaire was conducted between April 2012 and September 2012. Both the interviews and the questionnaire provided information about the state of the art of using open public sector data in general, and problems that are experienced in this regard. The questionnaire also asked for activities related to open data use that people desired to conduct and to which extent they were able to conduct their activities, and also to which extent they found them important and useful. In total 111 people completed the questionnaire fully. Furthermore, four workshops were conducted at international events to gather information about requirements for a second generation open data platform. The workshops aimed at engaging various open data users from different countries, so that different types of requirements can be identified. The workshops were conducted between May 2012 and September 2012 and involved 65
3. **Artefact design and development.** The previous step of the DSRM led to the design of the Web 2.0 OGD platform, which is described in the following chapter 5. The design uses the scenario-based design methodology. 10 scenarios were developed by the research team. These scenarios were assessed by OGD experts offering their perspectives towards the proposed functionality.

4. **Artefact demonstration.** A first prototype of the artefact was created and the artefact was publicly demonstrated. The platform was presented to open data users at several events (e.g. conference workshops and presentations), and also via Twitter, LinkedIn, Facebook and newsletters.

5. **Evaluation.** An evaluation framework was developed combining both qualitative and quantitative methods of evaluation as it is presented in chapter 6. A thorough research on Information Systems Evaluation Models was conducted in order to create the evaluation model scoping the OGD infrastructures and formulate the evaluation tools. This evaluation framework was applied on the first two versions of the platform driving the next developments. Furthermore, a value model was created and applied leading to the final version of the platform. Within a 2-years period, the first prototype of the artefact was evaluated by all the target groups (citizens, researchers, civil servants, journalists and computer science students and OGD experts). Most participating users had followed lectures on open data and were familiar with the topic. All evaluations consisted of an online questionnaire, a usability test and a qualitative discussion. In each of the evaluation sessions, the participants were asked to conduct a number of tasks that represented open data use on a Web 2.0 OGD portal. We refer to the whole of these tasks as a usage scenario. More features were added after each iteration of the platform development, based on the evaluation results. The findings of each evaluation were used to further specify the requirements for this new generation of Web 2.0 platforms and to further improve it.

6. **Communication of the artefact.** The artefact was communicated to potential open data users by giving presentations at conferences, organizing workshops, writing publications, sending newsletters to a large network of open data users and using social media.

The next figure illustrates the overall methodology of the dissertation.
Figure 2-1: Overall Methodology

RQ1: Which are the main challenges and opportunities in the development of OGD e-Infrastructures and how can we map them?

RQ2: What is the current state of play?

RQ3: How we can create a more advanced OGD infrastructure?

RQ4: Which are the existing evaluation models and how to evaluate an OGD initiative?

RQ5: How to create value for Collaborative and Individual use of OGD?
3. Background and literature review

3.1. A taxonomy of open government data research areas

The opening of government data have both social and economic value generated from them and 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 chapter, a detailed taxonomy of research areas and corresponding research topics of the OGD domain is presented: it includes 4 main research areas (ODG management & policies, infrastructures, interoperability and usage & value), which are further analysed 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 combination of relevant knowledge from three different kinds of sources: important relevant government policy documents, research literature and experts. For each of these 35 research topics we identified, its research literature has been summarized and main research objectives and directions have been highlighted.

3.2. Methodology for creating the OGD Research Areas Taxonomy

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 (OGDRAT), 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 et al. (2003) and Sujatha et al. (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.

Figure 3-1: Taxonomy development approach

In particular, the methodology we followed for the development of the OGDRAT 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 Figure 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, 2010; 2011a; 2011b; 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 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 section 3 (while we refrain from presenting a review of the relevant government policy documents on OGD identified and analysed in step 1, in order to keep the study 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 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 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 above 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 realisation of the fourth step resulted in the second version of the OGDRAT.

6. A workshop was organised for the discussion, evaluation and validation of the above 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 kinds of 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 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 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 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 3.5.
3.3. Literature Review

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

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. (2014a, 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 (Conradie et al. 2012; Janssen K. 2011a; Janssen M. et al. 2012; McDermott 2010; Barry et al. 2014). 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 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 et al. 2008; Yang et al. 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 above literature review it has been concluded that although there are some previous studies that propose categorisations 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.

3.4. 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 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’ section 2; however, for some research topics were proposed changes of the research area they were associated with. The OGDRAT has been constructed in an online tool “mind42.com” and is available for commenting.

9http://mind42.com/public/f2a7c2f6-63ec-475f-a848-7ed5abe6c5a4
3.4.1. 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 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 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.

Figure 4: Research topics for the OGD Management & Policies research area

Table 3-1: Description of the research topics of OGD Management & Policies research area

<table>
<thead>
<tr>
<th>Research Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Policy &amp; Legal Issues for OGD</td>
<td>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 &amp; Craglia, 2006; European Commission, 2013a and 2013b; Zuiderwijk &amp; Janssen, 2014c). Formulating an OGD policy is a complex multidisciplinary problem, and as such it is associated with many of the following research topics.</td>
</tr>
<tr>
<td>1.2 OGD Anonymisation Methods</td>
<td>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 data to be opened. Privacy-preserving data publishing (PPDP) provides methods and tools for publishing useful information while preserving data</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>1.3 OGD Cleaning Methods</td>
<td>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 et al. 2010).</td>
</tr>
<tr>
<td>1.4 OGD Quality Assessment Frameworks</td>
<td>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 et al. 2005).</td>
</tr>
<tr>
<td>1.5 OGD Visualisation methods and tools</td>
<td>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 et al. 2013).</td>
</tr>
<tr>
<td>1.6 OGD Linking</td>
<td>The principles, frameworks, techniques and tools for OGD linking are the subjects of this research area (Kalampokis et al. 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 et al., 2009). The</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>1.7 OGD Publishing</th>
<th>The OGD Publishing research deals with and investigates all the issues of the publishing workflow and its involved actors (Bizer et al. 2009; Dawes, 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.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.8 OGD Mining</td>
<td>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 et al. 2013).</td>
</tr>
<tr>
<td>1.9 OGD Rating and Feedback</td>
<td>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 above purposes are also investigated in this research area, e.g. through Web 2.0 oriented mechanisms (Alexopoulos et al. 2014; Charalabidis et al., 2014).</td>
</tr>
</tbody>
</table>
3.4.2. 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 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 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 also with representative literature from the EGRL.

![Figure 5: Research topics for the OGD Infrastructures research area](image)

### Table 3-2: Description of the research topics of the OGD Infrastructures research area

<table>
<thead>
<tr>
<th>Research Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 OGD Portals Architecture</td>
<td>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 et al., 2014; Helbig et al. 2012). Various types and generations of architectures are proposed and discussed from various...</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>2.2 Open Web Services / APIs</td>
<td>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 et al. 2003).</td>
</tr>
<tr>
<td>2.3 OGD User Profiling and Service personalisation</td>
<td>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 et al. 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 and El-Masry, 2013).</td>
</tr>
<tr>
<td>2.4 OGD Long-term Preservation</td>
<td>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 et al. 2000).</td>
</tr>
<tr>
<td>2.5 OGD Storage</td>
<td>This research topic concerns the optimization of OGD storage, combining knowledge from various domains, such as databases and algorithms.</td>
</tr>
<tr>
<td>2.6 Cloud computing for OGD</td>
<td>The use of private and public cloud computing technologies and services (Lewis 2013) for hosting and providing OGD is</td>
</tr>
<tr>
<td>2.7 Citizen-generated open data</td>
<td>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\textsuperscript{10} as an open encyclopaedia, or OpenStreetMap\textsuperscript{11} as an open topographic dataset of the world (Richter et al. 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).</td>
</tr>
<tr>
<td>2.8 Sensor-generated open data</td>
<td>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 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).</td>
</tr>
</tbody>
</table>

\textsuperscript{10} http://en.wikipedia.org/wiki/Main_Page
\textsuperscript{11} http://www.openstreetmap.org/
3.4.3. 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 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 et al. 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 annotation, ontologies and controlled vocabularies and codelists, and also on OGD platforms technical interoperability, services interoperability standards and organizational 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.

Figure 6: Research topics for the OGD Interoperability research area

Table 3-3: Description of the research topics of the ‘OGD Interoperability’ research area

<table>
<thead>
<tr>
<th>Research Topic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Metadata for</td>
<td>This research topic includes various OGD metadata related</td>
</tr>
</tbody>
</table>
OGD 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 et al. 2009; Zuiderwijk et al. 2012; Robertson et al. 2001).

### 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 et al. 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) (Dong-Po 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 et al. 2009; Dill et al. 2003). The major objective of this research is the development of algorithms and tools for semantic integration (Bergamaschi et al. 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 et al. 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. 2010; 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 et al. 2008; Jardim-Goncalves et al. 2013).

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 (Sarantis et al. 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).

3.4.4. OGD Usage and Value

The fourth research area of OGDRAT 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 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.
Figure 7: Research topics for the OGD Usage and Value research area

Table 3-4: Description of the research topics of the ‘OGD Usage and Value’ research area

<table>
<thead>
<tr>
<th>Research Topics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Skills Management for OGD</td>
<td>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.</td>
</tr>
<tr>
<td>4.2 Reputation Management</td>
<td>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 et al. 2013; Hansson et al. 2013).</td>
</tr>
<tr>
<td>4.3 OGD Use</td>
<td>It includes studies that describe and analyse examples, ways and paradigms of OGD use for various purposes, not</td>
</tr>
</tbody>
</table>
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. (2011) combined social data and ODG for participatory decision-making in government).

<table>
<thead>
<tr>
<th>4.4 OGD-based Entrepreneurship</th>
<th>This research topic concerns mainly business models for exploiting the potential value of ODG and initiating OGD value chains (Ferro et al. 2012; 2013).</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5 OGD Value and Impact Assessment</td>
<td>The current OGD research on this topic focuses on analysing OGD initiatives that have led to the generation of some kind of public value (Davies et al. 2013; Jetzek et al. 2012; 2013; Charalabidis et al., 2014), analysing the positive – and sometimes the negative as well - aspects of OGD use and impacts.</td>
</tr>
<tr>
<td>4.6 OGD Needs Analysis</td>
<td>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 et al. 2012).</td>
</tr>
<tr>
<td>4.7 OGD-based Accountability</td>
<td>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 (Böhm et al. 2012; Alon 2011).</td>
</tr>
<tr>
<td>4.8 OGD Readiness</td>
<td>The main objective of this research is to develop</td>
</tr>
</tbody>
</table>
Assessment

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, 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 et al. 2014; Alexopoulos et al. 2013; Kalabokis 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. 2014a). According to this literature, OGD innovation concerns mainly three domains: (a) research, (b) business and (c) transparency (Jetzek et al. 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.

3.5. Answering Research Question 1

Open Government Data has been recognized as a strategic tool for Governments all over the world, in their efforts to increase citizen trust, transparency and collaborative action. Despite its significance and the political support at EU level, many challenges remain open for member states, in their effort to provide on-line services for the discovery and use of public sector datasets, especially towards scientists of non-ICT domains. Most digital repositories offer very little in the way of advanced data services, and are usually limited to search, retrieval and exportation of data sets and documents.

Given the fact that different disciplines have different uses for OGD and recognizing that current capacities to conduct research with OGD remain lacking, the main focus is to support the research community by providing a state-of-the-art research data management and service infrastructure. An infrastructure should
integrate multiple and diverse tools and resources and will enhance thus the effectiveness and quality of e-government research, offering a single point of access for a huge volume of data that assist in navigating, searching, understanding and using the information within it.

As mentioned in the introduction, the OGD research domain is still in its early stages, so it is important to identify the challenges and opportunities by developing a taxonomy of its main research areas and topics. The major contribution of this chapter is that 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 towards higher levels of maturity), through extraction and combination of relevant knowledge from three different 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 analysed, which enables a better understanding of them and their main research objectives and directions.

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 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. 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.

However, 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 section 3.2). 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.

3.5.1. Limitations

A limitation of our study is that for practical reasons we organized only one workshop (though 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 analysed 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. Also more research is needed in order to map the multiple OGD research projects which are currently in progress (e.g. supported by European Commission or USA 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.

Most of the above-mentioned research areas provide insights to the development of an advanced OGD infrastructure. The following research areas left outside the analysis since they are criticized as out of scope for the development of such an infrastructure:

1. Policy & Legal Issues for OGD. Even though the policy and legal issues are characterised of major importance when a worldwide infrastructure is designed and even if they have been recognised and followed, they are left outside the analysis because our major scope is to define the solution conforming to the existing policy and legal framework and not to be part of their renegotiation.

2. Data Anonymisation Methods are not included in the analysis since the inception of the OGD infrastructure precludes the existence of non-anonymised datasets publication. We consider this prerequisite has been met from the design of the datasets from the source that publish them.
4. Current State of Play in OGD infrastructures

This chapter is dedicated to answering research question 2 through literature review and desk-based research upon the OGD initiatives. There has been a long standing movement towards the ‘Open Government’ paradigm, aiming at increasing interactions and improving relations between government agencies and citizens, and increasing the trust of the latter towards the former; the main components of this open government paradigm according to the extensively debated and highly influential ‘Open Government Directive’ of USA (Obama, 2009; Executive Office of the President, 2009; McDermott, 2010; Lathrop & Ruma, 2012; Bertot et al., 2014) are transparency, participation of citizens and collaboration with them. The opening of various categories of data possessed by government (e.g. government spending, economic and business, legal, social, environmental, agricultural and tourism data) to the citizens constitutes a central element of the first component, and at the same time a critical precondition for the achievement of the other two. For this reason Open Government Data (OGD) initiatives have proliferated over the last years worldwide, not only in developed countries, but also in developing ones as well (Huijboom & Van Den Broek, 2011; Harrison et al., 2012a; European Commission, 2013). According to the Open Knowledge Foundation as open data are defined data that are freely accessible online, available without technical restrictions to re-use, and provided under open access license that allows the data to be re-used without limitation (OKF, 2012). Geiger and v.Luecke (2012) define as OGD “all stored data of the public sector which could be made accessible by government in a public interest without any restrictions for usage and distribution”. A variety of government organizations with different strategies and technical capacities, and under different social, political and legal conditions, are embracing this concept and open big numbers of datasets they own to the society, aiming at the achievement of both social and economic benefits, the former associated with government transparency and accountability, and citizens’ participation, and the latter with development of economic activity in new value added e-services based on the combination of various types of OGD (and possibly private data as well) (Jetzek et al. 2013a and 2013b).

Though the initial motivation for opening government datasets was to promote government transparency and accountability, subsequently there has been extensive
interest in the potential of Public Sector Information (PSI) as an engine for business innovation and job creation, by opening appropriate parts of it as OGD to the society, and creating complex ecosystems of public and private actors around them, in order to exploit them and generate social and economic value from them (Harrison et al., 2012b; Chan, 2013). According to the Digital Agenda Assembly (2011) “the direct annual turnover in the PSI Sector is estimated to be around EUR 28 Billion, with annual growth of around 8% and this makes the PSI Sector one of the faster growing sectors of the EU economy”, while it is estimated that “the aggregated macroeconomic footprint of the PSI industry is EUR 140 Billion. This means that the ‘spill’ or overflow from the PSI industry is even bigger than the sector itself”.

A recent study by the McKinsey Global Institute showed that more than forty countries have developed OGD platforms, over one million datasets have been published worldwide, which is expected to lead to the creation of actual value of about three trillion US dollars in seven industries (education, transportation, consumer products, electricity, oil and gas, health care and consumer finance) (Manyika, J., 2013). For all the above reasons there is widespread and growing debate on OGD among government practitioners and academics, resulting in a continuously growing research literature (e.g. Allan, 2009; Meijer & Thaens, 2009; Robinson et al., 2009; Lathrop & Ruma, 2010; Parycek & Sachs, 2010; Maier & Huber, 2011; Janssen et al., 2012; Kassen, 2013; Zuiderwijk et al., 2014; Mellouli et al., 2014), and also in the establishment of opening government datasets as a ‘political orthodoxy’ in numerous countries worldwide (e.g. in the USA (Obama, 2010), in the UK (Cameron, 2010), in Australia (AGIMO, 2010) and across Europe (European Commission, 2013 and 2014)).

However, despite the big investments that have been made for the development of ‘OGD sources’, defined as various types of portals enabling access to government datasets by the public through the Internet and providing various capabilities/functionalities in this direction (see section 4.1 for more details), by a variety of government organizations with different strategies and technical capacities, and under different social, political and legal conditions worldwide, limited research has been conducted on these OGD sources in order to understand better their main characteristics from various perspectives, and identify their strengths and weaknesses. This study contributes to filling this research gap, by presenting an analysis of the characteristics of OGD sources. These OGD sources constitute a new type of information
systems (IS), so according to previous relevant research on IS Success (DeLone & McLean, 1992 and 2003; Urbach & Mueller, 2012) their success relies critically on three main characteristics of them: their ‘information quality’ (i.e. the quality of the information they provide), their ‘system quality’ (i.e. their quality viewed as technological systems) and their ‘service quality’ (i.e. the support provided to its users, such as training, helpdesk, etc.). Our study focuses on the first two of them. In particular, this first analysis of OGD sources focuses on two important dimensions of their ‘information quality’: the thematic width of the datasets they provide, and their semantic capabilities; and also on two important dimensions of their ‘system quality’: the functionality they offer and their technological characteristics.

Therefore the research objective of this study is to analyse the characteristics of OGD sources from functional, semantic and technological perspectives, in order to identify the current landscape. This is quite interesting, as previous research has revealed the importance of the national context on government information and knowledge sharing (Gharawi and Dawes, 2010; Dawes et al., 2011).

4.1. Background and key concepts

4.1.1. Open Government Data recent history

The potential of a significant part of PSI to be re-used beyond the public sector for various both commercial and non-commercial social purposes has been recognised more than a decade ago. In 2003, the European Union (EU) adopted the ‘Directive on the Re-use of Public Sector Information’\(^\text{12}\) (European Commission, 2003), which encourages the member states to make as much of the information they possess available for re-use as possible. It establishes a minimum set of rules and also the practical means for facilitating this reuse, focusing mainly on its economic aspects. An on-line public consultation on this PSI Directive was launched in September 2010, leading to a revision of this Directive\(^\text{1}1\) (European Commission, 2013), and also signalling its inclusion as one of the key actions of the ‘Digital Agenda for Europe’\(^\text{13}\). The new Directive highlights the importance of PSI as a vast, diverse and valuable pool of resources that can benefit the knowledge economy, and encourages the proliferation of

\(^{13}\) http://ec.europa.eu/information_society/digital-agenda/index_en.htm
OGD portals (Zijlstra et al., 2013). It includes policies for "encouraging the wide availability and re-use of PSI for private or commercial purposes, with minimal or no legal technical or financial constraints, and promoting the circulation of information not only for economic operators but also for the public, which can play an important role in kick-starting the development of new services based on novel ways to combine and make use of such information, stimulate economic growth and promote social engagement". Also, in July 2014 the European Commission published guidelines in order to help the member states transpose the revised rules and to propose best practices concerning several important aspects for PSI re-use (European Commission, 2014).

As it is already mentioned, the ‘Open Government Directive’ of USA was highly influential for opening PSI to the society (Obama, 2009; Executive Office of the President, 2009; McDermott, 2010; Lathrop & Ruma, 2012; Bertot et al., 2014), which states that “to increase accountability, promote informed participation by the public, and create economic opportunity, each agency shall take prompt steps to expand access to information by making it available online in open formats”, adding also that “with respect to information, the presumption shall be in favour of openness (to the extent permitted by law and subject to valid privacy, confidentiality, security, or other restrictions)”. This was followed by the launch of the ‘Open Government Partnership’ on 2011, when the eight founding countries (Brazil, Indonesia, Mexico, Norway, Philippines, South Africa, United Kingdom, United States) endorsed an Open Government declaration, announced their countries’ action plans, and welcomed the commitment of 38 more countries to join the partnership. A study of the OGD strategies of five countries (Australia, Denmark, Spain, the United Kingdom and the United States) conducted by TNO (an independent Dutch public law organization aiming to enable and support business and government to apply and exploit various types of new knowledge – see https://www.tno.nl/en/about-tno/) concludes that “in an increasing number of Western countries, ‘open data’ is being placed on the political and administrative agenda’, and that “the focus of strategies is currently on fostering innovation and strengthening democratic participation, whereas some evidence indicates that open data could also contribute to enhancing law enforcement’ (Huijboom & Van Den Broek, 2011). Another study concerning ‘Open Data in Developing Countries’ (Schwegmann, 2012) conducted under the auspices of the

14 http://www.opengovpartnership.org/
European Public Sector Information Platform concludes that “Open Data seems to be high on the agenda not only in Western countries, but also in developing countries”, as civil society organisations and external partners (including the abovementioned ‘Open Government Partnership’) of developing country governments are encouraging the use of open data, aiming mainly to increase transparency, accountability and citizen participation. Recently, the new ‘Open Data Policy’ Memorandum has been issued in USA (Executive Office of the President, 2013), stating that ‘Information is a valuable national resource and a strategic asset to the Federal Government, its partners, and the public’, and directing all executive departments and agencies to assign high priority to the management of the information artefacts they generate and maintain as important assets throughout their life cycle, having as main objectives the promotion of openness and interoperability; it also stresses that this will increase operational efficiencies, reduce costs, improve services, support mission needs, and increase public access to valuable government information.

Jetzek et al. (2013a and 2013b) argue that there are four types of value that can be generated from the OGD, which differ in the sector generating the value (public or private), and also in the kind of generated value (economic or social): i) transparency related value (public sector organizations generate social value by offering increased transparency into government actions, which reduces ‘information asymmetry’ between government officials and citizens, and therefore misuse of public power for private benefits and corruption), ii) efficiency related value (public sector organizations generate economic value through OGD by increasing internal efficiency and effectiveness), iii) participation related value (private sector firms generate social value through participating and collaborating with government), iv) innovation related value (private sector firms generate economic value through the creation of new products and services). It is gradually realised that OGD create big opportunities for both private and public sector innovation (see Editorial of Special Issue on Innovation through Open Data of the Journal of Theoretical and Applied Electronic Commerce Research by Zuiderwijk et al. (2014)), being quite importance for the development of information and knowledge economy. However, a study conducted by Janssen et al (2012), based on data collected through interviews and workshops, warns that “a conceptually simplistic view is often adopted with regard to open data, which automatically correlates the publicizing of data with use and benefits”; in particular, they identify five ‘myths’ that
have been gradually developed with respect to OGD: ‘The Publicizing of Data will Automatically Yield Benefits’, ‘All Information Should Be Unrestrictedly Publicized’, ‘It Is a Matter of Simply Publishing Public Data’, ‘Every Constituent Can Make Use of Open Data’ and ‘Open Data Will Result in Open Government’. They finally conclude that the success of the developed OGD infrastructures requires more than the simple provision of access to data: it is necessary to make progress towards the improvement of the quality of government information, the creation and institutionalization of a culture of open government, and the provision of the tools and instruments with which to use the data. The realisation of the ‘Open Government’ paradigm in general seems to be a difficult and complex task, which requires combined efforts of multiple actors, not only from the public sector, but also from the private sector as well, and gradual development of 'open government ecosystems' (Harrison et al., 2012b).

4.1.2. OGD Thematic Categories

According to the abovementioned EU ‘Directive on the Re-use of Public Sector Information’ (2003) and the relevant report of the OECD Working Party of Information Economy (2006) a wide variety of OGD thematic categories are opened by government agencies and offered as OGD through the Internet, and the most important of them are:

- **Economic and Business Information**, including financial, public spending, economic activity and statistics, industry and trade data, as well as, official business registers and public tender databases.

- **Geographic Information**, including cartographic, cadastral, spatial (geographical coordinates), administrative and political boundaries, topographical, public buildings (geo-coordinates) and elevation data.

- **Legal information**, including crime/conviction data, laws, rights and duties, legislation and treaties, judicial and patent and trademark information.

- **Meteorological and Environmental Information**, including oceanographic, hydrographical, environmental (quality), atmospheric and meteorological (weather) data.

- **Social Information**, including demographic, attitude, health, education and labor data.

- **Traffic and Transport Information**, including traffic congestion, work on roads, public transport, vehicle registration, transport networks and transport statistics.
- **Tourist and Leisure Information**, including tourism, entertainment and hotel related data.

- **Agricultural, Farming, Forestry and Fisheries Information**, including cropping/land use, farm income/use of resources, fish farming/harvest and livestock data.

- **Natural Resources Information**, including biological, ecological, geological and geophysical data, as well as, energy resource/consumption and geodetic networks information.

### 4.1.3. OGD Sources Capabilities/Functionalities

Research carried out as part of the European project ENGAGE (see [www.engage-project.eu](http://www.engage-project.eu), [http://www.engagedata.eu/](http://www.engagedata.eu/)) has revealed that two distinct types of OGD sources/portals have been developed with respect to the capabilities/functionalities provided to the user (Mouzakitis et al., 2011):

i) **OGD direct provision portals**: this is the main category of OGD portals, which are ‘primary sources’ of OGD, publishing original government datasets provided by either one government agency, or a small number of similar government agencies (who are the legal owners/licensers of the data); they usually offer a wide range of functionalities supporting the whole lifecycle of OGD, from the creation of datasets to the update and finally to the archiving of them.

ii) **OGD aggregators**: this category includes OGD aggregator portals, which are ‘secondary sources’ of OGD coming from a big number of government agencies, publishing and maintaining lists of other ‘primary’ OGD catalogues and links to them. They constitute single access points to multiple OGD direct provision portals, and make it easier for a user to locate the OGD he/she is interested in. Usually they include descriptive information about datasets and sources, which is quite useful for the users in order to get a first impression of what is available. Many of them act as highly structured registries of OGD primary sources and datasets, which store structured and machine processable information, and provide 'index'-like features, such as automated registration and discovery of OGD. Some prominent examples are the widespread CKAN data hub portal15, the Open Government Data Initiative16 (OGDI) - (a Microsoft initiative

---

15 CKAN The Data hub, [http://thedatahub.org/](http://thedatahub.org/)
16 Open Government Data Initiative, [http://ogdisdk.cloudapp.net/DataCatalog/DataSetList](http://ogdisdk.cloudapp.net/DataCatalog/DataSetList)
/ online tool to publish and use a wide variety of public data from government agencies), the European Union Open Data portal\textsuperscript{17} and the Freebase\textsuperscript{18}.

4.1.4. Semantic Technologies and OGD

The use of Semantic Web technologies (such as “Uniform Resource Identifiers” (URI), the “Resource Description Framework” (RDF), vocabularies and ontologies) in OGD enables a more effective browsing and discovery of datasets through distributed SPARQL queries, and also linking and combining OGD from multiple sources across the Web, which can increase significantly the usefulness of the OGD and the value generated from them (e.g. it allows discovering new correlations and gaining deeper insights, or developing new advanced value-added e-services by combining different datasets from multiple OGD sources); also, the value of any kind of data (including OGD) increases each time it is being re-used and linked to another resource, and this can be facilitated and triggered by providing informative and explanatory data about each available dataset (i.e. metadata). These ideas gave rise to the development of the Linked Open Government Data (LOGD) concept (an application/elaboration of the more general ‘Linked Data’ concept for the OGD) (Wood, 2011; Geiger and v.Luecke, 2012; Bauer and Kaltenböck, 2012; Zuiderwijk et al., 2012a; Zuiderwijk et al., 2012b). Geiger and v.Luecke (2012) define LOGD as “all stored data of the public sector connected by the World Wide Web which could be made accessible in a public interest without any restrictions for usage and distribution”, and argue that “the cross linking of Open Data via the Internet and the World Wide Web as “Linked Open Data” (LOD) offers the possibility of using data across domains or organizational borders for statistics, analysis, maps and publications”, which can lead to the generation of more insight, knowledge and innovation from OGD. The W3C Government Linked Data Working Group\textsuperscript{19} is developing standards and best practices in order to assist government agencies to open and publish their data as effective and usable Linked Data using Semantic Web technologies (for more details on methodologies for this see Wood (2011) and Bauer and Kaltenböck (2012)).

\textsuperscript{17} European Union Open Data portal, https://open-data.europa.eu/en/data/
\textsuperscript{18} Freebase http://www.freebase.com
\textsuperscript{19} http://www.w3.org/2011/gld/charter.html
4.2. Research Method

The research method we adopted for conducting a multi-perspective analysis of the Greek OGD Sources landscape consists of four distinct stages: i) Identification of OGD Sources, ii) Formulation of Analysis Framework (= definition of the analysis perspectives, and the particular indicators of each of them), iii) Data Collection, and iv) Data Processing. These four research stages are described in the following paragraphs.

Our study has been based on quantitative techniques, as, according to the relevant literature (Maylor and Blackmon, 2005; Ragin and Amoroso, 2011), they are the recommended approach for collecting data from a large quantity of units (in our case the whole population of OGD sources in Greece – see following description of stage 1) concerning a limited number of features of them (in our case the selected thematic, functional, semantic and technological characteristics of these OGD sources - see following description of stage 2), and then condensing/summarizing these data into a few numbers (e.g. relative frequencies).

**Stage 1: Identification of OGD Sources**

The first stage of our research method refers to the identification of the OGD Sources in EU. For this purpose we adopted a top-down approach based on desk-based research, starting from EU central points of access, World Bank, OECD and Ministries websites, in which we searched for published OGD datasets, and then following links to other government organizations (e.g. national or Regional Administrations etc.), in which we performed similar searches for finding OGD datasets, and links to other government organizations, etc., in an effort to identify websites and portals containing and providing OGD. As a result of this stage we identified 34 OGD sources, which are shown in the Appendix (for each of them we can see the corresponding government agency, a description, the thematic categories of the OGD it provides and finally its URL).

**Stage 2: Formulation of Analysis Framework**

The second stage of our research method is the construction of the analysis framework. As mentioned in the Introduction OGD sources constitute a new type of IS, so for analysing them we can use as our theoretical foundation the models that have been developed in previous research on IS Success (DeLone & McLean, 1992 and 2003;
Urbach & Mueller, 2012). According to them the success of an IS relies critically on three main characteristics of it: its ‘information quality’ (i.e. the quality of the information it provides to its users), its ‘system quality’ (i.e. its quality viewed as a technological system), and its ‘service quality’ (i.e. the support provided to its users, such as training, helpdesk, etc.); these affect ‘users satisfaction’ and the ‘actual use’ of the IS, and finally its ‘individual impact’ and ‘organizational impact’. The above model has been used extensively for studying various types of IS (with appropriate elaboration and adaptation to the characteristics and objectives of each), including government IS providing various electronic services to citizens (e.g. Wang & Liao, 2008; Rana et al., 2013; Chena et al., 2015). Our analysis of Greek OGD sources has focused on the first two of the above critical characteristics, the information quality and the system quality, elaborating and adapting them to the particular characteristics and objectives of the OGD sources. In particular, we examine two important dimensions of the OGD sources’ ‘information quality’: i) their semantic capabilities (since as mentioned in the previous section the use of Semantic Web technologies in OGD can significantly increase their usefulness and value). We also examine two important dimensions of the Greek OGD sources’ ‘system quality’: ii) the functionality they offer (i.e. to what extent are offered on one hand the basic and on the other hand the advanced features mentioned in the previous section, which is quite important for the users); iii) their technological characteristics (as they are critical for the performance of the OGD sources, for their adaptability to new needs, and also for the efficient provision of high quality functionality to the users). Therefore our analysis framework consists of the above four analysis perspectives (thematic, functional, semantic and technological), which are described in more detail in the following paragraphs.

I) Functional Analysis Perspective: It includes analysis of the functionalities provided by the OGD sources. It has been conducted using five indicators: three of them concern fundamental functionalities of the OGD sources and have been previously used in the above-mentioned analysis of the European OGD sources by Petychakis et al. (2014): they concern datasets’ discovery, datasets provision and also multi-linguality related capabilities. The other two indicators correspond to two important features of the new and more advanced generation of OGD sources mentioned in sub-section 2.3: visualization and users’ feedback related capabilities. Therefore the corresponding five indicators of the functional analysis of the OGD sources (it should be noted that their
possible values have been defined based on the study of Petychakis et al. (2014), and also the analysis of a representative subset of the identified Greek OGD sources, from which the main possible values have been determined, and then these have been enriched with new values we found during the process of analysing the identified OGD sources (in stage 3):

- Datasets Discovery: it concerns the tools provided for discovering the datasets the user is interested in; its main possible values (not mutually exclusive) were: simple document list, free text search, browsing through categories, browsing through filters, browsing through interactive map and SPARQL search.
- Data Provision: it concerns the ways of data provision to the users; its main possible values (not mutually exclusive) were: download file, online view of dataset and on map.
- Language: it concerns the language(s) the user interface is available in; its main possible values are (not mutually exclusive) Greek, English, French and German.
- Visualizations: it concerns the datasets’ visualization capabilities provided; one possible value is ‘not existing’, while other main possible values (not mutually exclusive) are visualizations in charts and visualizations in maps.
- Feedback: it concerns the existing tools allowing feedback from OGD users to the providers; its two main possible values were ‘not existing’ and ‘existing’ (the latter has been further refined during the process of analysing the identified OGD sources in stage 3).

II) Semantic Analysis Perspective: It includes analysis of the use of Semantic Web technologies for the representation and structure of OGD. In particular, it includes initially an overall assessment of the OGD sources using the well-established 5-stars Berners Lee’s rating system for open data (Bauer and Kaltenböck, 2012), and then an analysis of the metadata (which are critical for linking OGD and obtaining higher levels of value from them as mentioned in the previous section) and of licence information. Therefore the corresponding four indicators of the semantic analysis of the OGD sources are:

- Sources rating according to the 5-stars Berners Lee’s Rating Scheme for open data20:

---

* make your stuff available on the web (whatever format)
** make it available as structured data (e.g. excel instead of image scan of a table)
*** using non-proprietary format (e.g. csv instead of excel)
**** use URLs to identify things, so that people can point at your stuff
***** link your data to other people's data to provide context

- Sources metadata rating according to the 5-stars Maturity Scheme of Metadata Management (ISA, 2011).
  * Metadata Ignorance
  ** Scattered or Closed Metadata
  *** Open Metadata for Humans
  **** Open Reusable Metadata
  ***** Linked Open Metadata

- RDF-compliance: it concerns the use of technologies that support RDF, including technical products of open data initiatives publishing structured data in a way that it can be interlinked, which as mentioned in the previous ‘Background’ section is quite important for enabling more effective browsing and discovery of datasets, and for linking and combining OGD from multiple sources (e.g. see Wood (2011); Bauer and Kaltenböck (2012)); it is a binary indicator.

- Data License: it concerns license information related to the use of the published datasets; this is one of the most important characteristic of OGD sources, since it defines the allowed ways of OGD use and exploitation for generating various types of social and economic value, and reduces all relevant legal uncertainties and risks (e.g. see Wood (2011); Bauer and Kaltenböck (2012)); its two main possible values were 'not existing' and 'existing' (the latter has been further refined during the process of analysing the identified OGD sources).

III) Technological Analysis Perspective: It includes analysis of the technologies and products that have been used for the development of the OGD source at the main technological layers: web server, Content Management System (CMS) or platform, user interface, data format and API. The five indicators of the technological analysis are (their possible values have been formulated during the process of analysing the identified OGD sources):

- Web Server: it concerns the web server that the OGD site is hosted on
- CMS/Platform: it concerns the CMS or Platform that the OGD source has been based on.
- User Interface (UI): it concerns the technologies used for the presentation layer of the OGD source.
- Data Format: available data representation formats of the published information
- API: available Application Programming Interfaces (API) and Web Service interfaces.

Stage 3: Data Collection

This third stage includes the collection of all the necessary data for the analysis defined in stage 2 from the OGD sources identified in stage 1. For the first two analysis perspectives (functional and semantic) a desk-based research approach was used, which included manual examination of all OGD sources and assessment of all indicators. In order to ensure the validity of the results two different experts made these assessments independently, and then compared their results; in cases of differences (there were differences in less than 5% of the results), they were discussed by the whole authors’ group, which made the final decisions. For the fourth analysis perspective (technological) we used the software tool Wapallyser (https://wappalyzer.com/) in order to collect from all examined Greek OGD sources all the required information for assessing its five indicators. The collection of all these data from the OGD sources, and also the access to all websites mentioned in the previous ‘Background’ section and in the References section, has taken place in January 2014.

Stage 4: Data Processing

In the final stage statistical analysis of the data collected in stage 3 was conducted, which included calculation of various descriptive statistics, such as frequencies and relative frequencies of all values for each of the abovementioned indicators, and construction of various charts, using the Excel software. These statistics were discussed extensively among the whole authors’ group, and conclusions were drawn from them, as well as recommendations for enhancing OGD provision in Greece.
4.3. Open Data Sources

In this context, the following Public Sector Information Sources and Open Government Data initiatives have been selected for further analysis. Full lists of the European Initiatives and the Greek OGD Sources are provided in Appendices D and E respectively.

Table 4-1. Indicative Public Sector Information Sources and Open Data initiatives (alphabetical order)

<table>
<thead>
<tr>
<th>ID</th>
<th>Title</th>
<th>Description</th>
<th>Information</th>
<th>Coverage</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Borough of Trafford (UK) Open Data Catalog</td>
<td>Open Data Catalog for information about city's services</td>
<td>Various / Government data</td>
<td>UK</td>
<td><a href="http://www.trafford.gov.uk">www.trafford.gov.uk</a></td>
</tr>
<tr>
<td>2</td>
<td>Central Bureau of Statistics</td>
<td>This OGD site provides statistics of Israel.</td>
<td>Statistics</td>
<td>Israel</td>
<td><a href="http://www.cbs.gov.il">www.cbs.gov.il</a></td>
</tr>
<tr>
<td>3</td>
<td>Data Gov UK</td>
<td>Datasets from all central government departments and a number of other public sector bodies and local authorities for the UK</td>
<td>Various / Government data</td>
<td>UK</td>
<td>data.gov.uk</td>
</tr>
<tr>
<td>4</td>
<td>Data.gov</td>
<td>The purpose of Data.gov is to increase public access to high value, machine readable datasets generated by the executive branch of the federal government of US.</td>
<td>Various / Government data</td>
<td>US</td>
<td><a href="http://www.data.gov">www.data.gov</a></td>
</tr>
<tr>
<td>5</td>
<td>Deutschland API</td>
<td>Collection of election, electoral ward and politician data.</td>
<td>Governance</td>
<td>DE</td>
<td><a href="http://www.deutschland-api.de/Hauptseite">www.deutschland-api.de/Hauptseite</a></td>
</tr>
<tr>
<td>6</td>
<td>Diavgeia Gov (Cl@rity)</td>
<td>Laws / Regulations / Ministry Decisions &amp; Public Bodies Spending.</td>
<td>Financial / Governance</td>
<td>GR</td>
<td>et.diavgeia.gov.gr</td>
</tr>
<tr>
<td>7</td>
<td>EUR-Lex</td>
<td>EUR-Lex provides free access to European Union law and other documents considered</td>
<td>Law / Governance</td>
<td>EU</td>
<td>eur-lex.europa.eu</td>
</tr>
<tr>
<td>ID</td>
<td>Title</td>
<td>Description</td>
<td>Information</td>
<td>Coverage</td>
<td>URL</td>
</tr>
<tr>
<td>----</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>----------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>8</td>
<td>European Environment Agency Data Service</td>
<td>A European information source for environmental data.</td>
<td>Environmental science</td>
<td>EU</td>
<td>dataservice.eea.europe.eu/ dataservice/default.asp</td>
</tr>
<tr>
<td>10</td>
<td>Geospatial Governmental Data</td>
<td>The first try (beta version) to make geospatial data available to all citizens.</td>
<td>Geographical / Spatial</td>
<td>GR</td>
<td>geodata.gov.gr</td>
</tr>
<tr>
<td>11</td>
<td>Global Health Data Exchange</td>
<td>GHDx provides data for world’s health and demographic datasets</td>
<td>Health</td>
<td>Worldwide</td>
<td><a href="http://www.ghdx.org">www.ghdx.org</a></td>
</tr>
<tr>
<td>12</td>
<td>Hellenic Statistical Authority</td>
<td>Greece various statistics per business sector</td>
<td>Statistics</td>
<td>GR</td>
<td><a href="http://www.statistics.gr">www.statistics.gr</a></td>
</tr>
<tr>
<td>13</td>
<td>INSPIRE Geoportal</td>
<td>The INSPIRE geoportal provide the means to search for spatial data sets and spatial data services, and subject to access restrictions, view and download spatial data sets from the EU Member States</td>
<td>Geographical / Spatial</td>
<td>EU</td>
<td><a href="http://www.inspire-geoportal.eu">www.inspire-geoportal.eu</a></td>
</tr>
<tr>
<td>14</td>
<td>Israel Government Sharing Site</td>
<td>Sharing (discussions, votes) from the public on government policy and agenda topics</td>
<td>Governance</td>
<td>Israel</td>
<td><a href="http://www.shituf.gov.il/">www.shituf.gov.il/</a></td>
</tr>
<tr>
<td>16</td>
<td>Israel Mapping Center (Survey Of Israel)</td>
<td>National Agency for Geodesy, Cadastre, Mapping and</td>
<td>Geographical / Spatial</td>
<td>Israel</td>
<td><a href="http://www.mapi.gov.il">www.mapi.gov.il</a></td>
</tr>
<tr>
<td>ID</td>
<td>Title</td>
<td>Description</td>
<td>Information</td>
<td>Coverage</td>
<td>URL</td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>17</td>
<td>Israel Ministry of Environmental Protection</td>
<td>Environmental data overlayed on a map</td>
<td>Environmental</td>
<td>Israel</td>
<td>gis.sviva.gov.il/interactiveMap1.htm #9$0&amp;997</td>
</tr>
<tr>
<td>18</td>
<td>Israel Ministry of Finance</td>
<td>Israel Budget</td>
<td>Finance</td>
<td>Israel</td>
<td>budget.msh.gov.il</td>
</tr>
<tr>
<td>19</td>
<td>Lichfield District Council</td>
<td>Open data for Lichfield District Council</td>
<td>Various / Government data</td>
<td>UK</td>
<td>www2.lichfielddc.gov.uk/data/</td>
</tr>
<tr>
<td>21</td>
<td>London Datastore</td>
<td>The London Datastore is an initiative by the Greater London Authority (GLA) to release as much of the data that it holds as possible.</td>
<td>Various / Government data</td>
<td>UK</td>
<td>data.london.gov.uk</td>
</tr>
<tr>
<td>23</td>
<td>Manchester City Council Open Data</td>
<td>Open Data Catalog for information about city’s services</td>
<td>Various / Government data</td>
<td>UK</td>
<td><a href="http://www.manchester.gov.uk/info/500215/open_data">www.manchester.gov.uk/info/500215/open_data</a></td>
</tr>
<tr>
<td>ID</td>
<td>Title</td>
<td>Description</td>
<td>Information</td>
<td>Coverage</td>
<td>URL</td>
</tr>
<tr>
<td>----</td>
<td>-------</td>
<td>-------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----</td>
</tr>
<tr>
<td>26</td>
<td>offene Daten</td>
<td>ckan for german public data</td>
<td>Various / Government data</td>
<td>DE</td>
<td><a href="http://offenedaten.de">offenedaten.de</a></td>
</tr>
<tr>
<td>27</td>
<td>Organisation for Economic Co-operation and Development (OECD)</td>
<td>Includes statistical data and metadata for OECD countries and selected non-member economies.</td>
<td>Statistics</td>
<td>Worldwide</td>
<td><a href="http://stats.oecd.org">stats.oecd.org</a></td>
</tr>
<tr>
<td>28</td>
<td>Overheid.nl</td>
<td>ckan for dutch public data</td>
<td>Various / Government data</td>
<td>NL</td>
<td><a href="http://register.data.overheid.nl">register.data.overheid.nl</a></td>
</tr>
<tr>
<td>29</td>
<td>Pic and Mix</td>
<td>Pic and Mix aims to increase public access to Kent-related datasets including those generated by Kent County Council (KCC)</td>
<td>Various / Government data</td>
<td>UK</td>
<td><a href="http://picandmix.org.uk">picandmix.org.uk</a></td>
</tr>
</tbody>
</table>
4.4. Results

4.4.1. Functional Perspective Analysis

The above analysis indicates that, firstly, few OGD providers offer advanced data acquisition capabilities. The majority of data providers are internally linked to the relevant data repositories and provide only interfaces for data provision. That is especially common for organizations and agencies that are responsible for the complete life cycle of data (from creation to update/archiving), such as statistical offices. Furthermore, the majority of OGD providers offer simple free-text search and themebrowsing functions for the discovery and cataloguing of datasets, whereas only recent open data initiatives start to appreciate the advances of Semantic Web by providing semantically enriched discovery services such as performing SPARQL queries. Additionally, most local public agencies limit their data provision services to a simple download functionality whereas agencies addressed to a wider network (country-level or European level) typically include the capability to view datasets on a map or various types of charts. Nevertheless, the range of visualization facilities offered by each provider varies significantly. This is mainly due to the fact that during the last years visualization engines have become more comprehensive, flexible and light-weighted.

Lastly, few efforts concentrate on receiving the needs of users in a formal and systematic manner. In the majority of service providers comments and suggestions from users is limited to general-purpose feedback web forms that typically address comments on technical aspects of the site rather the actual datasets. On the other side, noteworthy efforts involve the inclusion of dataset rating and commenting, as well as viewing and voting users demands for specific datasets, that are not yet public or that follow strict data license.
4.4.2. Semantic and Data Perspective Analysis

The analysis shows that currently the majority of open data providers aim to adopt an already available metadata standard that fits within their context. Data providers that are based on the CKAN engine also adopt the CKAN metadata schema for the data catalogue and data discovery. Other governmental sites adopt a custom metadata schema for the data discovery and preserve the datasets in vertical-domain metadata standards. Noteworthy cases include open data initiatives that have developed metadata standards to become EU recommendations (e.g. INSPIRE). Furthermore, the majority of longstanding OGD sites indicate their intention not to follow the Linked Data paradigm, as opposed to more recent “data gov” efforts. Additionally, the analysis indicates that almost all initiatives (with the exception of EUR-Lex) limit their internationalization efforts (if any) to the user interface level. The data formats provided are more or less common between all initiatives, while the vast majority of OGD sites tend to provide data only in the format of the original source. Lastly, the analysis indicates that there is no common policy for license issues as the license for use and reuse of data significantly vary.

4.4.3. Technological Perspective Analysis

The analysis shows that there is a strong preference for open-source and free underlying platforms and content management systems in OGD sites with the exception of the Data.gov initiative which is based on the proprietary platform Socrata that receives widespread adoption in the US (State of Oregon, State of Oklahoma, City of Chicago, City of Seattle, etc). For data visualization, OGD sites are turning from heavy and proprietary engines to free and light-weighted javascript frameworks (Google charts, JQuery, JavaExts). Lastly relatively few data providers offer APIs for data and metadata interactions, whereas the paradigm of restful web services that output JSON objects is becoming the common approach.

4.5. Answering Research Question 2

The analysis of public sector data and knowledge resources strongly indicates the need for an infrastructure service platform for the collection, harmonisation, storage and dissemination of public sector information, coupled with powerful search,
visualisation and Semantic Web tools. The designing of a solution phase will address current challenges of OGD sites, while combining the best breed of practices, specifications, tools and standards currently adopted, for the planning, communication, and execution of the platform. The following table consolidates the utilities and functions of OGD portals examined in the analysis, in view of forming the base elements for the platform:

**Table 4-2. Consolidated view of utilities / functions of OGD infrastructures**

<table>
<thead>
<tr>
<th>Type</th>
<th>Function(s)</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data acquisition</strong></td>
<td>A “Submit Datasets” web form where users (authorized or not) can upload data to the portal</td>
<td>geodata.gov.gr</td>
</tr>
<tr>
<td></td>
<td>A Dataset Management System (DMS) to facilitate users to organize and maintain their dataset submissions via a web-based user interface</td>
<td>Data.gov</td>
</tr>
<tr>
<td><strong>Datasets Catalog / Discover</strong></td>
<td>Autocomplete free text search</td>
<td>Borough of Trafford (UK) Open Data Catalog</td>
</tr>
<tr>
<td></td>
<td>Browse through Categories</td>
<td>Israel Government Shring Site</td>
</tr>
<tr>
<td></td>
<td>Faceted Search</td>
<td>Data.gov</td>
</tr>
<tr>
<td></td>
<td>SPARQL query</td>
<td>Data Gov UK</td>
</tr>
<tr>
<td></td>
<td>Yahoo! Query Language (YQL) queries</td>
<td>Deutschland API</td>
</tr>
<tr>
<td></td>
<td>Advanced Query building (using all available data fields + conditional clauses)</td>
<td>EUR-LEX</td>
</tr>
<tr>
<td></td>
<td>Browse through interactive map</td>
<td>geodata.gov.gr</td>
</tr>
<tr>
<td></td>
<td>Browse data through charts</td>
<td>budget.msh.gov.il</td>
</tr>
<tr>
<td></td>
<td>Simple Document list</td>
<td>Luxembourg Stock Exchange</td>
</tr>
<tr>
<td></td>
<td>View popular / new searches</td>
<td>United Nations Data</td>
</tr>
<tr>
<td><strong>Data Provision</strong></td>
<td>Download data as a file</td>
<td>Israel Land Administration</td>
</tr>
<tr>
<td></td>
<td>Select Export file format (different from original format)</td>
<td>Organisation for Economic Co-operation and Development (OECD)</td>
</tr>
<tr>
<td></td>
<td>View data in HTML</td>
<td>Borough of Trafford (UK) Open Data Catalog</td>
</tr>
</tbody>
</table>
### Table 4-3. Consolidated view of technologies used by the analysed OGD portals

<table>
<thead>
<tr>
<th>Specification / Utility</th>
<th>Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata standards</td>
<td>Dublin Core</td>
</tr>
<tr>
<td></td>
<td>CKAN</td>
</tr>
<tr>
<td></td>
<td>INSPIRE</td>
</tr>
<tr>
<td>Specification / Utility</td>
<td>Technologies</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>ESRI ArcGIS Metadata – MXD</td>
</tr>
<tr>
<td></td>
<td>OECD metadata</td>
</tr>
<tr>
<td></td>
<td>Euro-SDMX</td>
</tr>
<tr>
<td></td>
<td>SDMX</td>
</tr>
<tr>
<td></td>
<td>CSMD</td>
</tr>
<tr>
<td></td>
<td>UN comtrade metadata</td>
</tr>
<tr>
<td></td>
<td>Health Data Exchange (GHDx)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Linked Data</th>
<th>RDF Support</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N-Triple stores</td>
</tr>
<tr>
<td></td>
<td>SPARQL Support</td>
</tr>
<tr>
<td></td>
<td>Custom tools by Rensselaer Polytechnic Institute (RPI) to transform CSV to RDF</td>
</tr>
<tr>
<td></td>
<td>Linked Data Mashups</td>
</tr>
</tbody>
</table>

| Supported Data formats       | Html, pdf, xls, doc, csv, xml, RDF, ppt, txt, JSON object, kml, kmz, mxd, tiff, png, ods, zip, tsv, txt |
| CMS / Platform               | Drupal, Socrata, MediaWiki, Wordpress, Plone, Joomla, PostGIS, CKAN engine |

| User Interface / Web Technologies | Html, Adobe Flash, JQuery, JavaExts, Struts MVC framework / Java, PHP (Joomla), Python, MooTools, SOBI 2, ExtJS, JSP, Prototype, script.aculo.us, Microsoft ASP.NET, Nginx, Varnish, Piwik, Adobe Flex |
| Mashups / Visualizations       | Google Maps, Bing Maps, ArcGIS, OpenLayers, GeoEx, OpenStreetMap, MapFish and MapServer, Google Charts |

<table>
<thead>
<tr>
<th>API</th>
<th>RESTful Web Services</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CKAN API</td>
</tr>
<tr>
<td></td>
<td>API query builder (code generation tool)</td>
</tr>
</tbody>
</table>

The investigation of current public data and knowledge sources leads to a number of key conclusions that will drive the principles and basic elements of the service infrastructure:
1. There is a vast amount of available OGD dispersed among a great number of constantly increasing Public Data and Knowledge resources (Public agencies and official, Open data initiatives, governmental portals, ministries, etc.). There is a great potential value for the effective exploration, management and distribution of this information towards the research communities in a unified manner.

2. The majority of OGD portals use non-standardized metadata, while there is no universally established metadata specification framework for OGD. There is a need for standardisation and harmonisation of metadata, semantics and ontologies, in order to ensure interoperability within and across Public Data Sources.

3. Searching and navigating through public sector data is difficult. State-of-the-art data discovery technologies need to be examined and applied to OGD portals.

4. The Linked Data paradigm is adopted only by recently developed initiatives, whereas traditional and longstanding public data sources are reluctant to adopt Linked Data or similar Semantic Web technologies.

5. Security threats such as data manipulation issues are not explicitly addressed in the majority of OGD portals, whereas security measures are limited to the typical authentication frameworks and configurations set in the web application server level or in the underlying platform level. The majority of OGD portals do not require registration or any special authentication and authorization to provide information, with a view to improving the user experience and increasing the user base.

6. Public datasets are provided in a national level only. There is a strong need to support cross-country datasets that would enable European and International researchers and citizens to utilize all available public datasets regardless of the language barrier.
5. Defining and Designing the Solution

Working on the outcomes of the previous chapter and working towards answering research question 3 we can recognise the following considering the opening of OGD:

In the last decade there has been an increase in activities and investments towards opening up of public sector information to the public, in order to be used for scientific, commercial and political purposes (McDermott, 2010; Janssen, 2011; Kundra, 2012; Commission of the European Communities, 2007; 2009; 2011; Luna Reyes et al., 2014; Zuiderwijk & Janssen, 2014; Dekkers et al., 2006; European Union, 2003). In particular, OGD can be valuable for scientific research in many different domains (e.g. in the social, political, economic, administrative and management sciences), and can contribute critically to the development of the ‘e-Science’ paradigm (Commission of the European Communities, 2007; 2009). OGD can be used by citizens and journalists for gaining better and deeper understanding of and insight into the activities and spending of government agencies. This should result in evidence-based, mature and effective political processes. Also, OGD can have a positive impact on innovation and economic growth, as they enable the development of new applications, products and services.

Many OGD platforms, often in the form of portals, have been developed and operated by government agencies. The existing first generation of OGD digital infrastructures offers mainly basic functionalities for searching and downloading data to the users of these data, and for uploading data by their providers. The majority of these portals offer simple free-text search and theme-browsing functions for the discovery of datasets. Only some portals have recently taken advantage of Semantic Web by providing semantically enriched discovery services such as performing SPARQL queries. Most OGD platforms limit their data provision services to simple download functionality, and only a few of them provide functionality to view datasets on a map or various types of charts.

There are no functionality for processing the datasets in order to improve them, adapt them to specialized needs, or link them to other datasets (public or private), and then for uploading-publishing new versions of them, or for uploading users’ own datasets. Furthermore, the ‘Linked Data’ paradigm is adopted only by some recently
developed initiatives, whereas traditional and longstanding public data sources are reluctant to adopt Linked Data and Semantic Web technologies. Also, only a few OGD platforms collect the needs of users for more datasets in a formal and systematic manner. The majority have only general-purpose feedback web forms for collecting comments and suggestions from users, which typically concern the technical aspects of the platform rather the actual datasets provided). Only some portals include datasets rating and commenting. Another important weakness is the limited functionality for networking, interaction and collaboration among users, in order to generate value from the provided datasets.

In general this first generation of OGD platforms follows mainly the Web 1.0 paradigm aimed at making OGD available. There is a clear distinction between content producers (public administrations) and content users (research communities, businesses and citizens), and limited interaction and collaboration among them. They place emphasis on the provision of large number of datasets, but offer no functionality for users supporting the generation of value. Our research aims at filling the above gaps and overcoming the above weaknesses, through the design of a second generation of OGD platforms, which combine and integrate opening data on one hand with exploiting the main characteristics of the social media in order to stimulate and facilitate value generation from the former.

In this chapter we will provide the list of requirements of this advanced infrastructure based on scenario building by experts which are validated and ranked by other experts and data researchers.

Finally, the infrastructure should focus on the scientific base of ICT-enabled governance for open data, thus harnessing open data sources and methodologies for annotating, visualising and making open data available to scientists and citizens.

It intends to provide an extensive data service suite consisting of:

- Visualisation tools
- Personalised services (user accounts, personalised / discipline specific / context specific workspaces and preferred data sets etc)
- Community building and collaboration tools
- Intuitive navigation and assistance especially for users that need to access unfamiliar data formats and content (e.g. cross disciplinary researchers, policy makers tackling unfamiliar problems, students and citizens)
Data mining and knowledge discovery tailored to user needs and specific data types

Having in mind the aforementioned aspects, the present chapter provides a number of different Use Cases and Service Scenarios, which are developed in strong cooperation with the experts from the THE PLATFORM project partners. Based on the Use Cases, Service Scenarios (e.g. e-government simulation scenarios) are defined and set the primary basis for defining requirements of the service infrastructure and the governance tools as well as categorise the main stakeholders and users of the project. These scenarios consist of different Use Cases based on the needs of the users and they provide the services design for effective and user-friendly participation, visualisation or further processing of governmental information and knowledge, enabling research, education, participative governance and administrative transformation.

5.1. Methodological Approach

5.1.1. Scenario Based Design

Scenario as a term lies his origin in the early theatre and film terminology. From that point of view scenario is a synonym to the screenplay or manuscript. The basic elements of a scenario include the actors (users), the scene (context), and the scheme (the story including the background, tasks, goals and action). These elements of a scenario are virtually the same both in the original notion and in the scenario definition applied as a method in user-centric system development process (Carroll, 1995).

In the context of this study, scenarios form a tool for user requirements capture, content of use analysis, concept definition and overall user involvement. This section will briefly outline the basic thinking, reason and role behind the utilisation of scenarios in system or service development work.

Scenario building is a widely accepted way to incorporate and generate design ideas for new systems and products. It is well suited to the design of new prototypes and concepts, where the context of use is variable (Rosson & Carroll, 2002). The value of the Scenario-based Design is that it concretises something for the purpose of analysis and communication. Concreteness enables designers and users to deal with complicated

---

22 [http://www.the-platform-project.eu](http://www.the-platform-project.eu)
and rich situations and behaviours in meaningful terms, and to understand better the implications of particular design solutions for performing realistic tasks (Carroll, 1995). The main elements a scenario should contain are:

- Description of a user / user group: the characteristics of those who will use the system.
- Contextual setting: where and how current or expected conditions could influence the use of the system.
- Circumstance: the trigger (the why) of using the system.
- Goals: motivations of users and critical success factors.
- Artefact: devices that might be used and how these might work.
- Time Frame: when and how long, at what frequency, and in which life time period.
- Qualities: the physical form factor of possible devices (portability, weight etc.)

Aggregately, scenarios as a design method could be used to:

- Describe problems to be solved.
- Catalyse interaction within a project and design teams, thus improving team building.
- Facilitate user involvement in the early design.
- Support collaborative design where all the participants don’t need to know the technology.
- Help in transferring and explaining tricky design ideas.

Finally, the analysis of scenarios provides room for discussion, help the involved persons prioritise the problem areas or potential solutions and aligns expectations and mindsets of the different disciplines and stakeholders within a project.

5.1.2. Description Template

The Scenario Template consisted of three sections of fields to be filled in, aiming at extracting useful information for defining requirements. The use of template should ensure the consistency of identified requirements covering all the project aspects have been defined till now. The template format has been designed based on the partners’
experience in such projects, having a common understanding of the used parts and terminology.

To identify use cases the following steps will be necessary:

- On a very abstract level: define logical system boundaries (Black-Box-View)
- Detect relevant roles of the systems and their communications
- Detect all types of usage, all types of operated infrastructures, incoming data etc. Every type of use became a use case
- Identify the beginning events or actions that start the use case.
- Define the conditions and constraints that will stop or end a use case.
- The Use Cases will be designed in UML and described textually.

Based on the Use Cases, Service Scenarios (e.g. e-government simulation scenarios) are defined.

**Section 1: General Information and Actors of the proposed scenario**

In this section every OGD expert proposes a descriptive scenario being aware of the context of users’ needs, services design provision for effective and user-friendly participation, visualisation or further processing of governmental information and knowledge, enabling research, education, participative governance and administrative transformation. In this section are also defined the actors involved as well as their roles.

- **Use Case Description**
  - Give a textual description of the use case
- **Goal**
  - What is the overall objective of the use case?
- **Topic**
  - What are the main subject areas of the use case? E.g. social science, environment?
- **Primary Actors and Roles**
  - Who are the people and organisations directly involved in the use case? What are their roles?
- **Secondary Actors and Roles**
  - Who are the people and organisations indirectly involved in the use case? e.g. subject groups, target audience What are their roles?
Section 2: Data, Metadata, Owners and Languages involved

In the present section is described all information needed in each scenario about data, metadata, languages, owners and rights.

- **Data sets**
  - What are the sets of data involved, with their size, location etc?

- **Data Services**
  - What are the services which provide the data, with their location? Is Linked Data involved?

- **Data owners**
  - Who has ownership and control of the data?

- **Data Access Rights**
  - What are the rights and conditions of use of the data?

- **Data Formats**
  - What are the data formats involved, e.g. XML, RDF, data models etc?

- **Metadata formats**
  - What are the metadata formats and standards used to describe the data?

- **Nationalities and Languages involved**
  - What countries or regions are involved in the use case? What natural languages are used for the data and metadata?

Section 3: Special Requirements of the proposed scenario

This section describes the additional requirements of the proposed scenario in order to clarify the main technological and other issues should be aware of the implementation process.

- **Trigger**
  - Why is the use case initiated by the stakeholder.

- **Pre-condition list**
  - What conditions must apply before the use case can start?

- **Post-conditions list**
  - What do we expect to be any new conditions resulting at the end?

- **Primary Flow**
  - What are the steps in the execution of the use case?
• **Error Flow**
  o What error states can arise? How are they managed?

• **Additional Requirements list**
  o Any additional requirements not covered above

• **Notes & Outstanding Issues**
  o Comments or issues arising in the use case which need further consideration

• **Use Case/ Flow Chart Diagrams (UML)**
  o A visual representation of the use case using one or more UML diagrams

5.2. **Usage Scenarios**

In total 10 scenarios were identified, analysed, deliberated during an open workshop\(^{23}\) commented online, and formally expressed, as following:

<table>
<thead>
<tr>
<th>Nr</th>
<th>Scenario</th>
<th>Public Servant</th>
<th>Researcher</th>
<th>Citizen</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC-01</td>
<td>Storing, Linking, Annotating and Visualising a PS data set</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC-02</td>
<td>Getting the metadata specifications (for applying them in my own systems)</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC-03</td>
<td>Getting useful information (through browsing datasets or visualisations)</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SC-04</td>
<td>Getting data for my research work</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SC-05</td>
<td>Linking my system with the platform, for uploading data</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>SC-06</td>
<td>Linking my system with the platform, for downloading data</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SC-07</td>
<td>Storing data in draft form (to be further curated)</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SC-08</td>
<td>Put data and annotate them according to platform standard and my needs</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>SC-09</td>
<td>Put my needs for Public Sector data</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>SC-10</td>
<td>Get information and training on metadata, open data, platform usage</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

---

\(^{23}\) At the IFIP eGOV Conference in Delft, The Netherlands, 29\(^{th}\) August 2011
The scenarios implement both public-sector-to-citizens and the opposite flow of information, as following:

<table>
<thead>
<tr>
<th>Information flow /Information Type</th>
<th>Number of Scenarios addressing the issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downwards: From Public Sector to society</td>
<td>7</td>
</tr>
<tr>
<td>Upwards: From Society to the Public Sector</td>
<td>4</td>
</tr>
<tr>
<td>Data specific issues: getting, putting, curating</td>
<td>8</td>
</tr>
<tr>
<td>Process and knowledge specific: learning, drafting, approving</td>
<td>2</td>
</tr>
</tbody>
</table>

5.2.1. Storing, Linking, Annotating and Visualising a PS data set

5.2.1.1. General Information and Actors

**Description:** A Public Sector Official is working in the Ministry of Citizen Protection and is in charge of evaluating public sector organizations. The minister delegates the evaluation of the police central departments' effectiveness in solved criminal cases to him. The Official learns about the existence of the platform by one of his colleagues who works in the technical department of the organization. So, they decide to use the platform to accomplish their purpose.

Searching in the platform for the necessary datasets, they find: criminal information per region in XLS format published from the Ministry of Justice and case solved per region in XML format published from the Police Headquarters. Looking into datasets they figured out that all the information they need was represented in those datasets, but they first have to face the problem of different dataset types. In order to achieve their purpose the platform overcomes this barrier by providing tools for the transformation of one of the existed datasets either XML to XLS or the opposite.

The Official decided that the latter is more convenient, because he wants to apply statistical analysis on those data. So, they used the XML to XLS parser function. Now they have the desired datasets in the same format and the only thing that remains is to combine the right columns and rows. The platform provides a function in order to achieve this combination by filtering the unwelcome columns and rows.
After the accomplishment of information acquiring, official decided that this information could be useful for the citizens and for that reason uploaded the transformed dataset in the platform by following the Scenario 3.5.

He also wants to find an easy way of demonstration (not only numbers) for saving minister’s time. For that reason he exploits the platform visualisation tool which provides categorization for map depiction (by country region, by country borough, by continent, etc). Thus, he chooses the “by region” map representation for his country.

**Goal:** Describing one aspect of Public Administration (PA) behaviour and how the platform corresponds within its functions.

**Topic:** Registrations that are used for policy questions.

**Primary Actors and Roles:** The primary actor is a Public Sector Official who needs information to make policy or to make a decision. The system also has a primary role in the scenario.

**Secondary Actors and Roles:** Secondary actors are the data owners of each PA (Ministry of Interior, Police Headquarters, Ministry of Justice). The Content Management System (CMS) expert and the Web editor for each PA have a secondary role.

### 5.2.1.2. Data, Metadata, Owners and Languages

**Data sets:** The datasets that are used by this scenario can be found at geodata.gov.gr:

1) Criminal information per region
2) Case solved per region
3) Geospatial data for Greek regions

**Data Services:** Data sets can be retrieved either manually (e.g. by following a link to the data set at the PA website and by downloading it) or programmatically (e.g. when the data source link to the data set points to a SPARQL endpoint of an RDF store). 1) criminal information per region published from the Ministry of Justice 2) case solved per region published from the Police Headquarters 3) Greek regions geospatial data from the ministry of interior. For the additional case of storing a physical (transformed) copy of the original data set, linked data is involved as it is the result format of the transformation and annotation.
Data owners: Ownership and control remain in each PA who publishes the datasets. The transformed dataset will hand over ownership to ministry of citizen's protection or to the platform.

Data Access Rights: The data sets and hence the access rights remain under control of each PA. The conditions of use are captured in a license. The license type may be part of the metadata.

Data Formats: The data formats are XLS, XML, KML.

Metadata formats: It will use the ‘predifined metadata’ formats for the final result. Each public source has its own metadata format (e.g geodata: metadata Standard Name ISO19115;)

Nationalities and Languages involved: The specific data are in Greek and the involved nationality is Greece. The metadata language has to be in English.

5.2.1.3. Special Requirements

Trigger: The need of a public sector to combine different datasets with an easy and acceptable way.

Pre-condition list: To complete this scenario we assume that functions of searching, parsing, visualising, uploading datasets are up and running. It is also needed to provide a well-documented API (the documentation of the API contains detailed information about the API protocol and the data formats (e.g. JSON/REST) to be applied) and all the necessary information to use the functionality of the platform. For Public Servant (PS), to complete the scenario with publishing the annotated and enriched data, we assume that has already an account and of course has the technical competencies to integrate with the platform.

Post-conditions list: The new information from combined datasets and the new dataset itself.

Primary Flow:

1) Public Sector Official searches for necessary datasets filling search criteria.
2) The platform returns the results with information for each of them.
3) Official select the datasets
4) Select final format ("Parse in XLS").
5) He presses the “Combine/Produce” button.
6) The platform using metadata schema (vocabulary) decides automatically how to combine the datasets and return the results into the selected format.

7) PS selects the right columns by filtering or using any aggregate functions.

8) PS presses the "Download dataset".

9) He uploads the final combined dataset after filling some more information into the platform.
   a. Use the SC-05 case

10) He presses the “Map Visualisation” tool.
   a. He follows the wizard and the instruction to show the data using maps per region.
   b. He chooses Greece and by region. He sees/download the final results.

**Error Flow:**

1) THE PLATFORM has not possessed the datasets for "by region" visualisation. The THE PLATFORM system suggests the user to select another visualisation method.

2) THE PLATFORM could not provide the “Convert” function on this particular dataset. The THE PLATFORM system suggests the user to select another conversion system can perform.

**Additional Requirements list:** None.

**Notes & Outstanding Issues:** None.
5.2.1.4. Use Case Diagram

Figure 5-1: Use Case Diagram – Storing, Linking, Annotating and Visualising a PS data set
5.2.1.5. Sequence Diagram

![Sequence Diagram](image_url)

Figure 5-2: Sequence Diagram - Storing, Linking, Annotating and Visualising a PS data set

5.2.2. Getting the metadata specifications

5.2.2.1. General Information and Actors

**Description:** The IT department of a Ministry decide to adopt the platform Metadata Specifications. They believe that it would be easier to communicate with platform and maybe to use a similar approach of data set combination in their own system in order to improve their services. So they get the specification from the portal. The platform provide this information to every public servant in order to reach the harmonization of metadata sets.

The Platform supports the PS with information (versioning, updates, successfully cases etc.) and explanation (full documented, help) about the specs and more over provides a specification search tool.
Public Servant (PS) creates an account and fills information about the use of metadata. The platform use this information to communicate with PS and to inform about any new changes or updates of the metadata specifications. The platform also provides a feedback form to PS to note possibly problems and to gather new proposals.

**Goal:** Accept the platform metadata as ad hoc standard

**Topic:** The use case is independent of a particular domain.

**Primary Actors and Roles:** The primary actor is any information specialist from IT department of a Ministry (more general any public sector) who decide to adopt the platform metadata schema.

**Secondary Actors and Roles:** Secondary actor is the The platform administrator who is responsible for update and monitoring of the metadata management.

### 5.2.2.2. Data, Metadata, Owners and Languages

**Data sets:** Any dataset.

**Data Services:** The services are provided by the platform.

**Data owners:** The data remains to the owner but the metadata to the platform.

**Data Access Rights:** In order to use metadata an open licence will be provided by the platform.

**Data Formats:** Any data format.

**Metadata formats:** The platform metadata will use XML.

**Nationalities and Languages involved:** English.

### 5.2.2.3. Special Requirements

**Trigger:** The Lack of common accepted specs in the public's servants applications and the need for advance services.

**Pre-condition list:** We assume that public sector actor has already registered to the platform. He has an account and fills a form with his interests. The public sector has to understand the technical instructions so it is assumed to have a basic knowledge about metaschema analysis. The PS also has to agree with the licence before adopt and use the platform metadata.

**Post-conditions list:** The metadata specs will be accepted as standard.

**Primary Flow:**
1. Communicate with the platform
2. Search for the specs using the tool.
3. PS registers to the platform (Create Account and fill Info)
4. PS follows the instructions
5. PS gets the documentation
6. PS ask for help/support
7. PS gets and finally integrates specs
8. Following the platform's notifications about specs
9. Try to combine data with the platform or other services
10. Send comments or suggestions to the platform (feedback)

**Error Flow:** There is a possibility the metadata to not cover all PS’s needs.

**Additional Requirements list:** Not Specified.

**Notes & Outstanding Issues:** Not Specified.
5.2.2.4. Use Case Diagram

![Use Case Diagram](image)

Figure 5-3: Use Case Diagram - Getting the metadata specifications
5.2.2.5. Sequence Diagram

![Sequence Diagram](image)

Figure 5-4: Sequence Diagram - Getting the metadata specifications

5.2.3. Getting useful information (through browsing datasets or visualisations)

5.2.3.1. General Information and Actors

**Description**: Joseph is an Israeli senior citizen who lives in the Haifa suburbs. Now having retired from work, Joseph has much more time to be active in environmental issues. Specifically he's bothered about the long-standing pollution issue of the nearby Kishon river.

Joseph would like to better understand the pollution problem, what facts have been published on it, what governmental decisions were made and have there been any progress since.
Joseph heard about the data platform and decides to try it out. Initially he is not sure what already exists and is interested to explore the repository. He starts with a simple keyword based search for "kishon + pollution". The search results are presented using facets plus an ordered list of dataset items. This presentation assists in understanding the inventory of related datasets. It also educates Joseph about the different dimensions (metadata) of the datasets within the result set, and further possible navigation and drill down opportunities to narrow down on specific dataset items.

Using hyperlinks and sliders he narrows down to find datasets from the last couple of years, generated by the Israeli Ministry of Environment with geospatial measurement data. He remains with a couple of datasets. One of them is a PDF report that summarizes the measurements from last year. The other dataset has a map icon. This is actually a machine readable version of the PDF information that was scraped/curated by a community of citizens. Joseph is happy to learn about the existence of this community and plans to join their facebook group. In addition he clicks on the map icon to visualize the dataset.

The visualization renders to pollution measurement data using points overlaid on map. The legend shows the colours and their semantics. Joseph navigates the map and focuses on the red area with higher colour intensity. This area is the main source of pollution. He wonders what other facilities and factories are nearby. He selects the area and hits the "search" button. This searches the repository for additional layers with geospatial information that is contained/intersects with the selected area. From the search results, he adds the factories layer. Now he can see the labels of the factories rendered as an overlaid layer on the map. It now becomes apparent that two factories are nearby. He selects these factories and hit the search button again. In addition to layers, the search results contains a section called 'Entities'. Under this section he sees the list of factories, for each a list of relationships, tied to a list of datasets. For example "factory1 is mentioned in Ministry of Justice - Court Decisions 2008-2010". This seems interesting to further explore... he open the dataset to read more...

In addition as he hovers over Factory2, additional information/pointers e.g. to Wikipedia around brought in a tooltip dialog. This was possible since a previous user has annotated Factory2 in a previous navigation.

**Goal:** Describing how a citizen may explore the repository visually to find available datasets for his needs.
**Topic:** The topic concerns environmental aspects.

**Primary Actors and Roles:** The primary actor is the citizen inquiring for data. The System has a primary role in serving the citizen exploration and navigation needs.

**Secondary Actors and Roles:** Secondary roles, are data curators who were involved in curation of the datasets, data owner role by public servants.

### 5.2.3.2. Data, Metadata, Owners and Languages

**Data sets:** This scenario involves 3 datasets.

1. PDF report of pollution data in Kishon River, published on the Ministry of Environment web site
2. Enriched/Curated machine readable geospatial dataset, was published by group of citizens based on the PDF report. Hosted on the site
3. The factories geo layer that is published by the Israeli Mapping Service and hosted on the data.gov.il site.

Both of these datasets are relatively small in Size (<2MB)

**Data Services:** Datasets are available through the following methods

4. Download via hyperlinks
5. APIs that other components can use (e.g. for visualization component that needs to bring information with respect to datasets)

Linked data is involved to capture relationships at the metadata level, and can be added as annotation by users, e.g. citizens adding linked data to dbpedia entities such as Kishon River, Factory2. For example an initial lookup based on the label "kishon" http://lookup.dbpedia.org/api/search.asmx/KeywordSearch?QueryClass=place&QueryString=kishon and marking the single query result into a linked entity.

**Data owners:** The ownership of the original pollution data is held by a public servant in the Ministry of Environment. The curated machine readable datasets is owned by the trusted community who helped curate and enrich the dataset. The factories geo layer is owned by the Israeli mapping service.

**Data Access Rights:** The access rights are documented in the metadata. In this case the pollution data has open data license. The geo layer has a free license for non-commercial use.

**Data Formats:**

1. PDF for textual data
2. RDF for the curated data
3. KML for layer information

**Metadata formats:** The platform common metadata will be used. For geo data bounding box information is needed, which is likely to be an extension of the common metadata or by referencing other standards for geo datasets.

**Nationalities and Languages involved:** English, Hebrew

### 5.2.3.3. Special Requirements

**Pre-condition list:** None.

**Post-conditions list:** None.

**Primary Flow:**

1. Joseph enters a keyword in the search box and presses enter.
2. Results are presented as an order lists of dataset items + links to facets that allow to drill down.
3. Joseph adjusts a date slider to narrow down on date.
4. Joseph clicks on environment link in the domain facet.
5. Joseph clicks on a dataset, an overview of the datasets details and main metadata attributes is presented. He sees the owner information of this dataset (PDF).
6. Joseph clicks on the "curated version" link to see a machine readable version of this dataset.
7. Joseph clicks on the map icon to visualize the dataset to view the dataset on a map.
8. Joseph uses interaction to navigate the map.
9. Joseph selects an area and clicks search. The search brings additional datasets within the selected area bounding box.
11. Joseph selects a couple of factories and clicks search by entity to find additional datasets that mention the factories.

**Error Flow:** Not needed.

**Additional Requirements list:** Datasets use the same coordination system. This may require transforming between various coordination systems used in geo datasets.

**Notes & Outstanding Issues:** This is a rather complex scenario that includes several use cases.
5.2.3.4. Use Case Diagram

Figure 5-5: Use Case Diagram - Getting useful information (through browsing datasets or visualisations)
5.2.4. Getting data for my research work

5.2.4.1. General Information and Actors

**Description:** Julia is a PhD candidate at the Department of Experimental Psychology of the University of Cambridge and wants to study the effect of natural environmental conditions on human behaviour. She currently wants to publish an academic paper that examines the relationship between the level of ultraviolet radiation (UV) and criminal activities. To validate this thesis, she will perform Hypothesis Testing with the method of Structural Equation Modeling analysis.

To find the related datasets she uses the search function of the platform. She finds a dataset with the levels of ultraviolet radiation in specific geographical areas for 3 years. She then uses the related function in the platform to "Combine" this data with the levels of crime rates in the same regions for the same time span (spatial query). The platform returns a combined view of the results. Julia downloads them in a csv format to open them in SPSS.

**Goal:** Retrieving data for scientific purposes.

**Topic:** The topic concerns Environment and Psychology.
**Primary Actors and Roles:** Julia, a researcher at the Department of Experimental Psychology of the University of Cambridge is the primary actor.

**Secondary Actors and Roles:** None.

5.2.4.2. **Data, Metadata, Owners and Languages**

**Data sets:**
The following datasets are involved:
1) Level of Ultraviolet Radiation per year / specific region
2) Level of crime rates per year / specific region

**Data Services:**
The data services provided are:
1) World Ozone and Ultraviolet Radiation Data Centre (WOUDC) - Canada
2) Eurostat – European Union

**Data owners:** The Data owners are the agencies that upload the data to the World Ozone and Ultraviolet Radiation Data Centre as well as the European Union (Eurostat).

Data Access Rights: Regarding Eurostat data downloading and reproduction of data/documents for personal use or for further non-commercial or commercial dissemination are authorised provided appropriate acknowledgement is given to Eurostat as the source, and subject to the exceptions/conditions hereinafter specified.

**Data Formats:** The involved data format is Microsoft Excel Format (xls).

**Metadata formats:** The metadata format used to describe the dataset is the Euro-SDMX Metadata Structure (ESMS).

**Nationalities and Languages involved:** The involved nationalities are United Kingdom, Greece and Germany. The language used is English.

5.2.4.3. **Special Requirements**

**Pre-condition list:** None.

**Post-conditions list:** None.

**Primary Flow:**
1) Julia searches for the UV data in specific regions for the last 3 years.
2) The platform returns the results.
3) Julia presses the "Combine" function.
4) She presses the "Map function".
5) She draws a rectangular around the same regions.
6) She presses "Search within this area".
7) She searches for the "Levels of crime rates" in the same regions for the last 3 years.
8) The platform combines the datasets.
9) She presses the "Download dataset".
10) She chooses "xls" format.

**Error Flow:** The platform cannot find data for "Levels of crime rates" for the specific region. The system suggests the user to broaden the selected region.

**Additional Requirements list:** None

**Notes & Outstanding Issues:** The complexity of the service that provides combination of spatial data could be high. How will the platform handle the linking of geodata sets when original data sets have been created under different measures (country level vs. specific region).

5.2.4.4. Use Case Diagram

![Use Case Diagram](image)

**Figure 5-7:** Use Case Diagram - Getting data for my research work
5.2.4.5. Sequence Diagram

![Sequence Diagram](image)

**Figure 5-8: Sequence Diagram - Getting data for my research work**

5.2.5. Linking my system for uploading data

5.2.5.1. General Information and Actors

**Description:** A public authority (PA), the Federal Statistical Office, frequently publishes structured data sets on its website ("www.statistics.xy"). The data sets are a valuable asset for various research communities.

The authority already has appointed an Open Data designee who has learned about the platform. The Open Data designee advises Mr. Smith of the IT service department to explore the technical means of hooking up with the platform in order to regularly publish statistical data not only on www.statistics.xy but to make it also automatically retrievable via The platform.

Mr. Smith visits the website and learns about the API - a Web-based open API for uploading metadata to the metadata registry which is at the heart of the Open Data platform. The documentation of the API contains detailed information about the API protocol and the data formats (e.g. JSON/REST) to be applied to upload metadata.

Having talked to his colleagues Mr. Smith comes up with the idea of extending the
traditional Web publishing process at the Federal Statistical Office by including the option to automatically publish metadata for the data sets on the platform.
The public servants that work as editors of the Statistical Office website make use of a central Content Management System (CMS) to publish information and to manage the content on www.statistics.xy. The CMS also supports an approval process as an additional step when publishing information on the website. Mr. Smith, who is technically experienced with the CMS, decides to extend the editorial forms that are used by the editors to upload datasets to www.statistics.xy. Having explored the metadata schema (i.e. the metadata for a data set), Mr. Smith notes that luckily several metadata elements, such as "author", "department" or "release date" are already part of the existing editorial forms and hence are already captured in the publishing process.
Mr. Smith extends the forms with the additional metadata elements, such as "format" and "geographic coverage". Mr Smith then extends the publishing routine of the CMS with an additional invocation of the open Publish API.
Whenever an editor fills the editorial forms and pushes the button to publish a data set (e.g. a spreadsheet) on the public Web server of the Federal Statistical Office, the metadata together with the link to the original data set on www.statistics.xy is also published in the metadata registry. Citizens and researchers visiting the data portal are now able to find the data sets of the Federal Statistical Office and access it at its original source. Government authorities keep their sovereignty over the data sets, e.g. in order to amend them whenever necessary.
In addition, the platform also provides facilities to upload the content data and not only the metadata as described above. This case might be useful (Carroll, 1995) when the organisation publishing the data does not have a Web site where to publish the data or (Rosson, M.B., Carroll, 2002) when content data should be provided in a semantically richer format such as linked data on the platform. In the second case, the editor uses a linked data transformation service provided by the platform. The services allows the editor to define a mapping from the existing data structure (e.g. column and row names) to a linked data set (RDF) and then store the transformation result as in the (triple) data store. Thus, not only the metadata of the data set is stored on the platform but also actual data sets (e.g the original data set transformed to RDF) is stored on the platform.
The more government authorities make use of the open Publish API in order to link their systems with the platform, the more becomes a one-stop shop for researches looking for open governmental data.

**Goal:** Provide public authorities with a means to link their data provisioning process with the platform.

**Topic:** The use case is independent of particular application domains or areas.

**Primary Actors and Roles:**
- Open data designee of the public authority
- Technical expert at the IT Service Department of the PA
- Public servant working as Web editor and/or data owner

**Secondary Actors and Roles:**
- The platform operator, e.g. for troubleshooting

### 5.2.5.2 Data, Metadata, Owners and Languages

**Data sets:** Any kind of data set can be made retrievable through the platform. The typical case would be that the data sets remains on the Web sites of the public authorities, only metadata (including the link to the original data) is transferred to the platform. In this case the size of the data sets depends on the capabilities of the PA systems (Web server).

For the additional case of uploading the data set itself (e.g. transformed to RDF), the platform should provide sufficient storage capacities.

**Data Services:** Search for a data set can be performed either by the end user using a Web-based front end or programmatically via the Search API. The data sets themselves can be retrieved/queried either manually (e.g. by following a link to the data set at the PA website and by downloading it) or programmatically (e.g. when the metadata points to a SPARQL endpoint of an RDF store).

**Data owners:** When uploading only the metadata including a link to the original data set, the ownership and control of the data remains at the original data owner, i.e. the PA. In the case of additionally storing the data itself on the platform, the ownership and control can remain as well at the original data owner, however under the conditions of the platform.

**Data Access Rights:** When uploading only the metadata including a link to the original data set, the data sets and hence the access rights remain under control of the PA. The
conditions of use are captured in a license. The license type may be part of the metadata. When storing a physical (transformed) copy of the original data set, the access rights are under control of the platform operator.

**Data Formats:** Any data format, for storing data on the platform linked data (RDF) is preferred.

**Metadata formats:** The metadata format is used to describe the data set.

**Nationalities and Languages involved:** Data and metadata most probably is in the national language of the PA. Having linked the PA system (e.g. the CMS) with the Publish API of the platform, the front-end for the public servants should present the metadata form in the national language while in the background the PA system (e.g. the extended CMS) maps it to the metadata schema.

### 5.2.5.3. Special Requirements

**Pre-condition list:**

- The Publish API (more specifically the service) is up and running
- The Publish API is well documented
- The PA has the technical competencies to integrate with the Publish API.

**Post-conditions list:**

Data sets published by the PA (either metadata or metadata and data) are becoming visible on the platform.

**Primary Flow:**

1. Open Data designee of the PA (here “Federal Statistical Office”) intends to link with the platform in order to make the PA’s data sets retrievable for the THE PLATFORM users.
2. Expert from the PA’s IT department accesses the Publish API documentation and the metadata documentation.
3. PA IT expert extends the publishing forms of the PA’s CMS with additional metadata fields as specified by the metadata specification.
4. IT expert extends the publishing routine of the CMS with a call of the Publish API.
5. (a) A public servant uses the CMS to publish a data set (e.g. a spreadsheet) on the PA’s Web server. In order to do so, the public servant fills the extended publishing form provided by the CMS. The CMS calls the Publish API. The metadata together with the link to the original data set on the PA’s website (e.g. www.statistics.xy) is published in the metadata registry.
(b, optional) The public servant uses tools and services (possibly also provided by the platform) to semantically enrich and transform the original data set to RDF. The public servant proceeds as described in (a). This time in addition to the metadata which is published to the metadata registry, the RDF data is also uploaded to the dedicated data store (triple store) of the platform.

6. The platform users find the PA’s public data sets by using the search for data sets.

**Error Flow:** To be specified when the scenario is further analysed on a more fine grained level.

**Additional Requirements list:** None.

**Notes & Outstanding Issues:** In addition to the provision of an API which allows for linking PA systems to upload metadata and data it should also be elaborated on how to update data/metadata. This will be analysed in later stages when the scenario will be analysed on a more fine grained level.
5.2.5.4. Use Case Diagram

Figure 5-9: Use Case Diagram - Linking my system for uploading data
5.2.5.5. Sequence Diagram

![Sequence Diagram](image)

Figure 5-10: Sequence Diagram - Linking my system for uploading data

5.2.6. Linking my system for downloading data

5.2.6.1. General Information and Actors

**Description:** "Scenarios of data from funding agencies on funding projects, outputs from research (e.g. papers, products, patents) etc.

**Researcher looks for prior art.**

A researcher (in academia or in the commercial sector) is looking for prior art related to an area of research. They seek to collate information from a number of linked data repositories, accessed via the portal. There are two stages:
1. Searching a CERIF (Common European research Information Format) CRIS (Current Research Information System) to identify projects which have been undertaken in a particular research area.

2. These provide links to papers describing work done. The researcher then accesses one or more linked data publication repository (or an aggregator service) to find the appropriate publications, accessing text as appropriate.

   The publications may have links to further publications, which may spark off further searches.

**Validation of results**

A researcher accesses a paper in a journal, and decides that results need to be examined in more detail to determine their validity.

1. She seek the supplementary material (within the journal system). This is not sufficient, but provide links to primary data within a facilities repository, accessed via the portal and CSMD metadata format.

2. She accesses the data via the linked data input – this requires that her registration ID is also sent with the request as the site allows free access to data, but insists on registration first.

3. Further, in order to re-analyse the data, the researcher accesses a software repository to determine the software version used.

4. She then downloads and installs the software and generates results. These confirm the results, but the researcher has some additional observations.

5. She then posts an annotation to the journal site, and on her research blog, with links to her new analysed data.

**Goals:**

- Discovery of linked data sources, and keyword/free text queries.
- Formulation of queries based on the results of previous queries.
- Accessing data sources which may require registration.

**Topic:** Academic research.

**Primary Actors and Roles:** Funders and aggregators (provide CRIS), science data managers (science data archives), journal providers (papers), librarians (papers, grey literature, publication repositories), Science Researchers (under take searches, provide annotations, secondary data).

**Secondary Actors and Roles:** Training Material Team.
5.2.6.2. Data, Metadata, Owners and Languages

**Data sets:** publication archives, raw data archives (could be large – TB), CRIS, software archives.

**Data Services:** this is an idealised scenario, and a number of data access and ingest services could be involved, including Linked Data.

**Data owners:** funding bodies, facilities, publishers, universities

**Data Access Rights:** open data and funding information – could require registration before free access. Open publication archives, though Journal publishers could add conditions.

**Data Formats:** XML, RDF, PDF. Science data could be in numeric or binary formats (e.g. HDF, CDF), but we would not necessarily expect it to be directly accessible - download viewers or converters.

**Metadata formats:** CERIF, bibliographic data (e.g. DC, MARC), science data (domain specific – e.g. CSMD.

**Nationalities and Languages involved:** Usually English. Data might be in numeric or binary format.

5.2.6.3. Special Requirements

**Pre-condition list:** linked data or other public access to sources, registered with the platform.

**Post-conditions list:** any new data/annotations stored within appropriate data provider.

**Primary Flow:** As defined in the description.

**Error Flow:** unregistered access to data source.

**Additional Requirements list:** Videos, presentations and all needed training material should be expressed in English as primary choice. The starting point would be to cover all languages of the involved countries.

**Notes & Outstanding Issues:** This is a rather complex scenarios that includes several use cases.
5.2.6.4. Use Case Diagrams

Figure 5-11: Use Case Diagrams - Linking my system for downloading data
5.2.6.5. Sequence Diagrams

![Sequence Diagrams](image)

**Figure 5-12:** Sequence Diagrams - Linking my system for downloading data
5.2.7. Storing data in draft form (in the buffer)

5.2.7.1. General Information and Actors

**Description:** Nikos lives on a big island in Ionian Sea and works for the Travel and Tourism industry. Nikos has noted down all the Scuba Diving Centers in Ionian Sea along with valuable related information (price, diving areas, equipment, etc) and wants to share this information with the rest of the world.

Nikos has prepared an excel sheet with the related data. He visits the The platform portal and uploads the excel sheet through the related The platform web page. The data will be published after they are approved and curated by the moderator.

**Goal:** The goal of this use case is to demonstrate citizens storing data in a draft form, to be further analysed and curated by data experts.

**Topic:** The main sub areas of this use case are Travel and Tourism.

**Primary Actors and Roles:** Nikos, a citizen that stores data to the platform.

**Secondary Actors and Roles:** The moderator that approves the data uploaded as well as owners of Scuba Diving Centers. Moreover visitors and tourists from all around the world that like scuba diving will enter the platform to find the location of diving centers.

5.2.7.2. Data, Metadata, Owners and Languages

**Data sets:** The data set involved in this use case is relative small (<100 rows) and involves location and Information of Scuba Diving Centers.

**Data Services:** No data services are available to provide this information.

**Data owners:** Nikos has the ownership of data as a user and the moderators have the control of the data.

**Data Access Rights:** The Data Access Rights are those specified by Nikos, which is the Data Access Rights assigned to the uploaded data.

**Data Formats:** The Data Format used is Microsoft Excel Format (xls).

**Metadata formats:** Custom, non-standardized metadata format.

**Nationalities and Languages involved:** The involved nationality and language is Greek.

5.2.7.3. Special Requirements

**Pre-condition list:** None

**Post-conditions list:** None
**Primary Flow:**
1. Nikos visits the portal.
2. Nikos uploads the excel file with the dataset through the corresponding web page.
3. The moderator views the pool of upload requests, approves the dataset and annotates as necessary.
4. The dataset is available on the platform.

**Error Flow:** The data uploaded is spam / contains illegal information / not interesting. The moderator disapproves the dataset and notifies the owner through email.

**Additional Requirements list:** None.

**Notes & Outstanding Issues:** None.

5.2.7.4. Use Case Diagram

![Use Case Diagram](image_url)

Figure 5-13: Use Case Diagram - Storing data in draft from (in the buffer)
5.2.7.5. Sequence Diagrams

![Sequence Diagram]

Figure 5-14: Sequence Diagram - Storing data in draft from (in the buffer)

5.2.8. Put data and annotate them according to the platform standard and my needs

5.2.8.1. General Information and Actors

**Description:** Annotation and linking of research. This scenario could be considered as continuation of scenario SC-06.

A researcher:
1. Carries out a search across a number of facilities primary data repositories (with registration) via the portal, each repository providing a link to suitable software for processing.

2. She downloads and uses the software, aggregating and comparing the results, writing a paper as a result.

3. She then deposits the paper in an open access repository prior to publication, and records the connections between the results in his electronic notebook.

**Goals:**

- Uploading of data and links to data.

**Topic:** Academic research.

**Primary Actors and Roles:** Funders and aggregators (provide CRIS), science data managers (science data archives), journal providers (papers), librarians (papers, grey literature, publication repositories), Science Researchers (under take searches, provide annotations, secondary data).

**Secondary Actors and Roles:** Training Material Team.

### 5.2.8.2. Data, Metadata, Owners and Languages

**Data sets:** publication archives, raw data archives (could be large – TB), CRIS, software archives.

**Data Services:** this is an idealised scenario, and a number of data access and ingest services could be involved, including Linked Data.

**Data owners:** funding bodies, facilities, publishers, universities

**Data Access Rights:** open data and funding information – could require registration before free access. Open publication archives, though Journal publishers could add conditions.

**Data Formats:** XML, RDF, PDF. Science data could be in numeric or binary formats (e.g. HDF, CDF), but we would not necessarily expect it to be directly accessible - download viewers or converters.

**Metadata formats:** CERIF, bibliographic data (e.g. DC, MARC), science data (domain specific – e.g. CSMD.

**Nationalities and Languages involved:** Usually English. Data might be in numeric or binary format.
5.2.8.3. Special Requirements

**Pre-condition list:** linked data or other public access to sources, registered with the platform.

**Post-conditions list:** any new data/annotations stored within appropriate data provider.

**Primary Flow:** As defined in the description.

**Error Flow:** unregistered access to data source.

**Additional Requirements list:** Videos, presentations and all needed training material should be expressed in English as primary choice. The starting point would be to cover all languages of the involved countries.

**Notes & Outstanding Issues:** This is a rather complex scenarios that includes several use cases.

5.2.8.4. Use Case Diagram

![UC-Annotation and Linking](image)

**Figure 5-15:** Use Case Diagram - Put data and annotate them according to the platform standard and my needs
5.2.8.5. Sequence Diagram

![Sequence Diagram](image)

Figure 5-16: Sequence Diagram - Put data and annotate them according to the platform standard and my needs

5.2.9. Put my needs for Public Sector data

5.2.9.1. General Information and Actors

**Description:** A European citizen is interested in making a new investment in a specific Greek region. He needs to get more details for economical and demographical data for the specific region. So as it is not possible for him to find all the datasets, he selects the platform page for new proposals. He fills the appropriate form and presses the apply button. The applied form goes directly to the possible data owners (public sector) which are selected from the citizen (The platform has a list with the registered data owners and the contact emails for every country).

**Goal:** Transferring the needs of researchers and citizens towards the data-suppliers

**Topic:** The use case is independent of particular domains.
**Primary Actors and Roles:** Citizen, Public Sector

**Secondary Actors and Roles:** N/A

### 5.2.9.2. Data, Metadata, Owners and Languages

In this scenario there is no information for these fields, as they do not refer to any dataset or service provided by datasets.

**Data sets:** None.

**Data Services:** None.

**Data owners:** None.

**Data Access Rights:** None.

**Data Formats:** None.

**Metadata formats:** None.

**Nationalities and Languages involved:** All nationalities and English language

### 5.2.9.3. Special Requirements

**Trigger:** Specific needs not covered by the existing datasets

**Pre-condition list:**

1. For each registered data owner The platform keeps contact information

**Post-conditions list:** None.

**Primary Flow:**

1. Connect to the Portal
2. Select Proposal’s Section
3. Fill form
   a. Put contact info
   b. Describe dataset
   c. Explain reason
   d. Select country
   e. Select possibly public sectors from a list
4. Preview
5. Press Apply button
6. The platform forward form to Public sector
7. Show succeed message

Error Flow: Not needed.

Additional Requirements list: In form you could have a choice to write in the same language with the possibly concerned public sector.

Notes & Outstanding Issues: The platform has not any responsibility for what the citizens write into the form.

5.2.9.4. Use Case Diagram

Figure 5-17: Use Case – Put my needs for Public Sector
5.2.9.5. Sequence Diagram

![Sequence Diagram](image)

Figure 5-18: Sequence Diagram - Put my needs for Public Sector

5.2.10. Get information and training on metadata, open data, platform

5.2.10.1. General Information and Actors

**Description:** A public sector servant wants to use the visualize tool. He doesn’t know how to do it and hasn’t any experience. He selects Training Section from Menu and he is looking for useful material. From topic categories select tools and more specific the visualization tools. The platform choices are to select Documentation, Presentation with voice, Videos, Samples and online lessons. He selects a presentation with voice and after checking documentation he downloads some files.

He continues checking the online lessons which need registration. The online lesson is like an e-class platform which you can follow the lessons as you want. (The lessons use the same material but organised in a more educational way).

He finally wants to give feedback to the platform about the training material and he fills the appropriate form.

**Goal:** Training all users.

**Topic:** The use case is independent of particular domains.

**Primary Actors and Roles:** All THE PLATFORM Users and Stakeholders.

**Secondary Actors and Roles:** THE PLATFORM Training Material Team.
5.2.10.2. Data, Metadata, Owners and Languages

In this scenario there is no information for these fields, as they do not refer to any dataset or service provided by datasets.

**Data sets**: None.

**Data Services**: None.

**Data owners**: None.

**Data Access Rights**: None.

**Data Formats**: None.

**Metadata formats**: None.

**Nationalities and Languages involved**: None.

5.2.10.3. Special Requirements

**Trigger**: Understanding and learning how to use the platform services.

**Pre-condition list**:

2. For each aforementioned scenario should be different and categorised material that shows the right way to achieve users needs.

3. To follow the online lesson a registration is needed.

4. List for feedback.

**Post-conditions list**: None.

**Primary Flow**:

8. Connect to The platform Portal

9. Select Training Section

10. Find Tools Category and Select Visualization Tool

11. Open Documentation (Read/Download)

12. Open Presentation with Voice

13. Open Videos

14. Select online Course
   a. Follow lessons
   b. Open Samples
   c. Fill evaluation

15. Send Feedback

**Error Flow**: Not needed.
**Additional Requirements list:** Videos, presentations and all needed training material should be expressed in English as primary choice. The starting point would be to cover all languages of the involved countries.

**Notes & Outstanding Issues:** There should be different training paths for different stakeholders and potential users.

5.2.10.4. **Use Case Diagram**

![Use Case Diagram](image)

*Figure 5-19: Use Case - Training all potential users*
5.2.10.5. Sequence Diagram

![Sequence Diagram](image)

Figure 5-20: Sequence Diagram - Training all potential users

5.3. Validation and Rating of the Results

More than 30 high level eGovernance, ICT and Policy Experts participated in the Open Data Workshop in Delft, the Netherlands, on August 29, 2011, in conjunction with the IFIP eGOV International Conference. The workshop focused on discussing the scientific base of ICT-enabled governance for open data, thus harnessing open data sources and methodologies for annotating, visualising and making open data available to scientists and citizens. Participants presented innovative approaches and provided their view on open data, including value-adding examples and best practices. The main topics of the workshop were structured around the state of the art, the visionary
scenarios, the research gaps and the future research challenges in the area of ICT for governance and public sector's open data.

The second half of the workshop was dedicated to gathering ideas from the audience on open data sources and curation. Two questionnaires were distributed and answered by participants.

- The first questionnaire targeted the rating of 9 (then) scenarios. From the results of the rating exercise, one more scenario was added to the scenarios list (the training and insight scenario) and the 10 final scenarios were prioritized and extended.

- A second questionnaire discussed the important features of a service infrastructure for open data, indicating that the three most important issues are:
  a. to ensure quality and authenticity of data
  b. to reach standardised metadata
  c. to convince Public Sector Officials to share knowledge

The results were then used in modifying, extending or focusing scenarios functionality.

Rating Open Data Usage Scenarios

Participants were given 9 scenarios to rank from most to less important. The resulting ranking of the services is as follows:

Table 5-3: Rating Open Data Usage Scenarios

<table>
<thead>
<tr>
<th>Usage Scenario</th>
<th>Stakeholder</th>
<th>Ranking (1 = most important)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting useful information (through browsing datasets or visualisations)</td>
<td>Citizen, Public Servant</td>
<td>1</td>
</tr>
<tr>
<td>Storing, Linking, Annotating and Visualising a PS data set</td>
<td>Public Servant</td>
<td>2</td>
</tr>
<tr>
<td>Getting data for my research work</td>
<td>Researcher</td>
<td>3</td>
</tr>
<tr>
<td>Getting metadata specifications (for applying them in my systems)</td>
<td>Public Servant</td>
<td>4</td>
</tr>
<tr>
<td>Linking my system for downloading data</td>
<td>Researcher</td>
<td>5</td>
</tr>
</tbody>
</table>
From the first discussion on key services, the audience gave the first five positions to the core services for posting, downloading, visualising and annotating metadata. They also stressed the need for research-oriented services, especially on metadata downloading. Also, the audience proposed another 2 services, some of them being integrated in existing services. The service that became a new (10th) scenario concerns user support:

- Usage guidelines (how to) on data & metadata levels, for all users

Two other ideas were handled as features in existing scenarios:

- Data Quality Rating
- Usage guidelines (how to) on data & metadata levels

5.4. Answering Research Question 3

5.4.1. Social Media Use in Governance

Web 2.0 constitutes a different Internet paradigm in comparison with its predecessor Web 1.0. Web 2.0 promotes the generation of content of various types by simple and non-expert users, the development of relationships and online communities
among them, and the extensive interaction, collaboration and sharing of content and information (O'Reilly, 2007). A big number of social media platforms have been developed adhering to these characteristics. The main capabilities of Web 2.0 social media are (Davis & Mintz, 2009):

i) User-generated social content: social media enable users to submit content which other users can access, rate and comment.

ii) Social networking: users of social media join together in online, which allow them to see profile information about the people to whom they are connected, and to share information and to have extensive interaction with them.

iii) Collaboration: users the platform in conversations, co-creation of content, collaborative problem solving, and collective action.

The above capabilities were initially exploited by private sector firms and later started being adopted and utilized by government agencies. Social media can offer government agencies significant opportunities for: i) increasing citizens' participation and the dialogue in public policy making, by providing to more groups a voice in discussions of policy development, implementation and evaluation; ii) promoting transparency and accountability, and in this way reducing corruption, by enabling governments to open up large quantities of activity and spending related data, and at the same time enabling citizens to collectively take part in monitoring the activities of their governments; iii) public services co-production, by enabling government agencies and the public to develop and design jointly government services; iv) crowdsourcing solutions and innovations, by exploiting public knowledge and talent in order to develop innovative solutions to the increasingly complex societal problems (Bertot et al., 2012; Bonsón et al., 2012; Chun & Luna Reyes, 2012; Margo, 2012; Criado et al. 2013; Bovaird, 2007; Torres, 2007; Lukensmeyer & Torres, 2008; Bertot et al., 2012; Linders, 2012).

5.4.2. Closing the feedback loop

In order be able to benefit from open government data and generate value from them, several researchers have argued that opening and using these data should be seen as an on-going process performed by an ecosystem of multiple collaborating entities (Pollock, 2011; Zuiderwijk, Janssen, Choenni, Meijer, & Sheikh_Alibaks, 2012). Pelet (2013) states that currently open data is an early experiment of a promising idea, and
that it is important to understand that an effective open data program requires time and patience to grow. The open data ecosystem is "a multi-level and multi-dimensional entity where raw material, as far as distribution and developing are concerned, is the target of cooperation" (Poikola, Kola, & Hintikka, 2011, p. 13). Open data ecosystems are characterized by the interaction of data producers, infomediaries as intermediate consumers of data or service providers and open data users (Ding, Peristeras, & Hausenblas, 2012; Ubaldi, 2013). They consist of multiple interdependent socio-technical levels, and elements.

One essential element of open data ecosystems concerns their development “through user adaptation, feedback loops and dynamic supplier and user interactions and other interacting factors” (Zuiderwijk et al., 2014). Open data ecosystems perform data production and usage cycles with feedback loops, sharing of data back to publishers and also with the so-called infomediaries (Pollock, 2011). However, discussion and feedback loops appear barely to be part of existing open data practices and infrastructures. Zuiderwijk and Janssen (2013) found that after open data have been used, the provision of feedback to data providers or a discussion with them is quite important by not facilitated by existing open data infrastructures, though such mechanisms might be useful for improving open data quality, data release processes and policies. Dawes and Helbig (2010) found that such mechanisms can help users to obtain insight in how they can use and interpret open government data and generate value from them.

At the same time another major trend in government agencies has been the exploitation of Web 2.0 social media for increasing citizens’ participation in the government decision and policy making processes, supporting networking, interaction and collaboration, and also collecting opinions, knowledge and ideas from citizens, and promoting government transparency and accountability (Bertot, Jaeger, & Grimes 2012; Bonsón, Torres, Royo, & Flores, 2012; Chun, & Luna Reyes 2012; Margo, 2012; Criado, Sandoval-Almazan, & Gil-Garcia, 2013). Therefore it would be interesting to investigate the use of Web 2.0 social media oriented capabilities in open government data platforms for the collection of feedback from their users, and in general for enabling and promoting discussion both between providers and users, and among users, in order to facilitate value generation from them and accelerate innovation.
5.4.3. Users and Final Requirements

This section provides the functionality of this platform to the two main stakeholders, the open data users and providers, is described, initially the ‘classical’ (basic functionality supported by most first generation OGD platforms) and then the novel Web 2.0 oriented one. The former includes mainly data publication-upload and modelling (metadata) for the data providers, and data search, visualization and download for the data users (see Table 5.4 for more details).

The novel Web 2.0 capabilities aim to support open data ‘prosumers’ (who are at the same time users of provided datasets, and also producers of new versions of them (through various types of processing), which are improved, enriched-extended or adapted for specific purposes, or even of new datasets), and also extensive interaction and collaboration among them. This novel functionality includes users groups formation and extensive communication and collaboration within them, data processing, enhanced data modeling (flat, contextual and detailed metadata), commenting existing datasets and expressing needs for new datasets, datasets quality rating, data linking, publication/upload of new versions of existing datasets, advanced data visualization (see Table 4-6 for more details).

Table 5-4: Classical Functionalities

<table>
<thead>
<tr>
<th>Functionality</th>
<th>Stakeholder</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Data Publication/upload</td>
<td>Provider</td>
<td>Support for publication/upload of datasets to the providers</td>
</tr>
<tr>
<td>2 Data Modeling</td>
<td>Provider</td>
<td>Capabilities of flat metadata descriptions (based on a specific metadata models)</td>
</tr>
<tr>
<td>3 Data Search</td>
<td>User</td>
<td>Simple search via keywords, resource format, publisher, topic categories and countries</td>
</tr>
<tr>
<td>4 Data Visualisation</td>
<td>User</td>
<td>Simple visualisation techniques on specific datasets (maps, charts)</td>
</tr>
<tr>
<td>5 Data Download</td>
<td>User</td>
<td>Data and metadata downloading capabilities – also provision of API for this purpose</td>
</tr>
<tr>
<td>Functionality</td>
<td>Stakeholder</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Grouping and Interaction</td>
<td>Provider/User</td>
<td>Capabilities for (a) searching for and finding other users/providers having similar interests with me in order to have information and knowledge exchange and cooperation, (b) forming groups with other users/providers having similar interests with me in order to have information and knowledge exchange and cooperation, (c) maintaining datasets/working on datasets within one group, (d) communicating with other users/providers through messages in order to exchange information and knowledge and (e) getting immediately updated about the upload of new versions and enrichments of datasets maintained/ worked on within the group, or new relevant items (e.g. publications, visualizations, etc.).</td>
</tr>
<tr>
<td>Data Processing</td>
<td>Provider/User</td>
<td>Capabilities for (a) data enrichment - i.e. adding new elements - fields, (b) for metadata enrichment - i.e. fill in missing fields, (c) for data cleansing - e.g. detecting and correcting ubiquities in a dataset, matching text names to database IDs (keys) etc., (d) converting datasets to another format, (e) submitting various types of items - e.g. visualisations, publications - related to a</td>
</tr>
<tr>
<td>3</td>
<td>Data Enhanced Modeling</td>
<td>Provider/ User</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------</td>
<td>----------------</td>
</tr>
</tbody>
</table>
| 4  | Feedback and Collaboration | Provider/ User | Capabilities (a) to communicate our own thoughts and ideas on the datasets to the other users and the providers of them through comments I enter on them,  
(b) to read interesting thoughts and ideas of other users on the datasets through comments they enter on them,  
(c) to express our own needs for additional datasets that would be interesting and useful to me,  
(d) to get informed about the needs of other users for additional datasets and  
(e) to get informed about datasets extensions and revisions. |
| 5  | Data Quality Rating | User | Capabilities to (a) communicate to the other users and the providers the level of quality of the datasets that I perceive,  
(b) get informed on the level of quality of the datasets perceived by other users through their ratings |
| 6  | Data Linking | Provider/ User | Capabilities of data and metadata linking to other ontologies in the Linked Open Data Cloud. Capabilities of querying data and metadata through Sparql Endpoints. |
| 7  | Data Versions | Provider/ | Support for publication/upload of new dataset and (f) datasets combination and Mash-ups. |

- **Data Enhanced Modeling**: Provider/User Capabilities for description of flat, contextual and detailed metadata of any metadata/vocabulary model.
- **Feedback and Collaboration**: Provider/User Capabilities (a) to communicate our own thoughts and ideas on the datasets to the other users and the providers of them through comments I enter on them,  
(b) to read interesting thoughts and ideas of other users on the datasets through comments they enter on them,  
(c) to express our own needs for additional datasets that would be interesting and useful to me,  
(d) to get informed about the needs of other users for additional datasets and  
(e) to get informed about datasets extensions and revisions.  
- **Data Quality Rating**: User Capabilities to (a) communicate to the other users and the providers the level of quality of the datasets that I perceive,  
(b) get informed on the level of quality of the datasets perceived by other users through their ratings.  
- **Data Linking**: Provider/User Capabilities of data and metadata linking to other ontologies in the Linked Open Data Cloud. Capabilities of querying data and metadata through Sparql Endpoints.  
- **Data Versions**: Provider/Support for publication/upload of new dataset and (f) datasets combination and Mash-ups.
<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Publication/upload</td>
<td>User</td>
<td>versions of the existing datasets, and connection with previous ones and initial datasets.</td>
</tr>
<tr>
<td>8</td>
<td>Advanced Data Visualisation</td>
<td>User</td>
</tr>
<tr>
<td>9</td>
<td>Advanced Data Search</td>
<td>User</td>
</tr>
</tbody>
</table>
6. Evaluation of an OGD e-infrastructure

This chapter is dedicated to answer the fourth research question of this dissertation: “How we can evaluate an OGD infrastructure?”. Answering the question we have to firstly identify the evaluation methods and techniques in order to construct an evaluation framework which will be consists of both qualitative and quantitative tools. The next phase is to construct the questions for the evaluation models and validate them. The final step is to evaluate the proposed infrastructure.

Furthermore, an algorithm for decision support and the identification of next development decisions is also developed and tested towards driving the identification of more advanced capabilities of an OGD infrastructure.


In this section, theories and practice in current ICT project evaluation will be detailed. This includes the introduction of both quantitative and qualitative concepts and methods for project evaluation. In addition, related works on evaluation criteria, as well as methodical concepts for performing project evaluations will be examined. A big challenge is how to evaluate and assess the value of projects on OGD e-infrastructures domain. There are numerous evaluation models and many more extensions of them. Many researchers have developed methodologies based on well-knowing models trying to shape them according to their specific objectives. To that extend it is absolutely necessary to review the existed methodologies in the fields concerning OGD e-infrastructures in order to conclude the model for the evaluation framework.

6.1.1. Theories for evaluation

In general, evaluation research can be seen as an application variant of empirical research methods for a special issue. Evaluation research address a variety of different evaluation objects, e.g. persons, environmental factors, products, techniques and methods, terms of reference, projects and programmes, systems and structures, as well as the evaluation of social intervention programmes (Bortz & Döring, 2002). Based on that, the following five fundamental but different objectives, purposes or functions of evaluation can be derived:
1. **Insights**: Collecting insights about attitudes and impacts of inventions
2. **Optimisation**: Identifying strength and weaknesses of interventions
3. **Controlling**: Evaluating the effectiveness and efficiency of the interventions
4. **Decision-making**: Shall a certain intervention be promoted, be implemented, be developed or be used?
5. **Legitimation**: supporting the legitimation of developments and implementations of an intervention to external stakeholder, in particular if it is about the deployment of public funds.

According to Smith (2001; 2006), evaluation in general is a process which includes gathering, ordering and assessing of information about an object of study in a methodical way. Evaluation is often applied in relation to monitoring. Both terms are sometimes used synonymously and even mixed up with assessment. To establish a clear-cut understanding among these terms: monitoring focuses on watching, observing or keeping an issue under review. As part of monitoring, performance indicators may be engaged to assess specific issues under consideration.

Evaluation is difficult to be expressed at once; reasons are that there is no standardised way of making good assessments, in particular developing appropriate measurable criteria and still keeping comparable means dealing with high complexity ((eParticipation) MOMENTUM, 2007). A further challenge of developing an evaluation methodology - if it is to have any meaning - is to involve targeted user groups, evaluation procedures and the expected results. The results have to stress both negotiation and consensus concerning the evaluation procedures, and the conclusions received.

In the settings of this deliverable, the definition of the Organisation for Economic Co-operation and Development (OECD) applies best to the objectives of the ENGAGE evaluation framework. OECD defines evaluations as ‘analytical assessments addressing results of public policies, organisations or programmes, that emphasise reliability and usefulness of findings. Their role is to improve information and reduce uncertainty; however, even evaluations based on rigorous methods rely significantly on judgement. A distinction can be made between ex-ante evaluations (or policy reviews) and ex-post evaluations’ (OECD, 1998).

Rogers and Smith (Rogers, 2003; Smith, 2001; 2006) understand evaluation as either providing evidence that something is working or needed, or improving and
strengthening practice, programmes or projects. The authors distinguish among program and project evaluation, and practice evaluation:

- **Program and project evaluation**: Program evaluation theory typically concerns judgements about the effectiveness, efficiency and sustainability of programs. The essential nature of this kind of evaluation is to serve as a management tool and therewith as part of decision support processes. This form of evaluation contributes to decision-making about whether the program or project should be refunded, cut or abandoned. Representatives of this approach are methods like the discrepancy evaluation model (Steinmetz, 2002; Cronbach et al., 1980; Weiss, 1972).

- **Practice evaluation**: Practice evaluation focuses on the improvement of work carried out by individuals and groups, and on the development of participants. This kind of evaluation is about situations and how people can learn to best possible act in a certain situation in future or further action. Hence, it is also strongly related to self-monitoring and -critic. People are encouraged to make judgements about the situation and evaluate their part in it. Here, learning is part of the process; for this reason this evaluation is also described as educative or pedagogical.

Another differentiation is to be addressed depending on the stage of development of a project. As far as concerns the specific stage of development, different objectives, methods, aspects and processes may be applied. Since the beginning of 1970s literature classifies types of evaluation by the specific aim and therefore distinguishes between formative and summative evaluation (Neumann, 2002):

- **Formative evaluation (the management view)**: This form of evaluation serves as improver and optimiser of programs. Thereby formative evaluation is typically informally arranged and applied. Consequently it uses rather qualitative than quantitative methods. Evaluation of this kind seeks to understand and evaluate the acceptance of the program, if impact and objectives match each other, and if efforts are justified. This form of evaluation is often related to the managerial definition of success.

- **Summative evaluation (the consumer view)**: This form of evaluation concerns the impact of different programs or projects often processed in comparison with each other. Objective of summative evaluation is to judge the occurrence or absence of expected internal program or project specific impacts, as well as external and mandatory impacts committed. Summative evaluation also assesses intensity of impact and duration until a program develops its effects. Therefore
qualitative methods are often applied. A typical example for summative evaluation is a user or consumer-centric evaluation. Users are only secondary interested ‘in improving the program or project (i.e. formative evaluation) as their primary interest is whether their own needs are met’ (Scriven, 2001).

Table 2-1 provides an overview of the crucial aspects, on which Wottawa (1998) distinguishes formative from summative evaluation.

**Table 6-1: Comparison of the crucial differences of formative and summative evaluation according to Wottawa (1998)**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Formative evaluation</th>
<th>Summative evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant target groups</td>
<td>• Program/project developer&lt;br&gt;• Program/project manager&lt;br&gt;• Program/project executives</td>
<td>• Politicians&lt;br&gt;• Interested parties/general public&lt;br&gt;• Investors</td>
</tr>
<tr>
<td>Dominant data collection</td>
<td>• Clarification of goals&lt;br&gt;• Kind of program/project procedures and implementation&lt;br&gt;• Clarification of implementation challenges and challenges when approaching results&lt;br&gt;• Analysis of implementation and results at the micro-level</td>
<td>• Documenting the results&lt;br&gt;• Documenting the implementation&lt;br&gt;• Analysis of implementation and results at the macro-level</td>
</tr>
<tr>
<td>Dominant role of program/ project developer and program/ project executives</td>
<td>• Contributer</td>
<td>• Data collector</td>
</tr>
<tr>
<td>Dominant role of the evaluator</td>
<td>• Interactive</td>
<td>• Independend</td>
</tr>
<tr>
<td>Typical</td>
<td>• Qualitative and</td>
<td>• Quantitative, sometimes</td>
</tr>
<tr>
<td>Methodology</td>
<td>Quantitative, but emphasis is on the former one</td>
<td>Enriched through qualitative methods</td>
</tr>
<tr>
<td>-----------------------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Frequency of data collection</td>
<td>• Continuous monitoring</td>
<td>• Limited in time</td>
</tr>
<tr>
<td>Dominant mechanisms of reporting</td>
<td>• Often throughout the term</td>
<td>• At the end</td>
</tr>
</tbody>
</table>
| Focus of reporting                | • Relation between process elements (micro-level)  
|                                   | • Relation between context and process        | • Implications for policy-making, administrative control and management |
|                                   | • Relation between process and result         |                                       |
|                                   | • Implications for program/project practices  |                                       |
|                                   | • Specific changes of operations              |                                       |
| Requirements for credibility      | • Commitment with developers and executives concerning reporting  
|                                   | • Backing and trust                           | • Scientific grounding               |
|                                   |                                               | • Objectivity                          |

Yet another well-known distinction is between qualitative and quantitative evaluation methods:

- **Quantitative research methods of evaluation** originally base on standardised questions, which are transferable to empirical data processed in statistic analysis. This results on one hand in the general assumption that quantitative research methods are more objective than qualitative ones. On the other hand, quantitative methods miss contextual details important for a deeper understanding of the issues studied. Examples of quantitative methods are survey methods, laboratory experiments, formal methods (e.g. econometrics) and numerical methods such as mathematical modelling.
Qualitative research methods of evaluation base on unstructured and non-numeric data, which enable researchers to investigate social and cultural phenomena (Myers, 1997). These data are usually gathered through observation and participant observation e.g. fieldwork, interviews and questionnaires, documents and texts, and the researcher’s impressions and reactions. Examples of qualitative methods are action research, case study research and ethnography. To underline the value and relevance of qualitative evaluation methods, Kaplan and Maxwell (1994) complained in their publication about quantitative methods that understanding an issue from participants perspective is for the most part lost when textual data are quantified.

A quite important distinction of types of assessments relevant for ENGAGE is the one among ‘project assessment’ and ‘tool assessment’. Project assessment covers the formative evaluation from the contracting entity point of view, whereas ‘tool assessment’ tackles the summative evaluation from the citizens and users perspective. The former type refers to criteria such as project duration, project costs, achievement of project results (Haering, von Arb, 2004), while the latter concerns criteria such as user-friendliness, transparency, acceptance, (Venkatesh, Davis, 2000; Venkatesh et al., 2003) which are frequently studied in technology acceptance and technology diffusion analyses.

To tackle the specific evaluation needs within ENGAGE, above mentioned theories set the ground for the evaluation framework. In the following section, methods for information gathering and analysis that can be employed in the evaluation of OGD e-Infrastructure project such ENGAGE will be introduced.

6.1.2. Methodical concepts for performing evaluations

Methodical concepts for performing project evaluations are manifold. This section introduces a set of common methods for information gathering and analysis that could be employed in the evaluation of OGD e-infrastructure projects and in particular of ENGAGE Evaluation Framework:

- **Quantitative surveys** own the typical attributes of a quantitative research method (Bortz & Döring, 2002). They gather highly standardised data on the base of closed-ended questions. These include standardised questionnaires and personal interviews. Respondents only can chose their answers within prescriptive answers fixed by the interviewer beforehand. For instance, the
question in how far citizens are satisfied with a certain e-participation tool can only be answered by the respondent within a fixed rating scala from ‘not at all satisfied’ to ‘absolutely satisfied’. Hence, quantitative surveys are simple feedback forms and therewith will be part of the evaluation methodology.

- **Web analytics** is the process of quantitative analysis of the behaviour of visitors to a web site by tracking different custom metrics. It aims at enabling e.g. a public agency to attract more citizens. In business settings web analytics are often used as part of customer relationship management. Logfile analysis (recording of all transactions made by a person) and page tagging (recording of any foreign server a person comes from and goes to) are two most popular technological approaches of web analytics. Such results are often depicted in the form of tables, charts, and graphs. In the context of this deliverable, web analytics is used as an appropriate tool for evaluating the impact of e-participation projects success in awareness raising towards their target groups because the official website of a project is one of the most important dissemination tools. Hence, from e.g. the number of visitors to a website the dissemination impact can be derived. Web analytics may be also used for discovering on the one hand if the target group is reached, and on the other hand if the target group is sustainably interested.

- **Usability testing** evaluates a product by testing it on users. Therefore it gathers data about participants’ success, speed of performance, and satisfaction. In practice it mainly refers to prototyping in an iterative process whereby the stage in the tool development cycle determines the technique applied. The findings include both quantitative data and qualitative observations. In OGD e-infrastructure settings, usability testing is usually not applied before the tool has been implemented. Hence, it is more a field test rather than a laboratory experiment. Data gathered by this approach provides detailed information about the usability of each tool and technology applied.

- **Focused (semi structured) questionnaire/interviews** is a general method of asking questions by the means of a prepared but dynamically applicable questionnaire which is convenient to any topic of interest (Bortz & Döring, 2002). This interview technique uses open-ended questions, of which some are pre-identified by the researcher and some occur naturally during the interview. The objective is that the respondent expresses her/his interpretations, opinions, evaluations, expectations, feelings, and values.

- **In-depth questionnaire/interviews** is a collective term for open-ended, discovery-oriented methods using open and semi-structured formats (Bortz &
Döring, 2002). The tool is used by researchers to seek understanding and interpretation whereby only the general topic of conversation is introduced by the researcher. Hence, the interviewer is constrained to avoid subjective hypotheses, understandings or specific questions. In-depth interviews emphasise recording responses, observations and reflections in order to identify unconscious motifs and processes.

- **Collective semi-structured questionnaire/interviews** is a version of the focused (semi-structured) interview method that allows researchers to generate answers to open questions (as in semi-structured interviews) from several respondents simultaneously. The main difference between individual and collective semi-structured interviews is the influence of group dynamics, which is possible when applying collective semi-structured interviews. On the one hand it may promote strategic and conformist responses from interviewees. On the other hand, it can encourage respondents to reflect upon their relations with one another (Bortz & Döring, 2002).

### 6.1.3. Review on Evaluation Models

#### 6.1.3.1. Technology Acceptance Models

TAM has been influenced by Theory of Reasoned Action (TRA) (Fishbein & Ajzen, 1975) and Theory of Planned Behavior (TPB) (Ajzen, 1991) and “posits that perceived usefulness and perceived ease of use determine an individual’s intention to use a system with intention to use serving as a mediator of actual system use”24. Perceived usefulness is also seen as being directly impacted by perceived ease of use. Researchers have simplified TAM by removing the attitude construct found in TRA from the current specification Venkatesh and Davis (2000) and Venkatesh et. al. (2003). Attempts to extend TAM have generally taken one of three approaches: by introducing factors from related models, by introducing additional or alternative belief factors, and by examining antecedents and moderators of perceived usefulness and perceived ease of use (Wixom & Todd, 2005). TRA and TAM, both of which have strong behavioural elements, assume that when someone forms an intention to act, that they will be free to act without limitation. In practice constraints such as limited ability, time, environmental or organisational

---

limits, and unconscious habits will limit the freedom to act is an information systems theory that models how users accept and use a technology. The model suggests that when users are presented with a new technology, a number of factors influence their decision about using it, but the two main factors are (Davis et al., 1989):

- **Perceived usefulness (PU)** - This was defined by F. Davis as "the degree to which a person believes that using a particular system would enhance his or her job performance".
- **Perceived ease-of-use (PEOU)** – F. Davis defined this as "the degree to which a person believes that using a particular system would be free from effort".

Each of these two factors can be developed into a detailed set of variables for each particular type of IS. Based on this framework extensive research has been conducted for understanding better and predicting user acceptance of various types of IS (Schepers & Wetzels, 2007; Hsiao & Yang, 2010). As it is referred, TAM is continued to expand, the two major upgrade being the TAM2 (Venkatesh & Davis, 2000) and the Unified Theory of Acceptance and Use of Technology (or UTAUT25, Venkatesh et al., 2003).

---

TAM2 explains perceived usefulness and usage intentions in terms of social influence and cognitive instrumental processes. Both social influence processes (subjective norm, voluntariness, and image) and cognitive instrumental processes (job relevance, output quality, result demonstrability, and perceived ease of use) significantly influenced user acceptance.

Venkatesh et al. (2003), Venkatesh and Zhang (2010) show that the theory of acceptance and use of technology (UTAUT) is useful to enrich one's understanding of research on technology adoption. The theory was developed through a review and consolidation of the constructs of eight models that earlier research had employed to explain information systems usage behaviour. The theory uses constructs of: theory of reasoned action, technology acceptance model, motivational model, theory of planned behavior, a combined theory of planned behavior/technology acceptance model, model of PC utilization, innovation diffusion theory, and social cognitive theory. UTAUT provides the rationale for the survey questions.

According to Venkatesh UTAUT identifies:

- 3 direct determinants of behavioral intention to use a technology
  - Performance expectancy (PE): the degree to which an individual believes that using the system will help him or her to attain gains in job performance
- Effort expectancy (EE): the degree of ease associated with the use of the system
- Social influence (SI): the degree to which an individual perceives that important others believe he or she should use the new system

- 2 direct determinants of technology use
  - Behavioral intention
  - Facilitating conditions (FC): the degree to which an individual believes that an organizational and technical infrastructure exists to support use of the system

- 4 contingencies
  - CG-1: Gender
  - CG-2: Age
  - CG-3: Experience with the technology
  - CG-4: Voluntariness of use (mandatory or voluntary setting)

Figure 6-3: Unified Theory of Acceptance and Use of Technology

TAM3 have also been proposed by Venkatesh and Bala (2008). They combine TAM2 and the model of the determinants of perceived ease of use (Venkatesh & Davis, 2000) to end to the following extended model.
Figure 6-4: Theoretical Model TAM3
6.1.3.2. Information Systems Success Models

This research stream that can provide useful elements to be taken into account for the evaluation of IS. The IS success theoretical model, was first developed by William H. DeLone and Ephraim R. McLean in 1992. The most widely used of them is DeLone and McLean model of IS success (DeLone & McLean, 2003).

It proposes seven IS success measures, which are structured in three layers: ‘information quality’, ‘system quality’ and ‘service quality’ (at the first layer), which affect ‘user satisfaction’ and also the ‘actual use’ of the IS (at the second level); finally these two variables determine the ‘individual impact’ and the ‘organizational impact’ of the IS. Seddon (1997) proposed a re-specification and extension of this model, which includes perceived usefulness instead of actual use.

From this research stream, it has been concluded that IS evaluation should adopt a layered approach based on the above interrelated IS success measures (information quality, system quality, service quality, user satisfaction, actual use, perceived usefulness, individual impact and organizational impact) and on the relations among them.
6.1.3.3. e-Services Evaluation

6.1.3.3.1. ServQual Model

SERVQUAL or RATER is a service quality framework. SERVQUAL was developed in the mid eighties by Zeithaml, Parasuraman & Berry (2002) and was initially used in a marketing context. Herschel et al. (2007) applied to IS as a measure of success.

The five gaps that organizations should measure, manage and minimize:

- Gap 1 is the distance between what customers expect and what managers think they expect - Clearly survey research is a key way to narrow this gap.
• Gap 2 is between management perception and the actual specification of
the customer experience - Managers need to make sure the organization
is defining the level of service they believe is needed.

• Gap 3 is from the experience specification to the delivery of the
experience - Managers need to audit the customer experience that their
organization currently delivers in order to make sure it lives up to the
spec.

• Gap 4 is the gap between the delivery of the customer experience and
what is communicated to customers - All too often organizations
exaggerate what will be provided to customers, or discuss the best case
rather than the likely case, raising customer expectations and harming
customer perceptions.

• Gap 5 is the gap between a customer's perception of the experience and
the customer's expectation of the service - Customers' expectations have
been shaped by word of mouth, their personal needs and their own past
experiences. Routine transactional surveys after delivering the customer
experience are important for an organization to measure customer
perceptions of service.

SERVQUAL model consists of 22 service quality measures that are organized in
five dimensions (Parasuraman et al., 1988):

• tangibles (appearance of physical facilities, equipment, personnel and
communication materials)

• reliability (ability to perform the promised service dependable and
accurately)

• responsiveness (willingness to help customers and provide prompt
service)

• assurance (knowledge and courtesy of employees and ability to convey
trust and confidence)

• empathy (provision of caring, individualized attention to customers)

Parasuraman et al. extent SERVQUAL for the evaluation service quality in web-based
environments. So they named E-S-Qual, e-service quality (Parasuraman et al., 2005).
E-S-QUAL Scale, consisting of 22 items on four dimensions:

• Efficiency: The ease and speed of accessing and using the site.
• Fulfilment: The extent to which the site’s promises about order delivery and item availability are fulfilled.
• System availability: The correct technical functioning of the site.
• Privacy: The degree to which the site is safe and protects customer information.

Parasuraman also try to measure the quality of recovery service provided by Web sites. The e-recovery service quality scale (E-RecSQUAL) consisting of 11 items on three dimensions:
• Responsiveness: Effective handling of problems and returns through the site.
• Compensation: The degree to which the site compensates customers for problems.
• Contact: The availability of assistance through telephone or online representatives.

6.1.3.3.2. E-government service quality models

E-government service quality models are mostly proposed by Parasuraman under the name of SERVQUAL model. Many research papers tried to expand or update the SERVQUAL model. Iwaarden et al. (2003) expanded the SERVQUAL model; the resulting model includes five quality dimensions corresponding to the ones of the initial SERVQUAL model, with their meaning adapted to the specificities of the websites. Papadomichelaki and Mentzas (2009) developed an e-government service quality model (e-GovQual) that includes 25 quality indicators classified into 4 quality dimensions:
• reliability
• efficiency
• user support
• trust

Liu et al (2010) established an e-government website evaluation index system using analytic hierarchy approach (AHP). The components of the index system are:
• content
• function
Fassnacht and Koese (2006) developed a broadly applicable hierarchical quality model for e-services. The model consists of three dimensions and nine sub-dimensions:

- environment quality
- delivery quality
- outcome quality

Halaris et al. (2007) model for assessing quality of e-government services consists of four layers:

- back office performance layer
- website technical performance layer
- website quality layer
- user’s overall satisfaction layer

ISO/IEC 9126 to evaluate e-service quality Behkamal et al. (2009) proposed six-quality dimensions:

- functionality
- reliability
- usability
- efficiency
- maintainability
- portability

Rowley (2011) argued that any successful e-government service should satisfy the following user benefits:

- easy to use
- accessibility and inclusivity
- confidentiality and privacy

Jaeger and Bertot (2010) argued that any attempt to create user-centered e-government services must account for a number of essential elements. These elements range from basic issues related to the ability to use e-government, to build trust to tie e-government to established social and institution requirements such as: access needs; information and service needs; technology needs; information and technology literacy; government literacy; availability of appropriate content and services; usability and
functionality; meeting user expectations; information concerns; social institutions providing access to e-government; trust; e-government 2.0; lifelong e-government usage; and understanding how users actually use e-government.

Osman et al. (2011) propose COBRAS evaluation framework (Cost, Opportunity, Benefit, Risk, Analysis for Satisfaction). The framework is developed based on the analogy to the SWOT analysis. Cost and Risk are negative for the Satisfaction while Benefit & Opportunity have positive impact. They group quality measures to four factors:

- Cost
- Benefit
- Risk
- Opportunity

Extensive reviews of e-service evaluation frameworks are provided in Rowley (2006) and Sumak et al. (2009). These frameworks suggest useful e-services evaluation dimensions and measures, with most of them assessing the quality of the resources and capabilities that the e-service provides to its users. Some others are assessing the support it provides to users for performing various tasks and achieving various objectives, or users’ overall satisfaction.

This observation leads to the categorisation of efficiency and effectiveness level of the e-service, in which there are no models that use both dimensions.

### 6.1.3.4. Value Measurement Models

According to Mechling and Hamilton (2002), the e-government Value Measurement Model (VMM) was introduced by Harvard University in response to a request to US government and was released by the Best Practices Committee of the US Federal CIO Council (2002) to assess of the value and usage of e-government websites and projects based on a multidimensional analysis of the cost/benefit, social, and political factors. The VMM framework includes five value factors:

- direct user value
- social/public value
- government financial value
- government operational/foundational value
- strategic/political value (Foley & Hamilton, 2006)
It starts with developing a set of values for each factor including: costs; risks; tangible returns; and intangible returns for each service. These values are measured through a set of dimensions/elements, and then assigned scores to each element/dimension.

Unlike VMM models, the DeLone & McLean models pay more attention to the quality of technology and user benefits with less attention to other dimensions such as cost; risk and opportunity that are very important to VMM users’ satisfactions Osman et al. (2011).

Loukis et al. (2012) proposed and verified a structured methodology for assessing and improving e-services developed in digital cities. The proposed methodology assesses the various types of value generated by an e-learning service and also the relationship among them, allowing a more structured evaluation and a deeper understanding of the value generation process. The proposed methodology estimates the value model of an e-learning service, which is structured in three layers, including the main dimensions and variables of the value it generates, and the relations among them. Figure 1 shows the three layers of the value model of an e-learning service as presented in the study.

![Value model definition of an e-learning service, adopted by [48]](image)

6.1.3.5. Web Quality Dimensions in Web 2.0 Environment

R.Saha & S.Grover (2011) they propose a model in their research, to measure website quality in WEB 2.0 environment. They end up with six dimensions; each of
these can be classified more. The above schema shows the dimensions in the form ‘cause and effect’ diagram.

![WebSite Quality Dimensions](image)

**Figure 6-8: WebSite Quality Dimensions, Source of image (2001)**

**Table 6-2: WebSite Quality Dimensions Description**

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface</td>
<td>Interface quality can be classified into proximity, compatibility, navigation, appearance, and layout dimensions.</td>
</tr>
<tr>
<td>Performance</td>
<td>The performance quality dimension may consist of availability, efficiency, reliability, security and collaboration.</td>
</tr>
<tr>
<td>Information</td>
<td>Information quality is important for success of the platform and classified into Completeness, Timeliness, Trustworthy, Presentation, Comprehensibility, Architecture, Search Capability.</td>
</tr>
<tr>
<td>Service</td>
<td>The service quality dimensions focus on how satisfactorily a website fulfills services. It includes customization, support, channel diversity, responsiveness, incentive, and compensation.</td>
</tr>
<tr>
<td>Emotional</td>
<td>The emotional quality dimension consists of factors as assurance, empathy, interaction, playfulness, and</td>
</tr>
</tbody>
</table>
6.1.4. Conclusions on Evaluation Models

For the development of our methodology we have taken into account approaches and frameworks developed from four relevant streams of previous IS research on: i) IS evaluation, ii) IS acceptance, iii) IS success and iv) e-services evaluation. A brief review of them is provided in this section.

Extensive research has been conducted on IS evaluation in the last twenty years (Smithson & Hirscheim, 1998; Farbey et al., 1999; Willcocks & Graeser, 2001; Irani, 2002; Gunasekaran et al., 2006; Irani & Love, 2008). Its main conclusion has been that IS evaluation is a difficult and complex task, since IS offer various types of benefits, both financial and non-financial, and also tangible and intangible ones, which differ among the different types of IS. Therefore each particular type of IS requires a different evaluation methodology, which takes into account its particular objectives and capabilities. Smithson and Hirscheim (1998) distinguish between two basic directions of IS evaluation. The first one is ‘efficiency-oriented’, evaluating IS performance with respect to some predefined technical and functional specifications; it focuses on answering the question of whether the IS ‘is doing things right’. The second direction is ‘effectiveness-oriented’, evaluating to what extent the IS supports the execution of business-level tasks or the achievement of business-level objectives; it focuses on answering the question of whether the IS ‘is doing the right things’. The conclusions of this research stream indicate that a comprehensive methodology for evaluating a particular type of IS should include evaluation of both its efficiency and its effectiveness, based on its particular objectives and capabilities.

Another central topic in IS research has been the identification of characteristics and factors of IS that affect the intention to use them and finally the extent of its actual usage. This research has led to the development and extensive validation of the Technology Acceptance Model (TAM) and its subsequent extensions (Davis, 1989; Venkatesh & Davis, 2000; Venkatesh et al., 2003; Wixom & Todd, 2005; Schepers & Wetzels, 2007). According to this model two characteristics of an IS, its perceived usefulness (= the degree to which users believe that using it will enhance their job
performance) and its perceived ease of use (=the degree to which users believe that using it would require minimal effort), are the main determinants of individuals’ intention to use it in the future and finally the actual use of it; perceived usefulness is also seen as being directly impacted by perceived ease of use. There have been many attempts to extend TAM, which have generally taken one of the following two approaches: i) introducing additional factors from related models, and iii) adding examining antecedents and moderators of perceived usefulness and perceived ease of use (Venkatesh & Davis, 2000; Venkatesh et al., 2003; Wixom & Todd, 2005; Schepers & Wetzels, 2007). The conclusions of this IS acceptance research stream indicate that a methodology for evaluating a particular type of IS should assess its ease of use, usefulness and users’ intention to use it in the future.

Another research stream that can provide useful elements is the IS success research (DeLone & McLean, 1992; 2003; Seddon, 1997). The most widely used IS success model has been developed by DeLone and McLean (1992). It proposes seven IS success measures, which are structured in three layers: ‘information quality’, ‘system quality’ and ‘service quality’ (at the first layer), which affect ‘user satisfaction’ and also the ‘actual use’ of the IS (at the second level); these two variables determine the ‘individual impact’ and the ‘organizational impact’ of the IS. Seddon, (1997) proposed a re-specification and extension of this model, which includes perceived usefulness instead of actual use. The conclusions of this research stream indicate that IS evaluation should adopt a layered approach based on the above interrelated IS success measures (information quality, system quality, service quality, user satisfaction, actual use, perceived usefulness, individual impact and organizational impact) and on the relations among them.

The emergence of numerous Internet-based e-services (e.g. information portals, e-commerce, e-banking, e-government, etc.) lead to the development of specialised frameworks for evaluating them (Lu & Zhang, 2003; Rowley, 2006; Fassnacht & Koese, 2006; Sumak et al., 2009; Saha & Grover, 2011). Extensive reviews of this research are provided by Rowley (2006) and Sumak et al. (2009). These frameworks suggest useful e-services evaluation dimensions and measures. Most of them assess the quality of the capabilities that the e-service provides to its users (being oriented towards the abovementioned efficiency evaluation). Some others assess the support it provides to users for performing various tasks and achieving various objectives (being oriented
towards the above-mentioned efficiency evaluation). However, most of the above frameworks do not include advanced ways of processing the evaluation data collected from the users, aiming to maximize the extraction of value-related knowledge from them. They include mainly simple calculations of average values of all evaluation measures and dimensions over all the users who evaluate the e-service; the relations among the proposed evaluation dimensions and measures, which could form the basis for advanced multi-dimensional statistical analysis, are not exploited all for drawing more conclusions and insights.

Only recently some research in this direction has been conducted. Pazalos et al. (2012) and Loukis et al. (2012) proposed and verified a structured approach for assessing and improving e-services, which is based on the estimation of value models of them from users’ ratings. Such a value model consists of a set of value measures, assessing different types of value generated by the evaluated e-service, and also the relations among them. These value measures are organized in three layers:

(a) Efficiency layer: it includes ‘efficiency’ measures, which assess the quality of the basic capabilities offered by the e-service to its users.
(b) Effectiveness layer: it includes ‘effectiveness’ measures, which assess to what extent the e-service assists the users for completing their tasks and achieving their objectives.
(c) Future behaviour layer: it includes measures assessing to what extent the e-service influences the future behaviour of its users (e.g. to what extent they intend to use the e-service again in the future, or recommend it to friends and colleagues).

The above value model shows how value generation starts through capabilities offered to the users, and then how this is transformed to support for completing users’ tasks and achieving their objectives, and finally how this affects their future behaviour; in this sense a value model enables a better understanding of the whole mechanism of value generation by the e-service. Also, it enables a rational definition of priorities for improvements in the capabilities it offers to users (in the first layer of the model), by giving highest priority to the improvement of the capabilities receiving lower users’ ratings and at the same time having higher impact on the measures of the higher level value (in the second and third layer). Such an approach can be useful for the development of a comprehensive methodology for evaluating second generation OGD infrastructures, after appropriate adaptations: inclusion in the first layer of quality measures of the main capabilities offered by these advanced OGD infrastructures, and in
the second layer of measures of the support they provide to their users for achieving their multiple objectives (associated with OGD 'pro-sumption').

6.2. Designing an Evaluation Framework

An important objective was to propose a unified evaluation framework that can be used for the evaluation of both traditional OGD e-infrastructures based on the web 1.0 paradigm and advanced ones based on the web 2.0 paradigm. As it is described above, traditional and advanced OGD e-infrastructures are both used by two main groups of stakeholders: the data providers and data users. In the case of traditional OGD infrastructures there are simpler functionalities corresponding either to the data users or to the data providers and these two groups of stakeholders have clearly divided roles. Consequently, different capabilities are provided to each group, so two corresponding evaluation sub-frameworks have to be built. On the other hand, advanced OGD e-Infrastructures are characterized by frequent and repeated role switching when users interact with the platform. The proposed evaluation framework consists of an integrated evaluation model and a comprehensive evaluation procedure that utilises both qualitative and quantitative methods in order to get meaningful insights based on the evaluation variables and measures were identified during the construction of the evaluation model.

For the above-mentioned reasons we developed an evaluation model for the most complex and difficult case of advanced OGD e-infrastructures 2.0, which is shown in figures 6.9 and 6.10.
Figure 6-9: Evaluation dimensions and criteria of the OGD e-Infrastructures evaluation model (continues)

It consists of evaluation dimensions (coming from sections 3.6, 4.5 and 5.4.2) which are further elaborated into evaluation criteria, and covers both data users’ and data providers’ perspectives. So, if a user interacts with the system taking only the role of data user/provider, he/she will assess only value dimensions/criteria corresponding to data users/providers (denoted with ‘U’/’P’ respectively in the following Figures 1&2), while if he/she has both roles, he/she will have to assess all value dimensions/criteria. Similarly, if the proposed evaluation framework is used for the simpler case of evaluating a traditional OGD e-infrastructures 1.0, it can be easily divided into two parts, one for the data users and one for the data providers.
6.2.1. Evaluation procedure

In order to finalise the construction of the evaluation framework, a comprehensive evaluation procedure was developed for the use of the presented evaluation model. The evaluation procedure includes both quantitative and qualitative evaluation methods to get deeper insights.

The evaluation approach is shown in Figure 6.11. It consists of internal evaluation (by individuals within the e-infrastructure development consortium) and external evaluation (by individuals not belonging to the development consortium). The internal evaluation includes: i) a semi-structured discussion in a group of experts, ii) Web Analytics and iii) a SWOT (Strengths-Weaknesses-Opportunities-Threats) analysis. The external evaluation includes: i) a similar semi-structured discussion in a group of social, political and management sciences researchers (potential users of the platform) not belonging to the project’s scientific committee, ii) a structured usability test measuring user’s performance on specific tasks (using a 5-Likert scale from very difficult to very easy), iii) a quantitative and structured questionnaire and iv) a Qualitative Discussion with users.

Figure 6-10: Evaluation dimensions and criteria of the OGD e-Infrastructures evaluation model (continued)
Figure 6-11: Evaluation Approach and Tools

The structured questionnaire follows the complete structure of the evaluation model and could be used for both traditional and advanced OGD e-infrastructures with the appropriate changes. The users who fill the questionnaire are asked to enter the extent of their agreement or disagreement with the statements it includes, answering the question: “To which extend do you agree with the following statements?” A five point Likert scale is used to measure agreement or disagreement with (i.e. positive or negative response to) such a statement (1= Strongly Disagree, 2= Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree).

The semi-structured discussions in groups of experts (e.g. researchers from the social, political and management sciences who would be interested in using public sector data), on the topics/questions shown in Figure 6.12.

<table>
<thead>
<tr>
<th>Internal assessment</th>
<th>External assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ENGAGE Consortium</strong> (including Experts Scientific Committee)</td>
<td></td>
</tr>
<tr>
<td>Semi-structured Experts Questionnaire</td>
<td>End user</td>
</tr>
<tr>
<td>Semi-structured SWOT-Analysis</td>
<td>Experts in the field</td>
</tr>
<tr>
<td>Web Analytics</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6-12: Topics/questions of the experts’ groups semi-structured discussions

The qualitative discussion can be based on the main evaluation dimensions and criteria of the OGD e-Infrastructures evaluation model described in the previous section, so it focus on the following topics/questions, presented in Figure 6.13:
Do you find the platform useful for downloading/using open governmental data?

Do you find the platform useful for publishing/uploading your own data, or improved/enriched versions of government data you have downloaded?

Which of the platform features (Searching, Metadata, Downloading, Comments and Rating, Linking, Uploading, Other) are useful for your needs?

Which of the platform features are not useful for your needs?

Which platform’s features are easy to use?

Which platform’s features are difficult to use?

Which features are not provided by the platform, but would be useful for your needs?

Do you currently participate in communities’ formation (based on shared interest in particular types of government data) within the platform?

Would you like to participate in communities’ formation within the platform in the future?

Would you be willing to contribute data to the platform?

Figure 6-13: Qualitative Discussion

For the usability test, a number of potential users are asked to perform a scenario (= a series of tasks) using the platform under evaluation. There are three possible ways of practically conducting the usability testing:

- Peer testing by users: whilst one user is testing the platform and performing the tasks, the other user is monitoring and providing feedback to the session leader by filling in the evaluation form.
- Session leaders evaluate a random sample of users according to their choice.
- Individuals test and evaluate themselves (self-organization principle): users perform the tasks and fill in the evaluation sheet themselves.

6.2.2. Application on the ENGAGE platform (version 2.0)

6.2.2.1. External Evaluation

6.2.2.1.1. Usability Test

In this section the results of the student usability test are described. First, the results that were derived from two Dutch civil servants are provided and, second, the results that were derived from students of Delft University of Technology are provided.

Usability test with civil servants of the Research and Documentation Centre, the Netherlands

On May 14, 2013, two civil servants were asked to perform three scenarios.
- Scenario 1: Searching, downloading, extending/visualizing/curating/linking and uploading interesting datasets;
- Scenario 2: Getting information about other open data websites and comparing them via the ENGAGE website;
- Scenario 3: Getting information about manuals, API’s and tutorials (training).

While the civil servants were performing the scenarios, they provided the following feedback.

- Registration with Google did not work.
- After I am registered I receive an e-mail about the registration. This e-mail could be more comprehensive. Right now it is not clear that it concerns my registration on the ENGAGE platform. In the e-mail you could also add other information (e.g. a reference to the platform and the project site or to social media).
- Multilingualism is a very strong selling point of the platform.
- Certain search terms do not give results. When I search in Dutch, the terms are probably not translated. It would be nice if the terms would be translated during the search. (especially because you claim to be multilingual).
- When I search for ‘boroughs’, I get two Greek datasets. But when I select ‘Greece’ in the ‘more options’ tool, these two datasets disappear.
- When I select ‘more options’, it is unclear whether ‘country’ means the country of the data publisher, the data author or the country covered by the dataset. This is unclear and confuses me.
- Sometimes when I click on ‘download dataset’, I am forwarded to another website (e.g. the website of DANS). I don’t like this, because I have to login again and I cannot use the ENGAGE tools then. I prefer to view the datasets on the ENGAGE platform.
- I tried to download three different datasets, and all three times it did not work. Usually I would have given up at that moment and go to another website to try to find open data. I could download the fourth dataset.
- In some datasets you cannot understand what it is about. I cannot read Greek so it would be nice if this would be translated too.
- Sometimes visualizations do not work. The respondents tried to visualize dataset 14, 22 and 26, which did not work. Eventually one of them managed to visualize another dataset.
- When I visualize a dataset (concerns dataset 8), I cannot save the visualization (nothing happens when clicking on ‘save’). I also cannot export the visualized dataset (nothing happens when clicking on ‘export’).
- When I use the chart designer to visualize a dataset I cannot go to the previous page within the chart designer. When I am in the chart designer and I click on 'previous page' I go back to the overview of the dataset, but not to the previous screen within the chart designer. This is inconvenient because now I have to filter and select again what I want to visualize.
- When I use the chart designer and I fill out the variables that I want to add to the chart, the pop-up screen of the chart designer is shown on top of the dataset. So when I select the chart designer, I cannot see the dataset anymore.
- The comparison and overview of open data sites is very interesting.
- In the overview of open data sites it would be very interesting to see the contact details of persons who maintain these sites.
- The datasets that are presented on the platform at this moment are not very valuable.
- When I visualize the licences overview for open data sites, half of the labels of the visualization is not shown (they “fall from the page”)

In sum, many comments were related to the lack of sufficient and interesting datasets on the platform and the difficulties with visualizing datasets.

**Usability test with students of Delft University of Technology, the Netherlands**

On May 27, 2013, 17 students were asked to perform the same three scenarios as the civil servants did. The scenarios were performed from different roles and perspectives (researchers, citizens, journalists, developers, civil servants, entrepreneurs and archivists/librarians). While the students were performing the scenarios, they provided the following feedback.

- There are way too few datasets on the ENGAGE platform. So there is not use in searching. (many students provided this comment)
- Visualizations do not work (many students provided this comment)
- I do not understand why I should upload a dataset. Could that really be useful to another person?
- Where can I find what the license of a dataset is?
- The overview of data sites is not up to data (e.g. some Dutch municipalities are missing)
- I cannot find useful datasets.
- I do not like to be referred to another website when I click on ‘view’ the dataset.
- My dataset does not fit in the categories for uploading certain datasets.
- Problems with date selection tool in the data upload menu.
- The open data site was not really working properly in Internet Explorer.
- Often when I pushed a button nothing happened and I can barely find any datasets via the search function.
- Visualization is not working very well. Also this website does not work well on Internet Explorer.
- Temporal Granularity is a term I’ve never heard of.
- The date selection tool is not working.
- There are not many open data sets available.
- Had some difficulty with visualizing some datasets.
- The editing tools are unresponsive and not intuitive. There aren’t a lot of datasets at the moment and the data-sets are quite large at times which makes editing a real nightmare.
- The main problem using these datasets is that is hard to find specific on the topic you’re searching for. There is not (yet) much information when you are looking for a very specific topic. If you do want to find proper information you have to be pretty vague in your search terms.
- Visualization didn’t work for me. And due to old excel files the uploading didn’t work to great.
- The layout of the page isn’t easy to read. I recommend making the text-boxes thinner to make the text less intimidating and maybe change the background of the text itself to make it visually more pleasing.
- The required load-time is a tad slow, but it works quite well.
- ERROR: stop running this script? A script on this page is causing your web browser to run slowly.

6.2.2.1.2. Online User Questionnaire

After the performance of the usability test, the civil servants and the students were asked to fill out the online user questionnaire. They were all provided with the link to the questionnaire and completed the questionnaire in about 15-20 minutes. In total, 2 civil servants and 15 students (with various roles) completed the questionnaire. One respondent is from Suriname, all the other respondents are from the Netherlands. Almost all questions were answered by all respondents. Most respondents used Mozilla Firefox to access the ENGAGE platform.

The number of respondents does not allow for advanced statistical analysis of the gathered data. For this reason, we only report on the data on a more aggregated
level. Although opinions are divided, most users stated that the ENGAGE platform is easy to use and it is easy to learn how to use the platform. The respondents state that the platform allows them to work in their own language and that it supports user account creation in order to personalize views and information shown.

The results from the online questionnaire are presented in more detail in section 6.2.2.3.

6.2.2.1.3. Qualitative discussion

About 16-19 people participated in a qualitative discussion during a workshop at the Conference on e-Democracy and Open Government (CeDEM13). The discussion took place on May 22, 2013 at Krems an der Donau, Austria. Most participants were researchers, some were civil servants and one was an entrepreneur. The participants came from over seven different countries.

Additionally, twelve people participated in the Open Space Programme (workshop) at the Conference on e-Democracy and Open Government (CeDEM13) on May 24, 2013 at Krems an der Donau, Austria. Most participants were researchers, some were civil servants and one was an entrepreneur. They came from five different countries. These people provided the following comments and suggestions during a qualitative discussion.

Moreover, two civil servants of the Research and Documentation Centre, Ministry of Security and Justice, The Netherlands, participated in a qualitative discussion.

Finally, twelve students of Delft University of Technology in The Netherlands participated in the discussion in different roles (researchers, citizens, journalists, developers, civil servants, entrepreneurs and archivists / librarians).

The results of these discussions are as follows:

- The approach of the platform and the ideas behind it are very interesting.
- The platform looks really nice and the functionalities would be very useful if they would all work.
- You can see that the platform has just been released, it still needs to develop further.
- If the bugs on the platform would be removed, the website would become very interesting for the developments of the scientific world, especially for people
who just start working as a scientist (e.g. master students writing their thesis). I think it would not be used by large research organizations.

- If the platform would work, it would be easy to use and user friendly.
- Advantages of the platform are that the datasets are available for free, can be obtained in an easy way, data are provided about different topics.
- Additional features that would be nice to have on the platform is more information about the context in which the dataset was created. Especially the source is important. I would also like more information in the context of law: terms of use (These are not part of the scenarios, so they were not evaluated by the respondents, but they are available) and datasets related to law and crime. Furthermore, maybe you should block uploading a dataset if insufficient metadata are provided.

To summarize, most comments and suggestions concern assessing datasets (reliability / trustworthiness), a lack of (interesting) datasets and problems with visualizations. These issues were taken up by the project partners who are responsible for them.

6.2.2.1.4. **SWOT Analysis**

In this section the results of the SWOT analysis by the ESC members are presented. The following sections show the SWOT analysis of Grigoris Antoniou, Spyros Blavoukos and Enrico Ferro.

**Grigoris Antoniou** is a professor at the University of Huddersfield. He is an expert in Web-based systems (Semantic Web, Web services, context-aware systems), e-Commerce and Ambient Intelligence. Professor Antoniou provided the following input.
### (a) Technical aspects (e.g. metadata approach, data processing, development)

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reputation management</td>
<td>Reliance on external tools</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(e.g. for curation etc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use of rich metadata</td>
<td>Too broad scope?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### (b) Sustainability aspects (e.g. business cases, exploitation, dissemination)

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government involvement/support</td>
<td>RDA</td>
<td></td>
<td>Other competitor platforms</td>
</tr>
<tr>
<td></td>
<td><a href="http://rd-alliance.org/">http://rd-alliance.org/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardization activities</td>
<td>PRELIDA</td>
<td></td>
<td>Sufficient completeness</td>
</tr>
<tr>
<td></td>
<td><a href="http://www.prelida.eu">www.prelida.eu</a></td>
<td></td>
<td>of data repository</td>
</tr>
</tbody>
</table>

---

**Spyros Blavoukos** is a lecturer at the Department of International and European Economic Studies at the Athens University of Economics and Business. He performs political research. Spyros Blavoukos provided his view on the "sustainability" aspects.

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geographical and data scope</td>
<td>All-encompassing target group does not help easy association of a profession with the project platform (‘everybody’ turned to ‘nobody’)</td>
<td><strong>Greek Presidency of the EU (first semester of 2014):</strong> great opportunity to spread the word; try to use available dissemination means and transgovernmental communication channels to promote ENGAGE</td>
<td>Early stages: data uploaded may shape the direction/ ‘image’ of the project and attract expert-groups, diluting the intention of broad geographical and issue area coverage. Needs balance and some kind of control (to avoid over-representation of particular sectors/ countries/issuing bodies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Build on the ‘request’ function – at the beginning at least, until the social science community gets actively involved, it would be good to strengthen this component to attract ‘customers’ and help</td>
<td>Needs constant flow of data inclusion to avoid stagnation and convey dynamism; are there adequate resources in the beginning until it becomes self-sustainable?</td>
</tr>
<tr>
<td>The cross-fertilization of data sets needs further illustration (it is necessary to construct ‘successful examples’/ case studies to show how useful it is!)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---
Social scientists in each respective field are already using their own databases (e.g. economic databases). Why care for ENGAGE? What's the added value? (It’s only science that is inter- and cross-disciplinary that could profit from such a venture – this is a weakness but also a challenge for ENGAGE)

Special ‘info day’ at Universities with strong social science departments and research expertise to present the platform/project

Watch out not to overemphasize at this stage technical aspects at the expense of data availability; you may end up with a ‘perfect’ platform technically-speaking but not able to ride the wave!

<table>
<thead>
<tr>
<th>Strengths</th>
<th>Weaknesses</th>
<th>Opportunities</th>
<th>Threats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Being the gatekeeper of a community of data scientists</td>
<td>Still unclear understanding of the targeted customer segments and their willingness to pay</td>
<td>Connect with the open access initiative for science (<a href="http://blog.okfn.org/2013/03/21/we-are-entering-an-era-of-open-science-says-eu-vp-neelie-kroes/">http://blog.okfn.org/2013/03/21/we-are-entering-an-era-of-open-science-says-eu-vp-neelie-kroes/</a>)</td>
<td>Verticalization on scientific data may not be sufficient for sustainability</td>
</tr>
<tr>
<td>Possibility to become a platform for third parties developers</td>
<td>Need to specify differential value proposition to existing experiences</td>
<td>Add information about EU grants allocation by organization and list of projects financed (this is highly valuable information researchers may be willing to pay for as it is functional for writing and winning new research money)</td>
<td>Legal barriers to data restitution to PAs</td>
</tr>
<tr>
<td>European scope</td>
<td>Emergence of data centrality in science, policy making and society</td>
<td></td>
<td>Harmonization of EU diversities (legal and technical)</td>
</tr>
</tbody>
</table>

**Enrico Ferro** is the head of the Business Model & Policy Innovation Unit at ISMB, in charge of the research and policy intelligence activities. *He is an expert on business models and exploitation strategies for Public Sector Information. Enrico is actively involved in the organization of a number of international academic events dealing with eGovernment and eInclusion. Enrico provided his view on the "sustainability" aspects.*
6.2.2.1.5. **Experts Questionnaire**

In addition to an internal expert questionnaire, an external expert questionnaire was conducted during the Conference on e-Democracy and Open Government (CeDEM2013). In total 12 external experts filled out the questionnaire. The results can be found in the two tables below.

The external experts mainly stated to which extend they found the USP's of the ENGAGE-project important. Some experts provided additional comments. The tables show that the external experts disagree on the importance of many USP's. In general, the experts agree that ENGAGE should be a marketplace for open data and collaboration (USP2), ENGAGE should put emphasis on rich metadata (USP 4), ENGAGE should provide a rich collection of data visualisation tools (USP 6), ENGAGE should develop cases of using the platform (USP 7), ENGAGE should provide a full API for machine-to-machine operation (USP 10), ENGAGE should target multiple-nationalities (USP 11).

<table>
<thead>
<tr>
<th>The model requires to collaborate with a network of organizations that may not be ready or willing to open their data</th>
</tr>
</thead>
</table>

The model requires to collaborate with a network of organizations that may not be ready or willing to open their data.
<table>
<thead>
<tr>
<th>USP 1: ENGAGE should be a marketplace for open data and collaboration</th>
<th>Expert 1</th>
<th>E2</th>
<th>E3</th>
<th>E4</th>
<th>E5</th>
<th>E6</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/7. What do you mean by marketplace? In the sense of &quot;portal&quot; / platform or a selling point?</td>
<td>6/7</td>
<td>6/7</td>
<td>4/7</td>
<td>6/7</td>
<td>4/7</td>
<td></td>
</tr>
</tbody>
</table>

| USP 2: ENGAGE should be an open public data reputation management system | 7/7 | 3/7 | 6/7 | 4/7 | 4/7 | 4/7 |

| USP 3: ENGAGE should put emphasis on rich metadata. | 4/7. Hard to answer. Google/Bing/... is doing an excellent job by inspection the actual data, scheme.org helps Google/Bing/... so we do not know yet if metadata will have that importance in 10 years as it has now. | 6/7 | 4/7 | 4/7 | 5/7 Metadata is considered to be one of the most important aspects in the Open Government Data community | 6/7 |

| USP 4: ENGAGE should provide a rich collection of data curation tools. | 7/7 | 5/7 | 5/7 | 4/7 | 4/7 | 5/7 |

| USP 5: ENGAGE should provide a rich collection of data visualisation tools. | Importance: 7 out of 7. | 4/7 | 6/7 | 7/7 | 6/7 in order to overcome a possible digital data divide. visualisation is crucial in any open data project | 7/7 Visualization is probably the best instrument for effective reuse of open data |

| USP 6: ENGAGE should develop cases of using the platform. | 7/7 A case tremendously helps to "sell" the idea to those who are affected by a case. Maybe go to those who already employ a national open data umbrella portal for their benefits and model the use cases from that feedback. | 7/7 | 6/7 | 4/7 | 5/7 | 6/7 |

| USP 7: ENGAGE should focus on a standardization | 4/7 | 4/7 | 6/7 | 6/7 | 5/7 | 6/7 |
## Proposal for Annotating Open Data Sets for Scientific Usage

| USP 8: ENGAGE should provide a complete data repository. | 6/7. Are there any good? | 3/7 | 6/7 | 6/7 | 4/7 | 5/7 |
| USP 9: ENGAGE should provide a full API for machine-to-machine operation. | 7/7 | 6/7 | 5/7 | 6/7 | 5/7 | 6/7 |
| USP 10: ENGAGE should target multiple nationalities. | 5/7 | 6/7 | 7/7 | 6/7 | 3/7 should not be priority | 6/7 Important for reuse in EU |

### What should be the ENGAGE dissemination steps towards sustainability?

- Integrate it into Joinup / Make it interoperable with joinup.
- Connect with the already existing community and focus on meta data and visualization.
- Try to play a role in the EC plans of changing the Psi directive into a Psi regulation with the result that public bodies may publish data wherever they want, but with the obligation to publish their open datasets on Engage.

### Which other steps/domains/USP’s should ENGAGE target according to you?

- Deportement of Information & Communication Systems Engineering
<table>
<thead>
<tr>
<th>E7</th>
<th>E8</th>
<th>E9</th>
<th>E10</th>
<th>E11</th>
<th>E12</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>USP 1: ENGAGE should be a marketplace for open data and collaboration</strong>&lt;br&gt;4/7 For sustainability purposes.</td>
<td>7/7</td>
<td>If it is a marketplace, then will the goods be priced? Who decides on the price?</td>
<td>5/7</td>
<td>6/7 As a marketplace, it should make users aware of what’s on offer and facilitate its distribution.</td>
<td>5/7 Marketplace provision requires a high degree of accessibility design to facilitate strong use.</td>
</tr>
<tr>
<td><strong>USP 2: ENGAGE should be an open public data reputation management system</strong>&lt;br&gt;4/7 On the other hand it should be a one stop center for high quality data to everyone.</td>
<td>7/7</td>
<td>6/7 The reputation management should focus on quality of goods and service delivery and data source rather than users and participation.</td>
<td>5/7</td>
<td>6/7 The reputation of the data source will be exposed as reputable or not by this means. If claims are made which cannot be substantiated when the data is examined by third parties, these will come out by user rating.</td>
<td>4/7</td>
</tr>
<tr>
<td><strong>USP 3: ENGAGE should put emphasis on rich metadata.</strong></td>
<td>7/7</td>
<td>5/7</td>
<td>-</td>
<td>5/7</td>
<td>2/7 This kind of information can be sourced separately, from the source itself. Why does ENGAGE have to re-interpret this info? This is a risk that it gets the context wrong, or the context changes and ENGAGE does not update its files. It seems to be an unnecessary burden.</td>
</tr>
<tr>
<td><strong>USP 4: ENGAGE should provide a rich collection of data curation tools.</strong></td>
<td>7/7</td>
<td>5/7</td>
<td>1/7</td>
<td>5/7</td>
<td>6/7 The search tools are key. Indexing and sorting of data is a fundamental task. How do we look for something if we do not know if it exists?</td>
</tr>
<tr>
<td><strong>USP 5: ENGAGE should provide a rich collection of data visualisation tools.</strong></td>
<td>7/7</td>
<td>6/7</td>
<td>7/7 See comment on USP 8</td>
<td>6/7</td>
<td>7/7 If ENGAGE is to be of interest to anyone beyond the computer nerds, then this is key.</td>
</tr>
<tr>
<td><strong>USP 6: ENGAGE</strong></td>
<td>7/7</td>
<td>6/7</td>
<td>2/7</td>
<td>7/7</td>
<td>5/7 Useful for promoting your</td>
</tr>
</tbody>
</table>
should develop cases of using the platform.

**USP 7: ENGAGE should focus on a standardization proposal for annotating open data sets for scientific usage.**

<table>
<thead>
<tr>
<th>3/7</th>
<th>7/7</th>
<th>5/7</th>
<th>7/7</th>
<th>3/7</th>
<th>2/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/7</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1/7</td>
<td>1/7</td>
</tr>
</tbody>
</table>

**USP 8: ENGAGE should provide a complete data repository.**

<table>
<thead>
<tr>
<th>7/7</th>
<th>7/7</th>
<th>7/7</th>
<th>7/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**USP 9: ENGAGE should provide a full API for machine-to-machine operation.**

<table>
<thead>
<tr>
<th>6/7</th>
<th>6/7</th>
<th>6/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>7/7</td>
</tr>
</tbody>
</table>

**USP 10: ENGAGE should target multiple nationalities.**

<table>
<thead>
<tr>
<th>7/7</th>
<th>7/7</th>
<th>7/7</th>
<th>5/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**What should be the ENGAGE dissemination steps towards sustainability?**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Depends on who should be the users. Need to make it well-known. Elasone its dissemination of this, researcher networks, research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It needs to be free to use, funded by advertising or the EU or by academic research councils.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Articulation of commercial value and targeting smart cities.</td>
</tr>
<tr>
<td><strong>Which other steps/domain s/USP’s should ENGAGE target according to you?</strong></td>
<td><strong>I think usability and more dissemination are important.</strong></td>
<td><strong>Legal issues are not clear to me as far as the information provision is concerned. To what extent has the ENGAGE-project focused on this issue? Which partner of the consortium is responsible for that?</strong></td>
<td><strong>Common format service (REST wrappers). High-value domain specific (agriculture / transport / energy) Geo-referring (OGD standards / GNSS frameworks)</strong></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6.2.2.2. Internal Evaluation

6.2.2.2.1. Experts Questionnaire

The five project partners that are involved in Work Package 6 filled out the experts questionnaire. The questionnaire asked for their opinion on the importance of the Unique Selling Points (USP’s) and a few other questions on sustainability and dissemination. The results of the questionnaire can be found in the table below.

The table below shows that the project partners euroCRIS, TU Delft, Intrasoft and AEGEAN have a very similar and complementary perspective on the ENGAGE-project. In sum, they state that rich metadata (USP 3), the development of use cases (USP6) and a full API for machine-to-machine operation (USP9) are the most important USP’s of the ENGAGE-project. Fokus Fraunhofer seems to agree with some answers of these four project partners, but has a different view on USP1 (ENGAGE should be a marketplace for open data and collaboration), USP3 (ENGAGE should put emphasis on rich metadata), USP5 (ENGAGE should provide a rich collection of data visualisation tools), USP6 (ENGAGE should develop cases of using the platform), USP7 (ENGAGE should focus on a standardization proposal for annotating open data sets for scientific usage), USP8 (ENGAGE should provide a complete data repository), USP9 (ENGAGE should provide a full API for machine-to-machine operation). The differences in opinions about the ENGAGE-project need to be discussed internally, to make sure that all activities are geared to one another.
| USP 1: ENGAGE should be a marketplace for open data and collaboration | Importance: 6 out of 7. Other data.gov sites do not provide this facility so it could be another unique selling point for ENGAGE. The combination of many datasets, rich metadata, some processing facilities and – specially – associated social networking / cooperative support give ENGAGE a lead. | Importance: 6 out of 7. Different stakeholders of publishing and using open data are currently not able to collaborate on any existing data.gov website. ENGAGE could add value by providing mechanisms for collaboration. | Importance: 6 out of 7. Important in terms of data exchange, but not in the literal sense of a "marketplace" (i.e. payment for datasets etc.) | Importance: 7 out of 7. That is the most important for ENGAGE. | Importance: 4 out of 7. Scientists working with data imho usually know their reliable data sources. Data journalists can be more interested in data marketplace than researchers. |
| USP 2: ENGAGE should be an open public data reputation management system | Importance: 6 out of 7. The subject of reputation or quality of datasets is difficult but should be attempted in ENGAGE by a combination of provider assertions, peer review and end-user rating. | Importance: 6 out of 7. By giving credits to users of the ENGAGE-platform and by showing activities performed by users of the ENGAGE-platform, ENGAGE could play a role in reputation management. For instance, a rating system for assessing the quality of data is very important, but has many difficulties that need to be examined. | Importance: 6 out of 7. Yes, it should provide the ability to rate open datasets based on their (perceived) quality and usage. But it is important for ENGAGE overall to have a reputation as a platform that deals with high-quality datasets. | Importance: 7 out of 7. | Importance: 6 out of 7. It is great for any portal or organization to become a reputation management system. However, it is difficult to achieve this goal. |
| USP 3: ENGAGE should put emphasis on rich metadata. | Importance: 7 out of 7. Existing data.gov websites have very limited metadata. Even the few percent having CKAN or DC have variable quality. If ENGAGE obtains datasets from the scientific community (widest sense) then they come with much richer metadata. Furthermore ENGAGE could provide facilities for improving poor metadata (enrichment). The use of CERIF (an EU Recommendation to Member States) provides an appropriate technology from which one can generate other metadata formats thus improving interoperability and access. | Importance: 7 out of 7. The CERIF metadata approach provides rich metadata, including discovery, contextual and detailed metadata. By providing rich metadata, the ENGAGE platform goes further than other existing data.gov websites. In addition, the rich metadata approach helps in meeting many of the requirements that were identified by potential ENGAGE users. During the dissemination and evaluation activities, we noticed that rich metadata is really the key to many problems of existing data.gov websites. | Importance: 7 out of 7. This puts ENGAGE ahead of other platforms and it is very important for interoperability and sustainability. | Importance: 7 out of 7. | Importance: 2 out of 7. Imho researchers are interested in data and not in metadata. Rich metadata is a good thing for an Open Data search portal, which is not the case for Engage. |
This is another unique selling point of ENGAGE.

**USP 4: ENGAGE should provide a rich collection of data curation tools.**

**Importance: 6 out of 7.** The whole subject of data curation is fraught, not least because there is no accepted or acceptable business model (most equations fail if they include a term infinity!). However there are best practices proposed by the Alliance for Permanent Access to the Records of Science and if ENGAGE followed the guideline and implemented curation this would distinguish ENGAGE from other open data platforms.

**Importance: 6 out of 7.** Currently, open data users often have to find curation tools themselves. Integrating curation tools on the ENGAGE-platform provides added value compared to existing open data websites.

**Importance: 5 out of 7.** This is important but it should focus on the quality of the tools, rather than the number of tools.

**Importance: 6 out of 7.** Yes, if we want ENGAGE become more than a metadata/data storage and exchange platform and build a community around it.

**USP 5: ENGAGE should provide a rich collection of data visualisation tools.**

**Importance: 6 out of 7.** With data of even moderate complexity visualisation is a great help to understanding. Thus it should be possible in ENGAGE environment to visualise data. However there is a technical problem if the datasets (especially if large) are located physically on server(s) remote from the ENGAGE portal; it may be better to run the visualization software at the remote location and just present the resulting images at the ENGAGE portal. However, this implies ENGAGE having rights to run software on other servers. This area of facility in ENGAGE need much more careful research on how best to achieve it. For now visualisation of small datasets can be done locally on the ENGAGE server. This may be a subject for a later project.

**Importance: 6 out of 7.** Visualizations make it possible to analyze datasets in an easy way and they are therefore very important for the target groups of the ENGAGE-project.

**Importance: 5 out of 7.** It is important for people to understand existing data and visualise their own derived datasets, but it should not be the main focus of the project, given that this area is widely covered already.

**Importance: 6 out of 7.** Visualisations allow quicker decisions on if the data are of interest to the user.

**Importance: 6 out of 7.** There are many visualisation tools. Mashable visualisations can be of high interest for journalists, or even researchers. However, the Engage visualisation tools should be attractive for the users.

**USP 6: ENGAGE should**

**Importance: 7 out of 7.** The whole idea of ENGAGE is not necessarily easy for an end-user to understand.

**Importance: 7 out of 7.** During the dissemination and evaluation activities that were performed for

**Importance: 7 out of 7.** We should develop

**Importance: 7 out of 7.** The use cases

**Importance: 4 out of 7.** Cases are useful to p
| **USP 7:** ENGAGE should focus on a standardization proposal for annotating open data sets for scientific usage. | **Importance:** 5 out of 7. Certainly ideally the ENGAGE metadata standard should be standardized even more widely (it is already an EU recommendation to Member States) as should the proposed API(s) and the overall architecture. Standardisation increase the likelihood of market penetration and use. However, standardisation is a long process with many discussions and compromises and the end-result is not always that which was intended at the beginning! | **Importance:** 5 out of 7. It would be excellent if ENGAGE could play this important role, but it might be too ambitious. However, we are not the expert on standardization, so it is difficult for TU Delft to assess this. | **Importance:** 5 out of 7. This is an important subject, but developing a standardisation proposal that can achieve sustainable impact may be beyond the scope of ENGAGE at this stage. | **Importance:** 6 out of 7. It would assist users to understand better and more quickly how the platform functions and what are the potential usage. 1 use case per target group unveiling the true need for this target group. | **Importance:** 4 out of 7. IMHO that will not help the ENGAGE platform to become sustainable, but it is still interesting topic for the project. |
| **USP 8:** ENGAGE should provide a complete data repository. | **Importance:** 6 out of 7. ENGAGE should provide a catalog of metadata which describes a complete repository. The location of the actual datasets should not be on the ENGAGE platform (impractical in terms of resources, problematic in ensuring update, some difficult rights / licensing issues) but where they are generated or managed / curated so the originator has control of the access rights and also has the responsibility for update and curation. The coverage of datasets by ENGAGE | **Importance:** 6 out of 7. During the workshops, potential users of the ENGAGE-platform stated that they would like to use a one-stop-shop to find open data. ENGAGE should harvest the metadata of many datasets, so that it could function as a catalogue. ENGAGE can refer to the websites where the original datasets can be found. Functioning as a catalogue harvesting metadata enables potential users of ENGAGE to find many types of datasets on other websites via the ENGAGE platform. | **Importance:** 7 out of 7. ENGAGE must become the “go to” place for open datasets for social scientists. | **Importance:** 3 out of 7. ENGAGE consortium should concentrate on the other USPs more. There are a lot data repositories out there. | **Importance:** 2 out of 7. The idea is interesting but: 1) we need a target community requiring the data 2) ENGAGE project will be finished in 1 year. 3) We probably cannot judge the quality of data since we are working in another do |
**USP 9: ENGAGE should provide a full API for machine-to-machine operation.**

**Importance:** 7 out of 7. Absolutely; automated machine-to-machine processing is required and allows users to have their existing environments interact with ENGAGE rather than going through a human interacting with the ENGAGE portal.

**Importance:** 7 out of 7. Providing an API for machine-to-machine operations is important for meeting many of the requirements of the ENGAGE-project. Many potential users have asked for such an API, stating its importance.

**Importance:** 6 out of 7. Extremely important for developers and other platforms, but may not be so important for other users such as researchers and social scientists.

**Importance:** 4 out of 7. It does not differentiate Engage from other platforms.

**USP 10: ENGAGE should target multiple nationalities.**

**Importance:** 6 out of 7. For ENGAGE to differentiate from national (usually monolingual) portals multilinguality is important. The CERIF metadata standard supports multilinguality and the portal user interface is being configured appropriately.

**Importance:** 6 out of 7. The ENGAGE-project should not just focus on one nationality. Many existing data.gov websites already target one nationality, and the ENGAGE-project should go beyond that by enabling comparative research among countries. Currently there is not platform enabling comparisons of data from many different countries. ENGAGE should also welcome users from many different countries, and as a result, multilinguality is an important requirement from potential ENGAGE-users.

**Importance:** 6 out of 7. It can increase the market for engage.

**What should be the ENGAGE dissemination steps towards sustainability?**

Sustainability options for ENGAGE include (a) the consortium forming an organisation; (b) a subset of the consortium members forming an organization; (c) the consortium proposing a follow-on project funded by the EC; (d) a subset of the consortium proposing a follow-on project funded by the EC; (e) the consortium licensing the ENGAGE IP to existing national open data portals

Possible sustainability plans could be to create a follow-up project from ENGAGE (ENGAGE 2) or to create a business model.

Promote success stories. Target key press channels.

As a collaborator on space that provides the richer metadata. A standard specification will be very helpful to the establishment of the ENGAGE platform.

Establish contacts with research communities, help them to find and curate data for their research.
so they can be improved. (a) and (b) could be commercial operations or government funded or a public-private partnership. The likely outcome is (b) as public-private partnership. Once the consortium has decided, dissemination is essential to build the market demand.

| Which other steps/domains/USP’s should ENGAGE target according to you? | ENGAGE should concentrate on demonstrating its ability (a) to be a unique European (possibly intercontinental) portal; (b) for the portal to be replicated with national localization; (c) to replace existing data.gov portals | ENGAGE should focus on sustainable strategies, tools and USP’s. ENGAGE should show to the community how it goes further that other data.gov websites. Users should be engaged to get good uptake of the ENGAGE platform. | International organisations working with data like http://www.transparency.org/ |
6.2.2.3. **Web Analytics and Metrics**

In this section is about the quantitative results and it is divided into several sub sections. Quantitative results will be presented for accomplishing the technical objectives (section 6.2.2.3.1) and achieving sustainable interest (section 6.2.2.3.2).

6.2.2.3.1. **Accomplishing Technical Objectives**

The accomplishment of technical objectives is measures with the results of the online user questionnaire, the usability test and the expert questionnaire.

**Online Users’ Questionnaire (1st layer variables)**

All variables and indicators in the first layer of the evaluation model have been developed in section 6.2. In parallel these indicators are used for the Online Users’ questionnaire. All questions in the online users’ questionnaire have the form of statements; the respondents who fill the questionnaire are asked to report on the extent of their agreement or disagreement with the statements, answering the question: “To which extend do you agree with the following statements?”. A five point Likert scale is used to measure agreement or disagreement with such a statement (1= Strongly Disagree, 2= Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). Finally, calculating the averages (mean) of all variables we conclude with a percentage for the perceived rating (by the user) on the platform capabilities. This total score is included in the Accomplishment of Technical Objectives measurement. The Online Users’ Questionnaire has been answered by 18 open data users referring to ENGAGE.
## 1. Ease of Use (P + U)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Questions</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efri</td>
<td>The platform provides a user friendly and easy to use environment.</td>
<td>3,375</td>
</tr>
<tr>
<td>Eeff</td>
<td>It was easy to learn how to use the platform.</td>
<td>3,5625</td>
</tr>
<tr>
<td>Eaes</td>
<td>The web pages look attractive.</td>
<td>2,875</td>
</tr>
<tr>
<td>Enos</td>
<td>It is easy to perform the tasks I want in a small number of steps.</td>
<td>3,0625</td>
</tr>
<tr>
<td>Elin</td>
<td>The platform allows me to work in my own language.</td>
<td>4,3125</td>
</tr>
<tr>
<td>Eper</td>
<td>The platform supports user account creation in order to personalize views and information shown</td>
<td>3,875</td>
</tr>
<tr>
<td>Esup</td>
<td>The platform provides high quality of documentation and online help.</td>
<td>2,8125</td>
</tr>
</tbody>
</table>

## 2. Performance (P + U)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Questions</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pavv</td>
<td>The platform is up and available without any interruptions.</td>
<td>3,375</td>
</tr>
<tr>
<td>Peff</td>
<td>Services and pages are loaded quickly.</td>
<td>3,0625</td>
</tr>
<tr>
<td>Prel</td>
<td>I did not notice any bugs while using the platform.</td>
<td>2,625</td>
</tr>
</tbody>
</table>

## 3. Data Provision Capabilities (U)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Questions</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUqu</td>
<td>The platform provides a large number of datasets</td>
<td>2,1875</td>
</tr>
<tr>
<td>PUqa</td>
<td>The platform provides dataset that are useful to me.</td>
<td>2,625</td>
</tr>
<tr>
<td>PUco</td>
<td>The platform provides me with complete data with all required fields and detail</td>
<td>2,75</td>
</tr>
<tr>
<td>PUac</td>
<td>The platform provides accurate data on which I can rely for my studies</td>
<td>2,625</td>
</tr>
<tr>
<td>PUms</td>
<td>There are datasets from many different thematic areas (economy, health, education, etc.)</td>
<td>2,6875</td>
</tr>
<tr>
<td>PUmn</td>
<td>There are datasets from many different countries</td>
<td>2,375</td>
</tr>
<tr>
<td>PUi</td>
<td>The platform provides sufficiently recent data</td>
<td>2,8125</td>
</tr>
</tbody>
</table>

## 4. Data Download (U)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Questions</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUUse</td>
<td>The platform provides strong dataset search capabilities using different criteria.</td>
<td>2,8125</td>
</tr>
<tr>
<td>DUca</td>
<td>The platform provides several different categorizations of the available datasets, which assists significantly in finding the datasets I need.</td>
<td>3,1875</td>
</tr>
<tr>
<td>DUac</td>
<td>The platform enabled me to download datasets easily and efficiently.</td>
<td>3,4375</td>
</tr>
<tr>
<td>DUsf</td>
<td>The datasets are in appropriate file/data formats that I can easily use for covering my needs.</td>
<td>2,9375</td>
</tr>
<tr>
<td>DUme</td>
<td>The datasets have also appropriate and sufficient metadata, which allowed me to understand these data and also how and for what purpose they were collected.</td>
<td>2,9375</td>
</tr>
<tr>
<td>DUap</td>
<td>The platform provides strong API for downloading datasets (data+metadata)</td>
<td>2,75</td>
</tr>
</tbody>
</table>

## 5. Data Upload (P)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Questions</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPac</td>
<td>The platform enabled me to upload datasets easily and efficiently.</td>
<td>3,4375</td>
</tr>
<tr>
<td>UPme</td>
<td>The platform has a good metadata model applicable for all its datasets, that provides full understanding of the</td>
<td>3,1875</td>
</tr>
</tbody>
</table>
5.3 Extra Tools provided 1
5.4 Automation of data entry
5.5 Extra Tools provided 2
5.6 Automated data provision

6. Data Curation Capabilities (P)
6.1 Data Enrichment
6.2 Data Cleansing
6.3 Linking
6.4 Visualization

7. Users’ Feedback Capabilities
7.1 Feedback U1
7.2 Feedback U2

8. Users’ Processing Capabilities
8.1 Visualization

9. Providers’ Feedback Collection Capabilities
9.1 Feedback P

<table>
<thead>
<tr>
<th>Data Curation Capabilities (P)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Extra Tools provided 1</td>
<td>UPe1</td>
<td>The platform enabled me to prepare and add the metadata for the datasets I uploaded easily and efficiently.</td>
<td>2,8666</td>
</tr>
<tr>
<td>Automation of data entry</td>
<td>UPad</td>
<td>The platform provides good capabilities for the automated creation of metadata.</td>
<td>2,8125</td>
</tr>
<tr>
<td>Extra Tools provided 2</td>
<td>UPe2</td>
<td>The platform provides good capabilities for converting datasets’ initial metadata in the metadata model of the platform easily and efficiently.</td>
<td>2,875</td>
</tr>
<tr>
<td>Automated data provision</td>
<td>UPap</td>
<td>The platform provides API for uploading datasets (data+metadata)</td>
<td>3,125</td>
</tr>
<tr>
<td>Data Enrichment</td>
<td>CPen</td>
<td>The platform provides good capabilities for data enrichment (i.e. adding new elements - fields)</td>
<td>3,0625</td>
</tr>
<tr>
<td>Data Cleansing</td>
<td>CPCl</td>
<td>The platform provides good capabilities for data cleansing (i.e. detecting and correcting ubiquitous in a dataset)</td>
<td>2,8125</td>
</tr>
<tr>
<td>Linking</td>
<td>CPIi</td>
<td>The platform provides good capabilities for linking datasets.</td>
<td>3,0625</td>
</tr>
<tr>
<td>Visualization</td>
<td>CPvi</td>
<td>The platform provides me good capabilities for visualization of datasets before uploading.</td>
<td>2,6875</td>
</tr>
<tr>
<td>Feedback U1</td>
<td>FUf1</td>
<td>Platform provides good capabilities for giving feedback on the datasets I download, e.g. for rating datasets, for entering textual comments on them.</td>
<td>3,1875</td>
</tr>
<tr>
<td>Feedback U2</td>
<td>FUf2</td>
<td>Platform provides good capabilities for reading available feedback of other users of datasets I am interested in, e.g. ratings, comments.</td>
<td>3,4375</td>
</tr>
<tr>
<td>Visualization</td>
<td>VUvi</td>
<td>The platform provides me good capabilities for visualization of datasets.</td>
<td>3,0625</td>
</tr>
<tr>
<td>Feedback P</td>
<td>FPfe</td>
<td>The platform allows me to collect user ratings and comments on the datasets I publish, and also needs for further datasets provision.</td>
<td>3,375</td>
</tr>
</tbody>
</table>

Technical Objective: Platforms Capabilities Total Score

Usability Test

With the usability test we measured the usability of the ENGAGE platform. All these questions have the form of statements; the stakeholders who run the usability test are asked to enter the extent of the easiness or difficulty they perform a task, answering the question: “How easy was it for you to perform the specific tasks?”. A five point Likert scale is used to measure easiness with (i.e. positive or negative response to) such a statement (1= Very Difficult, 2= Difficult, 3=Neutral, 4=Easy, 5=Very Easy). Finally,
calculating the averages of the all tasks performed we measure the perceived usability of the platform providing the final percentage. This total score is included in the Accomplishment of Technical Objectives measurement. The Usability Test has been answered by 22 open data users referring to ENGAGE.

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Servant</td>
<td>1</td>
<td>4%</td>
</tr>
<tr>
<td>Researcher</td>
<td>9</td>
<td>39%</td>
</tr>
<tr>
<td>Business Employee</td>
<td>2</td>
<td>9%</td>
</tr>
<tr>
<td>Citizen</td>
<td>8</td>
<td>35%</td>
</tr>
<tr>
<td>Policy Maker</td>
<td>2</td>
<td>9%</td>
</tr>
</tbody>
</table>
Scenario 1: How easy was it for you to perform the specific tasks?

Task 1.1 Go to www.engagedata.eu 4.53
Task 1.2 Select the language you prefer on the top right side of the home page. 4.63
Task 1.3 Register yourself (if you have not done this already) and sign in (you can use your Twitter, LinkedIn, Google or Facebook account to register or sign in). 4.47
Task 1.4 Use the search technique to search for a dataset that is interesting to you. For instance, think of a type of data (e.g. crime, education, housing, transport) and fill the search criteria. 3.05
Task 1.5 View or download (one of) the dataset(s) that you found. 3.58
Task 1.6 Think of how you could reuse the downloaded dataset. Reuse the dataset that you viewed or downloaded. For instance, visualize the dataset, analyze the dataset, remove duplicated records, change the format of the dataset, link the dataset to another dataset or post a message in the discussion area below the dataset. 2.53
Task 1.7 Have a look at the possibilities of and required information for adding extended datasets on the ENGAGE website. 3.26
Task 1.8 Download a dataset from another open data platform (e.g. data.overheid.nl, data.gov.uk or data.gov) and upload this dataset to the ENGAGE website. Click on ‘add new dataset’ in the tab ‘datasets’ and go through the five steps that are indicated (fill in the appropriate metadata fields). 2.74
Task 1.9 Extend a dataset (e.g. by cleansing or filtering it) and upload the extended dataset to the ENGAGE platform. Select the original dataset, click on ‘extend’ in the dataset view and go through the five steps that are indicated (fill in the appropriate metadata fields). 2.71
Task 1.10 Express your data needs and let the community work for you. Click on 'request data' tab -> 'Place a new request' and fill in the appropriate fields. 3.43

Scenario 2: How easy was it for you to perform the specific tasks?

Task 2.1 Go to the home page (engagedata.eu). 4.42
Task 2.2 Click on the tab ‘open data sites’ on top of the home page. 4.42
Task 2.3 Click on your country, if available, and/or another country that might be interesting to you. You can also select several countries if you want to compare them. 4.05
Task 2.4 Have a look at the tabs that are provided on the top right side of the web page. Click on ‘data 4
sources types', 'licenses', 'language UI', 'data format' and 'provision'.

**Scenario 3: How easy was it for you to perform the specific tasks?**

| Task 3.1 | Go to the home page (engagedata.eu). | 4.47 |
| Task 3.2 | Click on Wiki/API on top of the home page. Have a look at the possibilities of this page. Click on the Wiki Categories: 'Data Curation Methods', 'ENGAGE API Reference' and 'ENGAGE Documentation' (below the title 'Frontpage'). | 4.68 |
| Task 3.3 |  | 4.10 |

**Technical Objective: Ease of Use Total Score**

| Total Score | 3.83 |
| or | 76.51% |
Experts' Questionnaire

With the Expert’s Questionnaire we measure the identified Unique Selling Points (USPs). We asked the opinion of the ENGAGE consortium experts responsible for the evaluation of the platform (AEGEAN, TUDELFT, INTRASOFT, FOKUS, NTUA – 5 votes on each question) and the opinion of external experts in the field of open data (12 votes on each question). A seven point Likert scale was used to measure the importance of ENGAGE USPs from 1=Very Unimportant to 7=Very Important. This total score is included in the Accomplishment of Technical Objectives measurement. In this study, internal and external experts have participated. The Experts' Questionnaire has been answered by 17 open data users in total, referring to ENGAGE 2.0.

Experts' Questionnaire

To which extend each USP is important (or unimportant) in order to promote and reach sustainable interest for ENGAGE?

<table>
<thead>
<tr>
<th>USP</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGAGE should be a marketplace for open data and collaboration.</td>
<td>5,41</td>
</tr>
<tr>
<td>ENGAGE should be an open public data reputation management system.</td>
<td>5,39</td>
</tr>
<tr>
<td>ENGAGE should put emphasis on rich metadata.</td>
<td>5,29</td>
</tr>
<tr>
<td>ENGAGE should provide a rich collection of data curation tools.</td>
<td>5</td>
</tr>
<tr>
<td>ENGAGE should provide a rich collection of data visualisation tools.</td>
<td>6,11</td>
</tr>
<tr>
<td>ENGAGE should develop cases of using the platform.</td>
<td>5,89</td>
</tr>
<tr>
<td>ENGAGE should focus on a standardization proposal for annotating open data sets for scientific usage.</td>
<td>4,83</td>
</tr>
<tr>
<td>ENGAGE should provide a complete data repository.</td>
<td>5,06</td>
</tr>
<tr>
<td>ENGAGE should provide a full API for machine-to-machine operation.</td>
<td>5,94</td>
</tr>
<tr>
<td>ENGAGE should target multiple-nationalities.</td>
<td>6,06</td>
</tr>
</tbody>
</table>

Technical Objective: USPs for ENGAGE Total Score 5,42 or 78,54 %

6.2.2.3.2. Achieving Sustainable Interest

In this section we report on the indicator results with regard to achieving sustainable interest.

Web Analytics

For the second year of ENGAGE we pursued some targets for our presence on the internet. These indicators are additional to the web analytics that have been identified
by the project management and will help us to understand the intention of (potential) open data users to use the ENGAGE platform.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Target</th>
<th>Achieved</th>
</tr>
</thead>
<tbody>
<tr>
<td># of unique visitors</td>
<td>1500</td>
<td>1526</td>
</tr>
<tr>
<td># of persons connected through SM</td>
<td>200</td>
<td>367</td>
</tr>
<tr>
<td># of Registered Users</td>
<td>80</td>
<td>96</td>
</tr>
<tr>
<td># of rated datasets by users</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td># of people addressed via workshops</td>
<td>300</td>
<td>320</td>
</tr>
<tr>
<td># of people addressed via evaluation procedures</td>
<td>200</td>
<td>307</td>
</tr>
<tr>
<td># of connected projects (MoUs and crosslinking, common workshops)</td>
<td>16</td>
<td>18 (5 MoUs)</td>
</tr>
</tbody>
</table>

**Achieving Sustainable Interest: Internet Presence Total Score**  
90 %

**Online Users’ Questionnaire**

The indicators for the second and third layer of the evaluation model developed have also been developed in D6.6.1. In parallel these indicators are used to structure the Online Users’ questionnaire. Again these questions have the form of statements. A five point Likert scale was used to measure agreement or disagreement with (i.e. positive or negative response to) such a statement (1= Strongly Disagree, 2= Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). Finally, calculating the averages (mean) of all variables we conclude in a percentage for the perceived general satisfaction (by the user) about the platform. This total score is included in the Achieving Sustainable Interest measurement.

The Online Users’ Questionnaire has been answered by 18 open data users in total, referring to ENGAGE.

**10. General Satisfaction**

| Mean. |  
|-------|---|
| AoU1  | 3,0625 | I think that using this platform enables me to do better research/inquiry and accomplish it more quickly  
| AoU2  | 2,9375 | This platform allows me to draw interesting conclusions on government activity |
This platform enables me to create successful added-value electronic services.

My initial motivations for selecting ENGAGE platform have proved to be truth.

The platform enables me to open and widely publish datasets I possess with low effort and cost.

I am in general highly satisfied with this platform.

I would like to use this platform again.

I’ll recommend this platform to my colleagues.

<table>
<thead>
<tr>
<th>Achieving Sustainable Interest</th>
<th>General Satisfaction</th>
<th>61,40%</th>
</tr>
</thead>
</table>

Usability Test

A seven point Likert scale is used to measure the usefulness of ENGAGE capabilities (Tasks in the Usability Test) from 1=Very Unuseful to 7=Very Useful. Respondents were asked to determine to which extent these activities could be useful for them. This total score is included in the Achieving Sustainable Interest measurement.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Tasks</th>
<th>Mean.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Searching, downloading, extending/visualizing/curation/linking and uploading interesting datasets</td>
<td>5,578947</td>
</tr>
<tr>
<td>2</td>
<td>Getting information about other open data websites and comparing them via the ENGAGE website</td>
<td>5,157895</td>
</tr>
<tr>
<td>3</td>
<td>Getting information about manuals, API's and tutorials (training)</td>
<td>5,684211</td>
</tr>
</tbody>
</table>

Achieving Sustainable Interest: Usefulness of platform

5,47

or 78,19 %

6.2.2.3.3. General Impact

- Having in mind the stages for General Impact measurement that were described in D6.6.1: The platform could neither attract sustainable interest nor accomplish its technical objectives → NO SUCCESS/IMPACT achieved.
- The platform could not attract sustainable interest but it met its technical goals → SMALL SUCCESS/IMPACT achieved.
The platform could not meet its technical objectives but it attracted sustainable interest → MEDIUM SUCCESSS/IMPACT achieved.

The technical objectives have been accomplished and the platform has attracted sustainable interest of the target group → HIGH SUCCESS/IMPACT achieved.

The analysis and measurement in shows that ENGAGE has accomplished its technical objectives by 71.85% and has managed to achieve sustainable interest by 76.53%.

<table>
<thead>
<tr>
<th>Accomplishing Technical Objectives</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGAGE USPs – Experts’ Questionnaire</td>
<td>78.54%</td>
</tr>
<tr>
<td>Usability of the ENGAGE platform – Usability Test</td>
<td>76.51%</td>
</tr>
<tr>
<td>Platform Capabilities – Online Users’ Questionnaire</td>
<td>60.5%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>71.85%</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Achieving Sustainable Interest</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Users’ Interest – Web Analytics</td>
<td>90%</td>
</tr>
<tr>
<td>Users’ General Satisfaction – Online Users’ Questionnaire</td>
<td>61.4%</td>
</tr>
<tr>
<td>Users’ Perceived Usefulness – Usability Test</td>
<td>78.19%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>76.53%</strong></td>
</tr>
</tbody>
</table>

The qualitative analysis, by our understanding, has minor deviation from the quantitative one. Consequently, the ENGAGE Platform has accomplished its technical objectives and has attracted sustainable interest achieving HIGH IMPACT. Of course more need to be done in order to increase these percentages. Insights from the whole evaluation procedure and future steps are provided in the conclusions section.

**The ENGAGE Platform:**

- The platform could neither attract sustainable interest nor accomplish its technical objectives → NO SUCCESS/IMPACT achieved.
- The platform could not attract sustainable interest but it met its technical goals → SMALL SUCCESS/IMPACT achieved.
- The platform could not meet its technical objectives but it attracted sustainable interest → MEDIUM SUCCESSS/IMPACT achieved.
- The technical objectives have been accomplished and the platform has attracted sustainable interest of the target group → HIGH SUCCESS/IMPACT achieved.
6.2.3. Summary of results

In this section we report on the application of the proposed evaluation framework and procedure to the first version of the open public sector data e-infrastructure developed in the ENGAGE project. In total, 21 students from Delft University of Technology and 33 students from the University of Aegean performed the student usability test by executing a series of tasks that require most of the available infrastructure functionalities, filled out the online user questionnaire and participated in a qualitative discussion. In addition, students of Delft University of Technology and the University of Aegean wrote reports about the usability of the ENGAGE e-infrastructure for conducting open data download, process and upload scenarios. Five project members conducted the internal experts SWOT analysis. The results of these evaluation activities are as follows.

The application of the applied evaluation framework and procedure showed that the current ENGAGE e-infrastructure (ENGAGE 1.0) provides a good basic e-infrastructure to start an open data project. The current e-infrastructure enables searching, downloading, analyzing, uploading and rating data and also the manual linkage of data. Moreover, the ENGAGE e-infrastructure provides a comprehensive overview of which open data e-infrastructures currently exist in Europe. In addition, the ENGAGE infrastructure enables the usage of Application Programming Interface (API) for searching data, as well as, of a Wiki including tutorials. This means that the ENGAGE infrastructure manage to accomplish its technical objectives till the pilot evaluation.

Nevertheless, the e-infrastructure could be improved in many ways. The application of the evaluation model and procedure showed that future versions of the ENGAGE e-infrastructure should mainly focus on 1) difficulties in searching, need for more information about datasets, multilinguality, capabilities for combining datasets, and also for doing more complex analysis and performance, 2) more information about the quality of the data, for instance by providing a comprehensive rating systems for the data and 3) more information about which data can be linked, in order to meet users’ expectations for using the platform. This means that this first version of the ENGAGE OGD e-infrastructure did not achieve to attract sustainable interest.

In general, the first application of the proposed evaluation framework showed that meaningful insights can be obtained by applying the proposed evaluation framework and infrastructure to such an advanced open public data e-infrastructure 2.0. The
insights concern both managerial (categorizing priorities, decision making) and technical aspects of the e-infrastructure. The evaluation results appeared to be useful for defining priorities for improvements and enrichments required for developing the future versions of the ENGAGE e-infrastructure. They provide very detailed information about the features of the e-infrastructure and information about possibilities for improvement. One advantage of the evaluation framework was that the evaluation results are based on users experience with the actual ENGAGE e-infrastructure, instead of feedback on mock-ups of the e-infrastructure and/or presentations about the e-infrastructure. In addition, the evaluation results clearly show that user experience is a critical success factor for the community uptake of the e-infrastructure (navigation, search, user interface, services and features for non-Information Technology savvy users). Furthermore, some ideas were mentioned to improve the e-infrastructure that the consortium had not thought of. For instance, the idea of a ‘shopping cart’, where the user can add different datasets to make the linking process easier, could be considered for future versions of the ENGAGE e-infrastructure. Finally, several other inventive recommendations and solutions were provided by the potential users of the infrastructure.
6.3. A value model for ODG e-Infrastructures

A methodology for evaluating this advanced second generation of OGD infrastructures was developed based on:

i) The approaches and frameworks from previous relevant IS research outlined in the previous section, concerning: IS evaluation (including in the methodology both efficiency and effectiveness measures), IS acceptance (including measures of ease of use, usefulness and future intentions), IS success (adopting a layered evaluation approach, and including measures of both information and system quality, and also of user satisfaction and individual impact) and e-services evaluation (including measures of both the quality of the capabilities offered to the users, and of the support provided to them for achieving their OGD related objectives).

ii) The results of the analysis of potential users' requirements conducted as part of the abovementioned ENGAGE project (which, as described in more detail in (Charalabidis et al., 2011; Zuiderwick et al., 2013), include data search, provision and download capabilities, data processing capabilities, data upload capabilities, and also users – providers communication capabilities).

iii) The high level technological aspects proposed in the methodologies for country and government agency level OGD initiatives' evaluation proposed in (Ubaldi, 2013) and Solar et al., (2013) respectively (such as data completeness, quality, quantity, format and metadata, search capabilities, users-providers communication capabilities, users satisfaction, platform availability).

Our methodology includes the definition of a value model for these advanced second generation OGD infrastructures, and also an algorithm for estimating this value model based on users' evaluation ratings, adopting the approach proposed by Pazalos, Loukis, & Nikolopoulos (2012) and Loukis, Pazalos, & Salagara, (2012), in which they are described next in 3.1 and 3.2 respectively.

6.3.1. Value Model Definition

The value model consists of the main dimensions of the value that these OGD infrastructures generate (from section 5.4.2), and the relations among them. We can see its definition in Figure 6.14.
We remark that these value dimensions are organized in three value layers, adopting the structure proposed by (Pazalos et al., 2012; Loukis et al., 2012), which correspond to efficiency (value associated with the capabilities it offers to the users), effectiveness (value associated with the support of users for achieving their user-level and provider-level objectives) and future behavior (value associated with users’ future behavior) respectively.

The first efficiency layer includes eight value dimensions in total. Three of them concern the user-level capabilities offered by the OGD infrastructure: data provision capabilities (based on the ‘information quality’ proposed by the the IS success model of DeLone and McLean (1992; 2003), and also on (Ubaldi, 2013) and (Charalabidis et al., 2011; Zuiderwick et al., 2013), data search and download capabilities (based on [1] and [7-9]) and user-level feedback capabilities (also based on (Ubaldi, 2013) and (Charalabidis et al. 2011; Zuiderwick et al., 2013)). These value dimensions are expected to affect the ‘support for achieving user-level objectives’ value dimension of the second layer (we can see the corresponding relations in the value model of Fig.1). The next three value dimensions of the first layer are ease of use (based on the TAM (Davis, 1989)), performance (based on the ‘system quality’ proposed by the IS success model of DeLone and McLean (1992; 2003), and data processing capabilities users (based on (Ubaldi, 2013) and (Charalabidis et al. 2011; Zuiderwick et al., 2013)). These value dimensions are expected to affect both the ‘support for achieving user-level objectives’ value dimension of the second layer and the ‘future behavior’ value dimension of the third layer.
objectives’ and the ‘support for achieving provider-level objectives’ value dimensions of the second layer (so we can see the corresponding relations of them with both these second layer value dimensions in Fig.1). The final two value dimensions of the first layer concern the provider-level capabilities offered by the OGD infrastructure: data upload capabilities (based on (Ubaldi, 2013) and (Charalabidis et al., 2011; Zuiderwick et al., 2013)) and provider-level feedback capabilities (based on (Ubaldi, 2013) and (Charalabidis et al., 2011; Zuiderwick et al., 2013; Solar et al., 2013)). These two value dimensions are expected to affect the ‘support for achieving provider-level objectives’ value dimension of the second layer (we can see the corresponding relations in the value model of Fig.1). The second effectiveness layer includes the abovementioned two value dimensions concerning the support provided by the OGD infrastructure for achieving user-level and provider-level objectives respectively. Finally, the third layer includes one value dimension associated with users’ future behavior (based on the the TAM (Davis, 1989)).

It should be noted that the value dimensions of the first efficiency layer are independent variables, which are under the direct control of the infrastructure developer, who can take direct actions for improving them if necessary. In contrast, the value dimensions of the other two layers (effectiveness and future behavior ones) are not under the direct control of the infrastructure developer, and are dependent on the first level ones.

The above eleven value dimensions were further elaborated, and for each of them a number of individual value measures were defined (again based on the main foundations of our methodology i to iii mentioned in the beginning of this section). Each of these value measures was then converted to a question to be included in a questionnaire to be distributed to users of the infrastructure (who act both as data users and providers). All these questions have the form of statements, and the above users are asked to enter the extent of their agreement or disagreement with them, answering the question: “To which extend do you agree with the following statements?”. A five point Likert scale is used to measure agreement or disagreement with (i.e. positive or negative response to) such a statement (1=Strongly Disagree, 2= Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). In Table 1 we can see the questions that correspond to the value measures of each value dimension.

Table 6-3: Questions for Value Measures
<table>
<thead>
<tr>
<th><strong>Data Provision Capabilities (DPV)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DPV1</td>
</tr>
<tr>
<td>DPV2</td>
</tr>
<tr>
<td>DPV3</td>
</tr>
<tr>
<td>DPR4</td>
</tr>
<tr>
<td>DPV5</td>
</tr>
<tr>
<td>DPV6</td>
</tr>
<tr>
<td>DPV7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Data Search and Download Capabilities (DSD)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>DSD1</td>
</tr>
<tr>
<td>DSD2</td>
</tr>
<tr>
<td>DSD3</td>
</tr>
<tr>
<td>DSD4</td>
</tr>
<tr>
<td>DSD5</td>
</tr>
<tr>
<td>DSD6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>User-level Feedback Capabilities (UFB)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>UFB1</td>
</tr>
<tr>
<td>UFB2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Ease of Use (EOU)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>EOU1</td>
</tr>
<tr>
<td>EOU2</td>
</tr>
<tr>
<td>EOU3</td>
</tr>
<tr>
<td>EOU4</td>
</tr>
<tr>
<td>EOU5</td>
</tr>
<tr>
<td>EOU6</td>
</tr>
<tr>
<td>EOU7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Performance (PER)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>PER1</td>
</tr>
</tbody>
</table>
PER2 | Services and pages are loaded quickly.
PER3 | I did not realize any bugs while using the platform.

**Data Processing Capabilities (DPR)**

DPR1 | The platform provides good capabilities for data enrichment (i.e. adding new elements - fields)
DPR2 | The platform provides good capabilities for data cleansing (i.e. detecting and correcting ubiquitous in a dataset)
DPR3 | The platform provides good capabilities for linking datasets.
DPR4 | The platform provides good capabilities for visualization of datasets

**Data Upload Capabilities (DUP)**

DUP1 | The platform enabled me to upload datasets easily and efficiently.
DUP2 | The platform enabled me to prepare and add the metadata for the datasets I uploaded easily and efficiently.
DUP3 | The platform provides good capabilities for the automated creation of metadata.
DUP4 | The platform provides good capabilities for converting datasets’ initial metadata in the metadata model of the platform easily and efficiently.
DUP5 | The platform provides strong API for uploading datasets (data and metadata)

**Provider-level Feedback Capabilities (PFB)**

PFB1 | The platform allows me to collect user ratings and comments on the datasets I publish.

**Support for Achieving User-level Objectives (SUO)**

SUO1 | I think that using this platform enables me to do better research/inquiry and accomplish it more quickly
SUO2 | This platform allows drawing interesting conclusions on past government activity
SUO3 | This platform allows creating successful added-value electronic services

**Support for Achieving Provider-level Objectives (SPO)**

SPO1 | The platform enables opening and widely publishing datasets with low effort and cost.

**Future Behaviour (FBE)**

FBE1 | I would like to use this platform again.
FBE2 | I’ll recommend this platform colleagues.

The above value model can be adapted based on the capabilities offered by a particular second generation OGD infrastructure to be evaluated (e.g. additional value dimensions can be added corresponding to additional capabilities it offers).
Furthermore, the above approach can be used for the evaluation of first generation OGD infrastructures as well, which are characterised by clear distinction between data providers and data users, by defining one value model for the former and one value model for the latter, and based on them formulate the corresponding two questionnaires. Each of these value models will be estimated separately (based on evaluation data from the corresponding group) using the algorithm described next in section.

6.3.2. Value Model Estimation Algorithm

The users’ evaluation data collected through the above questionnaire will be processed using the algorithm shown in Fig. 7.2, in order to estimate the value model of the OGD infrastructure and identify improvement priorities.

![Value Model Estimation Algorithm Diagram]

It consists of the following seven steps:

1. Initially for each value dimension we examine the internal consistency of its value measures by calculating the Cronbach Alpha of the variables corresponding to its value measures (Boudreau & Gefen, 2004). This coefficient quantifies to what extent a set of variables measure different aspects of the same single uni-dimensional construct, and is calculated as:

   \[
   \text{Alpha} = \frac{k}{(k-1)} [1 - \frac{(s^2_i)}{s^2_{\text{sum}}}] 
   \]
where the $s^2_i$ ($i = 1, 2, ..., k$) denote the variances of the $k$ individual variables, while the $s^2_{\text{sum}}$ denotes the variance of the sum of these variables. A widely accepted and used practical ‘rule of thumb’ is that values of Cronbach Alpha exceeding 0.7 indicate ‘acceptable’ levels of internal consistency of the variables (Boudreau & Gefen, 2004). Therefore if for a value dimension its calculated value of Cronbach Alpha exceeds 0.7, we can conclude that all its measures have acceptable internal consistency; if this does not happen, we can conclude that some of the measures are not sufficiently related to this value dimension (they can be detected if for each of the individual variables is calculated the Cronbach Alpha without it, which is a standard calculation offered by all statistical packages), so they must be removed and not taken into account, or probably that this dimension should be split into two or more sub-dimensions.

2. For each value dimension an aggregate variable is calculated as the average of its individual measures' variables.

3. Average ratings are calculated for all value measures and dimensions (using for the latter the aggregate variables calculated in the previous step); this allows us to identify ‘strengths’ and ‘weaknesses’ of the OGD infrastructure.

4. For each aggregate variable of the second and third layer, assessing one of the ‘dependent’ e-service value dimensions, we estimate a regression having it as dependent variable, and having as independent variables all the aggregate variables of the previous layers, in order to estimate to what extent this value dimension is affected by value dimensions of previous layers; this is quantified by the $R^2$ coefficient of the regression (Greene, 2011). If we find that all value dimensions of the second and third layer are affected to a large extent by the value dimensions of the previous layers (e.g. having $R^2 > 0.50$), then we can conclude that this value model is characterized by coherence among its layers, so we can proceed to the following stages. On the contrary, if some value dimensions of second and third layer are affected only to a small extent by the value dimensions of the previous layers, this indicates that probably some important value dimensions have been omitted in the previous layers, so we have to redefine the value model of the OGD infrastructure.

5. For each value dimension of the first level we calculate its impact on the higher level value dimensions (of the second and the third layers) (using again the aggregate variables calculated in step 2). For this purpose we can use the corresponding standardised coefficients of the regressions of the above step 4. However, according to
econometric literature (Greene, 2011), if there are high levels of correlation between the independent variables of a regression, then the estimated regression coefficients are not reliable measures of the impacts of the independent variables on the dependent variable (multi-collinearity problem). For this reason we decided to use correlations instead; so as measure of the impact of a first layer value dimension on a higher layer value dimension has been used the correlation coefficient between them. Furthermore we calculated the correlations of all first level value measures with all second and third layers’ value dimensions and measure, as measures of their impact on higher level value generation.

6. By combining the average ratings calculated in step 2 with the correlations calculated in step 3 we can construct one value model of the OGD infrastructure at the level of value dimensions, and also a more detailed one at the level of value measures. These models enable a deeper understanding of the value generation mechanism of the OGD infrastructure.

7. Finally the value dimensions and the value measures of the first layer, which are the only ‘independent variables’ within the control of the OGD infrastructure developer, are classified, based on their average ratings by users and their impacts on the value dimensions of the second and the third level, into four groups: low rating – high impact, low rating – low impact, high rating – high impact and high rating – low impact. The highest priority should be given to the improvement of the value dimensions and individual value measures of the first group, which receive low ratings and at the same time have a high impact on the generation of higher level value; so it is on them that we should focus our scarce human and financial resources.

6.3.3. Model Validation and Results

The proposed methodology has been applied for the evaluation of the first version of an advanced second generation OGD infrastructure under development in the abovementioned project ENGAGE. The evaluation questionnaire shown in Table 1 was filled by 42 postgraduate students of the University of the Aegean (Greece) and Delft University of Technology (The Netherlands (both partners of this project). These students previously had been trained on the capabilities of this infrastructure, and then used it for implementing a representative scenario (including both data user and data provider tasks).
Initially for each value dimension the Cronbach Alpha coefficient of the variables corresponding to its value measures was calculated (step 1), and the results are shown in Table 6.4.

**Table 6-4: Cronbach Alpha of Value Dimensions**

<table>
<thead>
<tr>
<th>Value Dimension</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Provision Capabilities (DPV)</td>
<td>0.834</td>
</tr>
<tr>
<td>Data Search and Download Capabilities (DSD)</td>
<td>0.805</td>
</tr>
<tr>
<td>User-level Feedback Capabilities (UFB)</td>
<td>0.770</td>
</tr>
<tr>
<td>Ease of Use (EOU)</td>
<td>0.716</td>
</tr>
<tr>
<td>Performance (PER)</td>
<td>0.719</td>
</tr>
<tr>
<td>Data Processing Capabilities (DPR)</td>
<td>0.811</td>
</tr>
<tr>
<td>Data Upload Capabilities (DUP)</td>
<td>0.858</td>
</tr>
<tr>
<td>Provider-level Feedback Capabilities (PFB)</td>
<td>-</td>
</tr>
<tr>
<td>Support for Achieving User-level Objectives (SUO)</td>
<td>0.843</td>
</tr>
<tr>
<td>Support for Achieving Provider-level Objectives (SPO)</td>
<td>-</td>
</tr>
<tr>
<td>Future Behaviour (FBE)</td>
<td>0.876</td>
</tr>
</tbody>
</table>

We remark that for all value dimensions Cronbach Alpha exceeds the ‘acceptable’ internal consistency level of 0.7. This indicates that for all our value dimensions their selected value measures are sufficiently relevant and measure different aspects of the same uni-dimensional construct. This allowed us to proceed to the calculation for each value dimension of an aggregate variable, which is equal to the average of the individual variables corresponding to its measures (step 2).

Next for all value measures and dimensions average rating over all respondent students were calculated (step 3), and the results are shown in the second column of Table 6.5 (results for value dimensions are shown in bold). We remark that according to the respondents the strongest points of this OGD are the provider-level feedback capabilities (for collecting ratings and comments on the datasets they publish from their users), its ease-of-use, and its data processing capabilities, all perceived between
moderate and good (average ratings 3.44, 3.35 and 3.27 respectively). Its weakest points is its performance (with respect to its availability, response time and bugs), which is perceived as problematic (average rating 2.15). The remaining four first layer value dimensions (i.e. data provision, data search and download, capabilities for user-level feedback, and data upload) are regarded as moderate (average ratings 3.03, 3.03, 2.97 and 2.93 respectively).

As a next step we examined to what extent the value dimensions of the second and third layer are affected by the ones of the first layer (step 4). For this purpose initially we estimated two regression models having as dependent variables the two value dimensions of the second layer SUO and SPO, and as independent variables the eight value dimensions of the first layer. Also, we estimated one regression model having as dependent variable the value dimension of the third layer FBE and as independent variables the two value dimensions of the second layer, and finally another similar regression model having as additional independent variables the eight value dimensions of the first layer. In Table 6.6 are shown the $R^2$ coefficients of these models.

<table>
<thead>
<tr>
<th>Measure/Dimension</th>
<th>Average ratings</th>
<th>Correl SUO</th>
<th>Correl SPO</th>
<th>Correl FBE</th>
<th>Average Correl.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DPV</td>
<td>3.03</td>
<td>0.639</td>
<td>0</td>
<td>0.511</td>
<td>0.383</td>
</tr>
<tr>
<td>DPV1</td>
<td>2.68</td>
<td>0.502</td>
<td>0</td>
<td>0.378</td>
<td>0.293</td>
</tr>
<tr>
<td>DPV2</td>
<td>3.00</td>
<td>0.537</td>
<td>0</td>
<td>0.426</td>
<td>0.321</td>
</tr>
<tr>
<td>DPV3</td>
<td>2.51</td>
<td>0.593</td>
<td>0</td>
<td>0.606</td>
<td>0.400</td>
</tr>
<tr>
<td>DPR4</td>
<td>3.02</td>
<td>0.544</td>
<td>0</td>
<td>0.375</td>
<td>0.306</td>
</tr>
<tr>
<td>DPV5</td>
<td>3.71</td>
<td>0.329</td>
<td>0</td>
<td>0.159</td>
<td>0.163</td>
</tr>
<tr>
<td>DPV6</td>
<td>3.37</td>
<td>0.148</td>
<td>0</td>
<td>0.226</td>
<td>0.125</td>
</tr>
<tr>
<td>DPV7</td>
<td>2.95</td>
<td>0.574</td>
<td>0</td>
<td>0.418</td>
<td>0.331</td>
</tr>
<tr>
<td>DSD</td>
<td>3.03</td>
<td>0.760</td>
<td>0</td>
<td>0.747</td>
<td>0.502</td>
</tr>
<tr>
<td>DSD1</td>
<td>2.68</td>
<td>0.516</td>
<td>0</td>
<td>0.520</td>
<td>0.345</td>
</tr>
<tr>
<td>DSD2</td>
<td>3.24</td>
<td>0.422</td>
<td>0</td>
<td>0.386</td>
<td>0.269</td>
</tr>
<tr>
<td>DSD3</td>
<td>3.24</td>
<td>0.598</td>
<td>0</td>
<td>0.662</td>
<td>0.420</td>
</tr>
<tr>
<td>DSD4</td>
<td>3.10</td>
<td>0.576</td>
<td>0</td>
<td>0.603</td>
<td>0.393</td>
</tr>
<tr>
<td>DSD5</td>
<td>2.90</td>
<td>0.589</td>
<td>0</td>
<td>0.549</td>
<td>0.379</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>DSD6</td>
<td>3.05</td>
<td>0.515</td>
<td>0</td>
<td>0.425</td>
<td>0.313</td>
</tr>
<tr>
<td>UFB</td>
<td>2.97</td>
<td>0.651</td>
<td>0</td>
<td>0.410</td>
<td>0.354</td>
</tr>
<tr>
<td>UFB1</td>
<td>2.90</td>
<td>0.622</td>
<td>0</td>
<td>0.284</td>
<td>0.302</td>
</tr>
<tr>
<td>UFB2</td>
<td>3.05</td>
<td>0.624</td>
<td>0</td>
<td>0.442</td>
<td>0.355</td>
</tr>
<tr>
<td>EOU</td>
<td>3.35</td>
<td>0.730</td>
<td>0.479</td>
<td>0.448</td>
<td>0.552</td>
</tr>
<tr>
<td>EOU1</td>
<td>3.39</td>
<td>0.684</td>
<td>0.362</td>
<td>0.430</td>
<td>0.492</td>
</tr>
<tr>
<td>EOU2</td>
<td>3.80</td>
<td>0.539</td>
<td>0.359</td>
<td>0.295</td>
<td>0.398</td>
</tr>
<tr>
<td>EOU3</td>
<td>3.00</td>
<td>0.515</td>
<td>0.311</td>
<td>0.378</td>
<td>0.401</td>
</tr>
<tr>
<td>EOU4</td>
<td>3.39</td>
<td>0.487</td>
<td>0.213</td>
<td>0.293</td>
<td>0.331</td>
</tr>
<tr>
<td>EOU5</td>
<td>3.61</td>
<td>0.193</td>
<td>0.190</td>
<td>0.196</td>
<td>0.193</td>
</tr>
<tr>
<td>EOU6</td>
<td>3.44</td>
<td>0.220</td>
<td>0.318</td>
<td>0.213</td>
<td>0.250</td>
</tr>
<tr>
<td>EOU7</td>
<td>2.83</td>
<td>0.634</td>
<td>0.356</td>
<td>0.592</td>
<td>0.527</td>
</tr>
<tr>
<td>PER</td>
<td>2.15</td>
<td>0.379</td>
<td>0.135</td>
<td>0.377</td>
<td>0.297</td>
</tr>
<tr>
<td>PER1</td>
<td>2.10</td>
<td>0.363</td>
<td>0.113</td>
<td>0.371</td>
<td>0.282</td>
</tr>
<tr>
<td>PER2</td>
<td>2.15</td>
<td>0.310</td>
<td>0.185</td>
<td>0.328</td>
<td>0.274</td>
</tr>
<tr>
<td>PER3</td>
<td>2.20</td>
<td>0.278</td>
<td>0.126</td>
<td>0.209</td>
<td>0.204</td>
</tr>
<tr>
<td>DPR</td>
<td>3.27</td>
<td>0.735</td>
<td>0.632</td>
<td>0.640</td>
<td>0.669</td>
</tr>
<tr>
<td>DPR1</td>
<td>3.29</td>
<td>0.483</td>
<td>0.395</td>
<td>0.460</td>
<td>0.446</td>
</tr>
<tr>
<td>DPR2</td>
<td>3.26</td>
<td>0.644</td>
<td>0.593</td>
<td>0.581</td>
<td>0.606</td>
</tr>
<tr>
<td>DPR3</td>
<td>3.17</td>
<td>0.599</td>
<td>0.488</td>
<td>0.652</td>
<td>0.580</td>
</tr>
<tr>
<td>DPR4</td>
<td>3.41</td>
<td>0.619</td>
<td>0.527</td>
<td>0.354</td>
<td>0.500</td>
</tr>
<tr>
<td>DUP</td>
<td>2.93</td>
<td>0</td>
<td>0</td>
<td>0.680</td>
<td>0.543</td>
</tr>
<tr>
<td>DUP1</td>
<td>2.92</td>
<td>0</td>
<td>0</td>
<td>0.566</td>
<td>0.433</td>
</tr>
<tr>
<td>DUP2</td>
<td>3.00</td>
<td>0</td>
<td>0</td>
<td>0.573</td>
<td>0.380</td>
</tr>
<tr>
<td>DUP3</td>
<td>2.89</td>
<td>0</td>
<td>0</td>
<td>0.445</td>
<td>0.210</td>
</tr>
<tr>
<td>DUP4</td>
<td>2.79</td>
<td>0</td>
<td>0</td>
<td>0.562</td>
<td>0.538</td>
</tr>
<tr>
<td>DUP5</td>
<td>3.08</td>
<td>0</td>
<td>0</td>
<td>0.544</td>
<td>0.515</td>
</tr>
<tr>
<td>PFB</td>
<td>3.44</td>
<td>0</td>
<td>0.307</td>
<td>0.291</td>
<td>0.199</td>
</tr>
<tr>
<td>SUO</td>
<td>3.17</td>
<td>-</td>
<td>0</td>
<td>0.624</td>
<td></td>
</tr>
<tr>
<td>SUO1</td>
<td>3.27</td>
<td>-</td>
<td>0</td>
<td>0.513</td>
<td></td>
</tr>
<tr>
<td>SUO2</td>
<td>3.17</td>
<td>-</td>
<td>0</td>
<td>0.570</td>
<td></td>
</tr>
<tr>
<td>SUO3</td>
<td>3.07</td>
<td>-</td>
<td>0</td>
<td>0.548</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SPO</td>
<td>FBE</td>
<td>FBE1</td>
<td>FBE2</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>3.12</td>
<td>0.624</td>
<td>0.472</td>
<td>0.702</td>
<td>0.540</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0.489</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Table 6-6: R² coefficients of second and third layer value dimensions

<table>
<thead>
<tr>
<th>Regression Models</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUO model (8 indep. variables)</td>
<td>0.776</td>
</tr>
<tr>
<td>SPO model (8 indep. variables)</td>
<td>0.599</td>
</tr>
<tr>
<td>FBE model (2 indep. variables)</td>
<td>0.412</td>
</tr>
<tr>
<td>FBE model (10 indep. variables)</td>
<td>0.647</td>
</tr>
</tbody>
</table>

We can see that the R² coefficients of the first two models are 0.776 and 0.599 respectively, indicating that both second layer value dimensions are affected to a large extent by the ones of the first layer. On the contrary the R² coefficient of model_3 third has the much lower value 0.412, indicating that the third layer value dimension is affected to a smaller extent by the ones of the second layer. However, the last model has a much higher R² coefficient 0.647, which indicates that the first and second layer value dimensions in combination affect to a large extent the one of the third layer; therefore the first layer value dimensions affect users’ future behavior not only through the value dimensions of the second layer, but also directly as well. From the above results we can conclude that this value model is characterized by high coherence among its layers.

Finally, we calculated the correlations of the first layer value dimensions and their value measures with the value dimensions of the second and third layer (step 5), and the results are shown in the third, fourth and fifth column of Table 3; for each first level entity – value dimension or measure - we can see its correlations with the second and third level dimensions it affects according to the value model definition shown in Figure 6-16 - e.g. for the DPV we can see its correlations with the SUO and the FBE – and 0s in the other cells. In the sixth column we can see for the first level value dimensions and measures the average of their correlations with SUO, SPO and FBE, as an indicator of its overall impact on higher level value generation. We remark that with respect to the support of user-level objectives by the OGD infrastructure, the data search and download capabilities, the data processing capabilities and the ease of use are the first layer value dimensions that have the strongest impact on it (correlation coefficients
0.760, 0.735 and 0.730 respectively), while the performance has the weakest impact on it (correlation coefficient 0.379). With respect to the support of provider-level objectives by the OGD infrastructure, the data upload and the data processing capabilities are the first layer value dimensions that have the strongest impact on it (correlation coefficients 0.680 and 0.632), while the performance has the weakest impact on it (correlation coefficient 0.135). Finally looking at the last column of Table 6.5, we remark that the first layer value dimensions having the strongest overall impact on higher level value generation are the data processing capabilities (a novel capabilities set that characterise this second generation of OGD infrastructures) and the ease of use (correlation coefficients 0.669 and 0.552); the performance and the provider-level feedback capabilities have the weakest impacts (correlation coefficients 0.297 and 0.199).

Using the average ratings and correlations shown in Table 6.5 we can construct the value model of the OGD infrastructure (step 6) at the level of value dimensions, which is shown in the Figure 6.10 (while similarly we can construct a more detailed value model at the level of value measures). It provides a compact visualization of the main dimensions/types of value generated by this e-service (quantified through the corresponding average users’ ratings) and the relations among them (quantified through the corresponding correlation coefficients). This enables a better understanding of the value generation mechanism of OGD infrastructure, as it shows how value of one layer is transformed to value of higher layers, and also the origins of higher layers’ value.
Furthermore, based on these average ratings and correlations of Table 6.5 improvement priorities were identified (step 7). For this purpose we classified the first layer value dimensions into two groups according to their average rating: a higher ratings group and a lower ratings group (Table 6.7). Also, we classified them into two groups according to their average correlation with second and third layers’ value dimensions: a higher impact group and a lower impact group (Table 6.8). From these two classifications we can conclude that our highest priority should be given to the improvement of the data upload and data download capabilities, since they received low ratings from the users, and at the same time they have high impact on higher layers’ value generation. Similarly we can identify improvement priorities at the more detailed level of value measures.

Table 6-7: Classification of first layer value dimensions according to their average ratings by the users

<table>
<thead>
<tr>
<th>Lower Ratings Group</th>
<th>Higher Ratings Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6-8: Classification of first layer value dimensions according to their impact.

<table>
<thead>
<tr>
<th>Lower Impact Group</th>
<th>Higher Impact Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>data provision capabilities</td>
<td>data processing capabilities</td>
</tr>
<tr>
<td>data search-download cap.</td>
<td>ease of use</td>
</tr>
<tr>
<td>data upload capabilities</td>
<td>data search-download cap.</td>
</tr>
<tr>
<td>performance</td>
<td>data upload capabilities</td>
</tr>
<tr>
<td>user-level feedback capab.</td>
<td>user-level feedback capabilities</td>
</tr>
<tr>
<td>provider-level feedback cap.</td>
<td></td>
</tr>
</tbody>
</table>

6.3.4. Conclusions from the value model application

In the previous sections has been presented a methodology for evaluating the emerging second generation of OGD infrastructures, which has been influenced by the principles of the Web 2.0 paradigm, being oriented towards the elimination of the distinction between providers and consumers of such data, through the support of data ‘pro-sumers’ (i.e. users who both consume and produce such data). It is based on the estimation of value models of these advanced OGD infrastructures, which include assessments of the main types of value they generate, and also the relations among them. The proposed approach enables not only the identification of strengths and weaknesses, but also a deeper understanding of the value generation mechanism and a rational definition of improvement priorities. It should be noted that it can be used, with some adaptations, for the evaluation of the ‘traditional’ first generation OGD infrastructures as well.

An application of this approach was made for the evaluation of an advanced second generation OGD Infrastructure developed in the European project ENGAGE, leading to interesting insights. It has been concluded that the data processing capabilities, a key novel feature of this new generation of OGD Infrastructures, has the strongest impact on the generation of higher level value, associated with the achievement of fundamental objectives of users, and their future behaviour. Another novel feature, the user-level feedback capabilities (concerning rating and commenting datasets that users download and use, and also reading other users’ ratings and comments on datasets they are interested in), was found to have considerable impact...
on higher level value generation. Therefore, these novel Web 2.0 (active data pro-
sumers support) oriented capabilities seem to be valuable and promising.

6.4. Answering Research Question 4

Research question 4 has been answered in sections 6.1.4, 6.2.3 and 6.3.4. Section 6.1.4 presents the results and conclusions on the current theories of IS evaluation and how they can be used in the development of an Evaluation Framework. Section 6.2.3 presents the designed evaluation framework as well as the results from its application on the first version of the platform. Finally, section 6.3.4 define the value model for the final version of the designed infrastructure and provide the results from the application of the development priorities algorithm.
7. 3rd Generation of OGD Platforms

This chapter is dedicated to answer research question 5: “How to maximise value for Collaborative and Individual use of OGD and how to apply it to the Greek context?”. Many open data platforms are currently under development aiming to stimulate the potential advantages of the publication and use of open government data. In particular the development of open data platforms in the form of marketplaces, where open data providers and open data users trade and share data and data services, can stimulate the realisation of these advantages. Yet, only little research has been conducted on the development directions of open data platforms to realise such marketplaces. This chapter aims to identify elements for the development of future electronic open data marketplaces.

This aim is attained by investigating the literature and discussions with experts, which resulted in the following elements: 1) bring stakeholders together, 2) provide rich metadata, 3) enable data quality assessment, 4) ensure trust, security and critical mass, 5) have an appropriate revenue model, 6) provide use cases, training and support, 7) provide technical support: open data processing tools, 8) provide a full API for machine-to-machine operation and 9) target multiple nationalities. The results of this study can be used to develop and improve open data marketplaces to stimulate the realisation of open data advantages.

7.1. Literature Review

7.1.1. Market Places

Various studies have shown that the publication of public data has considerable potential to provide citizens, researchers, companies and other stakeholders with many advantages. These advantages include, but are not limited to, increased transparency (Bertot, Jaeger, & Grimes, 2010; McDermott, 2010), better services to citizens (Charalabidis, Ntanos, & Lampathaki, 2011), increased participation and interaction of stakeholders, empowerment of users and providers of open data (Neuroni, Riedl, & Brugger, 2013) and economic growth and value creation (Borzacchiello & Craglia, 2012).

To stimulate these potential advantages of the release and use of open government data, numerous platforms have been developed in the last decade. For
instance, open data platforms have been developed by national governments (e.g., Data.gouv.fr, 2013), local governments (e.g., Berlin Open Data, 2013; Open Government Wien, 2013), the European Commission (European Commission, 2012) and organisations and projects which are not country related (e.g., The ENGAGE project, 2014). These platforms have diverse characteristics, focus on different aspects and may complement each other (Zuiderwijk, Janssen, & Parnia, 2013).

7.1.2. Open Data Platforms

The literature suggests that open data platforms which facilitate interaction between open data providers and open data users could play an important role in stimulating the realisation of open data advantages (e.g. Chun, Shulman, Sandoval, & Hovy, 2010; Evans & Campos, 2013; Lee & Kwak, 2012; Maier-Rabler & Huber, 2011). Interaction between open data providers and users can be stimulated by so-called marketplaces. Marketplaces are places where suppliers and customers can meet each other (Henderson & Quandt, 1980) to indicate their intention to buy or sell certain products which eventually match and may be settled (Schmid & Lindemann, 1998). In the case of open data, open data providers and users can use a marketplace in order to interact and collaborate by trading and sharing open data and data services including advice and assistance in an open cooperative environment. As such, a marketplace can encompass various stakeholders and provide many types of data and numerous data services.

Despite the fact that numerous open data platforms have been developed, only few existing open data platforms actively stimulate the interaction between open data providers and open data users in the form of marketplaces. In addition, only few researchers have paid attention to the potential development directions of open data in general, and open data platforms in particular, in the near future (e.g., Lindman, Rossi, & Tuunanen, 2013; Vickery, 2011). Furthermore, these studies are mainly focused on high-level development directions for open data marketplaces, and do not pay attention to specific elements that these marketplaces could have, and which are necessary for progressing in this area. The objective of this study is to contribute to filling this research gap by identifying elements for the development of future electronic open data marketplaces. The identification of these elements can help in making better predictions.
for the evolution of marketplaces and in taking actions which may positively or negatively influence future developments in this area.

7.1.3. Open Web Services

The increasing release of public data is strongly correlated with the advent of openness in every aspect of software development. According to the Open Data Handbook, the definition of the Open Data (Dietrich et al., 2009) as “data that can be freely used, re-used and redistributed by anyone” is based on the principles of openness as a means to maximize interoperability (“The Open Definition”, 2014). The ability of diverse systems and datasets to inter-operate has been feasible through the emergence of web services and communication protocols among them, i.e. APIs (Bodle, 2011). Web services is the major technology being utilised for deploying automated interactions between distributed and heterogeneous applications and for enabling interoperability among business processes, which might span administration’s boundaries (Papazoglou & Georgakopoulos, 2003).

Open web services can contribute to the co-design and co-creation of applications in various domains, leveraging the vast amount of publicly available data. Therefore they can increase the innovation capacity and stimulate engagement of citizens, building upon the open data advantages (e.g. Chun et al., 2010; Evans & Campos, 2013; Lee & Kwak, 2012; Maier-Rabler & Huber, 2011). Despite the fact that numerous open data portals have been developed (Maier-Rabler & Huber, 2011; McDermott, 2010), only few existing open data platforms actively facilitate the interaction between open data providers and open data consumers in the form of service and applications creators. Most organizations providing governmental data merely publish this data on a platform, without being concerned on how this data can be used in business logic. Nonetheless, through proper APIs, services can be consumed by other people and be reused for building innovative applications. Furthermore, dependent services can be combined to each other. Therefore, it is of major significance to gather all available resources in a single repository, where citizens will be able to share information and collaborate.

Data and services coordination and integration demands time, effort and cost (Rolland et al., 2014). A system, whether it is between a government and its citizens (G2C) or any other cooperative system, has to bear the administrative burden for
collecting and providing such mechanisms (Roberts et al., 2011). Furthermore, there is a limit in how much information a person or an organisation can process (cognition wise) (Sack, 2006), even with the use of state-of-the-art infrastructures. There have been numerous studies concerning how a data repository should be properly designed (Bosley & Straub, 2002; Levi & Conrad, 2008; Frederickson-Mele, 1997), and their findings should be taken into consideration when targeting to a well-designed system.

There have been some attempts towards the reusability of the above resources to foster innovation (Bason, 2010; Borins, 2001; European_Commission, 2011a; 2011b; Hartley, 2005). However, they are mainly concentrated on research projects developing servers for facilitating the creation of mobile applications for governance (Paolucci, 2002), or supporting the co-creation in public services design and delivery and the enhancing access to Open Government Data and amalgamation of different open data sources. A lack of an appropriate ICT infrastructure for supporting open innovation in production of government applications has been identified, providing access to all reusable aspects, i.e. open data, services and applications (Kleijnen & Raju, 2003; Paolucci et al., 2002), which will act as a data and services registry. A registry provides a system the ability to record what is known about specific (e.g. open, application, interoperability (Charalabidis, Lampathaki, & Psarras, 2009) data and what is required to understand that data. Using the registry, data developers and collectors can document data for which they are responsible for, so that there is no need to repeatedly explain the data to each organization or individual interested in the data (unless the system changes).

In the current chapter (section 7.5.2), we present a recent initiative carried out by the University of the Aegean and Microsoft Innovation Center, in Greece, which develops a registry of data and services with the aim to support and foster development of applications for governance. The so-called Gov4All is a repository where all open data stakeholders could meet and access open data sources, services and applications developed upon open data, as well as, provide their own enhanced open datasets, services and applications. It must be noted here that the Gov4All acts as a directory of

26 http://opendata.okfn.gr/
27 http://www.engagedata.eu
28 http://yourdatastories.eu
29 http://stack.lod2.eu/blog/
30 http://www.icsd.aegean.gr/islab/
31 https://www.microsoftventures.com/locations/mic
open data repositories providing only references to external sites and datasets, and as such, data storage and curation is out of the scope of its infrastructure. The aim is to present and analyse the added value of this approach, while in parallel identify any shortcomings and limitations and finally to provide general guidelines with respect to necessary features of such marketplaces and repositories.

7.1.4. Current status in open data and open services repositories

There have been a lot of portals that provide open data and open services repositories. They can be distinguished in three types: (i) portals offering repositories on national level\textsuperscript{32} \textsuperscript{33} \textsuperscript{34} \textsuperscript{35}, (ii) ones that categorize their repositories according to their content topic\textsuperscript{36} \textsuperscript{37} (geospatial data, health data, etc.) (iii) portals that act as search engines, where a user can enter any search term (Canada, health) and get open data or services results \textsuperscript{27} \textsuperscript{38}. These portals are being enriched by experts or governmental bodies.

Most of the portals in all categories mentioned above provide a public API for information extraction and as well as lists of already deployed applications, taking advantage of a certain dataset, or open service. They present a menu option for listing applications containing government, community and business tools. The majority of the examined sites support user profiling. Content can be personalised, based on user custom preferences, and scalable, by providing a registration – login process.

Most sites offering data services (Tammisto & Lindman, 2012; Charalabidis, Loukis, & Alexopoulos, 2014), are focusing on a specific topic only (i.e. City information\textsuperscript{39}) and hence provide an API for gaining information on this topic. There are also other sites that provide services as well as open data on a specific term\textsuperscript{40}. Only a few portals\textsuperscript{41} have been identified that act as open service repositories, including a search engine, where a user can search any open service, regardless the topic or nationality.

\textsuperscript{32} http://open-data.okfn.gr/
\textsuperscript{33} http://open.canada.ca/
\textsuperscript{34} http://www.data.gov/
\textsuperscript{35} http://labs.europeana.eu/
\textsuperscript{36} http://inforumweb.umd.edu/econdata/econdata.html
\textsuperscript{37} http://geodacenter.asu.edu/
\textsuperscript{38} https://www.quandl.com/
\textsuperscript{39} http://www.citysdk.eu/
\textsuperscript{40} http://www.citadelonthemove.eu/
\textsuperscript{41} http://iserve.kmi.open.ac.uk/
The above analysis has revealed the absence of a combined approach in the design and features of open data portals. In particular there doesn’t exist a portal that contains a repository for both open datasets and open services, including also a list of public APIs for the above data or services and already deployed applications that utilize such APIs. Apart from that, they do not provide use cases, training material and support, which, according to literature (Zuiderwijk et al., 2013; Alexopoulos, Loukis, & Charalabidis, 2014; Zuiderwijk et al., 2015; Janssen, Charalabidis, & Zuiderwijk, 2012) consist basic elements for an envisioned open data repository. Moreover, existing approaches lack functionality that has been characterised as essential for the creation of such infrastructures (Zuiderwijk et al., 2013) in terms of: (a) open services repositories providing connections to consumed data, (b) citizens’ application repository, (c) classification schemes for apps and services in order for the user to locate more easily and (d) capabilities for users to provide input, needs, upload datasets, applications and services themselves.

As far as Greece is concerned, there are a lot of web applications providing information derived from a certain dataset (e.g. open data), however there is a lack of publicly available web services for developers or citizens to consume. The most active and innovative Ministries offering such web services are the Finance Ministry, the Ministry of Interior, the Ministry of Infrastructure, Transport and Networks and the Ministry of Citizen Protection (Alexopoulos, Spiliotopoulou, & Charalabidis, 2013). Some indicative examples of web services offered are the service providing information about a company based on their Tax Identification Number42 the Electronic submission of declarations for the Customs Office43 (ICISnet) and the Recovery of vehicle owners data (not publicly available).

In contemporary governance at any level (local, regional, national, European, as well as international), all open data should be publicly distributed through a web service. Developers would be able to consume these services and a lot of new and diverse web applications would then be offered to citizens.

Following a holistic approach, all previously described features should be incorporated in a single portal, that will allow users to choose between the data, the service or even a predeployed application to reuse, depending on their needs.

42 http://www.gsis.gr/gsis/info/gsis_site/PublicIssue/wnsp/wnsp_pages/wnsp_ver2.html
43 https://portal.gsis.gr/portal/page/portal/ICISnet/services?serid=10346626&adreseelID=10026938
(Alexopoulos et al., 2014). The Gov4All portal proposed in the subsequent sections, should it be developed, will fill the gap in this area.

7.2. Research approach

The approach of our research can be divided into three main steps. First we identified the developments influencing open data marketplaces. This was done by investigating the literature in the field of open data. Articles were found in databases such as Science Direct, Scopus, ACM Digital Library, IEEE Xplore and Google Scholar. We sought for journal and conference articles, books, governmental and non-governmental reports and other information in various databases. Eventually 20 papers were used to provide an overview of the main developments in the field of open data platforms.

Second, the general developments which may influence the development of open data marketplaces were organized by discussing them with fourteen experts from various fields, including the field of open government and e-democracy, public administration and engineering. Table 1 provides an overview of the expert details. Experts from various countries were consulted because we wanted to take into account open data marketplace developments in various contexts. Expert discussions took place during a workshop at the Conference on E-Democracy and Open Government (CeDEM) in 2013 in Austria, and in addition we consulted experts via e-mail.

Table 1: An overview of the experts consulted for this research.

<table>
<thead>
<tr>
<th>Expert number</th>
<th>Occupation</th>
<th>Sector</th>
<th>Country</th>
<th>Method of engagement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>Civil servant</td>
<td>Information Architect</td>
<td>Netherlands</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 2</td>
<td>Civil servant</td>
<td>Information manager</td>
<td>Netherlands</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 3</td>
<td>Consultant</td>
<td>Information Technology and Services</td>
<td>England</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 4</td>
<td>Researcher</td>
<td>E-government</td>
<td>Austria</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 5</td>
<td>Executive director</td>
<td>Information Technology and Services</td>
<td>Uganda</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 6</td>
<td>Senior Program Officer</td>
<td>Nonprofit Organization Management</td>
<td>Uganda</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 7</td>
<td>Researcher</td>
<td>E-government</td>
<td>Austria</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 8</td>
<td>Researcher, course director</td>
<td>E-government</td>
<td>Austria</td>
<td>Focus group</td>
</tr>
<tr>
<td>Expert 9</td>
<td>Researcher and Information Technologies &amp; New</td>
<td>Greece</td>
<td></td>
<td>Focus group</td>
</tr>
</tbody>
</table>
Third, based on the identified developments influencing open data marketplaces and on the expert discussions, nine elements for the development of future open data marketplaces were identified.

7.3. Identifying developments influencing open data marketplaces

In this section we define the main concepts used in this study and we describe important developments which influence future open data marketplaces.

7.3.1. Electronic marketplaces

Schmid and Lindemann (1998) write that historically marketplaces have evolved as institutions which allow "customers and suppliers to meet at a certain place and a certain time in order to announce buying or selling intentions which eventually match and may be settled" (p. 193). The evolution of information and communication technologies has led to the development of electronic marketplaces, which have enabled buying and selling at various times and spaces in the most efficient manner (ibid). Electronic marketplaces are virtual, technology-enabled trading spaces (Matook & Vessey, 2008), which can also be seen as intermediaries (Matook, 2013). Electronic markets emerge in various fields nowadays, such as stock exchange (e.g. NASDAQ Stock Exchange) and goods exchange (e.g. eBay). These electronic marketplaces support the exchange of numerous types of products and services with different types of actors (Schmid & Lindemann, 1998) and information exchange and payments among buyers and sellers (Matook & Vessey, 2008).
A number of electronic data marketplaces has already been developed in the field of open data. For example, InfoChimps\textsuperscript{44} focuses on obtaining business value from Big Data. A part of the open government data could be considered to be big data as well, but this does not count for most open government data. As a consequence, this marketplace does not well address other stakeholders than businesses and it does not focus on open government data in particular. Other examples of existing marketplaces are the Windows Azure Marketplace\textsuperscript{45}, and Timetric\textsuperscript{46}, which are also focused on commercial use. For instance, users of the Windows Azure Marketplace have to pay when they desire more than ten data transactions per month.

7.3.2. Open data platforms and marketplaces

Open government data are released increasingly (Whitmore, 2012) and many open data platforms have already been developed by governments at various levels (e.g. federal, ministerial, municipal) in all over the world. In this section we consider open data platforms to be platforms that are often owned by a single party (in this case governments) and that are used to make open government data available to the public. Various actors are involved in publishing and using open data on these open data platforms (Dawes & Helbig, 2010; Helbig, Cresswell, Burke, & Luna Reyes, 2012), such as open data providers, open data legislators, open data facilitators and many different types of open data users (e.g. citizens, researchers, journalists and developers).

The literature shows that there are considerable differences in the development of open data platforms (Braunschweig, Eberius, Thiele, & Lehner, 2012). For instance, the 50 repositories surveyed by Braunschweig et al. (2012) showed many differences in terms of openness. Zuiderwijk et al. (2013) investigated 35 functionalities of three open data platforms and also found that the investigated open data infrastructures are very diverse and focus on different aspects. The literature also provides some insights about similarities regarding open data platforms. After data providers have published their data on open data platforms, open data users may find these data, although most open data portals lack enhanced search capabilities (Tinholt, 2013). Furthermore, some open data portals redirect users to the websites of specific governmental organizations, which makes it far more cumbersome to obtain these data compared to data that can be

\textsuperscript{44} www.infochimps.com/marketplace
\textsuperscript{45} http://datamarket.azure.com/
\textsuperscript{46} https://timetric.com/
obtained from a central repository (ibid). The impediment of data fragmentation has also been found in other articles (e.g., Conradie & Choenni, 2012).

Many open data platforms lack standards and Application Programming Interfaces (APIs) and much open data is not machine readable or the data are provided in a proprietary format (Braunschweig et al., 2012). Kuk and Davies (2011) also emphasize the importance of APIs for machine-to-machine operations for open data. Open data providers usually publish their data without having contact with the data users. For instance, most open data portals lack tools for an effective dialogue with and for participation of users (De Cindio, 2012; Tinholt, 2013). On the other hand, research of Loukis, Charalabidis, and Alexopoulos (2014) has shown that open data platforms increasingly provide a wider range of open data marketplace functionalities, influenced by the principles of the Web 2.0 paradigm, and oriented towards the elimination of the clear distinction between providers and consumers of such data, and the support of data ‘pro-sumers’ (i.e. users who both consume and produce such data).

It was found that current open data portals often do not provide guidance to open data users on how to assess the data relevance and to investigate its feasibility to formulate positions (Evans & Campos, 2013). Usually only discovery metadata are provided with open government data and there is a lack of rich contextual metadata (Zuiderwijk, Jeffery, & Janssen, 2012), which are important for the interpretation of open data correct interpretation of open data and distilling knowledge from them (Foulonneau & Cole, 2005; Jeffery, 2000; Schuurman, Deshpande, & Allen, 2008; Vardaki, Papageorgiou, & Pentaris, 2009).

After the data users have found the data, they are usually not able to use the same open data platforms where they found the data to analyse, visualise, cleanse, curate, combine or link the data. There is some debate in the literature about whether this should be enabled by governments. Robinson, Yu, Zeller, and Felten (2009) argue that governments should provide simple open data platforms which are mainly focused on providing data. They state that private organizations could act as intermediaries which take the governmental data from these platforms and provide the data to citizens in an understandable way. On the other hand, this means that people who do not want to make use of intermediaries often have to search somewhere else for the tools to use open data. As a consequence, only the users with the appropriate technical skills can use
the data. Open data use by people with less developed skills is therefore less stimulated by these platforms.

7.4. Organizing developments: expert discussions

In this step the developments that were identified in the previous step are organized with the aim to end up with only a limited number of principal ones, which can then be used to derive development directions. The following statements and questions were presented to the experts.

1) The main target group of [the open data platform] should be the scientific communities.
2) [The open data platform] should be a marketplace for open data and collaboration.
3) [The open data platform] should be an open public data reputation management system.
4) [The open data platform] should put emphasis on rich metadata.
5) [The open data platform] should provide a rich collection of data curation tools.
6) [The open data platform] should provide a rich collection of data visualisation tools.
7) [The open data platform] should develop cases of using the platform.
8) [The open data platform] should focus on a standardization proposal for annotating open data sets for scientific usage.
9) [The open data platform] should provide a complete data repository.
10) [The open data platform] should provide a full API for machine-to-machine operation.
11) [The open data platform] should target multiple-nationalities.
12) What should be the [open data platform's] dissemination steps towards sustainability?
13) Which other steps/domains/USP’s should [the open data platform] target according to you?

According to their preferences, the experts received these statements and questions either on paper or online. In this survey the experts were asked to state to which extent they found this statement important on a range from one (very unimportant) to seven (very important). Thereafter the experts were asked to explain their opinion in a text box. Subsequently, a group discussion took place in which the experts were asked to explain what they had written in the online and paper surveys.
After the statements had been discussed, the more general questions (12 and 13) were discussed. The experts were also asked if there were other elements that were not included in the statements but that they assessed as important for a specific open data platform. Notes have been taken during the discussions. The whole session, including a general introduction, completing the surveys and the subsequent discussions, lasted for approximately 75 minutes.

Additionally, three experts were consulted via e-mail. They were also provided with the online forms with statements and questions and they were also asked to explain their assessments of the statements. No group discussions took place with these experts. Based on the discussions and the e-mail consultations, development directions for open data marketplaces were identified.

7.5. Answering Research Question 5

7.5.1. Identifying elements for open data marketplaces

In this section we discuss the elements of future open data marketplaces. In contrast with Robinson (2009), who argues that governments should provide open data platforms which are mainly focused on providing data, our discussions with experts revealed that future open data marketplaces should stimulate the interaction of open data providers and users. For example, expert 12 stated that “other data.gov sites do not provide this facility so it could be [a] unique selling point [...]”. This implies that the aspects of an open data marketplace could be integrated into open government data portals. Based on the discussions with the experts, an envisioned future open data marketplace was developed (see Figure 1).
Figure 1 represents an envisioned open data marketplace which integrates the identified elements for open data marketplaces. The database in this open data marketplace could merely contain metadata, but would preferably also contain the data themselves. Although the experts sometimes disagreed on the importance of certain developments, in general we could identify the following nine elements for the envisioned open data marketplace.

### 7.5.1.1. Bring stakeholders together (match supply and demand)

Even though markets are conventionally understood as places of commercial exchange and competition (e.g., Schmid & Lindemann, 1998), the expert discussions showed that open data markets could also be places of collaboration. Six out of the fourteen experts emphasized the need of an open data marketplace to provide mechanisms for interaction and collaboration, such as exchanging messages and data. This is in line with Davies (2012) who has also argued in favour of collaborating on open data as a common resource. The envisioned open data marketplace enables open data providers and users to find each other and interact. Thus, such marketplaces combine social aspects (e.g. user interaction) and technical aspects (e.g. the metadata and data system for supplying and demanding data). Research of Mayer-Schönberger and Zappia (2011) shows that there is currently a lack of evidence of collaboration between open data users and Velikanov (2010) writes that the regulation and facilitation of participation may become problematic when the number of participants...
becomes enormous. More research needs to be conducted on how collaboration in open data marketplaces could be stimulated.

Although this was not mentioned by the experts, in our opinion the provision of data should be connected to requests for data, in this way gearing the provided data to the needs of open data users. While the data are originally provided by governmental authorities, private organisations and individuals can curate and extend these datasets and uploading the curated and extended dataset. In this way the government can make use of the knowledge of the crowd and use new knowledge to improve their data provision, as well as their policy making and decision making. In the envisioned open data marketplace intermediaries can also offer data services to open data users (Davies, Perini, & Alonso, 2013; Mayer-Schönberger & Zappia, 2011). The users can decide whether they want to make use of the service. One of the experts stated that an open data marketplace “should make users aware of what’s on offer and facilitate its distribution”.

Quoting from one of the experts, “marketplace provision requires a high degree of accessibility design to facilitate strong use”. The reason for this is that in an open data marketplace for collaboration, different stakeholders involved in publishing and using open data are connected, such as civil servants and citizens. The marketplace should make it possible for the stakeholders to collaborate as also suggested by the literature (Davies, 2012; De Cindio, 2012), for instance, by creating groups and working on or using datasets together or by helping each other in finding certain data.

7.5.1.2. Provide rich metadata

With regard to the essential elements and development directions of an open data marketplace, expert 12 emphasized that the envisioned open data marketplace should provide a “combination of many datasets, rich metadata, some processing facilities and – specially – associated social networking/cooperative support [...]”. Also four other experts expressed that rich metadata are important for future open data platforms, whereas expert 10 disagreed with this and the other experts did not have a strong opinion about this. According to the experts in favour of rich metadata, emphasis should be put on rich metadata by combining discovery, contextual and detailed metadata (Zuiderwijk et al., 2012), so that advantages such as increased interoperability, better interpretation and better organisation can be achieved.
(Berners-Lee, 2009; Duval, Hodgins, Sutton, & Weibel, 2002; National Information Standards Organization, 2004). For instance, the CERIF metadata model could make it easier to interpret datasets by providing information about how the data were created, by who, when, and other information. This model also makes datasets more interoperable, as it allows for interconverting common metadata formats used in open data using CERIF as the superset exchange mechanism (Zuiderwijk et al., 2012). CERIF can be mapped to various other metadata models commonly used on open data platforms. Various initiatives have already initiated the harmonization of metadata between data catalogues. For example, the World Wide Web Consortium is also working on the Data Catalog Vocabulary (DCAT) for metadata about structured data resources (W3C, 2013).

7.5.1.3. Enable data quality assessment

Experts 8, 12 and 13 noted that the envisioned open data marketplace should provide information about the quality of datasets, as several types of open data users need this information to assess whether a particular dataset is appropriate for their purpose. The open data platform of the United Kingdom is already using such a data quality rating system (Read, 2012). Based on a quality rating users can decide whether they will use the dataset in a certain way. A rating system could make it easier to use data and generate value from it. Moreover, open data providers can use the wisdom of the crowd to learn from their feedback and to improve their datasets and policies. Based on the rating, they could perform further research and improve datasets. Nonetheless, it should be noted that user ratings are also contentious and subjective, since different users may assess the quality of one dataset in different ways. Furthermore, someone could organize a number of people to assign high data the highest quality, which is also known as the “claque effect” (Velikanov, 2010). A contribution to a partial solution to this problem could be the development of a framework on how the data quality could be assessed. For instance, the literature on Information Quality (IQ) can be used to select the appropriate quality aspects that need to be rated and develop a quality assessment framework (Batini, Cappiello, Francalanci, & Maurino, 2009). Yet, the problem of subjectivity cannot be solved completely.
7.5.1.4. Ensure trust, security and critical mass

De Cindio (2012) writes that a climate of mutual trust can be stimulated by fostering peer-to-peer, public dialogue among participants. Moreover, to create a position of trust in the open data marketplace, a critical mass of users is needed. For instance, many users are needed to assess the quality of the data to make the quality rating system useful. If only few people rate the quality of datasets, the reliability of this rating is low, while rating from many users could reinforce trust. Moreover, a critical mass of users is needed with regard to bringing together suppliers and users of datasets. If there are not enough stakeholders demanding data, the suppliers would not use the marketplace and trust would be low. At the same time, if many people would demand datasets, but there would be only few suppliers, trust in the marketplace would also be low. Trust should also be increased by clearly showing where certain data are coming from and how they were created. Related to trust, open data marketplaces need to ensure security. For instance, security can be created by using authorisation systems for users, by using secure payment systems to pay for open data services and by clearly explaining to the users which licenses and conditions apply to the use of specific datasets.

7.5.1.5. Have an appropriate revenue model

Various revenue models can be used for open data marketplaces (Ferro & Osella, 2013). One possible model is that an open data marketplace is funded by one or more governments, which makes it possible to provide data for free to the users of the platform. Free open data provision or open data provision at no more than a reasonable reproduction cost is often seen as one of the core principles of open data (e.g., Open Knowledge Foundation, 2005; Sunlight Foundation, 2013; Tauberer, 2012). On the other hand, one could also think of a revenue model in which individual users or private organization to a certain extent pay for open data use. For instance, users could be asked to donate money to support the platform or can be asked to pay for additional services that are offered, such as guarantees about when data are published or guarantees about quality checks of the data. Moreover, entrepreneurs can create applications based on open data and earn money by exchanging this product or service for money (Ferro & Osella, 2013). The same can be done with other open data services, such as selling services such as aggregating, comparing, analysing and visualising data.
by intermediaries. It is also possible to provide basic services for free, while asking money for more advanced services. One of the experts expressed that private sector innovation is one of the most important aspects of the envisioned marketplace.

7.5.1.6. Provide use cases, training and support

The open data marketplace can stimulate the advantages of open data by developing exemplar cases of using the platform and providing training and support for open data providers and users. Eight experts stated that this is an important aspect, and this could also stimulate the interaction between various stakeholders involved in the publication and use of open data. Use cases can provide an example of how open data platforms can be used, in this way also providing help, training and support. Open data users should be provided with support to build and sustain useful tools and services (Davies, 2012). One of the experts stated that support should also be provided with regard to providing clear information about the licenses that apply to open data.

7.5.1.7. Provide technical support: Open data processing tools

Future open data marketplaces can provide services which help with analysing, visualising, cleansing, curating, combining or linking the data on the marketplace itself, so that the users do not have to search for tools for performing the above types of data processing from other sources. This characteristic of the marketplace enables and assists open data users with less technical knowledge and skills to use open data. The data can then not only be found in the open data marketplace, but they can also be used and discussed there. Public dialogue among participants is often lacking in open data portals (De Cindio, 2012). Additionally, tools for visualizing data should be provided, which can make it easier to understand and interpret the data. Eleven experts emphasized the importance of visualizing open data, however, one expert also pointed at the problems that can arise when users try to visualize data that should not be visualized. Another expert pointed at the idea to warn users if certain use behaviour does not make sense, for instance, if a user tries to visualize certain values that should not be visualized. Moreover, during the discussions one expert suggested to have different options for technical and non-technical users. The system as a whole should be very simple and easy to use for non-technical people, but it should be possible for technical people to use more complicated options.
7.5.1.8. Provide a full API for machine-to-machine operation

Nine experts stated that it is important that the envisioned open data marketplace provides a full API for machine-to-machine operation. Such an API can be used, for example, to enable automated search to find datasets or to make the publication process easier. Furthermore, API’s allow for the development of mashups that combine data from different sources (Bizer, 2009) and the development of value added services. In this way APIs may provides technical support for the better and more effective use of open data platforms. Rich metadata enables such APIs.

7.5.1.9. Target multiple nationalities

Eleven of the thirteen experts expressed that it is very important that multiple nationalities are targeted. Open data platforms should make it possible to collaborate in an international level. Although the integration of datasets from different countries is very complex, the use of thesauri, lexicons and multilinguality might stimulate international collaboration in the use and exploitation of open data. Especially the comparison of heterogeneous data from different countries poses a risk, as there may be differences between these data which complicates their interpretation. Once again rich metadata (including multilingual ontologies) is the underpinning.

7.5.2. Realising an Open Data and Open Services Marketplace for Greece

7.5.2.1. The proposed Architecture

In order to fill the gap described in the previous section, the current study presents a complete governance framework in order to illustrate the clear benefits of ICT-driven public sector innovation and the take-up of the open and participative governance model (Osimo David, 2007) where both governments and third parties can collaborate and share responsibilities in designing, producing and delivering personalized services of public value, according to the accepted principles of subsidiarity (Botterman Maarten, Millard Jeremy et al., 2009). The proposed governance framework follows open innovation principles and is capable of guiding public administrators in adopting the ICT-driven public sector innovation approach to collaborative development of public services, as it provides the following:
An analytical guide for understanding how governments can involve citizens, businesses, CSOs in public service delivery, including planning and change management, success factors and risks in adopting public service co-production.

Identification of categories of promising public services and the potential benefits of their co-production.

A cost-and-benefit analysis for the adoption of public service co-production by the public sector.

Best practice co-production engagement strategies and initiatives for both public administrators and third parties.

Specifications for an integrated ICT infrastructure capable of supporting open innovation and co-production of personalised public services in an effective and efficient way.

A sustainable business model that offers financial incentives to citizens, businesses-SMEs, CSOs, and, public administrations for innovation generation through collaborative personalisation of public services.

The first priority for the Gov4All\textsuperscript{47} initiative is to establish the necessary substantiation for the case of collaborative production and delivery of public services, targeting, mainly, at mobilizing the public sector, but also the third parties to be

\textsuperscript{47} \url{http://gov4all.azurewebsites.net/en}
partnered, i.e., citizens, businesses, and CSOs. In other words, Gov4All aims to promote open and collaborative governance in Greece, through the development and dissemination of a modern and open platform for the voluntary development and utilization of governance applications from individuals and businesses, with the ultimate aim of improving the quality of citizens’ life and growth of Greek economy, saving resources and protecting the urban and natural environment.

Figure 1 presents the Gov4All governance framework in terms of involved stakeholders, provided assets and user actions and the overall concept towards the creation of new web services based on open data stimulating public value.

To achieve this, Gov4All has come up with an appropriate ICT infrastructure for supporting open innovation in collaborative production of public services. This constitutes a realization of the specifications found in the governance framework that can be achieved through integrating and customizing the results of relevant research and innovation projects. The proposed Gov4All ICT infrastructure consists of two main components: the Data, Services and Applications Directories; and the Community Engagement Environment. In particular:

**The Data, Services and Application Directories** provide access to open data, open services and applications offered by all levels of governance (central, regional and local) which have, fully or partially, adopted the open governance model. Access to available open data and open services is offered through web services APIs. Furthermore, the Data, Services and Applications Directories constitute a single access point to open data and open services for Greece adhering to the vision of the connected, networked and fully joined-up European public sector. Gov4All itself doesn’t collect, store or process any of these datasets but rather lists data and services that are already available in other places. As it refers only to open data, which by definition is freely available to everyone to use and republish as they wish without restrictions, Gov4All fully abides with data privacy and confidentiality principle. Users register their apps on their own and thus give their consent to publish this information adhering to a prescribed by the Greek data privacy regulation. Finally, the Data, Services and Applications Directories provide access to commercial services, such as GoogleMaps, which could be combined with open public services and data in order to create innovative services of public value.
The Community Engagement Environment provides a Web 2.0 participative environment that allows the users of the proposed open innovation platform, including citizens, businesses-SMEs, NGOs, and public administrators themselves, to engage in the specification of new personalized services of public value. The forms of engagement will vary as the community of users will be able to record the needs of a new service of public value or even needs for opening specific datasets by public sector providers. The users of the platform will be inspired by the discussions about public service delivery going on in the social media in a comprehensive way. Analytics about the engagement of users in the creation of new personalized services of public value will be produced and will be reported to the public sector policy makers.

Figure 2 illustrates the Gov4All basic modules and functionality. There are 5 basic categories of functionality consisting of additional sub-modules: (I) Declaration of needs in terms of (a) data, (b) web services and (c) applications; (II) Subscription of user developed assets or user identified assets, in terms of: (a) data, (b) web services and (c) applications; (III) Registration of (a) data, (b) web services, (c) applications, (d) semantic assets and (e) users; (IV) Addition of extra material (Use Cases, Educational Material and list of relative sites) and (V) Design of the application development. These five categories and their modules have been transformed into four major pillars (Applications, Services, Open Data and Community) in order to provide a friendly user interface according to (Charalabidis, Loukis, & Alexopoulos, 2014; Zuiderwijk et al., 2014; Alexopoulos et al., 2014) and it is depicted in Figure 3. Furthermore, Gov4All has been designed in order to meet the following technical characteristics:

- Web and mobile operation: compatibility with all platforms
- Mobile friendliness for all operating systems and devices
- Greek as basic language. English version is under development
- Social media-like user interface
- Maintenance without further development
• Three user roles: visitors, registered users, moderators
• Rating and comments capabilities for applications, data and web services from registered users and visitors
• Integration with Facebook, Twitter and LinkedIn social media platforms (users can use their own accounts for publishing their comments).

7.5.2.2. Design and Development

The Gov4All prototype has been designed with the aim to provide capabilities of interest for all involved stakeholders. The Governance Framework, including all relevant actors and their connections towards the creation of new web services based on publicly available data, is presented in Figure 1. Starting from this framework and its unique characteristics and moving forward to an application model of the previously described architecture, the Gov4All initiative provides the following re-usable assets and functionalities:

• Registration of open government applications developed by citizens and businesses, through equal and open participation of groups, individuals, businesses or organizations.
• Gathering of available open data and web services provided by the public sector, aiming to the voluntary development of open, free applications.
• Promotion of new technologies in open government, especially regarding mobile devices (mobile platforms), open web services and open governmental data.
• Provision of training on the use of open data and services for applications development.
• Supporting cooperation between citizens, government and businesses in order to promote the above objectives.

As an ongoing initiative, Gov4All will continue to be developed and engage stakeholders among open data and open service communities providing incentives and a space for service and application co-production, as well as new business models and guidance.

From a technological point of view, the Gov4All platform is based on Microsoft Azure Cloud Platform48. Microsoft Azure offers virtual machines for advanced computing,

web and mobile services hosting and support, unlimited data storage and stream analytics for real-time processing. These features are essential for the development, scaling and expansion of the Gov4All solution. The Gov4All initiative maintains its own directories of open data, open services and applications regarding Greece at the moment. It has developed custom metadata schemata for the subscription and hosting of open data, open services and applications.

Figure 3: Gov4All Home Screen Pillars

Services

Figure 4: Registered Taxation and Economy Services and Service submission Functionality

Finally, Gov4All provides a unique user interface, as well as, guidance for using it through the provision of training material, scenarios and use cases. Through these and the usability requirements, listed in the previous section, Gov4All attempts to address
the challenges of the integration of ICT infrastructures with human infrastructures. Such challenges are identified in (Procter et al., 2013) concerning research practices and their alignment with technical capabilities. However, the same applies on Gov4all, which can form an e-Infrastructure with the potential to be expanded and empower research community to access resources and services. As an open data repository, it can foster the emergence of e-Science and promote open collaboration between researchers and scientists across different disciplines.

Gov4All has been launched after meeting the following milestones:

- Name, URL, Facebook, Twitter, LinkedIn and Google+ reservation
- Basic functionality for all modules
- Basic codelists for metadata fields descriptions
- 50 registered apps following the developed metadata prototype
- 15 registered web services following the developed metadata prototype
- 22 registered open data sources following the developed metadata prototype
- 50 registered users
- Cooperation with other initiatives, developers, research centres, companies, public sector bodies, other communities etc.

As presented in the figure 3, the overall concept has been implemented on four major pillars: (a) Applications, (b) Services, (c) Open Data and (d) Community, thus, realising the proposed governance framework. Figure 4 presents the services pillar filtered for the registered taxation and economy services along with the services submission button (Gov4All maintains its own metadata schemata for the detailed description of services, open data and applications during their registration). Different information for the registered services is provided, such as: supported protocols, security types and data formats. Figure 5 shows the open data pillar with information about the datasets: to which categories it belongs; its data format; its data source; which is its rate by the users; how a user can submit a new dataset and a new resource; and how a user requests information from public authorities. Finally, figure 6 presents an active user’s profile with all the available information: how many and which services, applications and open datasets they have registered and finally, which public activities they have performed.
7.5.2.3. Further Development Steps

In the previous sections, Gov4All has been presented as a unique and innovative initiative, merging the two worlds of open government data and open government services within a single governance framework. As part of it, the Gov4All governance framework has been designed, which aims at stimulating the creation, delivery and use of new services on a variety of devices, utilising new web technologies, coupled with open public data. It also supports the creation of more personalised public services that better suit the needs of users through the collaboration space offering advanced web 2.0 – communication and rating – functionality. The Gov4All initiative makes a step forward to enable users develop and publish their own assets of public value. This step is focused on providing ease of development by addressing phased difficulties by mainly creating a single user interface and classification schemes for open data, open services and applications, hosted in a cloud database.

Moreover, Gov4All has developed a set of reusable and innovative assets that can be summarised in the following four points: (a) architecture for governance framework implementation, (b) functionality capturing and supporting citizens’ needs towards minimising administrative burden and maximising transparency, (c) managerial processes for the maintenance and management of such initiatives and (d) implementation of the quadruple helix involving academia, business, public sector and citizens in the service co-creation, overcoming the resistance to open innovation.

Gov4All platform is characterised by some limitations and shortcomings in order for the initiative to proceed on full scale: (a) there are only a few web services deployed in Greece, (b) there are only a few datasets with actual value for building added-value services in Greece (c) there is lack of interoperability mentality in the public sector and (d) there exist multiple platforms for developers. Furthermore, it is widely accepted [46], that to improve the innovation capacity of the public sector it is not enough to simply create a state-of-the-art technological solution – for the simple reason that without tangible evidence of success, governments will be reluctant to adopt collaborative open innovation – no matter how good the platform or its marketing. In the end, public administration should embrace this initiative by opening more data and providing better metadata upon them in order to strengthen this action and provide more incentives for users.
In order to address them, Gov4All foresees a three-prong approach to achieve its objectives and plan its further developments, illustrated also in figure 7:

1. Provide the means to demonstrate the clear benefits, including cost and benefit analysis, of ICT-driven public sector innovation. This will be achieved by creating an effective methodology or **Governance Framework** for engaging stakeholders in the innovation process in a scalable and cost-effective manner. Additionally, this will be enriched by the creation of more educational material.

2. Deploy this Governance Framework in conjunction with a state-of-the-art ICT infrastructure for open innovation in 3 pilot locations to substantiate and prove the **Business Case** for the collaborative production and delivery of public services. This will be achieved by organising targeted competitions / hackathons on specific open services provided by public administration, in order for users to develop applications.

3. Use tangible results from the deployment of the Framework and ICT Framework to drive the **dissemination activities** of the final Gov4All solution. This step will convince public administration to develop and publish new web services.

**Figure 7: Gov4All’s Virtuous Circle of Innovation Capacity**

Finally, the next steps of the initiative will be the development of the designing studio, the implementation of the public value services marketplace and the engagement activities towards (a) the public sector stimulating the web services development through the communication of citizens’ needs and (b) other communities, such as open software, in order for third-party applications to be realised.

**The Collaborative Service Design Studio** will allow the design and implementation of new personalised public services. It will offer to its stakeholders the opportunity to:
(a) design and create their own applications and (b) to create mashup applications, combining already existing applications and web services. The user, which could be either a citizen or a business or a public body, will be able to connect different provided services in order to create one of its own interest and, of course, publish it back for others to use. This will be achieved by analysing APIs of public sector online services, providing an extremely easy process for any user to create a devoted online service in one step. Thus, making citizens’ every day transactions with the government less time consuming and more effective.

*The Public Value Services Marketplace* offers created personalized services under commercial terms. It implements the vision where governments, citizens and businesses in Europe mutually benefit from services of public value. The proposed marketplace will offer services of public value created collaboratively by the public and private sector, within a flexible and sustainable business model. It will demonstrate the exact economic impact and financial sustainability of the proposed Gov4All governance framework.
8. Conclusions

Open Government Data (OGD) 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, due to its widely recognised potential to generate public value through driving innovation and economic growth, and also scientific research, and by promoting transparency and substantial evidence-based political dialogue.

However, this potential is to this day non-systematically exploited, as there are significant barriers that hinder the effective exploration, management and distribution of the vast amounts of available public sector information towards the research communities (e.g. operational data, financial performance data, process-related information, indices and metrics, key performance indicators, tacit knowledge). Furthermore, there is a shortage of experimental methods and tools that would allow effective knowledge mining, visualisation or further computation, empowering the integration of information and communication technologies into government practices and their adoption by the public. In the context of governance, multiple aspects have to be taken into account (e.g. financial, social, political, administrative, legal), indicating that in order to make administrative transformation a success, a multi-disciplinary approach has to be adopted, as opposed to the sole use of ICT.

Given all those facts, an intriguing opportunity emerges: there would be significant added-value from the deployment of an ICT platform for the collection, harmonisation, storage and dissemination of OGD, coupled with powerful search, visualisation and analytics tools according to the needs of its user-base (primarily for researchers, but also including citizens, industry and government).

In general, in order to cope with this challenge, a set of questions (section 1.3) were formulated resulting an advanced OGD infrastructure based on the design science approach (chapter 2). A scenario based design including users’ opinions were conducted in order to develop the first version of the infrastructure (chapter 5). An evaluation framework and a value model led to the next versions of the infrastructure (chapter 6). More specifically, the questions and their major results are presented in the next section.
8.1. Which are the main research challenges and opportunities for OGD and how we can map them?

The OGD research domain is still in its early stages, so it is important to identify the challenges and opportunities by developing a taxonomy of its main research areas and topics. The major contribution of this chapter is that 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 towards higher levels of maturity), through extraction and combination of relevant knowledge from three different 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 analysed, which enables a better understanding of them and their main research objectives and directions.

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 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. 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.

However, 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. 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.

8.2. What is the current state-of-play in the OGD infrastructures development?

The analysis of public sector data and knowledge resources strongly indicates the need for an infrastructure service platform for the collection, harmonisation, storage and dissemination of public sector information, coupled with powerful search, visualisation and Semantic Web tools. The designing of a solution phase will address current challenges of OGD sites, while combining the best breed of practices, specifications, tools and standards currently adopted, for the planning, communication, and execution of the platform.

The investigation of current public data and knowledge sources leads to a number of key conclusions that will drive the principles and basic elements of the service infrastructure:

1. There is a vast amount of available OGD dispersed among a great number of constantly increasing Public Data and Knowledge resources (Public agencies and
official, Open data initiatives, governmental portals, ministries, etc.). There is a great potential value for the effective exploration, management and distribution of this information towards the research communities in a unified manner.

2. The majority of OGD portals use non-standardized metadata, while there is no universally established metadata specification framework for OGD. There is a need for standardisation and harmonisation of metadata, semantics and ontologies, in order to ensure interoperability within and across Public Data Sources.

3. Searching and navigating through public sector data is difficult. State-of-the-art data discovery technologies need to be examined and applied to OGD portals.

4. The Linked Data paradigm is adopted only by recently developed initiatives, whereas traditional and longstanding public data sources are reluctant to adopt Linked Data or similar Semantic Web technologies.

5. Security threats such as data manipulation issues are not explicitly addressed in the majority of OGD portals, whereas security measures are limited to the typical authentication frameworks and configurations set in the web application server level or in the underlying platform level. The majority of OGD portals do not require registration or any special authentication and authorization to provide information, with a view to improving the user experience and increasing the user base.

6. Public datasets are provided in a national level only. There is a strong need to support cross-country datasets that would enable European and International researchers and citizens to utilize all available public datasets regardless of the language barrier.

8.3. What are the requirements and features for a new more advanced generation of OGD platform?

In the context of this study, scenarios form a tool for user requirements capture, content of use analysis, concept definition and overall user involvement. The value of the Scenario-based Design is that it concrete something for the purpose of analysis and communication. In total 10 scenarios were identified and analysed, based on the findings of research questions 1 and 2, deliberated during an open workshop[^49], commented online, and formally expressed, as following:

<table>
<thead>
<tr>
<th>Nr</th>
<th>Scenario</th>
<th>Public Servant</th>
<th>Researcher</th>
<th>Citizen</th>
</tr>
</thead>
</table>

[^49]: At the IFIP eGOV Conference in Delft, The Netherlands, 29th August 2011
More than 30 high level eGovernance, ICT and Policy Experts participated in the validation and rating of these scenarios resulting the requirements and capabilities of the first version of the platform.

The functionality of this platform to the two main stakeholders, the open data users and providers, is described, initially the ‘classical’ (basic functionality supported by most first generation OGD platforms) and then the novel Web 2.0 oriented one. The former includes mainly data publication-upload and modelling (metadata) for the data providers, and data search, visualization and download for the data users.

<table>
<thead>
<tr>
<th>Classical Functionality</th>
<th>Stakeholder</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong> Data Publication/upload</td>
<td>Provider</td>
<td>Support for publication/upload of datasets to the providers</td>
</tr>
<tr>
<td><strong>2</strong> Data Modeling</td>
<td>Provider</td>
<td>Capabilities of flat metadata descriptions (based on a specific metadata models)</td>
</tr>
<tr>
<td><strong>3</strong> Data Search</td>
<td>User</td>
<td>Simple search via keywords, resource format, publisher, topic categories and countries</td>
</tr>
<tr>
<td><strong>4</strong> Data Visualisation</td>
<td>User</td>
<td>Simple visualisation techniques on specific</td>
</tr>
</tbody>
</table>
The novell Web 2.0 capabilities aim to support open data ‘prosumers’ (who are at the same time users of provided datasets, and also producers of new versions of them (through various types of processing), which are improved, enriched-extended or adapted for specific purposes, or even of new datasets), and also extensive interaction and collaboration among them. This novel functionality includes users groups formation and extensive communication and collaboration within them, data processing, enhanced data modeling (flat, contextual and detailed metadata), commenting existing datasets and expressing needs for new datasets, datasets quality rating, data linking, publication/upload of new versions of existing datasets, advanced data visualization.

<table>
<thead>
<tr>
<th>Novel Functionality</th>
<th>Stakeholder</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Grouping and Interaction</td>
<td>Provider/User</td>
<td>Capabilities for (a) searching for and finding other users/providers having similar interests with me in order to have information and knowledge exchange and cooperation, (b) forming groups with other users/providers having similar interests with me in order to have information and knowledge exchange and cooperation, (c) maintaining datasets/working on datasets within one group, (d) communicating with other users/providers through messages in order to exchange information and knowledge and (e) getting immediately updated about the upload of new versions and enrichments of datasets maintained/worked on within the</td>
</tr>
</tbody>
</table>
3 Data Processing  Provider/User  Capabilities for
(a) data enrichment - i.e. adding new elements - fields,
(b) for metadata enrichment - i.e. fill in missing fields,
(c) for data cleansing - e.g. detecting and correcting repetitions in a dataset, matching text names to database IDs (keys) etc.,
(d) converting datasets to another format,
(e) submitting various types of items - e.g. visualisations, publications - related to a dataset and
(f) datasets combination and Mash-ups.

4 Data Enhanced Modeling  Provider/User  Capabilities for description of flat, contextual and detailed metadata of any metadata/vocabulary model.

5 Feedback and Collaboration  Provider/User  Capabilities
(a) to communicate our own thoughts and ideas on the datasets to the other users and the providers of them through comments I enter on them,
(b) to read interesting thoughts and ideas of other users on the datasets through comments they enter on them,
(c) to express our own needs for additional datasets that would be interesting and useful to me,
(d) to get informed about the needs of other users for additional datasets and
(e) to get informed about datasets extensions and revisions.
6. Data Quality Rating

User

Capabilities to (a) communicate to the other users and the providers the level of quality of the datasets that I perceive, (b) get informed on the level of quality of the datasets perceived by other users through their ratings.

7. Data Linking

Provider/User

Capabilities of data and metadata linking to other ontologies in the Linked Open Data Cloud. Capabilities of querying data and metadata through Sparql Endpoints.

8. Data Versions Publication/upload

Provider/User

Support for publication/upload of new versions of the existing datasets, and connection with previous ones and initial datasets.

9. Advanced Data Visualisation

User

Advanced visualisation techniques on specific datasets and/or datasets mash-ups (maps, charts, plots and other).

10. Advanced Data Search

User

Multilingual search based on thesauri and advanced metadata schema.

8.4. Which are the existing evaluation models and how to evaluate an OGD Initiative?

For the development of our methodology we have taken into account approaches and frameworks developed from four relevant streams of previous IS research on: i) IS evaluation, ii) IS acceptance, iii) IS success and iv) e-services evaluation.

An evaluation framework were designed based on qualitative and quantitative evaluation methods. The evaluation procedure consists of internal evaluation (by individuals within the e-infrastructure development consortium) and external evaluation (by individuals not belonging to the development consortium). The internal evaluation includes: i) a semi-structured discussion in a group of experts, ii) Web
Analytics and iii) a SWOT (Strengths-Weaknesses-Opportunities-Threats) analysis. The external evaluation includes: i) a similar semi-structured discussion in a group of social, political and management sciences researchers (potential users of the platform) not belonging to the project’s scientific committee, ii) a structured usability test measuring user’s performance on specific tasks (using a 5-Likert scale from very difficult to very easy), iii) a quantitative and structured questionnaire and iv) a Qualitative Discussion with users.

<table>
<thead>
<tr>
<th>Internal assessment</th>
<th>External assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGAGE Consortium (including Experts Scientific Committee)</td>
<td>End user</td>
</tr>
<tr>
<td>Semi-structured Experts Questionnaire</td>
<td>Structured SWOT-Analysis</td>
</tr>
<tr>
<td>Web Analytics</td>
<td>Experts in the field</td>
</tr>
</tbody>
</table>

The application of this framework identified the strengths and weaknesses of the first version of the platform and led to the next versions. The basic outcome is that for the first version, the platform achieved to meet the technical objectives but it not manage to attract sustainable interest. The very next version (2.0) achieved both targets. During the application of those tools to a very large and diverse audience, we manage to capture new needs from our target groups and identify the current gaps in the development of such an infrastructure as well as to quantify them and make decisions for the next versions.

The current e-infrastructure enables searching, downloading, analyzing, uploading and rating data and also the manual linkage of data. Moreover, the ENGAGE e-infrastructure provides a comprehensive overview of which open data e-infrastructures currently exist in Europe. In addition, the ENGAGE infrastructure enables the usage of Application Programming Interface (API) for searching data, as well as, of a Wiki including tutorials.

Nevertheless, the e-infrastructure could be improved in many ways. The application of the evaluation model and procedure showed that future versions of the ENGAGE e-infrastructure should mainly focus on 1) difficulties in searching, need for more information about datasets, multilinguality, capabilities for combining datasets, and also for doing more complex analysis and performance, 2) more information about the quality of the data, for instance by providing a comprehensive rating systems for the
data and 3) more information about which data can be linked, in order to meet users’ expectations for using the platform.

In general, the first application of the proposed evaluation framework showed that meaningful insights can be obtained by applying the proposed evaluation framework and infrastructure to such an advanced open public data e-infrastructure 2.0. The insights concern both managerial (categorizing priorities, decision making) and technical aspects of the e-infrastructure. A detailed presentation of the results and their usefulness exist in section 6.2.2.

Moreover, a structured approach was developed for assessing and improving e-services towards the final version of the platform, which is based on the estimation of value models of them from users’ ratings. Such a value model consists of a set of value measures, assessing different types of value generated by the evaluated e-service, and also the relations among them. These value measures are organized in three layers:
(a) Efficiency layer: it includes ‘efficiency’ measures, which assess the quality of the basic capabilities offered by the e-service to its users.
(b) Effectiveness layer: it includes ‘effectiveness’ measures, which assess to what extent the e-service assists the users for completing their tasks and achieving their objectives.
(c) Future behaviour layer: it includes measures assessing to what extent the e-service influences the future behaviour of its users (e.g. to what extent they intend to use the e-service again in the future, or recommend it to friends and colleagues).

The developed model has been influenced by the principles of the Web 2.0 paradigm, being oriented towards the elimination of the distinction between providers and consumers of such data, through the support of data ‘pro-sumers’ (i.e. users who both consume and produce such data). It is based on the estimation of value models of these advanced OGD infrastructures, which include assessments of the main types of value they generate, and also the relations among them. The proposed approach enables not only the identification of strengths and weaknesses, but also a deeper understanding of the value generation mechanism and a rational definition of improvement priorities. It should be noted that it can be used, with some adaptations, for the evaluation of the ‘traditional’ first generation OGD infrastructures as well.

An application of this approach was made for the evaluation of the final version of the advanced OGD Infrastructure developed in the European project ENGAGE, leading to interesting insights. It has been concluded that the data processing capabilities, a key
novel feature of this new generation of OGD Infrastructures, has the strongest impact on the generation of higher level value, associated with the achievement of fundamental objectives of users, and their future behaviour. Another novel feature, the user-level feedback capabilities (concerning rating and commenting datasets that users download and use, and also reading other users’ ratings and comments on datasets they are interested in), was found to have considerable impact on higher level value generation. Therefore, these novel Web 2.0 (active data pro-sumers support) oriented capabilities seem to be valuable and promising.

8.5. How to maximise value for Collaborative and Individual use of OGD and how to apply it to the Greek context?

Four types of OGD value generation mechanisms have been identified: i) efficiency mechanisms (public sector organizations generate economic value by increasing internal efficiency and effectiveness), ii) transparency mechanisms (public sector organizations generate social value by offering increased transparency into government actions, which reduces ‘information ‘asymmetry’ between government officials and citizens, and therefore misuse of public power for private benefits and corruption), iii) innovation mechanisms (private sector firms generate economic value through the creation of new products and services), iv) participation mechanisms (private sector firms generate social value through participating and collaborating with government).

Electronic markets emerge in various fields nowadays, such as stock exchange (e.g. NASDAQ Stock Exchange) and goods exchange (e.g. eBay). These electronic marketplaces support the exchange of numerous types of products and services with different types of actors and information exchange and payments among buyers and sellers.

Many open data platforms are currently under development aiming to stimulate the potential advantages of the publication and use of open government data. In particular the development of open data platforms in the form of marketplaces, where open data providers and open data users trade and share data and data services, can stimulate the realisation of these advantages. Yet, only little research has been conducted on the development directions of open data platforms to realise such marketplaces. The elements for the development of future electronic open data
marketplaces were identified by investigating the literature and discussions with experts, which resulted in the following: 1) bring stakeholders together, 2) provide rich metadata, 3) enable data quality assessment, 4) ensure trust, security and critical mass, 5) have an appropriate revenue model, 6) provide use cases, training and support, 7) provide technical support: open data processing tools, 8) provide a full API for machine-to-machine operation and 9) target multiple nationalities. The results of this study can be used to develop and improve open data marketplaces to stimulate the realisation of open data advantages.

Furthermore, open web services can contribute to the co-design and co-creation of applications in various domains, leveraging the vast amount of publicly available data. Therefore they can increase the innovation capacity and stimulate engagement of citizens, building upon the open data advantages. Despite the fact that numerous open data portals have been developed, only few existing open data platforms actively facilitate the interaction between open data providers and open data consumers in the form of service and applications creators towards exploiting the four types of value generation in the domain of OGD.

In order to apply this notion to the Greek reality, we developed an appropriate ICT infrastructure for supporting open innovation in collaborative production of public services. This constitutes a realization of the specifications found in the governance framework that can be achieved through integrating and customizing the results of relevant research and innovation projects. The proposed Gov4All ICT infrastructure consists of two main components: the Data, Services and Applications Directories; and the Community Engagement Environment. In particular:

**The Data, Services and Application Directories** provide access to open data, open services and applications offered by all levels of governance (central, regional and local) which have, fully or partially, adopted the open governance model. Access to available open data and open services is offered through web services APIs. Furthermore, the Data, Services and Applications Directories constitute a single access point to open data and open services for Greece adhering to the vision of the connected, networked and fully joined-up European public sector. Gov4All itself doesn’t collect, store or process any of these datasets but rather lists data and services that are already available in other places. As it refers only to open data, which by definition is freely available to everyone to use and republish as they wish without restrictions, Gov4All
fully abides with data privacy and confidentiality principle. Users register their apps on their own and thus give their consent to publish this information adhering to a prescribed by the Greek data privacy regulation. Finally, the Data, Services and Applications Directories provide access to commercial services, such as GoogleMaps, which could be combined with open public services and data in order to create innovative services of public value.

**The Community Engagement Environment** provides a Web 2.0 participative environment that allows the users of the proposed open innovation platform, including citizens, businesses-SMEs, NGOs, and public administrators themselves, to engage in the specification of new personalized services of public value. The forms of engagement will vary as the community of users will be able to record the needs of a new service of public value or even needs for opening specific datasets by public sector providers. The users of the platform will be inspired by the discussions about public service delivery going on in the social media in a comprehensive way. Analytics about the engagement of users in the creation of new personalized services of public value will be produced and will be reported to the public sector policy makers. There are 5 basic categories of functionality consisting of additional sub-modules: (I) Declaration of needs in terms of (a) data, (b) web services and (c) applications; (II) Subscription of user developed assets or user identified assets, in terms of: (a) data, (b) web services and (c) applications; (III) Registration of (a) data, (b) web services, (c) applications, (d) semantic assets and (e) users; (IV) Addition of extra material (Use Cases, Educational Material and list of relative sites) and (V) Design of the application development.

8.6. Discussion and further research

Addressing the five basic research questions of this dissertation, different issues accrued and handled that are of major importance for the development of an OGD
infrastructure. These issues will be discussed in the next sections and will be assimilated into the greater area of OGD research articulating the future research.

8.6.1. An extended OGD Life-Cycle

A few models have been designed in order to identify the open data life cycle. Each model focuses on different perspectives of open data. Two are the basic aspects on which different life-cycle models were created: (a) the Linked Open Data Life-Cycle and (b) the Data Curation Life-Cycle.

As the open government data evolving during the years there were many attempts of describing their life cycle as well as standardised it. Hyland et al. describe the open data life cycle which consists of seven discrete steps towards their publishing by the corresponding public body (Identify → Model → Name → Describe → Convert → Publish; and closing the loop → Maintain). Villazon-Terrazas et al. make a step further including exploitation after publishing and their model consists of five steps (Specify → Model → Generate → Publish → Exploit). Going a bit further we will meet the Datalift vision in which there is the additional step of interlinking after publication and then the exploitation one. After all these attempts for establishing a model describing the supply side, there are two more attempts trying to cover the demand side of the open government data life cycle. Hausenblas et al. include three steps for the side of data users (Discovery → Integration → Use Case). What is more the LOD2 project describes the linked open data life cycle considering both the supply and the demand side including tools for utilisation in each step and the Linked Open Data life cycle is formulated as follows: Manual Revision /Authoring → Interlinking / Fusing → Classification/Enrichment → Quality Analysis → Evolution/Repair → Search/Browsing → Extraction → Storing/Querying. Finally, in order to support the full life cycle of open data the Open Data Support Working Group resulted in the steps presented in Figure 8.1.

DCC Curation Lifecycle Model provides a graphical, high-level overview of the stages required for successful curation and preservation of data from initial conceptualisation or receipt through the iterative curation cycle\(^50\). It is important to note that the model presented in Figure 8.2 is an ideal. In reality, users of the model may enter at any stage of the lifecycle depending on their current area of need. For

\(^{50}\) [http://www.dcc.ac.uk/resources/curation-lifecycle-model#sthash.FnrCA3Kf.dpuf]
instance, a digital repository manager may engage with the model for this first time when considering curation from the point of ingest. The repository manager may then work backwards to refine the support they offer during the conceptualisation and creation processes to improve data management and longer-term curation. The model enables granular functionality to be mapped against it: to define roles and responsibilities and build a framework of standards and technologies to implement as it is showed in the above figure.

![Figure 8-1: OGD Life Cycle captured from Open Data Support Working Group](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)

![Figure 8-2: Data Curation LifeCycle Model](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)
A hybrid model has been produced incorporating steps from both its predecessors. Actually, the whole curation life cycle is embedded in the “Curate” and “Pre-process” steps of the ENGAGE Open Data Life Cycle and as far as the linked open data life cycle, it has been used for the rest. The basic development of the ENGAGE project since its conception is the collaboration step which is not included in any one of the above models. This is a result of the ENGAGE advanced functionality and web 2.0 capabilities which in fact provide a solid solution towards the realisation of the HORIZON 2020 vision concerning the e-infrastructures development for new workflows and collaboration introducing a crowdsourcing-based platform for data processing and data exchange among users.

The methods and tools used for each life cycle stage are presented in the following table.

<table>
<thead>
<tr>
<th>Life Cycle Stage</th>
<th>Tools</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create / Gather</td>
<td>Sensors, RFID, IS, Human, Connection with already gathered open data</td>
<td>Automated data creation Manual data entry Linking with Open Data Portals</td>
</tr>
<tr>
<td>Pre-process</td>
<td>Detailed Metadata Standards Evaluation Metrics and Models Maturity Matrices</td>
<td>Conceptualisation Structuring Anonymisation Evaluation</td>
</tr>
<tr>
<td>Curate</td>
<td>LOD Refine External Tool</td>
<td>Metadata Refinement</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td></td>
<td>Individual / Native Tools</td>
<td>Change Data Format</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data Cleansing</td>
</tr>
<tr>
<td>Store / Obtain</td>
<td>Repository and Data Centre</td>
<td>Versioning</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data Linking</td>
</tr>
<tr>
<td>Publish</td>
<td>Upload Capability</td>
<td>Licensing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Intellectual Property Rights</td>
</tr>
<tr>
<td>Retrieve / Acquire</td>
<td>Multilingual Search techniques</td>
<td>Open Access</td>
</tr>
<tr>
<td></td>
<td>Download capabilities</td>
<td>3-layer Metadata Schema</td>
</tr>
<tr>
<td>Process</td>
<td>External Data Processing tool</td>
<td>Data enrichment</td>
</tr>
<tr>
<td></td>
<td>LOD Refine External Tool</td>
<td>Create Linked Open Data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Different Datasets combination</td>
</tr>
<tr>
<td>Use</td>
<td>Internal Visualisation tool</td>
<td>Statistical Analysis</td>
</tr>
<tr>
<td></td>
<td>External Visualisation tool</td>
<td>Map Visualisation</td>
</tr>
<tr>
<td></td>
<td>Statistical Package</td>
<td>Chart Visualisation</td>
</tr>
<tr>
<td></td>
<td>Linking with external artifacts</td>
<td>Plot Visualisation</td>
</tr>
<tr>
<td>Collaborate</td>
<td>Collaboration Space and Workflow</td>
<td>Exchange</td>
</tr>
<tr>
<td></td>
<td>Web 2.0 Capabilities and Tools</td>
<td>notes/emails/ideas</td>
</tr>
<tr>
<td></td>
<td>Declare Need</td>
<td>Data Quality Rating</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Create Groups of common interests</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Requests on Open Data</td>
</tr>
</tbody>
</table>

### 8.6.2. Open Data Stakeholders

Within the scope of this dissertation, data users (scientific communities and citizens) and data providers (civil servants) are the primary identified users of an OGD platform. Besides that fact the research conducted in this dissertation provides a more general identification of users’ communities. The study was based on extensive literature review, study of relevant use case scenarios, interviews, workshops,
consultation and conversations between the ENGAGE project consortium and representatives of the various end-users' groups and brainstorming amongst experts. This study introduces the general user communities for Open Government Data - more specifically public data end-users – are as follows.

<table>
<thead>
<tr>
<th>Stakeholder Group</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Academia and Researchers</strong></td>
<td>Academia includes all stakeholders working in the frames of Universities and other post-graduate organisations. Indicative categories of academia stakeholders are: Professors; senior researchers; post-graduate students; pre-graduate students; PhD candidates in any academic field. Researchers constitute another, yet closely coupled with academia, category. Thus, they are reported as an integrated category of end-users. Researchers are persons that perform any type of research. Researchers might be working in academia (universities), in public administrations (e.g. ministries, provinces, municipalities) or private research institutions, commercial organisations etc. Academia and researchers can also be met in other categories of end-users (e.g. citizens, developers, businesses etc.), yet with different and not necessarily connected needs. It can be definitely claimed that end-users that belong either to Academia and/ or Researchers can use open/ public data on a daily basis; open/ public data might even constitute one of the main assets of their work and/ or research.</td>
</tr>
<tr>
<td><strong>Citizens</strong></td>
<td>Citizens, also referred to as “general public” and/ or “open public”, are persons that use open/ public data in various ways and for various purposes (<a href="http://openhalton.ca/">http://openhalton.ca/</a>). There is definitely an overlap amongst citizens and all other categories; citizens constitute a super set (from an entity scope) of all other categories. However, in the following analysis we consider citizens as end-users, not citizens as researchers or citizens as developers. The main difference between citizens and the other target groups is that citizens (as users) do not use open data for their work, but in their spare time, out of curiosity (e.g. hobby), for getting information etc.</td>
</tr>
</tbody>
</table>
| **Developers** | Developers are persons that usually want to reuse interesting open/ public governmental data for various purposes. Developers could be:
Persons working in their spare time (e.g. as a hobby); employees of businesses and self-employed persons; individuals working on start-ups; students developing mobile applications. Of course, as in previous categories, there is an overlap amongst this category and many of the others. For instance, a developer may work for a company and develop applications for financial gains. An indicative difference between developers and researchers is that developers may be interested in real-time data or data which is regularly updated (in the sense that few “added value” applications can be made using only static (e.g. geospatial) data), whereas researchers are less interested in real-time data. In addition, developers usually utilise data for commercial purposes, while this is not the case for most of the researchers. Developers can use OGD in their (daily) operation in work or in spare time. For instance, developers can build an application on top of the combined dataset, so that services to citizens are provided or money can be made.

| **Journalists** | Journalists are persons that want to reuse open/ public government data in order to report on them in various types of media they work for including: Academic/ scientific publications; television; magazines; radio; web/ online channels; social media etc. There is an overlap amongst the categories journalists, businesses, researchers and citizens. For instance, a journalist may do research by performing statistical analysis and report on this in a paper or on the internet. Journalists can use open data in their daily operation in work. Sometimes they even do not know which specific data they are looking for, or how the datasets will be eventually utilised. |
| **Civil Servants** | OGD can be of great value to administration organisations of every level. For instance, persons (such as policy makers and data managers) working in federal, ministerial, provincial or local government organisations, can benefit from available datasets (Saran, 2012). For the purpose of the ENGAGE project, civil servants that are willing to publish government data to the ENGAGE platform and that may also be willing to enrich existing datasets, comment on their quality and perform some other activities that the platform could benefit from, are more than important. There is an obvious overlap amongst the category civil servants and other categories (e.g. |
researchers, citizens). For instance, a policy researcher working for a governmental organisation may belong to both the category of researcher and civil servant. Civil servants can use open government data in their daily operation in work. For instance, they can use it to provide OGD, for reasons of transparency.

### Business Employees

Businesses are private commercial organisations that may use the ENGAGE platform to obtain economical/strategic benefits. There is an overlap amongst the categories businesses, researchers, developers and citizens. For instance, businesses may employ researchers that perform commercial or non-commercial research (e.g. market research). Indicative (yet not exhaustive) examples of research-intensive commercial organisations might include: Large industries; Large enterprises; SMEs. Businesses can profit by generating or consuming open data through providing data-centric applications and services.

### Archivists and librarians

Archivists and librarians are persons, organisations, or units within the larger organizations that archive or curate datasets produced by national and local governments, other public bodies and organizations, businesses, and researchers. Archivists and librarians, as it will be analysed in the next Chapter, mainly want to use a platform to curate their own data or link them to other open datasets, to monitor the open data dynamics and modes of use, to enhance/enrich datasets, manage digital preservation, as well as to produce reports and recommendations.

### 8.6.3. A maturity model for OGD e-Infrastructures

During the last years the number of published Public and Open Data portals has significantly increased, since the promise of open government and the potential for increased awareness by citizens of government activity and participation in the institutions of governance have never been greater. The next list outlines the existing categories of OGD portals based on the research conducted in chapter 4.

i) OGD direct provision portals: this is the main category of OGD portals, which are ‘primary sources’ of OGD, publishing original government datasets provided by either one government agency, or a small number of similar government agencies (who are
the legal owners/licensers of the data); they usually offer a wide range of functionalities supporting the whole lifecycle of OGD, from the creation of datasets to the update and finally to the archiving of them.

ii) **OGD aggregators**: this category includes OGD aggregator portals, which are ‘secondary sources’ of OGD coming from a big number of government agencies, publishing and maintaining lists of other ‘primary’ OGD catalogues and links to them. They constitute single access points to multiple OGD direct provision portals, and make it easier for a user to locate the OGD he/she is interested in. Usually they include descriptive information about datasets and sources, which is quite useful for the users in order to get a first impression of what is available. Many of them act as highly structured registries of OGD primary sources and datasets, which store structured and machine processable information, and provide 'index'-like features, such as automated registration and discovery of OGD. Some prominent examples are the widespread CKAN data hub portal, the Open Government Data Initiative (OGDI) – (a Microsoft initiative / online tool to publish and use a wide variety of public data from government agencies), the European Union Open Data portal and the Freebase.

Another categorization of the OGD sources with respect to the capabilities/functionalities offered can be made based on the web paradigm they are based on (the ‘traditional’ Web 1.0 paradigm, or the more recent Web 2.0 paradigm) (Alexopoulos et al., 2013; Alexopoulos et al., 2014):

**The ‘traditional’ first generation OGD portals**, which have been influenced by the Web 1.0 paradigm, in which there is a clear distinction between content producers and content users. They are characterized by datasets publishing in non-machine-processable formats (i.e. PDF), without providing any contextual information or linkage capabilities to other datasets; also, they are limited to offering basic functionalities to data users (consumers) for datasets’ downloading, and to data providers for uploading datasets. They do not support improvements of their published datasets by their users (e.g. through cleaning and further processing), or feedback provision by datasets’ users to their providers so that the latter can understand better the needs of the former.

**The advanced second generation Web 2.0 OGD portals**: The advent of the Web 2.0 paradigm, which facilitates the generation of content of various types by simple and non-expert users, the development of relationships and online communities among them, and the extensive interaction, collaboration and sharing of content and
information, has led to the emergence of a second generation of OGD portals, which have been influenced by these Web 2.0 principles. They provide (in addition to the abovementioned basic functionalities of the traditional first generation OGD portals) functionalities for commenting and rating datasets, forming groups around common interests, visualizing and processing datasets, improving or adapting them to specialized needs, and then publishing them again, uploading new datasets, enabling OGD users to become data ‘pro-sumers’ (both consuming and producing datasets). Their main objective is to support and facilitate extensive communication between OGD users (citizens, journalists, businesses, scientists, etc.) and providers (government agencies), and also collaborative value generation from OGD.

Based on the above categorisations and following the developments presented in this study we are in place to create a maturity model categorising the capabilities of OGD infrastructures.

<table>
<thead>
<tr>
<th>Time</th>
<th>Traditional OGD Infrastructures</th>
<th>Advanced OGD Infrastructures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet presence</td>
<td>OGD existence in silos accessed by application</td>
<td>OGD web presence</td>
</tr>
<tr>
<td>Functionality</td>
<td>N/A</td>
<td>Basic Web 1.0</td>
</tr>
<tr>
<td>Type</td>
<td>N/A</td>
<td>OGD direct provision portals</td>
</tr>
<tr>
<td>Stakeholders</td>
<td>Distinction between Data Providers and Data Users</td>
<td>Distinction between Data Providers and Data Users</td>
</tr>
<tr>
<td>License</td>
<td>Custom or N/A</td>
<td>Custom or N/A</td>
</tr>
<tr>
<td>Format</td>
<td>.xls, .pdf</td>
<td>html,.xls,.pdf</td>
</tr>
<tr>
<td>Metadata</td>
<td>Metadata Ignorance or Closed Metadata</td>
<td>Metadata Ignorance or Closed Metadata</td>
</tr>
<tr>
<td>Open Government level</td>
<td>Initial: Information broadcasting</td>
<td>Data Transparency: processes and performance</td>
</tr>
</tbody>
</table>
8.6.4. Multi-disciplinary Research Based on OGD

Many times in this study, it was emphasized that the most important and socially beneficial 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 the following Table.

<table>
<thead>
<tr>
<th>Societal Challenge</th>
<th>ICT-enabled Governance Research Topic</th>
<th>OGD Research Topic</th>
<th>Neighbouring Scientific Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language divide and lack of cross-communities communication</td>
<td>• Language and Cultural Interoperability</td>
<td>• Metadata for OGD</td>
<td>• Information Intelligence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Multilinguality</td>
<td>• Computer Science (Translation tools)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Controlled Vocabularies and Codelists Preservation</td>
<td>• Behavioural sciences</td>
</tr>
<tr>
<td>Anticipating unexpected crises</td>
<td>• Social – Economic Simulation Models</td>
<td>• Semantic Annotation</td>
<td>• Social and economic sciences</td>
</tr>
<tr>
<td></td>
<td>• Policy Modelling</td>
<td>• Organisational Interoperability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Process Optimization for OGD (Accurate provision)</td>
<td>• Sensor-generated open data</td>
<td></td>
</tr>
<tr>
<td>Enhanced collective cognitive intelligence (human / ICT-enabled) for better Governance</td>
<td>• Modelling and Simulation</td>
<td>• OGD Mining</td>
<td>• Economics</td>
</tr>
<tr>
<td></td>
<td>• Policy Analysis</td>
<td>• Citizen-generated open data</td>
<td>• Mathematics</td>
</tr>
<tr>
<td></td>
<td>• Identity Management</td>
<td>• Visualization</td>
<td>• Sociology</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Information Management</td>
<td>• Computer Science</td>
</tr>
</tbody>
</table>
Appendix A: References


Behkamal, B., Kahani, M., and Akbari, M. K. Customizing ISO 9126 quality model for evaluation of


CodagnoneChristiano& Osimo David (2008)


European Commission (2009). ‘Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – ICT


European Commission (2011a), Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. Open data. An engine for innovation, growth and transparent governance, European Commission, Brussels, COM(2011) 882 final, 2011.


http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf


Harrison, T. M., S. Guerrero, G. B. Burke, M. Cook, A. Cresswell, N. Helbig, J. Hrdinova, and T. Pardo


Janssen, K., (2011 a), "Legal interoperability – barriers to the harmonization of licences" presented at
the ICRI - Share PSI workshop, Brussels 2011


International Conference on Dublin Core and Metadata Applications (pp. pp-193).


The Open Definition (2014), http://opendefinition.org/


717–733.


Electronic Commerce Research, Vol. 9, No 2, pp. I-XIII.


Appendix B: Author’s Publications

B1. Journal papers


B2. Conference Papers


B3. Workshop papers


B4. Chapters in Books


B5. Book Authorships

Appendix C: Scenarios based Design Template

<table>
<thead>
<tr>
<th>Use Case Description: <em>give a textual description of the use case.</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Goal: <em>what is the overall objective of the use case?</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Topic: <em>what is the main subject areas of the use case? E.g. social science, environment?</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Data sets: <em>what are the sets of data involved, with their size, location etc?</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Data Services: <em>what are the services which provide the data, with their location? Is Linked Data involved?</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Data owners: <em>who has ownership and control of the data?</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Data Access Rights: <em>what are the rights and conditions of use of the data?</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Data Formats: <em>what are the data formats involved, e.g. XML, RDF, data models etc?</em></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Metadata formats: <em>what are the metadata formats and standards used to describe the data?</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nationalities and Languages involved: what countries or regions are involved in the use case? What natural languages is used for the data and metadata?</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Primary Actors and Roles: who are the people and organisations directly involved in the use case? What are their roles?</td>
</tr>
<tr>
<td>Secondary Actors and Roles: who are the people and organisations indirectly involved in the use case? e.g. subject groups, target audience What are their roles?</td>
</tr>
<tr>
<td>Trigger: why is the use case initiated?</td>
</tr>
<tr>
<td>Pre-condition list: what conditions must apply before the use case can start?</td>
</tr>
<tr>
<td>Post-conditions list: what do we expect to be any new conditions resulting at the end?</td>
</tr>
<tr>
<td>Primary Flow: what are the steps in the execution of the use case?</td>
</tr>
<tr>
<td>Error Flow: what error states can arise? How are they managed?</td>
</tr>
<tr>
<td>Additional Requirements list: any additional requirements not covered above</td>
</tr>
<tr>
<td>Notes &amp; Outstanding Issues: comments or issues arising in the use case which need further consideration</td>
</tr>
<tr>
<td>Use Case/Flow Chart Diagrams (UML): a visual representation of the use case using one or more UML diagrams</td>
</tr>
</tbody>
</table>
Appendix D: List of European Open Government Data Initiatives

<table>
<thead>
<tr>
<th>Initiative</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cities</td>
<td></td>
</tr>
<tr>
<td>Arvada</td>
<td><a href="http://arvada.org/opendata/">http://arvada.org/opendata/</a></td>
</tr>
<tr>
<td>Baltimore, Maryland</td>
<td><a href="http://data.baltimorecity.gov">http://data.baltimorecity.gov</a></td>
</tr>
<tr>
<td>Boston</td>
<td><a href="http://www.cityofboston.gov/mis/databoston/data.aspx">http://www.cityofboston.gov/mis/databoston/data.aspx</a></td>
</tr>
<tr>
<td>Calgary</td>
<td><a href="http://www.calgary.ca">www.calgary.ca</a></td>
</tr>
<tr>
<td>Chicago</td>
<td><a href="http://data.cityofchicago.org">http://data.cityofchicago.org</a></td>
</tr>
<tr>
<td>City &amp; County of Honolulu</td>
<td><a href="http://can-do.honolulu.gov/">http://can-do.honolulu.gov/</a></td>
</tr>
<tr>
<td>Edmonton</td>
<td><a href="http://data.edmonton.ca/">http://data.edmonton.ca/</a></td>
</tr>
<tr>
<td>Helsinki Region, Finland</td>
<td><a href="http://www.hri.fi/en/">http://www.hri.fi/en/</a></td>
</tr>
<tr>
<td>Hong Kong</td>
<td><a href="http://www.gov.hk/en/theme/psidatadownload">http://www.gov.hk/en/theme/psidatadownload</a></td>
</tr>
<tr>
<td>Lexington, KY</td>
<td><a href="http://www.lexingtonky.gov/">http://www.lexingtonky.gov/</a></td>
</tr>
<tr>
<td>Lichfield, UK</td>
<td><a href="http://lichfielddc.gov.uk">http://lichfielddc.gov.uk</a></td>
</tr>
<tr>
<td>London</td>
<td><a href="http://data.london.gov.uk/">http://data.london.gov.uk/</a></td>
</tr>
<tr>
<td>London, ON</td>
<td><a href="http://www.opendatalondon.ca">http://www.opendatalondon.ca</a></td>
</tr>
<tr>
<td>Manchester, UK</td>
<td><a href="http://www.manchester.gov.uk/opendata">www.manchester.gov.uk/opendata</a></td>
</tr>
<tr>
<td>Montevideo, Uruguay</td>
<td><a href="http://www.imm.gub.uy/node/13757">http://www.imm.gub.uy/node/13757</a></td>
</tr>
<tr>
<td>Montreal</td>
<td><a href="http://montrealouverture.net">http://montrealouverture.net</a></td>
</tr>
<tr>
<td>Nanaimo, B.C.</td>
<td><a href="http://www.nanaimo.ca/datafeeds/">http://www.nanaimo.ca/datafeeds/</a></td>
</tr>
<tr>
<td>Ottawa</td>
<td><a href="http://ottawa.ca/online_services/opendata/info/index_en.html">http://ottawa.ca/online_services/opendata/info/index_en.html</a></td>
</tr>
<tr>
<td>Paris, France</td>
<td><a href="http://opendata.paris.fr/opendata">http://opendata.paris.fr/opendata</a></td>
</tr>
<tr>
<td>Philadelphia</td>
<td><a href="http://opendataphilly.org/">http://opendataphilly.org/</a></td>
</tr>
<tr>
<td>Portland</td>
<td><a href="http://civicapps.org/datasets">http://civicapps.org/datasets</a></td>
</tr>
<tr>
<td>Initiative</td>
<td>URL</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
</tr>
<tr>
<td>San Francisco</td>
<td><a href="http://datasf.org/">http://datasf.org/</a></td>
</tr>
<tr>
<td>Seattle</td>
<td><a href="http://data.seattle.gov/">http://data.seattle.gov/</a></td>
</tr>
<tr>
<td>Toronto</td>
<td><a href="http://toronto.ca/open">http://toronto.ca/open</a></td>
</tr>
<tr>
<td>Vancouver, B.C.</td>
<td><a href="http://data.vancouver.ca/">http://data.vancouver.ca/</a></td>
</tr>
<tr>
<td>Warwickshire, UK</td>
<td><a href="http://www.warwickshire.gov.uk/home">http://www.warwickshire.gov.uk/home</a></td>
</tr>
<tr>
<td><strong>States, Provinces, and other Regions</strong></td>
<td></td>
</tr>
<tr>
<td>Basque Government, Spain</td>
<td><a href="http://opendata.euskadi.net/w79-home/es/">http://opendata.euskadi.net/w79-home/es/</a></td>
</tr>
<tr>
<td>British Columbia, Canada</td>
<td><a href="http://data.gov.bc.ca/">http://data.gov.bc.ca/</a></td>
</tr>
<tr>
<td>California, USA</td>
<td><a href="http://data.ca.gov">http://data.ca.gov</a></td>
</tr>
<tr>
<td>Colorado, USA and Overall summary with Chief Data Officer</td>
<td><a href="http://coloradoopendata.org/">http://coloradoopendata.org/</a></td>
</tr>
<tr>
<td>Greater New Orleans Community Data Center</td>
<td><a href="http://www.gnocdc.org/">http://www.gnocdc.org/</a></td>
</tr>
<tr>
<td>Illinois Data Exchange Affiliates including Metro Chicago Information Center</td>
<td><a href="http://data.cmap.illinois.gov/">http://data.cmap.illinois.gov/</a></td>
</tr>
<tr>
<td>Indiana Coalition for Open Government (ICOG)</td>
<td><a href="http://www.indianacog.org/main.php">http://www.indianacog.org/main.php</a></td>
</tr>
<tr>
<td>King County, WA, USA</td>
<td><a href="http://www.datakc.org/">http://www.datakc.org/</a></td>
</tr>
<tr>
<td>Maine, USA</td>
<td><a href="http://www.maine.gov/cgi-bin/data/index.pl">http://www.maine.gov/cgi-bin/data/index.pl</a></td>
</tr>
<tr>
<td>Massachusetts Campaign</td>
<td><a href="http://www.maopengov.org/">http://www.maopengov.org/</a></td>
</tr>
<tr>
<td>Initiative</td>
<td>URL</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>for Open Government</td>
<td><a href="http://wiki.state.ma.us/confluence/display/data/Data+Catalog">http://wiki.state.ma.us/confluence/display/data/Data+Catalog</a></td>
</tr>
<tr>
<td>Massachusetts, USA</td>
<td><a href="http://dwexternal.co.mecklenburg.nc.us/ids/Nav_Reports.aspx/">http://dwexternal.co.mecklenburg.nc.us/ids/Nav_Reports.aspx/</a></td>
</tr>
<tr>
<td>Mecklenburg County, NC, USA</td>
<td><a href="http://www.michigan.gov/som/0,1607,7-192-29938_54272---,00.html">http://www.michigan.gov/som/0,1607,7-192-29938_54272---,00.html</a></td>
</tr>
<tr>
<td>Michigan, USA</td>
<td><a href="http://www.mo.gov/data/">http://www.mo.gov/data/</a></td>
</tr>
<tr>
<td>Missouri, USA</td>
<td><a href="http://www.nebraska.gov/data/">http://www.nebraska.gov/data/</a></td>
</tr>
<tr>
<td>Nebraska, USA</td>
<td><a href="http://www.nj.gov/transparency/">http://www.nj.gov/transparency/</a></td>
</tr>
<tr>
<td>New Jersey, NJ</td>
<td><a href="http://www.nj.gov/transparency/">http://www.nj.gov/transparency/</a></td>
</tr>
<tr>
<td>North Dakota, USA</td>
<td><a href="http://www.nd.gov/gis/">http://www.nd.gov/gis/</a></td>
</tr>
<tr>
<td>Oklahoma, USA</td>
<td><a href="http://www.ok.gov/about/data.html">http://www.ok.gov/about/data.html</a></td>
</tr>
<tr>
<td>Open New York Forum</td>
<td><a href="http://opennyforum.org/">http://opennyforum.org/</a></td>
</tr>
<tr>
<td>Oregon, USA</td>
<td><a href="http://data.oregon.gov/">http://data.oregon.gov/</a></td>
</tr>
<tr>
<td>Rhode Island, USA</td>
<td><a href="http://www.ri.gov/data/">http://www.ri.gov/data/</a></td>
</tr>
<tr>
<td>Texas, USA (State Comptroller) and Texas, USA (Texas.gov)</td>
<td><a href="http://www.texas">http://www.texas</a> transparency.org/opendata/</td>
</tr>
<tr>
<td>Utah, USA</td>
<td><a href="http://www.utah.gov/data/">http://www.utah.gov/data/</a></td>
</tr>
<tr>
<td>Washington Coalition for Open Government</td>
<td><a href="http://www.washingtoncog.org/">http://www.washingtoncog.org/</a></td>
</tr>
<tr>
<td>Washington, USA</td>
<td><a href="http://data.wa.gov/">http://data.wa.gov/</a></td>
</tr>
<tr>
<td>Nations</td>
<td><a href="http://data.australia.gov.au">http://data.australia.gov.au</a></td>
</tr>
<tr>
<td>Australia</td>
<td><a href="http://data.gc.ca">http://data.gc.ca</a></td>
</tr>
<tr>
<td>Canada</td>
<td><a href="http://data.wa.gov/">http://data.wa.gov/</a></td>
</tr>
<tr>
<td>Canada</td>
<td><a href="http://datadotgc.ca/">http://datadotgc.ca/</a></td>
</tr>
<tr>
<td>Denmark</td>
<td><a href="http://digitaliser.dk/ressourcer">http://digitaliser.dk/ressourcer</a></td>
</tr>
<tr>
<td>Ireland StatCentral and</td>
<td><a href="http://www.statcentral.ie/">http://www.statcentral.ie/</a></td>
</tr>
<tr>
<td>Initiative</td>
<td>URL</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>Open Data Northern Ireland</td>
<td></td>
</tr>
<tr>
<td>Israel</td>
<td><a href="http://meida.org.il/">http://meida.org.il/</a></td>
</tr>
<tr>
<td>Japan</td>
<td><a href="http://openlabs.go.jp">http://openlabs.go.jp</a></td>
</tr>
<tr>
<td>Kenya</td>
<td><a href="http://www.opendata.go.ke/">http://www.opendata.go.ke/</a></td>
</tr>
<tr>
<td>New Zealand</td>
<td><a href="http://www.data.govt.nz/">http://www.data.govt.nz/</a></td>
</tr>
<tr>
<td>Norway</td>
<td><a href="http://data.norge.no/">http://data.norge.no/</a></td>
</tr>
<tr>
<td>OpenGovData.ru and GovWeb.ru in Russia</td>
<td><a href="http://opengovdata.ru">http://opengovdata.ru</a></td>
</tr>
<tr>
<td>Slovakia</td>
<td><a href="http://datanest.fairplay.sk/pages/index">http://datanest.fairplay.sk/pages/index</a></td>
</tr>
<tr>
<td>Sweden</td>
<td><a href="http://www.opengov.se/">http://www.opengov.se/</a></td>
</tr>
<tr>
<td>Switzerland</td>
<td><a href="http://opendata.ch/">http://opendata.ch/</a></td>
</tr>
<tr>
<td>United Kingdom</td>
<td><a href="http://data.gov.uk">http://data.gov.uk</a></td>
</tr>
<tr>
<td>United States of America</td>
<td><a href="http://www.data.gov/">http://www.data.gov/</a></td>
</tr>
</tbody>
</table>
## Appendix E: List of Greek Open Government Data Sources

<table>
<thead>
<tr>
<th>No.</th>
<th>Government Agency</th>
<th>Description</th>
<th>Thematic Categories</th>
<th>URL</th>
</tr>
</thead>
</table>
| 1   | Transparency Program                                   | Laws / Regulations / Ministry Decisions & Public Bodies Spending             | Economic and Business Information; Legal Information | http://diavgeia.gov.gr  
                               |                                                        |                                              |                                                                           | http://greek-lod.math.auth.gr/diavgeia/ |
| 2   | General Secretariat of Information Systems             | Taxation Statistics for Greece                                              | Economic and Business Information          | http://www.gsis.gr                                                   |
| 3   | Geodata project                                        | A project funded by Greek government in order to host geospatial public cross-sector information | Geographic Information                    | http://geodata.gov.gr                                                  |
| 4   | Hellenic Statistical Authority                         | The official source of statistical information for almost every activity of Greek economy | Social Information; Economic and Business Information | http://www.statistics.gr                                               |
| 5   | Ministry of Finance                                    | Country's Economic Data                                                    | Economic and Business Information; Legal Information | http://www.minfin.gr                                                  |
| 6   | Ministry of Citizen Protection - Greek Fire Department | Provide information on fires that have been recorded in Greece              | Legal Information; Geographic Information  | http://www.fireservice.gr                                              
                               |                                                        |                                              |                                                                           | http://greek-lod.math.auth.gr/fire-brigade/ |
| 7   | Ministry of Labour and Social Insurance                | Sharing legal information on employment rights and decisions by the ministry | Legal Information                          | http://www.ypakp.gr                                                   |
| 8   | Ministry of Education, Lifelong Learning and Religion  | Provide information of the location of all schools in Greece                | Geographic Information                    | http://www.minedu.gov.gr/                                             |
| 9   | Ministry of Citizen Protection - Greek Police Department| Sharing information on traffic accidents and criminality                    | Legal Information; Geographic Information  | http://www.astynomia.gr                                              
                               |                                                        |                                              |                                                                           | http://greek-lod.math.auth.gr/police/     |
| 10  | Ministry of Interior, Decentralisation and e-Government| Census about Public Sector Employees                                        | Social Information; Economic and Business Information | http://www.ydmed.gov.gr/                                             
<pre><code>                           |                                                        |                                              |                                                                           | http://greek-lod.math.auth.gr/kalikratis/ |
</code></pre>
<table>
<thead>
<tr>
<th></th>
<th>Ministry of Infrastructures and Transportations</th>
<th>Detail Invoices concerning Public Works</th>
<th>Economic and Business Information; Legal Information; Traffic and Transport Information</th>
<th><a href="http://www.yme.gr/">http://www.yme.gr/</a></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ministry of Health</td>
<td>Provide aggregated information about surgeries and tests conducted in hospitals</td>
<td>Social Information</td>
<td><a href="http://yyka.gov.gr">http://yyka.gov.gr</a></td>
</tr>
<tr>
<td></td>
<td>Hellenic Ministry of Rural Development and Food</td>
<td>Subsidy provision per prefecture and financial affairs</td>
<td>Economic and Business Information</td>
<td><a href="http://www.minagric.gr">http://www.minagric.gr</a></td>
</tr>
<tr>
<td></td>
<td>Ministry of Environmental Protection and Climate Change</td>
<td>Information about protected areas, such as Mountain Protection Zones, National Parks and Forests Borderline, Controlled Hunting Regions, etc.</td>
<td>Meteorological and Environmental Information; Geographic Information; Natural Resources Information</td>
<td><a href="http://www.ypeka.gr">http://www.ypeka.gr</a></td>
</tr>
<tr>
<td></td>
<td>Observatory for Digital Greece</td>
<td>Key point of reference for accurate and up-to-date information on Information Society indicators</td>
<td>Economic and Business Information; Social Information</td>
<td><a href="http://www.observatory.gr/">http://www.observatory.gr/</a></td>
</tr>
<tr>
<td></td>
<td>National Observatory for SMEs</td>
<td>Various types of data on SMEs</td>
<td>Economic and Business Information; Social Information</td>
<td><a href="http://pforumgr.eommex.gr">http://pforumgr.eommex.gr</a></td>
</tr>
<tr>
<td></td>
<td>Greek LOD Cloud</td>
<td>First attempt for provision of Open Linked data (not only governmental) (Police, Fire-Brigade, Diavgeia (Transparency programme), Kalikratis)</td>
<td>Social Information; Geographic Information; Economic and Business Information; Legal Information</td>
<td><a href="http://greek-lod.math.auth.gr/">http://greek-lod.math.auth.gr/</a></td>
</tr>
<tr>
<td></td>
<td>Athens Urban Transport Organisation</td>
<td>Public transport routes and stops</td>
<td>Traffic and Transport Information; Economic and</td>
<td><a href="http://www.oasa.gr">http://www.oasa.gr</a></td>
</tr>
<tr>
<td>Page</td>
<td>Government Department</td>
<td>Business Information</td>
<td>Data Source</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----------------------</td>
<td>----------------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Archaeological Cadastre</td>
<td>Quantitative and Geospatial Data on Archaeological Sites</td>
<td>Geographic Information</td>
<td><a href="http://archaeocadastre.culture.gr/el/data">http://archaeocadastre.culture.gr/el/data</a></td>
</tr>
<tr>
<td>21</td>
<td>Ministry of Aegean</td>
<td>Information on various topics concerning Aegean islands</td>
<td>Natural Resources Information; Agricultural, Farming, Forestry and Fisheries Information; Tourist and Leisure Information</td>
<td><a href="http://www.ypai.gr/">http://www.ypai.gr/</a></td>
</tr>
<tr>
<td>22</td>
<td>Ministry of Development</td>
<td>Information about public works</td>
<td>Economic and Business Information</td>
<td><a href="http://www.ypoian.gr/">http://www.ypoian.gr/</a></td>
</tr>
<tr>
<td>23</td>
<td>General Secretariat for Development</td>
<td>Information about public works</td>
<td>Economic and Business Information</td>
<td><a href="http://www.gsrt.gr/">http://www.gsrt.gr/</a></td>
</tr>
<tr>
<td>27</td>
<td>Ministry of Education and Religious Affairs</td>
<td>Information about educational system, public schools, public universities, educational laws</td>
<td>Legal Information</td>
<td><a href="http://www.ypepth.gr/">http://www.ypepth.gr/</a></td>
</tr>
<tr>
<td>28</td>
<td>Ministry of Maritime Affairs, islands and fisheries</td>
<td>Information about Greek islands, fisheries and sea affairs</td>
<td>Agricultural, Farming, Forestry and Fisheries Information</td>
<td><a href="http://www.yen.gr">http://www.yen.gr</a></td>
</tr>
<tr>
<td>29</td>
<td>Ministry of foreign Affairs</td>
<td>Information on foreign individuals, information about other countries</td>
<td>Legal Information</td>
<td><a href="http://www.mfa.gr/">http://www.mfa.gr/</a></td>
</tr>
<tr>
<td>31</td>
<td>Ministry of Macedonia and Thrace</td>
<td>Information about various topics concerning North Greece</td>
<td>Economic and Business Information</td>
<td><a href="http://www.mathra.gr/">http://www.mathra.gr/</a></td>
</tr>
<tr>
<td></td>
<td>Ministry of culture and Tourism</td>
<td>Information about Greek culture and tourism</td>
<td>Economic and Business Information; Traffic and Transport Information; Tourist and Leisure Information</td>
<td><a href="http://www.culture.gr">http://www.culture.gr</a></td>
</tr>
<tr>
<td>---</td>
<td>--------------------------------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>33</td>
<td>Ministry of the Press and Media</td>
<td>Information about Media</td>
<td>Economic and Business Information</td>
<td><a href="http://www.minpress.gr/">http://www.minpress.gr/</a></td>
</tr>
<tr>
<td>34</td>
<td>Athens' Prefecture</td>
<td>Information about various topics concerning Athens</td>
<td>Economic and Business Information</td>
<td><a href="http://nom-athinas.gov.gr">http://nom-athinas.gov.gr</a></td>
</tr>
<tr>
<td>35</td>
<td>West Macedonia's Region</td>
<td>Information about various topics concerning West Macedonia</td>
<td>Economic and Business Information</td>
<td><a href="http://www.grevena.gov.gr/">http://www.grevena.gov.gr/</a></td>
</tr>
<tr>
<td>36</td>
<td>East Macedonia's and Thrace's Region</td>
<td>Information about various topics concerning East Macedonia and Thrace</td>
<td>Economic and Business Information</td>
<td><a href="http://www.pamth.gov.gr/">http://www.pamth.gov.gr/</a></td>
</tr>
<tr>
<td>37</td>
<td>South Aegean's Region</td>
<td>Information about various topics concerning South Aegean</td>
<td>Economic and Business Information</td>
<td><a href="http://www.notioaigaio.gr/per/">http://www.notioaigaio.gr/per/</a></td>
</tr>
<tr>
<td>38</td>
<td>Epirus' Region</td>
<td>Information about various topics concerning Epirus</td>
<td>Economic and Business Information</td>
<td><a href="http://www.epirus.gov.gr">http://www.epirus.gov.gr</a></td>
</tr>
<tr>
<td>39</td>
<td>Thessaly's Region</td>
<td>Information about various topics concerning Thessaly</td>
<td>Economic and Business Information</td>
<td><a href="http://www.thessalia.gov.gr/">http://www.thessalia.gov.gr/</a></td>
</tr>
<tr>
<td>40</td>
<td>Ionian Islands' Region</td>
<td>Information about various topics concerning Ionian Islands</td>
<td>Economic and Business Information</td>
<td><a href="http://www.pin.gov.gr/">http://www.pin.gov.gr/</a></td>
</tr>
<tr>
<td>41</td>
<td>West Greece Region</td>
<td>Information about various topics concerning Western Greece</td>
<td>Economic and Business Information</td>
<td><a href="http://www.pde.gov.gr/gr/index.php">http://www.pde.gov.gr/gr/index.php</a></td>
</tr>
<tr>
<td>42</td>
<td>Attica's Region</td>
<td>Information about various topics concerning Attica</td>
<td>Economic and Business Information</td>
<td><a href="http://www.attiki.gov.gr/">http://www.attiki.gov.gr/</a></td>
</tr>
<tr>
<td>43</td>
<td>Peloponnese Region</td>
<td>Information about various topics concerning Peloponnese</td>
<td>Economic and Business</td>
<td><a href="http://ppel.gov.gr/">http://ppel.gov.gr/</a></td>
</tr>
<tr>
<td>No.</td>
<td>Region or Area</td>
<td>Information about various topics concerning...</td>
<td>Economic and Business Information</td>
<td>Website Address</td>
</tr>
<tr>
<td>-----</td>
<td>------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>44</td>
<td>North Aegean's Region</td>
<td>Information about various topics concerning North Aegean</td>
<td>Economic and Business Information</td>
<td><a href="http://www.northaegean.gr/">http://www.northaegean.gr/</a></td>
</tr>
<tr>
<td>45</td>
<td>Crete's Region</td>
<td>Information about various topics concerning Crete</td>
<td>Economic and Business Information</td>
<td><a href="http://www.crete-region.gr/">http://www.crete-region.gr/</a></td>
</tr>
<tr>
<td>46</td>
<td>Central Macedonia's Region</td>
<td>Information about various topics concerning Central Macedonia</td>
<td>Economic and Business Information</td>
<td><a href="http://www.pkm.gov.gr/">http://www.pkm.gov.gr/</a></td>
</tr>
<tr>
<td>47</td>
<td>Municipality of Keratsini and Drapetsona</td>
<td>Information about various topics concerning Keratsini and Drapetsona</td>
<td>Economic and Business Information</td>
<td><a href="http://www.keratsini-drapetsona.gr/">http://www.keratsini-drapetsona.gr/</a></td>
</tr>
<tr>
<td>48</td>
<td>Municipality of Athens</td>
<td>Information about various topics concerning Athens</td>
<td>Economic and Business Information</td>
<td><a href="http://www.cityofathens.gr/">http://www.cityofathens.gr/</a></td>
</tr>
<tr>
<td>49</td>
<td>Municipality of Thessaloniki</td>
<td>Information about various topics concerning Thessaloniki</td>
<td>Economic and Business Information</td>
<td><a href="http://www.thessaloniki.gr">http://www.thessaloniki.gr</a></td>
</tr>
<tr>
<td>50</td>
<td>Municipality of Chania</td>
<td>Information about various topics concerning Chania of Crete</td>
<td>Economic and Business Information</td>
<td><a href="http://www.chania.gr/">http://www.chania.gr/</a></td>
</tr>
<tr>
<td>51</td>
<td>Municipality of Rethymno</td>
<td>Information about various topics concerning Rethymno of Crete</td>
<td>Economic and Business Information</td>
<td><a href="http://www.rethymno.gr/">http://www.rethymno.gr/</a></td>
</tr>
<tr>
<td>52</td>
<td>Municipality of Heraklion</td>
<td>Information about various topics concerning Heraklion of Crete</td>
<td>Economic and Business Information</td>
<td><a href="http://www.heraklion.gr/">http://www.heraklion.gr/</a></td>
</tr>
<tr>
<td>53</td>
<td>Municipality of Larissa</td>
<td>Information about various topics concerning Larissa</td>
<td>Economic and Business Information</td>
<td><a href="http://www.larissa-dimos.gr/">http://www.larissa-dimos.gr/</a></td>
</tr>
<tr>
<td>54</td>
<td>Municipality of Messini</td>
<td>Information about various topics concerning Messini</td>
<td>Economic and Business Information</td>
<td><a href="http://www.messini.gr">http://www.messini.gr</a></td>
</tr>
<tr>
<td>55</td>
<td>Municipality of Kozani</td>
<td>Information about various topics concerning Kozani</td>
<td>Economic and Business Information</td>
<td><a href="http://www.kozanh.gr">http://www.kozanh.gr</a></td>
</tr>
<tr>
<td>56</td>
<td>Municipality of Samos</td>
<td>Information about various topics concerning Samos</td>
<td>Economic and Business Information</td>
<td><a href="http://www.vathi.gr/">http://www.vathi.gr/</a></td>
</tr>
<tr>
<td>No.</td>
<td>Municipality</td>
<td>Information</td>
<td>Economic and Business Information</td>
<td>URL</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>57</td>
<td>Municipality of Chios</td>
<td>Information about various topics concerning Chios</td>
<td>Economic and Business Information</td>
<td><a href="http://www.chioscity.gr/">http://www.chioscity.gr/</a></td>
</tr>
<tr>
<td>58</td>
<td>Municipality of Mytilene</td>
<td>Information about various topics concerning Mytilene</td>
<td>Economic and Business Information</td>
<td><a href="http://www.mytilene.gr/">http://www.mytilene.gr/</a></td>
</tr>
<tr>
<td>59</td>
<td>Municipality of Halkidiki</td>
<td>Information about various topics concerning Halkidiki</td>
<td>Economic and Business Information</td>
<td><a href="http://www.halkidiki.gov.gr">http://www.halkidiki.gov.gr</a></td>
</tr>
<tr>
<td>60</td>
<td>Municipality of Drama</td>
<td>Information about various topics concerning Drama</td>
<td>Economic and Business Information</td>
<td><a href="http://www.dimos-dramas.gr">http://www.dimos-dramas.gr</a></td>
</tr>
</tbody>
</table>

About the Author

Charalampos Harris Alexopoulos is a Research Associate in the Information Systems Laboratory of the Department of Information and Communications Systems Engineering at the University of the Aegean, working on European and National funded research and pilot application projects (ENGAGE, SHARE-PSI 2.0, EU-COMMUNITY, PADGETS, NOMAD, NET-EUCEN, PLUG-IN) for governments and enterprises and publishing at several conferences and journals on open government data, IS evaluation, interoperability and e-government.

In parallel, he is responsible for the International Relations of the Innovation and Entrepreneurship Unit of the University, coordinating collaboration activities. Harris serves as Programme and Organisation Committee Member of the annual Samos Summit on ICT-enabled Governance. He is also a constant Programme and Organisation Committee Member of the “Open & Collaborative Governance” and “Big Data Analysis on Earth Sciences” summer schools. Harris is a computer science graduate from the University of Peloponnese with an MSc in Management Information Systems from the University of the Aegean.

In 2015, Harris was ranked as one of the most prolific researchers in open data research by Hossain, Dwivedi and Rana (2015). He was a WeGov 2010 Awards Finalist in the category of best idea within the PCI 2010 14th PanHellenic conference on informatics with the “Public Services Transformation Methodology” and a WeGov 2011 Awards Finalist in the category of best application, with 4-Delta Application: Production and Management of Public Sector Evaluation Procedures.