LAPPEENRANTA UNIVERSITY OF TECHNOLOGY
School of Business and Management
Degree Program in Computer Science

Master's Thesis

Norismiza Ismail

A SYSTEMATIC MAPPING STUDY ON OPEN DATA

Lappeenranta, October 19, 2015

Supervisors: Associate Professor, Ph.D. Uolevi Nikula
Researcher, Ph.D. Andrey Maglyas

ABSTRACT

Lappeenranta University of Technology
School of Business and Management
Degree Program in Computer Science

Norismiza Ismail

MASTER's THESIS
A SYSTEMATIC MAPPING STUDY ON OPEN DATA

2015

104 pages, 24 figures, 19 tables and 2 Appendixes

Supervisors: *Associate Professor, Ph.D. Uolevi Nikula*
Researcher, Ph.D. Andrey Maglyas

Keywords: Open Data, Systematic Mapping Study, Systematic Literature Review

This thesis presented the overview of Open Data research area, quantity of evidence and establishes the research evidence based on the Systematic Mapping Study (SMS). There are 621 such publications were identified published between years 2005 and 2014, but only 243 were selected in the review process. This thesis highlights the implications of Open Data principals' proliferation in the emerging era of the accessibility, reusability and sustainability of data transparency. The findings of mapping study are described in quantitative and qualitative measurement based on the organization affiliation, countries, year of publications, research method, star rating and units of analysis identified. Furthermore, units of analysis were categorized by development lifecycle, linked open data, type of data, technical platforms, organizations, ontology and semantic, adoption and awareness, intermediaries, security and privacy and supply of data which are important component to provide a quality open data applications and services. The results of the mapping study help the organizations (such as academia, government and industries), researchers and software developers to understand the existing trend of open data, latest research development and the demand of future research. In addition, the proposed conceptual framework of Open Data research can be adopted and expanded to strengthen and improved current open data applications.

Preface

This Master's thesis was carried out during 2014-2015 at Lappeenranta University of Technology, Finland.

Acknowledgements

This work would have been impossible without the help and guidance of several people, whose contribution I would like to acknowledge. First of all, I would like to express my deepest gratitude to my supervisor, Associate Professor Dr. Uolevi Nikula for accepting and giving me this wonderful research topic. Your encouragement and enthusiasm have been source of inspiration which kept me going forward and at the same time provided much appreciated freedom and support to explore new ways and concepts. I am also thankful to the co-supervisor, Dr. Andrey Maglyas for reading the thesis and I appreciate your valuable comments. I am grateful for all the supports they provided during this entire research and the opportunity given to work on this interesting Master's Thesis of Open Data.

Many thanks to Suvi Tiainen, Arttu Hanska and all wonderful people at Department of Innovation and Software for the pleasant support.

I would like to express my appreciation to Universiti Malaysia Perlis (UniMAP), Brig. Gen. Datuk Professor Emeritus Dr. Kamarudin Hussin, Datin Noridah Yangman, Professor Datuk Dr. Zul Azhar, Professor Datin Dr. Zuraidah Mohamad Zain and Miss Norsyahiza Hamzah for the encouragement of my studies towards Master's Degree in Finland.

Special thanks to all my friends and their families, Professor Aki Mikkola, Madam Hanna Lommi, Dr. Norsuria, Dr. Azremi, Dr. Rafi, Normiza, Dr. Behnam, Zahra, Ummi, Azhan, Aida, Vina, Jussi, Roziyah, Norlaila, Yongyi, Fahad, Sahar, Mihai, Saiida, Meharullah and staffs of Malaysian Embassy for their outstanding support, encouragement and concerns during my stay in Finland.

I extend my gratitude to my beloved parents for their continuous prayers, encouragement and support, Haji Ismail, Hajah Umi, Haji Baharudin and Madam Sofiah. My brothers and sisters, Izwan, Hafizul, Fadhli, Ahmadi, Fitin, Niza, Aliaa' and Miza for their support.

Last but not least, I would like to express my heartfelt gratitude to my beloved family, wonderful husband Mohamad Ezral Baharudin and beautiful children, Eryssa Nur Iman and Ezz Eilman for their unconditional love and endless supports. Their patience made it all possible and without them I wouldn't be where I am now.

All the praises and thanks to the almighty Allah, The Most Gracious, The Most Merciful, who guide me in every step I take.

Lappeenranta, October 19th, 2015

Norismiza Ismail

Table of Content

Abbreviations

API	Application Programming Interface
CKAN	Comprehensive Knowledge Archive Network
CSV	Comma Separated Values
DaPaaS	Data and Platform-as-a-Services
GUI	Graphical User Interface
HTML	HyperText Markup Language
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
LOD	Linked Open Data
NDSAP	National Data Sharing and Accessibility Policy
OGD	Open Government Data
OGDI	Open Government Data Initiative
OWL	Ontology Web Language
PDF	Portable Document Format
RDBMS	Relational Database Management Systems
RDF	Resource Description Framework
SLR	Systematic Literature Review
REST	Representational State Transfer
RTF	Rich Text Format
SDK	Software Development Kit
SKOS	Simple Knowledge Organization System
SMS	Systematic Mapping Study
SOAP	Simple Object Access Protocol
SPSS	Statistical Package for the Social Science
SQL	Structured Query Language
TXT	Text File
URI	Universal Resource Identifiers
W3C	World Wide Web Consortium
XML	Extensible Markup Language

1 Introduction

1.1 Background

This research is based on the proliferation of Open Data concept that contributes to the emerging era of data transparency recently. Open Data principals exploring the potential implementations of accessibility, reusability and sustainability of transparent datasets in standardized formats, no restrictions, participation and engagement of humans and machines; which is so called interoperability, allowing different components, systems and organizations working together worldwide [1,2].

The implementation of Open Data in a useful manner benefits the society by increasing the transparency level, reproducibility and hence more efficient scientific process can be produced [4]. The Open Data movement had grown remarkably since 2009 when the United States Government decided to implement openness principle by releasing thousands of their datasets. Later, followed by European Commission, Mexico, and Singapore opening the spigots of readily usable public data [5, 6, 7, 8, 11]. Besides, the emerging Open Data increase the benefits for industries and academic discussions especially in services environment [13]. This can be observed from the Open Data services value of network such as co-creation by saving costs, new services or user interface creation by utilizing different sources, raw or community processed Open Data and active Data Visualization.

The important aspect of Open Data is trustworthiness as it has been shown that trust itself is the trans-disciplinary result of technical, sociological and legal aspects [14]. As the result of the overwhelming adoption of Internet technologies in almost every private, public, economic and social sector, trust and trustworthiness are the central notions in such networked environments.

Various types of Web 2.0 based technologies have been used, including raw data download, open Application Programming Interface (API) and Linked Open Data (LOD) [12] to achieve the concept of Open Data. Technically Open Data needs to be in a linked

format or computer-readable format such as comma-separated values (.csv), Excel spread sheet (.xls) or PC-axis (.px), websites and text documents. However, scanned documents (.pdf) or image files are not considered as machine-readable in the 5 Star data definition [24]. Semantic web technologies and Linked Data have been seen the best approach in publishing (expose, share and connect) a large amount of data on the web based on World Wide Web Consortium (W3C) standards [15].

The accessibility, reusability and sustainability of open data in computing evolution are the advantages of this research. Furthermore, latest technology from the catalyst of open data such as Semantic Web, Linked Data and Cloud Computing has emerged for new proximity services. As an example, the data transparency and new mobility technologies like so called Data and Platform-as-a-Services (DaPaaS) which the research funded by the European Commission between 2013 and 2015 [156]. DaPaaS approach goal is to make efficient the Open Data publication and consumption. In addition, the developers can publish and host data-sets and data-intensive applications which lead being accessed by the end-user applications in a cross-platform manner.

1.2 Motivation

In this research, the determinant factors of Open Data innovation are identified based on a Systematic Mapping Study, which is done by developing a framework on how to produce a good Open Data systems or applications. Units of analysis can be identified from the mapping study. After further research, many Open Data platforms, applications or systems still observed as poor and not good enough for providing quality data and services to the public which led to a failure system [62]. The main motivation of the study is the need to improve the novelty of Open Data innovation and identify areas within this topic from primary studies by doing an extensive literature study. The results are expected to help practitioners and researchers by providing them with more information.

1.3 Objectives and Restrictions

The result expected in this research project would give real impact to the computing society, in particular for those who are dealing with the open data revolution. The knowledge and solution approach were obtained from previous researches that were carried out in the consortium. Hopefully, the result and knowledge of this research would be distributed widely to get more feedback for improvement.

This study is based on the systematic literature review guidelines defined by Kitchenham and Charters [29] to identify the state of the current research. In detail, the review can be categorized as a Systematic Mapping Study (SMS) or scoping study. While mapping study as the goal, the quantity of evidence can be indicated by implementing a review process which presents the whole picture of the research area provided and the researchers pieces of evidence exist on this topic were established [29]. In addition, while identifying the research areas where primary studies are required, it provides an exploration of the existing studies on Open Data.

The review process starts by developing the review protocol in which all steps, research questions, inclusion and exclusion criteria, and analysis procedures are included. The search strategy identified 621 publications, of which 243 publications were included in the review as primary study papers and 11 studies of these publications were from conference, symposia or workshop. The selected papers were published between 2005 and 2014. The articles have been studied and analysed to answers the research questions. Furthermore, potential validity to threats were also identified and assessed. The results of the review are presented with suggestions.

1.4 Structure of the Thesis

This thesis consisted of eight chapters. The structure of this thesis is as follows. The second chapter devoted the basic concept of Open Data based on the background and motivation literature. More details on the research method, implementation of Systemat-

ic Mapping Study (SMS) to configure and initiate Open Data are presented in the third chapter. In the fourth chapter, the results are analysed. Quantitative Assessments is presented in the fifth chapter and chapter six highlights the Qualitative Analysis. Chapter seven presents further discussions on the findings of supporting literature, the limitations, threats to validity, and future research work. Finally, Chapter eight concludes the results of the whole thesis.

2 Open Data Concepts

Concepts that are required to understand the thesis have been defined and explained in this chapter.

2.1 What Is Open Data?

2.1.1 Definition

Open Data is generally defined as *"the idea that certain data should be freely available to everyone to use and republish as they wish, without restrictions from copyright, patents or other mechanisms of control"* [50]. Similarly, from the Open Data Handbook documentation from Open Knowledge Foundation project, Open Data is a type of data which can be used, reused and redistributed without any limitations by anyone [2]. In the same way, open data has been defined by W3C eGov Interest Group as publishing data in its raw format, machine-readable and can be reused in any applications developed by others [3]. Furthermore, licenses are applied to all of the datasets to guarantee their originality and control the data usage in the future. The primary usage of the data may also help the data owner to monitor the potential for data reuse, influence a change in position on data reuse permission and easily to be accessed by the developers.

Additionally, Open Data works of literature that have been published on this topic, previously, academicians and practitioners such as Government itself, Industrial, and public agencies defined Open Data from their different perspectives. Open Data definitions evolved from interdisciplinary areas such as business, cultural, science, finance, statistics, weather, environment and transport.

Berners-Lee has outlined a set of 'rules' which is so called 'Linked Data Principles' as the basic guideline for publishing and connecting data [17]. These data are published and connected by using the web infrastructure while adhering to its standards and architecture.

A single global data space can be achieved by the rules as follows:

i. *Use URIs as names for things*
ii. *Use HTTP URIs so that people can look up those names*
iii. *When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)*
iv. *Include links to other URIs so that they can discover more things*

According to the study, Open Source, Open Access and Open Innovation were influenced by the intellectual roots of Open Data movement [31]. Besides, Open Source and Open Data are different between each other which Open Source more on the openness of the applications and source code. However, common implementations of Open Data are dedicated to any applications as well as mashups and visualizations [39]. Whereas in Open Government Data, five distinct processes were suggested as *data to fact* (particular facts of interest identified), *data to information* (blogs or infographics), *data to interface* (interactive visualisations), *data to data* (sharing derived or combined data), and *data to service* (data for broader application) [56].

The Comprehensive Knowledge Archive Network (CKAN) [54], a web-based data management platform of open source maintained by the Open Knowledge Foundation provides a Registry of Open Data and content packages. These registries contain most open datasets that are available publicly by the variety of authors that powering data hubs and data portals. It has been used by numerous governments, organizations, and communities. For example, YAGO, which is a knowledge base derived from Wikipedia and GeoNames, and Freebase, which provided by Google [57].

2.1.2 Standard Data Type and Datasets in Publishing Open Data

The standard formats for representing Open Data are Extensible Markup Language (XML), text file (TXT), JavaScript Object Notation (JSON) and Comma Separated Values (CSV). However, Resource Description Framework (RDF), Application Programming Interface (API) and the Ontology Web Language (OWL) which are more specified as semantic web formats were used [9,10]. Open Data can be published from

different sources, but with the certain data type or datasets standardization such as by using Metadata. In all datasets, Metadata should be linked and included to describe the Open Data and its specific content such as some basic information before they were released into a platform. Some of the metadata were included with an additional description on Open Data license and the data rights of ownership which handles the license details.

The standardization of datasets format is very important to integrate different government Open Data into clean, concise and well-structured datasets. As an example, GovWild tool is a search engine-like web application that allows other applications or platforms browsing and querying the collected open data interactively which were organized in JSON format [104,156,157]. The example of JSON format as shown in Figure 1 which presents how the raw data from ec.europe.eu is transformed to the generic JSON format with specific standard structure.


```
{ "_id" : "euFinance#28994",
"year" : 2008,
"nameOfBeneficiary" : "ROBERT BOSCH GMBH*",
"coordinator" : false,
"countryTerritory" : "Germany 70049 STUTTGART",
"coFinancingRate" : "67,51 %",
"amount" : 3199959.00,
"commitmentPositionKey" : "F13.A22622.1",
"subjectOfGrantOrContract" : "MULTISPECTAL TERAHERTZ, INFRARED, VISIBLE IMAGING",
"responsibleDepartment" : "Information Society and Media",
"budgetLineNameAndNumber" : "Support for research cooperation in the area of information and
communication technologies(ICTs-Cooperation)(09.04.01.01)."
}
```


Figure 1: JSON format raw data example

Another example is demonstrated by the French passenger transport services [104,158] which different kind of functionalities and the multiple datasets can be used simultaneously by defining datasets in a well-structured format CSV and XML as shown in Figure 2 and 3.

```
Sheet number;Service Name;Coverage service;Region;Department;City;Modes of transport;Type of service;Network ccessibility for disabled person;Land informations;Website; Website accessibility for disabled person;Information points ;Re-mark; Comments; Sms;Mobile application;List of cities covered (Postal code);Sheet created;Sheet modi_ed
```

Figure 2: CSV format raw data example

```
<ChouettePTNetwork>
<ChouetteLineDescription>
<StopPoint>
<objectId>NINOXE: StopPoint :15577811
</objectId>
<objectVersion>0</objectVersion>
<creationTime >2007-12-16T14:2 6:1 9.000+01:00
</creationTime>
<longitude >5.7949447631835940</ longitude>
<latitude>46.5263907175936000</ latitude >
<longLatType>WGS84</longLatType>
<containedIn>NINOXE: StopArea :1557779
</containedIn>
<name>CimetieredesSauvages (A)</name>
</StopPoint>
</ChouetteLineDescription>
</ChouettePTNetwork>
```

Figure 3: XML format raw data example

DBpedia is another good example of presenting formats standardization by extracting structured information from Wikipedia and publish on the web [159]. RDF links have been set to DBpedia to produce the web of data environment as shown in Figure 4.

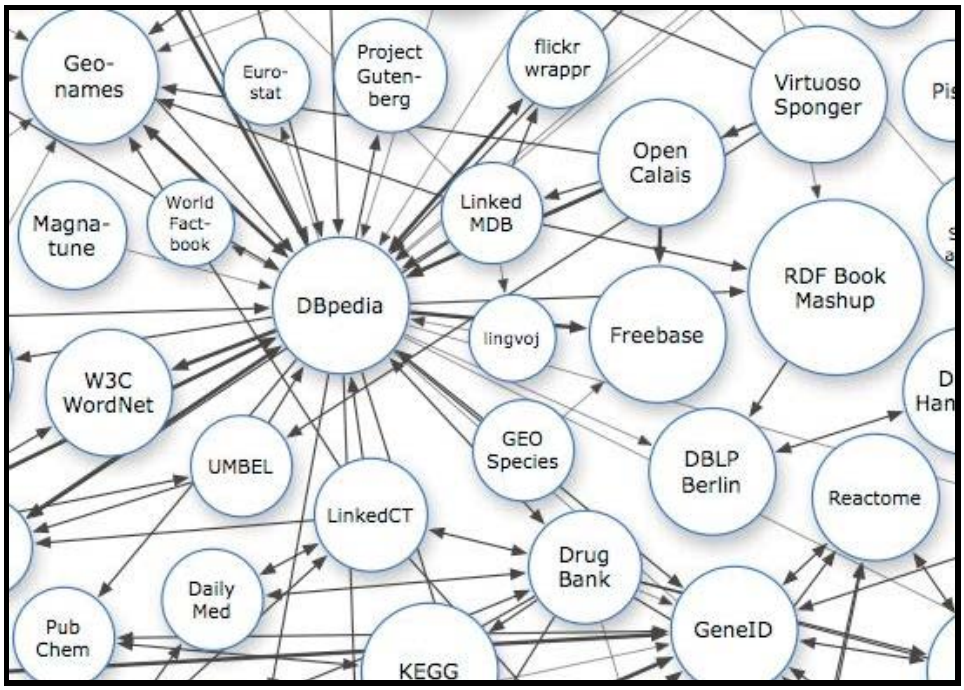


Figure 4: Dbpedia linked datasets using RDF

2.1.3 Semantic Web Technology and Linked Data

Semantic Web Technology (SWT) concept that is from the evolution of document-based web into a data-based web contributes to the data flexibility by the publishing, integration and interpretation process. The fundamental of SWT building blocks consists of Universal Resource Identifiers (URIs), Resource Description Framework (RDF), Simple Knowledge Organization System (SKOS), Ontology Web Language (OWL), SPARQL, Vocabularies and Linked Data [15]. The Semantic Web Layer Cake is as described in Figure 5.

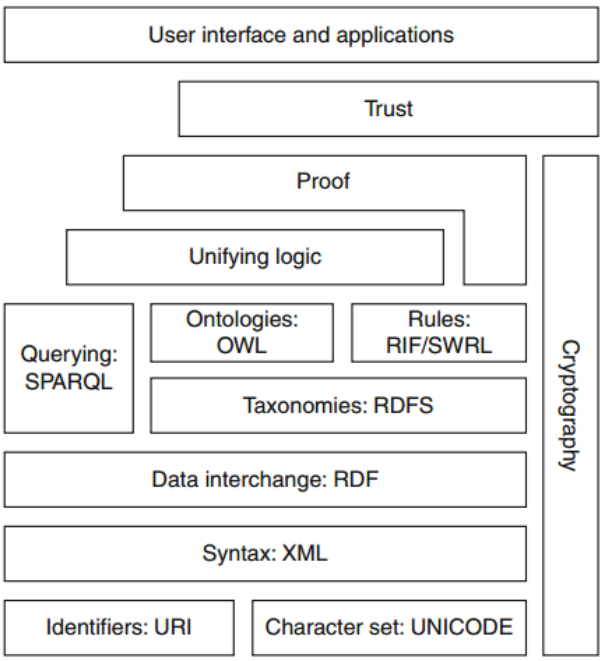


Figure 5: The Semantic Web Layer Cake [16]

In general, the relationship between Semantic Web, Linked Data, and Open Data is ambiguous. Open Data is simply 'data on the web', whereas Linked Data is a 'web of data'. In addition, from the Tim Berners-Lee presentation of Linked Data in 2008 [160], the Linked Open Data movement related to the principles set out by Tim Berners-Lee where "Linked Data is the Semantic Web done right" and the "Combination of openness with data + open standards". According to Miller, *"Linked Data may be Open, and Open Data may be Linked, but it is equally possible for Linked Data to carry licensing or other restrictions that prevent it being considered Open, or for Open Data to be made available in ways that do not respect all of Berners-Lee's rules for Linking."* [162]. However, there is the idea that Data.gov.uk is the combination of Open Data and Linked Data as the priority is to publish as much as Open Data and later by linking it all [161,163].

2.1.4 Linked Open Data Cloud (LOD Cloud)

The structured data such as RDF and XML can be published by the Linked Open Data Cloud. In the vision to create an integrated data space globally, the Linked Data princi-

ples are formulated by establishing links between related resources, especially in the Linking Open Data initiative. Linked Open Data Cloud has been published based on LOD principles. In 2011, LOD Cloud comprised of more than 200 datasets that are interlinked by RDF links from various domains of range around 504 million RDF links, interlinked by around 31 billion RDF triples [19]. However, this figure has grown to roughly 1000 datasets (as of April 2014) which the datasets cover various particular domains, which social media, government data, and metadata about publications being the most prominent areas [45,178]. Compared to the 2011 LOD Cloud, which dbpedia.org as a central linking hub, a more decentralized graph structure with geonames.org and dbpedia.org being linked from many datasets have been discovered in 2014 [178]. Figure 6 shows part of the diagram presenting published datasets and their inter-linkage relationships. The full diagram of LOD cloud diagram by particular domain as shown in Appendix I.

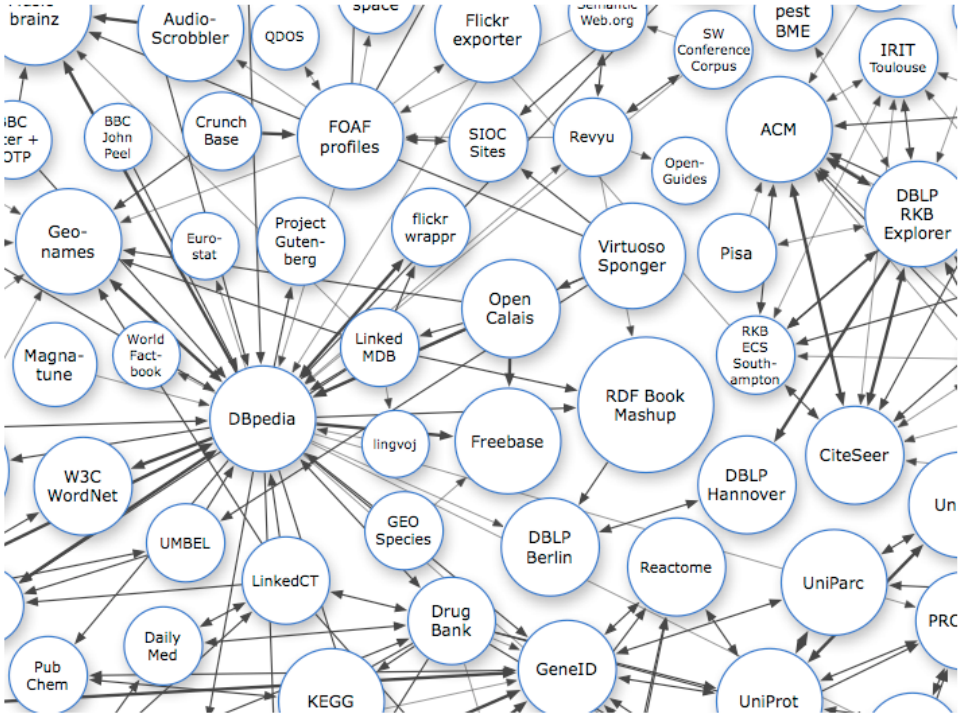


Figure 6: Part of LOD Project Cloud diagram [19]

2.1.5 Linked Data Applications

At present, there are three categories that classify the applications development by manipulating the published data: *Linked Data browsers, Linked Data search engines, and domain-specific Linked Data applications* [20]. A user browses through data sources by tracking links presented as RDF triples in Linked Data browser, the same way as the traditional web browsers navigation that is following hypertext links. The search engines of the linked data and indexes crawls the web by following RDF links from the beginning of navigation process and provide query intelligence over aggregated data. These services can be categorized into the Human-oriented search engine and application-oriented indexes. However, compared to both kinds of applications, a domain-specific application is the most sophisticated which 'mashing up' data from various Linked Data sources such as DBpedia mobile, the location-based Linked Data browser [176], BBC programs and Talis Aspire, the university web-based resource list management [177] applications.

2.1.6 Storage

The ultimate goal of Linked Data is to be able to use the web as a single global database [20]. Tim Berners-Lee once highlighted design issues on the socially-aware cloud storage. In addition, on top of a read-write storage, there are applications (desktop or Web application) working in few existing web protocols and architecture [22]. As a result, the storage becomes more commodity and independent. For example, In Open Government Data Initiatives (OGDI), the government and public data can be published based on cloud computing, Microsoft Windows Azure cloud storage [21]. The data in the cloud can be queried and browsed in interactive views by the users by its quickness and efficiency.

2.2 Open Data Principles

2.2.1 Star Rating

The 5-star development scheme for Open Data has been developed by Tim Berners-Lee, the Web inventor and Linked Data initiator in World Wide Web Consortium (W3C) as shown in Figure 7 [17, 24, 26]. This star rating was invented to encourage data owners especially the government to implement a good linked data. In general, Linked Data is not necessary to be opened, and 5-star Linked Data can be achieved without it being open. However, if it confirms to be Linked Open Data, then it has to be opened to get any star [17]. If the information has been made public, and it has an open license, one star is achieved. Besides, the more stars achieved, the more powerful and easier data access for users.

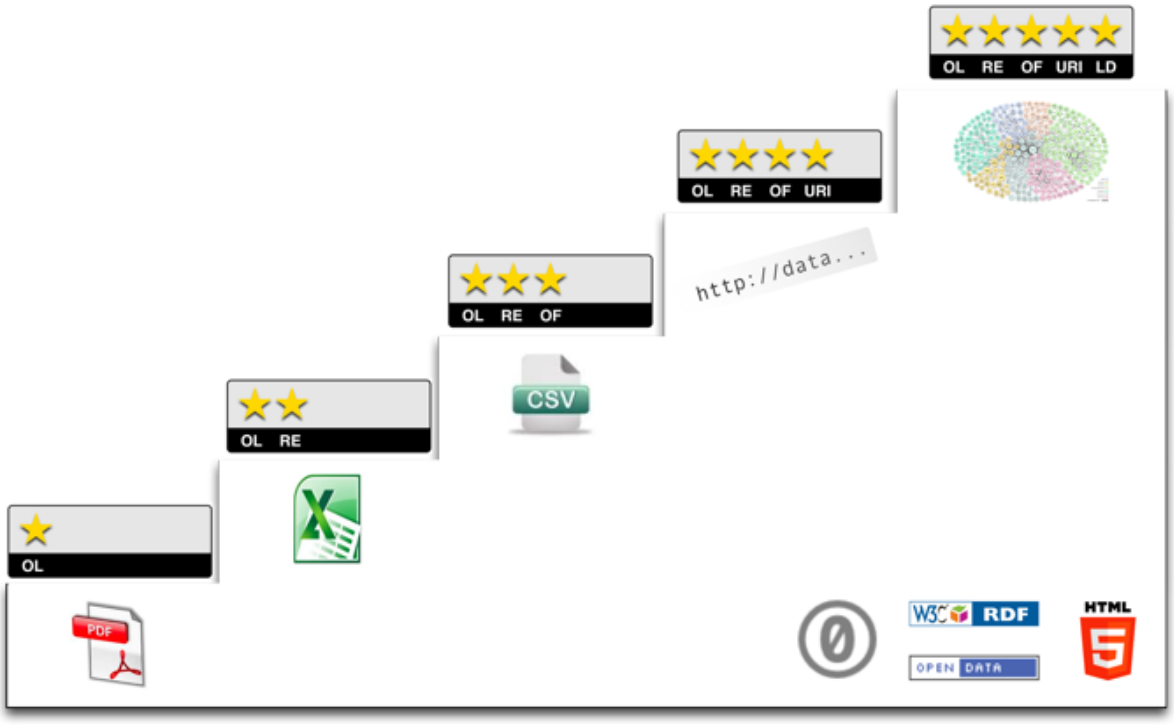
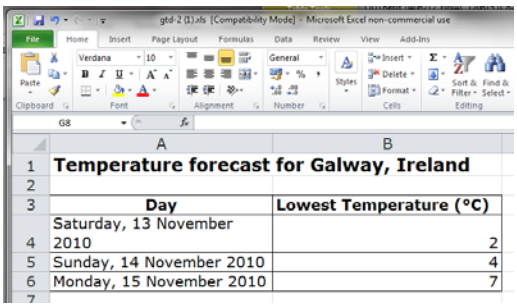
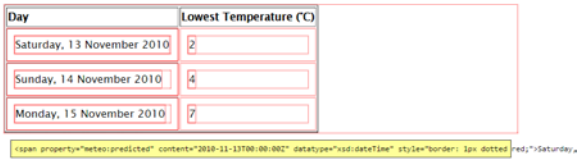


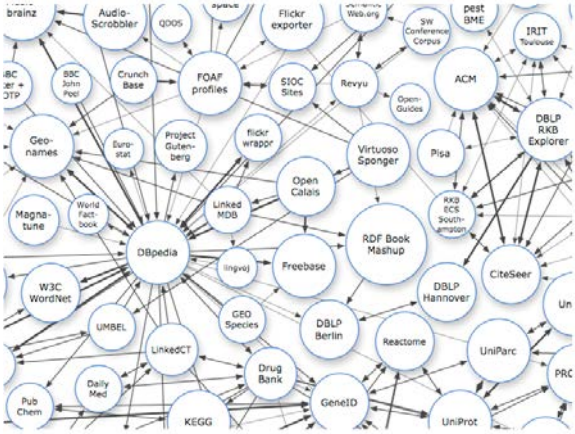
Figure 7: Star rating illustration

This 5-star system would help the public sector bodies much to achieve re-use facilitation by allowing the access of re-use propitiousness and a structured improvement guid-

ing star creation. These standards encourage data owners to publish their data according to Linked Data principles by asking: *Is your Linked Open Data 5 Star?* [27]. The star rating levels is summarized as shown in Table 1 below.

Table 1: Summarize of Star Rating Descriptions

Stars	Description	Example
★	Data available on the web under an open license in any formats. Human readable.	PDF Documents and HTML tables [24,25]. **Temperature forecast for Galway, Ireland** **Day** | **Lowest Temperature (°C)** Saturday, 13 November 2010 | 2 Sunday, 14 November 2010 | 4 Monday, 15 November 2010 | 7
★★	Data in a more structured and predictable format. Machine readable.	Microsoft Excel - only can be read by Microsoft Excel software  **Temperature forecast for Galway, Ireland** **Day** | **Lowest Temperature (°C)** Saturday, 13 November 2010 | 2 Sunday, 14 November 2010 | 4 Monday, 15 November 2010 | 7
★★★	Data is converted into a non-proprietary formats that can be accessed by any software. Machine readable.	Comma Separated Values (CSV) – each row contains one record with multiple pieces of data separated by a comma. Can be read by any spreadsheet software. Day, LowestTemperature Saturday,13 November 2010, 2 Sunday, 14 November 2010,4 Monday, 15 November 2010, 7
★★★★	Data is using standard formats like URIs or RDF triples to denote things. Machine readable.	 Day | Lowest Temperature (°C) Saturday, 13 November 2010 | 2 Sunday, 14 November 2010 | 4 Monday, 15 November 2010 | 7

		
★★★★★	Link Data to provide context (Linked Open Data concept) to the data published. Connecting related data across multiple data sources. Machine readable.	

2.2.2 Five Data Openness Levels

Basically, there are three categories of data classification: unstructured data, semi-structured data, and structured data [9]. However, different concepts related to Open Data were specified as Structured Data, Linked Data, Real-time data and Open Data itself [10]. Interoperability is the main characteristic of data which described its availability and accessibility, reuse and redistribution of dataset and everyone universal participation [2]. The openness level of data can be clearly defined when the two open datasets from different sources can be combined into a larger systems. In addition, openness defined as programmability by other digital objects such as computer programs create a platform where data and applications cooperate with each other [31].

The awareness of data owners can be raised by data openness based on 5-star rating Tim Berners-Lee [164]. Distribution of content types can be varied from non-XML machine readable data formats (Excel and CSV), image formats (JPEG and GIF), XML formats, text formats (Word, RTF or plain text), HTML, pdf, ppt and zip format. The 1-star lev-

el depends on the formats of the data available on the web such as .pdf or image formats. Whereas 2-star level requires the machine-readable structured data such as Excel. Furthermore, XML and CSV, the non-proprietary formats were considered as the 3-stars openness level. The 4-star openness levels require the use of W3C open standards such as HTML, XML, and RDF. Linked Data API, RDF, and JSON are the types of 5-star level of data openness.

2.2.3 Metadata

Metadata is more about data of data and the description of what, where, how and to whom the data itself are provided. In addition, metadata can contain information about which license is used, the geographic context and how many times it was updated. With the sharing, reusing and understanding aims for heterogeneous datasets, metadata often describes the principal aspect of data [58]. Later, Tim Berners-Lee was urged to add the new requirement for the data especially in the government data perspective by storing the metadata of the dataset itself in the star-rating mainframe. The metadata should be available from the major catalog or registered at their advanced registries such as ckan.net for any open dataset, data.gov.uk and data.gov for the United Kingdom or the United States government datasets respectively. The registry has a three folded goal: to be the reference for Open Datasets concerning the organization, to involve on-going and future projects and to raise data awareness within the organization [48].

Metadata becomes valuable information especially in the Open Government Data where many precious data were shared. In case of missing data, the metadata is automatically and the annotations manually responsible for working around collecting all information about the information to process useful knowledge out of it [59]. The main purpose of data catalog or data registries is to gather all metadata centralized where data seekers or consumers can benefit them with the “one-stop-center” experience. In addition, the federations of different catalogs are enabled which catalogs operated on different levels of administration [60]. The metadata and semantic research become an important area that aims to provide a seamless view of all information on the web through Linked Open Data (LOD).

2.2.4 Open or Free Licenses

Licensing on Data is necessary for the sake of clarity and preventing third party from using, reusing and redistributing data without explicit permission [51]. Besides, it is more important to Open Data format to avoid any unnecessary technological obstacles to the performance of the licensed rights with certain conditions [52].

2.3 Organizations and Workgroup

Besides Government, Academia and Industries, there are also several non-profit organizations that are dedicated focusing on any open data activities and discussions. These organizations provide the opportunity for people with the similar interests working together. As an example, Open Data Foundation (ODaF) is playing a role in the area of education, healthcare, social science, labor, economics, finance, development, technology, agriculture, and the environment [29]. It is focusing on enhancing data together with metadata accessibility and overall quality to encourage research, policy making, and transparency. In similar, Open Knowledge Foundation (OKF) provide the opportunity for the people to get involved in openness exploration by one of their working groups and services [1].

2.4 Open Data Consumption

In open data consumption, even though the data application is very open and sophisticated, it is of no use and meaningful if it cannot be utilized by the users [61]. The data consuming concepts are the ability of the users to identify process and generate information and knowledge from it. In addition, the open data consumption can be increased by the tools and applications of scalable mass appeal mobilization and civic engagement. However, there are some arguments between the public transparency and business profitability to this matter [35]. The separation between the supply (the side makes the data available) and demand (the side builds something useful on the data) were notified.

However, open data services allow end-users to access open data constructing a platform or infrastructure. In developing open data services, there are three requirements for the data openness. These are technical openness issues related to interfaces and standards), legal openness (contracts, copyright, licenses, privacy and data protection) and commercial openness (free of charge or pay the requested fee) of data [31].

2.4.1 Open Data Ecosystem (ODE)

Open Data Ecosystem (ODE) is a data cycle that comprised of Infomediaries to publish what they produce, Data packaging and patching for publish and share data format, and Publisher notification of patches with integration tools [53]. As an example, a government publish data to the world after it is processed by intermediaries such as app creators or analysts and then consumed by the end users. These data was cleaned, integrated or packaged into the ecosystem and become more valuable than the source. The examples of these tools are Google Refine, Scraperwiki, and CKAN Data Management System.

2.4.2 Open Data Platform

Open Data platform was developed and accessed directly via open protocols as an integrated solution to publish the non-proprietary format of data either in public agencies or private institutions. This cataloged datasets can be efficiently used, navigated, accessed and reused by the users from the consolidated view and minimal efforts of developers [60]. Besides, storing and keeping datasets in Open Data platform are allowed which it can be turned to use in the applications and services instead of just as metadata handling. More importantly, with the strength of linked data, the data quality and its application or services usability can be utilized and improved. However, most of Open Data platform still play a passive role as data providers, rather than a much more active roles as a data coordinators by providing API for application development [62].

2.4.3 Open Data Initiatives

Open Data Initiatives (ODI) were intended to provide government information becomes more transparent, participative and collaborative which led more interests and trusts of citizens, investors and public relations [63]. In addition, data collection, analysis and application can be done with public engagement that helps to decrease government expenses and improving their efficiency. Moreover, Open Data become the new source of economic growth such as for the European Union from data transactions and indirect contribution from information services innovation.

2.5 Summary

In summary, the whole concept of Open Data was discussed in this chapter in order to understand the topic well before the Systematic Mapping Study process begins. This up-to-date information covered from its definition, principles, organizations and workgroups responsible and how Open Data has been consumed.

The main reason of this study is to explore how to produce a quality Open Data application without failure. As at the end of April 2014, the Linked Open Data (LOD) cloud datasets increased from 200 datasets in 2011 to roughly 1000 datasets. It can be seen that how Open Data has positive impacts in Open Data or Linked Data applications. However, the standardization of data types and datasets are important while the software development lifecycle begins to integrate with other datasets. Another important aspect in Open Data is the 5-star rating system, Tim Berners-Lee to encourage data owners to implement a quality linked data and towards 5 stars rating. The same way goes to the Data openness with 1 to 5 rating.

To produce Open Data, applications are as important as the consumptions. The sophisticated open data applications only can be meaningful if it can be consumed by the users. This can be measured from the technical (interfaces and standards), legal (licenses, copyright, contracts, and privacy), and commercial (free of charge or pay the requested fee)

openness of data. Open Data Ecosystem (ODE), Open Data Platform (ODP) and Open Data Initiatives (ODI) are some examples of providing and publishing Open Data.

In the next chapter, the Open Data Systematic Mapping Study is presented based on formal process by Kitchenham and Charters to explore the existing studies on Open Data.

3 Research Method

This chapter contributes a general outlook of research method has been used; Systematic Mapping Study (SMS) approach followed the formal process described by Kitchenham and Charters [29].

3.1 Research Process

In this research, in order to access the number of potential articles and to understand the literature in different research streams, the Systematic Literature Review (SLR) has been used as the research approach of the study [29]. In particular, this review can be categorized as a Systematic Mapping Study (SMS) or scoping study process [20, 21]. SMS is useful to explore and present the whole picture of a research area, indicate quantity of the evidence and establish any research evidence exists in the context of Open Data. In fact, the number of published articles in this area is high; this system helps to identify the quality, potential and relevance of articles used in this review process. Based on the primary studies required, the results of mapping study help to identify research interest within this topic.

In general, the review process progresses with the following steps [29, 30]:

1. *Protocol preparation which include defining*
    a) *the process*
    b) *the research questions*
    c) *the inclusion and exclusion criteria*
    d) *the analysis procedure*
2. *Conduct of a pilot study*
    a) *defining search queries*
    b) *choosing the digital libraries and other sources of materials*
    c) *searching*
    d) *reviewing the results*
    e) *summarizing and analyzing the results*

f) refining the queries for the actual search

3. Conduct of an actual search

a) selection of databases and search queries based on the pilot study results

b) searches

c) removal of duplicates

d) application of inclusion and exclusion criteria

e) classification of excluded articles

f) summary and analysis of the results

4. Data extraction

a) review of the articles

b) gathering information from the articles

c) classification of the articles

d) identification of primary studies

5. Quantitative Assessment Study

6. Analysis of the results

7. Qualitative Analysis

8. Development of Conclusions

9. Reporting

From the process described, some steps have been excluded such as data synthesis which is unsuitable for mapping study [30]. However, Qualitative analysis remained in order to explore the potential new findings or solutions of Open Data.

To ensure consistency during the included and excluded articles classification, the publications have been reviewed twice according to Budgen et. al [23]. The first round of review was done by focusing on the pilot and actual search by using titles, abstracts and keywords to identify the fundamental topic of studies, the unit of analysis [41]. The publications where the units of analysis were not related to Open Data were excluded for the next round. In the second round, the full texts of publications have been reviewed and eliminate which papers should be excluded and extract any additional and useful information related such as research questions and research methods. The identi-

fied units of analysis were checked correctly and refined them as appropriate. By formulating the correct way to find suitable publications, the process of reading and identifying the research gaps is more straightforward. Furthermore, the Quantitative Assessment, a detailed meta-analysis of Open Data has been carried out to provide the current state and trend of the study.

3.2 Research Questions

The mapping study main objective is to find all relevant studies to the research questions. At the beginning of the research process, the research questions and findings concerning on: “What research questions in open data are being addressed?”; “What original research exists in the study of open data”; and “What areas of open data novelty research?”

Then, the more specific Research Questions (RQs) were developed as below:

1. How many publications have been published in the field of Open Data, and how has the production changed over the years?
2. What have been the most popular databases for Open Data publications, and how has this changed over the years?
3. What organizations have been responsible for writing Open Data publications?
4. What are the original researches have been addressed by the existing study of Open Data?
5. What areas of open data require further research?

Specifically, research questions 1, 2, 3 are answered in chapter 5 (Quantitative Assessments), Research question 4 is answered in Chapter 6 (Qualitative Analysis) and research question 5 in Chapter 7 (Discussions).

3.3 Selection of Database and Search Queries

The literature study results are heavily influenced by the keywords and digital databases used in the searches [29]. To get the idea of articles available, the review began with a quick search that covered the nature of Open Data based on the keyword search by accessing Google Scholar, Citation Databases, Scopus; and Information Discovery Tools, WorldCat and Web of Sciences as shown in Table 2. These databases have been selected to get an overview of the search relevance trends among databases.

Table 2: Quick Search Results

Search Keywords	**Google Scholar**		**WorldCat (Articles)**		**Scopus**		**Web of Science**	
	All range	2012-2014	All Range	2012-2014	All Range	2012-2014	All Range	2012-2014
Open Data	4,780,000	662,000	396,876	73,964	191,369	31,268	72,333	18,393
“Open Data”	113,000	17,600	1,929	1,155	1,582	889	269	172

Approximately 5,440,578 articles were found that include articles, patents, citations, etc. by the search keywords *open data* as shown in Table 3. The keywords decrease the number of hits to 116,780 by putting quotes around the keywords indicating that *“open data”* as a single concept has raised considerably less interest than *open* and *data* as distinct keywords. Overall from the resulting process, in the total of 785,625 publications matched with the Open Data and 19,816 publications matched with the “Open Data” within the scope of the area between 2012 and 2014. From the search result, the academic literature studies on Open Data are accumulating for the last few years. However, this topic of Open Data is still developing and fragmented. In addition, by narrowing down the search query by having the keyword in title appeared to be a reasonable basis for the literature search [30].

Table 3: Quick Search Result of Publications Found

Search Keywords	**All Range**	**2012-2014**
Open Data	5,440,578	785,625
“Open Data”	116,780	19,816

In order to get more precise results on the nature of Open Data, the searches continued in various scientific databases. The databases are IEEE, ACM, SpringerLink Ejournals and ScienceDirect, which are the most useful in the Computer Science and Information Technology [29] from the prior reports. Besides, a new digital database such as Ebsco-Host was added. The reason for choosing IEEE is that it is a significant innovative association for excellence in the field of technology. ACM remains world's largest database for computer science. Springerlink was chosen because it coordinates with the academicians and authors in the scientific community. The following were the electronic databases used:

a. IEEE Xplore (*http://ieeexplore.ieee.org/Xplore/home.jsp*)
b. ACM Digital Library (*http://dl.acm.org*)
c. Science Direct (*http://www.sciencedirect.com*)
d. SpringerLink (*http://link.springer.com*)
e. EbscoHost (*http://search.ebscohost.com/*)

Specific conferences, symposiums and workshops have been searched by Internet browsing and electronic databases organizer such as IEEE database, Open Data foundation [28] by looking at the main topic of interest in open data. The results are presented as below:

a. 2013 IEEE International Conference on Big Data
b. 2014 International Conference on Big Data and Smart Computing (BIGCOMP)
c. 2013 IEEE International Congress on Big Data
d. WOD '12 : Proceedings of the First International Workshop on Open Data
e. WOD '13 : Proceedings of the 2nd International Workshop on Open Data

The publication searches were conducted in two phases: Pilot Search and Actual Search. The Pilot Search was carried out to select the relevant sources of publications as many as possible and refine the search queries as the basis to be explored in the Actual Search.

3.4 Study Selection Criteria

The resulting of literature search can be scattered from different organizations and working groups. To identify only relevant articles, this study was limited to “Computer Science and Information Technology” area. Consequently, the journals or proceedings were selected with the keywords “Open Data”. From the results, the abstract of the top articles, which were cited the most are reviewed to eliminate those that were not related to the topic.

3.4.1 Inclusion Criteria

i. Include primary studies related to the research questions
ii. Research article or journal topic closely related to the subject of the research question
iii. Publications explaining “open data”
iv. Studies or research conducted in industry, government and any academic environment
v. Publication’s full text is available

3.4.2 Exclusion Criteria

i. A duplicate copy of the same research study
ii. Publications that do not describe open data
iii. Publications that were written in languages other than English
iv. Business Articles (general business point of view)

4 Results

The resulting pilot and actual searches within the research process are described in this section.

4.1 Pilot Search

The search process is carried by using five specific digital databases, IEEEXplore, ACM, Science Direct, EbscoHost and SpringerLink Ejournals as shown in Table 4 consists of Journals, Book Chapters and Proceedings. From the search process, 7,304 publications were matched with the specific search keywords by the title "open data" within all range. From this figure, more than 40% of the publications were carried out between 2012 and 2014 with 3,161 publications.

Table 4: Table of Search Keywords "open data" Result by Digital Databases, IEEEXplore, ACM, Science Direct, Ebsco and SpringerLink Ejournals (Journals, Book Chapters and Proceedings)

	IEEEXplore	**ACM**	**Science Direct**	**EbscoHost**	**SpringerLink Ejournals**	**Publications Found**
All Range	247	1,229	2,030	217	3,581	7,304
2012-2014	141	693	620	184	1,523	3,161

For the conference proceedings results, 1,289 publications have been produced and more than 50% from the result were between 2012 and 2014 with 734 publications as shown in Table 5.

Table 5: Table of Search Keywords "open data" Result by Digital Databases, IEEEXplore, ACM, Science Direct, EbscoHost and SpringerLink Ejournals (Proceedings)

	IEEEXplore	**ACM**	**Science Direct**	**Ebsco**	**SpringerLink Ejournals**	**Publications Found**
All Range	222	1,067	-	-	-	1,289
2012-2014	125	609	-	-	-	734

Table 7 shows the whole proceedings from the related conferences. However, only 21 literatures are related with Open Data research topic.

Table 7: Table of Search Proceedings of Open Data Conferences

Conferences	Publications Found	Publications Included
2013 IEEE International Conference on Big Data	262	2
2014 International Conference on Big Data and Smart Computing (BIGCOMP)	64	1
2013 IEEE International Congress, Big Data	70	0
WOD '12 : Proceedings of the First International Workshop on Open Data	8	8
WOD '13 : Proceedings of the 2nd International Workshop on Open Data	10	10
Total :	414	21

As the starting point, the search keywords are explored based on the particular interest in the implementation of Open Data within working groups' research interest in Open Knowledge Foundation. The working group consists of System, User, Link, Computing, Network, Government, Platform, Device, Health, Cloud, Sensor, Game, Geodata And Sustainability [1] and types of open data with potential uses and applications, Cultural, Science, Finance, Statistics, Weather, Environment And Transport [2].

Figure 8 shows the number of found papers based on the pilot search in digital databases. The search was based on the combined keywords, and the more detailed data as shown in Appendix II.

The keywords in the searches were:

- "open data" AND (system OR science OR user OR link OR computing OR network OR government OR platform OR environment OR cloud OR statistics OR health OR device OR sensor OR transport OR cultural OR weather OR game OR finance OR sustainability OR geodata)
- Within Title OR Abstract
- Within the range of the year 2012 and 2014.

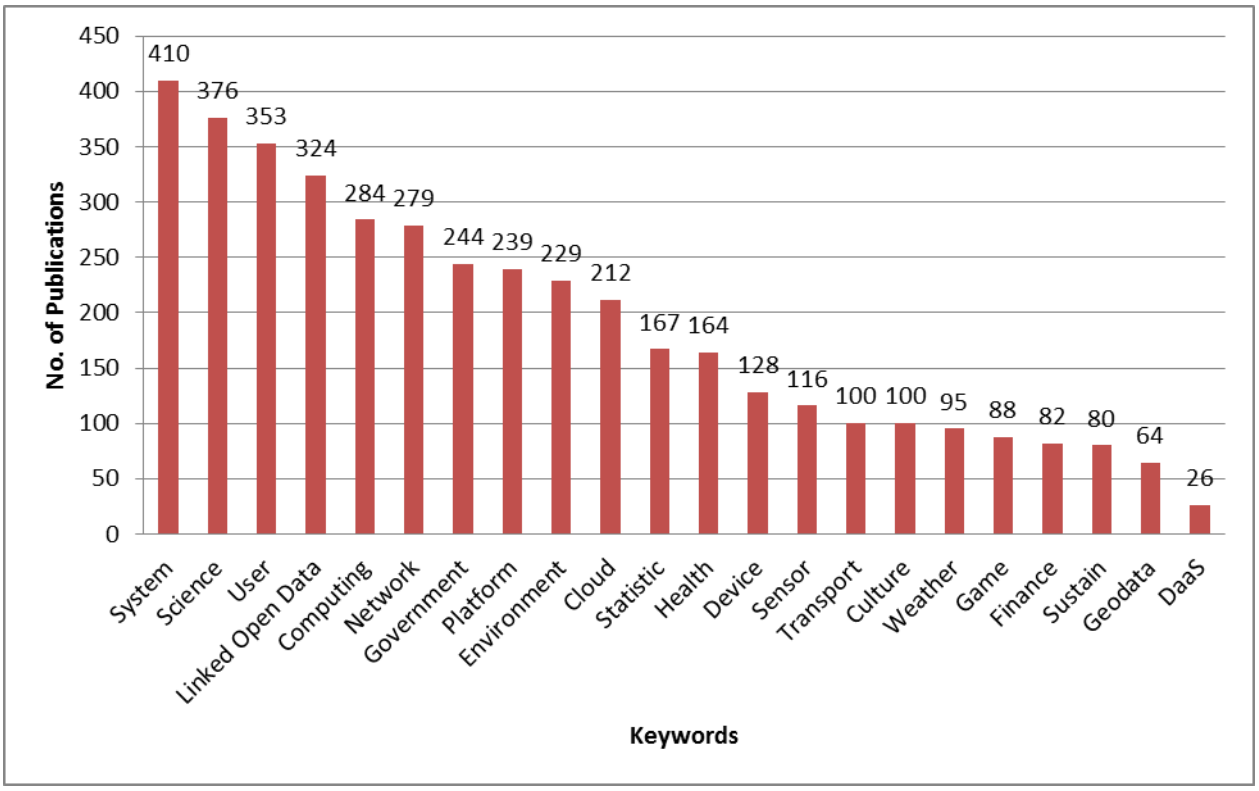


Figure 8: The Pilot Search Number of Publications on Open Data And the Keywords in Digital Databases

In order to identify the research trend and novelty of research topics, the random results showed in Figure 8 were clustered into the more narrowed focus of scopes. The scopes that relevant were:

- Type of Data (Transport, Environment, Geodata, Culture, Science, Finance, Statistic, Weather, Health, Government, Gamification, Sustainability)
- Systems Platform (Computing, System, Platform, Device, Cloud and DaaS)
- Network and Sensor
- User (Social Integration)
- Linked Open Data

The result percentages of the narrowed scopes are as shown in Figure 9 which Type of Data was the highest percentage with 43% and followed by Systems Platform with 31%. However, Network and Sensor, User and Linked Open Data were at the bottom of not more than 10% of the percentage.

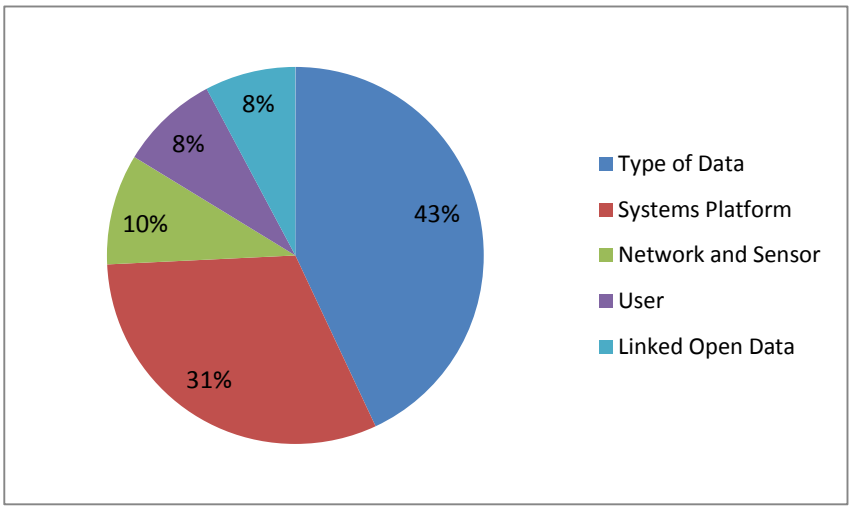


Figure 9: The percentage of results categorized by focused scope of research

4.2 Actual Search

A quantitative assessment study is considered within the actual search process to present the advances of an objective overview and trend.

4.2.1 Main Databases

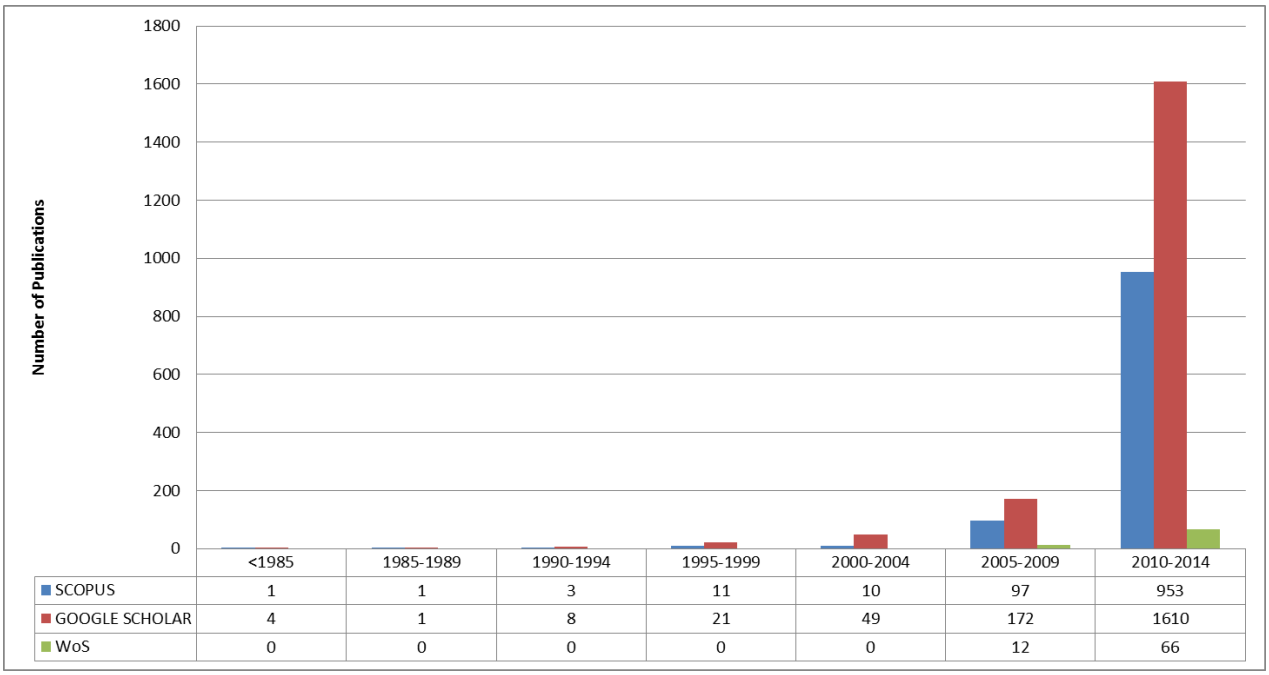
Extending from the Pilot Search, Actual search was conducted on May 15th, 2014 in the digital databases as below:

i. Google Scholar (search engines)
ii. Scopus
iii. Web of Sciences

These databases were selected as starting point to explore the trend of Open Data publications distributions. Scopus was chosen due to its largest abstract and citation database of peer-reviewed literature and provided a comprehensive the world's research output overview [179]. It is the same way for Web of Science that has been recognized as the most comprehensive and versatile research platform available [180]. Google scholar has

been chosen due to its accessibility web search engine that can provide indexes the full text or metadata of scholarly literature across most peer-reviewed online journals especially in Europe and America's largest publishers [181]. In addition, Google Scholar search results are efficient and also present similar articles related to the original results.

As a result, the total of 3019 Open Data publications were identified spanning the years from 1979 through 2014. The publications distribution for the five-year intervals is shown in Figure 10. The Open Data research area has exhibited exponential growth since 2005. The rapid growth of Open Data might be from the growing awareness of the Open Data novelty in the research area.



	<1985	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014
SCOPUS	1	1	3	11	10	97	953
GOOGLE SCHOLAR	4	1	8	21	49	172	1610
WoS	0	0	0	0	0	12	66

Figure 10: Quantity of publications in Domain of Open Data (Scopus, Google Scholar and World of Science Databases)

Based on the results in Figure 10, some decisions were made to narrow down the search criteria for the next process. In conclusion, the search queries keywords were decided in the range of years within 2005 and 2014 by followings criteria:

- ”Open Data” within:
    i. Article or Document Title AND
    ii. Abstract
- Computer Science and Information Technology field of research

4.2.2 Specific Databases

Based on the findings from the prior reports on Systematic Literature Review (SLR) [23,29,30], they had found five digital databases as the most useful ones. Thus the initial lists of digital libraries (Journals, Book Chapters, and Proceedings) are used as follows:

i. IEEE Xplore *(http://ieeexplore.ieee.org/Xplore/home.jsp)*
ii. ACM *(http://portal.acm.org/dl.cfm)*
iii. Science Direct *(http://www.sciencedirect.com)*
iv. EbscoHost - Academic Database *(http://search.ebscohost.com)*
v. SpringerLink Ejournals (*http://link.springer.com)*

Figure 11 shows the distributions of publications that are using the specific digital databases in the total of 295 Open Data publications. The highest publications were from SpringerLinkEjournals with 109 publications and followed by ACM with 62 publications. Meanwhile, IEEE Xplore and Ebscohost are sharing the same figure with 55 publications. Surprisingly, Science Direct is the lowest with only 14 publications.

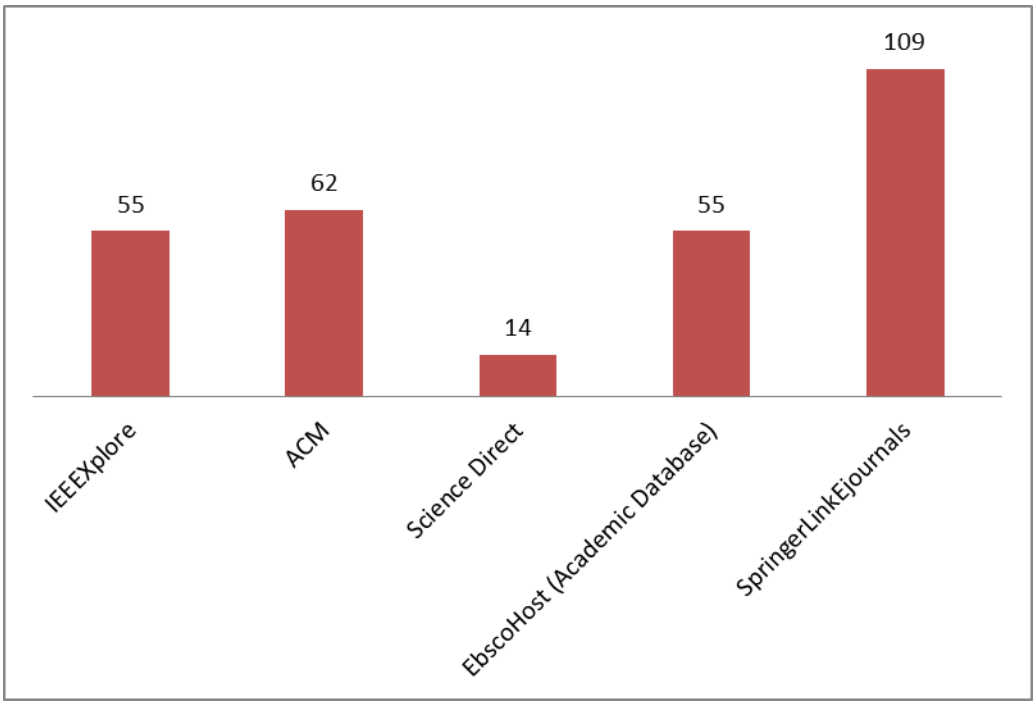


Figure 11: Open Data search results by specific digital databases (n=295)

The search results by publication types as shown in Figure 12. More than 40% of the items were published in conferences, more than quarter in Book Chapter and about a quarter were released in Articles (Journal or Magazine Articles).

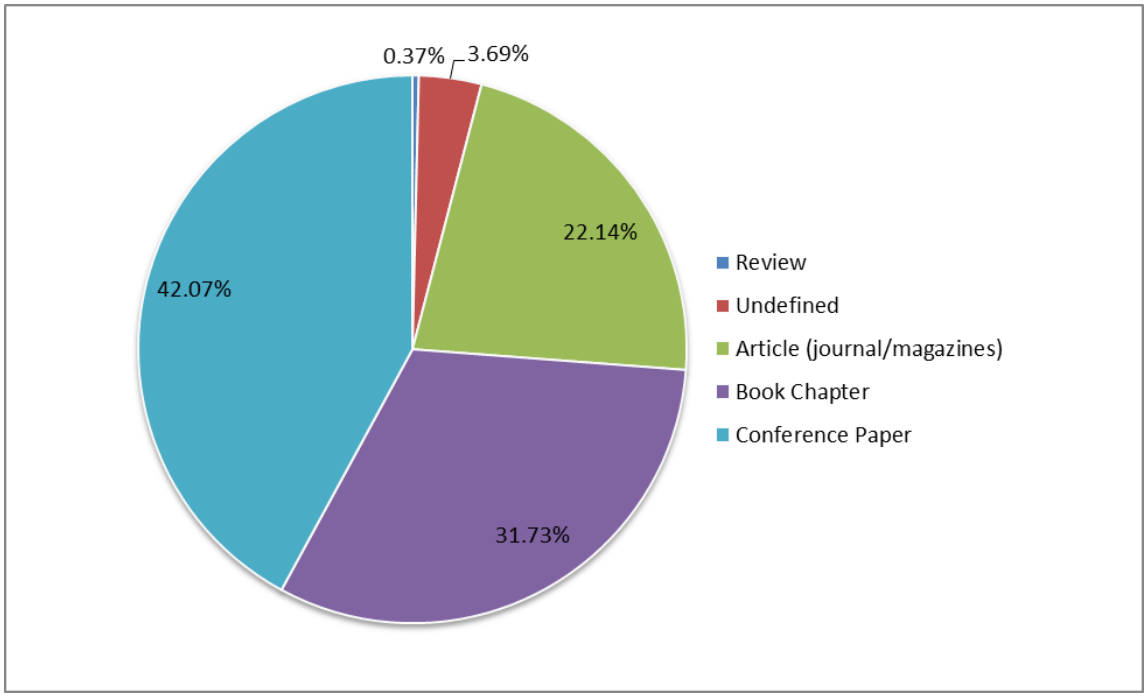


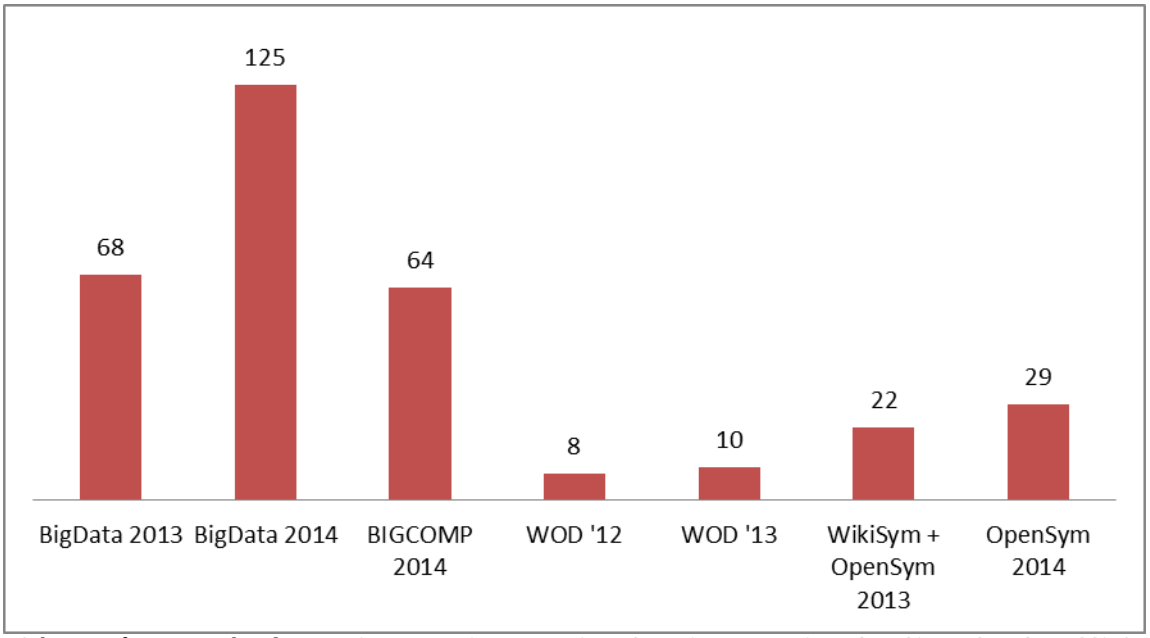
Figure 12: Open Data search results by publication types

4.2.3 Conferences, Symposia and Workshop

The selection of the conferences, symposium and workshop to be included in the review process are much influenced by the proceeding papers, workgroup main topic of interests and research by using Google. The conferences, symposium, workgroup and workshop were included as below:

i. WOD '12:Proceedings of the First International Workshop on Open Data
ii. WOD '13:Proceedings of the 2nd International Workshop on Open Data
iii. 2013 IEEE International Congress on BigData
iv. 2014 IEEE International Congress on BigData
v. Big Data and Smart Computing (BIGCOMP), 2014 International Conference
vi. WikiSym + OpenSym 2013 [34]
vii. OpenSym 2014

Based on the pilot search, the new addition of symposium related to Open Data, OpenSym2014 was included in the actual search as shown in Figure 13 with 326 publications in total.



**Other conferences related, WWW'11, WWW'12, WWW'13, SWIM'12, DDFP'13, AKBC'13, ISEMANTICS'10, ICTD'10, ISVC'11, dg.o'13, WebSci'12 and ISEMANTICS'11. But not included due to few number of proceedings.*

Figure 13: Open Data search results by Open Data Conferences, Symposiums and Workshops (n=326)

4.2.4 Results Included

The quality of publications that should be included in the study was restricted according to the title, abstract and availability. The richness of papers, from across research themes, involved the judgmental process and reading before the articles were selected. In other words, the selected articles were only accepted when the abstract contained related to Open Data and was in English. From the final review, overlapping and not related publications were excluded from the result as presented in Table 7.

Table 7: Results of the Actual Search

		Number of
Digital Databases	**Publications Found**	**Publications Included**
IEEEXplore	55	51
ACM	62	46
Science Direct	14	14
EbscoHost (Academic Database)	55	15
SpringerLinkEjournals	109	100
Total:	295	226

		Number of
Conferences/Symposium/ Work-shop	**Publications in Proceeding**	**Publications Included**
2013 IEEE International Congress on BigData	68	0
2014 IEEE International Congress on BigData	125	1
Big Data and Smart Computing (BIGCOMP), 2014 International Conference	64	0
WOD '12:Proceedings of the First International Workshop on Open Data (ACM)	8	6
WOD '13:Proceedings of the 2nd International Workshop on Open Data (ACM)	10	3
WikiSym + OpenSym 2013	22	3
OpenSym 2014	29	4
Total:	**326**	**17**

Total Publications:	**621**	**243**

The relevant titles and abstracts of the 621 publications were selected during the final synthesizing process; however only 243 articles are recorded as relevant. The relevant articles were chosen after reading the full text of the publications and based on selection criteria at section 3.2.3. After selection, 17 studies were the conference, symposia or workshop publications and the remaining studies (226) were published in journals.

4.2.5 Data Extraction Process

The thorough review of the publications was based on the elicitation techniques which the empirical results derived from the systematic review [31, 32, 33]. As the result, we should be able to gather some empirical evidence from the selected studies. The piece of proof found constructed a relationship between the elicitation techniques that were tested in empirical study. To describe the findings, all the articles are measured based on the Publication Types, Organization Affiliation, Countries, Year of Publications, Star Rating and Units of Analysis. For each category, the theoretical perspective, themes, empirical support and discussion of identified factors that related to Open Data were identified.

The results of 243 articles included in the review were analysed and presented in Chapter 5, Quantitative Assessment and 6, Qualitative Analysis which covered different study area and research interest.

5 Quantitative Assessments

The results of the Quantitative Assessments are described in this chapter.

5.1 Year of Publication

In this section, quantitative assessments of the findings based on the articles in Open Data have been done. It is interesting to analyze if there has been an increasing or decreasing trend by year. The distribution of all 246 publications identified is spanning the years from 2006 through 2014 as shown in Figure 14. Looking for the early year of analysis, from 2006 to 2008, the numbers of relevant articles published in Open Data is small; only six relevant articles were found. However, from the figure, it has been shown that Open Data research area has exhibited exponential growth since 2009. To conclude, this means more studies or research has been conducted in this period. This exponential growth is based on many supports and motivations either from the government or the practitioners themselves.

It can be seen that the rapid growth of Open Data as a research area in this period might be motivated by several supports from the government policy and initiatives themselves. Firstly, the United States government openness philosophy and transparency that government data can be accessed to public since 2009 [64, 6, 40]. Secondly, the European Commission has issued the new release of Public Sector Information Directive, which included particular cultural heritage data in public data accessible openly by Europe public institutions in 2013 [8]. Thirdly, in May 2013, the White House ordered federal agencies to create more open data and machine-readable government information such as public APIs that can be implemented by the government and private developers to access data [42, 43]. An instruction was announced in 2012 to open up government systems with public interfaces for commercial application developers. Moreover, the Germany policy-makers, public administrators, private sectors and researchers adopted the Dresden Agreement at the 5th National Summit in December 2010 which states in response to users' interest in convenient, standardized and user-friendly

access to Open Government Data, a centralized accessible Open Government Platform was developed [44].

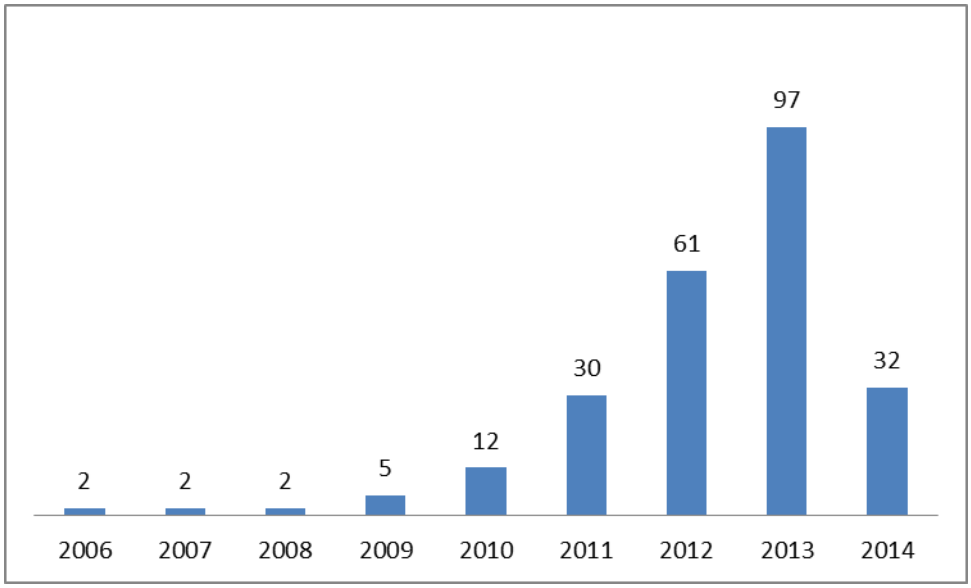


Figure 14: Publications Distribution by Years (n=243)

5.2 Research Methods

The sources of publication types were drawn from the databases, and the content were classified based on three Research Methodology paradigms [36, 65, 66] as shown in Table 8. They are Qualitative, Quantitative or Mixed methods. The definition of these three research methods can be clarified as below [66]:

- Quantitative method "*is one in which the investigators primarily uses postpositive claims for developing knowledge (i.e., cause and effect thinking, reduction to specific variables and hypotheses and questions, use of measurement and observation, and the rest of the theories), employs strategies of inquiry such as experiments and surveys, and collect data on predetermined instruments that yield statistics data*"

- Qualitative method "*is one in which the inquirer often makes knowledge claims based primarily on constructivist perspectives (i.e., the multiple meanings of individual experiences meanings socially and historically constructed, with an intent of developing a theory or pattern) or advocacy/participatory perspectives (i.e., political, issue-oriented, collaborative, or change-oriented) or both. It also uses strategies of inquiry such as narratives, phenomenologies, ethnographies, grounded theory studies, or case studies. The researcher collect open-ended, emerging data with the primary intent of developing themes from the data"*

- Mixed method "*is one in which the researcher tends to base knowledge claims on pragmatic grounds (e.g., consequence-oriented, problem-centered, and pluralistic). It employs strategies of inquiry that involve collecting data either simultaneously or sequentially to understand best research problem. The data collection also involves gathering both numeric information (e.g., on instruments) as well as text information (e.g., on interviews) so that the final database represents both quantitative and qualitative information"*

As a result, 86 studies (35.4%) were classified as Mixed Methods that remain the highest compared to the Qualitative (33.7%) and Quantitative Methods (30.9%). The growth of articles in the last 10 years show that the Mixed and Qualitative Methods articles widely published and contribute to the innovation of Open Data. Consequently, it shows that mixed methods become popular and qualitative methods falls more of the major approaches [66]. Even though mixed methods research uses both qualitative and quantitative, but the qualitative data are analysed qualitatively, and quantitative data are analysed quantitatively [65].

Table 8: Summary of Research Methods (n = 243)

	Research Methods		
	Qualitative	Quantitative	Mixed
No. of Publications	82	75	86
Percentage (%)	33.7	30.9	35.4

Another essential point from the Research Methods is which research strategies have been used in these publications. As mentioned in the definitions before, the major strategies of Qualitative research are varied from Phenomenological, Ethnography or Participatory Research which involved the authors' experiences, review, and observation; Case Study, Grounded Theory; and Archival research [65].

In contrast, the Experimental approach that are related to Laboratories work and internal validity such as project and prototype development where computation or formulation was needed and Non-experimental such as Surveys in which structured interviews, questionnaires and statistics involved for external validity are referred to Quantitative method [37]. Quantitative method is working more with such variables and hypothesis. Whereas, Mixed method which researchers mix both qualitative and quantitative research strategies.

Table 9 shows the type of research strategies identified which is used at least once in the publications. Most of the publications (57.2%) were using Experimental/Computation/Formulation as the research strategy which remained the highest among the categories. The second and third most common were Case Study with 94 publications (38.7%), while 75 publications (30.9%) were Phenomenology, Ethnography or Participatory Research strategy. Most of the case studies, phenomenology, ethnographies and participatory researches were carried out from the implementation of Open Data in the organizations such as governments or countries involved. 46 studies (18.9%) identified were surveys and most of the surveys are conducted from the per-

spective of stakeholders, users and organizations on open data tangibility and intangibility. Total of 20 publications (8.2%) were Archival Research which are analyzing secondary data such as Open Data government documents, administrative records, policies, and literatures [65]. Few studies were using Grounded Theory with only 10 publications (4.1%).

Table 9: Summary of Research Strategies

	Research Strategies					
	Experimental/Computation/Formulation	Case Study	Phenomenology/Ethnography/Participatory Research	Survey	Archival Research	Grounded Theory
No. of Publications	139	94	75	46	20	10
Percentage (%)	57.2	38.7	30.9	18.9	8.2	4.1

5.3 Organization Affiliation

Table 10 shows the categorization of the 243 publications based on the authors' affiliation: Academia, Industry, Government, Unknown and; Academia and Industry or Government or Unknown. The figure demonstrates that a majority 213 (87.7%) of the studies have author affiliated with academic institution. 3 studies (1.2%) were analysed and found that have authors from both academia and either one of industry or government. This small number might explain the difficulties experience in technology transfer by all the academicians from the lack of collaboration between academia and practitioners [165].

Table 10: Organization of Affiliation Types (n = 243)

Organization	ACM	Science Direct	Ebsco Host	Springer	IEEE	Conferences /Symposium /Workshop	Total	Percentage (%)
Academia	40	14	13	88	42	16	213	87.7
Industry	1	0	1	3	8	1	14	5.8
Government	4	0	0	7	1	0	12	4.9
Unknown	0	0	1	0	0	0	1	0.4
Academia and either Industy or Government	1	0	0	2	0	0	3	1.2

Another significant and interesting observation is to see if there is an increasing or decreasing trend in each of these affiliation categories by year as shown in Figure 15. The involvement of Academic organizations was so much higher and increased steadily to 90 publications in 2013. Meanwhile, the involvement by Industry or Government organization has remained fairly low consistently. However, Industry was increasing a bit higher than government organization but decreased in 2013. The involvement from both Academia and either of Industry or Government only can be seen from 2011 but consistently remained low. From this analysis, we can conclude that, the growth of articles on Open Data increased since 2010. However, more study or research has been conducted by Academicians in this period compared to the other affiliations.

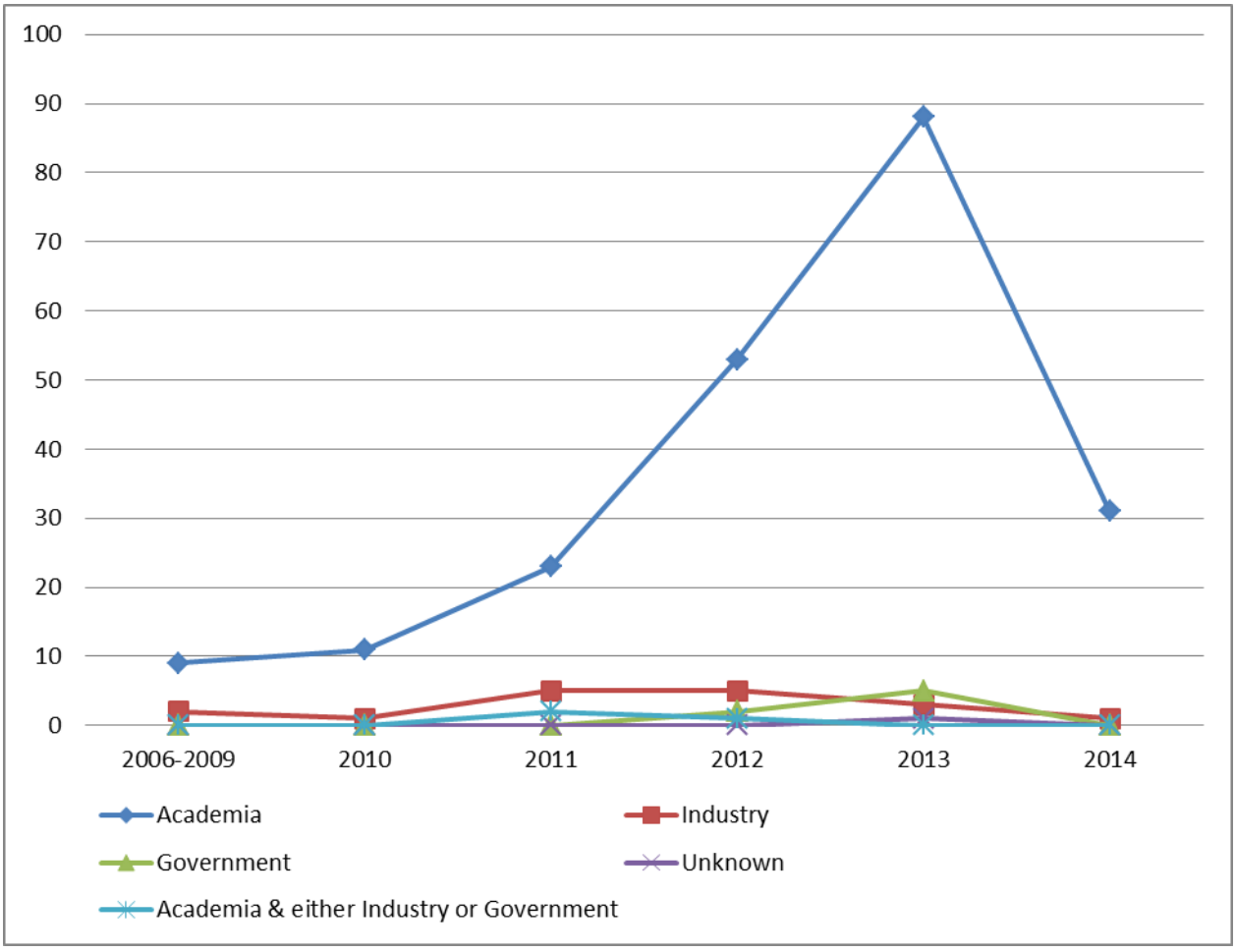


Figure 15: Organization of Affiliation by Year (n = 243)

5.4 Countries

The publications included were analysed on the Open Data publication-producing countries to see their level of contributions and activities have changed by the countries with respect to Open Data publications over the past 9 years. Furthermore, the more detailed analysis was done to see the trends between the cumulative countries contribution for the first 5 years, from 2006 to 2010 as shown in Table 11 and the remaining years, since 2011 as of mid of 2014 in Table 12. Even though there were some publications with more than one author came from different countries involved, only the first author (country) was recorded.

Table 11: Open Data Publications-Producing Countries 2006-2010 (n = 23)

Rank. No.	Country	No. of Publications	%
1	United States	6	26.1
2	Germany	5	21.7
3	United Kingdom	3	13.0
4	Austria	2	8.7
5	Italy	2	8.7
6	Japan	2	8.7
7	China	1	4.3
8	Finland	1	4.3
9	Ireland	1	4.3

**For any publications that have more than one country involved, the first author (country) was included.*

Table 12: Open Data Publications-Producing Countries 2011-2014 (n = 220)

Rank. No.	Country	No. of Publications	%
1	Germany	29	13.2
2	United States	18	8.2
3	Italy	17	7.7
4	Japan	12	5.5
5	Netherlands	12	5.5
6	Greece	11	5.0
7	United Kingdom	11	5.0
8	France	7	3.2
9	Finland	6	2.7
10	Spain	6	2.7
11	Australia	5	2.3
12	Austria	5	2.3
13	Brazil	5	2.3
14	Canada	5	2.3
15	Czech Republic	5	2.3
16	Ireland	5	2.3
17	Taiwan	5	2.3
18	Belgium	4	1.8
19	China	4	1.8
20	India	4	1.8
21	Sweden	4	1.8
22	Colombia	3	1.4
23	Singapore	3	1.4
24	Switzerland	3	1.4
25	Thailand	3	1.4
26	Unknown	3	1.4
27	Hungary	2	0.9
28	Kenya	2	0.9
29	Luxembourg	2	0.9
30	Mexico	2	0.9
31	Serbia	2	0.9
32	South Korea	2	0.9
33	Chile	1	0.5
34	Cyprus	1	0.5
35	Denmark	1	0.5
36	Ecuador	1	0.5
37	Israel	1	0.5
38	New Zealand	1	0.5
39	Pakistan	1	0.5
40	Poland	1	0.5
41	Portugal	1	0.5
42	Russia	1	0.5
43	South Africa	1	0.5
44	Venezuela	1	0.5
45	vietnam	1	0.5

**For any publications that have more than one country involved, the first author (country) was included.*

Contrasting Table 11 and 12 shows some remarkable trends as follows. In comparison, the number of countries researching on Open Data have been increased drastically more than 80% from 9 countries (2006-2010) to 45 countries (2011-2014). The growth of the publications-producing countries in the last four years might be motivated from some government memorandums [42, 43] and guidelines of Open Data principles [46] which are believed to promote openness and interoperability of public information without barriers of its reuse and consumptions. Germany and United States remained at the top of the ranking competitively for the past 9 years. However, the ranked list has a bit changed which the United States where on the highest of the list in Table 11 and Germany in Table 12.

In order to see the changes of the individual countries have been contributed to this trend, the top nine Open Data publication-producing countries (Germany, United States, Italy, Japan, United Kingdom, Finland, Austria, Ireland and China) in Table 11 have been investigated with these same countries in Table 12 by calculating the output percentage of those countries as shown in Table 13.

Table 13: Output Percentage of top ten ranked Publications-Producing Countries (n: 243)

Country	Output (%)
Germany	11.9
United States	7.4
Italy	7.0
Japan	4.9
United Kingdom	4.5
Finland	2.5
Austria	2.1
Ireland	2.1
China	1.6

Overall, the top nine ranked countries has shown so many increases in output as shown in Table 13 especially for Germany with almost 12%. The United States increases more than 7% at the second place and Italy about 7.0% at the third place. Japan and the Unit-

ed Kingdom increase by more than 4%. Meanwhile, Finland, Austria and Ireland remained fairly lower increases not more than 2.5% and China at the bottom with 1.6%.

In the same way, it is interesting to make a comparison between the results of the latest ranking organized by Open Knowledge Foundation [38] for 2014 which United Kingdom has the best score as shown in Figure 16. The United States is ranked only in the 8th country and is followed by Germany at the 9th of the ranking. However, Finland has shown an excellent improvement at the 4th ranked. These comparisons have shown totally different result because the Open Knowledge Foundation is the community-based surveys that could be not showing the whole picture due to the lack of government openness and sufficient civil society engagement. The countries are ranked accordingly to their information availability and accessibility in ten key areas. The key areas are Transport Timetable, Government Budget, Election Results, Government Spending, Emission of Pollution Levels, National Statistics, Company Register, Postcodes, Legislation, and National Map criteria.

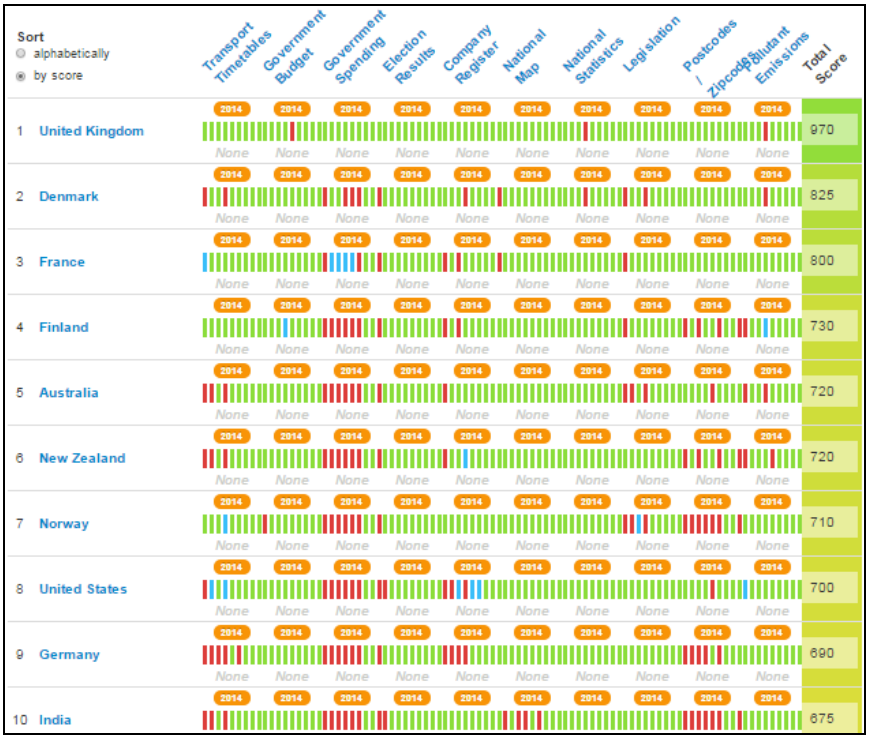


Figure 16: The Global Open Data Index 2014 [38]

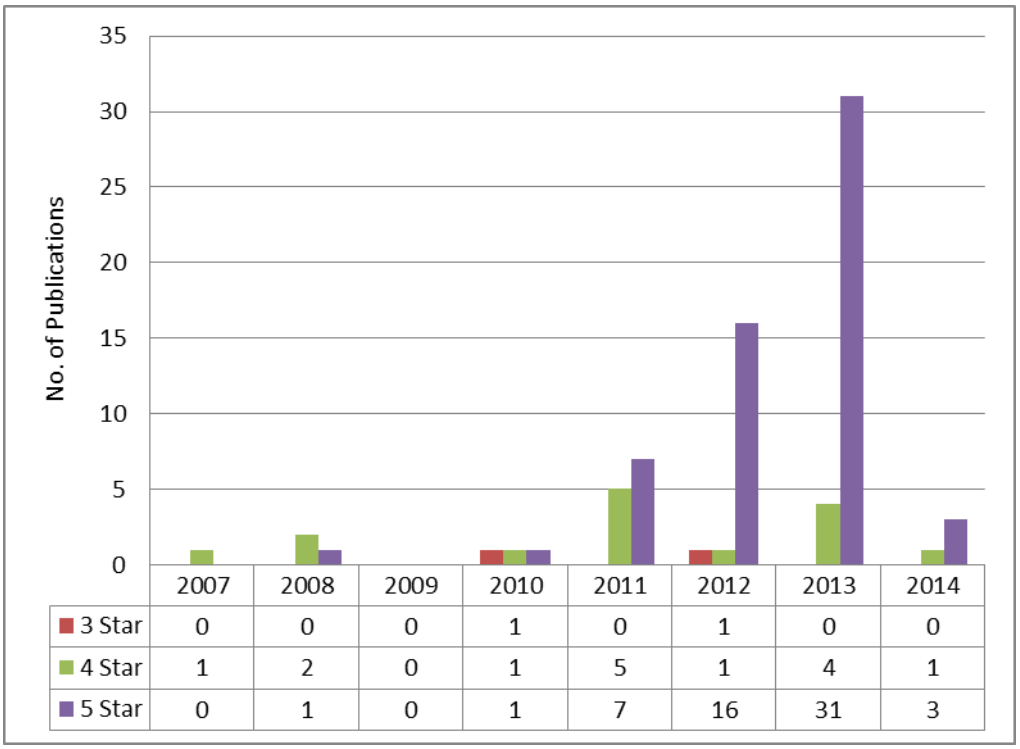
5.5 Star Rating

Table 14 shows the publications distribution by star rating which based on Tim-Berners Lee 5 star deployment scheme for Open Data [24, 25, 26]. The star rating can be identified based on the software development processes that have been mentioned in the publications. As a result, only 76 publications have fulfilled the 5 Star Rating criteria that 59 publications were rated as 5 Stars, 15 publications 4 stars and only 2 publications with 3 stars. However, none of the publications were rated with 2 or 1 stars.

Table 14: Number of Publications by Star Rating

Databases	Conferences/ Symposium/ Workshop	EbscoHost	ScienceDirect	ACM	IEEE	SPRINGER	No. of Publications
1 Star	0	0	0	0	0	0	0
2 Star	0	0	0	0	0	0	0
3 Star	0	0	0	1	1	0	2
4 Star	1	2	0	6	1	6	15
5 Star	1	0	0	5	29	23	59

Figure 17 demonstrates that 5-stars publications have the most exponent growth since 2011 compare to the other ratings. This is perhaps most of the publications that research on Open Data showed their best quality of Open Data practices and their usability improvement through their prototypes, applications or systems developed [62].



	2007	2008	2009	2010	2011	2012	2013	2014
3 Star	0	0	0	1	0	1	0	0
4 Star	1	2	0	1	5	1	4	1
5 Star	0	1	0	1	7	16	31	3

Figure 17: Publications Star Rating (n=76)

In addition, from the resulting publications by Star Rating, it is interesting to investigate the star ratings based on the authors' countries involved due to some issues on open data have attracted lots of government attention in the past few years. As an example, this deficiency is much serious in Taiwan since open data is still in its infancy and far from perfect [62]. The governments started to emphasize that the quality of open data, both in provider and data aspects, should be justified to ensure better usability and outcomes.

Table 15 shows the percentage of star rating by authors' countries which were included from the resulting publications. To no surprise, Italy is at the top of the ranking at 14.7% and followed by Japan (10.7%) and Germany (9.3%). In their research work, they have presented the solutions to improve the data usability, presentation, applicability, privacy, as well as other related issues to maintain the quality of Open Data.

Table 15: Authors's Countries Star Rating

No.	Country	No. of Publications with Star Rating	Percentage (%)
1	Italy	11	14.7
2	Japan	8	10.7
3	Germany	7	9.3
4	Austria	4	5.3
5	Greece	4	5.3
6	United States	4	5.3
7	Brazil	3	4.0
8	China	3	4.0
9	Netherlands	3	4.0
10	Thailand	3	4.0
11	United Kingdom	3	4.0
12	Australia	2	2.7
13	Canada	2	2.7
14	Ireland	2	2.7
15	Belgium	1	1.3
16	Chile	1	1.3
17	Cyprus	1	1.3
18	Czech Republic	1	1.3
19	Ecuador	1	1.3
20	Finland	1	1.3
21	Hungary	1	1.3
22	Kenya	1	1.3
23	Pakistan	1	1.3
24	Russia	1	1.3
25	Singapore	1	1.3
26	South Africa	1	1.3
27	South Korea	1	1.3
28	Spain	1	1.3
29	Sweeden	1	1.3
30	Taiwan	1	1.3

5.6 Units of Analysis

In order to summarize the included articles, different units of analysis were identified from the thorough review of the publications. These units serve as the basis for the analysis.

The formation of units of analysis was done by identifying the theme emerging from each selected study. Each of the included publications was reviewed by analyzing the context of the article's title, keywords, research methods, research interest, conclusion and empirical confirmation of the results. Then, each of the publications has clustered accordingly through the observation and which finally intended to define the particular unit of analysis of the reviewed research paper. The identified and clustered units based

on the summary of the publications and observations during the final review are shown below:

1. Development Lifecycle
2. Linked Open Data
3. Type of data
4. Technical Platforms
5. Organizations
6. Ontology and Semantic
7. Adoption and Awareness
8. Intermediaries
9. Security and Privacy
10. Supply of Data

Subsequently, the 243 publications were analyzed based on the Units of Analysis and then summarizes all the information of the publications selected and included. Figure 18 shows the distribution of publications that related to the different main Units of Analysis.

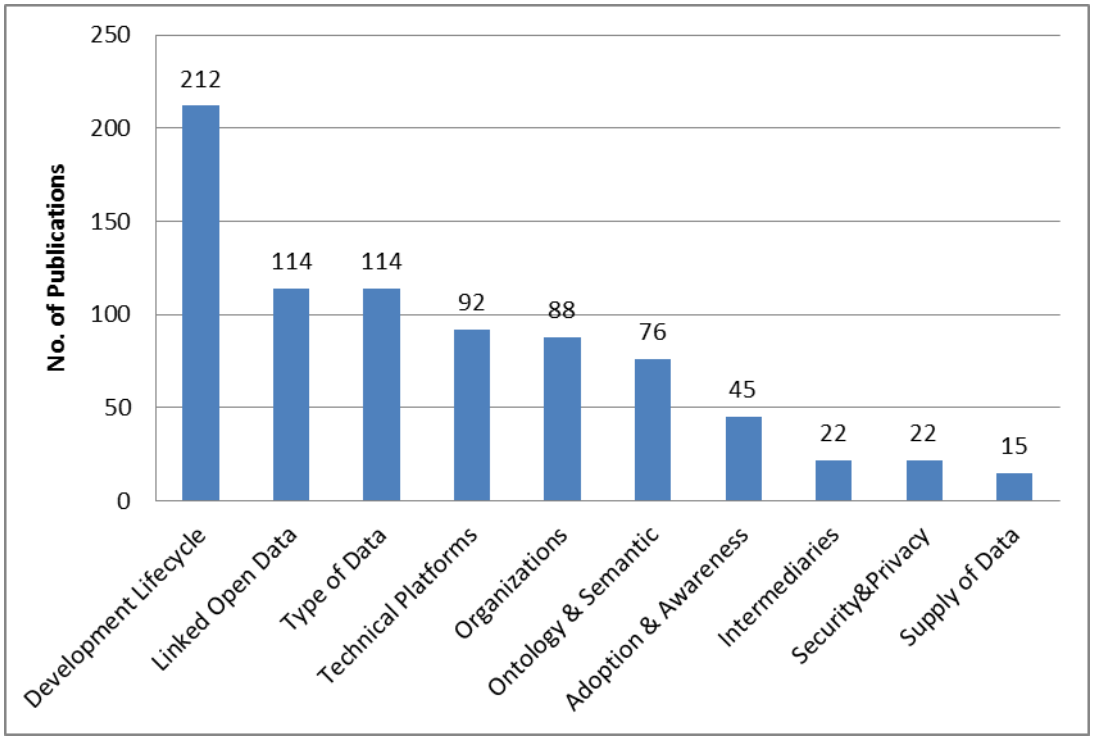


Figure 18. Distribution of Publications clustering by Units of Analysis

5.6.1 Development Lifecycle

From the distribution of publications in Figure 18, we can see that the most popular issues of Open Data are on Development Lifecycle. It can be seen that one of the challenges in Open Data implementation was on the technological aspects that all the development lifecycle or processes complexity were not managed and integrated well with data, tools and systems [47]. As an example, any Open Data that can be transformed and published is so-called Open Data infrastructures [67]. However, as many as Open Data infrastructures exist, users as well as policy-makers difficult to choose the most appropriate infrastructure for certain Open Data and its purposes. For this reason, the high level of enterprise architecture developed in the early stages of the development processes is believed to be useful [47]. Therefore, this issue was explored by identifying how a good Open Data software or applications were developed with different solutions and processes involved in the publications.

The core activities in development lifecycle include Requirements and Specifications, Design and Architecture, Construction, Deployment, Testing and Maintenance processes. The natures of these publications were mainly based on Experimental (Project, Computation or Formulation), Case Study and Survey Research Approach. In addition, some of the publications were presenting prototypes and examples to support their ideas. The detailed analysis and distribution of these activities is shown in Figure 19.

From the finding, most of the publications, 134 described on how the Design and Architecture of an applications, system or prototype in their research. 100 publications presented how the software was constructed, and 83 publications demonstrated the Deployment and Implementation of the software. 78 publications explored both on the Testing and Maintenance processes. Finally, more than 60 publications were presented the Requirements and Specifications of the software.

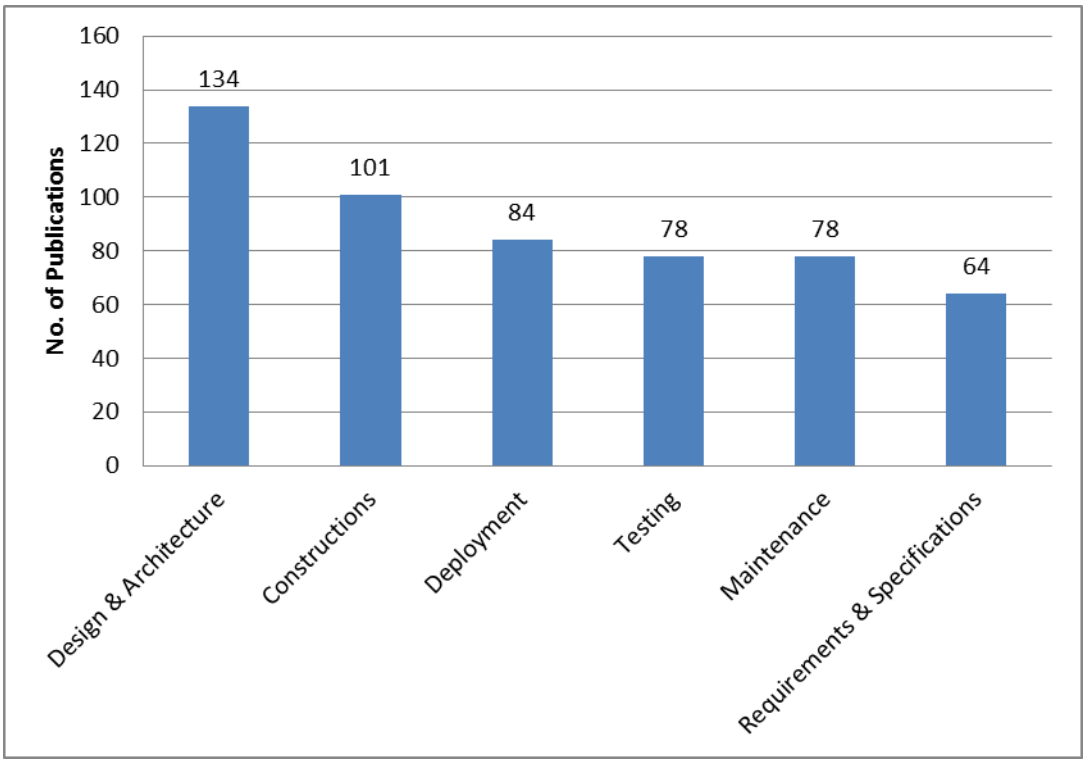


Figure 19: Distribution of Publications by Development Lifecycle Activities

5.6.2 Linked Open Data

Based on the results, Linked Open Data (LOD) was considered as the second top most popular research interest with 114 publications after development lifecycle category. LOD were popular research interest because these linked open data innovation led to a more sophisticated and viable market for LOD applications serve as a "public infrastructure" for the data publishers and businesses [68]. LOD principles or initiatives can be manipulated for data distribution with a better interface such as LOD cloud, visualization, statistics or crowdsourcing.

5.6.3 Type of Data

Equally important that there are about 112 publications described on the type of data which it is has a potential to be used in many uses and applications. In addition, the type of data has contributed to the Open Government Data (OGD) Higher Impact based on

the survey from consultations with stakeholders and OGD experts by 15 countries of four continents (America, Asia, Australia and Europe) [69].

Table 17 shows the distribution of the type of data reviewed and Governmental data remained the highest by almost 30% (32) of the publications. Governmental data were from the scope of government policies, regulations, socials, political or community. The second highest was Educational data with 14 publications (12.5%). The ranges of the scope of the educational data were the publications that included the teaching and learning process, educational institutions, expert profiles and researcher networking. Then, science followed as the third highest with 10 publications (10.7%) which data were produced as part of scientific research such as biologists, chemistries and life sciences. Total of 8 publications (7.1%) were Cultural data that related to cultural works and artefacts included humanity, linguistics and arts. The Geographical and Healthcare type of data were sharing the same number with 6.3%, 7 out of 112 publications. Most of the applications presented in the publications were related with geographical mapping data which were using geo-data or geo-spatial. However, healthcare data was used for some clinical and medical measurement. The remains out of 112 publications were from Environmental, Multimedia, Statistical, Transportation, Entrepreneurship, Computing, Gamification, Tourism, Agricultural, Crime and Sensor.

Environmental data was information that related to natural environments such as the presence of disaster management, the level of pollutants and the environmental quality. On the other hand, Multimedia type of data consisted of images and movies information that some were used in entertainment industry. Furthermore, Statistical type of data was defined for the data produced by demographic census and socioeconomic key indicators. Entrepreneurship data which included the enterprises, business and economy information led to the importance of certain countries benefits or advantages in income growth. Few of the publications were researched on the transportation data such as timetables, routes, on-time statistics or surveillances and some of applications were produced. Another interesting data was Gamification which some publications described the applications which were using game data for entertainment which were sharing the

same percentage 1.8% with Computing and Tourism Categories. Agricultural, Crime, and Sensor category were at the bottom lowest compare to the other categories.

Table 17: Distribution of Publications by Type of Data

Type	No. of Publications	%
Governmental	32	28.6
Educational	14	12.5
Science	12	10.7
Cultural	8	7.1
Geographical	7	6.3
Healthcare	7	6.3
Environmental	6	5.4
Multimedia	5	4.5
Statistical	5	4.5
Transportation	4	3.6
Entrepreneurship	3	2.7
Computing	2	1.8
Gamification	2	1.8
Tourism	2	1.8
Agricultural	1	0.9
Crime	1	0.9
Sensor	1	0.9
	112	100.0

5.6.4 Technical Platforms

92 publications were discussing on the technical platforms of Open Data, which commonly explained on how Open Data can be stored, the data repositories and how the data were made available to the others to use. However, it is not an easy matter for releasing large datasets, as it comes with several challenges including data storage [49]. Furthermore, the early planning of Open Data is very important as making data available can be complicated such as making data anonymous, writing metadata and transforming data into the machine-readable format. In publishing open data, interoperability is an important aspect and integrated system in a framework. From the literature review, Open Data platforms can be identified either from the project or research infrastructure.

The challenge of Open Data platforms is how to make the data available in bulk and a useful format.

Overall, Linked Open Data (LOD) Cloud is the main data linking principle that has been addressed a lot in the publications. With more than 200 datasets interlinked by RDF links in 2011 such as DBpedia (Wikipedia), Geonames, NYTimes, LinkedMDB and Billion Triple Challenge (BTC). As an example, dumped Wikipedia content was used in the prototype platform to observe the success of linking through DBpedia [70].

5.6.5 Organizations

It is interesting to see the publications' rights of ownership for the data generated by a project (systems or applications). As the result, 86 publications datasets identified were related to any organization's ownership, government, industry, academia or public agencies as shown in Table 18. To no surprise, the majority (50%) of the publications data were owned by the government (Governance divisions, Policies and etc.). More than 20 publications from academic (Universities, Research Institutions, Libraries and etc.) and 17 publications were public agencies such as maritime, transport, museum and biologist agencies. From this table, data owned by the industries show the lowest numbers with only 4 publications.

Table 18: Distribution of Publications by Organization (n=86)

Organization	**No. of Publications**
Government	43
Academic	22
Public Agencies	17
Industries	4

5.6.6 Ontology and Semantic

76 publications focused on the Ontology and Semantic research interest. Ontologies are critical since they are associating precisely defined semantics with the content on the web allowing obtaining more rigorous and wider information. Different ontologies were used specifically by the datasets of LOD to describe instances that cause the ontology heterogeneity problems [94]. Heterogeneous Ontologies handling were challenging and learning LOD big ontologies manually was time-consuming. These problems can be decreased by integrating related ontology classes and property automatically from interlinked instances. In addition, integrated ontology approach helps Semantic Web application developers query easily on various datasets. As an example, for United Kingdom Publicspending.gr ontology was introduced which was developed from basic by reusing some components from the corresponding Payments Ontology of the British established vocabularies and “Opening up government” project, data.gov.uk [68]. Specifically, the transparency and efficiency in the economic domain would improve by enabling the Open Data of semantic processing public budgeting and spending, e-procurement and business information.

5.6.7 Adoption and Awareness

There are also challenges related to community or organizational awareness of Open Data [47, 104]. For example, the majority of people prefer a closed version of data and most of typical organizational culture characterized as top-down, hierarchical, command-and-control, and siloes [47]. However, the trend for Open Data that can be accessed by third parties such as developers has been growing recently. As an example, Open Data Institute (ODI) which encourages others to release their information has fully supported from the government in the United Kingdom [49]. For Adoption and Awareness, 34 publications (74%) were related with at least one of Open Data Policy, Initiatives, Directives or Strategies topic interest. Only 12 publications (26%) were presenting Open Data challenges and opportunities as shown in Figure 20.

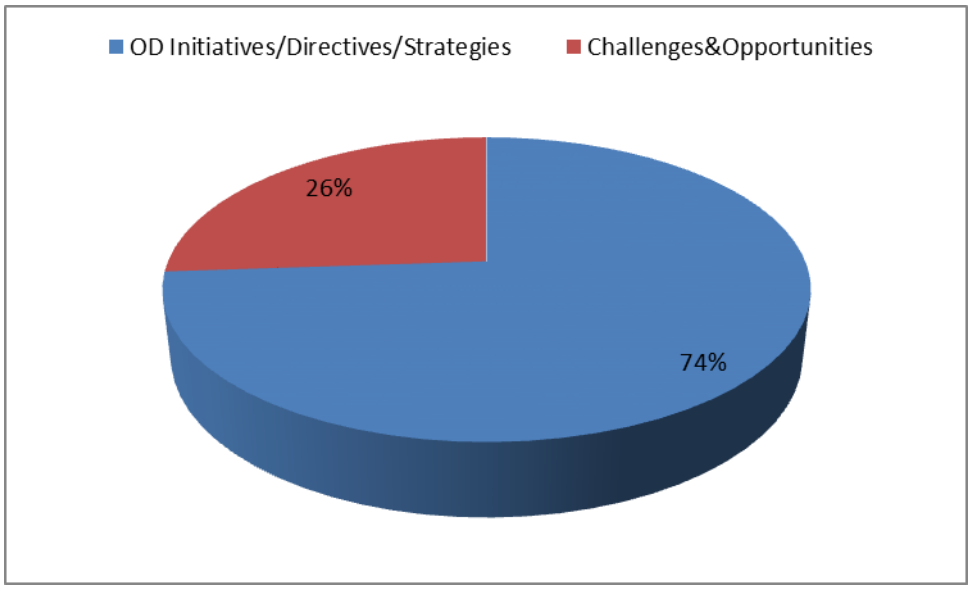


Figure 20: Number of Publications with Adoption and Awareness sub-categories

5.6.8 Intermediaries

Only 22 publications were identified within this unit of analysis. Data Intermediaries' sub-categories can be from the activities of Data Extraction, Data Mining, Data Visualization and Data Mapping. Data intermediaries were important which people and tools can use data and visualize them in the most useful way [166]. This criterion is also important to avoid any incomplete, poor quality and inaccessible Open data.

5.6.9 Security and Privacy

Security and Privacy category focused on how the data are protected. Only a few publications were focusing on the security and privacy issues with only 22 publications. Enhancing security and privacy of Open Data information is important and often late in a software development even though for certain applications and initiatives because risks can be perceived to outweigh potential benefits [47]. The concerns can be the systems vulnerability to acts of hacking, Denial of Services (DoS) attacks and intrusion of spyware and malware, and the risks of accidentally disclosing confidential information such as personally identifiable information.

5.6.10 Supply of Data

The least number of publications were researching on the supply of data with only 15 publications. These types of data supply issues can be from the crowdsourcing, crowdsensing or websensing technologies. Some more research should be done in this area as applications or software with sensing ability is quite trendy nowadays.

6 Qualitative Analysis

The results of the Qualitative Analysis are described in this chapter.

6.1 Development Lifecycle

The success of a product depends on the activities in Software Development Lifecycle involved. Further inspection has been done to identify the connection based on open data quality (5-star rating) and the development lifecycle of the software as shown in Table 19.

Table 19: Distribution of 5 Star rating Publications clustering by Development Lifecycle

Development Lifecycle Covered	**No. of publications**
Requirements & Specifications, Constructions, Design & Architecture, Deployment, Testing and Maintenance	8
Four of the processes	18
Three of the processes	16
At least one or two of the processes	15
None of the processes	2

As the result, from 59 publications that were rated with 5 star rating, eight of the publications presented all the full development processes involved consist of Requirements and Specifications, Constructions, Design and Architecture, Deployment, Testing and Maintenance. Furthermore, 18 publications presented four out of five processes. Meanwhile, 16 publications are having three of the processes implemented. Similarly, 15 publications involved at least one or two of the processes. It can be seen that only two publications with 5-star rating have not involved in any of the processes. Considering these factors and very few of the publications were not implementing software development lifecycle in producing open data applications, the quality of Open Data applications can be achieved by following software lifecycle management. Subsequently, this is important in order to ensure Open Data applications success specifically to identify from which process the defects and failures arise.

Another interesting finding identified is the potential of the Software Development Lifecycle linking together with the Data Management Lifecycle [25]. Data management lifecycle is a policy-based approach to managing the flow of an information system data throughout it lifecycle: from creation and initial of storage to the time when it becomes obsolete and is deleted. There are numerous versions of data management lifecycle but are not limited to the basic steps of Plan, Collect, Quality Control, Document, Preserve, and Use. One of the examples is as shown in Figure 21 [75].

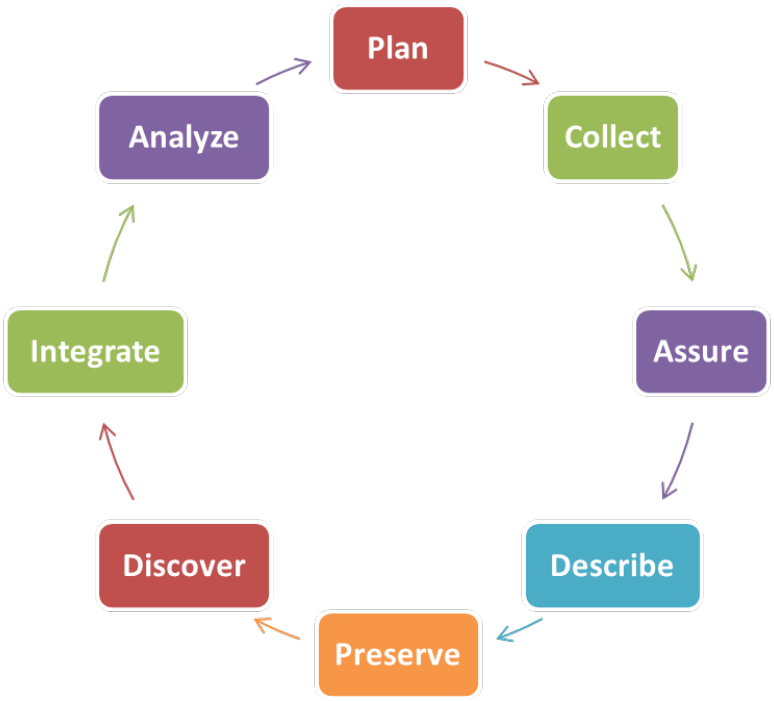


Figure 21: Data Management Life Cycle [75]

However, after linking the Software Development Lifecycle and Data Management Lifecycle together, it shows some steps that are sharing certain symmetry during quality assessment as indicated by red arrows in Figure 22 [25].

From this comparison, further discussions can be explored to produce sustainable scientific software [25]. Firstly, about the metadata that there has increasing value added through the creation as data moves through the lifecycle. Secondly, the preservation management principles which is important in the context of validation of science soft-

ware. The third is the explicit connections between data flows into or out of science software. And finally, standards which are recognized internationally for both data and software. These factors indirectly can contribute to develop useful Open Data applications that the software and data are maintained together throughout the whole lifecycle.

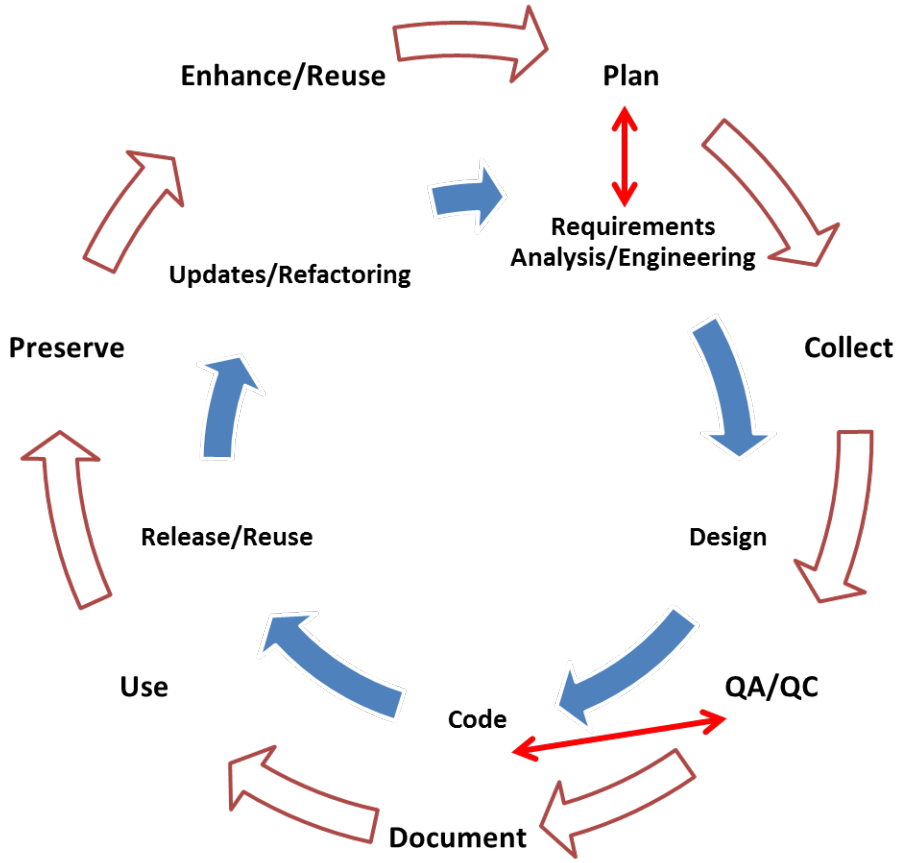


Figure 22: Linking Data Management Life Cycle and Software Development Lifecycle [25]

6.2 Technical Platforms

From the literature analysis, six platforms of Open Data were identified. The seven platforms are comprised of Linked Open Data (LOD) Cloud, Open Datasets Concept and databases as storage, Triplestore Repository, Comprehensive Knowledge Archive Network (CKAN), Application Programming Interface (API) and Virtuoso and are explained in the next sections.

6.2.1 Linked Open Data (LOD) Cloud as a database

LOD cloud as a single data-space or web-scale database is purposely to make relational databases comparison. However, it is challenging to understand LOD as a database and to identify LOD specialised architectures requirements to enable the essential principles of Relational Database Management Systems (RDBMS). National Data Sharing and Accessibility Policy (NDSAP) encourage any organization to launch their Open Data Initiative to facilitate the cycle cloud-based solution by requesting over the cloud as SaaS [71]. In the same way, Television Linked to the Web (Linked TV) is a ubiquitous online cloud of networked Audio-Visual content decoupled from a device, place or source [73]. The growing online datasets that were linked data were presented in the Linked Data Cloud. Different Web sources, for example the BBC and Musicbrainz, have embraced Linked Data and begun to publish their data according to Linked Data principles. They are making their content (e.g. different BBC TV programmes, or music artists) by unique reference ability (URI) and to provide their machine understandable information. In addition, for environmental databases such as SILO, AWAP, ASRIS, CosmOz and MODIS were also using LOD Cloud, which implementing Semantic Machine Learning approach [72].

6.2.2 Open Datasets Concept and Databases as Storage

Some platforms or infrastructure were releasing their open datasets. The database was working with SQL query for information retrieval engine that integrated with Database Management System (RDBMS) [76]. Numerous open datasets from SQL databases were made public through the internet [77]. Some researches were integrating various types of databases using related ontologies [78]. The multiple related LOD can be obtained as a whole by users while presenting the relationships between multiple LOD through ontological hierarchical structure as well as by semantic link relation. It is included search service of abbreviations and long forms automatically extracted from repositories [79]. In “Hub and Spoke” approach, different databases, OCW (OpenCourseware) and OER (Open Educational Resource) repositories were using the same vocabulary modeling in Linked Open Data environment by running queries of repositories content through SparQL-Endpoint, APIs and web services [80]. SPARQL Query Endpoint, which masked with Graphical User Interface (GUI) linked Internal (back-end services) and external (front-end services) to the repositories [74].

6.2.3 Triplestore Repository

The triple store repository commonly was exposed by using SPARQL End-point, which is accessible over HTTP. The triplestore is a framework used for querying and storing RDF data in persistent storage and access of RDF graphs [58]. From the literature, there was major development initiative in triplestore technologies, query processing and access protocols. “Sesame triplestore” is the storage inference and RDF data query open source framework based on triple latest ontology. This sesame matches the feature of Jena, a Java framework for building Semantic Web applications which once the API connection and inference support for multiple back-ends are available similarly like MySQL and Postgres. The integrated knowledge of RDF files was stored into the triplestore which all functionalities of browsing, querying, exporting data in various formats were provided. As an example, while D2R server9 was used, the generated triples are loaded in a Jena TDB instance, a native high performance triple store accompanied by a Fuseki component which expose triples as SPARQL end-point accessible over HTTP [81].

Another example of triplestore ontology, there was a web-based prototype, ProLOD using profile triplestore in a relational database to run the basic database operations [83]. Several computations involved in the pre-processing on the input data. Firstly, the triples are processed and stored in a normalized database schema. Secondly, each triple is enriched with the data type and pattern information. And finally, an initial clustering is computed and cluster-labels are determined. Another way of incorporating between two or more datasets can be done by a small programming scripting such as Python, Ruby, Visual Basic and Bash. The problems with fields can easily be corrected using a text editor or a data cleaning program such as a freely available open-source data manipulation or cleaning program, Google Refine [82]. There is a research that examines the novel data integration method in the case of *publish-time.* The publisher is suggested with options on how to integrate the new dataset with the corpus as a whole, without resorting to the created mediated schema or ontology for the platform manually [85]. The concept is called Open Data Repositories, where the new dataset attribute names are aligned with the corpus to which the dataset should be added. While the users publish the data to these platforms continuously, this led to the steady increment of schema vocabulary.

6.2.4 Comprehensive Knowledge Archive Network (CKAN)

Comprehensive Knowledge Archive Network (CKAN) was considered as a powerful and open data world's leading open-source data portal platform that have been implemented worldwide. CKAN is a tool to streamline publishing, sharing, searching and using of data, including data storage and robust data APIs provision. CKAN were used widely in governments, industries and organizations [54]. CKAN Meta portal was used for storing metadata information that can be linked to Open Data resources and Web Portal interfaces [10]. The metadata included the data, purpose, maintainer and licensing descriptions. The CKAN's approach is ambivalent which offer well-documented REST API and web access based on JSON (JavaScript Object Notation) interfaces [60]. With CKAN, the databases for structured storage of data were with a powerful web-accessible Data API.

6.2.5 Application Programming Interface (API)

Application Programming Interface (API) can be also considered as a new alternative way of making Open Data available especially on the Cloud hosting [1]. As an example, Publicspending.gr and openspending.org provide Open Data APIs that can be queried on demand and results are returned in JSON and XML formats [68]. In addition, web APIs open a new possibility for TV data services by giving access to the other repositories with Linked Open Data [73]. APIs server-less front-end was also used where Flash proxy system or JavaScript handles the back-end services communication [74]. In Real-time Open Sensor Network Data platform, an API was provided to download all the standard formats of structured data which were stored into a high-speed database by different city sensor networks [10]. In general, API is purposely to handle any hierarchical data structure [86].

6.2.6 Virtuoso

Virtuoso is also a tool to publish linked data and provide SPARQL End Point. Virtuoso is considered as the unique hybrid server architecture which offering RDF Data Management, XML Data Management, Relational Data Management, Document Web Server, Linked Data Server, Free Text Content Management & Full Text Indexing, Web Application Server, Web Services Deployment, SOAP or REST [55]. Virtuoso can be queried using SPARQL to store RDF descriptions (graphs) and provides the corresponding APIs for issuing SPARQL queries on RDF graphs. As an example, OpenCityGuide application which was developed on top of the Open Data platforms with different cities offer different touristic data as LOD [60]. By using SPARQL queries over the Virtuoso interface, the application accesses the touristic data located at the various open data platforms and aggregates them in a single application. In addition, the *OpenCityGuide* consists of several backend and frontend components which include a database where all the data generated by end-users such as comments, ratings, and check-ins are stored. Virtuoso also provides JSON REST interface that touristic data cache was synchronized with the integrated open data platforms which can be updated on-demand or periodically.

There was also research discussing how data is stored either by distributed (decentralized) or centralized approach which were using data release from six local public services City Development (Stadsontwikkeling), City Works (Gemeentewerken), City Archive (Gemeentearchief), Library (Bibliohteek), regional Police and Centre for Research and Statistics (Centrum voor Onderzoek en Statistiek) [87].

Another storage was introduced which was built on the distributed framework called Virtual Repository (VR) which serve supplement information sources. Automatic Disaster Alert System for Tourists (ADAST) is an example of an application that is using proof-of-concept VR prototype [88]. With ADAST VR tools and processes, all or several data in each source was published by following LOD practice. Depending on the VR applications, the translation layer stored in computers and devices produce the data links and linked data. However, some or all of LOD are also uploaded to the VR core in a cloud-based server. In the similar approach, virtual RDF which accessing the non-RDF database was using the RDF view. This approach enables the access of non-RDF legacy databases without need to replicate the whole database into RDF. As an example, two microRNA biomolecules LOD databases, miRNA and miRBase were built using D2R LOD infrastructure which imitates the Virtual RDF approach for publishing the content of relational databases on semantic web [89]. Databases content is mapped to RDF using the D2RQ declarative mapping language that captures mapping between database schemes and RDFS/OWL schemas.

There was also a framework for integrating data both from structured data sources (LOD) and unstructured data sources (news reports and web articles). This can be done by populating a knowledge base (KB) with Semantic Web Technology Evaluation Ontology (SWETO) instances and store the resulting Resource Description Framework (RDF) 2 triples [90].

Another alternative on publishing Open Data is by using a ready-to-deploy server and data repository that is so called Aggregate Framework concept. The ODK Aggregate server is an example which the framework can be deployed on Google's App Engine to enable users to get running without facing the complexities of setting up their scalable web service [91]. This aggregate server can also be installed locally on a Tomcat server

backed with a PostgreSQL or MySQL database server. It also provides a platform to manage collected data, visualize those using simple graphs and maps, and export them as CSV files or other spreadsheet file format published to external systems. The other examples are the aggregation framework based on DBpedia and Yago datasets [92, 93]. In the initial proof of concept application, two knowledge bases were created, DBpedia Property Dump and a Yago Classification Dump. Before aggregating all the distinct property, SNORQL query explorer (SPARQL Endpoint of DBpedia) querying each type of person to build DBpedia Property Dump.

6.3 Organizations

Open Data Initiatives was considered as a type of cross-boundary information sharing between government agencies and public including industries, individuals and non-profit organizations [170]. Cross-boundary information sharing was defined as the data sharing between entities such as groups, departments, and organizations by interconnection and collaboration of different telecommunication technologies and information systems [171]. The current research mainly focuses on inter-agency information sharing as Open Government has become important policy among government administrations around the world. Another highest impact from the partnership of government with NGO's or private organizations was the popularization of Open Data on mobile devices and social media platforms and tools [172]. Any Open Data applications developed not just for sale but obtaining more efficient productivity and businesses between partners.

The government has to invent or innovate creative solutions in promoting ecosystems where all vendors, developers, and citizens could manipulate government available public information or regulations as a solution to any public issues. One of the aims in The European e-Government Action Plan 2011-2015 is to harness citizens and businesses involvement in the policy-making process in 2010 [173]. This collaboration enhances the government efficiency by supporting cooperation and partnership with Federal Government, across levels of government, and between the government and

private sectors [174]. As an example, there was a project in India to aggregate Indian government, the private sector, research institutes and NGOs open data in a one-stop entry to all the aspects of the entities [175].

6.4 Ontology and Semantic

In ontology and semantic findings, a good example of a framework called Montage was introduced which creates domain ontologies in the LOD Cloud datasets available [95]. This framework allows a domain ontology schema was created by a domain expert by defining its properties (data and object) and classes. Then, the defined schema was mapped onto the selected LOD datasets. The new ontology classes and properties were populated by acquiring data from selected datasets in the form of SPARQL CONSTRUCT.

Another example is in the medical research where a semantic web-based knowledge representation that integrates information about clinical trials from some LOD sources [96]. These representations are augmented by linking medical concepts to UMLS (Unified Medical Language System) medical ontology to enable more powerful queries over the clinical trial information with consistent semantics across concepts used in different sources. With powerful patient user interface semantic search capabilities, the navigation from one concept to another can help users with visual search and exploration of clinical trials and related information.

Furthermore, there was a research proposed a new approach based on LOD principles that the automatic retrieval of Learning Objects can be done from local or external LOD repositories [97]. The approach was innovated to interlink the external knowledge with domain ontology in the LOD cloud. The ontology is also needed to overcome the difficulties of integrating heterogeneous datasets due to different vocabularies which described the same type of data especially in Open Data platforms [98]. This can be done which an ontology is corresponding URIs that express the same relation and allowed the Users or Administrators to add or edit entries in the ontology to maintain the good

quality. The Broadly Integrated Ontological Linked Open Data (BioLOD.org) is an example of ontology-based integrated database that offers downloadable OWL/RDF graph files of mutually linked public biological data organized as a semantic web using World Wide Web Linking Open Data Consortium project standardized formats [78]. BioLOD.org mined numerous semantic links from primary databases and re-classified them into graph files based on ontology classifications. The connections between the files were clearly and mutually referred so it is easy to find other files associated by semantic links included in specific data instances.

Often, the lack of integrated schema of ontology in large corpora of dataset happened by the inconsistency of names for an attribute with the same meaning. This factor contributes to the difficulty in the discovery of connections between datasets as well as their reusability [85]. As an alternative solution, the novel method of data integration was investigated on how to integrate the new dataset with the corpus as a whole based on the data-driven algorithms. LOD is part of Semantic Web, and the content data can be queried easier than crawler compared to the other knowledge bases [99]. With Semantic technologies, semantic metadata matching mechanism led the data providers, application developers, and the general public accessing and downloading data and integrate automatically [61, 72]. As an example, the development of i-EKbase according to semantic data integration and machine learning recommendation which provided great flexibility in terms of data, knowledge and provenance integration [72]. Qualifying Knowledge Acquisition and Inquiry (qKAI) is an application framework that narrowing some remaining gaps between the hands-on web 2.0 and sophisticated semantic web, enabling loose-coupled knowledge and information services [59].

In several Open Data Initiatives, there are various formats of data services, such as SaaS, PaaS and IaaS which were published online based on cloud computing era nowadays. However, it is challenging to utilize fully the data by communicating and answering queries from different sources automatically, dynamically and in more meaningful manners [100]. To overcome these challenges, semantic web ontologies were used to annotate the data service semantically, and semantic-based data integration was achieved by implementing service composition technique. Several issues were consid-

ered from the interface designing aspect between the Semantic Web and its users [93]. The issues consisted of potential resources from LOD were discovered, structuring the discovered resources and the structured information was presented in an easily understandable way. Semantic (metadata, linked data and multilingualism) technologies play a significant role in easing integration and published them as Linked Data focused on quality and openness [101]. As the result of the spreading Semantic Web and LOD initiative recently, a vast amount of RDF data was published in freely accessible datasets that were connected with each other to form LOD Cloud. These data can be utilized successfully to build Recommender System (RS) that relies exclusively on the information encoded in the Web of Data [102]. Several works on Ontological RS which based on the ontologies and semantic technologies led to boost collaborative filtering systems or to build smarter content-based systems [103].

6.5 Adoption and Awareness

With the citizen involvement, Open Data is recognized as a solution to improve economic welfare, policymaking and public decision-making. The new transformation of open government and open data initiatives led the access of public data not limited only to a few key decision makers but to more participatory and democratic access of data between government and society actors [105]. In 2010, the European e-Government Action Plan 2011-2015 initiated by European Commission to harness Information Technology for public information easy access improved transparency and increased citizens and businesses participatory in the policy-making process. However, the challenges happened at the social, use, political, and technical level.

The new strategy of Open Data initiatives in the aspect of participation, collaboration, transparency and empowerment can be done by combining the high diversity of datasets, sophisticated information technologies and powerful analytical methods [105]. In addition, decision-makers, policymakers and scientists were interested in combining datasets that led to scale Open Data applications, predefined data silos, to produce large numbers of datasets having heterogeneous data formats. These challenges can be man-

aged firstly by delivering incentive policy guidelines to stimulate the centralization of open data repositories and to revise the fragmentation [106]. Secondly, by creating Open Data access enabled to users. Thirdly, to guarantee easy discovery and recognition of its potential, a structured metadata should be added while making data available by creating interoperability. And finally, this challenge can be handled by creating a good infrastructure. There were differences Open Data policies implementation in different organizations especially in the guidelines and principles for publishing data, the amount and type of data [107]. This was shown through a framework which focusing more on different kinds of policy-relevant information such as policy's problems, actions, outcomes and its performance.

There are several challenges and opportunities for the services of Open Data. The use of open data sources was complicated by the heterogeneous Open Data, non-standard APIs and varying licensing conditions in commercial applications and services [108]. In the same way, business opportunities were found such as selling data-based knowledge, providing information that competitors do not have, integrating data and data mining. Furthermore, there are increasing need for raw data, information, knowledge, predictions and market information. In addition, there are strong dependencies between an application or services in different regions and countries due to non-standard APIs. In order to provide public sector information transparency, new platforms and tools are needed to suit with the emerging technical challenges [60]. Only a better understanding of these challenges and opportunities can ultimately lead to more and better services for citizens, consumers, as well as organizational users.

For the case study in Taiwan, the development of Open Government Data (OGD) "value-added" government information concept was investigated [109]. The six criteria of Open Data policy were observed and considered for OGD progress. They consisted of legislation and policy, types of open data, technological standard and data format, open data promotions, single open data platform and appropriate licensing but with some challenges. One of the challenges comes from the gap of mutual understanding from the government agencies. Government agencies pay less attention on the reusability of government information and data; value added features and the needs of the public. Imma-

ture operational model is another challenge for the government agencies while stick to the respective regulations and act inflexibility towards value added purposes in releasing information and data to public. In addition, high transaction cost is an issue happened when dealing with several government agencies to obtain any data or information. The last challenge is in asymmetry information that both public and private organizations do not know that government data and information are available and permitted for value-added purposes. However, the encountered challenges can be reduced by strengthening the Open Data policy by the six dimensions of the value-chain creation which reutilizing government information and data.

6.6 Intermediaries

Intermediaries such as data analysis and visualization are important especially in exploring large-scale datasets or Linked Open Data [167]. In addition, numerous of government data available on the websites can be both useful and complicated because not all of the data itself is necessary to be used by every user [168]. It is important to create new useful information and knowledge from the data itself that can be done by implementing data mining and data analysis techniques. In acquiring useful information, the open data in CSV format that transformed into applications such as Excel and SPSS can be analysed statistically. Besides obtaining new knowledge, analysis leads to establish concepts for processing the data that identify new methods and tools to collect broader new information. In the same way, data mining techniques are used to analyse open data and provide a user-oriented approach to any novel and hidden patterns of data to create a richer environment of knowledge. Data mining has earned significant attention lately and improved from the computing technology advances and information digitalization.

An excellent tool or platform help the programmers to bear on the aforementioned data analysis and visualization challenges [167]. From a report on the experiences of pioneer entrepreneurs of Open Data which companies that offering for opening datasets, most of the companies performed their data extraction and transformation themselves before analysing process [13]. This is because more time is needed in finding, extracting and

transforming data compared to perform the actual analysis. Data analyser's roles are to create visualizations or algorithms based analysis of data and usually cross-analysis data from multiple sources to present powerful results. The data needs to be extracted from its original source and transformed to an analysable format before it can be analysed. Still, in government Open Data the application of data mining must care so much especially on the data integrity, incorrect analysis of data during processing and improper usage. These obstacles could cause scepticism of end-users and additional costs to the Government of the user's damage.

6.7 Supply of Data

In the new world nowadays, data is characterised by a new series of information suppliers or infomediaries for citizens, who are able to mash up information from various sources (not only from government) to create applications which deliver services (either free or for payment) to citizens, and provide alternative sources of information to the mass media. This is how crowdsourcing emerges refers to a new distributed social collaborative problem-solving model in which a crowd of undefined size is engaged to solve a complex problem through an open call [57]. Smartphones and mobile devices can unfold the full potential of crowdsourcing. Crowdsourcing also offer benefits by saving cost by releasing private data. Open Government Initiatives was launched with the rise of crowd-based initiatives and networks [117]. Crowdsourcing approach is a solution for the tasks that cannot be solved by using automatic methods which now is gaining interest in the data integration community [154]. A pilot survey was conducted among memory institutions (galleries, libraries, archives and museums) to observe the trend of Open Data and crowdsourcing strategies [155]. With regard to crowdsourcing the data suggest that the adoption process is slower than for open data.

7 Discussions

This chapter presents the answer to the Research Question 5 defined in the earlier chapter and the trends in the studies of Open Data were considered as well as a conceptual framework was proposed in order to organize the whole ideas or principal findings of the study. The limitations of the review, validity to threats and future research works are also identified and discussed.

7.1 Conceptual Framework for Research

The conceptual distinctions and idea of research were organized in a conceptual framework to examine the importance of Open Data based on the examples of readiness assessment and implementation evaluations listed in Open Data in Developing Countries (ODDC) Conceptual Framework for Open Data, Governance and Emerging Impacts [147].

7.1.1 Open Data, Governance and Emerging Impacts

In order to identify the commonalities between open data publications and use cases, the conceptual framework of ODDC Open Data, Governance and Emerging Impacts was adopted as shown in Figure 23 [147]. In the same time, the impacts that open data might be involved should be defined. To distinguish this, the best option for Research in Open Data according to Helbig et al. recommendations is to investigate the context and dynamics of embedded open data within, by developing a framework of an ‘information polity’[148]. This framework consists of all the stakeholders, data sources, information flows data resources, governance relationships involved in the provision, and governmental and non-governmental data sources of use.

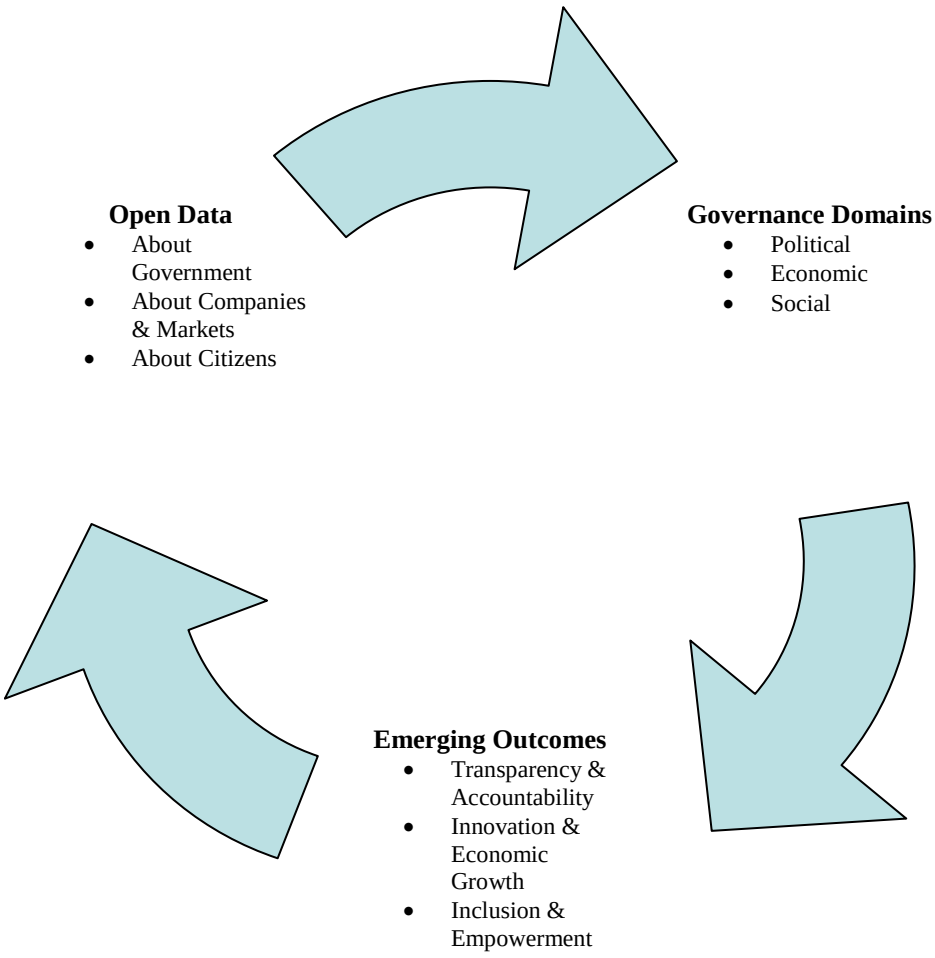


Figure 23: ODDC Open Data, Governance and Emerging Impacts Conceptual Framework [147]

The three key terms are as followings:

- **Open Data** – the importance to understand the subject, structure and status of datasets in creating an Open Data account. The emerging impacts of Open Data are affected by the role of Data, Data Licenses, Data Standards and Technical and non-technical intermediaries' activities. In addition, the open data applications developed provide a magnificent impact on citizens by enjoying the services created by the developers and transparency by the government [10].
- **Governance Domains** – the importance to explore how governance is operating in each of these domains and to identify different 'theories of change' or hypothesis of political, economic and social aspects. From the political aspects, it focuses on the state power exercise, control and shaping which valuable to explore the balance between institutions, citizens and media or journalism scholars in opening data. How-

ever, for economic it focuses the attention on the efficiencies for government better data use by the governance tool (distributing decision through markets) and on the market regulation or promotion. While governance of the economic domain (usually driven from the political domain) may be imposed upon all actors in a market, and the rules set through these processes affect market outcomes, governance carried out through the economic domain is distributed without central control or 'designed' outcomes. In social aspect governance, it focuses on the inclusion of marginalized groups, and on the capacity of individuals and communities at the grassroots to exercise influence over their own lives, without necessarily deploying either political or market power.

- **Emerging Outcomes** – the importance of change key theory that emerging impacts of Open Data may emerge from the followings [147]:
    - **Transparency and accountability**: provide greater accountability from greater transparency of key actors which led in making decisions and applying rules in the public interest;
    - **Innovation and economic development:** enable non-state innovators to improve public services or build new products and services with social and economic value. As a result, open data shift certain decision making from the state into the market;
    - **Inclusion and empowerment:** remove power imbalances that resulted from asymmetric information, and bring new stakeholders into policy debates, giving marginalized groups a greater say in the creation and application of rules and policy;

7.1.2 Proposed Research Conceptual Framework

To examine the importance of studying the emerging impacts of Open Data with broad range of research agendas, the research conceptual framework was developed as illustrated in Figure 24. In general, the development of this research involved several open data capabilities from the existing empirical evidence that have been highlighted in the previous chapters.

This proposed conceptual framework shows the stage of analysis from the strategy formulation that related to the three main components, Development Lifecycle, Seven core case components and open data emerging impacts contributions iteratively; and Star Ratings. Based on the existing literature, Star Rating components emerge from the Systematic Literature Review specifically Systematic Mapping Study which gets significant attention because of its characteristics. Thus, the rest of the components emerged from the units of analysis.

In this conceptual framework, the first element of Development Lifecycles presents on how the applications and systems were developed by using at least one of the processes that are important in any software project management. The services from the unplanned and failed to follow the process required for producing applications or systems might lead to the delayed or unsuccessful applications or systems to end-users.

The case components that were units of analysis identified from the Open Data research included Context, Supply of Data, Technical Platforms, Data Standards, Adoption and Awareness, Security and Privacy and Intermediaries as presented in section 5.6. From the existing empirical evidence, 30 characteristics of Open Data were identified and presented in the conceptual framework in Figure 22. These features summarized the highlights project purposes and research data usage, the rights of ownership and usage to the data used and generated by the project, how the research infrastructure was stored and made available, and how the data was protected in the aspect of security and privacy.

In producing an excellent Open Data applications or systems, this conceptual research model aims to achieve Open Data Quality measurement of Star Rating. In order to achieve the standards of the openness levels, the applications or systems are evaluated towards more 5-star rating of open data.

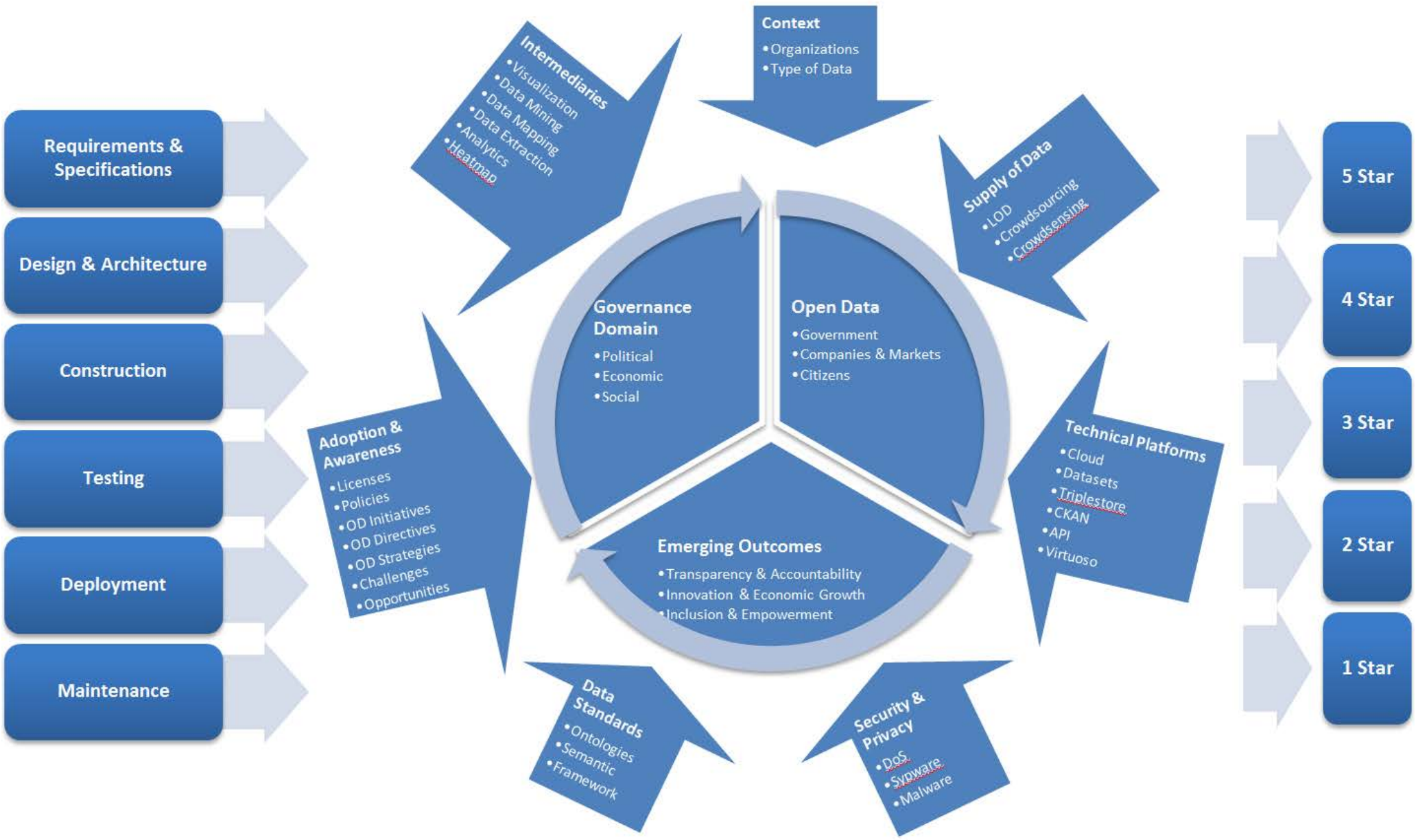


Figure 24: The Proposed Research Conceptual Framework

7.2 What Areas of Open Data Require Further Research?

The functions presented in Open Data conceptual framework in Figure 22 can be divided into two groups: Resources (intangible/tangible) and External Factors [149]. Intangible resources are referring to 'invisible asset' which means difficult to imitate and non-purchasable. In contrast to intangible, 'tangible' can be defined as something that is able to grasp, visible and touched. External factors are defined as any external sources that can provide or support for improvement. Taking this into account, only two of the case components from the existing literature, "Adoption and Awareness" and "Star Rating" are related to the external factors. Other studies are categorized as the intangible or tangible resources in the studies. As the result, more studies are required in the external factors research which needs to be explored and extended. As an example, based on the publication in 'Researching the Emerging Impacts of Open Data: ODDC Conceptual Framework' [147], there are no rigorous extensive studies on the impact of Open Data, and most work still at the level of ad-hoc and isolated case studies or anecdotes. There is the lack of evidence in the intended and unintended consequences for those who are documenting the experience of using Open Data. In addition, for the specific government setting, the context including a description and history of the issues in focus, details of key stakeholders, and analysis of how data plays a potential role in this setting are still an immature research area. More case study should be conducted in the Actions and Impacts of Open Data [150].

Open Data in the Intermediaries (22 articles), Security and Privacy (22 articles) and Supply of Data (15 articles) are the broader areas where the lack of research evidence. In general, more research is required in the aspect of Intermediaries to enable more sustainable interfaces and visualizations for the wider audience or users. As an example, with a strong user intermediaries or interface built upon a linked open dataset, users with different level of expertise can access and analyze information in a very comprehensive graphical view.

Security and privacy are the other issue that is required more research especially in producing robust and efficient applications or systems. In the information age nowadays,

the amount of data generated by larger companies and institutions daily is measured by terabytes and surely difficult to improve efficiency and deliver better services without appropriate security [151].

Data security also poses severe limitation for application developers in order to turn any idea into deployable application [152]. Trust is the other issue that is related to security and privacy [153]. Trust is important not just being the only driver for transparency and open data applications, but also important in many countries that have lately experienced scandal or corruption. Thus, the crowdsourcing strategies contributed as the value-added in improving the government information [153]. However, supply of data including crowdsourcing or crowdsensing that are an important aspect of the Internet of Things (IoT) nowadays are also the other unexplored issue in Open Data studies.

7.3 Limitations

There are several limitations of this study that should be taken into consideration while working with the reported findings:

- Only five digital libraries were used in the study.
- Only research or scientific publications were included in the study.
- The review did not include books about Open Data.
- The review only included publications that were accessible and available in full texts.

7.4 Threats to Validity

Threats to validity constitute factors that can influence the accuracy of research in a negative way. For this reason, it is important to identify and explain these threats to make the review results as reliable as possible. The study has several threats to validity, which are divided into four categories as follows.

7.4.1 Biases Related to Search, Researcher, and Publication

Search bias happened while the digital databases provide different options for searches conditions and search engine work differently. As a result, it might be some of the publications were not found. Therefore, the query for each digital database was conducted by adapting the search engine options as similarly as possible.

Since the review was conducted by an individual researcher, there is a higher potential for threats to validity in comparison with a review conducted by several researchers. In order to overcome this bias, the tasks were carried out twice or more to ensure the quality of the work. As an example, the author reviewing the abstracts was conducted twice to minimize possible mistakes, and all the reviewed information are recorded in a database (Microsoft Excel).

Publication bias is associated to the study that arises in the published article that unrepresentative of the population of completed studies [29]. There is a chance that the researcher decides the opposite conclusion about what the research presents while the particular research has different result compared to the other results of all the research has been done in the same field. In order to mitigate this problem, the pilot search was performed and followed by the actual search. Then, the unsuitable publication should be excluded. Defining the search strategy in the review protocol also helped to minimize the publication bias.

7.4.2 Biasness Related to Primary Studies

In order to minimize the threats to the identification of primary studies, the search strategy defined in the review protocol was used to cover as many studies as possible; the search string was applied in several well-known databases. In addition, the titles and the abstracts were read several times to include the right studies.

7.4.3 Data Extraction Process and Results

Another threat to validity is related to data extraction phase. The author defined the data extraction form and data extract process while designing the review protocol and followed that to record information about studies. This procedure helped to minimize the data extraction process bias.

7.5 Future Research Work

Even though the result of this Systematic Mapping Study presents the whole overview of Open Data, but some suggestions for further research work has been mentioned in Section 7.2 in this thesis. Furthermore, a survey should be done to identify the issues or challenges from the real experience of Open Data users such as organizations (Academia, Industries and Government) and developers. This survey can enhance the results in order to increase the success and reduce the failure of an Open Data application.

The whole study covered all of Open Data scope but the latest repository of Github (*https://github.com*) was not included in this research investigation due to the time limitation but is an excellent resource for the Open Data research. Github is a web-based Git public repository that is created purposely to be utilized by government and private developers to help agencies improve the management and release of open data by providing plug-and-play open source tools. These tools were aimed to accelerate the adoption of open data practices. Furthermore, from the Open Data project, anyone, from government agencies to private citizens can freely use and integrate with these tools.

8 Conclusions

This thesis systematically presents a mapping study in the context of Open Data by following the guidelines of Kitchenham and Charters and it fulfilled all the requirements of the standard systematic review. Based on the predefined search strategy, 621 publications published between years 2005 and 2014 were identified, but 243 were included in the review process.

The findings were described in quantitative assessments and qualitative analysis. All the publications were measured based on the Organization Affiliation, Countries, Year of Publications, Research Method, Star Rating and Units of Analysis. For each category, the theoretical perspective, themes, empirical support and discussion of identified factors that related to Open Data were identified. Units of Analysis were identified by the theme emerged from the selected studies categorized by Development Lifecycle, Linked Open Data, Type of Data, Technical Platforms, Organizations, Ontology and Semantic, Adoption and Awareness, Intermediaries, Security and Privacy and Supply of Data.

The broad overview of the research area, the quantity of evidence and the research data existed were established and described. In specific, Open Data topic was explored and discussed the trend of the researches. The importance of how existing empirical evidence highlights the applications purposes and research data usage was highlighted. The rights of ownership and usage to the data used, generated by the project was explained in the thesis. In addition, how the Open Data was stored and made available, and in the aspect of security and privacy how the data was protected were found from these researches.

From the theoretical view, this study provides the insight implications of Open Data principals' proliferation in the emerging era of the accessibility, reusability and sustainability of data transparency. The results of the mapping study could help organizations (such as academia, government and industries), researchers and software developers by providing them the information on the existing trend of Open Data, latest research development and the demand of future research. The capability and novelty of Open Data

innovation has been identified in order to produce useful Open Data applications. Meanwhile the quality of Open Data applications is presented based on the Tim Berners-Lee 5 star ratings criteria. The findings of this mapping study highlight the important of Open Data to improve services and to support open data sharing with more efficient and valuable.

In addition, the proposed conceptual framework of Open Data research is outlined based on the units of analysis identified from the mapping study. From the framework, it can provide guidelines to the managers, software developers and organizations to be adopted and expanded to strengthen and improved current Open Data applications.

REFERENCES

[1] Open Knowledge Foundation (OKFN). 2014. What is Open Data? [Online Document]. [Accessed 1st April 2014]. Available at http://okfn.org/opendata/

[2] Open Data Handbook. 2012. What is Open Data? [Online Document]. [Accessed 1st April 2014]. Available at http://opendatahandbook.org/en/what-is-open-data/index.html

[3] W3C(e-Gov). 2009. eGovernment at W3C: improving access to government through better use of the web. [Online Document]. [Accessed 1st April 2014]. Available at http://www.w3.org/2007/eGov/

[4] Molloy, JC. 2011. The Open Knowledge Foundation: Open Data Means Better Science. PLoS Biol 9(12): e1001195.

[5] Foulonneau, M., Martin, S. & Turki, S. 2014, February. How Open Data Are Turned into Services?. In Proceedings of 5th International Conference, IESS 2014. Geneva, Switzerland. Springer International Publishing. Pages 31-39.

[6] Obama, B. Jan. 21, 2009. Memorandum on Transparency and Open Government. [Online Document]. [Accessed 1st April 2014]. Available at http://www.whitehouse.gov/the-press-office/freedom-information-act

[7] Office of Management and Budget's (OMB). Dec. 8, 2009. Memorandum M-1 0-06, Open Government Directive. [Online Document]. [Accessed 1st April 2014]. Available at available at http://goo.gl/LcxbZE

[8] Directive 2013/37/EU of the European Parliament and of the Council. 26 June 2013. Amending Directive 2003/98/EC on the re-use of public sector information known as the "PSI Directive". [Online Document]. [Accessed 1st April 2014]. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp207_en.pdf

[9] Machado, A.L. & Parente de Oliveira, J.M. 2011, August. DIGO: An Open Data Architecture for e-Government. In 15th IEEE International Enterprise Distributed Object Computing Conference (EDOCW), 29 August – 2 September 2011, Sao Paulo, Brazil. IEEE. Pages 448 – 456.

[10] Oliver, M., Palacin, M., Valls, V. & Domingo, A. 2012, July. Sensor Information Fueling Open Data. In IEEE 36th Annual Computer Software and Applications Conference Workshops (COMPSACW), 16-20 July 2012, Barcelona, Spain. IEEE. Pages 116-121.

[11] Insights & Publications. January 2014. What Executives Should Know About Open Data. [Online Document]. [Accessed 1st April 2014]. Available at http://www.mckinsey.com/insights/high_tech_telecoms_internet/what_executives_should_know_about_open_data

[12] Song, S. H., Kim, T. D. & Won, M. 2013, January. A Study On The Open Platform Modeling For Linked Open Data Ecosystem In Public Sector. In 15th International Conference on Advanced Communication Technology (ICACT), 27-30 Jan. 2013. IEEE. Pages 730-734.

[13] Lindman, J., Kinnari, T. & Rossi, M. 2014, January. Industrial Open Data: Case Studies of Early Open Data Entrepreneurs. In 47th Hawaii International Conference on System Sciences (HICSS). IEEE. Pages 739-748.

[14] Schieferdecker, I. 2012, July. Trustworthiness of Open Source, Open Data, Open Systems and Open Standards. In IEEE 36th Annual Computer Software and Applications Conference Workshops (COMPSAC), 16-20 July 2012, Barcelona, Spain. IEEE. Page 82.

[15] O'Riain, S., Curry, E., & Harth, A. 2012, June. XBRL and Open Data For Global Financial Ecosystems: A Linked Data Approach. International Journal of Accounting Information Systems, Volume 13, Issue 2. Pages 141-162.

[16] Domingue, J., Fensel, D. & Hendler, J. A. (Eds.). 2011. Handbook of semantic web technologies, Vol. 1. Springer Science & Business Media.

[17] Berners-Lee, T. 2009. Linked Data - Design Issues. [Online Document]. [Accessed 1st April 2014]. Available at http://www.w3.org/DesignIssues/LinkedData.html

[18] Haslhofer, B & Isaac, A. 2011, September. Data.europeana.eu : The Europeana Linked Open Data Pilot. In International Conference on Dublin Core and Metadata Applications. Pages 94-104.

[19] Schwarte, A., Haase, P., Hose, K., Schenkel, R. & Schmidt, M. 2011. FedX: Optimization Techniques for Federated Query Processing on Linked Data. In The Semantic Web – ISWC 2011, Vol. 7031. Pages 601-616.

[20] Heath, T., Berners-Lee, T. & Bizer, C. 2009. Linked Data : The Story So Far. Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS), Vol. 5(3). Pages 1-22.

[21] Open Government Data Initiative (OGDI). 2013. Welcome to Open Government Data Initiative (OGDI) [online document]. [Accessed 1st April 2014]. Available at http://datadotgc.cloudapp.net

[22] Berners Lee, T. Design Issues. 2011. Socially Aware Cloud Storage. [online document]. [Accessed 1st April 2014]. Available at http://www.w3.org/DesignIssues/CloudStorage.html

[23] Budgen, D., Burn, A.J., Brereton, O.P., Kitchenham, B.A. & Pretorius, R. 2011. Empirical Evidence About The UML: A Systematic Literature Review. Software: Practice and Experience, Vol. 41(4). Pages 363–392.

[24] Five Star Open Data. 2012. [online document]. [Accessed 1st April 2014]. Available at http://5stardata.info/

[25] Lenhardt, W. C., Ahalt, S., Blanton, B., Christopherson, L. & Idaszak, R. 2014. Data management lifecycle and software lifecycle management in the context of conducting science. Journal Of Open Research Software, 2(1), E15.

[26] EPSIplatform. May 2010. The Five Stars of Open Data. [Online Document]. [Accessed 1st April 2014]. Available at http://www.epsiplatform.eu/sites/default/files/The%205%20stars%20of%20Open%20Data_MdV_PR2.pdf

[27] Janowicz, K., Hitzler, P., Adams, B. , Kolas, D. & Vardeman, C. 2014. Five Stars of Linked Data Vocabulary Use. Semantic Web Journal, Vol. 5. Pages 173-176.

[28] Open Data Foundation (OdaF). [Online Document]. [Accessed 1st April 2014]. Available at http://www.opendatafoundation.org/

[29] Kitchenham, B. & Charters, S. 2007. Guidelines For Performing Systematic Literature Reviews In Software Engineering. Keele University & Durham University Joint Report.

[30] Maglyas, A., Nikula, U. & Smolander, K. 2011, August. What Do We Know About Software Product Management? - A Systematic Mapping Study. In Fifth International Workshop on Software Product Management (IWSPM). IEEE. Page 26-35.

[31] Lindman, J., Rossi, M. & Tuunainen, V.K. 2013, January. Open Data Services: Research Agenda. In 46th Hawaii International Conference on System Sciences (HICSS), 7-10 January 2013. IEEE. Pages 1239-1246.

[32] Davis, A. & Hickey, A. 2009. A Quantititative Assessment of Requirements Engineering Publications – 1936-2006. Proceeding of 15th International Working Conference, REFSQ 2009, Amsterdam, The Netherlands. Pages 175-189.

[33] Davis, A., Dieste, O., Hickey, A., Juristo, N. & Moreno, A.M. 2006, September. Effectiveness of Requirements Elicitation Techniques: Empirical Results Derived from a Systematic Review. In 14th IEEE International Conference on Requirements Engineering, 11-15 Sept. 2006. IEEE. Page 179-188.

[34] OpenSym 2014, the 10th International Symposium on Open Collaboration. [Online Document]. [Accessed 26 February 2015]. Available at http://www.opensym.org/os2014/submission/research-track-calls/open-data/

[35] Gurstein, M.B. 2011. Open Data: Empowering the Empowered or Effective Data Use for Everyone?. First Monday, 16(2).

[36] Burke Johnson, R. 2007. Educational Research : Quantitative, Qualitative and Mixed Research. [Online Document]. [Accessed 26 February 2015]. Available at http://www.southalabama.edu/coe/bset/johnson/lectures/lec2.htm

[37] ProfEssays.com. 2009. Data Analysis Methodology [Online Document]. [Accessed 26 February 2015]. Available at http://www.professays.com/essay/essay-methodology-example/

[38] Global Open Data Index. 2014. Survey. [Online Document]. [Accessed 26 February 2015]. Available at http://global.census.okfn.org/

[39] Friberger, M.G. & Togelius, J. 2012, September. Generating interesting Monopoly boards from open data. In IEEE Conference on Computational Intelligence and Games (CIG), 11-14 Sept. 2012. IEEE. Pages 288-295.

[40] Office of Management and Budget's (OMB), Memorandum M-1 0-06. Dec. 8, 2009. Open Government Directive. [Online Document]. [Accessed 26 February 2015]. Available at http://goo.gl/LcxbZE

[41] Yin, R.K. 2013. Case study research: Design and methods. Fifth Edition. Sage Publications.

[42] The White House. 2013. Executive Order : Making Open and Machine Readable the new default of Government Information. [Online Document]. [Accessed 26 August 2014]. Available at http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

[43] ArsTechnica. 2013. Obama Orders Agencies to Make Open, machine-readable by default. [Online Document]. [Accessed 26 August 2014]. Available at http://arstechnica.com/tech-policy/2013/05/obama-orders-agencies-to-make-data-open-machine-readable-by-default/

[44] Federal Ministry of the Interior. 2012. Open Government Data Germany [Online Document]. [Accessed 26 August 2014]. Available at http://www.bmi.bund.de/SharedDocs/Downloads/DE/Themen/OED_Verwaltung/ModerneVerwaltung/opengovernment_kurzfassung_en.pdf?__blob=publicationFile

[45] State of the LOD Cloud. 2014. LOD Cloud Diagram [Online Document]. [Accessed 26 February 2015]. Available at http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/

[46] Sunlight Foundation. 2013. Open Data Policy Guidelines. [Online Document]. [Accessed 26 August 2014]. Available at http://sunlightfoundation.com/opendataguidelines/

[47] Lee, G. & Kwak, Y. H. 2011. An Open Government implementation model: moving to increased public engagement. IBM Center for the Business of Government, USA. Pages 26-29.

[48] Colpaert, P., Sarah, J., Peter, M., Mannens, E., & de Walle, R.V. 2013. The 5 stars of open data portals. In 7th International Conference On Methodologies, Technologies And Tools Enabling eGovernment (MeTTeG) Conference Proceedings. Pages 61–67.

[49] v3.co.uk. 2014. V3 Storage Summit: Open Data Storage and Access Are Key To Success, Says ODI. [Online Document]. [Accessed 26 August 2014]. Available at http://www.v3.co.uk/v3-uk/news/2324074/open-data-storage-and-access-are-key-to-success-says-odi

[50] Wikipedia. Open Data. [Online Document]. [Accessed 26 August 2014]. Available at http://en.wikipedia.org/wiki/Open_data

[51] Open Definition. Guide to Open Data Licensing. [Online Document]. [Accessed 26 August 2014]. Available at http://opendefinition.org/guide/data/

[52] Open Definition. [Online Document]. [Accessed 26 August 2014]. Available at http://opendefinition.org/od/

[53] Open Knowledge Blog. 2011. Building the Open Data Ecosystem. [Online Document]. [Accessed 26 August 2014]. Available at http://blog.okfn.org/2011/03/31/building-the-open-data-ecosystem/

[54] CKAN. The Open Source Data Portal Software. [Online Document]. [Accessed 26 July 2014]. Available at http://ckan.org/

[55] Virtuoso Open-source Wiki. 2009. [Online Document]. [Accessed 26 July 2014]. Available at http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/

[56] Davies, T. 2010. Open data, democracy and public sector reform. A look at open government data use from data.gov.uk. Master's thesis, Oxford Internet Institute, University of Oxford.

[57] Larkou, G., Metochi, J., Chatzimilioudis, G. & Zeinalipour-Yazti, D. 2013, June. CLODA: A crowdsourced linked open data architecture. In IEEE 14th International Conference on Mobile Data Management (MDM), 3-6 June 2013, Vol. 2. Pages 104-109.

[58] Morshed, A., Aryal, J. & Dutta, R. 2013, July. Environmental spatio-temporal ontology for the Linked Open Data cloud. In 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), 16-18 July 2013. IEEE. Pages 1907-1912.

[59] Steinberg, M. & Brehm, J. 2009, January. Towards utilizing Open Data for interactive knowledge transfer. In International Conference on Mobile, Hybrid, and On-line Learning, 2009. ELML '09, 1-7 Feb. 2009. IEEE. Pages 61-66.

[60] Lapi, E., Tcholtchev, N., Bassbouss, L., Marienfeld, F. & Schieferdecker, I. 2012, July. Identification and Utilization of Components for a Linked Open Data Platform. In IEEE 36th

Annual on Computer Software and Applications Conference Workshops (COMPSACW), 16-20 July 2012. IEEE. Pages 112-115.

[61] Theocharis, S.A. & Tsihrintzis, G.A. 2013, July. Open data for e-government the Greek case. In Fourth International Conference on Information, Intelligence, Systems and Applications (IISA), 10-12 July 2013. Pages 1-6.

[62] Lin, C. S., & Yang, H. C. 2014. Data Quality Assessment on Taiwan's Open Data Sites. In Multidisciplinary Social Networks Researc. Springer Berlin Heidelberg. Pages 325-333.

[63] Ren, G.J. & Glissmann, S. 2012, September. Identifying information assets for open Data: the role of business architecture and information quality. In IEEE 14th International Conference on Commerce and Enterprise Computing (CEC), 9-11 Sept. 2012. IEEE. Pages 94-100.

[65] Saunders, M. N., Saunders, M., Lewis, P., & Thornhill, A. 2014. Research methods for business students. Rotterdam: Erasmus University. Pages 106-153.

[66] Creswell, J.W. 2013. Research design: Qualitative, quantitative and mixed methods approaches. Sage publications. Pages 4-27.

[67] Zuiderwijk, A., Janssen, M. & Parnia, A. 2013, June. The complementarity of open data infrastructures: an analysis of functionalities. In Proceedings of the 14th Annual International Conference on Digital Government Research (dg.o '13), ACM, New York, NY, USA. Pages 166-171.

[68] Vafopoulos, M.N. & Meimaris, M. 2012, December. Weaving the Economic Linked Open Data. In 7th Semantic and Social Media Adaptation and Personalization (SMAP) Workshop, 3-4 Dec. 2012, Luxembourg City, Luxembourg, Vol. 3. Pages 92-97.

[69] Solar, M., Meijueiro, L., & Daniels, F. 2013. A guide to implement open data in public agencies. In Electronic Government. Springer Berlin Heidelberg. Pages 75–86.

[70] Meij, E., Bron, M., Hollink, L., Huurnink, B., & de, R. M. 2011. Mapping queries to the Linking Open Data cloud: A case study using DBpedia. Web Semantics: Science, Services and Agents on the World Wide Web, 9(4). Pages 418-433.

[71] Verma, N. 2013, August. Open data for inclusive governance. In Joint Proceedings of the Workshop on AI Problems and Approaches for Intelligent Environments and Workshop on Semantic Cities (AIIP '13). ACM, New York, NY, USA. Pages 5-5.

[72] Morshed, A., Dutta, R. & Aryal, J. 2013, April. Recommending environmental knowledge As linked open data cloud using semantic machine learning. In 2013 IEEE 29th International Conference on International Conference on Data Engineering Workshops (ICDEW). IEEE. Pages 27-28.

[73] Nixon, L.J. 2012, September. How open data can enhance interactive television. In 5th Joint IFIP Wireless and Mobile Networking Conference (WMNC), 19-21 Sept. 2012. Pages 118-121.

[74] Oomen, J., Tzouvaras, V. & Hyyppä, K. 2013, May. Linking and visualizing television heritage: the EUscreen virtual exhibitions and the linked open data pilot. In Proceedings of the 22nd international conference on World Wide Web companion (WWW '13 Companion). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland. Pages 481-484.

[75] Dataone. Data Life Cycle. [Online Document]. [Accessed 15 October 2015]. Available at http://openresearchsoftware.metajnl.com/articles/10.5334/jors.ax/

[76] Eberius, J., Thiele, M., Braunschweig, K. & Lehner, W. 2012. Drillbeyond: Enabling Business Analysts to Explore the Web of Open Data. Proceedings of the Vldb Endowment, 5(12). Pages 1978-1981.

[77] van Schalkwyk, F. 2013, October. Supply-side variants in the supply of open data in university governance. In Proceedings of the 7th international Conference on Theory and Practice of Electronic Governance, Seoul, Republic of Korea, October 22 – 25 2013. Pages 334-337.

[78] Nishikata, K, & Toyoda, T. 2011, December. BioLOD.org: ontology-based integration of biological linked open data. In Proceedings of the 4th International Workshop on Semantic Web Applications and Tools for the Life Sciences (SWAT4LS '11). ACM, New York, NY, USA. Pages 92-93.

[79] Yamamoto, Y., Yamaguchi, A. & Yonezawa, A. 2011, December. Building linked open data using approximate string matching methods and domain specific resources. In Proceedings of the 4th International Workshop on Semantic Web Applications and Tools for the Life Sciences (SWAT4LS '11). ACM, New York, NY, USA. Pages 121-122.

[80] Piedra, N., Tovar, E. Colomo-Palacios, R., Lopez-Vargas, Jorge & Janneth Chicaiza. 2014. Consuming and producing linked open data: the case of Opencourseware. Program: Electronic Library and Information Systems. Vol 48(1). Pages 16-40.

[81] Dimou, A., De Vocht, L., Van Grootel, G., Van Campe, L., Latour, J., Mannens, E., & Van de Walle, R. 2014. Visualizing the Information of a Linked Open Data Enabled Research Information System. Procedia Computer Science. 33. Pages 245-252.

[82] Waldman, M. 2013. Keeping it real: utilizing NYC open data in an introduction to database systems course. Journal of Computing Sciences in Colleges, 28(6). Pages 156-161.

[83] Bohm C., Naumann F., Abedjan Z., Fenz D., Grutze T., Hefenbrock D., Pohl M., & Sonnabend D. 2010, March. Profiling linked open data with ProLOD. In 2010 IEEE 26th International Conference on Data Engineering Workshops (ICDEW). IEEE. Pages 175-178.

[85] Eberius, J., Damme, P., Braunschweig, K., Thiele, M & Lehner, W. 2013, June. Publish-time data integration for open data platforms. In Proceedings of the 2nd International Workshop on Open Data (WOD '13). ACM, New York, NY, USA, Article 1.

[86] Kuhn, A., 2012, June. IDEs need become open data platforms (as need languages and vms). In Proceedings of the 2nd International Workshop on Developing Tools as Plug-ins (TOPI), 3-3 June 2012. IEEE Press. Page 31-36.

[87] Conradie, P., & Choenni, S. 2014. On the barriers for local government releasing open Data. Government Information Quarterly, 31, S10-S17.

[88] Lai, Y.A., Ou, Y.Z., Su, J., Tsai, S.H., Yu, C.W. & Cheng, D. 2012, December. Virtual disaster management information repository and applications based on linked open data. In 5th IEEE International Conference on Service-Oriented Computing and Applications (SOCA), 17-19 Dec. 2012. IEEE. Pages 1-5.

[89] Dalamagas, T., Bikakis, N., Papastefanatos, G., Stavrakas, Y. & Hatzigeorgiou, A.G. 2012, May. Publishing life science data as linked open data: The case study of miRBase. In Proceedings of the First International Workshop on Open Data (WOD '12). ACM, New York, NY, USA. Pages 70-77.

[90] Gupta, A., Viswanathan, K.K., Joshi, A., Finin, T. & Kumaraguru, P. 2011, March. Integrating linked open data with unstructured text for intelligence gathering tasks. In Proceedings of the 8th International Workshop on Information Integration on the Web: in conjunction with WWW 2011 (IIWeb '11). ACM, New York, NY, USA, Article 3.

[91] Macharia, P., Muluve, E., Lizcano, J., Cleland, C., Cherutich, P. & Kurth, A. 2013, May. Open Data Kit, A solution implementing a mobile health information system to enhance data management in public health. In IST-Africa Conference and Exhibition (IST-Africa), 29-31 May 2013. IEEE. Pages 1-6.

[92] Latif, A., Afzal, M.T., Saeed, A.U., Hoefler, P. & Tochtermann, K. 2009, July. CAF-SIAL: Concept aggregation framework for structuring informational aspects of linked open data. In First International Conference on Networked Digital Technologies, 28-31 July 2009. IEEE. Pages 100-105.

[93] Raza, M.A., Afzal, M.T. & Khanum, A. 2011. Discovering, structuring and visualizing organizations from Linked Open Data. In 6th International Conference on Computer Sciences and Convergence Information Technology (ICCIT), Nov. 29 – Dec. 1 2011. IEEE. Pages 738-744.

[94] Zhao, L. & Ichise, R. 2012. Graph-based ontology analysis in the linked open data. In Proceedings of the 8th International Conference on Semantic Systems (I-SEMANTICS '12). ACM, New York, NY, USA. Pages 56-63.

[95] Dastgheib, S., Mesbah, A. & Kochut, K. 2013. mOntage: building domain ontologies from linked open data. In IEEE 7th International Conference on Semantic Computing (ICSC), 16-18 Sept. 2013. Pages 70-77.

[96] MacKellar, B., Schweikert, C. & Soon Ae Chun. 2013, July. Patient-oriented clinical trials search through semantic integration of Linked Open Data. In 12th IEEE International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC) 16-18 July 2013. IEEE. Pages 218-225.

[97] Yoosooka, B. & Wuwongse, V. 2011. Linked open data for learning object discovery: adaptive e-learning systems. In 3rd International Conference on Intelligent Networking and Collaborative Systems (INCoS), Nov. 30 2011-Dec. 2 2011. Page 60-67.

[98] Hagedorn, S. & Sattler, K.U. 2013, March. Discovery querying in linked open data. In Proceedings of the Joint EDBT/ICDT 2013 Workshops (EDBT '13). ACM, New York, NY, USA. Pages 38-44.

[99] Kabutoya, Y., Sumi, R., Iwata, T., Uchiyama, T. & Uchiyama, T. 2012, December. A Topic Model for Recommending Movies via Linked Open Data. In IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), 4-7 Dec. 2012. Pages 625-630.

[100] Feng, Y., Veeramani, A. & Kanagasabai, R. 2012, December. Enabling On-Demand Mashups of Open Data with Semantic Services. In IEEE 18th International Conference on Parallel and Distributed Systems (ICPADS), 17-19 Dec. 2012. Pages 755,759.

[101] Hoxha, J., Brahaj, A. & Vrandečić, D. 2011. open.data.al: increasing the utilization of government data in Albania. In Proceedings of the 7th International Conference on Semantic Systems (I-Semantics '11), Chiara Ghidini, Axel-Cyrille Ngonga Ngomo, Stefanie Lindstaedt, and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA. Pages 237-240.

[102] Noia, T.D., Mirizzi R., Ostuni, V.C., Romito, D. & Zanker, M. 2012. Linked open data to support content-based recommender systems. In Proceedings of the 8th International Conference on Semantic Systems (I-SEMANTICS '12), Harald Sack and Tassilo Pellegrini (Eds.). ACM, New York, NY, USA. Pages 1-8.

[103] Ostuni, V.C., Noia, T.D., Sciascio, E.D. & Mirizzi, R. 2013. Top-N recommendations from implicit feedback leveraging linked open data. In Proceedings of the 7th ACM conference on Recommender systems (RecSys '13). ACM, New York, NY, USA. Pages 85-92.

[104] Vosough, P. 2013. Implementing An Open Data System And Showing Its Benefits, MSc. Thesis, Department of Software Engineering and Information Management, School of Industrial Engineering and Management, Lappeenranta University of Technology, Finland.

[105] Puron-Cid, G., Gil-Garcia, J.R. & Luna-Reyes, L.F. 2012. IT-enabled policy analysis: new technologies, sophisticated analysis and open data for better government decisions.

In Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o '12). ACM, New York, NY, USA. Pages 97-106.

[I06] Zuiderwijk, A., Janssen, M., van den Braak, S. & Charalabidis, Y. 2012. Linking open data: challenges and solutions. In Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o '12). ACM, New York, NY, USA. Pages 304-305.

[107] Zuiderwijk, A. & Janssen, M. 2012. A comparison of open data policies and their implementation in two Dutch ministries. In Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o '12). ACM, New York, NY, USA. Pages 84-89.

[108] Immonen, A., Palviainen, M. & Ovaska, E., 2014. Requirements of an Open Data Based Business Ecosystem. In Access, IEEE , Vol.2. Pages 88-103.

[109] Yang, T.M., Lo, J., Wang, H.J. & Shiang, J. 2013, October. Open data development and value-added government information: case studies of Taiwan e-Government. In Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance (ICEGOV '13). ACM, New York, NY, USA. Pages 238-241.

[110] Clarke, C. 2009. A resource list management tool for undergraduate students based on linked open data principles. In The Semantic Web: Research and Applications. Springer Berlin Heidelberg. Pages 697-707.

[111] Song, S.H., Kim, T.D. & Won, M. 2013, January. A study on the open platform modeling for linked open data ecosystem in public sector. In 15th International Conference on Advanced Communication Technology (ICACT), 27-30 Jan. 2013. Pages 730, 734.

[112] Silva, T., Wuwongse, V. & Sharma, H.N. 2013. Disaster mitigation and preparedness using linked open data. In Journal of Ambient Intelligence and Humanized Computing 4 (5). Pages 591-602.

[113] Chen, B., Ding, Y. Wang, H., Wild, D.J., Dong, X., Sun, Y., Zhu, Q. & Sankaranarayanan, M. 2010, August. Chem2Bio2RDF: A linked open data portal for systems chemical biology. In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Aug. 31 2010-Sept. 3 2010, Vol.1. IEEE. Pages 232-239.

[114] Fanizzi, N., dAmato, C. & Esposito, F. 2012, September. Mining Linked Open Data through Semi-supervised Learning Methods Based on Self-Training. In IEEE Sixth International Conference on Semantic Computing (ICSC), 19-21 Sept. 2012. IEEE. Pages 277-284.

[115] Latif, A., Afzal, M.T. & Tochtermann, K. 2010, October. Constructing experts profiles from linked open data. In 6th International Conference on Emerging Technologies (ICET), 18-19 Oct. 2010. Pages 33-38.

[116] Le, T., Vo, B. & Duong, T.H. 2012, September. Personalized Facets for Semantic Search Using Linked Open Data with Social Networks. In Third International Conference on Innovations in Bio-Inspired Computing and Applications (IBICA), 26-28 Sept. 2012. Pages 312-317.

[117] Cannataro, M., Guzzi, P.H. & Veltri, P. 2013, December. Using open data in health care and tourism. In IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 18-21 Dec. 2013. IEEE. Pages 30-33.

[118] Niu, X., Sun, X., Wang, H., Rong, S., Qi, G. & Yu, Y. 2011. Zhishi.me: weaving chinese linking open data. In Proceedings of the 10th international conference on The Semantic Web (ISWC'11). Springer Berlin, Heidelberg. Pages 205-220.

[119] Oomen, J., Tzouvaras, V., Verbruggen, E. & Hyyppä, K. 2013, October. Television heritage linked and visualized: The EUscreen virtual exhibitions and the Linked Open Data pilot. In Digital Heritage International Congress (DigitalHeritage), Oct. 28 2013-Nov. 1 2013, Vol.2. IEEE. Pages 393-396.

[I20] Buchanan, F., Capanni, N. & Gonzalez-Velez, H. 2011, June. Fine artists of the world unite: Bridging heterogeneous distributed open data sources of fine art. In International Conference on Information Society (i-Society), 27-29 June 2011. IEEE. Pages 224-229.

[121] Zablith, F., Fernandez, M., & Rowe, M. 2012, January. The OU linked open data: production and consumption. In Proceedings of the Workshop on eLearning Approaches for the Linked Data Age, 8th Extended Semantic Web Conference (ESWC2011), 29 May 2011, Heraklion, Greece. Springer Berlin Heidelberg.

[122] Kuznetsov, K.A. 2013. Scientific data integration system in the linked open data space. Programming and Computer Software 39(1). Pages 43-48.

[123] Swezey, R.M.E., Sano, H., Hirata, N., Shiramatsu, S., Ozono, T. & Shintani, T. 2012. An e-participation support system for regional communities based on linked open data, classification and clustering. In IEEE 11th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), 22-24 Aug. 2012. IEEE. Pages 211-218.

[124] Kashyap, A. & Galarza, G. 2011. Building sustainable healthcare knowledge systems by harnessing efficiencies from biomedical Linked Open Data. India Conference (INDICON), 2011 Annual IEEE, 16-18 Dec. 2011. Pages 1-5.

[125] Callahan, A., Cruz-Toledo, J. & Dumontier, M. 2013. Ontology-based querying with Bio2RDF's linked open data. Journal of Biomedical Semantics, April 2013, 4 (Suppl 1), S1.

[126] Sorrentino, S., Bergamaschi, S., Fusari, E. & Beneventano, D. 2013. Semantic annotation and publication of linked open data. In Computational Science and Its Applications (ICCSA 2013), Vol. 7975. Springer Berlin Heidelberg. Pages 462-474.

[127] Balena, P., Bonifazi, A. & Mangialardi, G. 2013. Smart Communities Meet Urban Management: Harnessing the Potential of Open Data and Public/private Partnerships Through Innovative E-Governance Applications. In Computational Science and Its Applications (ICCSA 2013), Vol. 7975. Springer Berlin Heidelberg. Pages 528-540.

[128] Shiramatsu, S., Swezey, R.M., Sano, H., Hirata, N., Ozono, T. & Shintani, T. 2012. Structuring japanese regional information gathered from the web as linked open data for use in concern assessment. In Electronic Participation, Vol. 7444. Springer Berlin Heidelberg. Pages 73-84.

[129] Yamamoto, Y., Yamaguchi, A., & Yonezawa, A. 2013. Building Linked Open Data Towards Integration of Biomedical Scientific Literature with DBpedia. Journal of Biomedical Semantics 4 (8). PMC. Web. 14 Oct. 2015.

[130] Karam, R. & Melchiori, M. 2013. A Crowdsourcing-Based Framework for Improving Geo-spatial Open Data. In IEEE International Conference on Systems, Man, and Cybernetics (SMC), 13-16 Oct. 2013. Pages 468,473.

[131] Pinheiro, V., Furtado, V., Pequeno, T. & Ferreira, C. 2012. Towards a common sense base in portuguese for the linked open data cloud. In Computational Processing of the Portuguese Language, Vol. 7243. Springer Berlin Heidelberg. Pages 128-138.

[132] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. 2007. Dbpedia: A nucleus for a web of open data. The Semantic Web, Vol. 4825. Springer Berlin Heidelberg. Pages 722-735.

[133] O'Boyle, N. M., Guha, R., Willighagen, E. L., Adams, S. E., Alvarsson, J., Bradley, J. C., & Murray-Rust, P. 2011. Open Data, Open Source and Open Standards in Chemistry: The Blue Obelisk five years on. Journal of Cheminformatics 3 (37). PMC. Web. 14 Oct. 2015.

[134] Nečaský, M., Knap, T., Klímek, J., Holubová, I., & Vidová-Hladká, B. 2013, January. Linked open data for legislative domain-ontology and experimental data. In Business Information Systems Workshops. Springer Berlin Heidelberg. Pages 172-183.

[135] Kalampokis, E., Hausenblas, M., & Tarabanis, K. 2011. Combining social and government open data for participatory decision-making. In Electronic participation, Vol. 6847. Springer Berlin Heidelberg. Pages 36–47.

[136] Izzi, F., La Scaleia, G., Buono, D. D., Scorza, F. & Las Casas, G. 2013. Enhancing the spatial dimensions of open data: Geocoding Open PA information using geo platform fusion to support planning process (ICCSA 2013). Springer Berlin Heidelberg. Pages 622–629.

[137] Kobayashi, N. & Toyoda, T. 2011. BioSPARQL: ontology-based smart building of SPARQL queries for biological linked open data. In Proceedings of the 4th International Workshop on Semantic Web Applications and Tools for the Life Sciences (SWAT4LS '11). ACM, New York, NY, USA. Pages 47-49.

[138] Nonthakarn, C & Wuwongse, V. 2012. Linked OpenScholar: A Researcher Network Using Linked Open Data. The Outreach of Digital Libraries: A Globalized Resource Network, Vol. 7634, Springer Berlin Heidelberg. Pages 325–328.

[139] Masuya, H., Takatsuki, T., Makita, Y., Yoshida, Y., Mochizuki, Y., Kobayashi, N., Yoshiki, A., Nakamura, Y, Toyoda, T. & Obata, Y. 2013. Development of Linked Open Data for Bioresources. Semantic Technology, Vol. 7774. Springer Berlin Heidelberg. Pages 350–355.

[140] Colpaert, P. 2014. Route Planning Using Linked Open Data. In The Semantic Web: Trends and Challenges, Springer International Publishing. Pages 827–833.

[142] Gombos, G., & Kiss, A. 2014. SPARQL Processing over the Linked Open Data with Automatic Endpoint Detection. In Advanced Approaches to Intelligent Information and Database Systems, Springer International Publishing. Pages 183-192.

[143] Damjanovic, V., Glachs, D., Tcholtchev, N., Ras, E., & Tobias, E. 2014. EAGLE–Open Data and Linked Data Architecture of an Enhanced Government Learning Platform. In Open Learning and Teaching in Educational Communities, Springer International Publishing. Pages 558-559.

[144] Hienert, D., Wegener, D., & Schomisch, S. 2013. Making Sense of Open Data Statistics with Information from Wikipedia. In Availability, Reliability, and Security in Information Systems and HCI. Springer Berlin Heidelberg. Pages 329-344.

[145] Groen, M., Meys, W., & Veenstra, M. 2013, September. Creating smart information services for tourists by means of dynamic open data. In Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication. ACM. Pages 1329-1330.

[146] Chan, C. M. 2013, January. From open data to open innovation strategies: Creating E-services using open government data. In 46th Hawaii International Conference on System Sciences (HICSS), IEEE. Pages 1890-1899.

[147] Open Data Research Network. 2014. Researching The Emerging Impacts Of Open Data - ODDC Conceptual Framework. [Online Document]. [Accessed 1st August 2014]. Available at http://www.opendataresearch.org/content/2013/480/publication-%EF%BF%BCresearching-emerging-impacts-open-data-oddc-conceptual-framework

[148] Helbig, N., Cresswell, A. M., Burke, G. B., & Luna-Reyes, L. 2012. The dynamics of opening government data. Center for Technology in Government. [Online Document]. [Accessed 1st August 2014]. Available at http://www. ctg. albany. edu/publications/reports/opendata.

[149] Grant, R.M. 2002. The resource-based theory of competitive advantage. California Management Review. Pages 114–135.

[150] The Department of Public Expenditure and Reform.2014. Open Data Ireland: Evaluation Framework. [Online Document]. [Accessed 1st February 2015]. Available at http://www.per.gov.ie/en/open-data/

[151] Milić, P., Veljković, N. & Stoimenov, L. 2012, September. Framework for open data mining in e-government. In Proceedings of the Fifth Balkan Conference in Informatics (BCI '12). ACM, New York, NY, USA. Pages 255-258.

[152] Masip-Bruin, X., Ren, G.J. , Serral-Gracia, R. & Yannuzzi, M. 2013, July. Unlocking the Value of Open Data with a Process-Based Information Platform. In 15th Conference on Business Informatics (CBI). IEEE, 15-18 July 2013. Pages 331-337.

[153] O'Hara, K. 2012, June. Transparency, open data and trust in government: shaping the infosphere. In Proceedings of the 4th Annual ACM Web Science Conference *(WebSci '12)*. ACM, New York, NY, USA. Pages 223-232.

[154] Eberius, J., Braunschweig, K., Thiele, M., & Lehner, W. 2012, May. Identifying and weighting integration hypotheses on Open Data Platforms. In Proceedings of the First International Workshop on Open Data (WOD '12). ACM, New York, NY, USA. Pages 22-29.

[155] Estermann. B. 2013, August. Are memory institutions ready for open data and crowdsourcing?: results of a pilot survey from Switzerland. In Proceedings of the 9th International Symposium on Open Collaboration (WikiSym '13). ACM, New York, NY, USA, Article 29.

[156] GovWild.2012. Government Web Integration for Linked Data. [Online Document]. [Accessed 1st February 2015]. Available at http://govwild.hpi-web.de

[157] Böhm, Ch., Naumann, F., Freitag, M., George, S., Höfler, N., Köppelmann, M., Lehmann, C., Mascher, A. & Schmidt, T. 2010, September. Linking open government data: what journalists wish they had known. In Proceedings of the 6th International Conference on Semantic Systems (I-SEMANTICS '10). ACM, New York, USA.

[158] Plu, J. & Scharffe, F. 2012, May. Publishing and linking transport data on the Web. In Proceedings of the First International Workshop on Open Data (WOD '12). ACM, New York, USA. Pages 62-69.

[159] Bizer, Ch., Lehmann, J., Kobilarov, G., Auer, S., Becker, Ch., Cyganiak, R. & Hellmann, S. 2009. DBpedia - A crystallization point for the Web of Data. Web Semantics: science, services and agents on the world wide web, Vol. 7, Issue. 3. Pages 154–165.

[160] Linked Data on The Web (LDOW2008). 2008. Linked Data on the Web. [Online Document]. [Accessed 1st February 2015]. Available at http://events.linkeddata.org/ldow2008/slides/LDOW2008-Workshop-Intro.pdf

[161] Readwrite. 2010. It's All Semantics: Open Data, Linked Data & The Semantic Web. [Online Document]. [Accessed 1st July 2015]. Available at http://readwrite.com/2010/03/31/open_data_linked_data_semantic_web

[162] Linked Data on The Web (LDOW2008). 2008. Open Data Commons, A License for Open Data. [Online Document]. [Accessed 1st April 2014]. Available at http://events.linkeddata.org/ldow2008/papers/08-miller-styles-open-data-commons.pdf

[163] JISC cetis. 2012. The Semantic Web, Linked and Open Data. [Online Document]. [Accessed 1st April 2014]. Available at http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf

[164] Martin, S., Foulonneau, M. & Turki, S. 2013. 1-5 Stars: Metadata on the Openness Level of Open Data Sets in Europe. In Metadata and Semantics Research. Springer International Publishing. Pages 234-245.

[165] Hickey, A., Davis, A. & Kaiser, D. 2013. Requirements elicitation techniques: Analyzing the gap between technology availability and technology use. Comparative Technology Transfer and Society 1, No. 3. Pages 279-302.

[166] Devex. 2000. Global Development's 'Data Intermediaries' and their critical role post-2015. [Online Document]. [Accessed 31st July 2015]. Available at https://www.devex.com/news/global-development-s-data-intermediaries-and-their-critical-role-post-2015-86511

[167] Downie, M., Kaiser, P., Enloe, D., Fox, P., Hendler, J., Ameres, E. & Goebel, J. 2011, October. Evolving a rapid prototyping environment for visually and analytically exploring large-scale Linked Open Data. In IEEE Symposium on Large Data Analysis and Visualization (LDAV). Pages 139-140.

[168] Milić, P., Veljković, N. & Stoimenov, L. 2012, September. Framework for open data mining in e-government. In Proceedings of the Fifth Balkan Conference in Informatics (BCI '12). New York City. Pages 255-258.

[170] Yang, T.M. 2012, October. From inter-agency information sharing to open data: a case study of Taiwan E-Government. In Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance (ICEGOV '12). ACM, New York, NY, USA. Pages 477-478.

[171] Barki, H. & Pinsonneault, A. 2005. A model of organizational integration, implementation effort, and performance. Organization Science, 16 (2). Pages 165-179.

[172] Sandoval-Almazan, R., Ramon Gil-Garcia,J., Luna-Reyes,L.F., Luna, D.E. & Rojas-Romero, Y. 2012, October. Open government 2.0: Citizen empowerment through open data, web and mobile apps. In Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance (ICEGOV '12). ACM, New York, NY, USA. Pages 30-33.

[173] European Commission, 2010. The European e-Government Action Plan 2011-2015. Harnessing ICT to promote smart, sustainable and innovative Government. Brussels, COM (2010) 743.

[174] Puron-Cid, G., Ramon Gil-Garcia, J. & Luna-Reyes, L.F. 2012, June. IT-Enabled Policy Analysis: New Technologies, Sophisticated Analysis And Open Data For Better Government Decisions. In Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o '12). ACM, New York, NY, USA. Pages 97-106.

[175] Agrawal, S., Deshmukh, J., Srinivasa, S., Jog, C., Bhavaani, K. S. & Dhek, R. 2013, October. A survey of Indian open data. In Proceedings of the 5th IBM Collaborative Academia Research Exchange Workshop (I-CARE '13). ACM, New York, NY, USA, Article 2.

[176] Becker, C., Bizer, C. 2008. DBpedia mobile - a location-aware semantic web client. In Proceedings of the Semantic Web Challenge (ISWC 2008). Pages 13-16.

[177] Clarke, C. 2009. A resource list management tool for undergraduate students based on linked open data principles. In The Semantic Web: Research and Applications. Springer Berlin Heidelberg. Pages 697-707.

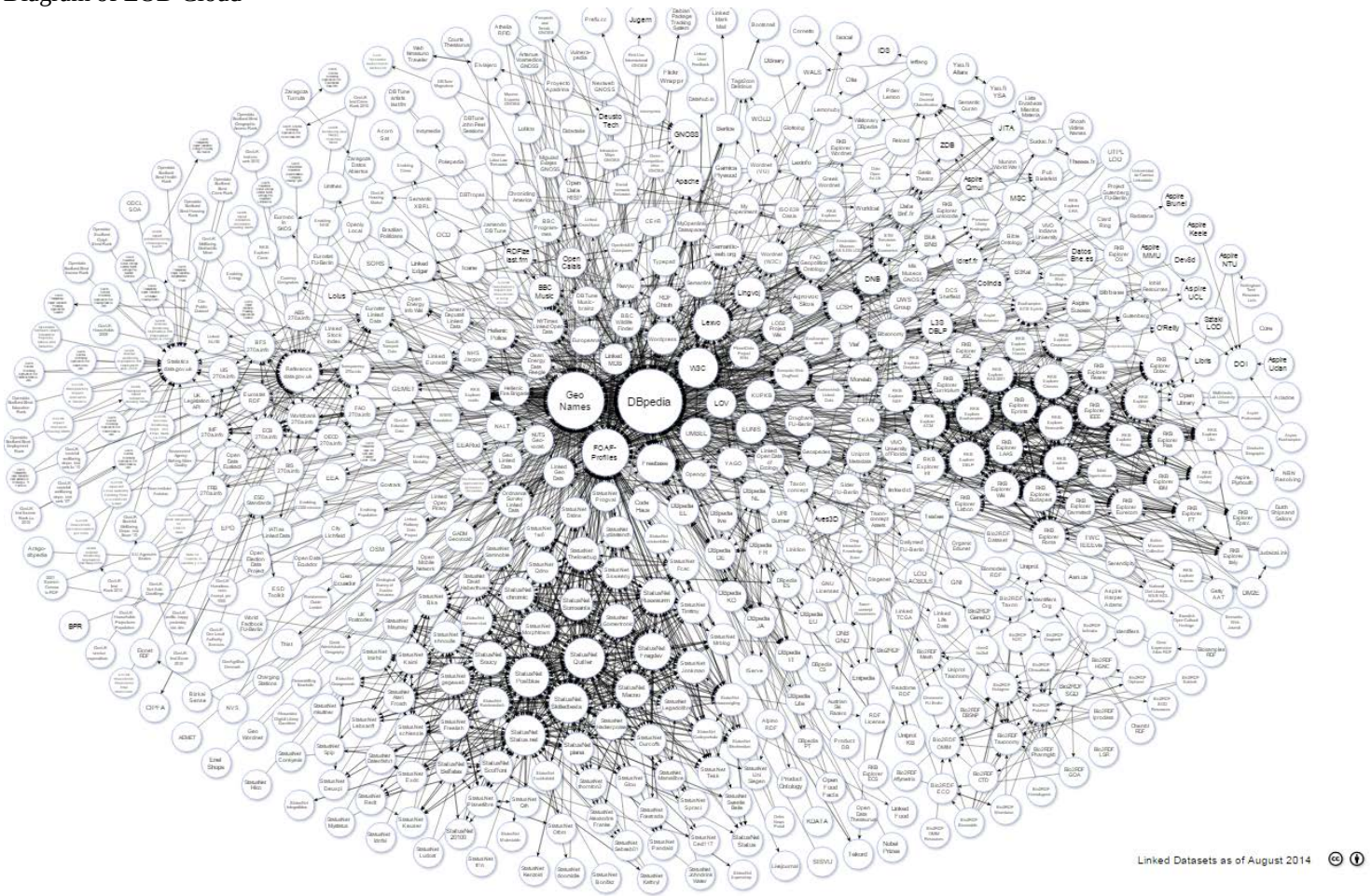
[178] Schmachtenberg, M., Bizer, C. & Paulheim, H. 2014. Adoption of the linked data best practices in different topical domains. In The Semantic Web–ISWC 2014. Springer International Publishing. Pages 245–260.

[179] Scopus. About Scopus. 2014. [Online Document]. [Accessed 1st July 2015]. Available at http://www.scopus.com

[180] Web of Sciences. 2014. [Online Document]. [Accessed 1st July 2015]. Available at https://woninfo.com

[181] Google Scholar. 2014. [Online Document]. [Accessed 1st July 2015]. Available at https://scholar.google.fi/

APPENDIX I:
Diagram of LOD Cloud



APPENDIX II:

Table of Search Keywords “open data” Result by Digital Databases, IEEEXplore, ACM, Science Direct, Ebsco and SpringerLink Ejournals (Search Categorization : Title OR Abstract, 2012-2014)

Search Keywords	IEEEXplore	ACM	Science Direct	EbscoHost	SpringerLink Ejournals	Total	Percentage(%)
“open data” AND system	30	144	58	102	76	410	9.86
“open data” AND science	19	117	52	106	82	376	9.04
“open data” AND user	31	131	48	70	73	353	8.49
“open data” AND link	63	96	35	56	74	324	7.79
“open data” AND computing	54	92	31	64	43	284	6.83
“open data” AND network	31	84	36	69	59	279	6.71
“open data” AND government	28	66	35	73	42	244	5.87
“open data” AND platform	17	73	38	71	40	239	5.75
“open data” AND environment	14	71	44	65	35	229	5.50
“open data” AND cloud	23	67	29	58	35	212	5.10
“open data” AND statistic	1	62	28	69	7	167	4.01
“open data” AND health	7	37	27	71	22	164	3.94
“open data” AND device	11	30	22	56	9	128	3.08
“open data” AND sensor	9	22	15	56	14	116	2.79
“open data” AND transport	2	20	10	54	14	100	2.40
“open data” AND culture	0	19	11	56	14	100	2.40
“open data” AND weather	1	16	14	57	7	95	2.28
“open data” AND game	4	16	6	56	6	88	2.12
“open data” AND finance	2	11	9	54	6	82	1.97
“open data” AND sustain	1	8	13	53	5	80	1.92
“open data” AND geodata	0	5	1	53	5	64	1.54

"open data" AND DaaS	1	0	0	18	7	26	0.63
Total	349	1187	562	1387	675	4160	100.00