User Interface design for open data platforms

Master’s Thesis

Armin Parnia
User Interface design for open data platforms

Thesis

Submitted in partial fulfillment of the

Requirements for the degree of

MASTER OF SCIENCE

In

COMPUTER SCIENCE

By

Armin Parnia

Born in Tehran, Iran

TU Delft
Faculty EEMCS, Delft University of Technology
Delft, The Netherlands
User interface design for open data platforms

Author: Armin Parnia
Student ID: 4138260
Email: parniaa@gmail.com

Abstract
Open data can become available by open data infrastructures. In recent years, different governmental and private organizations have published their data through open data platforms to accelerate knowledge and information sharing. They are open data platforms, which are not particularly developed by any public or private organization. These universal open data platforms facilitate combining data various data types from different data resources. The aim of this study is to create integrated user interfaces for different groups of users, which stimulate services for users Open Data (OD), including the processes of importing, curating, processing and visualizing OD. The process of Open Data consists of following steps; Data production, publish data, Look for data, use data, feedback. The process focuses on providing data that are not accessible, to be published. For this research, various research methods were used: literature research has been conducted. Next we followed requirements engineering approach to elicit and analyze requirements by using scenario and use case techniques. We have created prototypes by visualizing use case flows in BPMN diagram. Finally, we have conducted an evaluation by user testing in the scenario-based environment, and walk-through technique.

In this research, we focus on technological barrier of open data. Many technological impediments restrict users to have easy access to open data. The most significant impediments are Data deposit, Data access and Data use, impediments. Also, different requirements are identified for open data platforms. These requirements are categorized in five subcategories. Namely, Standardization, API, Materialization, Policies and integration. Data publishers usually only make data available and often do not consider the user perspective. One aspect of the user perspective that has received very little attention is the user interface of open data platforms. We have performed requirements analysis by using scenario and use case techniques. We have designed 6 scenarios that cover all steps of open data process. Next, we have created BPMN diagram to represent the flow of use cases. Finally, we have created UI mockups for Open Data platform.

This research contributes in adding more benefits from open data by designing the user interface to increase the user’s contribution, as well as contributing into studies in the field of open data. This research involved a usability test in scenario-based environment as well as inspection. A total of 16 usability tests was conducted. The usability test was intended to verify the proposed UI. Participants were selected from different age ranges, countries, and study fields. Therefore, we conclude that users’ success rates (performing tasks correctly) with the all the three scenarios were considerably high. However, a disadvantage regarding the dataset extension is reported. Lastly, most of the tasks were performed correctly and according to users most of the tasks were easy to perform.

Chair: Prof.Dr.ir. G.J.P.M. Houben
Head of Web Information Systems (WIS) at the Software and Computer Technology (ST) department at EEMCs of Delft University of Technology (TU Delft)

First supervisor: Prof. Dr. ir. M. F. W. H. A. Janssen
Professor, Information and Communication Technology department at ICT Group of the Department of TPM of the Delft University of Technology

Second supervisor: Dr. ir. Jan Hidders
Web Information Systems (WIS) at the Software and Computer Technology (ST) department at EEMCs of Delft University of Technology (TU Delft)

Third supervisor: Anneke Zuiderwijk-van Eijk, M. Sc
Information and Communication Technology at ICT Group of the Department of TPM of the Delft University of Technology
To my beloved parents Roya and Rahman,

for their unconditional sacrifice, love, and support.

To my dear sister Eileen, as a good friend in my life.
Primarily, I would like to express my gratitude to my supervisor Prof. Marijn Janssen. The idea of this thesis topic has primarily been brought up by Prof. Marijn Janssen. From the start of this project, I have learnt a lot from him. Thanks to Marijn in both personal and academic aspects, he provided me continuous support in many ways during my project while guiding me to find out the way on my own.

I would also like to thank Anneke Zuiderwijk for her patience, assistance and guidance from an early stage of this project, and providing me feedback in all steps of my thesis project.

My sincere thanks also goes to Prof. Geert-Jan Houben and Dr. Jan Hidders for their support and help during my studies and in the process of applying for PhD. Position.

Also, I would like to thank the participants in my questionnaire, who have willingly shared their precious time during the process of evaluation.

Moreover, I would also like to thank my dear friend Ir. Nami Naserazad for his kind academic support in many moments during my studies and my project. Special thanks to Ir. Samaneh Tajalizadehkhooob for her kind support during my studies. Also, Special thanks goes to Ir. Leila Dehghani, Ir. Delaram Azari, and Shahab Shariat Torbaghan who helped me by proof reading this thesis.

Last, but not least, During the period of my masters studies, many friends are helpful to color my life here in Delft. I thank all of them who supported me during this journey. Pejman, Farshad, Siavash, Siamak, Armin, Mahsa, Dena and some of whom help sustained me with hope and passion during my studies in the Netherlands.
Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>II</td>
</tr>
<tr>
<td>FOREWORD</td>
<td>IV</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>VIII</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>X</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.2 PROBLEM STATEMENT</td>
<td>2</td>
</tr>
<tr>
<td>1.3 OUTLINE</td>
<td>4</td>
</tr>
<tr>
<td>2 RESEARCH APPROACH</td>
<td>6</td>
</tr>
<tr>
<td>2.1 INTRODUCTION</td>
<td>6</td>
</tr>
<tr>
<td>2.2 RESEARCH OBJECTIVE</td>
<td>6</td>
</tr>
<tr>
<td>2.3 RESEARCH QUESTION</td>
<td>6</td>
</tr>
<tr>
<td>2.4 RESEARCH METHODOLOGY</td>
<td>8</td>
</tr>
<tr>
<td>2.5 PRACTICAL AND SCIENTIFIC RELEVANCE</td>
<td>12</td>
</tr>
<tr>
<td>3 LITERATURE OVERVIEW</td>
<td>13</td>
</tr>
<tr>
<td>3.1 INTRODUCTION</td>
<td>13</td>
</tr>
<tr>
<td>3.2 OPEN DATA</td>
<td>13</td>
</tr>
<tr>
<td>3.2.1 LINKED DATA AND LINKED OPEN DATA</td>
<td>16</td>
</tr>
<tr>
<td>3.2.2 META DATA</td>
<td>18</td>
</tr>
<tr>
<td>3.2.3 OPEN GOVERNMENT DATA MODELS</td>
<td>19</td>
</tr>
<tr>
<td>3.2.4 OPEN DATA PROCESS</td>
<td>22</td>
</tr>
<tr>
<td>3.3 OPEN DATA IMPEDIMENTS</td>
<td>23</td>
</tr>
<tr>
<td>3.4 OPEN DATA REQUIREMENTS</td>
<td>27</td>
</tr>
<tr>
<td>3.5 USER INTERFACE DESIGN</td>
<td>33</td>
</tr>
<tr>
<td>3.5.1 WEB SITE DEVELOPMENT PROCESS</td>
<td>34</td>
</tr>
<tr>
<td>3.5.2 REQUIREMENT ENGINEERING PROCESS AND TECHNIQUES</td>
<td>35</td>
</tr>
<tr>
<td>3.5.2.1 Requirement Elicitation</td>
<td>37</td>
</tr>
<tr>
<td>3.5.2.2 Requirements Specification</td>
<td>39</td>
</tr>
<tr>
<td>3.5.2.3 Requirements Validation</td>
<td>39</td>
</tr>
<tr>
<td>3.5.3 SCENARIOS AND USE CASES FOR REQUIREMENT ELICITATION AND ANALYSIS</td>
<td>40</td>
</tr>
<tr>
<td>3.5.4 USER INTERFACE DESIGN GUIDELINE</td>
<td>41</td>
</tr>
</tbody>
</table>
3.6 **CHAPTER SUMMARY**

4 **ANALYSIS & DESIGN OF USER INTERFACE**

4.1 **INTRODUCTION**

4.2 **REQUIREMENT ANALYSIS**

4.2.1 **ENGAGE OPEN DATA PLATFORM**

4.2.1.1 Major characteristics and objectives of the platform

4.2.1.2 Involved Users and responsibilities

4.3 **SCENARIOS**

4.3.1 **SCENARIO 1: LOOK FOR DATASETS, ANALYSE, SAVE AND EXPORT THE RESULTS**

4.3.1.1 Scenario Description

4.3.1.2 Use Case diagram

4.3.1.3 Use cases

4.3.1.4 BPMN Diagrams

4.3.1.5 User interface design

4.3.2 **SCENARIO 2: SEARCH, SUBMIT AND REPLY DATASET REQUEST**

4.3.2.1 Scenario Description

4.3.2.2 Use case diagram

4.3.2.3 Use cases

4.3.2.4 BPMN diagram

4.3.2.5 User interface design

4.3.3 **SCENARIO 3: DOWNLOADING, PROCESSING, COMBINING DATASETS AND PROVIDING FEEDBACK**

4.3.3.1 Scenario Description

4.3.3.2 Use case diagram

4.3.3.3 Use cases

4.3.3.4 BPMN diagram

4.3.3.5 User interface design

4.3.4 **SCENARIO 4: MANAGING USER ACCOUNT AND COMMUNICATIONS.**

4.3.4.1 Scenario Description

4.3.4.2 Use case diagram

4.3.4.3 Use cases

4.3.4.4 BPMN diagram

4.3.4.5 User interface design

4.3.5 **SCENARIO 5: PREPARING, UPLOADING AND MODIFYING DATASET AND METADATA**

4.3.5.1 Scenario Description

4.3.5.2 Use case diagram

4.3.5.3 Use cases

4.3.5.4 BPMN diagram

4.3.5.5 User interface design

4.3.6 **SCENARIO 6: USING DOCUMENTATION/TUTORIALS**

4.3.6.1 Scenario Description

4.3.6.2 Use case diagram

4.3.6.3 Use cases

4.3.6.4 BPMN diagram
List of Figures

Figure 1-1 Chapter overview ........................................................................................................... 5
Figure 2-1 Research questions and related corresponding chapter .................................................. 8
Figure 2-2 Requirements engineering process (Sommerville, 2010, p. 99) ........................................... 10
Figure 2-3 Research flowchart ....................................................................................................... 12
Figure 3-1 Open Government Data Stage Model (Kalampokis et al., 2011, p. 241) ....................... 20
Figure 3-2 Open data process (Janssen & Zuiderwijk, 2012) .......................................................... 23
Figure 3-3 Possible features of open data platform (Braunschweig et al., 2012, p. 4) ...................... 28
Figure 3-4 Possible features of open data dataset (Braunschweig et al., 2012, p. 2) ....................... 28
Figure 3-6 Requirement Engineering Process .................................................................................. 37
Figure 4-1 Research methodology .................................................................................................. 45
Figure 4-2 Relation between open data process and scenarios ........................................................ 48
Figure 4-3 Use case diagram of scenario 1 ....................................................................................... 50
Figure 4-4 BPMN diagram: Task flow of use case 1.1 ................................................................. 56
Figure 4-5 BPMN diagram: Task flow of use case 1.2 ................................................................. 56
Figure 4-6 BPMN diagram: Task flow of use case 1.3 ................................................................. 57
Figure 4-7 BPMN diagram: Task flow of use case 1.4 ................................................................. 57
Figure 4-8 BPMN diagram: Task flow of use case 1.5 ................................................................. 58
Figure 4-9 Wireframe Search result page ....................................................................................... 59
Figure 4-10 Wireframe Output type ............................................................................................... 59
Figure 4-11 Wireframe search result analysis ............................................................................... 60
Figure 4-12 Wireframe Saved searches .......................................................................................... 60
Figure 4-13 Use case diagram of scenario 2 .................................................................................. 63
Figure 4-14 Use case diagram of scenario 2b ................................................................................ 68
Figure 4-15 BPMN diagram: Task flow of use case 2.1 ............................................................... 68
Figure 4-16 BPMN diagram: Task flow of use case 2.2 ............................................................... 69
Figure 4-17 BPMN diagram: Task flow of use case 2.3 ............................................................... 69
Figure 4-18 BPMN diagram: Task flow of use case 2.4 ............................................................... 70
Figure 4-19 Wireframe dataset request result ............................................................................... 70
Figure 4-20 Wireframe add new dataset request .......................................................................... 71
Figure 4-21 Wireframe Dataset request preview .......................................................................... 71
Figure 4-22 Wireframe dataset request page and feedback ............................................................ 72
Figure 4-23 Use case diagram of scenario 3 .................................................................................. 74
Figure 4-24 BPMN diagram: Task flow of use case 3.2 ............................................................... 79
Figure 4-25 BPMN diagram: Task flow of use case 3.3 ............................................................... 79
Figure 4-26 BPMN diagram: Task flow of use case 3.4 ............................................................... 80
Figure 4-27 Wireframe Dataset main page .................................................................................... 80
Figure 4-28 Wireframe dataset use ............................................................................................... 81
Figure 4-29 Wireframe dataset page (continue) .............................................................................. 81
Figure 4-30 Wireframe quick extend ............................................................................................. 82
Figure 4-31 Wireframe dataset visualization tool ........................................................................... 82
Figure 4-32 Use case diagram of scenario 4 ................................................................................. 85
Figure 4-33 Use case diagram of scenario 4b ............................................................................... 86
Figure 4-34 BPMN diagram: Task flow of use cases 4.1 & 4.2 & 4.3 ............................................. 91
Figure 4-35 BPMN diagram: Task flow of use cases 4.4 ............................................................... 92
Figure 4-36 BPMN diagram: Task flow of use cases 4.5 ............................................................. 92
Figure 4-37 Wireframe User profile main page .............................................................................. 93
List of Tables

Table 3-1 Open Data Definition ................................................................. 14
Table 3-2 Open data users(Wood et al., 2010) .................................................. 16
Table 3-3 “5 Star” Scheme proposed by Berners-Lee (2009) .................................... 18
Table 3-4 OGD model summary (Kalampokis et al., 2011) ..................................... 21
Table 3-5 Data access impediments (Zuiderwijk, Janssen, Choenni, et al., 2013) ....... 24
Table 3-6 Data use impediments(Zuiderwijk, Janssen, Choenni, et al., 2013) .......... 25
Table 3-7 Data Deposition impediments(Zuiderwijk, Janssen, Choenni, et al., 2013) ... 27
Table 3-8 Open data requirements (access) ..................................................... 29
Table 3-9 Open data requirements (Searching) ................................................... 29
Table 3-10 Open data requirements (Navigation) ................................................ 29
Table 3-11 Open data requirements (Uploading) ................................................... 30
Table 3-12 Open data requirements (Downloading) ............................................. 30
Table 3-13 Open data requirements (Data Quality) ............................................. 30
Table 3-14 Open data requirements (Analysing) .................................................. 30
Table 3-15 Open data requirements (Visualizing) ............................................... 31
Table 3-16 Open data requirements (Linking and combining data) ......................... 31
Table 3-17 Open data requirements (Collaboration) ............................................. 31
Table 3-18 Open data requirements (Support and Help) ........................................ 32
Table 3-19 Open data requirements (Feedback) ................................................... 33
Table 4-1 Use case 1.1 description .................................................................... 51
Table 4-2 Use case 1.2 description .................................................................... 52
Table 4-3 Use case 1.3 description .................................................................... 53
Table 4-4 Use case 1.4 description .................................................................... 54
Table 4-5 Use case 1.5 description .................................................................... 55
Table 4-6 Use case 2.1 description .................................................................... 64
Table 4-7 Use case 2.2 description .................................................................... 65
Table 4-8 Use case 2.3 description .................................................................... 66
Table 4-9 Use case 2.4 description .................................................................... 67
Table 4-10 Use case 3.1 description ................................................................... 75
Table 4-11 Use case 3.2 description ................................................................... 76
Table 4-12 Use case 3.3 description ................................................................... 77
Table 4-13 Use case 3.4 description ................................................................... 78
Table 4-14 Use case 4.1 description ................................................................... 86
Table 4-15 Use case 4.2 description ................................................................... 87
Table 4-16 Use case 4.3 description ................................................................... 89
Table 4-17 Use case 4.4 description ................................................................... 90
Table 4-18 Use case 4.5 description ................................................................... 90
Table 4-19 Use case 5.1 description ................................................................... 97
Table 4-20 Use case 5.2 description ................................................................... 98
Table 4-21 Use case 5.3 description ................................................................... 99
Table 4-22 Use case 5.4 description .................................................................. 100
Table 4-23 Use case 6.1 description .................................................................. 107
1 Introduction

1.1 Introduction

Nowadays, information systems are growing fast in terms of size and complexity. Technological developments have been directed towards producing data in massive scale (worldwide), and increasing data storage capacity and communication bandwidth (Davies & Edwards, 2012). It was estimated by Hilbert and López (2011) that the global information that was digitally stored in 1986 was less than 1 percent. Surprisingly, it is assumed that in 2007 94% of data was digital. Besides, in recent years there is a movement towards making different kinds of data publicly accessible, because the availability of such data can provide benefits for both data providers and - users (Fioretti, 2011).

Governments are a massive source of data. They produce and collect a wide range of data. Moreover, technological evolutions have exceptionally modified the role of public sector information (PSI). PSI sources are valuable when they contain accurate, up-to-date and easy accessible content. The technological advancements enables digitizing the existing or produced data resources, and making those resources easily accessible (Vickery, 2011). Combing and re-using PSI with other sources of information is a developing industry. It was estimated that direct PSI re-use market in Europe is worth 27 billion euros (Dekkers et al., 2006). This emphasizes the significance of providing access to PSI (Lundqvist, 2013). Therefore, different countries have focused to publish data for public in order to provide benefits for their citizens.

The announcement of President Barack Obama’s Administration regarding initiation of high level of openness to ensure public trust and increase transparency, public participation, and collaboration is considered as the starting point of Open Data movement (McDermott, 2010). In the same announcement, Obama also stated “Openness will strengthen our democracy and promote efficiency and effectiveness in Government”. (Obama, 2009, p. 1). Furthermore, this global movement engendered the launch of the Open Government Partnership (OGP) in September 2011. The OGP aims to protect commitments from governments in order to ensure transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance. The OGP consists of 63 government partners. One of the responsibilities of the OGP members is to increase the availability of information about governmental activities. The data will be gathered by governments on behalf of public ("Open Government Partnership (OGP)," 2013).

Open data can help to accelerate knowledge and information sharing. Knowledge sharing can deliver potential benefits for organizations on several levels. To put it briefly, knowledge sharing can provide more efficient data management, develop information infrastructure, create more integrated services, and improve interactions among the involved organizations (Dawes, 1996; Landsbergen & Wolken, 1998; McCaffrey et al., 1995; Zhang et al., 2005).

Information sharing is a possible way to reduce the replication of data collection and data handling. Also, information sharing can provide opportunities to improve technological
infrastructure, enhance related business processes, and increase quality of information (Dawes, 1996). In practice, it can provide integrated services in a more effective, efficient and responsive way (Edmiston, 2003; Landsbergen & Wolken, 1998). Information sharing can help to increase trust among different involved organizations, and between governments and people (Cresswell et al., 2002). Having access to the entire information can help organizations through decision-making about prospect progress and system design (McCaffrey et al., 1995). Moreover, having access to complete information by the public can increase transparency in the governments’ decisions and make governments more reliable and efficient (Wong & Welch, 2004).

Even though organizations benefit from open data, but sustaining the information sharing systems is a challenging task, because barriers are attached to social, economic, and political principles and values of organizations (McCaffrey et al., 1995). Zhang et al. (2005) summarized the barriers of Open Data into three main categories namely technological, organizational, and legal and policy barriers (Arzberger et al., 2004; Sayogo & Pardo, 2012). To support information sharing and provide superior value for exchanged information, the contributors need to adopt consistent data standards and appropriate infrastructure (technological barriers) (Dawes, 1996; Zhang et al., 2005).

In this research, we focus on technological barriers of knowledge sharing. More precisely, we focus on Linked Open Data (LOD). The idea of publishing data in a way that is re-usable and in a machine-readable format is called linked data (Berners-Lee, 2009). Linked Data is simply about (re-)using the Web to create typed links between data from different sources (Christian Bizer et al., 2009). Linked open data is about publishing open data in a more useful way. This can be done by interlinking datasets in a semantic way, which enables data sources to be connected and queried (Yu, 2011). In the remaining of this thesis, Open Data is used instead of the Linked Open Data. Converting public sector information (PSI) into LOD is a process, which consist of 5 steps (Zuiderwijk, Jeffery, et al., 2012a). These steps will be discussed later in chapter 3.2.4. To enable this process, relevant data (meta-data) about the PSI should be published to re-use the data and link them.

Currently considerable research is being conducted in the field of open data. Linking social media information with datasets in open data platforms improves user modeling and recommender systems (Abel et al., 2012). “Recommender systems are intelligent applications, which assist users in their information-finding tasks, that best suit their needs and preferences” (Tariq & Francesco, 2009, p. 5).

1.2 Problem Statement

Open data can become available by open data infrastructures. In recent years, different governmental and private organizations have engendered open data platforms such as US Government open data platform and UK government open data platform ("UK government open data platform," 2013; "US government open data platform," 2013). Moreover, there are
start-up companies that are providing services based on open data. For instance Buienradar in the Netherlands provides weather forecast platform based on an open data from Dutch government (Fioretti, 2011).

Additionally, Other than governmental and public agencies, they are other type of open data platforms that are not particularly developed by any public or private organization ("Engage open data platform," 2013). These universal open data platforms facilitate combining data from several data resources, and combining various data types (Zuiderwijk, Janssen, & Parnia, 2013). Also, these platforms have different types of users (e.g. researchers, citizens, journalists, developers, public servants). Different users use open data for various purposes, and they are interested in different services of open data platforms and therefore they have different needs.

Open data is almost a new field, and many impediments restrict users to have easy access. Several impediments for the use of open data are mentioned in literature. The most important impediments are categorized by (M. Janssen et al., 2012; Zuiderwijk, Janssen, et al., 2012) are: Data deposit impediments that refer to difficulties with the deposition of data, for instance data uploading is limited to one data format. Data access impediment is about difficulties of obtaining data and lack of central data center. Data use impediments; Due to lack of knowledge using data may be restricted. Besides relating to the user interface impediments in open data platform, accessing data requires registration or data request. Open data platforms provide support for a limited group of users and even they have access they will face with difficulties to search and browse data (Zuiderwijk, Janssen, & Parnia, 2013).

Data publishers usually only make data available and often do not consider the user perspective. It was argued in (M. Janssen & Zuiderwijk, 2012) data publishers mainly focus on publishing data, although the open data process should cover more processes, so that it can bring added value for both users and the public sector (Davies & Edwards, 2012). Hence, there is a need for research to address the demand for open data from different stakeholder perspective.

One aspect of the user perspective that has received very little attention is the user interface of open data platforms. More specifically, a coherent and integrated set of user interfaces, based on services that can be used to create a user-driven process for dealing with open data, is lacking. A good UI can encourage participation and communication of users; also it enables ease of use of the platform, which can guarantee the service-use and citizens engagements. Besides the hidden advantages of user involvement is increasing the Big Data knowledge of users. Big data is the word for a collection of large and complex data sets, which is difficult to process using traditional data processing applications (Gordon, 2013).

Customized UI’s can help in providing various services for particular open data users. In the requirement analysis, we will identify the potential users of open data and categorize them based on their requirements. Various groups of users have different demands from the platform. Each group can gain more benefits by interacting with the platform, and this enables ease of use for them. Introducing the real advantage and benefits of open data, and attracting
the users to use this service improves the process of opening the data. This is beneficial for both public sector and the users.

UI should be designed in a way that it can fulfill gaps between users and data public sector. Main concern of the public sector is about publishing data and privacy issues related. However, there should be added value for both government and public. Also, there is no mechanism to monitor data use, and feedback system to increase the open data process efficiency (M. Janssen & Zuiderwijk, 2012).

The platform enables users to access data in different formats. The purpose is to enable users to process and reuse data in their desired way. This can be done by, providing an overview of dataset, and to enabling data visualization. The two steps mentioned above give users the opportunity to check whether data is of interest. Afterwards, users can access data in their preferred format (for instance XLS, comma-separated value (CSV), Cascading Style Sheets (CSS), resource description framework (RDF)).

1.3 Outline

This thesis consists of 6 chapters. We describe contents of all chapter and Figure 1-1 represents chapters overview. The outline of this thesis will be as follows:

• Chapter 1 – Introduction

This chapter gives a detailed explanation about the subject and point outs the background of the problem and the context of the problem that is researched. Also, the outline of thesis is provided in this section.

• Chapter 2 -Research approach

This chapter contains, a clear explanation of the research objective that will be addressed and research questions that become from the problem identified. The methodologies that will be applied to answer the research questions will also be stated based on the appropriate approaches to reach the answer. Also, the practical and scientific contribution of the research that is conducted is included in this chapter.

• Chapter 3 – Literature overview

This section will further detail explain the concept of open data and benefits and barriers of open data. Brief descriptions of Linked Open Data, metadata, Open Government Data models, and Open data Process is given. Next a discussion about impediments and requirements of UI of Open Data platform is provided. Following that in the second part of this chapter the user interface design process is included. Briefed description of Interface Design needs, Requirement engineering processes and techniques, Scenarios for requirement analysis, and User Interface design principles is stated.
• Chapter 4 – Analyze requirements and design of User Interface

This section provides detail about the functionality of open data platform. The scenarios are developed for requirements analysis. Thereafter use cases are developed for different scenarios, and detailed description of use cases is presented in tables. BPMN diagrams are used to represent the flow of each use cases. Finally, the wireframes designed for the platform to represent the user interface. This section will then provide the result of the analysis and address the main features of the platform and different audience classification. Finally, The facade use case is used to represent the main function of the platform.

• Chapter 5 – Evaluation

Based on the findings from chapter 3 and analysis and design in chapter 4, this section evaluates the proposed design. This will be done by conducting usability test. First we describe the method of evaluation by introducing the participants, defining the measures and describing the procedure of the evaluation. Finally, a descriptive analysis of the result is presented.

• Chapter 6 – Conclusions and Discussion

This section imitates the key discoveries and conclusions to the research question on user interface design of open data platform. The user interface (mockups) and detail functionality will be applied to propose a new user interface for open data platform and also further research on the topic.

Figure 1-1 Chapter overview
2 Research approach

2.1 Introduction

This chapter first describes the research objective based on the problem statement. Next, based on the objective the main question and the subsequent research sub-questions are formulated, 2.2. The Research methodology that is used in this research is discussed in, 2.3. Finally practical and scientific relevance is addressed in section 2.4.

2.2 Research Objective

The aim of this thesis is to create integrated user interfaces for different groups of users, which stimulate services for user of Linked Open Data (LOD), including the processes of importing, curating, processing and visualizing LOD. Integration of user interface is about integrating the UI components by mixing their presentation front-ends, instead of their application logic or data. When comparing UI integration with other forms on integration, it is important to realize the main difference with other type of integration, but also there is numerous lessons that can be learned from these differences and understanding the characteristics of UI integration. The goal of UI integration is to create composite applications that control the components’ individual UIs to make richer, composite UI applications (Daniel et al., 2007).

The focus is on service engineering approach to Identify and design user-services, which are necessary for dealing with different types of users of LOD. There is no overview of services available and the processes needed for processing linked open data yet. Current infrastructures take a narrow view and focus on publishing data, but do not support processing, visualizing, interpreting of open data. Meta data are vital for the platform in a way that it can link user area of interest to the proper data and related services. Of course the backbone of this platform is datasets, which can be provided by public sector and even the users.

2.3 Research Question

To achieve the research objective, set in section, research questions are developed consisting of a main question and sub-questions. These sub-questions divided into the specific aspects that are addressed in the principle question. Since different users involved in this platform an integrated user interface will cover different user interactions in a way that user’s requirements fulfilled. This will bring following research questions. The main research question is:
**Main Question.** How can user interface improve the process of open data?

The outcome of this research is design of User Interface for an open data platform and evaluation of proposed mockups; this can help to guarantee user engagement and gain more benefits for both public sector and users. It will be attained by developing different scenarios based on the requirements, and next, designing the User Interface based on the scenarios. This will be accomplished by answering the following sub-questions.

**Sub-Question1.** What are the impediments of interfaces of open data platforms?

Identifying impediments regarding different user's need provides better understanding of real system requirements. And it leads to design a user favorable interface. To answer the question research will be conducted about different user groups need and their constraints to interact with the platform. This research question is answered by conducting literature survey about impediments of current open data platform. The expected outcome of this sub-question is an overview of impediments of UI of open data platforms. The result of the literature review is grouped in section 3.3.

**Sub-Question2.** What are the requirements for user interfaces of open data websites?

We need to clarify open data platform requirements as an input for the UI design, in order to design a user-centered platform, which can guarantee users engagement. This leads to identify different users group’s needs, and public sector requirements, to provide requirement table as an input for next phase of design. To answer this question literature review will be conducted about open data requirements and requirement analysis methods, to categorize open data platform requirements according to the analysis method. The expected outcome of this sub-question is an insight into open data requirements, and User Interface Design. Section 3.4 contains the requirement table of open data platform, and section 3.5 provides an overview about the process of designing an User Interface.

**Sub-Question3.** How User Interface can increase the use of open data?

This research question will be answered by using the requirements identified in the previous sub-questions, to design a User Interface for Open Data platform. This will be done by developing scenarios, which will cover earlier identified requirements. The expected outcome of this research question is mockups that represent the requirements identified in previous research question. Section 4.4 contains answer to this sub-question.

**Sub-Question4.** What is the value of the proposed User Interface of open data?

Evaluating the mockups of the open data platform provides a better overview about the real value of the proposed UI. To evaluate the proposed UI we will design a usability test. This will assess by filling out a questionnaire by various users, and inspecting them during evaluation process. Finally we will describe the aspects that will be mentioned by participants to answer research sub-question. This provides an opportunity to realize the strengths and weaknesses of the open data platform. Finally, it is expected to deliver recommendation for possible enhancement to the platform. This sub-question is answered in section 5.3.
Figure 2-1 provides, an overall overview about research question and sub-question, methodologies that are used to answer to each sub-questions, and corresponding chapter regarding each sub question.

![Diagram](Image)

**Main question:** How can user interface improve the process of open data

**Sub question 1:** What are the impediments of interfaces of open data platforms?
**Literature review**

**Sub question 2:** What are the requirements for user interfaces of open data websites?
**Requirement analysis**

**Sub question 3:** How user interface can increase the use of open data?
**Scenario, Use case, BPMN diagrams**

**Sub question 4:** What is the value of the proposed user interface open data?
**Usability test, requirements test**

Section 3.2
Section 3.3 & 3.4
Chapter 4
Chapter 5

**Figure 2-1 Research questions and related corresponding chapter**

### 2.4 Research Methodology

This research focuses on the field of Open Data, to get insight into the potential benefits of User Interfaces for open data platform. In order to answer the proposed research questions, various methodologies are applied; literature review, developing scenarios, questionnaire, and inspection, based on proposed User Interface. The information gathered via these methods was deployed to design several User Interface mockups and assess the value created by the user interface. Below, each of the techniques is described in detail.

A literature review was conducted to collect information from journals and scientific articles about open data. Literature search conducted at search databases, using different keywords regarding, impediment and requirements of open data, user interface design, and requirement engineering process. Since open data is relatively novel and developing field most significant criteria to assess the relevance of the articles, is whether they are up to date, or not.
When finding relevant articles the references used in these articles were also studied, in some cases this resulted in finding other related articles. The literature research gathered data and information about impediments and requirements of UI of Open data platform, and UI design process.

A systematic approach was deployed in the literature search, keeping track of the keywords, sources and the similar results. First of all, a literature research is conducted to find impediments of user interface of open data platform. Search terms include, ‘open data impediments’, and ‘knowledge sharing into databases such as Google Scholar, Scopus, and TU Delft repository. We have searched for literature based on their keyword. Primary “Open Data” searched in whole parts of the papers and around 1150 results were returned. Afterwards, the search for “Impediments” and “Knowledge sharing” were restricted to keywords, and the result was around 180 papers. At the last stage “User Interface” were added to the keywords and total of 40 related papers were found. Second, a literature research is conducted to acquire knowledge about requirements of user interface of open data platform. Next, Search terms that are entered to search databases. At the beginning we have searched for ‘open data requirements’ and total of 600 results were returned. Next the search was restricted to the keywords ‘User Interface’ and the result was about 140 papers. Finally, we have added ‘requirements analysis’ to the keywords and the results were around 35 papers.

Further literature study is conducted to attain knowledge about User Interface Design process and requirement engineering process. Furthermore, literature that is more focused on dimensions developed.

To design proper services for open data users, we will follow the requirement engineering approach. The complete requirement engineering cycle is followed from deriving requirements, designing services, and user-interface, to evaluation of proposed design. This approach focus on a real system needs by deriving requirements in an iterative cycle. Besides, the process of opening data is an iterative process, which can be organized around spiral model (Robertson & Robertson, 1999; Sommerville, 2010). Requirements engineering process consists of four high level activities that results into requirements document. Detailed information regarding requirements engineering process can be found in section 3.5.2. These four activities will be organized in a way that, they support process of opening data. Description about different steps of open data process can be find in section 3.2.4. Meanwhile, open data process requires close interactions between data users and data providers, due to its iterative essence; Open data process will be matched to spiral model.

Requirements engineering process consists of four high level activity (Sommerville, 2010). A spiral view of the requirements engineering process is represented in Figure 2-1.

1. First activity of this process is a feasibility study, which is about focusing on assessing whether or not, system supports different users interactions.
2. Elicitation and analysis are the second activity that defines the application domain and in order to realize what services should be provided. This activity aims to specify exact system definition and required performance of the system by understanding end-users needs.
3. Converting the requirements into standard form is covered in *specification* phase.
4. The last and fourth activity is about checking whether or not these requirements define system that the customer wants.

![Figure 2-2 Requirements engineering process (Sommerville, 2010, p. 99)](image)

Third research sub-question is answered in two steps. First, we will derive requirements, and second we will design UI based on requirements. We have performed Requirements analysis by developing several scenarios and correspondence use cases. Scenarios are developed in a way that they cover all steps of open data process. By addressing the user interface requirements in previous sub question, and providing a list of requirements of open data we will develop scenarios to cover different steps of open data process.

Scenario explains relations between system and users. They are defined as sequence of actions and events (Some, 2005). They are stories that contains situation state in which one or more actors with personal motivations and knowledge, and different tools and objects are involved in. The scenarios represent sequence of actions and events that end up to a result. These actions and events are correlated in a way that system is used in a specific episode (Rosson & Carroll, 2009). Six scenarios will be developed to cover all steps of open data process.

For each scenario, we will develop use case diagram to represent basic flow of each task. They are used to represent the interactions between the system and the external entities of the system. External entities are actors, other information systems or external events. Meanwhile,
scenarios are not highly detailed and they cannot reflect the whole story. Therefore we have developed different use cases to cover various aspects of the system in order to represent detailed interactions.

To design UI prototypes, we will visualize the flow of use cases in BPMN diagrams. BPMN provides a standard way to define and analyze business processes and represents the flow and sequence of the actions. Moreover, It provides a standard notation that is readily understandable by managers, analysts and developers. BPMN was developed to narrow communication gaps between the different departments within an organization. Finally we represent the corresponding mockup for each use case.

To answer fourth research sub-question and evaluate the value of the proposed UI for open data platform two types of validation techniques is used. First, walk-through technique is used to match the requirements and proposed prototypes. Second, Usability testing is used to evaluate the value of design.

In Web page design, a design evaluation of iterative design process is adopted to improve Web usability. The traditional usability engineering emphasizes the precise definition of Usability metrics. (Yan & Guo, 2010). Usability tests are divided into three main categories (Battleson et al., 2001): inquiry, inspection, and formal usability testing. The first and the last categories require real user participation, but the second does not. Besides Different usability evaluation method for web user interface have been introduced and according to Tan et al. (2009) user testing and heuristic analysis are two of the most popular ones. In this thesis, we have deployed user testing usability test. This is performed by using user-testing technique. User testing is conducted in a scenario-based environment, and inspecting participants during the evaluation process. 3 scenarios consisting of approximately 15 tasks are developed and participants performed the evaluation by finding and conducting tasks on screen, and next rating task how easy/hard was to perform. Finally, Result of this evaluation is represented by performing qualitative descriptive, method.

By using mentioned above methods information was gathered about: Impediments and Requirements of Open Data, Requirement engineering process, User Interface design and evaluation of proposed prototype.

Figure 2-3 represent the research flow and methodologies that are used in each step. To answer first two-research question, literature research has been conducted. Next we followed requirements engineering approach to elicit and analyze requirements by using scenario and use case techniques. We have designed prototypes by visualizing use case flows in BPMN diagram. Finally, we have conducted an evaluation by user testing in the scenario-based environment, and walk-through technique.
2.5 Practical and Scientific relevance

This master thesis focuses on both practical and scientific aspects of the research. We focus gaining more benefits from open data by designing (improving) the user interface to increase the user’s contribution, as well as contributing into current scientific studies in the field of open data.

Open data is a relatively new field, and at the moment there is still very little literature available on requirements of open data, especially from the perspective of users and user interface requirements. Existing literature emphases on ways to produce and publish open data and mostly on ways to make data available. Although, the users of open data are in the target and have a crucial role to attain benefits of open data. Without intense user involvement, the goals cannot be achieved. This research focus on different users’ needs while analyzing functional requirements of open data platform to increase functionality and user engagement. This result in scientific contribution of the research that brings a comprehensive source of user requirements of open data.

Governmental organizations can gain different benefits from open data. Moreover, opening data can increases the transparency for the governmental organizations, and brings a significant financial value for citizens (users). Thus, focusing on the users’ needs and providing a good user interface can increase the engagement of the users. This research will contribute to address user requirements to design a user interface for different users to gain the benefits. Also, the user interface presented in this research will be used to design user interface for open data platforms.
3 Literature overview

3.1 Introduction

As it is discussed in the introduction chapter, in this thesis we are going to design an integrated user interface for open data platform. This chapter is divided into two main sections. In the first section, primarily we will provide an overview of open data. Next we will discuss open governmental model, metadata. Moreover we provide an overview of open data process. Finally we have detailed discussion about open data impediments and we list the open data platform requirements. The second section contains a quick review on user interface design and requirements engineering process, and User Interface design principles. In this section, we aim to find out to what extent these theories can help us to create a requirement list for open data platform for addressing the research problem that is introduced earlier in the chapter.

This chapter covers the first two sub questions of research, which are “What are the impediments of interfaces of open data platforms?” and “What are the requirements for user interfaces of open data websites?” Literature review is deployed to answer the first sub-question to present the impediments of open data platform user interface. And the second sub-question is countered based on service engineering approach is to gather the requirements of open data platform.

3.2 Open data

This section discusses the literature found relevant to this research: Open data definition, Linked data and Linked Open Data, Meta data, Open Government Data (OGD) models, and Open Data process. These elements are relevant to this research as they discuss parts of the research sub-questions 2. They describe the core of the technological advancement of open data platforms. All of the topics are vital to consider while designing the open data platform. This literature is used address the primary requirements of open data platform.

There are various definitions of Open data in previous research work. Different explanation of open data is because of inconsistency in the source of data, format of data, and data users. In order to determine the definition that is mostly related to this research work, we will provide an overview of different definitions of Open Data. Table 3-1 provides various definitions of Open Data in previous research and by authorities.
Table 3-1 Open Data Definition

<table>
<thead>
<tr>
<th>Definition</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>“A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.”</td>
<td>((OKF), 2013a)</td>
</tr>
<tr>
<td>“Open Data is the Public Sector Information (PSI) in standard formats, open and interoperable, facilitating their access and enabling reuse.”</td>
<td>(Rojas et al., 2013, p. 51)</td>
</tr>
<tr>
<td>“All stored data which could be made accessible in the public interest without any restrictions on usage and distribution.”</td>
<td>(Geiger &amp; Lucke, 2011, p. 184)</td>
</tr>
<tr>
<td>“Data based on following criteria:</td>
<td>Open Data White Paper(UK, 2012, p. 8)</td>
</tr>
<tr>
<td>• Accessible at, no more than cost of reproduction, without limitations based on user identity or intent, in a</td>
<td></td>
</tr>
<tr>
<td>• Digital, machine readable format for interoperation with other data, and</td>
<td></td>
</tr>
<tr>
<td>• Free of restriction on use or redistribution in its licensing conditions”</td>
<td></td>
</tr>
<tr>
<td>“Open data is non-personally identifiable data produced in the course of an organization’s ordinary business, which has been released under an unrestricted open license.”</td>
<td>Adelaide City Council (Technology, 2013)</td>
</tr>
<tr>
<td>“Data that is freely available and can be used as well as republished by everyone without restrictions from copyright or patents”.</td>
<td>(Braunschweig et al., 2012)</td>
</tr>
</tbody>
</table>

By comparing each of the definitions presented above we will come to the fact that the main idea is quiet identical. The definition by Open Data White Paper contains a detailed description of Open Data. Also, it is indicated that the data needs to be machine readable and easily accessed. In contrast, the definition that is provided other sources mostly focus on the data use and accessibility and do not focus on the precise requirements of data. Thus, the definition open data in this research is based on the by Open Data White Paper explanation.

The objective of Open Data is to open all non-personal and non-commercial data, particularly all information gathered and handled by government organizations. The idea is much alike to the Open Source or Open Access developments (Braunschweig et al., 2012).
The instance of contents of open data can be train schedule, radio and television programs, education material, geo data, statistics, traffic data, scientific publications, or medical studies (Geiger & Lucke, 2011). They are not the information like the individual’s name, addresses that can be related to specific person (Kulk & Van Loenen, 2012). The concept of open data is about the data that is most anonymous and cannot be related to individuals, the Open Knowledge Foundation express open data and “private data” as a diverse concept without common subsets.

The published data spreads an extensive variety of areas. Data can be combination of PSI data or even data from businesses, universities, news agencies or non-profit- organizations (NGOs) (Geiger & Lucke, 2011). Publishers can be single government agencies or suppliers of bigger repositories that gather public datasets and make them accessible in an centralized and standardized way (Braunschweig et al., 2012).

The present state of Open Data requires citizens who wish to access and utilize this data, to first recognize relevant datasets manually. This involves finding organization and agencies, which they publish open datasets on platforms that gives central and responsive entry point to search for data. Different organizations most likely has their own publishing policies, the data can be exceptionally heterogeneous. Software tools can uphold clients with the integration of their data. There are two opposite styles to publish data in current Open Data platforms. In some platforms, data are in human-readable format while in others are in machine-readable format (Braunschweig et al., 2012).

The first approach provides human readable information rather than raw-data. This naturally indicates that existing raw data is transformed to make data conceivable. But these types of data do not allow re-use of data. The procedure of recovering machine-readable data from human-readable documents is extremely challenging (Braunschweig et al., 2012).

The second approach is publishing the raw data in a machine-readable format to empower software tools to process it. As a substitute for preprocessing the data, users can straightforwardly access the raw data and modify the data for their own particular needs. Besides It can be used for different use cases without extra processing(Braunschweig et al., 2012).

Braunschweig et al. (2012) Discussed that the second approach doesn't restrain the amount of possible use cases, also it does support reusability. Consequently, a perfect Open Data platform ought to be optimized towards technical users and programmatic reuse.

Vast variety of users has been identified as potential users of Open Data. The most important users categorized by Wood et al. (2010) are; citizen, funders and policy makers, researchers and people from industry. Table 3-2 summarize benefits for potential beneficiaries.
Table 3-2 Open data users (Wood et al., 2010)

<table>
<thead>
<tr>
<th>Beneficiaries</th>
<th>Benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citizens</td>
<td>- Perceive the result and benefits arising from research.</td>
</tr>
<tr>
<td></td>
<td>- Find the proper answer to their questions, based on real evidence.</td>
</tr>
<tr>
<td></td>
<td>- Enable the knowledge and experience flow.</td>
</tr>
<tr>
<td>Funders and Policy Makers</td>
<td>- Decision making based on real evidence.</td>
</tr>
<tr>
<td></td>
<td>- Avoid unnecessary work duplication.</td>
</tr>
<tr>
<td></td>
<td>- More profit and greater return on investment.</td>
</tr>
<tr>
<td>Researcher</td>
<td>- Can have their required data and tools to increase the productivity.</td>
</tr>
<tr>
<td></td>
<td>- Multi-disciplinary environment provides new insight resulting to new solutions.</td>
</tr>
<tr>
<td>Enterprise and Industry</td>
<td>- Use the best available information for R&amp;D, increasing productivity.</td>
</tr>
<tr>
<td></td>
<td>- Create new knowledge, markets and job opportunities.</td>
</tr>
<tr>
<td></td>
<td>- Increase opportunities for mobility and knowledge exchange.</td>
</tr>
</tbody>
</table>

3.2.1 Linked Data and Linked Open Data

The Idea of World Wide Web was about creating hyperlink between web documents to change the way of sharing information. The hyperlinks enables integration of all web documents into a global information space (C. Bizer, 2009). Recently Web is changing from a global information space of linked documents to a virtual place where both documents and data are linked (Christian Bizer et al., 2009). According to Berners Lee’s definition, Publishing and connecting structured data on the web is known as Linked Data (Berners-Lee, 2009).

“Linked Data is basically about using the Web, to create typed links between data from different sources” (Christian Bizer et al., 2009, p. 2). In summary linked data provides guidelines to set data level links from different resources by use of web technologies. Linked data is based on two fundamental technologies; Uniform Resource Identifiers (URIs) and Hypertext Transfer Protocol (HTTP)(Christian Bizer et al., 2009) Technologically, Linked data used HTTP URIs to identify real-world entities as well as web documents (C. Bizer, 2009). Resource Description Framework (RDF) is deployed to represent data about these entities. The Web of Linked Data can be considered as a layer above the classical Web which Interoperates with that and has many of that functionality (C. Bizer, 2009).
In recent years, an increasing number of data providers have started to adopt the Linked Data principles, accelerating the formation of worldwide data space, holding great attestations about people, books, scientific publications, geographic data, movies, music, and companies. Public organizations produce an plenty of quite important information ranging from investment statistics, the land register, and crime statistics, to your local agent's voting record (C. Bizer, 2009). Providing the simple access to this data empowers more accountability, assists individuals settle on logical decisions, and lets third parties make tools to work with the data. Besides, numerous public-sector organizations pushed to provide data from their operations in public. Linked Data is mainly based on open Web standards; data users can utilize generic tools to access, mash-up, and envision data. Also, Web search engines can grab data and utilize it to provide better services for their users (C. Bizer, 2009).

The important advantage of Linked Data from the individual perspective is the provision of integrated use of data from a wide variety of distributed and heterogeneous data sources. Linked Data browsers will also need to offer intuitive and effective mechanisms for adding and removing data sources from an integral, entity-centric view. (Christian Bizer et al., 2009)

Geiger and Lucke (2011, p. 184) has defined the Linked Open Data (LOD) as “all stored data connected by the World Wide Web which could be made accessible in the public interest without any restrictions for usage and distribution”.

LOD facilitates using data through different domains or organizational boundaries for statistics, analysis, maps and publications. Added value can be created by linking unconnected data (Geiger & Lucke, 2011; Zuiderwijk, Jeffery, et al., 2012b).

Tim Berners-Lee (2009) has developed a rating system to inspire public and private organizations to provide a good linked data. According to Berners-Lee (2009), Linked Open Data (LOD) is “Linked Data which is released under an open license, which does not impede its reuse for free”. Linked data can be used within a group or organization, and even for personal use. In general linked data doesn’t have to be open. Nevertheless if it claims to be LOD, openness is one of its necessities to get any star in rating system. Table 3-3 represents rating scheme proposed by Berners-Lee (2009). The first star can be attained when information has been made public in any format. The number of stars can be increased by providing easy access to the published information.
Table 3-3 "5 Star" Scheme proposed by Berners-Lee (2009)

<table>
<thead>
<tr>
<th>Definition</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available on the web (whatever format) but with an open license, to be Open Data.</td>
<td>★</td>
</tr>
<tr>
<td>Available as machine-readable structured data (e.g. excel instead of image scan of a table).</td>
<td>★★★</td>
</tr>
<tr>
<td>as (2) plus non-proprietary format (e.g. CSV instead of excel)</td>
<td>★★★★</td>
</tr>
<tr>
<td>All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff.</td>
<td>★★★★★</td>
</tr>
<tr>
<td>All the above, plus: Link your data to other people’s data to provide context.</td>
<td>★★★★★</td>
</tr>
</tbody>
</table>

Data publishers publish data on the web to make data accessible and usable by different applications based on the linked data principles. The following three step are essential to publish dataset as Linked Data on the web. First step is assigning URIs to the entities, which are described by the dataset. Next step is to set RDF link to other data sources on the web. This enables users to navigate on the web of linked data. The last step is about providing metadata about the published dataset. Metadata facilitates the quality assessment of data. (Christian Bizer et al., 2009). There should be a metadata about data and it should be accessible from a major list of data. And different countries are expected to create their own registries for the related government data (Berners-Lee, 2009). We will discuss and define metadata in next section.

3.2.2 Meta data

As it was mentioned above Linked Data should be published alongside of metadata to increase the value of data for its consumers. Metadata empowers data users to measure the quality of data and whether it is trustable or not. The metadata should contain information about data creator, its creating data, and creation method (Hartig, 2009).

Numerous definition of metadata is published in the research field. The most frequent one defines Metadata as “data about data” (Bargmeyer, 2000, p. 2; Jeffery, 2000, p. 2; Schuurman et al., 2008, p. 925). Moreover, Ma (2006) has examined twenty seven definitions of metadata and selected the explanation of ALCTS (2000) as best definition. It defines metadata as
“structured, encoded data that describe characteristics of information-bearing entities to aid in
the identification, discovery, assessment, and management of the described entities”.

Metadata is data that is used to describe other data, so the usage turns it into metadata. It is
machine-readable data about a dataset that contains information about the context and the
detailed use of data for the discovery (Bargmeyer, 2000). Absence of metadata, may limit
data discovery (Schuurman et al., 2008).

Metadata is used in different ways and it can be deployed to describe different characteristics
constrains the integrity of associated data, Navigational metadata facilitates the access to the
data addressed by schema metadata, and Associate metadata are classified into, descriptive,
restrictive and supportive categories.

In the context of LOD Zuiderwijk, Jeffery, et al. (2012b) have identified three types of
metadata. Discovery (flat) metadata contains data about title, creator, publisher, country, etc.
of dataset and they are descriptive and navigation metadata that enable the discovery of
related open data, Contextual metadata are data that provide rich information about
organizations, persons, projects, etc. that are descriptive, restrictive and navigational, and
Detailed metadata mostly contain data about specific domain and quality of data, they use
schema metadata to ensure quality of dataset.

The current state of metadata provision is far from the ideal state. Mainly contextual metadata
is missing and specially contextual metadata is vital to ensure reuse of LOD (Tenopir et al.,
2011). There is not much effort on metadata provision. However numerous benefits about
metadata is identified, but the exact benefits is not visible and the way that metadata can be
(re) used in LOD architectures is not clear(Zuiderwijk, Jeffery, et al., 2012b).

3.2.3 Open Government Data Models

“Public sector information” is broadly defined as “information, including information
products and services, generated, created, collected, processed, preserved, maintained,
disseminated, or funded by or for the Government or public institution”(OECD, 2008).

Henninger (2013, p. 78) provides a complete definition of PSI as “information, including all
information products in any format, and services, generated, created, collected, processed,
preserved, maintained, disseminated, or funded by or for public entities in all branches and at
all levels be presumed to be in the public domain, unless another policy option is adopted and
clearly documented, preventing it from being freely accessible to all”. Longworth (2005, p. 4)
defined PSI as “any information that is produced by a public sector entity”. The public entity
is a national or local government body, or in some cases an international organization.
The content of Public Sector Information (PSI) and Open Government Data is any kind of information that is produced and/or collected by a public sector that has of authorized organization role (Vickery & Wunsch-Vincent, 2006, p. 7).

The Open Government Data (OGD) follows the concept of the Open Data, to publish data for public freely without restriction. According to OGD principles governments act as data provider and private organizations create product and service to deliver interactive access for people (Kalampokis et al., 2011). In principle, a data is “open” if it can be freely used, reused and redistributed by anybody and it is called government data if it is produced or commissioned by government or government controlled entities. This is the definition of Open Government Data (OGD) by (OKF, 2013b).

According to Commercial Exploitation report (2000) governmental organizations are the largest data producers of information in Europe. This makes OGD a valuable source of information. Different organizations and researchers have been suggested a number of models and scheme in order to represent a road map for eGovernment development (Andersen & Henriksen, 2006; West, 2004).

Recently Kalampokis et al. (2011) presented a classification scheme for (OGD) regarding the relevant technological approaches as well as limitations in current initiatives. Four classes are presented to describe all related initiatives. The model’s purpose is to: first, to provide a roadmap for OGD reuse and second, to facilitate assessment of different initiatives. Figure 2-1 represents the technological and organizational complexity that is involved with the possibility of creating added value on the basis of provided data.

![Figure 3-1 Open Government Data Stage Model (Kalampokis et al., 2011, p. 241)](image-url)
The OGD stage model covers different aspects related to OGD. The model first describes different stages and its objectives. Next it provides the technological requirements of each class. Third it addresses the possible benefits related to different classes of the model.

Table 3-4 OGD model summary (Kalampokis et al., 2011)

<table>
<thead>
<tr>
<th>Classes (stages)</th>
<th>Description</th>
<th>Technological perspectives</th>
<th>Benefits</th>
<th>Objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Aggregation of Government Data</td>
<td>Opening up data, publishing data online for re-use aggregating data from different sources.</td>
<td>Publish downloadable files in formats such as CSV, XML, KML Publish data using the linked data without linking to other data sets.</td>
<td>Public gains access to a wealth of valuable data</td>
<td>Provide data online for anyone to reuse.</td>
</tr>
<tr>
<td>2. Integration of Government Data</td>
<td>Government data integration across public administration</td>
<td>Data schemas standardization. Identifiers standardization</td>
<td>Provision of a unified view of government data from different sources. complete and concise integrated government data.</td>
<td>Provide integrated government data across every public agency.</td>
</tr>
<tr>
<td>3. Integration of Gov Data with Non-Gov Formal Data</td>
<td>Integration of government data with the non-government formal data.</td>
<td>Linked Data. Establishment and maintenance of links between government and non-government datasets. Richer metadata should be included</td>
<td>Provision of richer information to data consumers. Complex queries answering.</td>
<td>Connecting relevant government data with source of non-governmental formal data</td>
</tr>
<tr>
<td>4. Integration of Gov Data with Non-Gov Formal and Social Data</td>
<td>Integration of government data with not only non-government formal data but also social data on the Web</td>
<td>Linked data platform. Mechanisms to facilitate government and social data integration.</td>
<td>Innovative services in which government data provide context of interpretation for social data. Allow citizens to deliberate in social media about public administration related real-world</td>
<td>Enables governments to consider citizens’ opinion expressed through social media in governmental decision-making processes</td>
</tr>
</tbody>
</table>

Table 3-4 represent the summary of the OGD stage model and correspondence issues related to each class. According to table the proposed scheme focus on two main technological aspects. Providing accessible OGD on the web for download in well-known format, and publishing OGD as Linked Data through a REST full API, SPARQL search endpoints or RDF dumps. Moreover, from domain specific aspect the two major organizational approaches are identified in the model. A central OGD platform publishes data coming from different organizations (direct provision). In the other approach different organization publish their own data in decentralized way and the OGD platform act as central point to provide linking mechanisms or metadata to identify datasets (indirect provision) (Kalampokis et al., 2013).
Indirect provision of OGD will facilitate information integration around real issues, namely public agencies, laws and services. In this research we focus on the requirements for decentralized provision of OGD for a central platform, which represent datasets from different OGD portals and other dataset resources.

3.2.4 Open data Process

Increasing amounts of unstructured and semi structured data that are made available raises the question about how to bring added value for the society and government by using them. A key condition for this is a clear support of the process of using open data. Considering these steps as a process creates a frame for these actions. Figure 3-2 shows the open data process. As it is visible in the, figure both users and public sector’s contribution is necessary to complete the process.

Data production is the primary step of this process as long as there is no data whole cycle is useless (step 1). Thus data quality improvement is vital requirement for the process cycle. In the next step of the process data is published for users. Nevertheless as mentioned in the text, public sector mainly focus on publishing data and they are less concerned about the data use (step 2). Next step will be done by users by looking after data sets. Users can find their desired data set by using Metadata, which will be provided by data publishers (step 3). Finding the data is not sufficient for the users, they need to use the data in a way that they can gain benefits for their purposes. Tools for data use should be provided in step 2 to enable user to achieve their goals (step 4). The focus of this research is on this step to enable data use for different user groups. The last step of the process focuses on improving the quality of data. Feedback that is provided by the data user can help both government and users (step 5). By participating of government in this iterative cycle possibility of improving the policies or updating data publishing process can be brought (M. Janssen & Zuiderwijk, 2012). For the purpose of this research we will focus on all steps of processes for analysis. The analysis will be done in Chapter 4 by designing scenario for all the steps. This approach provides an opportunity to increase the functionality of user interface of open data platform.
3.3 Open data impediments

Open data can bring several benefits. Nevertheless the open data process is faced with many socio-technical impediments (Dekkers et al., 2006; Zuiderwijk, Janssen, Choenni, et al., 2013).

Major impediments for using open data mentioned in Chavan and Ingwersen (2009) are lack of an infrastructural, technical, socio-cultural, policy-political, and economic supports. Open data policies, can result in impediments for the open data process and numerous barriers is identified related to the policies. The impediments related to the open data policies are out of focus of this research and we mostly focus on technological barriers of open data and mainly those that are related to the user interface of open data platform.

Currently different concerns about the technological impediments like data deposition exist. Lack of unified standard platform and difficulties involved in data retrieval are the most important shortcomings. Some of the platforms require registration and related action to give users permission to access data.

Zuiderwijk, Janssen, Choenni, et al. (2013) Categorized open data impediments in three main groups. Moreover each category is divided into different sub-categories. Data access impediments are concerns regarding creating, opening, finding and obtaining data. Data Use Impediments are about the limitation that restricts the use of open data. And the Data deposition impediments are about the difficulty of data storage and providing feedback on datasets. In the following we focus on each impediment category in detail. Also Zuiderwijk, Janssen, Choenni, et al. (2013) have made wide list of impediments based on the mentioned above three categories. Since the presented lists of impediments, are quite broad we have customized list of impediments in a way that is most relevant to the subject of this research.
Data access Impediments

data access impediments are mostly happen while creating, opening, finding and obtaining the data. (Zuiderwijk, Janssen, et al., 2012). Data access impediments are divided into two subcategories; Availability and Fundability.

1) Availability and access; datasets that are only available as group downloads this brings difficulties for those who are only interested part of the dataset. Since open access is related to publication, one of the primary obstacles to creating open access is transfer of rights to publishers, a concept that has been failed to be accomplished.

2) Find ability; users are not able to find proper data they need, mainly because there are no or few central website to have access to data. Users are confused and there is no support tooling for guiding the users. Sometimes using same dataset, results in contradicting outcomes, which reduces the reliability of public datasets. Furthermore, some websites provide only single datasets, while combined data sets are more valuable most of the time (Geiger & Lucke, 2011).

<table>
<thead>
<tr>
<th>Impediments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 PSI is hard to find due to publication in several resources.</td>
<td>(Vickery &amp; Wunsch-Vincent, 2006)</td>
</tr>
<tr>
<td>2 It’s hard to gain access to appropriate datasets.</td>
<td>(Conradie &amp; Choenni, 2012)</td>
</tr>
<tr>
<td>3 The provided data is incomplete.</td>
<td>(M. Janssen et al., 2012)</td>
</tr>
<tr>
<td>4 Data is temporarily available on the website.</td>
<td>(Veljković et al., 2011)</td>
</tr>
<tr>
<td>5 No access to raw data, processed data is only available</td>
<td>(M. Janssen et al., 2012)</td>
</tr>
<tr>
<td>6 To access data, user are required to take specific actions. For instance registration, giving permission.</td>
<td>(Blakemore &amp; Craglia, 2006; M. Janssen et al., 2012; Meijer &amp; Thaens, 2009; Murray-Rust, 2008)</td>
</tr>
<tr>
<td>7 There is no standard licensing type for re-use of open data.</td>
<td>(M. Janssen et al., 2012; Judge, 2010; Vickery &amp; Wunsch-Vincent, 2006)</td>
</tr>
<tr>
<td>8 data access, is restricted to certain group of users.</td>
<td>(“Junar Open data platform,” 2013)</td>
</tr>
<tr>
<td>9 Poor search possibility due to lack of index.</td>
<td>(M. Janssen et al., 2012)</td>
</tr>
<tr>
<td>10 Various publishers set copyright over data and restrict the re-use without permission.</td>
<td>(Murray-Rust, 2008)</td>
</tr>
<tr>
<td>11 Difficulty to access data due to administrative overhead, terms of access and other barriers</td>
<td>(Charalabidis et al., 2011; Commission, 2011)</td>
</tr>
<tr>
<td>12 Lack of information about the certain data availability.</td>
<td>(Commission, 2011; Zuiderwijk, Janssen, Choenni, et al., 2013)</td>
</tr>
<tr>
<td>13 It’s not clear who are the data owners</td>
<td>(Commission, 2011)</td>
</tr>
<tr>
<td>14 There is no research on real needs of data users.</td>
<td>(Zuiderwijk &amp; Janssen, 2012; Zuiderwijk, Janssen, &amp; Parnia, 2013)</td>
</tr>
</tbody>
</table>
Lack of world-wide open data policies.  
(Schellong & Stepanets, 2011; Zuiderwijk & Janssen, 2012)

Too many information to process, and complexity of finding proper data.  
(M. Janssen et al., 2012)

Data Use Impediments

According to Zuiderwijk, Janssen, Choenni, et al. (2013) data use barriers are divided into six sub-categories: Usability, understandability, quality, linking and combining data, comparability and compatibility, and metadata. First, We discuss each aspect in the following and next we provide list of data use impediments. Data use impediments directly hinders use of open data. Using datasets is only limited for some researchers most of the times and for ordinary user numbers of datasets are limited. Today's Open Data movement motivates governments to open their data based for some fundamental ideas. Government data is paid by the citizens and should therefore be obtainable by the public co-ownership (Kulk & Van Loenen, 2012), having a strong legal framework in order to go toward transparency is needed to be taken into account.

Data users do not have a capabilities and knowledge to extract and use the datasets (understandability). This can happen because there is no description for the usage of datasets. For ordinary users with low statistical knowledge, it is tough to use datasets with unfamiliar data formats. While metadata provides order within datasets (Duval et al., 2002)and make data storage and preservation (King et al., 2011)easier, lack of metadata makes the usage and tracking of data source hard. Another shortcoming is that institutions that provide datasets do not comment on user input so there is no feedback for users.

Quality impediments are mostly regarded in the following domains. The accuracy and reliability of information are not clear for users. Information may appear to be irrelevant based on categorization in websites. Sometimes users get confused because of quantity of the data they need to process and overload of information, which at the end leads to missing the essential data. Last but not least, it is hard to measure the data quality from combined outputs (Smith, 2011).

Table 3-6 Data use impediments(Zuiderwijk, Janssen, Choenni, et al., 2013)

<table>
<thead>
<tr>
<th>Impediments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Difficulties with measurement of data quality</td>
<td>(Smith, 2011)</td>
</tr>
<tr>
<td>2 Open data platforms are diverse and each of them cover different aspect of open data use.</td>
<td>(Zuiderwijk, Janssen, &amp; Parnia, 2013)</td>
</tr>
<tr>
<td>3 To ensure the quality of information, supervision is required before dataset publication.</td>
<td>(M. Janssen et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>Lack of accuracy of information</td>
</tr>
<tr>
<td>---</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>5</td>
<td>Lack of users knowledge about compare, link and re-use data</td>
</tr>
<tr>
<td>6</td>
<td>Linking and combing data by user required much more effort than embedding data in application which, requires good knowledge of users</td>
</tr>
<tr>
<td>7</td>
<td>The available PSI needs to be converted to other formats in order to enable re-use the data and data comparison</td>
</tr>
<tr>
<td>8</td>
<td>No standard policy for data publication</td>
</tr>
<tr>
<td>9</td>
<td>Due to license restriction users may not be able to edit data.</td>
</tr>
<tr>
<td>10</td>
<td>Lack of agreement on URIs matching for resources, which can result in missing dependency between different resource during the integration</td>
</tr>
<tr>
<td>11</td>
<td>Condition of re-use are not clearly defined for use of data</td>
</tr>
<tr>
<td>12</td>
<td>There is out-dated and invalid data, because they are not regularly updated</td>
</tr>
<tr>
<td>13</td>
<td>Lack of accurate and updated data</td>
</tr>
<tr>
<td>14</td>
<td>Data are typically available in heterogeneous format, which requires huge human workload to make it clean and machine readable.</td>
</tr>
<tr>
<td>15</td>
<td>Lack of tooling support and helpdesk</td>
</tr>
<tr>
<td>16</td>
<td>Inactive user might be blocked due to the lack of intensive or added value for other users</td>
</tr>
<tr>
<td>17</td>
<td>Datasets are used as an individual dataset, however combination of various datasets can bring the added value</td>
</tr>
</tbody>
</table>

**Data Deposition Impediments**

Data deposit impediments are related to store, discuss and provide feedback on datasets. In the research of Zuiderwijk, Janssen, Choenni, et al. (2013) data deposit barriers are divided into two categories; Interaction with data provider and Opening and uploading. According to the feedback loop stated by Zuiderwijk, Janssen, et al. (2012) it is obvious that government plays an important role in making data available for all the users. Basically, accessing and storing data is hard due to existence of different data formats. High barriers hamper the effective exploration, management and distribution of the large amounts of available public sector data. Data are provided in different formats so this makes it annoying for users to
deposit and use the datasets. Users should adopt themselves to use different dataset formats and standards (Judge, 2010).

<table>
<thead>
<tr>
<th>Impediments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Legacy systems cause complication in data publication</td>
<td>(M. Janssen et al., 2012)</td>
</tr>
<tr>
<td>2 Users registration are required</td>
<td>(van der Graaf et al., 2011)</td>
</tr>
<tr>
<td>3 Quality of data published by individuals</td>
<td>(M. Janssen et al., 2012)</td>
</tr>
<tr>
<td>4 Open data use requires conversions in public sector organizations</td>
<td>(M. Janssen &amp; Zuiderwijk, 2012)</td>
</tr>
<tr>
<td>5 The main concern of data publishers is just publishing the data rather than real use of data</td>
<td>(M. Janssen &amp; Zuiderwijk, 2012)</td>
</tr>
</tbody>
</table>

To sum up, from the literature review we realized that all current literatures focused on the barriers of open data from the data provider perspective while little attention has been given to the user’s perspective regarding the issue (Geiger & Lucke, 2011). However, current research by (Zuiderwijk, Janssen, Choenni, et al., 2013) provided a different view of analyzing the impediments of Open Data by focusing on user perspective. They have called it as socio technical impediments of open data from users perspective, and they have made a list of ten categories. Namely, 1) the availability and access, 2) find ability, 3) usability, 4) understand ability, 5) quality, 6) linking and combining data, 7) comparability and compatibility, 8) metadata, 9) interaction with data provider, and 10) opening and uploading.

### 3.4 Open data requirements

Several requirements are identified for open data platforms. These requirements are categorized in five sub categories. Namely, Standardization, API, Materialization, Policies and integration. Figure 3-3 represents an overview of the requirements. Standardization enables automatic processing of the data and facilitates automatic access. While Materialization guarantees better quality control of datasets to provide better services for the users. Integration is the concern regarding the combining dataset in a way that the uniformity provided among different datasets. And policies legalize access to the data by different users.(Braunschweig et al., 2012)
A Dataset is the main product of the open data platform. Braunschweig et al. (2012) identified five groups of requirements for a dataset (Figure 3-4). Dataset should be discoverable, without enough metadata, dataset cannot be found and it cannot provide added value for users. And a machine-readable format is severely essential to support the paradigm of Linked Data. Moreover validity, quality and granularity are important features that provide support for more use cases and bring more added values.

Numerous requirements have been identified for open data platforms. Main requirements are categorized in the following groups according to (Zuiderwijk, Janssen, & Jeffery, 2013; Zuiderwijk, Janssen, & Parnia, 2013). This classifies the requirements in twelve main categories. We will use these requirements in next chapter, as an input in the requirements analysis phase.

1. Access
2. Searching
3. Navigation
4. Uploading
5. Downloading
6. Data Quality
7. Analysis of data set
8. Visualization
9. Linking and combining data
10. Collaboration
11. Support and Help
12. Feedback
### Access

**Table 3-8 Open data requirements (access)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Provide a good overview of which data are available for reuse and make them accessible.</td>
</tr>
<tr>
<td>2</td>
<td>Preferably provide data free of charge.</td>
</tr>
<tr>
<td>3</td>
<td>Support the use of different languages for the same dataset.</td>
</tr>
<tr>
<td>4</td>
<td>Provide easy access on a daily basis (e.g. be reliable in terms of accessibility of the</td>
</tr>
<tr>
<td></td>
<td>website, response times, make sure that the webpages are loaded quickly)</td>
</tr>
<tr>
<td>5</td>
<td>Provide recent, less recent, old and historic data.</td>
</tr>
<tr>
<td>6</td>
<td>Provide datasets on different government levels, such as the local, national and</td>
</tr>
<tr>
<td></td>
<td>international level.</td>
</tr>
<tr>
<td>7</td>
<td>Pay attention to interoperability of the infrastructure with other systems.</td>
</tr>
<tr>
<td>8</td>
<td>Provide real-time data</td>
</tr>
<tr>
<td>9</td>
<td>Provide data that are interesting for building applications.</td>
</tr>
<tr>
<td>10</td>
<td>Provide a sustainable platform, so that users get the confidence that they can use the</td>
</tr>
<tr>
<td></td>
<td>platform for a long time.</td>
</tr>
<tr>
<td>11</td>
<td>Provide a large number of datasets.</td>
</tr>
</tbody>
</table>

### Searching

**Table 3-9 Open data requirements (Searching)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>Make sure that the platform can be found easily via Google and other websites is referred</td>
</tr>
<tr>
<td></td>
<td>to on websites of governmental organizations.</td>
</tr>
<tr>
<td>13</td>
<td>Provide good search functionality with advanced search fields.</td>
</tr>
<tr>
<td>14</td>
<td>Make it possible to search for/through metadata.</td>
</tr>
<tr>
<td>15</td>
<td>Provide tags that make search easier.</td>
</tr>
</tbody>
</table>

### Navigation

**Table 3-10 Open data requirements (Navigation)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Make sure that the features for the user interface are very clear and provide clear</td>
</tr>
<tr>
<td></td>
<td>navigation.</td>
</tr>
<tr>
<td>17</td>
<td>Show process directions already on the home page.</td>
</tr>
<tr>
<td>18</td>
<td>Make it possible to click on target groups or target group features on the home page,</td>
</tr>
<tr>
<td></td>
<td>that direct people to parts of the website where they get the support that they need.</td>
</tr>
<tr>
<td>19</td>
<td>Use different interfaces/models for different target groups. Within those target groups,</td>
</tr>
<tr>
<td></td>
<td>make different interfaces/models for different communities.</td>
</tr>
<tr>
<td>20</td>
<td>Create a personal website, to enable user authentication via different ways (e.g. Social</td>
</tr>
<tr>
<td></td>
<td>media).</td>
</tr>
</tbody>
</table>
• **Uploading**

<table>
<thead>
<tr>
<th></th>
<th>Table 3-11 Open data requirements (Uploading)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Allow for uploading data in different formats</td>
</tr>
<tr>
<td>22</td>
<td>Allow for uploading derived datasets (i.e. reused datasets are uploaded again, related to the original dataset)</td>
</tr>
<tr>
<td>23</td>
<td>Provide very clear instructions for uploading datasets.</td>
</tr>
<tr>
<td>24</td>
<td>Provide clear tutorials/videos about the risks and benefits of uploading datasets (e.g. what are the responsibilities of the data provider and user).</td>
</tr>
<tr>
<td>25</td>
<td>Governmental employees are very risk-averse, so pay much attention to data security.</td>
</tr>
<tr>
<td>26</td>
<td>Provide information about which datasets can be published and which cannot be published (e.g., provide a framework that helps identifying privacy sensitivity).</td>
</tr>
</tbody>
</table>

• **Downloading**

<table>
<thead>
<tr>
<th></th>
<th>Table 3-12 Open data requirements (Downloading)</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>Enable downloading data as well as downloading metadata.</td>
</tr>
<tr>
<td>28</td>
<td>Enable downloading data in different formats.</td>
</tr>
</tbody>
</table>

• **Data Quality**

<table>
<thead>
<tr>
<th></th>
<th>Table 3-13 Open data requirements (Data Quality)</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>Make clear to users what the quality of the data is and provide good quality data as much as possible. However, the definition of quality data depends on a person’s background. Certain open data can be of good quality for one purpose, but not for another. For those data that are not of good quality, make it possible to work with messy data.</td>
</tr>
<tr>
<td>30</td>
<td>Develop a good rating system for the data, using literature about information quality.</td>
</tr>
<tr>
<td>31</td>
<td>Make sure that different types of people rate the quality of the dataset (e.g. both providers and users).</td>
</tr>
<tr>
<td>32</td>
<td>Make sure that users can see the distribution of opinions about the quality of the dataset, not just a general average.</td>
</tr>
<tr>
<td>33</td>
<td>Provide considerable metadata, as this should help to assess the quality of the data.</td>
</tr>
</tbody>
</table>

• **Analyzing**

<table>
<thead>
<tr>
<th></th>
<th>Table 3-14 Open data requirements (Analysing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>Provide tools to clean messy datasets.</td>
</tr>
<tr>
<td>35</td>
<td>Provide tools to analyze the data.</td>
</tr>
<tr>
<td>36</td>
<td>Provide information about which analyses are relevant for the kind of dataset offered</td>
</tr>
<tr>
<td>37</td>
<td>Provide help and recommendations for the evaluation of policies and policy developments (e.g. an evaluation framework).</td>
</tr>
<tr>
<td>38</td>
<td>Make it possible to easily obtain information out of statistical analysis (e.g. download in a PDF).</td>
</tr>
</tbody>
</table>
39 Make it possible to convert unstructured data to structured data.
40 Make it possible to generate automatic reports
41 Make it possible to forecast future developments based on historical developments.
42 Assist in analysing policies across boundaries (by linking).
43 Provide information about national, ministerial and local open data policies related to this dataset

• Visualizing

**Table 3-15 Open data requirements (Visualizing)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>Enable visualizing data in tables. Make sure that they can be copied, so that they can be used in research (e.g. academic publications).</td>
</tr>
<tr>
<td>45</td>
<td>Enable visualizing data in maps.</td>
</tr>
<tr>
<td>46</td>
<td>Enable visualizing data in charts.</td>
</tr>
<tr>
<td>47</td>
<td>Provide geo referencing in combination with visualization tools.</td>
</tr>
<tr>
<td>48</td>
<td>Provide the possibility to visualize metadata, not just to visualize the data.</td>
</tr>
</tbody>
</table>

• Linking/Combining data

**Table 3-16 Open data requirements (Linking and combining data)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>Support data integration.</td>
</tr>
<tr>
<td>50</td>
<td>Provide the possibility to link metadata.</td>
</tr>
<tr>
<td>51</td>
<td>Recommend/advise to link with certain other datasets.</td>
</tr>
<tr>
<td>52</td>
<td>Warn if linking two datasets does not make sense.</td>
</tr>
<tr>
<td>53</td>
<td>Use a good URI strategy.</td>
</tr>
<tr>
<td>54</td>
<td>Use identifiers.</td>
</tr>
<tr>
<td>55</td>
<td>Use well-accepted vocabularies.</td>
</tr>
<tr>
<td>56</td>
<td>Use well-accepted thesauri.</td>
</tr>
<tr>
<td>57</td>
<td>Warn about linking when datasets have temporal aspects. Provide advice.</td>
</tr>
<tr>
<td>58</td>
<td>Provide a link with laws for specific law related datasets.</td>
</tr>
<tr>
<td>59</td>
<td>Monitor links between data and make sure that they are still up to date.</td>
</tr>
<tr>
<td>60</td>
<td>Make sure that linking is not just spatial, link to other domains as well.</td>
</tr>
</tbody>
</table>

• Collaboration

**Table 3-17 Open data requirements (Collaboration)**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>61</td>
<td>Develop a (market) place where researchers can collaborate with other people interested in open data, such as other researchers and public servants.</td>
</tr>
<tr>
<td>62</td>
<td>Make it possible to send personal messages to individual other users.</td>
</tr>
<tr>
<td>63</td>
<td>Provide discussion environments for each dataset/organization.</td>
</tr>
</tbody>
</table>
| 64 | Make it possible to write down which data or type of data users need and from which
organization in which country (data requests). Send these needs to public agencies that can provide those data.

65 Public servants should be linked to the platform, so that users are able to ask them questions about the datasets and get other types of help. Support interaction with data providers.

66 Enable researchers to find other researchers via the platform, so that the platform can serve as a place for finding research partners.

67 Enable commercial stakeholders to find other commercial stakeholders via the platform, so that the platform can serve as a place for finding project/business/sales partners.

68 People who can provide help with processing the data should be linked to the platform.

69 Support interaction for getting help.

70 Make use of social media. Provide links with Facebook and Twitter and make it possible to extract data from social media.

71 Provide a market place where different stakeholders can find each other to work on new projects, applications.

72 Provide a market place where stakeholders can offer services (e.g. developers offer making an application) and other stakeholders mention needs.

- Support and Help

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>Provide tutorials and show for each target group how they can use the data.</td>
</tr>
<tr>
<td>74</td>
<td>Make videos showing how the platform can be used for a real scenario for a specific community.</td>
</tr>
<tr>
<td>75</td>
<td>Provide information about licenses for the use of each specific dataset and explain what those licenses mean in practice.</td>
</tr>
<tr>
<td>76</td>
<td>Make version management very clear and understandable. Show who has done what and when to datasets.</td>
</tr>
<tr>
<td>77</td>
<td>Provide updates of datasets and inform users about when updates are published.</td>
</tr>
<tr>
<td>78</td>
<td>Provide the possibility to subscribe to e-mail or other messages when a new/curated version of a dataset has been uploaded.</td>
</tr>
<tr>
<td>79</td>
<td>Refer to other places where publications or other reports based on a certain dataset can be found.</td>
</tr>
<tr>
<td>80</td>
<td>Provide considerable metadata, as this should help to find and interpret the data.</td>
</tr>
<tr>
<td>81</td>
<td>Link the raw data to publications or analysis reports of these raw data.</td>
</tr>
<tr>
<td>82</td>
<td>Provide tutorials, videos and other help about statistical analysis.</td>
</tr>
<tr>
<td>83</td>
<td>Provide help with policy research. E.g. provide a general framework with effect indicators.</td>
</tr>
<tr>
<td>84</td>
<td>Provide help for reporting on / performing statistical analysis about policy effects.</td>
</tr>
<tr>
<td>85</td>
<td>Provide a good judicial framework and legal regulation for the reuse of datasets.</td>
</tr>
<tr>
<td>86</td>
<td>Provide an overview of examples of benefits of publishing data for different organizations.</td>
</tr>
<tr>
<td>87</td>
<td>Provide a tutorial stating for which purpose each dataset can be reused.</td>
</tr>
<tr>
<td>88</td>
<td>Provide examples of use cases for datasets.</td>
</tr>
<tr>
<td>89</td>
<td>Provide examples of business cases for datasets.</td>
</tr>
</tbody>
</table>

Table 3-18 Open data requirements (Support and Help)
Provide a clear governance model: show who does what with the data.

Provide tools to change the format of a dataset.

- Feedback

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 3-19 Open data requirements (Feedback)</strong></td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>Provide tools to analysis what users do with open data. This analysis can be based on people that are logged in and give permission to follow their actions (quantitative feedback).</td>
</tr>
<tr>
<td>93</td>
<td>Make it possible to contact users of the data of the provider and/or receive users’ opinion on the data.</td>
</tr>
<tr>
<td>94</td>
<td>Provide a good rating system for datasets and provide civil servants with the service to get regular updates about the ratings of their datasets.</td>
</tr>
<tr>
<td>95</td>
<td>Monitor who are the reusers of specific government data (link to provider).</td>
</tr>
</tbody>
</table>

3.5 User Interface design

This section discusses the literature found relevant to this research: Web site development process, Requirements engineering process and techniques, and User Interface Design principles. These elements are relevant to this research as they provide answer to third research sub-question. They provide an overview of techniques and process of Requirement elicitation and analysis, and User Interface Design. All of the topics are vital to consider while designing the open data platform.

In Human-Computer Interaction (HCI) field, measurements have generally been usability measures, such as efficiency, effectiveness, and satisfaction learnability, memorability, error prevention, and satisfaction effectiveness, learnability, flexibility, and attitude guess ability, learnability, experienced user performance, system potential, and re-usability. Learnability is regular component that is incorporated in all above measures of usability. When usability evolved to User Experience (UX), the measurements is widened from practical to experiential(Ketola & Roto, 2008).

All Web applications need requirements analysis. To develop usable Web applications that meet users ‘needs, we ought to capture the proper requirements. This process is generally known in software engineering as requirements engineering.

Requirements elicitation focuses not only on determining needs but also on understanding them. This gives the designer the knowledge necessary to establish application requirements and the information with which to perform analysis as decisions arise during the development process.
3.5.1 Web site development process

According to the findings of (Zhou & Stålhan, 2004). The development process in practice is divided into the following tasks:

1. **Requirements Capture and Analysis**: During an initial meeting with the client important issues of the future web-based system are identified: purpose, goals information needs, content and potential users of the web site. Then the web team analyses the business processes, defines the required functionality and chooses the web technology and development tools. Other activities that might be included in this phase are an analysis of the client’s organization brand, and an analysis of competitors’ web sites and marketing research. The deliverable of the phase are project documentation or a development plan with time estimates (Vukovac & Klicek, 2009).

2. **Design Phase**: In this phase, three activities can be distinguished
   - **Conceptual Design**: Includes the design of the information architecture of the application. During this stage, the structure of the database is planned and different diagrams are built.
   - **Navigational Design**: Includes the construction of models with different navigational paths
   - **User Interface Design**: Several templates of user interface are sketched and authored in HTML/CSS. In this stage, different multimedia elements are edited and optimized.

3. **Development or Programming Phase** (Implementation)
4. **Testing and Implementation of Subprojects**: Different types of testing are carried out, as well as the implementation of smaller subprojects. The web-based system is also upgraded on the basis of user evaluation.
5. **Maintenance**: The final phase comprises maintenance of the system and enhancements.

In this thesis we focus two primary phases. First we will capture the requirement. Second analyzing the requirements and finally designing the user interface.

Interface design is part of design phase and in this thesis we design a prototype of UI. A Web interface is a complex mix of text, links, graphic elements, formatting, and other aspects that affect the Web’s usability. According to (Yan & Guo, 2010) three types of design needs to be considered for a good design.

- **Readability Design**

With a few exceptions, people visit the Web for its utility, not its beauty. User chooses clarity over confusion. Three guidelines for better Web writing: skip the jargon, avoid acronyms, and bar sarcasm, subtle word play.
• Search Design.

Search is one of the most important design elements on a website. User wants the search on websites to work like that of their favorite major search engine.

• Page Design

Page design includes two parts, the first is graphic, and the second is HTML coding. Graphic design focuses on visual presentation. Experience design encompasses all two of these categories, as well as properties that affect the overall user experience (download time, ads, popup windows).

### 3.5.2 Requirement Engineering Process and techniques

A requirement is defined as a condition or capability that must be met or fulfilled by a system to satisfy a contract, standard, specification, or other formally imposed documents (IEEE Standard 610.12-1990).

The requirements defined for a system should be: correct, consistent, verifiable and traceable. Requirements engineering is the process of eliciting, understanding, specifying and validating customers’ and users’ requirements. Elicitation and specification of these requirements is a complex process as it is necessary to identify the functionality that the system has to fulfill in order to satisfy the users’ and customers’ needs (Escalona & Koch, 2004).

Deriving requirements follows requirement engineering approach. In this approach a clear distinction between different levels of description should be made. Sommerville (2010) Identified two terms that point out these differences. User requirement and System requirements are terms, which prevent unwilling results in the requirements engineering process. Moreover these requirements will be classified as functional and non-functional requirements. Sommerville (2010) argued, the most effort is spend on high-level business and non-functional requirements.

**Non-functional requirements** are the requirements, which act to limit the solution (e.g. usability requirements, performance requirements, portability requirements, availability requirements, etc.).

**Functional requirements** are capabilities a system must exhibit to be able to solve a problem, the functional requirement can be categorized to the following categories (Escalona & Koch, 2004).

- Data requirements (Conceptual requirements) indicate how information is stored and administrated by an application.
- Interface requirements (Interaction requirements) are the requirements that provide the way that the user interacts with web application.
• Navigational requirements are the requirements that provide the users’ navigation needs through the hyperspace.
• Personalization requirements (customization requirements) are the requirements that a web application has to adapt itself based on the user preferences.
• Transactional requirements (service requirements) are requirements that indicate internal computation of web application without paying attention to interface and interaction aspects.

Sommerville (2010) model of elicitation and analysis process is shown in the figure (2). This is an iterative process with feedback from each activity to other. Each rotation provides better understanding of requirements and finally, results to complete requirements document. Four types of process activities in this process are defined.

1. Requirement discovery is a process to discover the stakeholders’ requirements.
2. Next activity organizes the requirements into coherent clusters.
3. Requirements prioritization and negotiation enables prioritizing requirements according to multiple stakeholders’ requirements.
4. The last activity provides documentation of requirements as an input for next cycle of spiral.

![Figure 3-5 Elicitation and analysis process (Sommerville, 2010)](image)

Number of iteration in each cycle varies due to the structure of organizations or project type. In the last activity of requirements engineering process different validation checks will be done. Consistency check and validity check are the examples to ensure activity will guarantee the validity of requirements. Providing the final requirements document is outcome of these processes. However requirements management is required in order to understand and control the changes during the system usage. In order to identify user interface requirements for open data platform the mentioned process will be done. However following principles of user interface design is a clue to achieve the goal (Sommerville & Kotonya, 1998).

The iterative process of requirements engineering consists of three main activities:

• Requirements elicitation
• Requirements specification
• Requirements validation

Figure 3-6 shows the process of requirements engineering. It is represented as a UML activity diagram and is part of the iterative development life cycle, which in the case of Web applications has the tendency to continue during the whole life of the application.

The process starts with the requirements elicitation. The set of developers collect information from the users and customers. Information can be gathered from different sources, such as documents, legacy applications, interviews, etc. which are used in the preparation of the requirements catalogue. Finally, the requirements validation is performed to find out if there are some inconsistencies, mistakes or undefined requirements. The specification-validation process is iterative and may be executed several times in complex projects.

In the next sections, we briefly describe some classic techniques to elicit, specify and validate requirements. These techniques can be more or less suitable for requirements engineering in the Web environment. It is very difficult to establish precise criteria to select the most suitable techniques.

3.5.2.1 Requirement Elicitation

The capture or the elicitation of requirements is the activity by means of which the development team collects from any available source the functionality the system needs to provide to the future users. The process of requirements elicitation can be complex, mainly if
the problem domain is unknown for the analysts. Thus, a set of techniques have been defined and tested by requirements engineering experts to make this step more efficient and precise.

**Interviewing** is a traditional and frequently applied technique. By means of interviews analysts are able to understand the problem and get information about the objectives of the application to be developed. Basically, the interviewing process covers four steps: the identification of stakeholders for the interview, the preparation of the interview, the interview itself and the documentation of the results in form of an interview protocol.

**JAD (Joint Application Development)** can be regarded as an alternative to interviewing. It is a group technique that requires the participation of all stakeholders of a project, i.e. analysts, designers, users, system administrators and customers. The requirements are captured in a set of sessions over several days. In each session, the high level requirements are analyzed and the problem field and the documentation are established.

**Brainstorming** is also a group meeting technique similar to JAD. It consists of collecting non-evaluated ideas and information of all stakeholders of the project [30].

**Concept Mapping** is a technique by means of which concept maps are built. Concept maps are graphs with vertexes representing concepts and edges representing relationships between these concepts.

**Sketching and Storyboarding** is a technique frequently used by graphical designers in the development of Web applications. It consists of a schematic representation (usually on the paper) of the different user interfaces (sketches).

**Use Case Modeling** is a technique which was developed to define requirements not for capturing them. However, use cases are sometimes used by companies in the communication between developers and customers because they are an easy understandable graphical representation for customers and future users of a software application.

**Questionnaire and Checklist** is a technique that consists of preparing a document with questions for which only short and concrete answers or even with a limited choice of answers (checklist) is possible.

**Terminology Comparison** is a technique that does not resolve the problem of requirements elicitation on its own. Instead, it is a complementary technique used to overcome the communication difficulties, that may arise among developers and users, who do not use the same language.

The mentioned above requirements are the common techniques to capture requirements. However many different techniques has been identified.
3.5.2.2 Requirements Specification

Also for the requirement specification different methods have been identified, but in this part we have focused on the mostly wide used techniques.

**Natural Language** It is an ambiguous technique to define requirements. Requirements are described in natural language without any kind of rules.

**Glossary and Ontology** are used to define the terminology that should be used in every software project where stakeholders with different background work together. Therefore, many methodologies propose the use of a glossary in order to define and maintain the most important and critical concepts related to the application.

**Templates** They are used to describe the objectives and requirements using natural language, but in a structured way. A template is a table whose fields have a predefined structure and are filled in by the development team using the user’s terminology. Templates – also known as patterns – are less ambiguous than descriptions in natural language due to their structure.

**Scenarios** consist of the description of the characteristic of the application by means of a sequence of steps. Scenarios can be represented in different ways: as texts or in a graphical form, e.g. by use cases.

**Use Case Modeling** has been widely accepted as a technique to define requirements although it is also used in requirements eliciting as described in the previous section. However, it has the disadvantage that it is ambiguous when defining complex requirements.

**Formal description** is another important group of techniques that proposes in contrast to natural descriptions the use of formal languages to specify requirements.

**Prototypes** are a valuable tool for providing a context within which users are able to better understand the system they want to build. There is a wide variety of prototypes that range from mock-ups of screen designs to test versions of software products.

3.5.2.3 Requirements Validation

Once requirements are defined, they have to be validated. Through requirements validation the requirements specification is checked to correspond to the user’s needs and the customer’s requirements. Only few approaches provide techniques to validate requirements. Most of them only define some guidelines about how developers and customers should review the requirements specification in order to find inconsistencies and mistakes.

**Review or Walk-through** is a technique which consists in reading and correcting the requirements definition documentation and models. Such a technique only validates the good
interpretation of the information. The verification of documentation inconsistencies and the detection of missing information require more sophisticated methods.

**Audit** consists of a check of the results presented in the review documentation. The results are compared with a checklist predefined at the start of the process. It provides only a partial review of the information and results.

**Traceability Matrix** consists of a comparison of the application objectives with the requirements of the system. A correspondence is established between objectives and how they are covered by each requirement. This way, inconsistencies and non-covered objectives will be detected.

**Prototyping for validation** is a technique that consists in building tools based on the requirements specification, i.e. the developers’ interpretation of the systems requirements. These prototypes usually only implement a partial set of functional requirements but provide a global vision of the user interface.

### 3.5.3 Scenarios and Use cases for requirement elicitation and analysis

System analysis and design involves better understanding of users and how and why they perform a particular activity. One way to do this is clearly observing the significant user activities and documenting them. The mentioned above description is often called scenarios (Carroll, 2000). Scenarios explain interactions between system and users. Scenarios are defined as sequence of actions and events (Some, 2005). They are stories that contains situation state in which one or more actors with personal motivations and knowledge, and different tools and objects are involved in. The scenarios represent sequence of actions and events that end up to a result. These actions and events are correlated in a way that system is used in a specific episode (Rosson & Carroll, 2009).

Scenarios can be used in two steps of the development of a system. First they can be used during the requirements activity to provide feedback to both users and analysts, to reflect whether use cases are accurate reflections of user’s needs. Second they can be used during the testing activity to evaluate whether the system reflects the requirements (Kulak & Guiney, 2004).

“A use case represents a series of interactions between an outside entity and the system”(Kulak & Guiney, 2004, p. 35). They are deployed to represent the interactions between the system and the external entities of the system. External entities are actors, other information systems or external events. Since scenarios are not highly detailed they cannot reflect the whole story. Therefore we have developed different use cases to cover various aspects of the system in order to represent detailed interactions.

“Scenarios are instances of use cases that effectively test one specific path through a use case” (Kulak & Guiney, 2004, p. 49). Scenario-based design is a set of techniques that describes the
use of future system in detail at an early point in development process. The narrative description of scenarios, guides the development of the system. It shifts the focus of design work from defining system operations to describing the way user will use a system to perform tasks and other activities (Rosson & Carroll, 2009).

3.5.4 User Interface Design Guideline

We have reviewed the whole design process and now we need to focus more on the rules and principles for the design part. There are different models and guidelines for designing user interface but in general there are some guidelines that are discussed (Heim, 2008; O.Galitz, 2007) and need to be considered while designing a good user interface. We review detail of each factor in the following. This is deployed while designing UI mock-ups.

- Accessibility comes from the idea that a system should be available for people with disabilities afterwards it was expanded to cover user’s needs. Four main characteristics of accessible design are; perceptibility, operability, simplicity and forgiveness.
- Aesthetically pleasing or visually pleasing represent that in a UI design require a good graphical design because most of the human-computers interactions happens visually. Also there should be a contrast between different elements.
- Clarity; in the first looks UI should be look simple and understandable for users. It should be visually, conceptually and linguistically clear and includes visual elements, functions, metaphors and words.
- Compatibility; it should be compatibility with the user, task and job and product.
  - Design must be compatible with user needs and know the fact that users do not look alike and they have different needs.
  - Users should not force to navigate between many screens to complete a routine job and the task should be matched with specific job they willing to do.
  - Product compatibility is taking advantage of old system in order to increase the user understandings from the system.
- Configurability; easy personalization and customization leads to higher satisfaction, people have different reaction to configuration some accept main changes and some others try to use the first things they have learned and worked with
- Consistency; a system should be act, look and operate in a same way for similar options. System components should have similar look, similar user and operate similar. Consistency in design means uniformity in placement, appearance and behaviour. In consistency in design usually is caused by differences in people (designers) also it can occur do to the time limit.
  Developing and applying standards or guidelines can achieve design consistency. This guideline provides detailed information that is required for design. World Wide Web is consortium that is responsible for web interface standards.
- Control; users wants to control system this feeling can be presented to user when the system respond to users action. User has bad feeling when system is not acting as they wish. UI can increase the feel of control. This can be achieved when a user is able to how to do an action and what do with the system.
Some system actions can influence on control feeling for instance long delays in responses, surprising system actions and difficulties in obtaining necessary information and disability to achieve desired results.

- **Directness;** there should be a direct way to do a task, for this alternatives should be visible for the user. A task should be done by selecting an object, selecting an action and observing the action to be done and at end seeing the result. A UI according to user’s skills, habits and experience will provide better feeling for its users.

- **Efficiency;** anticipating user needs can increase the efficiency of the platform. Providing complete information during the process while doing an action can guide user to complete his/her desired action.

- **Familiarity;** using a known concept and language beside real-world and natural metaphors can make the interface more familiar to the users.

- **Flexibility;** a system should be sensitive and designed according to users knowledge, experience, personal preferences, habits and all around infective factors. People should be able to interact with the system according to their needs. Flexibility and usability are against each other usually a flexible system cause user to make errors and news user may not be able to get to use it, therefore flexible UI is recommended for well-experienced users.

  Also based on differences in personalities, some people prefer to use system as they first learnt. Thus the first view of interface can be very effective on users and can increase the loyalty of the users to the new platform.

- **Forgiveness;** errors usually happens by human and system should be able to cope with unwanted errors and be able to decrease errors by anticipating where mistakes can occur. Beside it should be able to prevent users from making errors and lead them to the right direction. Users should be able to review, change and undo actions whenever it’s required. In case of errors system should be able to lead user to error amendment.

- **Obviousness;** a system should be able to easily understandable and quick to be learnt. User should know what can he/she do, why to do and how to do a specific task. To achieve this goal a system should provide overview of steps to complete a task. Keep in mind that long descriptions will make system ambiguous and it should be avoided.

- **Operability;** the system should be usable by all the users without paying attention to their physical abilities; also supporting assistive technologies increases the system operability.

- **Predictability;** task and movements through a system should be anticipatable based on user knowledge and experience. Predictability decreases the mistakes and enables users to finish task more quickly. To achieve this goal user should know about the next steps while they are in a process. A system should avoid combining the tasks. Instead making an independent task is recommended. User should be able to combine the actions whenever they want.

- **Recovery;** system should support reversible actions. It should be able to reverse an action by supporting undo command. Recovery should be done automatically and easily. A system should ensure that users do not lose their works because of their own mistakes or due to software/hardware communication problems.

- **Responsiveness;** system should be rapidly responding to user’s requests. Users need to get feedback from the action they have performed and especially expert users need to get detailed version of feedback.
System should avoid displaying a blank page for a long period. In case of a long processing time a “in a progress” message should be displayed. To support all types of users the acknowledgment should be provided visually and auditorily.

- Safety; system should provide possibilities for user to be protected from mistakes. It should not rely on the human’s memory and should always support users in case of forgetting information by a user.
- Simplicity; a simple system is the one that understandable by all kind of users. To have simple system providing defaults, uniformity, consistency and elimination of unnecessary elements is recommended.
- Transparency; it is better to enable users to concentrate on their job, without any concerns about the technical parts of the interface. The working mechanism of UI should be invisible for the users.
- Visibility; a system can be considered as a visible system when it indicates their capabilities of what action they can perform, their status and the result of action after it have been done.

The guidelines mentioned above are resulted of book review, and they will guide us during designing the UI mock-ups for Open Data platform.

### 3.6 Chapter Summary

This chapter reviews the essential background information related to this research. It provides the answer to first two, research sub-questions. Conducting a literature research provides contents and discussions in this chapter. Literature review chapter consists of four parts: Open data, Open data Impediments, Open data Requirements, and User interface Design. In this chapter we gain knowledge about design of UI, besides we have provided lists of requirements for UI of Open Data platforms. This two is deployed to design UI for Open Data platform in next chapter.

Discussion in the first section begins with a brief description of Open Data. This chapter provides an overview of the different definitions of the Open Data. Next, we have reviewed related topics to Open Data paradigm such as; Linked Data, Meta data, Open Government Data Model, and Open Data Process. These topics are relevant to this research as they describe the core of the technological advancement of open data platforms. All of the topics are necessary to consider while designing the open data platform. However, Open Data process is vital to consider since we will use it in the next chapter as an input for requirements elicitation and analysis.

Second section discusses the current impediments of Open Data. Major impediments for using open data identified as; lack of an infrastructural, technical, socio-cultural, policy-political, and economic supports. In this research, we have mostly focused on technological barriers of open data and mainly those that are related to the user interface of open data.
platform. Next, we have categorized Technological impediments of open data platform into three groups; data access, data use, and data deposit impediments.

Third section of this chapter gathers information regarding Open Data platform requirements. It has identified 12 groups of requirements for Open Data platform, and each category consists of the list of requirements. The requirements provided in this section are an input for the design of UI of Open Data platform. We will input the requirements in the next chapter for requirements elicitation process. We will develop six scenarios in the next chapter to address the requirements.

The last section of the chapter reviews the topics that are vital for User Interface design as well as requirements engineering process. In order to give a general picture of UI design, this section has provided a general overview of Web site development process, and described requirements engineering process, and reviewed requirements elicitation and analysis process. Moreover, this chapter provides techniques related to each step of requirements engineering process. Since, in the next chapter we will deploy scenario and use case techniques for requirements analysis and elicitation. Finally, we have a review a brief summary of guidelines of designing a good User Interface.

The discussion in the chapter draws on the lessons learned during the short period that Open Data is begun. Also, introduces factors concerning technical impediments and UI requirements of Open Data platforms. Besides, discussion regarding the UI design process and guidelines is reviewed.
4 Analysis & Design of user interface

4.1 Introduction

This Chapter consists of three main parts. First section is about elicitation and analysis of requirements and next design of the user interface based on the analysis. Finally, providing the analysis result. This Chapter will address the third sub question of this research that is “How can user interface increase the use of open data? ”.

To achieve the objective several studies is conducted regarding the open data requirements and impediments. The result is designing different models and mockups, which will be used in the evaluation part to answer the research questions. At the last stage we evaluate the user interface. Figure 4-1 represents the methodologies that is used in the previous chapters and that will be used in this chapter.

4.2 Requirement Analysis

In this section we will use scenarios and use cases are used to analyse the requirements in order to evaluate the current user interface of open data platform. We develop scenario to increase the platform’s usability. This is done by conducting literature review about the requirements of the platform and performing different experiments (interaction) with the platform regarding the different function of the system. Literature review guides us through the shortcomings of the open data platform to consider a scenario which can contains impediment of current open data platforms.
Requirements are better communicated if they illustrated by scenarios as well as presented as an organized list. Scenarios are a narrative version of specifying the requirements. They help to provide a more detailed list of requirements (Benyon et al., 2005).

In order to have a better understanding of the system and to focus on the shortcomings of the system we need to provide a façade use cases of the open data platform. “Façade use cases show the basic, essential, high-level interaction that the application must support” (Constantine & Lockwood, 1999, p. 3). A façade use case outlines the scope of proposed system. In order to develop a façade use cases we have studied the platform’s documentation to provide overall view of the system. Thus, we will provide general description and objective of the system in the following.

4.2.1 ENGAGE open data platform

ENGAGE is an open data platform that aims to enhance open data use. In comparison with other open data platform ENGAGE mostly focus on the 5th step (feedback) of open data process. It enables all the citizens to become data user as well as data provider. Since ENGAGE is unique in this way among other platforms, and open data platforms can provide large number of benefits it requires improvement to create more advantages.

In order to improve the ENGAGE open data platform we focused on improving the user interface. Therefore, we use ENGAGE platform documentations as an input for the requirements analysis. This helps us to have a better understanding of the platform and create a rich user interface. Designing an integrated user interface improves the whole open data process and provides more benefits for its users.

4.2.1.1 Major characteristics and objectives of the platform

The main goal of the open data platform is “the development and use of a data infrastructure, incorporating distributed and diverse public sector information (PSI) resources, capable of supporting scientific collaboration and research, particularly for the Social Science and Humanities (SSH) scientific communities, while also empowering the deployment of open governmental data towards citizens” (Charalabidis et al., 2011; "Engage open data platform," 2013). This platform will enable users to accessing, downloading, processing and visualizing data for researchers, citizens, politicians and business people.

Key objectives of the platform are: creating open services, engage citizens, explore synergies between public sector and data users, contribute to the evolving standardization and empower researchers. Several target groups are identified as the stakeholders of the project, namely researchers, citizens, developers, journalists, civil servants, companies and archivists. The main focus is on researchers and citizens.
4.2.1.2 Involved Users and responsibilities

The open data platform is aimed to use by different type of users. The main stakeholders of the systems are identified as follows:

- A public sector organization (PSI) aims to make datasets available in public, either by storing the dataset in the open data platform, or linking external datasets that have been published in other open data platform.

- Citizens who want access to government information or wish to build their own innovative applications, visualizations and mash-ups based on open government data. In this regard, the platform will act as a central portal for the discovery and retrieval of public sector information. Furthermore, the platform gives citizens the opportunity to provide feedback to public sector organizations on the quality and usefulness of published datasets, as well as the ability to make suggestions for new datasets or mash-ups.

- Researchers who use government information for research and wish to share their own curated versions of government datasets. They have different needs and requirements. The platform can provide them vast variety of benefits.

- The open data platform staff whom are responsible for administrative tasks and the moderation, curation and enrichment of the submitted datasets. Furthermore, they provide documentation and assistance on tools and methods for public sector data re-use.

4.3 Scenarios

Scenarios are designed according to various user groups (persona) that are involved in the process of opening data. The main stakeholders that are identified as main actor of the platform are public servant, researchers and citizens. As described in section 3.2.4 the open data process consists of five steps, including the process of opening data. Two primary steps related to public sector are data production and Data publish. Other three steps regarding the users are looking for data, using data and providing feedback. We design a scenario in a way that it covers all steps of open data process since in this way we can make sure that all the requirements are covered. Figure 4-2 represent the connection between the scenarios that cover each step of open data process.
4.3.1 Scenario 1: Look for datasets, analyse, save and export the results

First scenario covers the ‘look for governmental data’ step of open data process. It describes a situation when a user needs to find a dataset. System enables search option for the user to help him/her to find the dataset. In this scenario the following requirements from section 3.3 is covered (1-11, 12-17, 28, 34-43, 44-48)

4.3.1.1 Scenario Description

A Researcher who is working in a psychology department wants to study the effect of using drugs on violation rate in different countries (cities). He wants to find violation in different cities as well as drug use rate in those countries.

To search the related datasets he uses the search function of the platform. He first input text ‘drog’ in the search field. Right after he finds that he inputs wrong phrase, thereafter he modifies the input text from modify search pane to ‘drug’. Second he finds about 400 results related to the query. Therefore he needs to narrow down the result into the domain of health. Since he is looking for the rate in different regions he can limit the result to different target countries. Besides he needs datasets for the period between 2000 and 2010, in result he limits
datasets to the mentioned above dates. Also he is looking for specific data formats therefore he restricts the results to XLS format. In this step he has 50 datasets, which he needs to sort them based on the popularity of datasets. He selects a dataset from results to obtain more details about it.

In the next steps he needs statistics from the results to present in his report. Thus he selects ‘analyze result’ tab to visualize results based on metadata (e.g. table and charts). Afterwards he exports the final result in PDF format for the documentation. Moreover he needs to focus on the violation rates therefore he save the search results in his user profile to be able to access it next phase of his research. At the end he performs the same steps to search for the violation rate. Also he access to a pane that represents the last viewed datasets.

Note that the researcher has been tried other search function of the platform (browsing categories and search using map) to query datasets, but he find the text based query the most accurate.
4.3.1.2 Use Case diagram

![Use case diagram of scenario 1](image)

Figure 4-3 Use case diagram of scenario 1

4.3.1.3 Use cases

We describe detail of main use cases in the following tables. Use cases are presented in the following format to provide flow in a same format for all the use cases. Basic flow describes the main flow of the use cases.
<table>
<thead>
<tr>
<th>Use Case 1.1</th>
<th>Search Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>The platform enables number of options to enable to search for a dataset.</td>
</tr>
</tbody>
</table>
| **Used by** | • Use case 1.2  
• Use case 1.3  
• User case 1.4 |
| **Actors** | A researcher |
| **Triggers** | A user who is looking for a dataset |
| **Preconditions** | The platform has provided sufficient amount of datasets. |
| **Basic flow** | 1. The researcher enters a keyword in search bar and selects search tab.  
2. Results are presented in an ordered list. Also a pane represents the facets that allow narrow down the results and modifying the search.  
3. The researcher navigates among different result pages.  
4. The system provides filtering functionality in order to narrow down the search result. The following field are provided to select and customize the search:  
   1. Category  
   2. Region  
   3. Related period  
   4. Resource Format  
   5. License  
   6. Publisher  
   7. Tags  
5. The researcher selects the desired category and can customize the search result by selecting the following options:  
   1. Limiting to the selected categories  
   2. Excluding the selected categories  
6. Results are presented in a customized list.  
7. The results can be sorted according to following criteria:  
   1. Relevance  
   2. Popularity  
   3. Title  
   4. Last update  
8. The researcher selects a sort criterion.  
9. Results are presented in customized order. |
| **Exceptions** | 1. The researchers insert a new text in the search bar.  
2. The researcher selects the modify tab.  
3. System returns to state 3 in basic flow. |
<p>| <strong>Post conditions</strong> | If a user views a dataset page, the system uses the information to show as previous viewed dataset. |
| <strong>Additional requirement</strong> | N/A |</p>
<table>
<thead>
<tr>
<th>Use Case 1.2</th>
<th>Output Search results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>When a user finds number of datasets. We predict she/he will need to export all/some of the result.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Actors</strong></td>
<td>A researcher</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>To create documentation of search result.</td>
</tr>
<tr>
<td><strong>Preconditions</strong></td>
<td>At least one result should be provided by the platform.</td>
</tr>
</tbody>
</table>
| **Basic flow** | 1. A researcher views customized list of datasets.  
2. The researcher changes number of desired result to be shown in one page. (15, 25, 50, 100).  
3. The researcher selects one/all of the result/s (maximum 100).  
4. The researcher chooses export option.  
5. The system responds by navigating to output result page.  
6. The system represents, the three following options as ‘output type’ of the search result:  
   a. To export the results, system represents two drop-down menu regarding ‘export format’ and ‘output’ to select.  
      1. Results can be exported in two formats:  
         1. PDF  
         2. TEXT  
      2. Results outputs are in three kind of detail  
         1. Dataset subject only  
         2. Basic information (Abstract format)  
         3. Detailed metadata  
         4. Complete format  
   b. To Print the search results directly, system represents a drop-down menu to select ‘output’. The output types are same as 5.1.2.  
   c. To Email the search results directly, system represent a drop-down menu to select ‘output type’ same as 5.1.2 and the following fields:  
      1. To (Email address of receiver)  
      2. Carbon copy receiver (C.C)  
      3. Subject  
      4. Body text  
      5. The email address of sender (your email address)  
      6. Email format (HTML or TEXT). |
| | 7. The researchers select the preferred output type.  
| | 8. The researcher fills in the required fields based on the output type.  
| | 10. The system provides the download link or Print |
preference, or sends the email.
11. The researcher downloads the file or prints the results.

Exceptions
3. The researcher doesn’t select any result.
4. The researchers select export option.
5. The system shows an error message ‘select one or more search result and then click export’.
6. The researcher confirms the error message.
7. The system returns to step 3 of basic flow.

Post conditions
The researcher navigates to search result page.

Additional requirement
N/A

<table>
<thead>
<tr>
<th>Use Case 1.3</th>
<th>Analyze search results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>The platform provides options of analyzing search results based on metadata.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Actors</strong></td>
<td>A researcher</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>To have a quick overview of search results.</td>
</tr>
<tr>
<td><strong>Preconditions</strong></td>
<td>At least two results should be provided by the system.</td>
</tr>
</tbody>
</table>
| **Basic flow** | 1. A researcher views customized list of datasets.  
2. The researcher selects analyze search results tab.  
3. The system navigates to analyze search result page.  
4. Results can be represented in charts based on the number of documents in following criteria:  
   1. Year  
   2. Categories  
   3. Maintainer  
   4. Publisher  
   5. Country  
   6. Dataset format  
   7. License  
5. The researcher selects a criterion.  
6. The system represents one of the following charts based on the mentioned above criteria:  
   1. Line chart  
   2. Pie chart  
   3. Column chart  
   4. Column chart  
   5. Column chart  
   6. Pie chart  
   7. Pie chart  
7. The system represents, the three following options to output the results analysis:  
   1. Export  
   2. Print |
3. Email
7.1 To export the results, system provides download link to upload the CSV format.
7.2 To print the results, system provides the print preferences.
7.3 To Email the results, system navigates to the ‘email page’ and represents a field to insert the email address. (a CSV file is attached to the Email).
8. The researchers selects the preferred output type
9. The system provides the download link or prints preferences, or sends the email.
9.1 the system informs the researcher that the selected action is performed.
10. The researcher downloads or prints, analysis result.

Exceptions
5. The researcher selects different criteria

Post conditions
The researcher navigates to result analysis page.

Additional requirement
N/A

Table 4-4 Use case 1.4 description

<table>
<thead>
<tr>
<th>Use Case 1.4</th>
<th>Save search results in user account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>When a researcher needs to access the search results next time the platform enables users to save the search results in user profile.</td>
</tr>
<tr>
<td>Used by</td>
<td>• Use case 4.1</td>
</tr>
<tr>
<td>Actors</td>
<td>A researcher</td>
</tr>
<tr>
<td>Triggers</td>
<td>Access to modified search results if needed in next interactions with platform.</td>
</tr>
</tbody>
</table>
| Preconditions| • At least one result should be provided by the platform.  
• Activation of this option requires user to login to user. |
| Basic flow   | 1. A researcher views the customized list of datasets.  
2. The researcher selects the ‘save results’.  
3. The system save the results in the user profile  
4. The system notifies the user that the search result has been saved in his/her user profile.  
5. The researcher selects ‘saved searches’ link  
6. The system navigates to ‘saved searches’ page.  
7. The system represents list of saved searches and the following information regarding each search:  
   1. Title of search (keywords input by user)  
   2. Last run date of query  
   3. Results number  
   4. View search result  
8. System represents the selecting the results and deleting them. |
9. The researcher selects a search to view it.
10. The system navigates to selected search option.

Exceptions
9. The researcher selects all data sets and selects ‘delete’
10. The system asks from the user to confirm the selected action.
11. The researcher confirms to delete the search results.
12. The system represents ‘no search result saved’.

Post conditions
Search result is saved in the user profile.

Additional requirement
N/A

<table>
<thead>
<tr>
<th>Use Case 1.5</th>
<th>Preview last viewed datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>The platform enables user to have a preview of the previous viewed dataset.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>A researcher</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>Navigating to dataset page</td>
</tr>
<tr>
<td><strong>Preconditions</strong></td>
<td>The user should have been navigated to a dataset page to be able to see the preview list.</td>
</tr>
</tbody>
</table>
| **Basic flow** | 1. A researcher views the search result.  
2. The system provides empty list of ‘viewed dataset’.  
3. The researcher selects a dataset.  
4. The system navigates to a dataset page.  
5. The system provides the list of ‘viewed dataset’, which contains the following details about the dataset that is just viewed:  
   1. Title  
   2. Quality ranking  
   3. Publish date  
   4. Remove dataset tab from list  
6. The system presents five previous viewed dataset and possibility to remove the entire list.  
7. The researcher select a dataset from ‘viewed dataset’  
8. The system navigates to the ‘dataset page’  
9. The researcher selects to return search result page.  
10. The system navigates to the search result page.  
11. The list of ‘viewed dataset’ is displayed. |
| **Exceptions** | 9. The researcher select a dataset from ‘viewed dataset’  
10. The system navigates to the ‘dataset page’ |
| **Post conditions** | The list is displayed in any ‘dataset page’ and ‘search result page’ as long as the user does not delete it. |
| **Additional requirement** | The list is presented based on the IP address of the user. The items in the list will be deleted by change of IP. |
4.3.1.4 BPMN Diagrams

- Use case 1.1: BPMN diagram represents the process flow of the use case 1.1.

![Figure 4-4 BPMN diagram: Task flow of use case 1.1](image)

- Use case 1.2: BPMN diagram represents the process flow of the use case 1.2.

![Figure 4-5 BPMN diagram: Task flow of use case 1.2](image)
• Use case 1.3: BPMN diagram represents the process flow of the use case 1.3. There is a repetitive step in use case 1.3, which is similar to use case 1.2 therefore the output the analysis is not represented in detail. To find out the process step we refer to the BPMN diagram of use case 1.2.

Figure 4-6 BPMN diagram: Task flow of use case 1.3

• Use case 1.4: BPMN diagram represents the process flow of the use case 1.4.

Figure 4-7 BPMN diagram: Task flow of use case 1.4
• Use case 1.5: BPMN diagram represents the process flow of the use case 1.5.

![BPMN diagram](image)

Figure 4-8 BPMN diagram: Task flow of use case 1.5

### 4.3.1.5 User interface design

We have developed five use cases According to the scenario 1 and we have the BPMN diagram for each use case to reflect the flow in each use case. The use cases and BPMN diagram are used in this step to design the user interface. We design wireframes to present UI. Wireframes are used to represent the page layout and arrangement of website’s content.

According to use cases and BPMN diagram we identified four main pages to cover the functions described in the scenario. Figure 4-9 is the wireframe of the ‘search result’ page. in represents a page when a user enters a text to search dataset.
Figure 4-9 Wireframe Search result page

- Output type select:

Figure 4-10 Wireframe Output type
• Analyze the search results

![Wireframe search result analysis](image1)

Figure 4-11 Wireframe search result analysis

• Saved search page

![Wireframe Saved searches](image2)

Figure 4-12 Wireframe Saved searches
4.3.2 Scenario 2: Search, submit and reply dataset request

In this scenario the following requirements from section 3.3 is covered (29-32, 61, 69, 71, 72, 94).

4.3.2.1 Scenario Description

An investor is looking forward to start up new company in her home town. She wants to open a new restaurant based on her townsman preference. She found a dataset from the city’s open data website. This dataset contains information about different type of restaurant in the city, which she is not sure about the accuracy of dataset. Besides she needs a dataset about the food preferences of her townsman. She has searched in different portals unfortunately she did not find any information regarding that issue. She has heard about the dataset request functionality of the open data platform from her colleagues.

Thus she enters website and selects data request tab. She gets an overview of latest requests, a search pane which enables searching through data sets, and a pane that enables to submit a dataset request. But to do this she needs to sign up. In the first steps she browses into the previous dataset requests by importing a text query to search in the title of the requests. Next she narrows down the result by specifying the request type (e.g. Looking for dataset, data cleansing, data enrichment etc.). Also she limits the results to the specific category (e.g. crime, health, education etc.). Afterwards she sorts the results based on the requests that has been replied. She selects three search results but, the platform doesn't provide what she is looking for.

Therefore she decided to submit a ‘dataset request’. To do this she clicks on the place a new request tab. Next she inputs request title as “number and type of restaurant in city A”. When she imports the request title the platform recommends set of requests, which have similar subject. This is done to prevent similar requests. Next she selects ‘data verification’ as a request type and inserts ‘restaurant’ and ‘food’ as keywords of this request, next she add short description about his request. In next step she needs to provide the link to the dataset or upload the file for verification. After importing the URL of dataset she selects the ‘Send me new responses to my posts via email.’ option and finally submits the request.

Also she is looking for datasets about food preference in city A. she performs the same steps above request but this time she selects ‘dataset request’ as a request type. When she imports the title, no similar request is recommended by system to her. She inserts ‘food preference’ and ‘city A’ as keywords of her request. Since the request type is different from the other request, there is no other step to finalize the request. She press the submit button as well as selecting the option of receiving email for new responses.

After few hours she receives a new email indicating that somebody has respond to the request. She clicks on the link that is included in the email and directs her to the request page. A researcher who is working in food industry replies the request and he provides a link to her
dataset request. This dataset is uploaded to the open data platform by the researcher. While she is looking through the datasets another user respond to her request with a new link about the food preference in city A. she finds the second respond more accurate and relevant. Therefore she accepts the second respond as the best answer. Also she rates the both respond as a useful answer by clicking on ‘plus’ sign.

She has received five answers regarding the request about data verification. Two of them are question about the request, which she needs to respond them. The platform hast the capability of commenting on each individual responds to separate the dialogue between users. Thus she comments on each respond and provides answers to their questions individually. After a long dialogue with a user B, she gets a respond with a link that verifies the datasets she was wondering about. Therefore she selects the respond by user ‘B’ as the best answer and rate it as a useful respond. Since other respond doesn’t satisfy what she was looking for she mark them with minus which represent that the answer was not useful.

A journalist who is working in a magazine wants to write an article about the relation of the food and weight of the people living in city ‘A’. He found the researchers request related to his work. Therefore he selects the request and finds out that there are five replies to her request. He wants the modified version of the dataset therefore she post a comment about the reply that is marked by the researcher as the best answer. He gets some answers from the respondents under his comments and rate them with ‘plus’ as a useful reply. Unfortunately the replies do not provide a suitable answer to his question. Suddenly he find a list in the right side of the page which is recommending related requests to the subject he is looking for. He finds that one of them is similar to what he is looking for. He navigates to the page containing the related request. Even the related request doesn’t provide the dataset for the journalist. He finds a link indicating ‘submit your own request’ and navigates through submitting the new dataset request and performs the same steps as the investor did.

4.3.2.2 Use case diagram
Figure 4-13 Use case diagram of scenario 2
### 4.3.2.3 Use cases

**Table 4-6 Use case 2.1 description**

<table>
<thead>
<tr>
<th><strong>Use Case 2.1</strong></th>
<th><strong>Search for a request</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>The platform enables users to search in the submitted request when looks for specific type of dataset that is not available. Also the submitted request may be asked to perform a task on the dataset.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td><strong>Actors</strong></td>
</tr>
<tr>
<td></td>
<td>An investor</td>
</tr>
</tbody>
</table>
| **Triggers**     | • When a user can’t find a dataset  
                    • When a user need to perform a task on the dataset, which requires other user participation. |
| **Preconditions** | • At least one request should be submitted in the platform |
| **Basic flow**   | 1. An investor enters a keyword in search bar and selects search tab.  
                    2. Results are presented in an ordered list. Also a pane represents the facets that allow narrow down the results and modifying the search.  
                    3. The investor navigates among different result pages.  
                    4. The investor uses the facet search functionality. The following fields are represented to narrow down and customized the search.  
                       1. Request type  
                       2. Category  
                       3. Tag  
                    5. The investor selects a field.  
                    6. The system shows the customized search.  
                    7. The results can be sorted according to following criteria:  
                       1. Relevance  
                       2. Popularity  
                       3. Title  
                       4. Last submit  
                    8. Results are presented in customized order.  
                    9. The researcher selects a sort criterion.  
                   10. Results are presented in customized order. |
| **Exceptions**   | 3. Based on request type next step is provided to enable uploading link or dataset. |
| **Post conditions** | N/A |
| **Additional requirement** | N/A |
## Use Case 2.2: Submit a request

### Description
Users who have user account can submit a request for a dataset. This enables user interactions to answer other member’s request.

### Used by
- Use case 2.3
- Use case 2.4

### Actors
An Investor

### Triggers
When a user can’t find a dataset in dataset list.

### Preconditions
N/A

### Basic flow
1. The investor navigates to submit dataset request page (requires user login).
2. System displays following fields to submit a dataset request.
   1. **Title.**
   2. **Request type**
      - a) Dataset request
      - b) Dataset verification
      - c) Data cleansing
      - d) Data enrichment
      - e) Metadata enrichment
      - f) Combination of datasets
      - g) Snapshots of real time data
      - h) Conversion to other formats
   3. **Request category**
      - a) Health
      - b) Geography
      - c) Crime
      - d) Financial
      - e) Education
      - f) Arts
      - g) Law
      - h) Agriculture
   4. **Tags**
   5. **Description.**
   6. **URL of dataset**
   7. **Notification option**
3. The investor inserts input as the title.
4. System recommends number of related requests based on the input for ‘title’.
5. The investor fills out the other fields and submits her request.
6. System navigates to preview dataset request page and shows the detail of request and provides the following options for the user:
   - a) Delete
   - b) Edit
   - c) Publish
7. The Investor Selects ‘Publish’.
8. System navigates to dataset request page and shows the detail of request and provides the following options:
   a) Delete
   b) Edit
9. The investor views the request page.

Exceptions
7. The investor selects ‘Edit’.
8. System navigates to submit dataset request page (1).
7. The investor selects ‘Delete’
8. System ask user to confirm the action.
9. User confirms to delete the submitted request.
10. System notifies user that the action is performed.
11. System navigates to the dataset request list page.

Post conditions
N/A

Additional requirement
Number of request categories should be designed in a way that it covers all requests.

<table>
<thead>
<tr>
<th>Use Case 2.3</th>
<th>Handle a request</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>When a user submits a request and receives answer from other users, system provides options to handle various replies.</td>
</tr>
<tr>
<td>Used by</td>
<td>An Investor</td>
</tr>
<tr>
<td>Triggers</td>
<td>When a user replies to a request.</td>
</tr>
<tr>
<td>Preconditions</td>
<td>At least one request should be submitted.</td>
</tr>
</tbody>
</table>
| Basic flow   | 1. The investor is informed that a user replied the request.  
2. The investor views the dataset request page. 
3. The investor views a reply from a user. 
4. System provides the following options to deal with a reply  
   1. Accept the answer as the best answer 
   2. Rate the answer as useful respond (Vote up and down) 
   3. Comment 
   5. The investor accepts (a) and responds (b) the answer. 
   6. System displays the performed task about the answer. | |
| Exceptions   | The investor posts a comment about the answer. System displays the comment of the investor. |
| Post conditions | The user who replied the request views the updated dataset request page. |
| Additional requirement | The system should handle multiple responds to a request. |
Use Case 2.4 | View and response to a request

**Description**
A user that has solution to a request can help the user that submitted a request. Platform facilitates users communication.

**Used by**
A journalist and a researcher

**Triggers**
When the journalist have the answer for the request

**Preconditions**
A dataset request should be submitted earlier. User should login to perform the tasks.

**Basic flow**
1. The journalist views the data request.
2. The journalist respond to the request by performing the following (requires user login):
   a) Insert a message.
   b) Provide a link.
3. The journalist posts a reply.
4. The system displays added reply and presents the updated request page.
5. The researcher views the request submitted by investor and a respond by a journalist.
6. The researcher perform the following task regarding the respond of the journalist (requires user login):
   a) Comment on the response of the journalist.
   b) Rate the respond of the journalist as an un-useful.
7. The researcher posts a reply.
8. The system displays the rating and the comment about the respond of the journalist and presents the updated request page.

**Exceptions**
The sequence of actions by journalist and researcher can be changed. Also unlimited number of users can contribute in this use case.

**Post conditions**
The dataset request page will be updated based on the latest responses.

**Additional requirement**
1 & 6. Both users have a preview pane, which recommends similar requests based on the subject of the request.
4.3.2.4 BPMN diagram

- Searching for a dataset request. This BPMN diagram is representing the process flow in use case 2.1

![Use case diagram of scenario 2b](image)

Figure 4-14 Use case diagram of scenario 2b

![BPMN diagram: Task flow of use case 2.1](image)

Figure 4-15 BPMN diagram: Task flow of use case 2.1
Figure 4-16 BPMN diagram: Task flow of use case 2.2

Figure 4-17 BPMN diagram: Task flow of use case 2.3
4.3.2.5 User interface design

Figure 4-18 BPMN diagram: Task flow of use case 2.4

Figure 4-19 Wireframe dataset request result
Figure 4-20 Wireframe add new dataset request

Figure 4-21 Wireframe Dataset request preview
4.3.3 Scenario 3: Downloading, processing, combining datasets and providing feedback

In this scenario the following requirements from section 3.3 is covered (27-33, 44-60, 92-95).

4.3.3.1 Scenario Description

A general physician, who is working in cancer research center, wants to research about the cancer rate in his country. For this purpose he focus on the reports of the cancer rate in different hospitals. To perform an accurate research he needs to have the detailed list of the hospitals and their corresponding cancer reports.

He uses the advanced search of the platform and finds about 50 results. Since he found large number of datasets. First he decides to access datasets directly from the search result page. He directly downloads five datasets, which have the highest ranking by clicking on the preferred format (two of them are CSV, one html and two others are XLS). He downloads the datasets for his research work.

Next he selects a datasets and enters to the dataset page. First he views the title, ranking and description of the dataset. He also previews quick summary and metadata consisting of
information about maintainer, publisher, tags, and timing about the dataset. From the quick summary he finds the dataset interesting and he share it in his social media account (LinkedIn).

Second he finds three formats that are available to use; PDF format, web page link and XLS format. He finds that he can use the XLS format in three ways; view, visualize and download it. He views the XLS format using the view button. Next he decides to visualize the file but unfortunately due to style of the file, the platform can’t visualize the datasets and presents an error message to the researcher to inform him about the happened error.

Third he views the detail metadata about the dataset which includes information about, maintainer, publisher, author, uploaded data, modified date, license, category and tags. Also he downloads the metadata relate to dataset. After performing the above actions the researcher finds the dataset useful for his research and clicks on the download link to download the XLS format. After downloading and using datasets and he finds that a column of dataset that contains important information is not filled out. Therefore he provides a feedback on the dataset page indicating the mentioned above problem and posts his comment. Also a cancer specialist posts a comment indicating the poor quality of datasets. After a while the dataset maintainer replies each of the feedbacks in a separate way (the general physician easily finds the reply related to his feedback). The feedbacks inform the maintainer that the dataset need to be completed. Thus the newer version is uploaded after a while, and the researcher downloads the updated datasets from the same dataset page. Also general physician can report a reply as containing incorrect information.

Afterwards he navigates to search page and finds datasets about the all the hospitals in the country. He downloads the dataset as described above. He finds the list very useful but the dataset doesn’t contain information regarding the hospitals in two major cities. He uses the finds two separate datasets that contains the missing information. He combines three datasets, which is completed version of the first dataset. He selects the link ‘quick extend’ and the type as ‘combination of dataset’ and the same title as the original data set. Also he selects to upload dataset to platform and submits his new dataset. Since the maintainer of the dataset is a researcher. This dataset can be displayed as a completed version after approving by the maintainer or the platform expert. Note that the general physician can select the extend link to publish a dataset directly and this make him as the maintainer of dataset (this is link to scenario 5 to upload a dataset through wizard).
4.3.3.2 Use case diagram

Figure 4-23 Use case diagram of scenario 3

[Use case diagram of scenario 3 with various use cases and relationships]
4.3.3.3 Use cases

Table 4-10 Use case 3.1 description

<table>
<thead>
<tr>
<th>Use Case 3.1</th>
<th>View dataset page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>A user can search for dataset and get detailed information about it.</td>
</tr>
<tr>
<td>Used by</td>
<td>• Use case 5.2</td>
</tr>
<tr>
<td></td>
<td>• Use case 4.4</td>
</tr>
<tr>
<td>Actors</td>
<td>A general physician</td>
</tr>
<tr>
<td>Triggers</td>
<td>When a user needs to find dataset.</td>
</tr>
<tr>
<td>Preconditions</td>
<td>User should login in order to be able to provide feedback and extend dataset.</td>
</tr>
<tr>
<td>Basic flow</td>
<td>1. User select a dataset from search list</td>
</tr>
<tr>
<td></td>
<td>2. System navigates to the dataset page.</td>
</tr>
<tr>
<td></td>
<td>3. User views the following parts related to the dataset.</td>
</tr>
<tr>
<td></td>
<td>1. Basic information</td>
</tr>
<tr>
<td></td>
<td>1. Title</td>
</tr>
<tr>
<td></td>
<td>2. Openness Rating</td>
</tr>
<tr>
<td></td>
<td>3. Favorite Mark</td>
</tr>
<tr>
<td></td>
<td>4. Publisher</td>
</tr>
<tr>
<td></td>
<td>5. License</td>
</tr>
<tr>
<td></td>
<td>6. Description</td>
</tr>
<tr>
<td></td>
<td>2. Quick summary</td>
</tr>
<tr>
<td></td>
<td>1. Title</td>
</tr>
<tr>
<td></td>
<td>2. Publisher</td>
</tr>
<tr>
<td></td>
<td>3. Release date</td>
</tr>
<tr>
<td></td>
<td>4. License</td>
</tr>
<tr>
<td></td>
<td>5. Country</td>
</tr>
<tr>
<td></td>
<td>6. Tags</td>
</tr>
<tr>
<td></td>
<td>7. Number of downloads</td>
</tr>
<tr>
<td></td>
<td>8. Sharing on social media</td>
</tr>
<tr>
<td></td>
<td>3. Resources</td>
</tr>
<tr>
<td></td>
<td>1. Title</td>
</tr>
<tr>
<td></td>
<td>2. Description</td>
</tr>
<tr>
<td></td>
<td>3. Size</td>
</tr>
<tr>
<td></td>
<td>4. Language</td>
</tr>
<tr>
<td></td>
<td>5. Upload date</td>
</tr>
<tr>
<td></td>
<td>4. Metadata</td>
</tr>
<tr>
<td></td>
<td>1. Title</td>
</tr>
<tr>
<td></td>
<td>2. Publisher</td>
</tr>
<tr>
<td></td>
<td>3. Maintainer</td>
</tr>
<tr>
<td></td>
<td>4. Author</td>
</tr>
<tr>
<td></td>
<td>5. License</td>
</tr>
<tr>
<td></td>
<td>6. Country</td>
</tr>
<tr>
<td></td>
<td>7. Geographical coverage</td>
</tr>
</tbody>
</table>
8. Temporal coverage
9. Release date
10. Update date
6. Extension graph (extend)
7. Files based on dataset (Quick extend)
   1. Type
   2. Title
   3. Publisher
   4. Upload date
   5. Download or URL link
8. Feedback

Exceptions
The presented schema for some datasets may contain no information and in this case the null filed will display deactive.

Post conditions -
Additional requirement -

Table 4-11 Use case 3.2 description

<table>
<thead>
<tr>
<th>Use Case 3.2</th>
<th>Use dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>User can view dataset directly in the open data platform. Also he can download and visualize dataset and provide feedback about it.</td>
</tr>
<tr>
<td>Used by</td>
<td>-</td>
</tr>
<tr>
<td>Actors</td>
<td>General physician</td>
</tr>
<tr>
<td>Triggers</td>
<td>When user wants to use a dataset</td>
</tr>
<tr>
<td>Preconditions</td>
<td>-</td>
</tr>
</tbody>
</table>
| Basic flow   | 1. User enters keyword in search bar.  
               2. System provides the search result.  
               3. User views the dataset lists with a provided download format and openness ratings for each dataset.  
                   a. User downloads a dataset directly from search result page.  
                      1. System provide dataset link for download by user.  
                   b. User selects a dataset and navigates to dataset page. Following actions can be done by user and system regarding the ‘resources’ section:  
                      1. User selects ‘View’  
                         i. System navigates to view dataset page.  
                         ii. Users views dataset directly in the platform (XLSX, PDF, CSV)  
                      2. User selects ‘Visualize’  
                         i. System navigates to visualize dataset |
ii. User can select the following options to visualize and filter dataset information.
   1. View as table
   2. View as chart
   3. View as map

iii. System represents the select visualized option.

3. User selects ‘Download’
   i. System provides the download link for download
   ii. User downloads the dataset.

4. User views metadata and download the metadata directly.
   i. System provides a download link for metadata.
   ii. User downloads the metadata.

### Table 4-12 Use case 3.3 description

<table>
<thead>
<tr>
<th>Use Case 3.3</th>
<th>Extend dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>System provides options for user to extend the existing dataset.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Actors</strong></td>
<td>A general physician, Specialist</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>When a user creates new item based on the dataset or want to update existing dataset.</td>
</tr>
</tbody>
</table>
| **Preconditions** | • User should login  
• This task can be deployed if a user navigates to dataset page |
| **Basic flow** | 1. User can perform following two actions to extend dataset.  
   1. User selects Quick extend  
   2. System provides quick extend pane.  
   3. User fills out the following fields and selects submit.  
   1. Type (Combined dataset, Publication, Application, web services, visualization image)  
   2. Title  
   3. URL or dataset file  
   2. System presents the extension on ‘files based on
dataset page’ |
### Use Case 3.4 Provide feedback

<table>
<thead>
<tr>
<th>Description</th>
<th>System provides option for users contributions.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Used by</td>
<td>General physician, Specialist</td>
</tr>
<tr>
<td>Triggers</td>
<td>When a user wants to discuss about the dataset</td>
</tr>
<tr>
<td>Preconditions</td>
<td>• User should login</td>
</tr>
</tbody>
</table>
| Basic flow  | 1. User can contribute in discussing about the dataset and provide feedback by performing the following actions:  

   1. User Posts a comment  

   2. User posts Comment on a reply by other users  

   3. User Rate a comment  

       i. System represents the selected action.  

   4. User Report a reply of other users regarding the following issues:  

       a. Offensive comment  

       b. Inappropriate content  

       i. System represents the notification message to user that report is submitted. |

| Exceptions  | User can select each of actions to provide feedback. There is no sequence in the presented actions. |
| Post conditions | -                                                  |
| Additional requirement | -                                                      |
4.3.3.4 BPMN diagram

Figure 4-24 BPMN diagram: Task flow of use case 3.2

Figure 4-25 BPMN diagram: Task flow of use case 3.3
Figure 4-26 BPMN diagram: Task flow of use case 3.4

4.3.3.5 User interface design

Figure 4-27 Wireframe Dataset main page
Figure 4-28 Wireframe dataset use

Figure 4-29 Wireframe dataset page (continue)
Figure 4-30 Wireframe quick extend

Figure 4-31 Wireframe dataset visualization tool
4.3.4  Scenario 4: Managing user account and communications.

In this scenario the following requirements from section 3.3 is covered (18-20, 61-72, 77, 78, 92-94.)

4.3.4.1  Scenario Description

A researcher who is working in physics department is active in using the open data platform. She had different interactions with the platform. She has uploaded 5 datasets, submitted 2 dataset requests, downloaded 8 datasets, provided feedbacks, responded to requests, sent private messages and etc. She has recently submitted a ‘dataset request’ and asked a question from a public servant who works in the ministry of science in a private message. She receives an email indicating that he received a private message from the public servant. She navigates to open data platform and log-in to the system using his social media account (LinkedIn). In the main page of the platform he views that he has two notifications as well as one message. She directly clicks on the message and navigates to the inbox and responds to the message.

Next she navigates to the user profile by clicking on her user account. She views two main parts. One part, which is private provides summary of basic information, links to notifications, edit profile, edit preferences, navigate to inbox, access to community groups, sign in logs, saved search results and logout. The other part, which is a public, provides the activity that she has performed. She can view summary of activities (uploads, downloads, request and etc.), maintaining datasets, dataset requests, responded requests, tags, downloaded datasets, feedbacks, favorite datasets and detail of activities.

There after she views in summary that they are two responds to each of the requests. She selects the ‘responds tab and preview that one her requests has been viewed by 150 users and two people has responded to her request. She finds that she can sort responds based on the ‘oldest’ and ‘newest’. She views the two answers in his user profile. In order to post a comment about one of the respond he clicks on the link of ‘dataset request’ and navigates to the ‘dataset request’ page and posts her comment. The researcher isn’t happy that she didn't receive any notification about the responds. Therefore she navigates to her user account page, selects preferences and finds that she hasn't tick ‘notification email’ feature. Thus she select this option, submit the ‘save setting’ button. She finds in her user account that information is incomplete. She selects edit profile tab and insert her background knowledge, her occupation, and her current location.

Next she selects the community tab and views that she is member of two communities ‘university ‘z’ employees’ and ‘ science’. She decides to create a new group (community) by selecting the ‘creating new group’ tab. She names the group as ‘physics’ and enters the name of group members and administrator. Next she sends a group message to the ‘science’ group
members asking about a researcher who is doing research in nanophysics. One of the group members replies a name ‘Smith’ who is a researcher working in that field and recently joined to the platform.

A university professor working in a physic department has received the invitation from the research to join to the new group named ‘physics’. He reads group description and reviews the group members. He finds the researcher as the group administrator. Therefore the professor navigates to the researcher’s profile. He views her public profile and find her works interesting and related. Thus he navigates back to the group management and accepts the researcher invitation to join to new group ‘physics’.

The researcher navigates to the platform home page and selects the user tab and selects communities from sub-menu. She views different communities but she doesn't find any group related to nanophysics. And she decides to directly search in the members. For this purpose she selects user tab in the main page and navigates to user’s page. Next she enters ‘Mr. Smith’ to search in the users list and views 50 results. Next she sorts the search results by selecting the category (new users, dataset requests, dataset uploads, responds to requests). She selects ‘new users’ as the preferred category and sorts them based on the ‘joined date’. Since she knows that Mr. Smith has joined recently (two days ago) she easily finds him and navigates to his profile.

She views his public profile summary. She views datasets uploaded by him and sort them based on the’ most viewed’ (most viewed, newest, favorite) datasets. She views a dataset about her interested field (nanophysics) and navigates to the dataset’s page and downloads it.
4.3.4.2 Use case diagram

Figure 4-32 Use case diagram of scenario 4
4.3.4.3 Use cases

Table 4-14 Use case 4.1 description

<table>
<thead>
<tr>
<th>Use Case 4.1</th>
<th>Access user account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Users can have user account and manage their preferences and interaction directly from there. Besides user account facilitates user’s interactions by different service (Messages, Communities).</td>
</tr>
<tr>
<td>Used by</td>
<td>• Use case 4.4</td>
</tr>
<tr>
<td>Actors</td>
<td>A researcher</td>
</tr>
<tr>
<td>Triggers</td>
<td>When user need to communicate and manage pervious interactions with platform</td>
</tr>
<tr>
<td>Preconditions</td>
<td>• User should login</td>
</tr>
</tbody>
</table>
| Basic flow   | 1. System provide following features divided into two main part  
|              | 2. Private part which is accessible for the account owner is divided into following sections:  
|              |   1. Basic information  
|              |   2. Edit profile  
|              |   3. Notifications  
|              |   4. Inbox  
|              |   5. Edit preferences |
6. Community groups
7. Saved search results
8. Sign in logs
9. Logout

b. Public part is accessible by all users and consists of two sections:

I. Data provider: represents activities of a user when he/she provides datasets and responds to dataset requests
   1. Summary
   2. Datasets
   3. Feedbacks
   4. Responded requests
   5. Tags

II. Data seekers: represent the activities of a user when he/she needs to use datasets.
   1. Summary
   2. Dataset requests
   3. Responds
   4. Downloaded datasets
   5. Favorite datasets

3. User views the features.

<table>
<thead>
<tr>
<th>Exceptions</th>
<th>User use different features provided by platform (This will be displayed in correspondence use cases).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post conditions</td>
<td>-</td>
</tr>
<tr>
<td>Additional</td>
<td>-</td>
</tr>
</tbody>
</table>

### Table 4-15 Use case 4.2 description

<table>
<thead>
<tr>
<th>Use Case 4.2</th>
<th>Manage private user account</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Platform enables managing different features from user profile</td>
</tr>
<tr>
<td>Used by</td>
<td>• Use case 4.1</td>
</tr>
<tr>
<td>Actors</td>
<td>A researcher</td>
</tr>
<tr>
<td>Triggers</td>
<td>When a user wants to manage private parts of user profile</td>
</tr>
<tr>
<td>Preconditions</td>
<td>User should login</td>
</tr>
<tr>
<td>Basic flow</td>
<td>1. User views the first page of the private part that displays the Basic information.</td>
</tr>
<tr>
<td></td>
<td>2. System displays basic information consisting of following fields</td>
</tr>
<tr>
<td></td>
<td>1. Name of the user</td>
</tr>
<tr>
<td></td>
<td>2. Country</td>
</tr>
<tr>
<td></td>
<td>3. Study field</td>
</tr>
<tr>
<td></td>
<td>4. Occupation</td>
</tr>
<tr>
<td></td>
<td>5. Contact information</td>
</tr>
</tbody>
</table>
6. Membership date
7. Number of profile views
2. User select ‘Edit’ Tab
3. System navigates to edit mode.
4. User edit different fields and adds contact information and saves the results.
5. System navigates to basic info mode.
7. System represents the latest interaction regarding public part (responds, feedbacks, etc.)
9. System represent the received messages, and provides following options:
   1. Sent messages
   2. Compose
10. User select compose tab
11. System navigate to send message page
12. User writes the subject, receiver (groups of individuals), and message text and submits send button.
13. System sends the message and navigates to inbox.
14. User selects Edit preferences
15. System enable users select to get email about the new notifications and new messages.
16. User activates to get email and save the results.
17. System displays the updated information.
18. User select community tab.
19. System navigates to community page and represents the Group name and role of the user. User can view all groups, groups that owns, groups that is member of. Also user can delete group that he/she owns, and create new groups.
20. User select create new group.
21. System represents create new group pane consisting of following fields:
   1. Group title
   2. Group members
   3. Administrators
22. User fills out the form and submits ‘create new group’ tap.
23. System represents the created group in the group list.
25. System represents saved search result. (if user saved previous search).
26. User select ‘Sign in logs’ tab
27. System represents the previous login of the user with detail date and time.
28. User select log out tab
29. System navigates to the home page of the platform.
### Use Case 4.3 Access public user account

**Description**
Platform provides a public access of user interactions from user profile.

**Used by**
- User case 4.1

**Actors**
A public servant

**Triggers**
When users want to have overview of their interaction with the platform and other users.

**Preconditions**
User should login

**Basic flow**
1. Use views the public section of the user profile
2. System represent the public section consisting of the following:
   1. Data provider: represents activities of a user when he/she provides datasets and responds to dataset requests
      1. Summary shows the latest (Datasets, Feedbacks, Tags)
      2. Datasets (Downloads, views, Rating)
      3. Responds (User replied, text of respond)
      4. Replies (DS request, and self replies)
      5. Tags (keywords used when uploading DS)
   2. Data seekers: represent the activities of a user
      1. Summary (DS requests, Downloaded DS, Favorite DS)
      2. Dataset requests (Views, replies, Rating)
      3. Feedbacks (User replied, Text of feedback)
      4. Downloaded datasets (Title, Ranking, viewed, downloaded)
      5. Favorite datasets (Title, Ranking, Date marked)
3. User can select each section and access to the desired information.
4. System represents the detail about each section in user profile page.

**Exceptions**
In the cases that there was no action performed by action the list will be displayed as an empty list.

**Post conditions**
- 

**Additional requirement**
- 

---

<table>
<thead>
<tr>
<th>Exceptions</th>
<th>There is no sequence between in performing different task user can do any task by selecting the related tab.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post conditions</td>
<td>-</td>
</tr>
<tr>
<td>Additional requirement</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 4-17 Use case 4.4 description

<table>
<thead>
<tr>
<th><strong>Use Case 4.4</strong></th>
<th><strong>Find User</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Platform facilitates user’s collaboration by providing features for users to find each other easily.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Actors</strong></td>
<td>Public servant</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>When a user needs to find a user</td>
</tr>
<tr>
<td><strong>Preconditions</strong></td>
<td>Users should login to be able to view user profiles</td>
</tr>
</tbody>
</table>
| **Basic flow**   | 1. The public servant first selects ‘user’ tab and next ‘profiles’ from home page of the platform  
2. Platform navigates to the user profile page and display list of users. Also it provides following features:  
   1. Search user’s profiles by inserting text.  
   2. Sort user profile based on the following criteria:  
      1. New users  
      2. Dataset requests  
      3. Dataset uploads  
      4. Responds to requests  
3. User inserts text in the search bar and submits it.  
4. System provides the list of users.  
5. User sort result based on one of the criteria  
6. System represent the result based on the selected criteria  
7. User finds the person he was looking for and selects the user’s profile.  
8. System navigates to user profile page.  
9. User finds a dataset from a user profile and selects it.  
10. System navigates to the dataset page. |
| **Exceptions**   | User may not use the search functions and navigates between different pages of user profile list. |
| **Post conditions** | - |
| **Additional requirement** | - |

### Table 4-18 Use case 4.5 description

<table>
<thead>
<tr>
<th><strong>Use Case 4.5</strong></th>
<th><strong>Find a community</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>Platform facilitates user’s collaboration by providing features for users to join a community and find different communities.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>-</td>
</tr>
<tr>
<td><strong>Actors</strong></td>
<td>A public servant</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>When user needs to find a community to join or find specific person.</td>
</tr>
</tbody>
</table>
Preconditions
User should login to have access to communities

Basic flow
1. The public servant first selects ‘user’ tab and next ‘communities’ from home page of the platform
2. Platform navigates to the user communities’ page and display list communities. Also it enables users to view all communities. Also user can leave the groups and send request to join groups.
3. User selects a group that is member of.
4. System displays the group page and list of its members. With an option to leave the group.
5. User selects a user account.
6. System navigates to the user profile.

Exceptions
3. User finds a group and sends a request to join the group.
4. Platform change the state of the status from ‘send request’ to ‘pending request’.

Post conditions
-

Additional requirement
-

4.3.4.4 BPMN diagram

Figure 4-34 BPMN diagram: Task flow of use cases 4.1 &4.2 &4.3
Figure 4-35 BPMN diagram: Task flow of use cases 4.4

Figure 4-36 BPMN diagram: Task flow of use cases 4.5
4.3.4.5 User interface design

Figure 4-37 Wireframe User profile main page

Figure 4-38 Wireframe User profile (datasets)
Figure 4-39 Wireframe User profile (Communities)

Figure 4-40 Wireframe Users profile list
4.3.5 Scenario 5: Preparing, Uploading and modifying dataset and metadata

In this scenario the following requirements from section 3.3 is covered (21-26, 73, 75, 76, 82-91).

4.3.5.1 Scenario Description

A public servant who is working in ministry of finance is responsible to upload a dataset (a PDF report) regarding the household income in different province in the country ‘Z’ to the open data platform.

He searches for the open data platform in a search engine and enters to the home page of the platform. He directly enters to dataset upload page. He selects the add dataset link but he needs to log in to be able to upload datasets. After logging in, he enters to the upload page.

He views the upload procedure is consisting of seven steps (prepare data, upload dataset, basic information, organization, coverage, scientific coverage and preview). In first step he views quick summary of the preferred data format by the platform. To navigate to the next step he asked to answer if the dataset is ready to upload (according to the guidelines provided in the wiki section). He selects negative as the answer, therefore he navigates to the ‘data preparation guideline’ page. After reviewing the guidelines he decides to upload CSV and XLS format of the result table.

He enters to the ‘dataset upload’ page and selects positive as the answer of the first step. In second step he enters title as ‘household income’, description, tags of dataset. Also he selects health as the category and the Netherlands as the country of the dataset. He select next to navigate to the third step but he is returned in the same step with an error messaging regarding the title of dataset (this title exist, please be more specific). He views an exclamation mark in front of the title bar. He clicks on the mark and gets a quick description about selecting a good title. This time he enters ‘household in come in different cities’ and navigates to the third step. In third step he uploads two datasets and selects the language of the datasets as ‘Dutch’ and selects ‘next’ to navigate to 4th step. In this step he needs to import temporal granularity but doesn’t know, he clicks on the exclamation mark to get more information. He finds that he needs to import the coverage date of datasets. He enters 2000 to 2010 as an input. Also he imports information related to the geographical information related to the datasets and navigates to the 5th step. In this step he enters information about author and publisher. He leaves the maintaining group empty since he doesn’t find any group in the recommendation list. He navigates to the 6th step and provides information about scientific context of dataset and selects next to navigate to the 7th step. In last step he previews all the information he provided and finds that he need to add ‘income’ as a tag. Therefore he selects the related tab and input the extra tag and navigates to the last step. He selects the ‘upload’ and publishes the dataset.

Since he is conducting research in this field after two weeks he ends up with more completed version of datasets therefore he decides to upload the completed dataset. To do this he
navigates to the uploaded dataset page. He selects the ‘edit resource tab’ and a new pane is provided for him to upload the dataset in a quick way. He chooses to upload the file and enters ‘updated’ as a short description and selects the language of dataset as ‘Dutch’. He selects the submit button and views the uploaded file right under the dataset. Note that he can upload another formats or URL link and edit metadata as well, through the same procedure, which he updated the previous dataset.

4.3.5.2 Use case diagram

![Use case diagram of scenario 5](image_url)

*Figure 4-41 Use case diagram of scenario 5*
4.3.5.3 Use cases

Table 4-19 Use case 5.1 description

<table>
<thead>
<tr>
<th>Use Case 5.1</th>
<th>Add dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>One of the main functions of the open data platform is facilitating dataset upload. This will be done by providing support and guiding the user during all steps of adding dataset.</td>
</tr>
<tr>
<td>Used by</td>
<td>Public servant</td>
</tr>
<tr>
<td>Actors</td>
<td></td>
</tr>
<tr>
<td>Triggers</td>
<td>When users or organizations aimed to publish dataset</td>
</tr>
<tr>
<td>Preconditions</td>
<td>• User should login</td>
</tr>
</tbody>
</table>
| Basic flow   | 1. A Public servant selects add dataset tab button from main page of the platform.  
  2. System navigate to add dataset page and provide the following sections:  
     1. Quick summary of dataset upload steps with a link to the tutorial of each step. The dataset |
upload steps are the following:

i. Prepare data
ii. Upload dataset
iii. Basic information
iv. Organization detail
v. Dataset Coverage
vi. Scientific coverage
vii. Preview

2. Asking user whether the dataset is ready for upload or not.
3. User confirms that the dataset is ready to upload.
4. System navigates to dataset upload wizard.
5. User completes the wizard steps and select submit.
6. System navigates to preview dataset upload page. Provides a preview of uploaded content. The following options is provided by the system:
   1. Edit
   2. Delete
   3. Publish
7. User selects to publish dataset
8. System navigates to the dataset page and presents the following options:
   1. Edit
   2. Delete

Exceptions

Two following exceptions regarding the action 3 are:

a.
3. User doesn't confirm that dataset is ready.
4. System navigate to ‘learn to interact with platform’ page.

b.
3. User select one of the links in the quick summary section
4. System navigates to the correspondence tutorial.

Post conditions

The dataset is uploaded

Additional requirement

-  

<table>
<thead>
<tr>
<th>Use Case 5.2</th>
<th>Edit dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td>After finishing upload wizard and after publishing the dataset user can edit dataset.</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>• Use case 5.1</td>
</tr>
<tr>
<td><strong>Actors</strong></td>
<td>Public servant</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>User wants to edit the uploaded content</td>
</tr>
<tr>
<td><strong>Preconditions</strong></td>
<td>User should have completed the upload wizard</td>
</tr>
</tbody>
</table>

Table 4-20 Use case 5.2 description
Basic flow

1. User views ‘dataset preview’ page.
2. User selects Edit tab.
3. System provides the following options to edit dataset
   1. Edit basic information
      System enables editing the basic information the fields that represented in Use case 3.1 can be edited
   2. Edit resources
      System enables editing the resource. Dataset can be deleted (at least one resource should be available) and new sources can be added by providing following details:
      - Title
      - Description
      - Resource (file, URL)
      - Language of dataset
   3. Edit metadata
      System facilitates updating metadata fields presented in use case 3.1
4. User edits a field and select save button.
5. System represents the updated dataset page.

Exceptions
At a mean time user can edit one field.

Post conditions
- 

Additional requirement
- 

<table>
<thead>
<tr>
<th>Use Case 5.3</th>
<th>Delete dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>After publishing a dataset user can delete dataset</td>
</tr>
<tr>
<td>Used by</td>
<td>• Use case 5.1</td>
</tr>
<tr>
<td>Actors</td>
<td>Public servant</td>
</tr>
<tr>
<td>Triggers</td>
<td>When a user wants to delete a published dataset</td>
</tr>
<tr>
<td>Preconditions</td>
<td>User should have published a dataset</td>
</tr>
<tr>
<td>Basic flow</td>
<td>1. User views the dataset page.</td>
</tr>
<tr>
<td></td>
<td>2. User selects delete button.</td>
</tr>
<tr>
<td></td>
<td>3. System represents the warning message and asks user if he/she wants to delete dataset.</td>
</tr>
<tr>
<td></td>
<td>4. User confirms deletion.</td>
</tr>
<tr>
<td></td>
<td>5. System navigates to datasets page and displays a notification that the dataset is deleted.</td>
</tr>
<tr>
<td>Exceptions</td>
<td>-</td>
</tr>
<tr>
<td>Post conditions</td>
<td>Dataset is removed from database.</td>
</tr>
<tr>
<td>Additional requirement</td>
<td>-</td>
</tr>
<tr>
<td>Use Case 5.4</td>
<td>Upload dataset Wizard</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td><strong>Description</strong></td>
<td>Platform facilitates the steps of uploading a dataset. The platform provides basic guides for each step by linking the steps to the related documentation in the platform</td>
</tr>
<tr>
<td><strong>Used by</strong></td>
<td>• Use case 5.1</td>
</tr>
<tr>
<td><strong>Actors</strong></td>
<td>A Public servant</td>
</tr>
<tr>
<td><strong>Triggers</strong></td>
<td>When user want to upload dataset and confirms to navigate to upload dataset wizard</td>
</tr>
<tr>
<td><strong>Preconditions</strong></td>
<td>User should have confirmed that the dataset is prepared according the platform requirements.</td>
</tr>
</tbody>
</table>
| **Basic flow** | 1. User views the upload dataset wizard page.  
2. Platform provides following steps to accomplish by user, to upload dataset.  
   1. Basic information  
      1. Title  
      2. Description  
      3. License  
      4. Country  
      5. Categories  
      6. Language  
   2. Upload Dataset (user can add/delete multiple resource)  
      1. Upload file or Provide URL  
      2. Description  
      3. Language  
   3. Dataset coverage  
      1. Temporal granularity  
      2. Geographical coverage  
   4. Organizational detail  
      1. Author  
      2. Publisher  
      3. Release date  
      4. Maintainer Group  
   5. Scientific coverage  
      1. Data collection type  
      2. Data collection description  
      3. Statistical software used  
      4. Analysis unit  
      5. Statistical methodology  
3. User fills out the first step of the wizard.  
4. System recommends similar datasets to avoid publishing similar datasets.  
5. User selects ‘next’  
6. System navigates to all other 4 steps of wizard.  
7. User fills out the 5th step and selects submit.  
8. System navigates to dataset preview page. |
| **Exceptions** | 5. User selects one of the recommended dataset. |
6. System navigates to the dataset page.
7. User views the dataset page.
8. User selects ‘expand’ and publish a dataset linked to the recommended dataset.

<table>
<thead>
<tr>
<th>Post conditions</th>
<th>System navigates to the dataset preview page.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Additional requirement</td>
<td>In all the steps of the wizard, system provides more information regarding each field of different steps of the wizard. User can have a preview of the guide and also user can navigate to detail description of each step by clicking the ‘more info’ link provided in each step.</td>
</tr>
</tbody>
</table>

### 4.3.5.4 BPMN diagram

![BPMN diagram](image)

Figure 4-43 BPMN diagram: Task flow of use case 5.1
Figure 4-44 BPMN diagram: Task flow of use case 5.2

Figure 4-45 BPMN diagram: Task flow of use case 5.3
4.3.5.5 User interface design
Figure 4-48 Wireframe dataset edit

Figure 4-49 Wireframe dataset edit (add resources)
Figure 4-50 Wireframe dataset page (Metadata edit)

Figure 4-51 Wireframe add dataset wizard
4.3.6 Scenario 6: Using documentation/tutorials

In this scenario the following requirements from section 3.3 is covered (3, 73-91).

4.3.6.1 Scenario Description

A public servant working in ministry of health is willing to upload dataset to the open data platform. She uses the platform to get more information about preparing the dataset to upload. She navigates to the home page of the platform and views different tabs. She is looking for general information about data and maybe tutorials. She views the help tab. She finds that the help tab has three other sections (frequently asked questions, tutorials and documentation). She selects ‘frequently asked questions’ and navigates to that page. She views a question ‘how should I prepare datasets before publishing?’ The answer provides a link that contains a video tutorial about that. She navigates to the link page and watches the tutorial video.

After wards she views list of related documentation and tutorials. She selects a link ‘uploading a dataset’ and navigates to the page. She finds the provided information useful and returns to the datasets in order to prepare dataset according to what she has learned.

4.3.6.2 Use case diagram
4.3.6.3 Use cases

Table 4-23 Use case 6.1 description

<table>
<thead>
<tr>
<th>Use Case 6.1</th>
<th>Learn to interact with platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Platform provides different resources to guide and educate users to deal with platform and open data related topics.</td>
</tr>
<tr>
<td>Used by</td>
<td>A public servant</td>
</tr>
<tr>
<td>Actors</td>
<td>When the user doesn’t know how to</td>
</tr>
<tr>
<td>Triggers</td>
<td>• User should login to contribute in improving the documentation.</td>
</tr>
<tr>
<td>Preconditions</td>
<td>Basic flow 1. User selects ‘help’ tab from the home page of the platform. 2. System present following three sections: 1. Documentation: users can contribute in enriching different parts of the documentation as well as using documentation. 2. Tutorials: consist of video tutorials and step-by-step task description. 3. Frequently Asked Questions (FAQ) 3. User selects documentation 4. System navigate to documentation part and for each</td>
</tr>
</tbody>
</table>
topic it represents following options:

1. View
2. Edit: user can contribute in writing documentation and adding more information regarding each section.
3. Updates: users can view previous editions of a document as well as document editor and time of update.
4. Attachments: user can Upload and download attachments related to documentation.
5. Notification: user can select to receive notification or email about a document.

5. User selects a document and select edit tab
6. User adds information regarding the documentation and save the edit.
7. System represents the updated document.

Exceptions

• User selects to navigate to FAQ section
• User selects to navigate to Tutorial section
• User selects to view Updates
• User selects notification and sets to receive notification

Post conditions -

Additional requirement -

4.3.6.4 BPMN diagram

![BPMN diagram: Task flow of use case 6.1](image)

Figure 4-54 BPMN diagram: Task flow of use case 6.1
4.3.6.5 User interface design

Figure 4-55 Wireframe Help page

Figure 4-56 Wireframe Documentation page
4.4 Analysis results

4.4.1 Audience Modeling

The aim of this thesis is to create user interfaces for different groups of users. To satisfy the goal of thesis we focus on audience-driven design approach. An audience-driven design idea reflects that the different target audiences and their requirements. The requirements are taken as starting point for the design and the basis of structure of the platform. Since one of the objectives of the thesis is about designing UI from different categories of users, we will focus on this method for initial requirement analysis of the open data platform.

The target users identified in the previous section are refined into audience classes. This is done by means of two sub-phases: the audience classification and the audience characterization. During audience classification, the different types of users are identified in more detail and classified into so-called audience classes. During audience characterization, relevant characteristics are specified for each audience class.
4.4.2 Audience classification and characterization methods

Target users classified into audience classes based on differences in their informational and functional requirements. All members of an audience class must have the same set of informational and functional requirements. Set of informational and functional requirements of one audience class is a subset of the set of requirements of other audience class.

Two alternative methods for discovering the audience classes identified. The first method uses the activities of the organization for which the Web system needs to be developed and the role people play in these activities. In summary the first method is as follows:

Step 1: Consider the activities of the organization related to the purpose of the Web system.
Step 2: For each activity,

1. Identify the people involved.
2. Restrict them to the target users.
3. Identify their requirements.
4. Divide them into audience classes based on different informational or functional requirements.
5. Decompose the activity if possible, and repeat Step 2.

This creates the hierarchy of audience classes. Audience class Visitor is always placed at the top of the audience class hierarchy. The Visitor audience class represents all target users. The requirements associated with the Visitor class are the requirements that are common to all users.

The second method starts with identifying all possible informational and functional requirements of the target users, without considering how to classify them in audience subclasses.

The algorithm to perform this method is formally specified in XXX (2001). In summary, it is as follows:

1. Construct the audience class matrix based on all the requirements formulated for the Web system to be built.
2. Determine the equivalence rows. Each set of equivalence row represents an audience class. The user requirements for this audience class are the requirements associated with the rows. Meaningful names should be given to the audience classes.
3. Identify subset relations between the rows of the different equivalence classes. These subset relations result in audience subclass relations.

4. Construct the integrated audience class hierarchy (using the audience classes and the audience subclass relations).

4.4.3 Overall description

To achieve the objectives that are addressed above, we focus on improving the user interface of current open data platform. This will result in designing new user interface for different group of users, which provide different services, including the process of importing, curating, processing and visualizing the data, based on the system requirements document of web-based portal. The major services and functions that the system provides for its users are:

- A web-based user interface, which provides easy access to open data for all citizens.
- A dataset management system that provides download function for users. The system enables dataset storage as well as data catalogue for external datasets.
- A dataset management system that provides Upload function for users. In order to enable users to upload a new datasets or modify or curate the existence datasets of the platform.
- A request handler that is responsible to manage dataset requests. It provides the possibility to submit a dataset requests. This enables the users not only to submit a request but also an advanced search possibility for data provides to reply and provide possible sources for data seekers.
- A user management system that provides user’s sign-up/ sign-in. and enables customization of user profile in order to provide more functions for various users categories.
- Advanced search mechanisms to boost up the access to datasets and ‘dataset requests’, such as faceted browsing and geo-spatial search. As well as search result-sorting possibility (e.g. downloads, quality ranking, recently uploaded and etc.)
- Visualization and processing tool support, which enables display of structured datasets. Users may need to view visualize datasets in online maps or chats.
- A two way communication system to provide feedback about the quality, validity and usefulness of specific dataset. This enable ranking the datasets based on different metrics (e.g. quality, usability, validity)
- A robust tutorial and documentation support to inform users about the functions of the platform by providing information about different tools and services that portal offers.
- API that facilitates the communication between the platform and third party applications. The API will include methods for the federation of content, in order to share datasets and dataset metadata with other data hubs and third party applications.
4.4.3.1 User Roles and categories

To provide the system objective we need to focus on separating concerns. Also we have identified the stakeholders and the main functionality of the system, we need to separate task and relate them to each group categories for this purpose, and first we need to categorize different users. Second we describe the related task to each category in order to provide a detailed description of main function of the system. In this case we limit the audience into three main categories.

1) **Visitors**: are users who either do not have a user account or are not currently logged in. Visitors have read-only access to published content through the platform, such as the provided datasets, user-submitted content and documentation also they can search for specific type of datasets as well as processing and visualizing datasets. Visitors may use the platform’s API for searching and retrieving datasets, however all unauthenticated API requests are

2) **Registered users** are users who created a user account, which allows them to access personalized services. Registered users can rate and comment on datasets, suggest the inclusion of new datasets and contribute to the platform’s Wiki-based documentation, however the platform staff moderates their submissions before being published publicly. Registered user are divided in two sub categories in the following:

   i)  **Data Providers** are registered users that are the main providers of datasets of the platforms. Thus they are able to upload dataset through a wizard, which guide them through the upload process. Also they are able to modify the uploaded datasets based on the user’s and platform maintenance group’s feedback and modified version by users. The other main function of data provider’s category is replying the data request, which is made by ‘data seekers’ category. They have advanced search option to search for a request based on the tags, popularity and etc. Public agents (servants), publishers and researchers (considered as data provider) are considered as member of this category.

   ii) **Data Seekers** are registered users that are looking for datasets. They have advanced searching options to look for their preferred datasets. They can comment about datasets, rate it and submit a modified version of the dataset. The modified version needs confirmation of platform maintenance group or data provider in order to republish. Data seekers are able to submit a data request based on the category, tag and detailed description of needed dataset. Also they have a possibility of accessing ‘data request’ based on the category, rating, date and tag of dataset in order to avoid reparative requests. Citizens and researchers (considered as data consumer) are categorized under this user group.

3) **Platform maintenance group** are responsible for the platform maintenance and administrator. This group is divided into following roles:

   i)  **Moderators** are responsible for reviewing user-submitted content, including comments, and suggestions, in order to modify or delete any material that violates the platform’s Terms of Use.
ii) **Platform Experts** provide assistance to ‘Data providers’ on using platform services to publish PSI and provide all relevant metadata, as well as ‘Data seekers’ on tools and methodologies to process the published datasets. This responsibility is primarily achieved through maintaining the platform’s documentation. Also, experts moderate and curate datasets submitted by ‘Data providers’.

iii) **Technical staff** configures, monitors and maintains the platforms and it’s supporting systems through the Content Management System, the technical staff is able to re-structure the website’s layout, add or remove user interface features and assign security privileges to users. Finally, the technical staffs are responsible for maintaining the infrastructure of the platform and provide technical support to users.

### 4.4.3.2 Audience modeling for open data platform

We now perform the audience classification phase based on the activity-based method (first method) for the open data platform. The groups involved in interacting with the platform are registered users, data seekers, data providers and platform maintenance group (Figure 4-58).

![Audience class hierarchy](image)

The general requirements for the target audience are specified in section 2.1.3. We will provide the tasks related to each user groups. Since our audience group has a hierarchy structure we will start the task description starting from bottom of hierarchy. Next we will elaborate more on the task and detailed description of task by use of use cases. The requirements for these users can be specified as follows:
• **Data seekers**
  1. To submit dataset requests through a wizard.
  2. To have overview of other dataset requests.
  3. To query submitted dataset requests.

• **Data Providers**
  1. To upload datasets.
  2. To modify uploaded datasets.
  3. To query submitted dataset requests.
  4. To reply the dataset requests.

• **Registered users**
  1. To have an user account
  2. To customize the user account based on their preference.
  3. To get notification about new uploads based on their preferences.
  4. To access the most important function of the platform based on their user group.
  5. To provide feedback for datasets.
  6. To communicate with other user (data providers and data seekers).
  7. To register in a community based on different user preferences.
  8. To rate the datasets.
  9. To modify the documentation.
  10. To access the platform’s API without any restriction.
  11. To log in.
  12. To log out.

• **Visitors**
  1. To query dataset.
  2. To search based on the metadata.
  3. To download datasets.
  4. To have an overview of datasets.
  5. To get an overview of available data, indicating which data is available for reuse.
  6. To have an overview of statistics of datasets.
  7. To access real-time data.
  8. To access the documentation.
  9. To access information about the platform.
  10. To access contact information of the project.
  11. To access term of use.
  12. To visualize datasets.
  13. To process datasets.
  14. To access the platform in their preferred language.
  15. To access platform easily from search engines.
  16. To have an easy access to platform.
  17. To register.
  18. To register in the platform by use of social media account.
  19. To access platform by social media account.
Figure 4-59 represents the relation between user groups and tasks. The figure presents that users group at the top of triangle have the maximum access to the platform and they gain more benefits than other user groups. This represents that, for instance Data providers not only are able to perform the task in their own stage, but also they have can conduct the task related to the lower stages. Thus in the figure visitors have limited access and they cannot contribute with the system and other users. Providing easy sign up leads users to maximum access as well as contribution.

4.4.4 Façade Use case
After identifying users and their related task we provide a façade use cases of the platform to represent overall system functions. Figure 4-60 represents main use cases of the platform that is addressed in the above scenarios.
4.5 Chapter Summary

The objective of this chapter is to apply the knowledge obtained in the previous chapter, and to design a good UI for Open Data platform. This Chapter has addressed the third sub question of this research by designing integrated UI mockups for Open Data Platform.

This chapter has three sections. First section is about elicitation and analysis of requirements of Open Data platform by using scenario and use case techniques. Next section focus on designing prototype of User Interface of Open data platform. We have used BPMN diagrams to present the flow of use cases and wireframes to represent the prototype of UI. Finally, last
section provides a summary of the requirements analysis and represents overall description of Open Data Platform.

We have reviewed major characteristics of open data platform, involved users, and responsibilities in the Open Data Platform to be able to design scenarios for requirements analysis. Furthermore, we have designed 6 scenarios in a way that they cover all steps of Open Data process. The scenario section consists of 6 sub-sections, and each sub-section contains the following topics. First, we have described the scenario, and next we presented use case diagram of the scenario. Next, we have described each use case in a table to show basic flow of the use case. Moreover, we have created BPMN diagram for each use case to visualize the flow of each use case. Finally, we have used analysis to design prototype of User Interface of Open Data platform.

Last section reflects results of the analysis. We have described user roles and created audience models. Next, we have categorized users into four main groups; Visitors, Registered users, Data providers, and Data seekers, and mentioned the requirements for each group. According to the audience model, Data providers and Data seekers can use all the functions of the Open Data platform. Finally, a facade use case represents a general overview of the main use cases of the platform.
5 Evaluation

5.1 Introduction

In this chapter, we answer third sub question of this thesis that is “What is the value of the proposed user interface open data?” This is done by conducting descriptive qualitative analysis. First we describe the method of evaluation by describing the participants, measures and procedure of the evaluation. Second we present the results of the evaluation.

5.2 Method

We have done UI evaluation by using a questionnaire and inspecting participants while completing the questionnaire. In this section first we identify a target group and the sample group that is used (sampling). Next we identify the measures for the evaluation and finally we describe the procedure of the evaluation.

The study uses a cross-sectional qualitative, descriptive design on a sample of 16 persons. All were master graduates from Delft University of technology. Cross-sectional studies are a type that involves data collection from a population at one point in time.

5.2.1 Participants

The participants were purposefully selected to evaluate the User Interface. Each participant has been contacted, and a convenient location and time will be determined for filling the questionnaires. Prior to interview, the researcher asked the participants to sign a consent form and completed a demographic form of relevant background data. We have used Purposeful sampling for the recruitment of participants for evaluation by selecting participants according to the needs of the study. The selection criteria for inclusion were graduated students or current students of Delft University of Technology in the age range of 25 to 30 both male and female, from different fields of study.

Individuals completed the questionnaires. Most of the participants were recruited from current students and graduate students of The Delft University of Technology. The evaluators have been chosen from different fields of study, in order to keep diversity in the evaluation process.

The sample consisted of 10 males and 6 females. Mean age was 27. We have asked Individuals to self-identify their occupation from the four categories (Student, Researcher, Business employee and other). Moreover, we have asked Participants to mention their field of
study and country of origin. More details about participants of the evaluation can be find in Appendix B: Evaluators detailed information.

5.2.2 Measures

Participants were administered to select how easy/difficult were to perform each task. Ordinal scale is assigned to measure usability of the proposed UI. They were asked to pick one of the five predefined options which were introduced as; easy, very easy, not easy/ not difficult, difficult, and very difficult. Moreover, in a separate table the organizer inspected each participant while performing each task and indicated whether task performed correctly or not. In addition to the above measures, participants were asked to provide additional comment about each scenario. A sample of the questionnaire is accessible at Appendix A: Questionnaire.

5.2.3 Procedure

The evaluation is conducted by using a questionnaire and direct inspection techniques. Questionnaire is designed in a way that it covers essential functions of the Open Data platform.

Prior to real evaluation process, a pretest is conducted by to test timing and terminology of test. Pretest is conducted by representing the mockups printed on paper and running through the questionnaires. We have modified evaluation process by decreasing numbers of tasks in each scenario, and representing the mockups on screen, Due to the extensive evaluation time.

The questionnaire consists of three scenarios. First two scenarios contain 17 tasks, and last scenario contains 13 tasks. We have expected to perform each scenario approximately takes 15 minutes, and the entire evaluation process takes about 45 minutes to be completed.

Primarily a general description about Open Data platform and its main functionality is given to all participants. Also, we have asked for permission to record sound during the evaluation process. Next participant asked to complete his demographic information and go through the instructions and review the sample of the evaluation procedure. Participants completed the questionnaire individually by presence of the organizer of this thesis.

The following procedure is performed to complete evaluation process.

1. Mockups is shown to participants on the screen by the organizer of the evaluation
2. A task will be given to the participant described in the questionnaire. Next he/she had to figure out how to perform the task.
3. The participant clicks on the part of the mockup that he/she would use to conduct the task.
4. Consequently, organizer observes whether he/she performed the task correctly or not.
5. Finally participant has filled out the questionnaire by selecting one of the options about, how easy or difficult it was for him/her to perform each task.

At the end of each scenario participant is asked to fill out the designated part if he/she has additional comment. Also, the conversation during the evaluation between the organizer and participants will be used for qualitative analysis in the next section. Lastly, participant has confirmed that the information provided by him/her can be published anonymously in this master thesis. Figure 5-1 provides an overview of the evaluation procedure.

**Figure 5-1 Evaluation procedure summary**

5.3 Results

In this section first we describe demographics about participants. Figure 5-2 represents the occupation of participants. 11 researchers, 2 business employee, and 3 students have been participated in the evaluation process. Among those, majority of the researchers has been claimed interacting with UI is easy form them since it is similar to other search databases. However, the other two groups have not previous experience interacting with the UI.

Also, participants selected from different nationalities; Iran, Netherlands, France, and India. Next we will describe the detail of the evaluation for each scenario. Results of the evaluation will be described separately for each scenario. For each scenario, first we point out the majority of responds and thereafter exceptions in each scenario. In the charts related to detail
about each scenario, we have abbreviated not easy/not difficult to E/D. Finally, we mention the user's feedbacks, which was inspected and recorded during the evaluation.

![Figure 5-2 Occupation of participants](image)

In the first scenario, more than 80% percent (14 persons) of participant indicated that Task 10 was very easy and easy to perform. According to 70 percent (13 person) of participants, it was difficult or very difficult to perform Tasks 13 and 14 and 14 out of 16 person did not correctly perform those tasks. Moreover, tasks 3, 10, and 11 according to five participants were difficult to accomplish. Also at least 2 participants did not complete tasks 3, 10, and 11 correctly. Figure 5-3 represents the percentage of participants that evaluate each task in the first scenario, and average for the scenario. More details about evaluation for all the scenarios can be find in Appendix C: Evaluation results.

![Figure 5-3 scenario1: percentage of participants evaluating each task](image)
In the second scenario, all of the participants indicated 11 tasks out of 17 tasks were very easy/ easy to perform. All tasks performed correctly by participants except the following; 2 participants did not perform task 5 and 6 correctly, and 1 participant did not perform task 9 correctly. Also, Tasks 5, 6, 7, 13 were difficult to perform for two participants, and task 3 and 11 was not easy/ not difficult to perform for one participant. Figure 5-4 represents the percentage of participants that evaluate each task in the first scenario, and average for the scenario.

![Graph: Scenario 2](image)

**Figure 5-4 scenario2: percentage of participants evaluating each task**

In scenario 3, all the participants indicated that 8 tasks were very easy and easy to perform. Two participants indicated that 5 tasks were neither easy nor difficult to perform. Tasks 3 and 10 reported by two participants as not easy/not difficult tasks, and tasks 2, 6, and 9 stated by one participant as no easy/, not difficult task. No participant reported any difficult tasks and all the participants accomplished the entire task correctly. Figure 5-5 represents the percentage of participants that evaluate each task in the first scenario, and average for the scenario.
5.3.1 User’s Feedback (Pros and Cons)

In the following, we will describe the feedbacks of users that were recorded during the evaluation process or it was written by the participants. We have characterized them into five categories. The categorization is based on the terms that are mentioned by participants. Feedbacks are coded into the following words; Tab, Color, User’s Profile, and Terms. The last category contains feedbacks that we were not able to place them in each of categories; therefore, ‘general’ title is assigned.

- Tabs

A participant reports download, save, export button was not accustomed for her, and that the reason she found it late. Four participants indicated that different colors for tabs in the same page can guide them through performing the task. Also, Some participants were vague as to how edit a dataset resource, “ well what if I submit the new dataset and I forgot to push the save button?”. All the participants cannot recognize how to extend a dataset “ extend/ quick extend tab distinction could be more clear”.

- Color

A Participant reported that due to the colour of tab/button he could not find how to perform the task. Additionally, Some participants indicated using colour in appropriate way can guide them to distinguish how to perform some tasks, “ the colours of mock-ups can be improved”.

Figure 5-5 scenario3: percentage of participants evaluating each task
• Users’ profile

A participant quoted that “user section it is better to be called social and under social we can have ‘users’ and ‘communities’ tab”. Also another participant recommended the notification preferences can be modified in the notification part in user profile, “this will give better impression that I can control and view my notification from one place”. Finally, A participant reported that representing links without underlines in user profile can carry more consistency in that section.

• Terms

A participant recommended that in using ‘as data provider/ as data seeker’ instead of ‘data provider/ data seeker’ in user’s profile could directly guide him to recognize difference between the mentioned parts. Moreover, Two participants specified understanding what to do is harder than coping with user interface, “using the UI is often easier that implied initially by the questions in the questionnaire”. Lastly, Two participants reported that using ‘help’ (term) instead of ‘how’ gives them better perception of help function.

• General

Two participants also find that ‘limit’ and ‘exclude’ options (tabs) not useful, “this makes me confused”. Furthermore, Some participants indicated that the preview list in dataset page look like an advertisement, “I do not pay attention to that side of screen usually, since I assume that section is designated for advertisements”. As a final point, seven participants after completing the evaluation procedure indicated the user interface was easy to use.

5.4 Chapter Summary

This chapter evaluates the value of the proposed UI for Open Data platform. We have conducted a usability test to evaluate the user friendliness of the mockups. First, we have described the method of the evaluation by specifying the participants, measures, and procedure of the evaluation. Second, we have described the result of the qualitative analysis. The general findings of the evaluation are described in the following. The first scenario was about finding and using dataset. According to respondents, most of the tasks were easy to perform, and 4 tasks were hard to complete. Besides, only 5 tasks out of 17 tasks were not performed correctly by all of the participants. Most indicated that they were not able realizing, how to extend the datasets. They also commented if they could have interact with real User Interface (not mockups) they would be able to recognize how to perform the task by clicking on more information link. The second scenario was about evaluating users’ interactions and finding and submitting a request for dataset. All of the tasks except 4, were easy to perform, and only 3 task out of 17 tasks was not performed correctly by all the participants. Most indicated that the user interface is easy to use. Finally, third scenario was about managing user account and using manuals and tutorials. According to respondents, all of the tasks were
easy to perform, and only 3 tasks out of 17 tasks was not performed correctly by all the participants.

In the last scenario, most of the participants indicated they have got used to work with UI. However, during the evaluation session we discovered a problem about test procedure. The test was highly descriptive and proved to be easy for most of the participants. This indicates that a more abstract task description can provide a more realistic test with more potential for usability errors.

This usability test provided many benefits, but there are few limitations in using this methodology, which should be stated. Firstly, testing is not good representative of the real situation. Also, People behave differently when they know they are being observed. Moreover, the usability test is qualitative; therefore it does not provide large samples of the feedback. This low size of sample makes the conclusion, not trustworthy. The test has the weakness of demanding users. Finding users to contribute in the evaluation process was a difficult task. For this test, the procedure took several days. Organizing schedules for the participants were also time consuming.

Another drawback that we observed during the evaluation process is limited in the way that not all functionality of the UI can be investigated. Also, the discovered problems were only reflecting what was tested in the scenarios. For instance, since we did not ask the participants to utilize 'back' function in any scenario, we were not capable of detecting a problem with that functionality.

The final disadvantage of the test identified during the evaluation dealt with the format of the prototype. In pretest, the prototypes had been on paper, we had a limited ability to identify navigation problems, and because going from one page to another would simply involve changing one paper prototype for another. However, in the real test we have performed the test on the screen, identifying navigational problems was not promising because the participants were not involved in the process of going from one page to another.
6 Conclusion and Discussion

This chapter will conclude our research. First a summary will be given in which the research objective and questions will be evaluated. Next, contribution to research will be stated. Subsequently, some limitations of this research will be highlighted and discussed, as well as reflection on the research. Furthermore, recommendation will be given. Finally, unsolved problems and uncovered topics are mentioned and proposed as future work.

6.1 Conclusion

In this chapter, first we will represent introduction, problem statement, and research objective. Next, we will recall the research question and will respond it by summarizing the theoretical and practical evidence presented in this report.

Open data can help to increase the speed of knowledge and information sharing. Knowledge sharing can provide potential benefits for organizations on several levels. It can deliver more efficient data management, develop information infrastructure, create more integrated services, and improve interactions among the involved organizations. While organizations benefit from open data, but sustaining the information sharing systems is a challenging task, because barriers are attached to social, economic, and political values of organizations. In this research, we have focused on technological barriers of knowledge sharing.

The universal open data platforms facilitate combining data from several data resources, and combining various data types. Basically they are not particularly developed by any public or private organization. Also, these platforms have different types of users. Different users use open data for various purposes and they have different needs.

While designing systems, the interface between human and system is not usually thought about. Instead, the human is often given wide access to arbitrary parts of the system, which results in many failures. Also, User Interfaces designed to hide irrelevant complexity can have a beneficial impact on the user experience, including improving the reliability of the total system. User interface design is the overall process of designing how a user will be able to interact with a system. It aims to enhance the visual, usability and technological qualities of an interface. It adds to the satisfaction of the person using a product or a service.

Traditional user interface design mainly concerns the situation of a single user without thinking about his preferences and background. In current applications the UI should include all aspects of communication between the user and other users of the system. It should also include different users preferences and priorities to ensure user engagement. In the new systems the idea is about supporting the user task, instead of giving access to the underlying system. The newer types of application convey another dimension of complexity
into view. People are collaborating in various ways facilitated by information technology. Collaboration via systems requires special aspects of functionality, both providing facilities for the integration and environments, providing capabilities to manage and coordinate, and providing interaction functionality. Moreover while designing the UI of Open Data platform not only we had considered those facilities, but also we had considered different group of users and their related requirements. This made the situation more challenging because in Open Data platform wide variety of users are involved and various facilities should be considered.

Several impediments restrict users to have easy access. Different impediments for the use of open data are mentioned in literature. The most important impediments are; Data deposit, Data access, and Data use impediments. Data publishers usually only make data available and often do not consider the user perspective. A user perspective aspect that has received less attention is the user interface of open data platforms. Specially, a coherent and integrated set of user interfaces, based on services that can be used to create a user-driven process open data, is lacking. Customized UI helps in providing various services for open data users. The aim of this thesis is to create integrated user interfaces for different groups of users, which, provides services for user Open Data, including the processes of importing, curating, processing and visualizing OD. The goal of UI integration is to create composite applications that control the components’ individual UIs to make richer, composite UI applications.

The focus is on service engineering approach to identify and design user-services, which are necessary for dealing with different types of users of OD. There is no overview of services available and the processes needed for processing open data yet. Current infrastructures do not support processing, visualizing, interpreting of open data. Also, Meta data are vital for the platform in a way that it can link user area of interest to the proper data and related services.

The main research question of this thesis has been formulated as:

**How can user interface improve the process of open data?**

The main research question is divided into four sub-questions. In this concluding chapter, we address the main findings of the thesis by formulating answers to these sub-questions. The first sub question was defined as:

1. What are the impediments of interfaces of open data platforms?

   Based on a literature study on impediments of open data platform a list of impediments has been identified in section 3.3. Major impediments for using open data are identified as lack of infrastructural, technical, socio-cultural, policy-political, and economic supports. To answer this research sub-question, we mostly focus on technological barriers of open data that are related to the user interface of open data platform.

   Technological impediments of open data are divided into three groups. Namely; Data access,
Data Use, and Data deposit impediments. Data access impediments are concerns regarding creating, opening, finding and obtaining data. They are divided into two sub-categories; Availability and Fundability. We have specified 16 impediments and resources related to data access impediments.

Data Use Impediments are about the limitations that restricting the use of open data. These impediments directly hinder the use of open data. They are divided into different categories; Usability, understandability, quality, linking and combining data, comparability and compatibility, and metadata. We have presented each aspect and related impediments to data use. Lastly, Data deposition impediments are about the difficulty of storing data and providing feedback on datasets. Data deposit barriers are divided into two categories; Interaction with data provider and, Opening and uploading. We have described each category and related impediments.

To summarize, there are several impediments identified in the literatures but current literature mostly focused on the impediments from data provider perspective while little attention has been given to the impediments form user’s perspective.

The second sub question was formulated as follows:

2. What are the requirements for user interfaces of open data websites?

To answer the second sub-question, we conducted literature research on the subject of requirements of Open Data platforms that are specifically related to UI. Several requirements are identified for open data platforms. These requirements are divided into five groups; Standardization, API, Materialization, Policies and integration. Standardization enables automatic processing of the data and facilitates automatics access. Materialization guarantees better quality control of datasets to provide better services for the users. Integration is the concern regarding the combining dataset in a way that the uniformity provided among different datasets. API and policies are out of context of this thesis. Moreover, dataset is the important product of Open Data platforms. Thus, we also have identified the requirements for dataset. The main requirements of dataset are identified as; Discoverability, format, validity, quality, and granularity.

Finally, we have gathered a long list of requirements that are divided into twelve groups. These requirements are categorized into following groups; Access, Downloading, Searching, Navigation, Uploading, Data Quality, Analysis of data set, Visualization, Linking and combining data, Collaboration, Support and Help, Feedback requirements. We have identified a total number of 95 requirements, and each requirement is assigned to one of the twelve categories. We have used the requirements to answer third sub-question that aims to design a prototype of UI of Open Data platform.
The third sub question was formulated as follows:

3. How User Interface can increase the use of open data?

To answer this research sub-question, we have applied the knowledge obtained in the previous chapters by designing a good UI for Open Data platform. We have performed requirements analysis by using scenario and use case techniques. Next, we have created BPMN diagram for each use case finally; we have designed UI wireframes. We have addressed the third sub question of this research by designing UI mockups for Open Data platform. We have created UI that covers the Open Data process as well as requirements that we have identified in previous research question.

Primarily, We have reviewed main characteristics of open data platform, involved users, and responsibilities in the Open Data Platform, to be able to design scenarios for requirements analysis. Furthermore, we have designed 6 scenarios in a way that they cover all steps of Open Data process. We have described the scenario, and next we presented use case diagram of the scenario. Following, we have described each use case to show basic flow of the use case. Additionally, we have created BPMN diagram for each use case to visualize the flow of each use case. Lastly, we have created the prototype of User Interface of Open Data platform.

Moreover, we have discussed results of the analysis. We have defined user roles and created audience models. Subsequently, we have characterized users into four main groups; Visitors, Registered users, Data providers, and Data seekers, and mentioned the requirements for each group. Giving the audience model, Data providers and Data seekers can use all the functions of the Open Data platform. Besides, a facade use case is created to provide a general overview of the platform and answer to the sub-question. The main result of this chapter is the UI mock-ups which can represents the main characteristics of the platform.

The fourth sub question was formulated as follows:

4. What is the value of the proposed user interface of open data?

Further research was conducted to discover the value of the proposed UI by asking users to perform a usability test. This research involved a usability test in scenario based environment as well as inspection. A total of 16 usability tests were conducted with participants that varied from researchers, students, and business employees. The usability test were intended at gaining additional vision and also verifying the proposed UI, which were created based on initial literature findings in comparison to reality.

To answer to the fourth sub-question we represent the general findings of the evaluation in the following. The result of evaluation for the first scenario indicated that the, most of the tasks
were easy to perform, and 4 tasks were hard to complete. Besides only 5 tasks out of 17 tasks were not performed correctly by all the participants. For the second scenario most of the participants indicated that the all of the tasks except 4, were easy to perform, and only 3 task out of 17 tasks was not performed correctly by all the participants. Furthermore, the result of third scenario was quite better than other two scenarios; all of the tasks were easy to perform, and only 3 tasks out of 17 tasks was not performed correctly by all the participants. Finally, we conclude from the result of the evaluation that Users’ success rates (performing tasks correctly) with the all the three scenarios were high. Also, according to the Most of the tasks were easy to perform except task related to extending the dataset and user account management. Increasing consistency in different elements of UI can help to solve the mentioned problems.

However, the following shortcomings are realized during the evaluation process. Most of the participants indicated that they were not able to realize, how to extend the datasets. Therefore, an improvement is required regarding dataset extension part. Moreover, for the last scenario, they have stated that the user interface is easy to use, and they have got used to work with UI. Thus, if we had a pretest before the real evaluation process, user could have got familiar with UI, and this could have help to improve the outcome of the first scenario.

Additionally users commented if they could have interact with real User Interface (not mockups) they would be able to recognize how to perform the task by clicking on more information link. This indicates that mockups decrease the level of usability of user interface. Also, Users were very successful with interacting with the user account service. Difficulties did arise when users had to figure out how change preferences for receiving notifications.

Finally, while administering the evaluation, we discovered a problem with the test design. The test procedure was highly descriptive and verified to be easy for the participants. This indicates that a more abstract task description can provide a more realistic test with more potential for usability errors. A detailed limitation of the method is provided in section 6.3.

6.2 Contributions to research and Implications

Open data is a relatively new field, and at the moment little literature is available on requirements of open data especially from the perspective of users. Existing literature emphases on ways to produce and publish open data and mostly on ways to make data available. Open Data users have a crucial role to attain benefits of open data. Considering them in the Open Data process they play an important role to increase the advantages of open data for themselves as well as governments.

Present research in the field of open data not only, focused on the use of open data from data provider perspective but also, there is also literature on data use. This research is focused on applying the knowledge in design and perform usability test. This helps to comprehend user needs and results increasing us of Open Data. This thesis also presents requirements analysis
that is performed in the scenario-based environment. This analysis provides a detailed overview of functionality of Open Data platforms.

The contribution of this study in the field of Information Architecture is designing User Interface for open data platform by analysis of requirements and open data process and applying to design good UI, which can help to gain knowledge about user’s requirements since users are involved in evaluating the prototypes. Since, Open Data is a new developing topic current study in this research area especially regarding the technological topics and Human Computer Interactions (HCI) is limited.

Additionally, this research related to the field of computer science since it is focusing on the requirements engineering, which is a crucial aspect of software engineering. Moreover, theoretical contribution of this research is providing a comprehensive study of current requirements of UI of Open data platforms as well as existing impediments of UI of Open Data platforms. Mean while practical contribution of this thesis in practice is increasing the user participations in deploying Open Data platforms, which can deliver added value for governments and users. Universal Open Data platforms can use the result of thesis to increase usability of their systems. In some cases this can be valuable for the governmental and private organization that publishing data.

The findings in this thesis can have practical implications for Open Data publishers (e.g. governments and private organizations). It gives insight into what are the user’s requirements and how they can improve the UI of their platform and continue development of Open Data platforms. Moreover, in this research we have focused on requirement engineering, which is an important part of the software engineering. Without analyzing the system and user requirements a software cant satisfy the users. Thus, the implication of this research for the field of the computer science is focusing on the subject related to software engineering.

6.3 Limitations and reflections,

In this section, fist limitation of the research is provided, and next, reflections on the research are presented. They were some limitations to research and should be considered when adopting the conclusions presented in this thesis. The limitations are discussed in the following.

Primarily, This research was aimed to be conducted in a limited period. The limited period also has implications on the strength of this research. In the event that there was more time to conduct the research the author would have tried to design interactive prototypes and conduct the evaluation with more participants. By conducting post-test and pre-test validation of the requirements the design could have a better and realistic results. This could have been done before stating the design phase by asking users opinions regarding their needs for using the Open Data platform. Next, this could have been re-validated by asking users to assess the importance of each function of the UI. Moreover, by adding more tasks to evaluate the
findings would have been more accurate, but to decrease the evaluation time we have been limited to certain tasks.

Furthermore, before starting the design phase, there was confusion about choosing a proper methodology to design UI. There was a threat of diving into high-level analysis. Since, the focus of this study was on applying requirements in designing of UI we did not focus on the technological aspect of UI design.

Additional implications that are founded in this research are the findings that were observed during the evaluation process. Firstly, testing was not good representative of the real situation because users do not have freedom in interacting with platform, and they have been restricted to the tasks that they have been asked to perform. Also, People behave differently when they know they are being observed. During the evaluation some participants were worried if their assessment can have negative effect on the final examination. This reflects that if they were not directly observed the evaluation could have been more realistic. Moreover, the usability test is qualitative and result can be used in a way that it describes the characteristics of the design and no mathematical calculation can be used.

Another limitation is about testing and investigation of all of the functionality of the UI. This reflects that the discovered problems were only representing what was tested in the scenarios. For instance, since we did not ask the participants to use 'back' function in any scenario, we were not capable of detecting a problem with that functionality.

The other limitation is regarding the format of the prototype of the evaluation. In the pretest process, the prototypes had been presented on paper, we had a limited ability to identify navigation problems, and because going from one page to another would simply involve changing one paper prototype for another. However, in the real test we have performed the test on the screen, identifying navigational problems was not promising because the participants were not involved in the process of going from one page to another.

Meanwhile, in this research we have performed a post-test evaluation and described the result. It would have been better if we have performed a pre-test evaluation. Also, having a randomized control group could have been helping us to compare the result of evaluation to make our design more accurate and acceptable.

6.4 Recommendation for further research

In this study, based on the analysis of the requirements it can be conclude that different usability issues exist with current Open Data platforms. In summary, Open Data platforms should add functionalities that support collaboration and enhance user interaction. In the following, we list the recommendations to improve UI of the Open Data platform. A detailed list of recommendation can be found in Appendix D. The Open Data platform should provide following possibilities for users:

• Advanced search possibility through the datasets and dataset requests.
• Easy download (dataset, metadata) possibility.
• A recommender system to prevent dataset duplication.
• Advanced dataset upload wizard with possibility of assigning tags.
• Advanced user account management and notification center.
• Simple navigation and help possibility.

We recommend separating and clarify each section of the platform. For this purpose, we recommend assigning following sections to any Open Data platform: Home, Data, Data requests, and Users. This categorization provides consistency and easy navigation for users. Also, ‘Help’ and ‘About’ section should be provided separate from the main sections of the platform. Additionally, the home page of the platform should provide an overview of functionalities of the platform. Providing too many-detailed information in the home page makes user confused.

Therefore facilitating and stimulating the mentioned above improvements should be established in the platform. According to evaluation, these changes will increase the usability of the Open data platform and consequently this results in improving the process of Open data.

This section recommends possible further research related to this field. Fast development of this Open Data provides potentials for future studies. Currently, development of open data platforms is in early phases, and studying users to increase the usability of these platforms can help to create platforms that can ensure user engagements and provide advanced services. More research can be conducted to find what extra functions attract users. Also, more research can be accomplished by focusing on the system and user requirements of Open Data platforms. This can be achieved by conducting interviews and questionnaires.

For the moment, this research is limited to study how can user interface be improved in ENGAGE open data platform. This can be studied further how to improve UI of other Open Data platforms based on the approach of this thesis.

Alternative point of discussion that can start from this research is the improvement of UI of other open data platforms. Different organizations are developing their own platforms this means there is still room for further development. This can be performed by, using the result of the analysis of this research. Also, developing a service that recommends datasets based on user preferences and metadata during the upload process can be a subject for future research. Further topics interesting to research might be:

• What kind of technologies can help to improve the open data platforms?

This can be an interesting research question since it aims to focus on analysis of technologies that can improve the usability of open data. In this thesis we have focused on designing UI mockups but implementing the platform. This will address more real usability problems and support future improvements.

• What kind of technological advancements can increase user engagements?
Meanwhile using open data can bring added value for both governments and users it is necessary to think about the solutions to ensure user’s engagements. According to open data objective, it is important to ensure user’s engagement and their active participation. This can be a subject of research question what kind of technological advancement can guarantee user engagements.

- What features can increase trust amongst data users?

During the evaluation process majority of the participants were concerned about the degree of trust ability of a dataset. A subjective of future research can be increasing the trust amongst data users. A future study can be conducted regarding the elements or features that can help to increase the datasets validity.
REFERENCES


Appendices

Appendix A: Questionnaire
Usability Test for Mockups of Open data platform

Name:                      Gender:       Age:                  Country:

Occupation:
   o Researcher
   o Student
   o Business Employee
   o Other:

Field of Study:             Date:

This questionnaire is part of a master thesis which aims to design a user interface for open data platforms. For this purpose, we have designed three scenarios and mockups related to each scenario. You are kindly requested to evaluate how easy it is to perform each task. The following actions should be performed:

1. Mockups will be shown to you on the screen by the organizer of the evaluation (Armin Parnia).
2. A task will be given to you described in a questionnaire. You have to figure out how to do the task.
3. You click on the task on the part of the mockup that you would use to conduct the task.
4. Fill in the questionnaire how easy or difficult it was for you to perform each task.

The following example makes mentioned above steps more clear.
This questionnaire consists of three scenarios and each scenario contains about 15 tasks and it approximately takes 15 minutes to perform each scenario. The entire test takes about 45 minutes to be completed.

- Please note that the mockups cannot be used for performing any real tasks. Therefore please imagine each mockup as a real web page and evaluate how easy or difficult it can be to perform each task.
- While performing each task, please provide feedback if you do not understand any task or if you have any suggestions. This will help us considerably to improve the user interfaces.
- By completing this questionnaire, you agree that the information you have provided is allowed to be used in a master thesis.

I would like to thank you for your cooperation and your valuable time that you spent to help me to evaluate the design.
**Scenario 1: Searching, downloading, extending, visualizing, curating, providing feedback, editing and linking datasets.**

<table>
<thead>
<tr>
<th>Step</th>
<th>Very difficult</th>
<th>Difficult</th>
<th>Not difficult/Not easy</th>
<th>Easy</th>
<th>Very easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. View home page of open data platform. Insert a keyword and search for a dataset that is interesting to you.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Sort search results based on your preference.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Filter the search results by selecting Geography for UK region and limit search results.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Save search results in your user account.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Export the search result in your preferred format.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Select three datasets and analyze the search results (compare them based on metadata).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Analyze search results based on different fields view table based on country or maintainer.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Download a preferred format dataset, directly from the search page.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Select a dataset that you found and view the dataset page.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Have a preview of your previous viewed datasets.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If you have any other comment that you think is not dealt with in the above scenario please write it down in the following section.
**Scenario 2: Using the platform for interaction with other members. Searching dataset request and providing feedback and place a new dataset request.**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
<th>Very difficult</th>
<th>Difficult</th>
<th>Not difficult/Not easy</th>
<th>Easy</th>
<th>Very easy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>View the home page of the open data platform. Go to the user section and subsequently ‘User profile’ to see the active users of the platform.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>2.</td>
<td>Search through the users, use sort functions and try to find a user you know by searching the names. Sort users by the data provider.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>3.</td>
<td>Find an open data user and select that user.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>4.</td>
<td>Send a message to the user that you are in his/her user profile.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>5.</td>
<td>In the data provider part view summary of it. Next sort the ‘datasets’ based on the highest votes.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>6.</td>
<td>In data seeker part select requests to view detailed information about the dataset requests.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>7.</td>
<td>Go to user section and subsequently ‘Communities’ section and view different groups.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>8.</td>
<td>View the groups that you are member of on top of the list, and then leave a group.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>9.</td>
<td>Imagine that you do not find a group, now create a new group. Use the help function to learn to how create a new group.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
10. Go to dataset request section.

11. Browse through older requests. Customize search result by selecting a request type, category or tag and then sort the results.

12. Specify the customized result list and detail about each request (Votes, answers, views).

13. Find and select a dataset request.

14. Leave a reply in feedback section. And report a reply to if it contains inappropriate contents.

15. Add a new dataset request.

16. Fill out the fields to submit a request and subsequently have a look at the requests that were published by other that is similar to your request.

17. Preview your request and publish or delete it.

If you have any other comment that you think it is not dealt with in above scenario please write down in the following section.
**Scenario 3: Uploading and editing a dataset, managing user account, and getting information about manuals and tutorials.**

<table>
<thead>
<tr>
<th>Step</th>
<th>Activity</th>
<th>Difficulty Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>View the home page of open data platform. Go to the ‘Data’ section and consequently to the ‘add dataset’ section.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Select each of fields in the upload wizard steps to get more information.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Confirm that your dataset is ready for upload. And fill out the fields in dataset upload wizard and submit your dataset.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Preview the uploaded dataset and then publish it.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Edit the dataset resources. Add a new dataset.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Fill out the fields in the new pane and submit it.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Edit metadata. Edit one field and save the result.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Select your user ID on top of the page.</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>View your notifications. Select a notification and view the correspondence task.</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Edit your preference by adding a new email address.</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Go to the help section from the main page of the open data platform.</td>
<td></td>
</tr>
</tbody>
</table>
If you have any other comment that you think is not dealt with in the above scenario please write it down in the following section.

- I agree that the information provided by me is allowed to be used and published anonymously in the master thesis of Armin Parnia.

Signature:                          Date:
### Appendix B: Evaluators detailed information

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Age</th>
<th>Occupation</th>
<th>Study field</th>
<th>Country of Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>28</td>
<td>Researcher</td>
<td>Power System Engineer</td>
<td>Iran</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>26</td>
<td>Researcher</td>
<td>Management of Technology</td>
<td>Iran</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>30</td>
<td>Business employee</td>
<td>Computer science</td>
<td>Netherlands</td>
</tr>
<tr>
<td>4</td>
<td>M</td>
<td>27</td>
<td>Business employee</td>
<td>Mechanical Engineering</td>
<td>Iran</td>
</tr>
<tr>
<td>5</td>
<td>F</td>
<td>25</td>
<td>Student</td>
<td>Sustainable Energy</td>
<td>India</td>
</tr>
<tr>
<td>6</td>
<td>M</td>
<td>26</td>
<td>Researcher</td>
<td>Power System Engineer</td>
<td>France</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>25</td>
<td>Student</td>
<td>Computer science</td>
<td>Netherlands</td>
</tr>
<tr>
<td>8</td>
<td>M</td>
<td>25</td>
<td>Researcher</td>
<td>System Engineering, Policy Analysis Management</td>
<td>Iran</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>28</td>
<td>Researcher</td>
<td>Microelectronics</td>
<td>Netherlands</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>26</td>
<td>Student</td>
<td>Management of Technology</td>
<td>Netherlands</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>28</td>
<td>Researcher</td>
<td>Transport &amp; Planning</td>
<td>Iran</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>25</td>
<td>Researcher</td>
<td>Microelectronics</td>
<td>Iran</td>
</tr>
<tr>
<td>13</td>
<td>F</td>
<td>28</td>
<td>Researcher</td>
<td>Transport &amp; Planning</td>
<td>Netherlands</td>
</tr>
<tr>
<td>14</td>
<td>M</td>
<td>27</td>
<td>Researcher</td>
<td>Sustainable Energy</td>
<td>India</td>
</tr>
<tr>
<td>15</td>
<td>M</td>
<td>28</td>
<td>Researcher</td>
<td>System and Control</td>
<td>Netherlands</td>
</tr>
<tr>
<td>16</td>
<td>F</td>
<td>27</td>
<td>Researcher</td>
<td>Engineering Policy Analysis</td>
<td>Iran</td>
</tr>
</tbody>
</table>
### Appendix C: Evaluation results

#### Scenario 1:

<table>
<thead>
<tr>
<th>Task</th>
<th>Very easy</th>
<th>Easy</th>
<th>Not easy/Not difficult</th>
<th>Difficult</th>
<th>Very Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>71%(13)</td>
<td>12%(2)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T2</td>
<td>71%(13)</td>
<td>19%(3)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T3</td>
<td>50%(8)</td>
<td>25%(4)</td>
<td>25%(4)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T4</td>
<td>57%(9)</td>
<td>31%(5)</td>
<td>6%(1)</td>
<td>6%(1)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T5</td>
<td>63%(10)</td>
<td>31%(5)</td>
<td>0%(0)</td>
<td>6%(1)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T6</td>
<td>57%(9)</td>
<td>25%(4)</td>
<td>12%(2)</td>
<td>6%(1)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T7</td>
<td>71%(13)</td>
<td>19%(3)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T8</td>
<td>63%(10)</td>
<td>31%(5)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T9</td>
<td>88%(14)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T10</td>
<td>39%(6)</td>
<td>12%(2)</td>
<td>31%(5)</td>
<td>12%(2)</td>
<td>6%(1)</td>
</tr>
<tr>
<td>T11</td>
<td>44%(7)</td>
<td>25%(4)</td>
<td>31%(5)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T12</td>
<td>71%(13)</td>
<td>19%(3)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T13</td>
<td>0%(0)</td>
<td>12%(2)</td>
<td>31%(5)</td>
<td>44%(7)</td>
<td>12%(2)</td>
</tr>
<tr>
<td>T14</td>
<td>19%(3)</td>
<td>38%(6)</td>
<td>12%(2)</td>
<td>31%(5)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T15</td>
<td>69%(11)</td>
<td>31%(5)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T16</td>
<td>71%(13)</td>
<td>6%(1)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T17</td>
<td>94%(15)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T2</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T3</td>
<td>88%(14)</td>
<td>12%(2)</td>
</tr>
<tr>
<td>T4</td>
<td>82%(13)</td>
<td>18%(3)</td>
</tr>
<tr>
<td>T5</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T6</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T7</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T8</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T9</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T10</td>
<td>82%(13)</td>
<td>18%(3)</td>
</tr>
<tr>
<td>T11</td>
<td>94%(15)</td>
<td>6%(1)</td>
</tr>
<tr>
<td>T12</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T13</td>
<td>6%(1)</td>
<td>94%(15)</td>
</tr>
<tr>
<td>T14</td>
<td>18%(3)</td>
<td>82%(13)</td>
</tr>
<tr>
<td>T15</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T16</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T17</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
</tbody>
</table>
Scenario 2:

<table>
<thead>
<tr>
<th>Task</th>
<th>Very easy</th>
<th>Easy</th>
<th>Not easy/Not difficult</th>
<th>Difficult</th>
<th>Very Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>69%(11)</td>
<td>31%(5)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T2</td>
<td>88%(14)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T3</td>
<td>88%(14)</td>
<td>6%(1)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T4</td>
<td>100%(16)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T5</td>
<td>25%(4)</td>
<td>63%(10)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>6%(1)</td>
</tr>
<tr>
<td>T6</td>
<td>57%(9)</td>
<td>25%(4)</td>
<td>6%(1)</td>
<td>6%(1)</td>
<td>6%(1)</td>
</tr>
<tr>
<td>T7</td>
<td>69%(11)</td>
<td>19%(3)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T8</td>
<td>94%(15)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T9</td>
<td>94%(15)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T10</td>
<td>88%(14)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T11</td>
<td>88%(14)</td>
<td>6%(1)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T12</td>
<td>75%(12)</td>
<td>25%(4)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T13</td>
<td>88%(14)</td>
<td>12%(2)</td>
<td>6%(1)</td>
<td>6%(1)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T14</td>
<td>69%(11)</td>
<td>31%(5)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T15</td>
<td>94%(15)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T16</td>
<td>94%(15)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T17</td>
<td>94%(15)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T2</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T3</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T4</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T5</td>
<td>88%(14)</td>
<td>12%(2)</td>
</tr>
<tr>
<td>T6</td>
<td>88%(14)</td>
<td>12%(2)</td>
</tr>
<tr>
<td>T7</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T8</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T9</td>
<td>94%(15)</td>
<td>6%(1)</td>
</tr>
<tr>
<td>T10</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T11</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T12</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T13</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T14</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T15</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T16</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T17</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
</tbody>
</table>
Scenario 3:

<table>
<thead>
<tr>
<th>Task</th>
<th>Very Easy</th>
<th>Easy</th>
<th>Not easy/Not difficult</th>
<th>Difficult</th>
<th>Very Difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>75%(12)</td>
<td>25%(4)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T2</td>
<td>44%(7)</td>
<td>50%(8)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T3</td>
<td>69%(11)</td>
<td>19%(3)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T4</td>
<td>94%(15)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T5</td>
<td>75%(12)</td>
<td>25%(4)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T6</td>
<td>88%(14)</td>
<td>6%(1)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T7</td>
<td>88%(14)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T8</td>
<td>88%(14)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T9</td>
<td>63%(10)</td>
<td>31%(5)</td>
<td>6%(1)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T10</td>
<td>63%(10)</td>
<td>25%(4)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T11</td>
<td>81%(13)</td>
<td>19%(3)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T12</td>
<td>88%(14)</td>
<td>12%(2)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T13</td>
<td>81%(13)</td>
<td>19%(3)</td>
<td>0%(0)</td>
<td>0%(0)</td>
<td>0%(0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T2</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T3</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T4</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T5</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T6</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T7</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T8</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T9</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T10</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T11</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T12</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
<tr>
<td>T13</td>
<td>100%(16)</td>
<td>0%(0)</td>
</tr>
</tbody>
</table>
Appendix D: Detailed recommendation for Open Data platforms

Open data platforms can facilitate the following options to increase usability of their websites:

- To search dataset in an advanced way: by customizing the search criteria and sorting search results.
- To export and save search result.
- To analyze search result based on metadata.
- To download dataset with different formats directly from the search page.
- To preview the latest viewed dataset.
- To upload dataset through a wizard that can provide progress and help.
- To get a recommendation about similar dataset while uploading a dataset.
- To be able to assign tags related to the dataset.
- To preview the uploaded dataset before publishing dataset.
- To view, edit, and download dataset resources and metadata as the dataset.
- To rate and report user feedbacks.
- To have advanced user account and manage different functionalities of user account.
- To access and preview to all user’s interaction from the user profile.
- To submit a dataset request and preview the request before submitting the dataset.
- To be able to assign tags related to the request.
- To have advanced notification center.
- To get guidelines before uploading a dataset. This feature can increase data consistency in the platform.
- To access documentation and tutorial from help tab.
- To provide Frequently Asked Question section in the platform.
- To view navigational pane in all part of the platform.