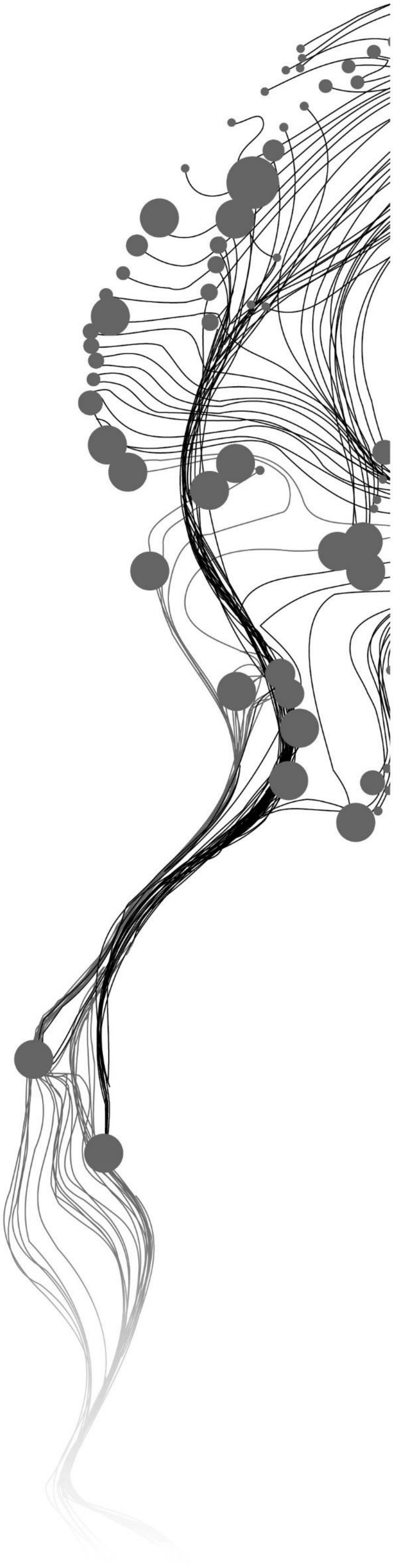
SEMANTIC PREVIEWS OF GEODATA

CHRISPER IMMACULATE ABUNI
February, 2017

SUPERVISORS:
Dr. Ir. R.L.G. Lemmens
Dr. C.P.J.M. van Elzakker



SEMANTIC PREVIEWS OF GEODATA

CHRISPER IMMACULATE ABUNI
Enschede, The Netherlands, February, 2017

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Dr. Ir. R.L.G. Lemmens
Dr. C.P.J.M. van Elzakker

THESIS ASSESSMENT BOARD:
Prof. Dr. M.J. Kraak (Chair)
Dr. C. Stasch (External Examiner, 52°North Initiative for Geospatial Open Source Software GmbH)

ABSTRACT

There is a lot of geodata that is made available through geoportals. With the large amount of data comes the challenge of finding data that is relevant for a particular purpose. Worse still is the problem of determining the suitability of the dataset for the particular purpose. People typically make use of metadata to assess the fitness for use of a dataset. However, the contents and the meaning of the dataset are usually not made clear to the user in the metadata. People need a more efficient way of assessing datasets for fitness of use before they have access to the dataset. This research proposes the use of semantic previews to make the meaning and contents of datasets explicit to the user.

First, to understand how current systems are being used, existing geoportals were evaluated. Requirement analysis was also done to determine what the users of these datasets are interested in when assessing the fitness for use of datasets. The results of the evaluation of geoportals and requirement analysis were used to create a prototype that uses semantic previews. The previews were showed on the dashboard as data tree structure, real-world representation, completeness of data attributes and map rendering. The dashboard also provided a link to workflow diagrams.

The prototype was evaluated by doing a usability test with test persons from ITC. The methods used were questionnaires, creation of tasks for the test persons, thinking aloud, audio and video recording. There were a total of 7 test persons and they carried out tasks using a geoportal and the prototype designed as a result of the research.

The results showed that the geoportal without the semantic previews did not make the data attributes of the dataset clear enough to the user as most of the test persons spent a lot of time trying to understand the contents of the dataset. Meanwhile with the prototype, almost all of them spent a shorter time with the prototype. The element of the prototype which was designed to enable the user to understand the completeness of the dataset was ignored by most of the test persons which shows that a more obvious solution should have been implemented. In conclusion, semantic previews have the potential to be useful in the field of geospatial data; however, the previews to be shown need to be carefully considered to reflect the contents of the dataset. The prototype could be incorporated into a catalogue service to provide the users with information that enables them to understand datasets by just looking at previews.

ACKNOWLEDGEMENTS

I would like to thank the Lord God Almighty for enabling me to get this far, without him, I really do not think I would have reached this far. My appreciation goes out to NFP –Nuffic for sponsoring me and granting me this opportunity to further my studies.

My sincere appreciation goes out to my first supervisor, Dr. Ir. R.L.G. Lemmens. Thank you for patience, encouragement and your guidance. To my second supervisor, Dr. C.P.J.M. van Elzakker, I am so grateful for your all your feedback and guidance as well. You were both great team to work with.

I appreciate all those who volunteered to take part in the usability testing, your contribution is very much appreciated.

My parents, my siblings especially my brother who always encouraged and prayed for me, I am so grateful. All my aunties, uncles and relatives, who checked in on me, thank you very much.

To my friends, thank you for walking this journey with me.

Finally to all my classmates, thank you for the experiences we all shared here.

TABLE OF CONTENTS

1. INTRODUCTION 1
   1.1. Motivation and problem statement 1
   1.2. Research identification 2
   1.3. Innovation 3
   1.4. Methodology 3
   1.5. Thesis structure 6
2. THEORETICAL FRAMEWORK 7
   2.1. Assessing fitness for use of a dataset 7
   2.2. Visualization 9
   2.3. Previews 11
   2.4. Dashboards 12
   2.5. Semantic Enrichment 13
   2.6. Geospatial datasets 14
   2.7. Users of geospatial data 15
   2.8. Conclusion 16
3. EVALUATION OF GEOPORTALS 17
   3.1. Framework for evaluation of geoportals 17
   3.2. Procedure for the evaluation of existing geoportals 17
   3.3. Results of evaluation 18
   3.4. Summary of results 19
   3.5. Conclusion 19
4. BUILDING SEMANTIC PREVIEWS 21
   4.1. User Centred Design 21
   4.2. Requirement analysis 22
   4.3. Design 25
   4.4. Prototyping 29
   4.5. Conclusion 35
5. USABILITY TESTING 37
   5.1. Objective of the usability testing 37
   5.2. Test preparation 37
   5.3. Test persons for usability testing 40
   5.4. Pilot test 40
   5.5. Scheduing of the tests 40
   5.6. Test set-up and execution 40
   5.7. Test results 40
   5.8. Conclusion 46
6. DISCUSSION 47
   6.1. Evaluation of geoportals 47
   6.2. Development of the prototype 47
   6.3. Usability testing 48
   6.4. Metadata handling and standards 49
   6.5. Conclusions 50
7. CONCLUSIONS AND RECOMMENDATIONS 51
   7.1. Overview 51
   7.2. Answers to the research questions 51
   7.3. Recommendations 53
REFERENCES 55

APPENDICES....59
Appendix 1 59
Appendix 2 60

LIST OF FIGURES

Figure 1: Metadata for a dataset in the National Dutch geoportal (http://www.nationaalgeoregister.nl) 2
Figure 2: The overview of the methodology and research objective achieved at each stage 4
Figure 3: The relationship between different concepts are related in this research...7
Figure 4: Graphical definition for fitness for use 8
Figure 5: An example hierarchy of data quality indicators (adopted from Devillers et al., 2002) 9
Figure 6: The interfaces and the contents in the atlas metaphor prototype are organized in accordance with the atlas information structure (adopted from Aditya & Kraak, 2006) 10
Figure 7: Figure 7(a) shows a Geolabel for a dataset with only six elements that comply with data management principles (DMP). 7(b) shows the DMP labels. 11
Figure 8: Figure 8(a) showing preview of the contents of the folders and Figure 8(b) shows the folders without preview 11
Figure 9: The Dublin dashboard (adopted from McArdle & Kitchin, 2016) 13
Figure 10: Figure showing part of OSMonto visualized with WebVowl 15
Figure 11: Figure showing the processes involved in UCD (adopted from Elzakker & Wealands, 2007). 22
Figure 12: Steps followed for requirement analysis 22
Figure 13: Graphical representation of the user profile 24
Figure 14: The relationship between the areas of interest, the question and the design solution in the prototype 26
Figure 15: Tag cloud from Inspire geoportal (http://inspire-geoportal.ec.europa.eu/discovery/)............. 27
Figure 16: Links within the different elements of the prototype.. 28
Figure 17: Figure showing a secondary road in Germany (left) and in Uganda (right) (www.wiki.openstreetmap.org, 2017) 29
Figure 18: The landing page of the workflow website... 31
Figure 19: Figure showing a workflow diagram in the website.. 32
Figure 20: The components of the data page... 32
Figure 21: The interface of the data page 33
Figure 22: Figure showing the data flow between the software components 33
Figure 23: Figure showing the steps followed for the usability testing.. 37
Figure 24: Graph showing the total time taken by each of the participants to complete the test................ 42
Figure 25: Figure showing the time taken to complete each of the tasks. 43

LIST OF TABLES

Table 1: Table showing the results of evaluation of geoportals 18
Table 2: Table showing methods and techniques used in UCD (adopted from Delikostidis, 2011) 21
Table 3: Table showing the characteristics of the user group 23
Table 4: Table showing the background information of the test participants 41
Table 5: Table showing data sources the participants used to obtain spatial data 41

1. INTRODUCTION

1.1. Motivation and problem statement

The emergence of the World Wide Web led to the development of web-based systems known as geoportals or geo-web portals. Geoportals enable users to access, view and manipulate various forms of data over internet. The data forms include; geospatial data (geodata), maps and web services among others (Athanasis et al., 2009). According to Vockner & Mittlböck (2014), geoportals are the main gateways to geographic information.

With the advancement in communication, technology and need for information sharing, there is a large amount of geo-data and services made available through geoportals in different formats and from various sources. This makes it difficult for users to quickly find the information they are looking for in the vast amount of data (Bogdanović, Stanimirović, & Stoimenov, 2015).

To facilitate the ease with which relevant information is found, geoportals typically make use of metadata catalogues. Metadata is defined as information that describes the characteristics of a dataset based on predefined standards (Maguire & Longley, 2005). Metadata catalogues contain a collection of published metadata and users obtain the relevant information by querying these catalogues (Athanasis et al., 2009). In order to find datasets that can be useful to them, users still have to manually browse through the search results thoroughly to determine which datasets are fit for use (Aditya & Kraak, 2006). This is not very effective and efficient especially with a large number of search results. In addition, metadata used in geoportals do not provide enough information about the dataset but for the user to clearly understand the contents without accessing the dataset (Klien, 2007). An example of metadata from a geoportal is shown in Figure 1. Information like for example data attributes are not explicitly defined in the metadata yet users always want to get as much information as possible at a glance as a kind of preview of the data or service that they need.

The definition of "Preview" by Merriam-Webster (2015) is: “to see beforehand”. Previews therefore enable users to understand the contents of a dataset before they actually access it. For spatial data attributes, previews can be presented as images, video clips among other options depending on the contents of dataset. In current geoportals, previews are usually presented as overview maps and thumbnails; however they do not clearly present the semantics and contents of a dataset.

Semantics has to do with meanings of words in a certain situations. One of the definitions given for “Semantics” by Merriam-Webster (2017) is: “the meanings of words and phrases in a particular context” which was the definition adopted for this research.

For this research, semantic preview is defined as: a graphical representation of contents about or within a dataset presented as previews that gives the potential user of the dataset useful information (in relation to their perception and experience) about the dataset like the contents, potential use or lineage of the dataset.

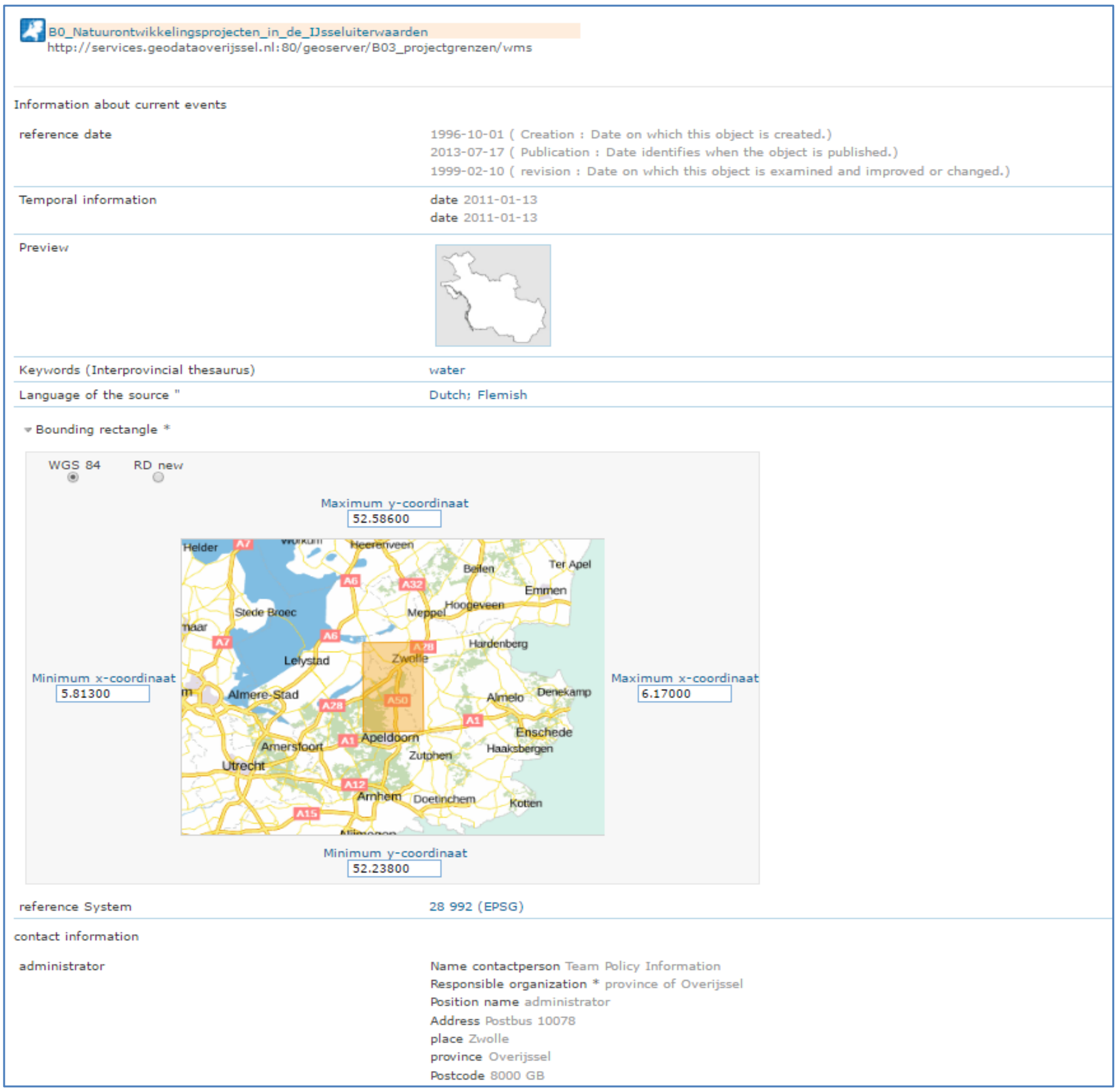


Figure 1: Metadata for a dataset in the National Dutch geoportal (http://www.nationaalgeoregister.nl)

1.2. Research identification

The main objective of this research is to develop a method that uses previews for spatial data users to quickly understand the contents of a geo-dataset and determine its fitness for use. This is broken down into sub objectives as listed in Section 1.2.1.

1.2.1. Research objectives

1. To evaluate existing geoportals in terms of users' ease in determining the fitness for use of a dataset
2. To identify users and perform user requirement analysis
3. To design a prototype that quickly makes the meaning of data attributes explicit to the user and evaluate it
4. To suggest improvements of current practice in metadata handling (services and standardisation) based on the results of this research.

1.2.2. Research questions

Questions for objective 1

1. What are the strengths and weaknesses of existing geoportals in terms of discovery of relevant data?
2. How can these geoportals be made more efficient and effective for discovery of relevant data?
3. What can be the role of (semantic) previews?

Questions for objective 2

1. Who are the users that the research is targeting and what are their requirements?
2. What kind of geo-datasets are of interest to these users?
3. How do users make use of previews?

Questions for objective 3

1. How can the previews be semantically enriched, e.g., by using Semantic Web technology?
2. What tools are suitable for implementing comprehensive previews?
3. What are the strengths and weaknesses of the prototype?
4. What can be done to improve the prototype?

Questions for objective 4

1. How can the prototype be integrated into a catalogue service?
2. How can metadata standards be improved, based on this research?

1.3. Innovation

This research proposes a new design of geoportals that makes use semantic previews to show the actual contents of the datasets and enable the users to quickly understand them.

1.4. Methodology

This section gives an overview of the steps followed in the research to answer the research questions and achieve the research objectives. The literature review was the initial stage of the methodology. Other processes included; evaluation of geoportals, requirement analysis, prototype development and usability testing. Figure 2 shows the graphical representation of the methodology and the different stages at which the research objectives were achieved. More details for each of the stages are described in subsequent chapters.

Literature review: this was a continuous process throughout the research period. This is covered in Chapter 2.

The research followed the User Centred Design (UCD) approach that involves requirement analysis, prototyping and Usability testing.

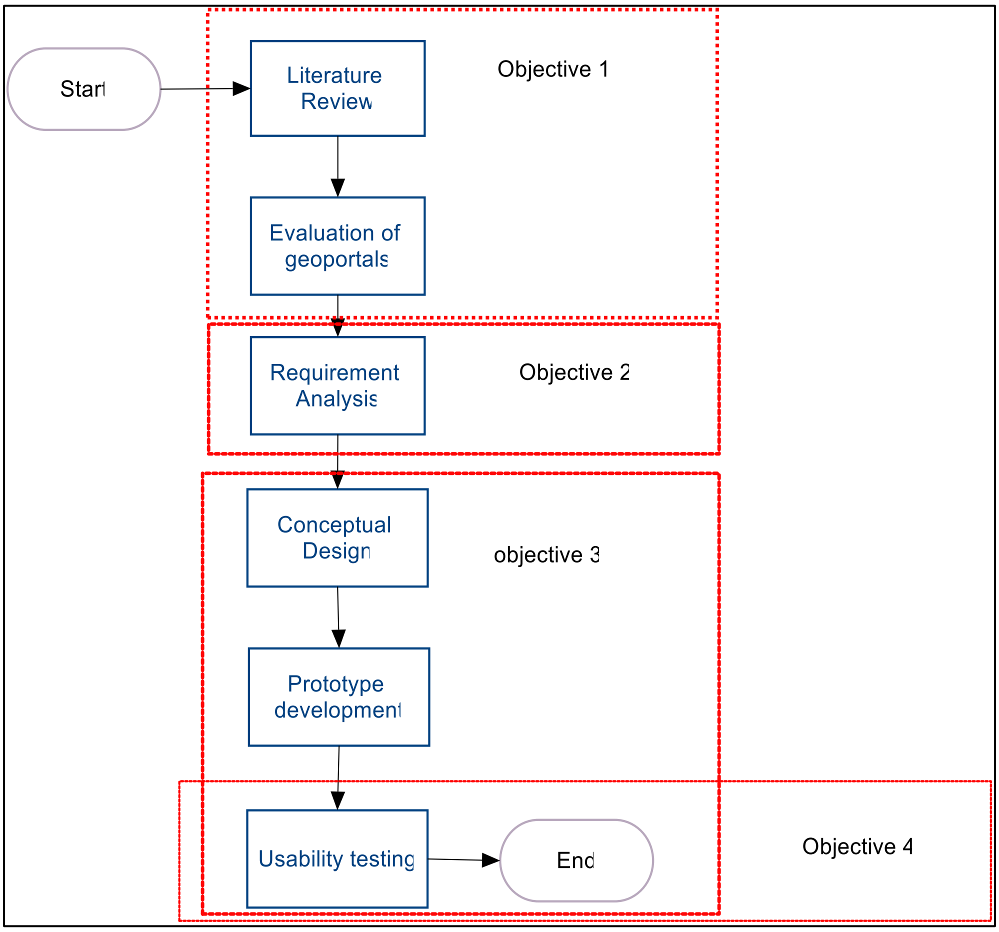


Figure 2: The overview of the methodology and research objective achieved at each stage

1.4.1. Evaluation of geoportals

The first objective of the research was to evaluate the existing geoportals in terms of assessing the fitness for use of a dataset. To achieve this objective, the existing geoportals had to be studied to determine how users are currently utilising them to determine the suitability of a dataset for a particular purpose. It also helped to determine whether existing geoportals made use of previews and if they did, whether the previews were sufficient for helping the users assess the fitness for use of the datasets. Based on literature, there are no standard frameworks developed for the evaluation of geoportals. Different frameworks have been used in various researches to achieve different goals.

1.4.2. User-centered design

The user-centred design (UCD) approach was adopted for this research. UCD is defined by Abras et al. (2004) as the design process in which the final users determine the design. UCD approach was followed because it improves the usability of the product since the focus is user satisfaction. Gulliksen et al. (2003) describes 12 key principles in the User-Centred System Design. Some of these principles include;

- Focus on the user which involves understanding needs, tasks and characteristics of the user
- Involvement of the user throughout the development stages of the system
- Presentations of designs in ways that can easily be understood by the users
- Prototyping and evaluation of designs with representatives of the target users

Since users are the focus for UCD, there are various methods that could be selected to involve them (Abras et al., 2004);

- Interviews and questionnaires in the early stages of the design to collect information about the users' experiences, expectations and how what tasks they perform with the product. It enables one to gain insight on what to include in the product.
- Focus group discussions, where different stakeholders meet and come up with their requirements. This is also done early in the design phase.
- On site observation in the early phases of the design phase, where information is collected about the environment in which the product will be used.
- Role playing and simulation; this can be done either at the beginning or mid-way through the design process.
- Usability testing; this is done at the final stage of the design cycle. It measures the usability of the product. Sometimes it may be done more than once to achieve a desirable product.
- Interviews and questionnaires at the final stage of the design phase to determine user satisfaction and find ways in which it can be improved.

1.4.2.1. Requirement Analysis

Techtarget.com (2017) defines Requirement Analysis as the approach taken to discover the needs of a user. It was done to achieve the second objective of the research which was to identify users and perform user requirement analysis. Requirement Analysis is important because it ensures that the user is considered from the initial design phase. To carry out Requirement Analysis, different methods could have been used, for example; use of interviews, questionnaires, focus-group discussions (brain-storming), on-site observation among others as shown in Section 4.1 (Table 2).

The initial target users of this research were researchers from the AfriAlliance project. The project deals mainly with innovation and technological solutions for water and climate change challenges (UNESCO-IHE, 2016). These researchers are interested in datasets that are water related because their interest is to find ways of providing safe drinking water to people especially in African Countries. Since the researchers were scattered in different parts of the world, the use of an online questionnaire was chosen to determine their requirements. Considering the situation, it was chosen because it was assumed that it would be time saving and accessible to many of the people representing the target users. A questionnaire was therefore developed using SurveyMonkey a free online tool for creating questionnaire.

This unfortunately was not successful because of the targeted people did not respond to the questionnaire. Because of this, a different case study was chosen with a fictitious use case scenario. The target users were still researchers but in the field of road safety. The requirement analysis was determined by defining the tasks and questions that they would be interested in. The use case scenario, user profile and required datasets are described in Chapter 4.

1.4.2.2. Prototyping

This was done to achieve the first part of the third objective of the research which was to design a prototype that quickly makes the meaning of data attributes explicit to the user and evaluate it. To develop the prototype, a conceptual design was created which took into consideration the results from the evaluation of geoportals and requirement analysis. Different tools and platforms were used for the implementation of the prototype and Chapter 5 describes them in more detail.

1.4.2.3. Usability testing

Bevan et al. (2015) defined usability as: "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". According to the same authors, effectiveness has to do with accomplishing a goal; efficiency is related to saving time and satisfaction is about the desire to use the system.

Usability testing enables one to understand whether the system can be used by the people for which it was designed. It is the best method of determining how actual users experience an application or website (Usabilityfirst.com, 2015). It is usually done at the last stage of the design process and can be more than once iteratively to achieve a desired outcome (Abras et al., 2004).

1.4.2.4. Methods used for usability testing

There are different methods that can be used to carryout usability testing depending on the product being tested. The result of the research that was being tested was a prototype in the form of a web application that shows previews of geospatial data. Pre-defined tasks were assigned to the test persons in combination with other usability testing methods, these are described below;

- Thinking aloud, according to Anandhan et al. (2006) is the prominent and widely used usability testing method. It was chosen because it enables the tester to know exactly what the user is thinking of the product which is the prototype in this case. It was coupled together with tasks assigned to the users.
- Screen recording was done to be able to see the mouse movements of the test person and be able to understand better how the user was interacting with the application. This provided more insight into the thoughts of the user.
- Audio and Video recording helped to capture the reactions of the test persons which also provided additional information about their experiences with the application. It also helped to note the amount of time the test person took to complete all the tasks.

The details of how the usability test was carried out are explained in Chapter 5.

1.5. Thesis structure

Chapter 1 is the introduction of the thesis which consists of the motivation and problem statement, the research identification, the innovation of the research and an over of the methodology followed for the research. **Chapter 2** is the Literature review which looks at the different studies that have been done that that are related to this research and the areas where this research builds on. In **Chapter 3** is the Evaluation of geoportals where the process followed and the results of the evaluation are covered. The steps followed to build a prototype with semantic previews are covered in **Chapter 4** and they include; requirement analysis and prototyping. The last step was the evaluation of the prototype which is covered in **Chapter 5. Chapter 6** contains the discussion and **Chapter 7** the conclusions and recommendations. Finally, the Appendices are shown.

2. THEORETICAL FRAMEWORK

This Chapter explains different concepts that provide the background of this research. The focus of the research is on the use of previews to show the contents of spatial data. There is not a lot of research that has been done in this field yet and therefore there is not a lot of related work presented here. However, different concepts have been borrowed from other fields. The first section explains some of the methods that are available to assess the fitness for use of datasets. The concept of visualization is then introduced and narrowed down to geovisualization. An argument for the use is preview with geospatial data is provided and examples of fields where they are currently being used are given. This Chapter also covers some literature about dashboards. Semantic enrichment is explained to give background in the use of semantics and clarify the meaning of semantic previews which is important for this research. Figure 3 shows how the different concepts in this research are related. With the different concepts are explained, Chapter 3 looks at evaluation of geoportals.

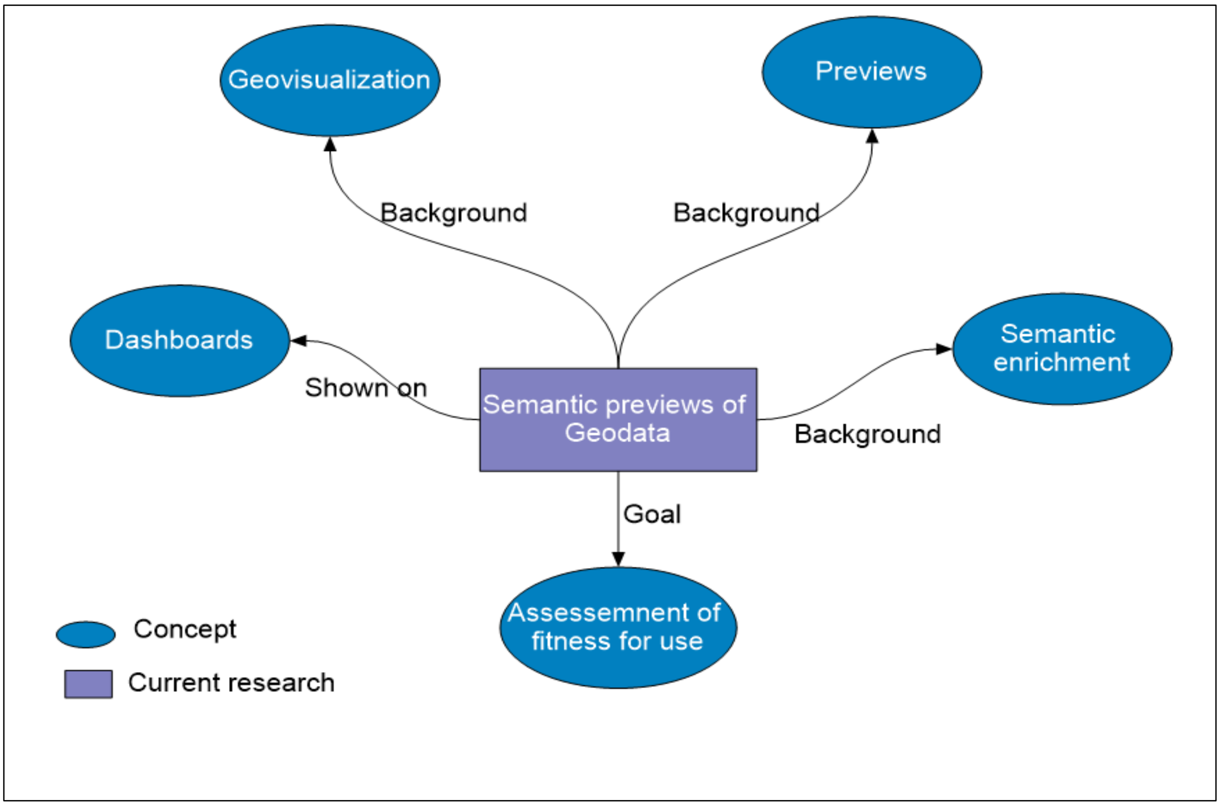


Figure 3: The relationship between different concepts are related in this research

2.1. Assessing fitness for use of a dataset

The ultimate goal of this research is to provide potential users of datasets with an effective way of assessing the fitness for use of a dataset. Fitness of use, also known as external data quality refers to the degree of overlap between the data characteristics and the needs of the users (Devillers et al., 2005).

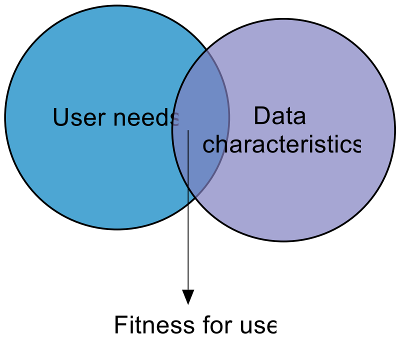


Figure 4: Graphical definition for fitness for use

Of recent, interest is growing in assessing the fitness for use of geographic datasets and the amount of research being done in this area is increasing. Some methods available are discussed below:

Bruin et al. (2001) proposed the use a decision analytical approach which compares the values two datasets. The approach proposed by the authors helps users to choose the optimum dataset from a range of alternatives for a particular use in relation to financial implication. The assessment of fitness for use is dependent on the reported error of the dataset by the producer. The uncertainty in each case is compared against a reference dataset for the intended use and financial loss resulting from the uncertainty in each dataset is computed statistically. The research targets the decision stage where the datasets that have the required attributes have already been identified and the user intends to select the best option from the available alternatives. This deviates from this research in that the latter focuses on helping the users in the quickly identifying datasets that could be useful for them depending on their intended use while the former focuses on the best available dataset from a list of identified datasets. In chronological order of workflow, the results of this research would be the alternatives available in the research proposed by Bruin et al., (2001).

Pôças et al. (2014) proposes the use metadata to analyse the similarity between the data properties and the expectation or the needs of the user. The framework proposed by this method creates a matrix based on the attributes found in the metadata and those specified by the users and presents the results so the user can decide more easily on the datasets that may be useful for their purpose(s). This method helps the user to refine the search results to those datasets that could of use to them. However, since the quality indicators are obtained from the metadata, some information relevant for assessment of fitness for use are not considered (Devillers et al., 2005), for example the attributes within the datasets are still not explicitly shown. This means that the user would still have to access the dataset to be able to obtain search information that is helpful in understanding the dataset better. Furthermore, this method does not provide previews which are the proposed means by which the potential user of a dataset assesses the fitness for use of a dataset.

Another method of assessing fitness for use of a dataset proposed by Devillers et al. (2002) is the use of Multidimensional User Manuals (MUM). In this method, the quality information of datasets is presented at different levels of abstraction to users. The quality is defined by the producers and this framework enables the users to traverse the different levels of quality presented assess the fitness for use of the datasets. Figure 5 shows an example of the levels of hierarchy. The indicators used in creating the hierarchy used for assessing fitness for use are obtained from mainly metadata components and feedback from previous users (Devillers et al., 2002) among other sources however the focus is not on the use of previews.

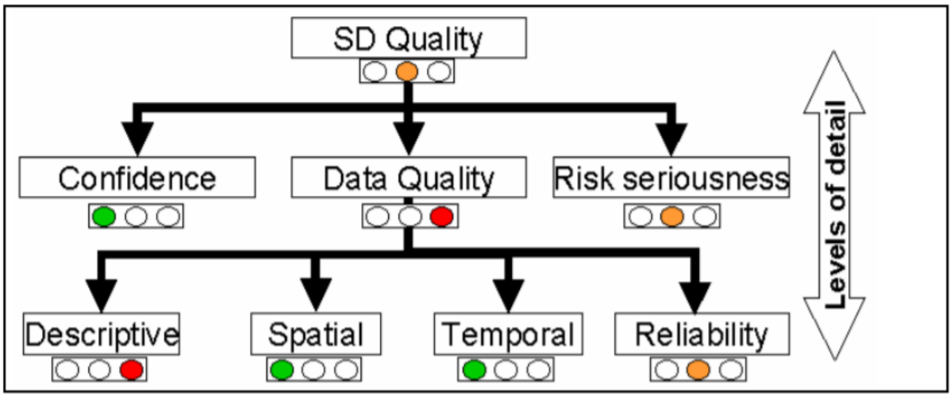


Figure 5: An example hierarchy of data quality indicators (adopted from Devillers et al., 2002)

Metadata is widely used in the assessment of fitness for use of a dataset. However, metadata do not often provide enough information relevant for the assessment of fitness for use of a dataset (Devillers et al., 2005). The focus of this research is the use of previews to present the contents of a dataset which are not made explicit in the metadata. It hopes to go a step further from the existing methods and incorporate semantic previews of data attributes to enable users better understand and assess the fitness of datasets.

2.2. Visualization

The use of previews in the geospatial field is still not common, although there has been research in the use of visualizations to make contents of spatial data easier to understand. In Mendoza et al. (2006) visualization is associated with showing data in a graphical picture or display to enable the user to visually see the graphical presentation of data. When visualization is used in geospatial data, it is known as geovisualization (Nöllenburg, 2006). Kraak (2003) refers to geovisualization as: "the use of visual geospatial displays to explore data and through that exploration to generate hypotheses, develop problem solutions and construct knowledge". The main importance of geovisualization is to stimulate visual thinking by the users (Kraak, 2003). The focus of this research was not to assist users of a dataset in exploration or provide alternative ways of displaying it, rather, to find a way of enabling potential users of dataset to understand its contents and meaning using visualizations. In addition to search techniques, proper visualizations are also key in helping potential users of a dataset during the data discovery stage (Aditya & Kraak, 2006). Data discovery stage here refers to finding a suitable dataset for a particular purpose. Aditya & Kraak (2006) continues to state that because of insufficient implementations with current systems, there are still challenges in deciding whether a dataset is suitable for a particular purpose or not. However, there are some examples of implementations where users have been provided with visualizations to make this process easier and more efficient.

Aditya & Kraak (2006) developed a method that improves the user experience by enabling them to easily understand the characteristics of a dataset. This method makes use of the different visualizations for example tables, charts, map and metadata visualizations shown in different views to enable users understand better a particular dataset and even enable comparison with other datasets to determine the most suitable for use for a particular purpose. The research focuses on the suitability of a dataset for a particular purpose for a specified purpose in relation to other datasets not on understanding the attributes within that dataset. In other words, contents within the datasets are still not explicitly defined rather different views of the dataset is given to enable the user make a decision on the fitness of the dataset in relation to other datasets. Figure 6 shows the interface of the prototype designed for this research.

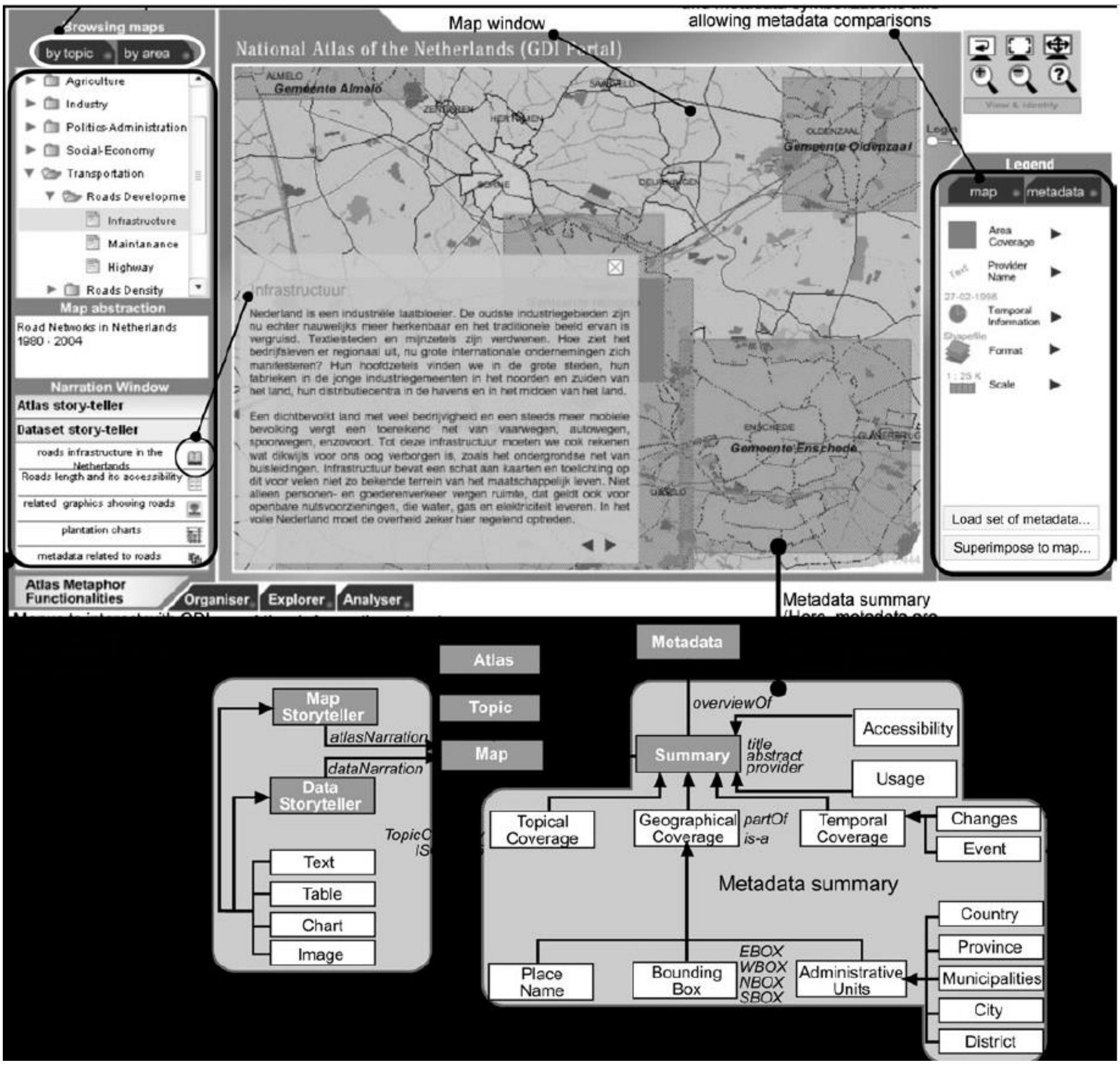
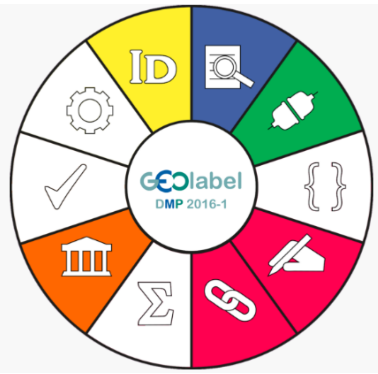


Figure 6: The interfaces and the contents in the atlas metaphor prototype are organized in accordance with the atlas information structure (adopted from Aditya & Kraak, 2006)

Another example of an application where visualization is used to enable potential users to understand the characteristics of a dataset is GeoViQua. This is a project associated with GEOSS (Global Earth Observation System of Systems) that is aimed at providing users with visualizations that give data quality indicators. The measure of data quality is obtained from metadata, data inter-comparison with other datasets, validation processes towards in-situ sensor data, provenance information and from user feedback (GeoViQua, 2014b). GeoViqua makes use of geo labels to create visualizations of quality indicators for each dataset (GeoViQua, 2014a). The geo-label consists of facets, each representing a component that can be used to give an indication of quality. They are generated depending on the attributes of the data compliant with the Data Management principles (DMP) in the project. These geo labels however do not explicitly focus on the contents of the dataset, rather on attributes like user feedback, comments from the producer and lineage of the data among others. Figure 7 shows a geolabel form GeoViqua and DMP principles.

(a) 

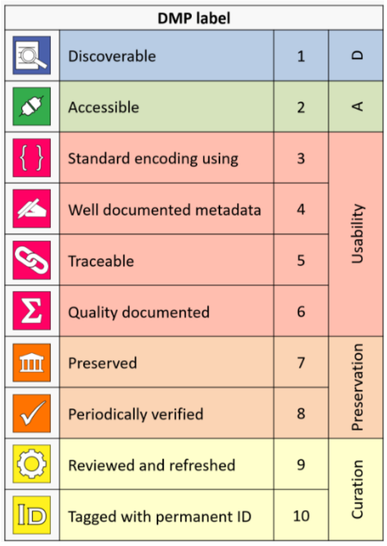
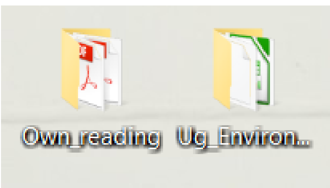
(b) 

Figure 7: Figure 7(a) shows a Geolabel for a dataset with only six elements that comply with data management principles (DMP). 7(b) shows the DMP labels.

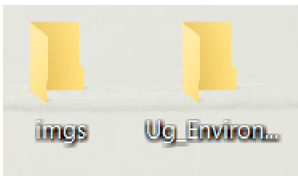
2.3. Previews

For this research, the visualizations were showed as previews. Previews are a key component of this research and Dictionary.com (2016) defines preview as: "anything that gives an advance idea or impression of something to come". They help to people to understand what something is about before they make use of or have access to the particular thing/application. Previews are graphical in nature and summarise the contents of in the particular context in which they are used. Photo apps and movie industry widely make use of previews to advertise their products to potential customers. For movies, trailers are always made which show the major actors, scenes etc. of the movie to promote it to potential audiences. The most important and interesting scenes of a movie are included in a preview to capture viewers (Rasheed & Shah, 2002). They are also helpful in the appropriate classification of movies (Rasheed & Shah, 2002).

In Windows operating systems, there is an option of showing thumbnails that when checked shows a preview of a photograph and contents of folders. This enables the user to know the main contents of a folder before opening it. Figure 8 illustrates the use of previews for folders in windows.



(a)



(b)

Figure 8: Figure 8(a) showing preview of the contents of the folders and Figure 8(b) shows the folders without preview

TechSmith (2016) gives some dos and don'ts for creating previews for applications, for example focussing on a few killer features when creating the preview. Four explanations are given for why app previews matter (TechSmith, 2016);

- Saving the customers time and money by enabling them to look into the app before downloading or purchasing it.
- Encouraging happy customers since they are aware of what the app does

- Showing the unique functionality of the app
- Encouraging the discovery of apps

As already mentioned, the use of previews in geospatial field is not very common. For spatial data, previews are usually shown as thumbnails of the dataset in the metadata and an overviews map to give the user an impression about the dataset. However, just like in the trailers of movies, previews can be used to provide more information to enable users understand the dataset better and be able to determine its suitability for use for a given purpose. The reasons for creating previews for apps listed earlier are also applicable in the field of spatial data, namely;

- Saving users of a dataset time in finding the appropriate dataset
- Enabling the customers to understand how the dataset can be used
- Encouraging the discovery of relevant datasets

For spatial data, different visualizations need to be provided at the same time to enable the potential user to get a lot of information about the dataset at once as a preview. This information can be shown on a dashboard where it can be viewed at the same time without causing "information overload" (Yigitbasioglu & Velcu, 2012).

2.4. Dashboards

A dashboard is defined as: "a visual display of the most important information needed to achieve one or more objectives: consolidated arranged in a single screen so that information is monitored at a glance" (Few, 2006). A dashboard is therefore expected to summarise and present the most important information to the user on a single screen (Yigitbasioglu & Velcu, 2012) similar to a car dashboard which provides the most important information for operating a car at a glance (Rivard & Cogswell, 2004).

Dashboards have been used in various fields. In Pauwels et al. (2009), dashboards are used in the business management field. Many examples of firms that use dashboards for increasing effectiveness in making and decision making are provided in Pauwels et al. (2009). Four factors driving the need for dashboards by managers are discussed(LaPointe, 2005 as cited by Pauwels et al., 2009), namely;

1. Poorly organised data for decision making
2. Biases in information processing and decision making by management
3. Increase in demand for market accountability
4. Need for inter- departmental integration

Pauwels et al. (2009) further explains four uses and purposes of dashboards; enforcing consistency in measures and measurement procedures, monitoring performance, planning for the future and communication to important stakeholders. These uses are not only limited to the field of business management but in the other areas like in city planning too where dashboards have been used. The term 'city dashboards' has been used to refer to online interactive data visualizations platforms that provide citizens with information about a city (Kitchin, Lauriault, & McArdle, 2015). In McArdle & Kitchin (2016), a dashboard was designed for a city to help its citizens to know and understand the city and various projects being undertaken. Figure 9 shows the interface of the Dublin dashboard. It provides an overview of the key knowledge materialising from urban data with the ability to explore the data further and to identify relationships between data (Rivard & Cogswell, 2004).

According to Kitchin et al. (2015), dashboards can be analytical, showing the current system, performance driven or used for comparing services or systems. Analytical dashboards do more than provide a summary but enables it users to drill down and explore the data in a single interface to identify relationships between them if they so desire. In the recent years, dashboards have become more interactive and enables users to interact with the data and query it compared to earlier dashboards that were rigid in their display (Kitchin et al., 2015). With a dashboard, different visualizations of the same data can be displayed simultaneously like in a table, a map and graph(Kitchin et al., 2015) in fact now it is common for dashboards to have an interactive and inter linked maps, graphs, gauges and indicators (Keim et al., 2010).

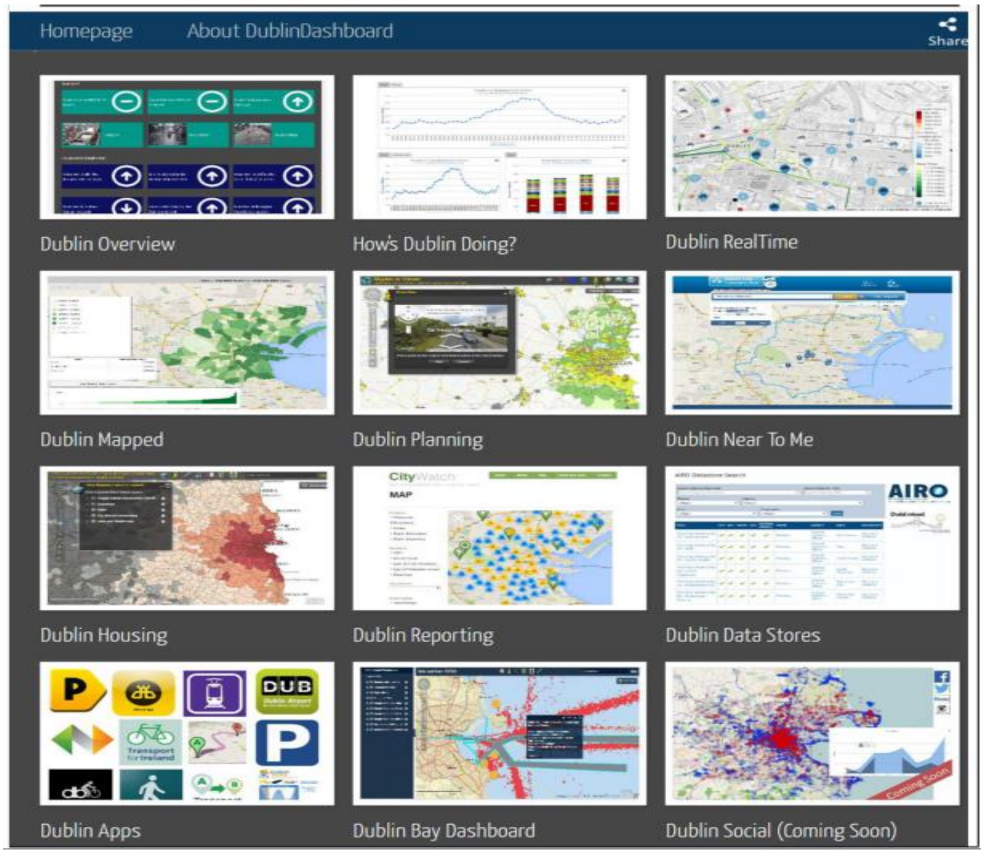


Figure 9: The Dublin dashboard (adopted from McArdle & Kitchin, 2016)

For this research, a dashboard was designed for geospatial data to show the contents of the data in different visualizations to facilitate its users to assessing fitness for use of the datasets. Just like in McArdle & Kitchin (2016), it was used by the citizens and personnel to understand the data about the city, in this research a dashboard is used with the hope that it will show its users, the contents of datasets and enable them to understand them.

2.5. Semantic Enrichment

Initially, geospatial data was produced and used within the same organisation and therefore, characteristics and production processes about the data was fully known by the users (Devillers et al., 2005). This means that the users fully understood the semantics of the datasets. The terms "Semantics" refers to "expressions in a language" according to Kuhn (2005) and Merriam-Webster (2016) dictionary defines it as: "the meanings of words and phrases in a particular context". For this research we shall consider the definition given by Merriam-Webster dictionary.

With the need to share data from heterogeneous sources, comes the problem of semantic interoperability. This is because attributes and data types understood by a user may vary from those implied by the initial producer (Kuhn, 2005). Two datasets about the same theme like land use may have attribute commercial building but when the description of commercial building varies between the two datasets.

Research has been done in the field of enrichment of data to address the issue of semantics. Semantic enrichment involves adding metadata to contents so that machines can make sense of it and use in appropriately in a given context (Clarke & Harley, 2014). Angeletou (2008) focussed on how to automatically enrich folksonomy tagsets with ontologies. Vockner & Mittlböck (2014) provide a method of enriching metadata with synonyms, toponyms and user defined keywords with the help of multilingual thesauri and ontologies. This facilitates improvement of the user experience while searching for information. Semantic enrichment has been mainly done for fields with textual content like search engines (google is an example where semantic enrichment is used) and it has also been used been used in the field of spatial data particularly with geoportals to facilitate search for geospatial data using keywords. Athanasis et al. (2009) proposed the use of ontology-based methods to organise metadata so as to allow for navigation and exploitation of data semantics to make information discovery more efficient. Ontology is defined in Gruber, (1993) as "an explicit specification of a conceptualization". The ontology- based method makes use of the Resource Description Framework (RDF) which is a semantic query language to enable users to discover data close to what they are looking for. Bogdanović et al. (2015) on the other hand proposed a method that enables users to discover data sources by simply providing in natural language the geo-information they are interested in. To facilitate this, the system was semantically enriched with different languages which enable connections to the datasets regardless of the language the user provides the search terms in. Ontologies provide a means for users of geospatial data from different fields to access and share information (F. Fonseca, 2013). Ontologies describe concepts that are related to actual positions in the real-world are known as geo-ontologies. These concepts have space as one of their characteristics (F. Fonseca, 2013). The author continues to emphasize that geo-ontologies describe both spatial and semantic relations between concepts. Therefore, geo-ontology enables spatial data enrichment with semantic information; this means that information about the meaning and context of the dataset is added to location based data. The two advantages of using geo-ontologies; it is supported by the current system of semantic web and the user of the data is provided with all the information necessary to complete a particular task, further (Mobasheri et al., 2013). One of the research areas in this field was by F. T. Fonseca & Egenhofer (1999) who proposed a framework that makes use of ontologies linked to GIS data sources to enable integration of data depending the meaning. There are other examples of research done in relation to this field; however, there is still limited application of semantic enrichment in the field of visualization particularly of previews.

The concept of semantic previews is introduced in this research. Semantic preview is defined here as: "a graphical representation of contents about or within a dataset presented as previews that gives the potential user of the dataset useful information (in relation to their perception and experience) about the dataset like the contents, potential use or lineage of the dataset". Instead of using text and dictionaries to enrich the meanings of terms to enable people of different backgrounds to understand the context in which it is used, this research proposes the use of previews.

Ristoski & Paulheim (2016) conclude that the full potential of Semantic Web technologies and linked data is yet to be fully utilized. This shows that there potential for using semantic web in metadata enrichment however it has not yet be used in creating previews for geospatial data. In this research, the attempt is to semantically enrich datasets with previews that provide information about the dataset so that the user can understand the context in which the data is being referred and make proper use of it. The idea of a dashboard was implemented so that the user can get as much information as possible as an overview.

2.6. Geospatial datasets

Geospatial data is defined as data relating to a position on the earth's surface. For many years, the production of geospatial data was left to skilled people like surveyors and therefore one needed to be skilled in order to participate in the production of geospatial data. However, with the removal of selective availability of GPS signals, position accuracy was improved for low-cost GPS receivers and therefore more people were able to participate in production of geospatial data (Haklay & Weber, 2008). With more unskilled people being able to create geospatial data, the concept of VGI(Volunteered Geographic Information) was developed. Geospatial data therefore are now being produced by both skilled people and unskilled or those with limited skills in this field. Goodchild (2007) describes VGI as a special form of User-generated content where private citizens often with limited formal qualifications providing geographic information. For this research we shall adopt the definition of VGI provided by Haworth (2016): "the widespread creation and sharing of geographic information by private citizens, often through platforms such as online mapping tools, social media, and smartphone applications". OpenStreetMap (OSM) is one of the most widely used VGI source. OSM is "a knowledge collective that provides user-generated street maps" (Haklay & Weber, 2008). This research makes use of OSM as a data source to create previews that were used in developing the prototype. The motivation for using OpenStreetMap data was the free access to current data, the wide and detailed coverage for many parts of the world. In addition, OSM data has be structure to created linked geodata. www.wiki.openstreetmap.org describes linked geodata as "a large spatial knowledge base which has been derived from OpenStreetMap for the Semantic Web". Semantic web is an extension of the World Wide Web and also sometimes referred to as a web of data. OWL (W3C Web Ontology Language) is one of the vocabularies used to enrich data for semantic web. One of the ontologies created for OSM data is OSMonto (Codescu et al., 2011). This ontology can be edited and queried with an ontology editor like protégé. Figure 10 shows a visualization of the OSMOnto with WebVowl, a visualization tool for ontologies.

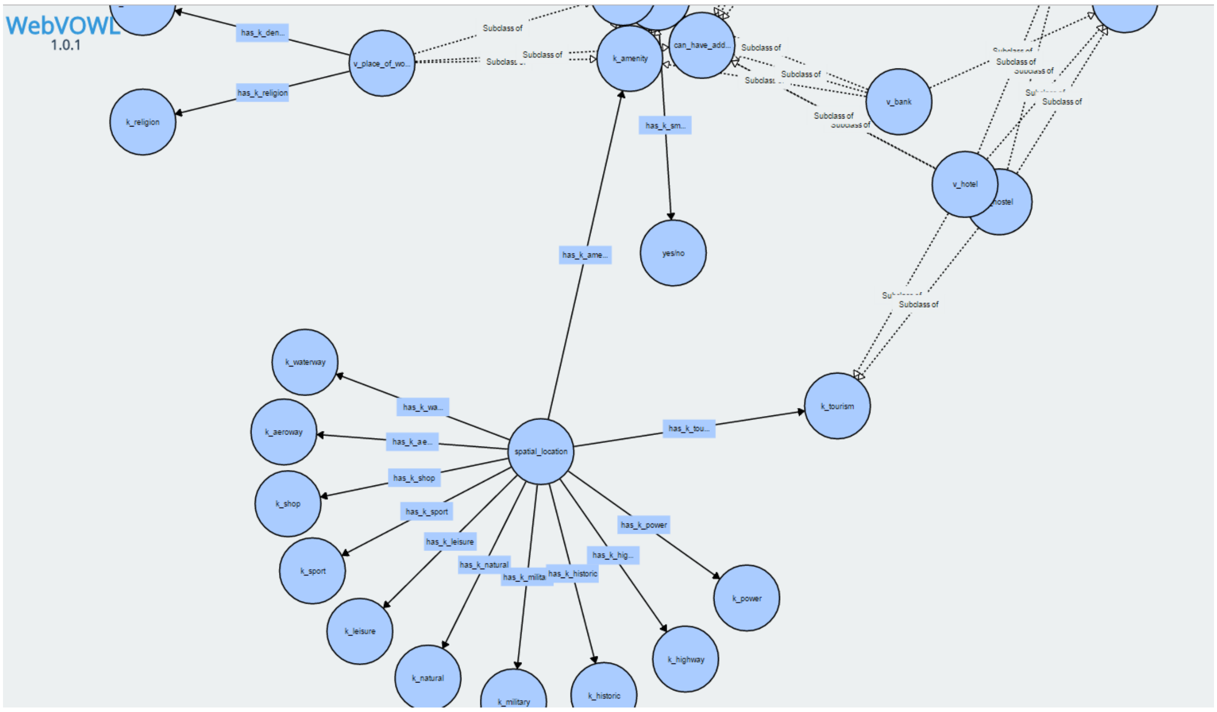


Figure 10: Figure showing part of OSMonto visualized with WebVowl

2.7. Users of geospatial data

Similar to the producers of Geospatial data, the users of geospatial data and applications were also mainly experts in this domain. However, with the increased access through the internet, there is an increasing number of people with limited expertise that are making use of geospatial data and applications (Devillers et al., 2005). This shows that there are two major classifications of geospatial data users: experts and non-experts. According to Kuhn (2005), geospatial semantics is about “understanding GIS contents and capturing it in formal theories”. The author continues to state that; “geospatial information is often based on human perception and social agreements”. Based on this, people within the same social circles or community will have a similar perception on something.

2.8. Conclusion

This Chapter provided a background for this research. Different concepts and work related to this research were covered. The chapter attempts to show how the different concepts are related to each other and their connection to this research. Chapter 3 covers the evaluation of geoportals to determine how the current systems are being used to assess the fitness for use of datasets.

3. EVALUATION OF GEOPORTALS

This Chapter covers the current state of affairs concerning geospatial data. Geoportals are widely used as a storage and access point for geospatial datasets. In this Chapter, we do an evaluation of the geoportals to determine how they make use of previews. This was done to achieve the first objective of the research which was the evaluation of geoportals. Some background information about evaluation of geoportals and the procedure that was followed to do the evaluation is described. The results of the evaluation and the summary of the results are shown in Table 1. The results of this evaluation along with the results from the requirement analysis (which is discussed in Chapter 4) are used for designing prototype.

3.1. Framework for evaluation of geoportals

There is no standard framework for evaluating geoportals; however, different authors have come up with different frameworks for the evaluation of geoportals for depending on the purpose of the evaluation. Braggaar (2016) developed an assessment framework (drafted form literature) to assess the current state of Dutch geoportals. The framework consisted of ideas drafted from theoretical perspectives and discussions from a team Geomatics experts. The evaluation was based on a desktop research. This enabled the identification of strengths and weaknesses and facilitated the suggestion areas of improvement in the Dutch portals to an ideal portal. In Aditya & Kraak (2006), geoportals were evaluated to determine whether the functionalities that are provided to users for defining queries and assessing the search results are sufficient. The evaluation was based on the 'four phase' framework proposed by Shneiderman, Byrd, & Croft (1997). The evaluation was based on tasks that were given to the users to perform and questionnaires were given to them to evaluate the geoportals based on he defined tasks. From the evaluation, it was discovered that the geoportals lacked appropriate navigation tools and facilitation to support user's understanding. For this research, an evaluation framework was constructed based on aspects relevant for the research.

3.2. Procedure for the evaluation of existing geoportals

For this research, the evaluation framework for existing geoportals was constructed based on the aspects that are important in this research, i.e. assessment of fitness for use of datasets and use of previews. Five geoportals were selected with different coverage to get a general picture of geoportals, these included; Inspire and GEOSS (Global Earth Observation System of Systems) which are regional geoportals, Geoplatform and Geoportal.de which are National Geoportals and IGAD-UNOSAT which is a project geoportal. The evaluation criteria consisted of availability of previews which is a key component of this research as already mentioned and different components of metadata since users mainly make use of metadata to assess fitness for use of datasets. The evaluation was a desktop evaluation because it gives more time to understand each geoportal compared to a static survey (Vandenbroucke et al., 2008).

3.3. Results of evaluation

Table 1: Table showing the results of evaluation of geoportals

		Geoportals				
No.	Criteria	Inspire geoportal	GEOSS portal	Geoplatform	Geoportal.de	IGAD-UNOSAT geoportal
1	Data coverage	Europe	World	US	Germany	Eastern Africa
2	Targeted user	Public sector(experts and non-experts in geospatial field)	Various users of spatial data especially for decision making	Expert and non-expert users of spatial data	Both expert and non-expert users	IGAD staff members (People with some GIS expertise)
3	**Keywords:**					
	Predefined	Yes	Yes	Yes	Yes	Yes
	User defined	No	Yes	Yes	Yes but predefined gives better results	No
4	**Overview of search results**	Yes	Yes	Yes	Yes	Yes
5	**Ranking of search results**	Yes	No	Yes	Yes	No
6	**Access to full metadata:**					
	One format	Yes, normal language	No	No	No	No
	Various formats	No	Yes, Xml and normal language	Yes; xml and normal language	Yes; Xml, normal language	Yes; xml and normal language
7	**Availability of Preview:**					
	Overview map	Yes	Yes	Yes	Yes	Yes
	Zoomed-in map to show details	No	No	No	No	No
	Static	No	Yes	Yes	Yes	No
	Dynamic	Yes, allows zooming in and out	No	No	No	Yes, allows for zooming in and out
8	Availability of dataset description	Yes as abstract	Yes	Yes	Yes	Yes
9	**Availability of lineage:**					
	Text	Yes	Included in the description	No	No	No

	Workflow diagram	No	No	No	No	No
10	**Explicit definition of the data attributes**	No	No	No	No	No
11	**Access to the actual dataset**	Yes; Download, WMS, WFS, API	Yes; Download	Yes; Download	Yes, download	Yes; Download
12	**Information on intended use of the datasets**	No	No	Yes	No	No

3.4. Summary of results

From the evaluation, it was determined that the geoportals only show an overview map of the dataset which are static in most geoportals however others allow for zooming in and out. The previews do not give a representation of the contents within the dataset. To get an idea of what is within the datasets, the users have to read the metadata which also do not provide enough information about the contents that within the dataset that the user can expect. Some of the geoportals like Inspire geoportal have keywords which could also be some of the attributes within the dataset but the users can only assume unless they have access to the dataset.

3.5. Conclusion

In this Chapter, the evaluation of geoportals was covered. It started with an introduction, available framework for the evaluation of geoportals, the procedure and finally the results of the evaluation. The results from this evaluation together with the requirement analysis were used in developing the design of the prototype which is described in Chapter 4.

4. BUILDING SEMANTIC PREVIEWS

This Chapter covers the steps followed to achieve the second and the first part of the third objective. It covers the requirement analysis and the development of the prototype. The approach used to design the prototype was the User Centred Design approach which is described in the next section.

4.1. User Centred Design

Delikostidis (2011) provides a review of different research methods and techniques used in UCD. They were categorized into three and are shown in Table 2. UCD involves three main processes; requirement analysis, prototyping and usability testing as illustrated in Figure 11. Requirement analysis of a design involves the identification of the users, the use context and the organisations (Delikostidis, 2011). The results of the requirement analysis are used to create a design solution (prototyping) which is then evaluated to determine whether it fulfils the purpose for which it was created. Evaluation and design solution is done iteratively until a satisfactory result is reached (Elzakker & Wealands, 2007).

Table 2: Table showing methods and techniques used in UCD (adopted from Delikostidis, 2011)

Analyse Requirements	Produce Design Solutions	Evaluate Designs
Survey / interview of existing users	Usability goal setting	Usability Inspections
User requirements interviews	Design guidelines and standards	Usability testing in the lab
User profiling		
Contextual Observations / interviews	Scenario-based design	Usability testing in the field
Diary keeping	Parallel design	Post-experience interviews
Task analysis	(paper- or working interface-) Prototyping	Heuristic evaluation
Competitive analysis	Card sorting	Focus groups
Card sorting	Focus groups	Satisfaction questionnaires
Personas	Individual interviews	Expert reviews
Scenarios of use	Surveys (online)	Surveys (online)
User / task models	Usability testing	Diagnostic evaluation
Interaction modelling		
Heuristic evaluation	Use cases	Performance testing
Usability testing	Style guide	Critical incidence technique
Evaluating existing system(s)	Wizard of Oz	Remote evaluation
Brainstorming	Heuristic evaluation	Logging
Affinity diagramming	Interface design patterns	
Requirements meeting	Rapid prototyping	

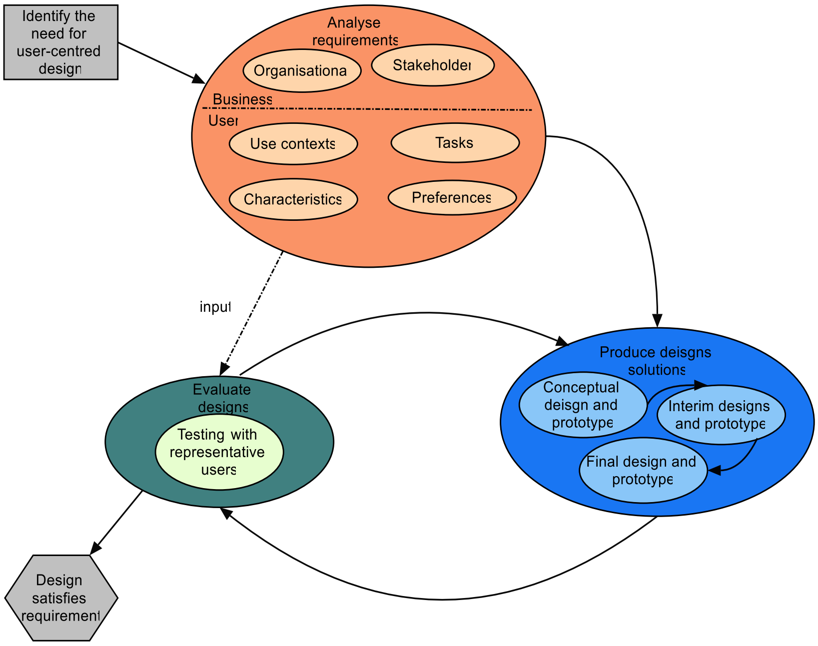


Figure 11: Figure showing the processes involved in UCD (adopted from Elzakker & Wealands, 2007)

4.2. Requirement analysis

In this section, a detailed description of the target user group and use environment is given. Requirement analysis was the second objective of this research. The requirements of the users are determined by studying the kind of information the user would be interested in while trying to determine the suitability of a dataset for a particular use. For this research, the methods selected for requirement analysis included; user profiling, defining a use case scenario use and task analysis based on Table 2. Figure 12 shows the steps taken for the requirement analysis. The datasets of interest to the users are also described in this chapter. The questions that the users may have while trying to find relevant datasets for a particular purpose were used in determining the tasks of the user. The conceptual design, which is described in Section 4.3, was developed based on these solutions for the user.

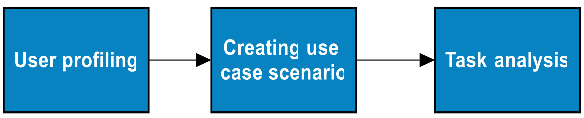


Figure 12: Steps followed for requirement analysis

4.2.1. User profile

The user profile describes the characteristics of the target users that were considered for this research. The age group, years of experience with geospatial data among others are described below. Table 3 gives an overview of these characteristics.

Table 3: Table showing the characteristics of the user group

Age	20-65
Nationality	Any
Field of study	Road safety
Years of experience in the field	More than three years
Other experiences	Has experience using geoportals and web applications for finding relevant datasets

The target users of this research are researchers in the field of road safety. They are people who have some years of experience working with geospatial data in this field. Their main interest is finding ways to reduce the number of road accidents in different parts of the world. The current focus is on third world countries so that they could help their governments cut the annual spending on road accidents. To do this, they always need to find geospatial data for road accidents and other road related datasets for example, road types, length, curvature and surface type. These researchers would like to spend as little time as possible in trying to find relevant datasets. Therefore a system that provides as much information as possible about a dataset at a glance is required to enable them to quickly understand a dataset. The system should be intuitive and easy to use so that the users do not have to spend a lot of time trying to understand what a dataset is about.

Since the researchers usually use large screens, the system should be usable on a desktop but could also be made accessible with mobile devices since they have gained popularity in recent time.

The information about a dataset provided in the overview should be the most relevant for someone assessing a dataset for fitness for use. The information seeking mantra of overview first and then details on demand as suggested by (Shneiderman, 1996) would also be helpful to the researcher because datasets not useful for a particular purpose would be eliminated quickly and more information could be obtained for datasets that might be useful.

The next section explains a use-case scenario and a user profile.

4.2.2. Use case scenario

This sets the use context and scenario in which it could be applied. A fictitious user that fits the description of the target user group and a use case scenario is described below.

Background

James is a foreign researcher in the field of traffic accidents and has about eight years of experience working in this field. He is familiar with geoportals and has used some of them in the past to find datasets for his research. He has used a few national geoportals, project specific geoportals and Inspire geoportal and therefore is familiar with how geoportals work. His current research project is to compare traffic accident Situations in Africa (Uganda) and in Europe (Netherlands). For Africa, he has selected Kampala Uganda as his case study. Figure 13 shows a graphical representation of the user profile

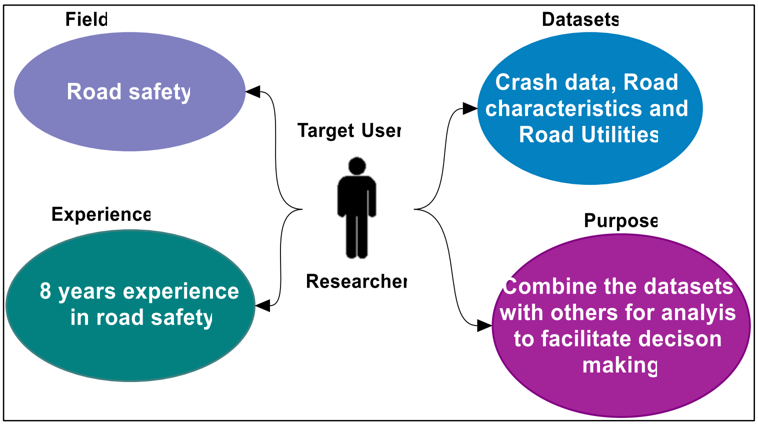


Figure 13: Graphical representation of the user profile

Datasets

The datasets that are of interest to him are;

- Crash data which is contains the number of the accidents along different roads. It is assumed that he already has this data and is interested in finding the rest.
- Road characteristics which include road type, length, curvature, maximum speed limit, and road surface.
- Road utilities like lighting on the roads, traffic signs and signals

Purpose

James wants to use this data to combine it with other data like crash data to carry out analysis on this data. Using the analysis, he wants to determine the roads with the highest accident rates, possible reasons for these accidents and also determine countermeasures to reduce them.

How to interact with the application

As he searches for the datasets to use, James wants to quickly understand the contents within the dataset to determine whether or not that particular dataset is useful for his research. When James opens an application, he types in a keyword that is relevant for the dataset he is interested in. when he searches for matches, he is presented with search results. On selecting a particular result, he is presented not only with the names of the datasets that match his query but with the contents within that dataset and previews to give more information about the attributes within the dataset.

4.2.3. Questions and tasks of the user

There are many questions that James would have while trying a find a suitable dataset for a given purpose; these questions are the ones that will be addressed in this research.

1. What is the dataset about?
2. For which area is the dataset?
3. What was the intended purpose of the dataset?
4. What are the attributes within the dataset?
5. In which context are they used in this dataset?
6. Are the attributes in the format I am interested in using?
7. How do the features appear on a map?
8. Are the attributes complete enough for the intended purpose I have?

For the user to answer the questions above, he/she should be able to perform certain tasks with the application. These include;

- Use keywords to search for dataset of interest which eliminates datasets that are not useful for a particular purpose
- Visualize the features of a dataset
- Answer the question about the completeness of the attributes within the datasets

The prototype (described in the Chapter 6) was designed to enable the user carry out these tasks.

4.2.4. Summary of user requirements

In conclusion, the users are interested in a system that gives as much information about a dataset as possible using previews because texts take longer to read. These previews should enable the researchers to decide whether or not a dataset is useful for their specific purpose. To be able to determine the suitability of a dataset for a particular use, the user should be provided with information about the attributes within the dataset, the completeness of these attributes, the context of the attributes and possible use of the datasets. The user also should be provided with the map rendering of the features and the representation of the location of the features. The conceptual design and the implementation of a prototype which shows the semantic previews are discussed in the next Chapter.

4.3. Design

This section makes use of results from the requirement analysis and evaluation of geoportal to answers some of the questions for the third objective; which was development of a prototype. It describes and justifies the design choices made for the prototype that was developed in this research. It covers the solutions that were implemented to help the user answer the questions that he/she is interested in while looking trying to understand the contents of a dataset. Different elements and how they interact with each other to help the user in understanding datasets are described here. The idea of a dashboard was used to show different visualizations of the data to help the users assess the fitness for use of a dataset. There are five dashboard development stages in Pauwels et al. (2009) which were obtained from Gulati & Oldroyd (2005), Reibstein et al. (2005) and Wyner (2008);

1. Selecting key metrics
2. Populating the dashboard with data
3. Establishing relationship between dashboard items
4. Forecasting and scenarios
5. Connecting to financial consequences

These stages were used in creating a dashboard for business field however the stages 1-4 were adopted and used for creating the dashboard in this research. These steps were modified for this research, the steps followed included;

1. Creating use scenarios
2. Selection of key elements to be shown on the dashboard.
3. Determining the data sources for the elements
4. Establishing interaction between the different elements
5. Finalising on the design choices and creating a conceptual design
6. Implementation of the design

4.3.1. Task description and system design requirements

The use case scenario was described in Section 4.2. The questions and tasks developed from the use case scenario were used in selecting the elements to include in the prototype. Figure 14 shows the relationship between the areas of interest, the questions and the solution that was provided by the system. The areas of interest indicate the kind of information the users of geospatial are interested in. The questions listed in the figure are the questions they ask and would like answers to as they search for relevant datasets. Solutions in the prototype show the capabilities provided by the prototype to enable users answer the questions they have about a dataset.

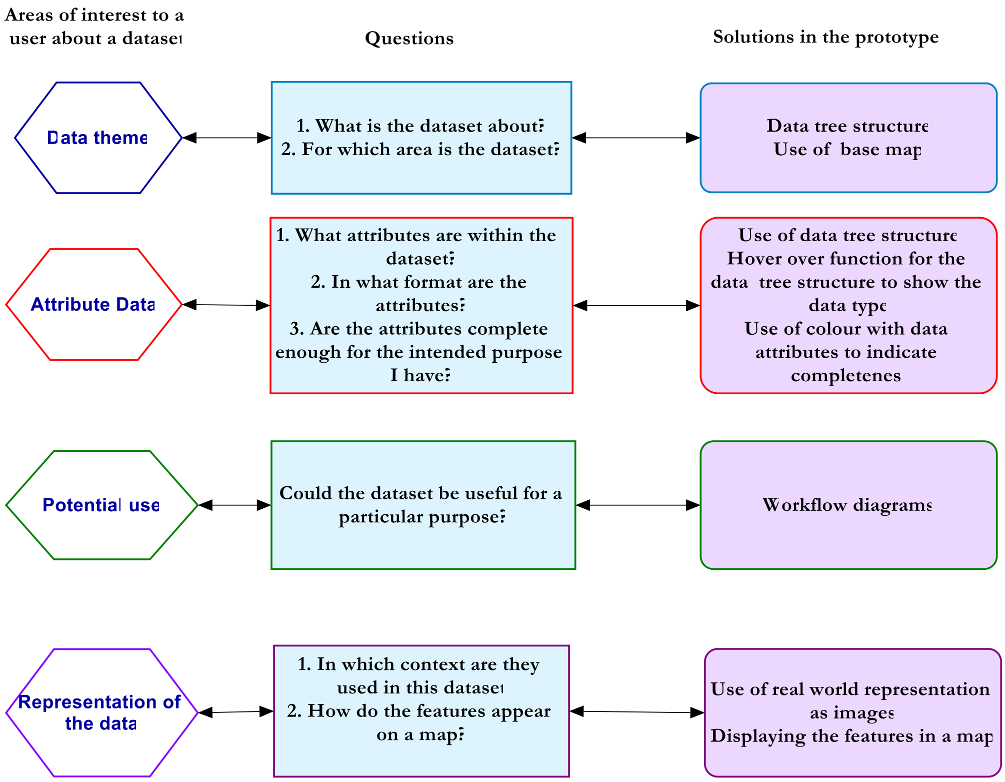


Figure 14: The relationship between the areas of interest, the question and the design solution in the prototype

4.3.2. Selection of dashboard elements

This section gives a description of the components that were selected or considered for the implementation of the prototype. The components that were considered were based on the potential to portray information at a glance. The options that were selected for implementation were chosen on the basis of feasibility of the implementation, usefulness in relation to the number of datasets selected for the prototype and relation to research in the field of semantic web. One of the options that were considered to help users in the search of datasets was the use of tag clouds; however it was not implemented because there were very few datasets used in the application which made it impractical. There are scenarios, however where the tag clouds would be useful. The use of tag clouds is discussed below;

4.3.2.1. Tag clouds

Tag clouds are defined as "visual representation of text, mostly to put metadata tags and keywords on websites in order to perceive freely without main text" (Mishra & Mishra, 2012). They are commonly used in websites to give the user information that is available on the website. It is used with text, which are clustered together into one word. In Mishra & Mishra (2012), the font size is commonly used to display the relevance or frequency of a word in the website, sometimes colour or weight is used for the same purpose. Most important or frequent tags are given a larger font, brighter colour or heavier weight. Tag clouds are often linked to resources which are used to find information (Sinclair & Cardew-Hall, 2007). In the field of geodata, an example of where a tag clouds is used is in the Inspire geoportal. Figure 15 shows the tag cloud from the Inspire geoportal. While tag clouds are useful in helping people find information when associated with links, they should not be the only navigation tool provided (Sinclair & Cardew-Hall, 2007). The authors continue to argue that users prefer tag clouds only for general search while for more specific search, they prefer typing in the search terms. Paulolyslager.com (2017) provides some information on when and how to use tag clouds. Since tags are often provided for data by different people, different tags may be assigned to the same or similar things (Begelman, Keller, & Smadja, 2006). The same author, proposed the use of clustering to make the use of tag clouds more relevant and helpful in finding data.

In summary, based on the literature, tag clouds are be used in websites that have datasets to show the datasets that are available. The font size, colour and weight can are useful in showing relevance and importance of tags. Colour needs to be used with caution otherwise only the tags that stand out will be seen by the user. The tags that are used also need to have a good balance between generalization and specification of tags. The use of semantically related tags would make data discovery more efficient since tags are usually provided by different people. It also helps in the generalization problem. In this research, tag cloud was not used because of the limited number of datasets that were used and the implementation would have required more time than was available for designing the prototype.

Figure 15: Tag cloud from Inspire geoportal (http://inspire-geoportal.ec.europa.eu/discovery/)

The parts that were considered for the design of the prototype and how each of the elements was related are shown in Figure 16. These elements included; the data tree structure, the map rendering, the data attributes, the real-world representation and a button that links to the workflows. Each of the components were selected to enable the users perform certain tasks to help them answer the questions about the suitability of a dataset for a particular purpose. The justification for each selection is discussed below.

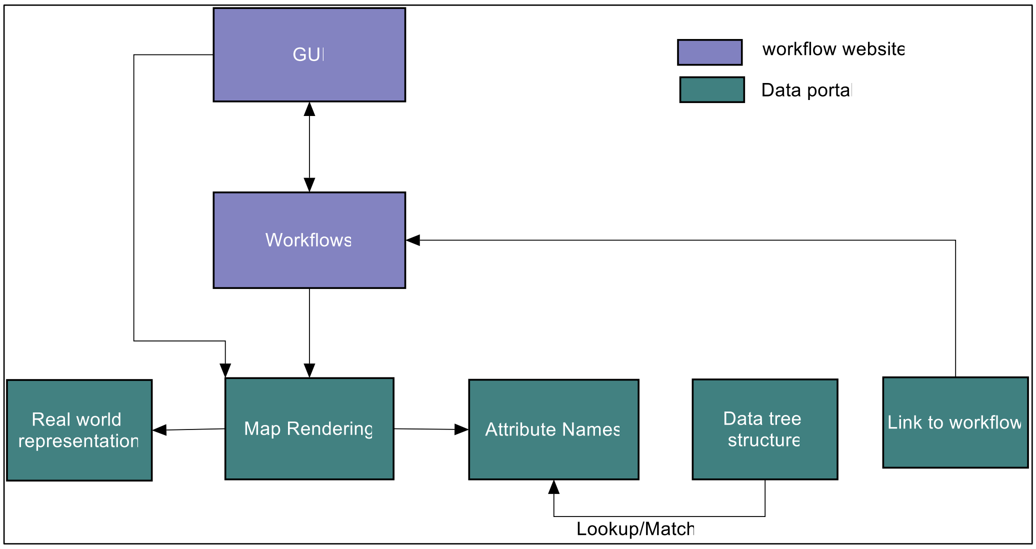


Figure 16: Links within the different elements of the prototype

4.3.2.2. Workflows

A workflow was defined by www.techtarget.com (2017) as: "the series of activities that are necessary to complete a task". The website continues to state that flowcharts are useful in enabling people to envisage the procedures followed in a workflow. According to Carline et al. (2016), the use of workflows can be used with geospatial data to visualize lineage. Lineage was defined as: "information describing source materials and transformations used to derive final digital cartographic data files" (Lanter, 1990). Lineage therefore describes the processes a dataset has undergone to arrive at the current state it is in. This information gives the potential user of a dataset an idea of how the dataset could be used. One of the user requirements identified was the need to understand how a dataset can be made use of. For this research, the use of workflows diagrams was selected. The use of workflows was thought to be easier to read and understand compared to the text provided in the lineage which is used in current metadata. In the implementation, the workflows used showed how a dataset could be used and not the processes the data had undergone. This was chosen to illustrate how potential use of a dataset could be implemented as a workflow. The use of workflows was chosen to enable the user determine the usefulness of a dataset for a particular purpose.

4.3.2.3. Map rendering

Elzakker (2004) defines a map as: "a simplification of reality". The author continues to state that maps are a useful and effective means of communication and analysis among other functions. Map rendering of features was therefore selected to show how the features of a dataset are display on a map. In addition, maps are the most common way of displaying geospatial data. To enable the potential user of a dataset to better understand it and therefore make an informed decision about the dataset, an interactive map was chosen. It enables the user to visualize the features as a map and enable them to interact with the datasets to be able to better understand them and therefore make a more informed decision as to whether to use the dataset or not. This helps the user to answer questions about spatial extent and the representation of the features.

4.3.2.4. Real-world representation

This was selected to give the potential user the real-world representations of the features in the dataset to enable him/ her understand the content in which it was being used. Two datasets with the same attribute may not necessarily represent the same features in the real-world. So to get an understanding of what features are being referred to, the presentation of real-world images was thought to be relevant. The principle was taken from OpenStreetMap in the features page where the same tag is used to show features that appear different in the real-world. For example, a highway in Europe may not look like a highway in Africa and therefore the user needs to be given a clear picture of what they mean by highway in the particular instance in which it is used. Figure 17 gives an illustration of such a scenario. The features page (www.wiki.openstreetmap.org, 2017) shows images that are different from those shown on the Africa classification of the same tag, Highways.

Figure 17: Figure showing a secondary road in Germany (left) and in Uganda (right) (www.wiki.openstreetmap.org, 2017)

This enables people who do not know much about the dataset to understand the context in which the attributes are referred in that particular case. This is also helpful in understanding the dataset better and therefore facilitates a more informed decision as whether to use a particular dataset or not.

4.3.2.5. Data tree

This component shows the hierarchical representation of the datasets as a tree structure to enable the user to know the contents within the dataset. The data tree enlightens the user on the attributes within the dataset and enables them to answer questions about the data attributes like for example what data types to expect. The data tree was inspired by ontology structure that exposes links within a dataset to enable users understand related concepts. Because the implementation on an ontology structure that was legible and could be queried was difficult, a tree structure was opted for instead. The purpose was to enable the users answer questions about the contents and data types of the attributes.

4.3.2.6. Completeness of data attributes

The objective of the component was to show the completeness of the data attributes within the datasets. For users, it is not enough to know that a datasets has particular attributes, the completeness of the data is also very important. Data with incomplete attributes may not be useful for a particular purpose and the user would also want to know this information as soon as possible so that they can look elsewhere for more appropriate datasets. Completeness of data is one of the attributes of data quality users of geospatial data consider when deciding on the suitability of a dataset for a particular purpose.

4.4. Prototyping

This section describes how the implementation of the prototype was done. The first section gives a description of the data used is the application. The tools and data sources used in the implementation are then explained. The evaluation of the prototype is presented in the next Chapter.

4.4.1. Datasets

The selected case study was in the field of traffic, particularly traffic accidents. Datasets of interest are therefore traffic related particularly road datasets. The datasets used to create the prototype in this research were obtained from OSM which is a widely used VGI source. The datasets were downloaded as shapefiles. This is because OSM has a lot of information about roads for the selected for the study area. In addition, OSM is very widely used and therefore many people are familiar with it and also they have ontologies that have been created for the dataset that can be re-used to provide more information about the attributes within the dataset and their semantic relationships for the users to quickly understand what the dataset is about. For the prototype, the data from OSM was modified to reduce the number of attribute values for simplicity. The attributes were modified as described below;

- Primary link was changed to Primary
- Secondary link was changed to secondary
- Tertiary_link was changed to tertiary
- Living_street was changed to service road

4.4.2. Implementation

The implementation of the design requirements was done in two parts. First a website for workflows was created and then the data page where the all data in the application could be accessed from was developed. The prototype was designed in that users accessed the data from the workflows but users could also go directly to the data page without going to the workflows first. First, a description of the data page is given after which the creation of data page (where all the components described above) is also described. Figure 16 shows how the two pages were linked.

4.4.2.1. Workflows website

The prototype was designed in such a way that the main entry point to the data portal is through the workflows website. The workflows enable users to understand how the dataset within the portal could be used. The implementation of the workflow website was created using HTML, CSS and Javascript platforms and libraries. Figure 18 shows the interface of the website landing page. The website can be accessed via this link: http://win371.ad.utwente.nl/student/s6025013/openlayers/simple-website/

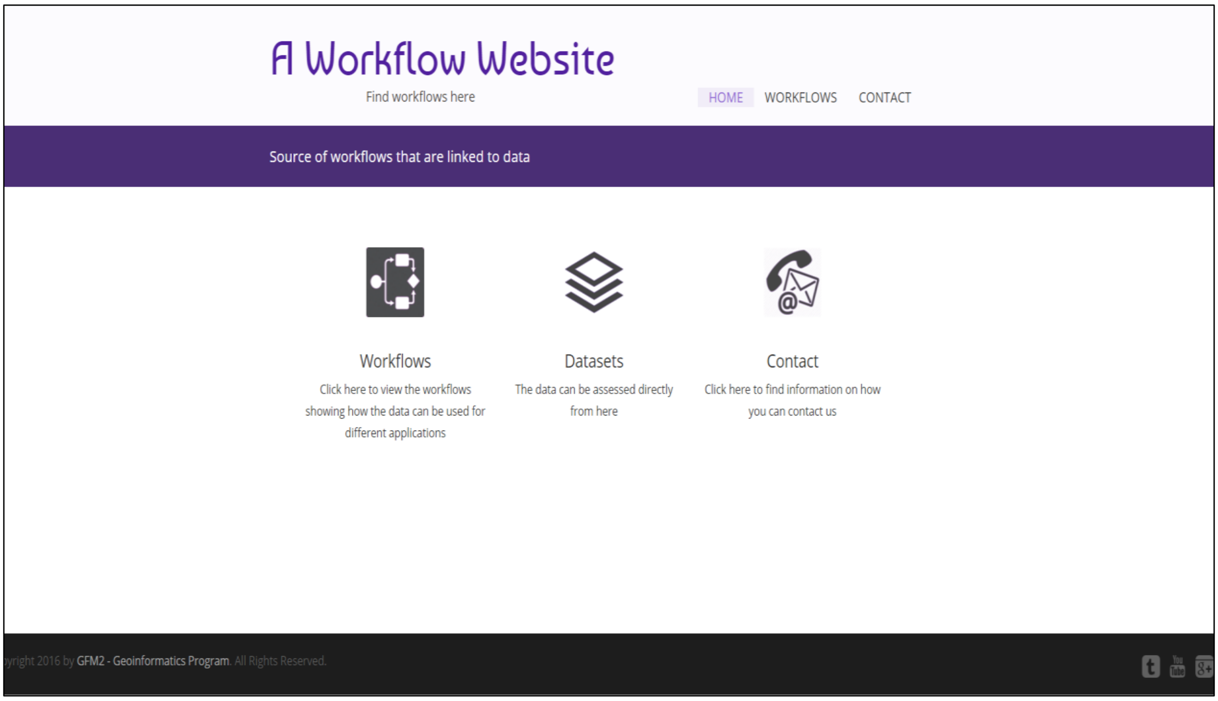


Figure 18: The landing page of the workflow website

The workflows were created separately with clickcharts software (www.nchsoftware.com, 2017) based on literature about road accident analysis. They were drawn following the framework provided by (Carline et al., 2016) The charts were then added to the website using HTML and JavaScript and styled with CSS. The boxes with red text showed the data in the workflow that could be obtained from the application. This helped the user to get an idea of the kind of data to expect. Current implementation showed only two workflows for which the data in the application could be useful. The purpose was to illustrate how workflows could be implemented in a system to help users understand how different datasets could be used. Figure 19 shows the interface where a workflow has been selected. In the future, possibilities to upload more workflows could be incorporated so that the users could upload new workflows of how they have been able to use the data to help other users or correct workflows that are incorrect.

By clicking in the red boxes of the workflows, the user could be redirected to the data page. Alternatively, the user could access the data directly by selecting datasets. The data page was linked to the two data access points to cater for both users who wanted to use the workflows as an access point to the datasets and users who already knew the kind of datasets in the application.

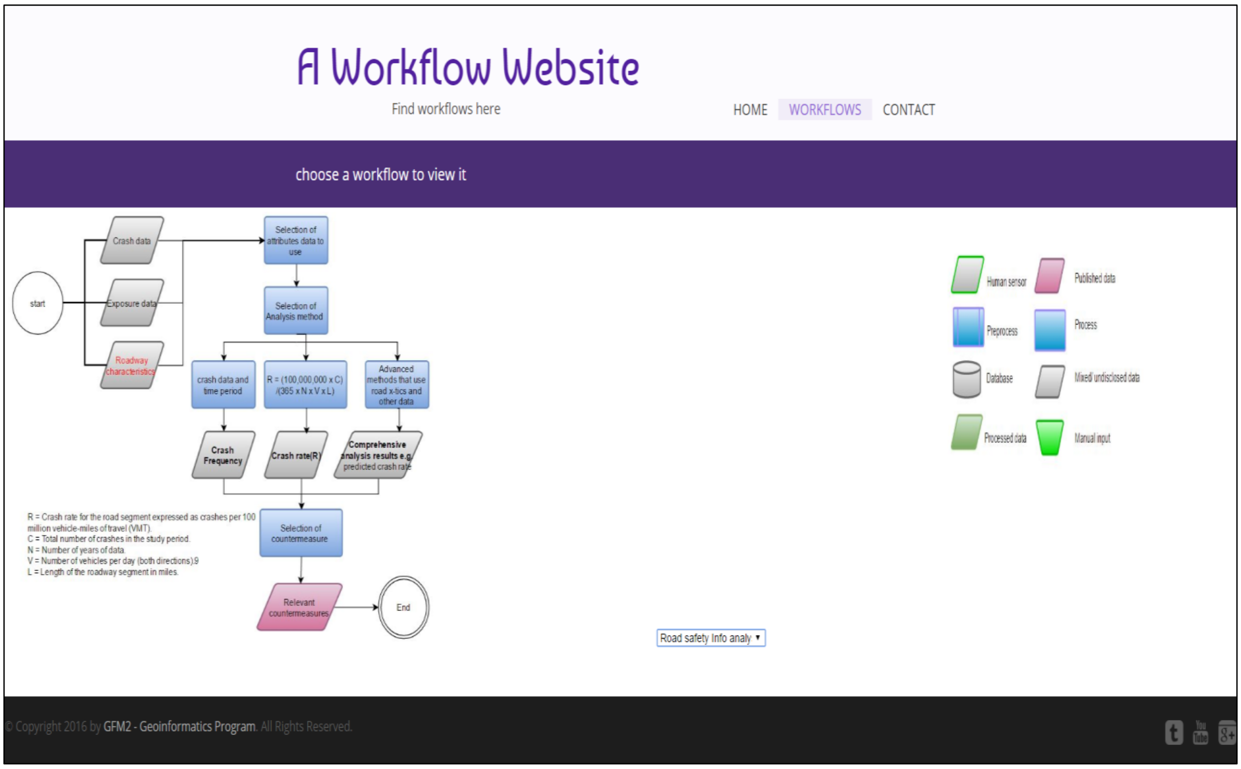


Figure 19: Figure showing a workflow diagram in the website.

4.4.2.2. Data page

This is the page where the all the available datasets in the application could be found. The description of how it was done is given below. Figure 20 give an illustration of the framework of the data page showing data sources, storage, tools used in implementations and the components of the interface. Figure 21 shows the interface of the prototype. The page can be accessed directly using this link: http://win371.ad.utwente.nl/student/s6025013/openlayers/scroll.html

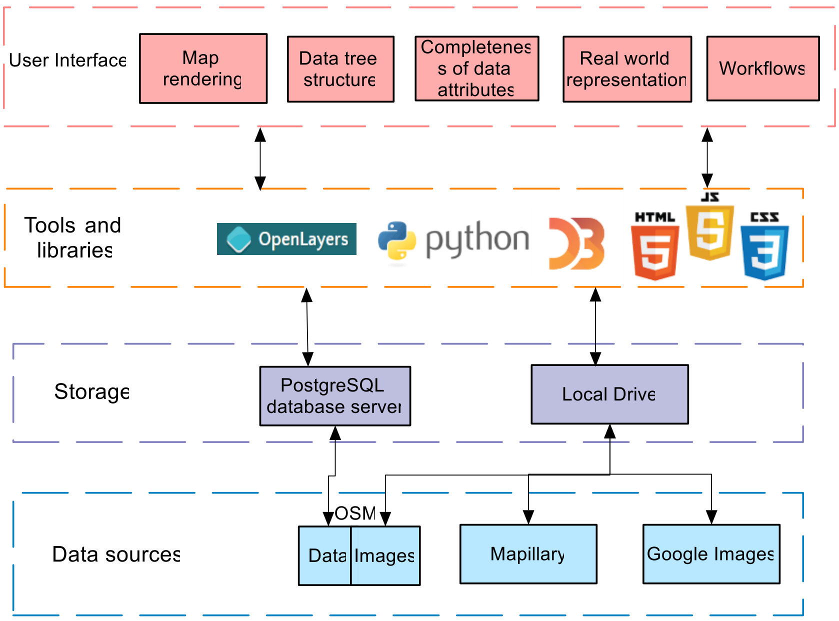


Figure 20: The components of the data page

Figure 21: The interface of the data page

Data tree

This was created based on the data structure of the datasets. The data structure that was followed to create the ontology was followed based on the OSMonto (Codescu et al., 2011). To construct this data tree structure, D3 library was used with Python script to retrieve the data and HTML and CSS to display it on the dashboard.

The data from postgreSQL database was first converted to json format to make the construction of the data structure less complicated. Using D3, hover over functionality was implemented to enable users to know the data type of attributes where applicable. This window was included in the interface to enable the user quickly understand the structure of the datasets and display the attributes with them.

Figure 22: Figure showing the data flow between the software components

Completeness of data attributes

The attributes were retrieved with python script from the PostgreSQL database and displayed with HTML, CSS and javascript. Each of the attributes was attached to a small circle that showed the completeness of the attributes. Red colour circles are for attributes that are incomplete and blue is for attributes that are complete. Partial red and blue shows the percentage of the attributes that are complete. In the implementation, the circles were created for each of the attributes and stored in a local drive and then attached to the appropriate attribute. In the future, this could be made more dynamic where more datasets are involved.

Map rendering

The base map used was from OpenStreetMap. The motivation for using OpenStreetMap data was the free access to current data, the wide and detailed coverage for many parts of the world. In addition, the datasets were also originally obtained from OpenStreetMap. The features displayed in the map were retrieved from PostgreSQL database where they are stored as shapefiles. A python script is used to retrieve them. OpenLayers is used to display the vector features in the map window. The features were only displayed after the user selected the layer for the features.

Real-world representation

The previews used in this research are real-world images obtained from various sources. The images were downloaded from OSM feature page, Mapillary and a few from Google images. Mapillary is an application that enables crowd-sourcing of street-level photos ("Mapillary," 2016). Images from here were used because they are freely accessible and they give are photos for the actual locations places, in this case roads and what they look like (assuming that they are current). The images shown were selected represent the attributes within the data as close to reality as possible. They were then stored in a local drive and retrieved using Javascript and displayed using HTML and CSS.

Workflow diagram

After finding the relevant dataset, the user could click the button of the workflows to check whether the datasets is useful for the particular purpose.

4.4.2.3. Interaction between the elements

The window labelled "Map rendering" displays the features of a selected dataset in the map. Selecting the feature in the map, shows the real-world display of the feature in the window labelled "Real-world representation". The legend for the particular layer is also displayed and the completeness of the data attributes as well. Selecting a feature in the map display also shows the name of the selected feature.

The data tree structure shows the attributes within the dataset of the application. Selecting a node of a particular dataset displays the "completeness of data attributes".

4.4.2.4. Tools and libraries

Openlayers

This is an open source JavaScript library that enables the display of maps in web pages. It was chosen to be used in the application because it is easy to use and can easily be integrated with other application (www.packtpub.com, 2017), for example like D3.

HTML, JavaScript and CSS

Since the prototype created was a web application, HTML, JavaScript and CSS were used. HTML describes the structure of the page, JavaScript is a programming language for web pages and CSS were used for styling the positioning the html elements.

Postgresql database

PostgreSQL was used to store the data for the application. It is easy to use especially in retrieving data for use in a web application.

Python

This is a programming language which was used to create scripts for retrieving data. The language was chosen because it was the researcher was familiar with.

4.5. Conclusion

This Chapter describes the User Centred Design approach which include; requirement analysis, the design and implementation of the prototype. In the requirement analysis the needs of the target users were determined and integrated with the results of the evaluation of geoportals to design the prototype. Different design choices were explained and justified. The tools used in creating the prototype were also discussed in this Chapter. The next stage of the UCD which is evaluation of the prototype is described in Chapter 5.

5. USABILITY TESTING

This Chapter describes the evaluation of the prototype with potential users. The purpose was to achieve the second part of the third objective. The content of this Chapter includes; the objective of the usability testing, the test persons, test setup and execution and the results of the evaluation. Figure 23 shows a flow of work for the usability testing.

5.1. Objective of the usability testing

There were three main objectives for the usability testing, these were;

- To determine whether the use of previews is helpful in providing more information about a dataset
- To determine whether the previews used are useful in determining the fitness for use of a dataset
- To find out how the prototype can be improved where applicable

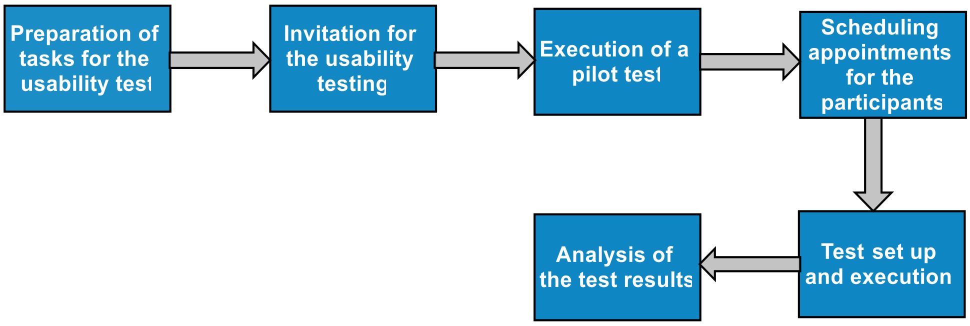


Figure 23: Figure showing the steps followed for the usability testing

5.2. Test preparation

The tasks for the usability testing were in three parts. The first part included questions that were directed towards finding out background information of the participants in regards to using geospatial data. The second part involved using the Inspire geoportal to carry out some tasks. The purpose of this part was to find out how the participants make use of the current geoportals and to find out from them the areas of improvements that they would like to be incorporated in the geoportals to make the assessment of fitness for use of datasets more efficient and effective. The Inspire geoportal was used here as a representative of the current geoportals. There were four questions attached to the first part. They are listed below and the purpose for each question is explained.

<u>**Question 1**</u>

By looking at the information provided about the dataset, can you identify attributes within this dataset?

The purpose of this question was to determine whether the geoportal makes the data attributes of the datasets explicit.

Expected answer: “Yes”

<u>**Question 2**</u>

If yes, which attributes can you identify and how were you able to identify them?

This question was asked to find out how the users identify the attributes for datasets

Expected answer: By reading the resource abstract. Some of the attributes within the dataset were mentioned, these were;

Road segments, road property, road surface, street names

<u>**Question 3**</u>

Do you think this dataset could be useful to you for the given task? Justify your reason.

The purpose of this question was to trigger the user to determine the suitability of the dataset for the given task.

Expected answer: I am not sure because it is not clear whether the attributes mentioned in the abstract are the only attributes within the dataset or there are others. The contents within the datasets are not clear.

<u>**Question 4**</u>

What areas of improvement can you suggest to help you understand the contents of the dataset better?

The purpose of this question was to find out from the users how they would want the geoportals to be implemented to make the data attributes explicit.

Expected answer: Have a list of attributes that stand out so that the user does not have to struggle with looking for where to find the attributes

Provide a clear preview of the dataset

The last part of the test involved tasks that required the participants to use the prototype to carry out some tasks. The aim of this part was to find out whether the prototype provided more information that helped users in determining the fitness for use of a dataset. The test was designed so that it could be completed within one hour.

The questions for this task and the reasons for including them are explained below;

<u>**Question 5**</u>

Does the workflow provide any useful information for you? Justify your answer.

The purpose of this question was to find out whether including workflows provided any additional information about the dataset that could be useful.

Expected answer: Yes, because it showed how a dataset could be used and gives an idea of what the contents could be to the potential user.

<u>**Question 6**</u>

Can you find any data that can be useful in this scenario?

This question was supposed to trigger look at the different previews and try to identify the attributes of the datasets.

Expected answer: Yes

<u>**Question 7**</u>

If yes, which data can you obtain with this application and how were you able to identify it?

This question was asked to identify the different areas in the prototype that the participants looked at to find the attributes within the dataset.

Expected answer: Road characteristics from the workflows

Max speed, traffic utilities, road types from the data tree structure or the completeness

Of the data attributes

__Question 8__

Judging from the information provided in the different previews about the data layers, do you think the data is appropriate for use in this scenario? Justify your answer. (Indicate which is appropriate for use and which is not)

This question tested whether the participant made use of the completeness of the data attributes.

Expected answer: Not all of them because the data for maximum speed is not complete.

__Question 9__

Looking at the track road and the secondary roads, which do you think is more likely to have more accidents on it? Why?

This question tested the relevance of the window entitled "real-world representation". It triggered the participant to use this window.

Expected answer: Secondary roads because they have more traffic on them compared to the track roads and in addition, the track roads seem to be find in residential areas.

__Question 10__

Do you think all the previews provide enough information that help in understanding the data? Explain.

This question was intended to trigger the user to assess the relevance of the different windows in providing more information about the datasets.

Expected answer: Yes, because each gives different information about the dataset

__Question 11__

Currently, the previews provided are static images of the real-world; do you think that dynamic data would be more useful? Explain.

This question was intended to trigger the user to think about ways of advancing the previews provided and determine their usefulness.

Expected answer: Yes, depending of the dataset. In this case more dynamic data would provide the situation as it is on ground which would be more useful in understanding the dataset.

__Question 12__

What suggestions would you give to improve the prototype?

The purpose of this question was to find out from the users how the prototype could be improved.

Expected answer: Linking the information to metadata so that other characteristics of the dataset are also accessible

__Question 13__

What did you think of the test in general? (You can just express your thoughts)

The question was intended to get the views of the participants about the test.

Expected answer: this was an open question. Any answer was possible.

The test was conducted with a laptop that had Camtasia software installed on it to do the audio, video and screen recording during the usability testing. Camtasia was used to record audio, video and the screen of the computer on which the participants were working. The audio provided the thoughts of the test participants as they were performing the tasks and answering the question. The video recording provided the time frame and the reactions of the participants. Screen recording showed how the participants were interacting with the applications. The result provided by Camtasia is a video which contains the audio, video, and screen recording for each participant.

5.3. Test persons for usability testing

After the tasks were prepared, an invitation was sent out to researchers at the ITC, University of Twente. These participants were MSc and PhD students who took on the role of potential users. The prototype was developed with the target users being Researchers in the field of road accidents, however since representatives from the target group could not be found in the given time to do the usability testing, the students at ITC, took on these tasks. The disadvantage with this is that, the results may not be very accurate because the people doing the usability test are not the people for whom it was designed. However, using the researchers still provides some useful information since it is the concept of using semantic previews that is being tested. An invite was sent to students via email for interested persons to take part in the usability testing.

5.4. Pilot test

The pilot test is the first usability test that is done to determine how the application is working. This is a very important step before the actual execution of the usability test. It enables the tester to determine the flows in the application, have better idea of the average time the test could take. The pilot test was done two days before the scheduled tests were to take place. Some flaws in the prototype and the questions were identified and fixed before the actual test. The pilot test took about 30 minutes.

5.5. Scheduing of the tests

The persons interested in participating in the usability testing were asked to choose the most suitable time for them in the time slots created with a Doodle poll. The tests were scheduled for 27/01/2017 and 30/01/2017 from 13:30 to 17:30 hrs. Each participant was allocated 45mins to complete the test.

5.6. Test set-up and execution

The test was done individually by seven (7) test persons at their scheduled time. The test went on smoothly though some of the test persons had to be reminded to think aloud and in some cases the participants had to be asked to follow the instructions provided. The script for the tester is shown in Appendix 1 and the tasks for the users in Appendix 2.

5.7. Test results

This section gives a summary of results of each part from the usability testing.

5.7.1. Part 1

The results shown in the tables are the summary of the background profile of the test participants. Table 4 shows the background information and Table 5 shows the how the test participants interacted with spatial data. This information is relevant in explaining how the participants carried out the tasks and the answered the questions that were asked.

Table 4: Table showing the background information of the test participants

TP	**Age**	**Country**	**Field of study**	**Education**	**Experience (spatial data)**
TP1	25-34	Ghana	Natural Resource Management	Bachelor	3-5 years
TP2	25-34	Kenya	Natural Resource Management	Bachelor	3-5 years
TP3	25-34	Ghana	Land Administration	Masters	3-5 years
TP4	25-34	Brazil	Applied Earth Sciences	Bachelor	Less than 3 years
TP5	35- 44	Ethiopia	Computer Science	Masters	3-5 years
TP6	25-34	Indonesia	Geography	Masters	5-10 years
TP7	18-24	Philippines	Machine Learning	Masters	Less than 3 years

Table 5: Table showing data sources the participants used to obtain spatial data

Question	**Answers**	**Test persons**
Data sources for spatial data that have been used	Geoportal	TP2,TP3,TP4, TP5, TP6, TP7
	Database	TP2, TP5,TP6,TP7
	Web browser	TP3,TP4,TP5, TP7
	Government institution	TP6
Experience with Geoportals	Inspire	TP3, TP7
	GEOSS	TP5
	REMRD	TP2
	New Orleans	TP4
	Philippines	TP7
	Indonesian	TP6
	National Georegister	TP7
How often do you use geoportals?	Weekly	TP4
	Less than once per month	TP7
	2-4 times per year	TP3,TP5,TP6
	Once per year or less	TP2
	Never	TP1
What do you consider to determine the fitness for use of a dataset	Data source	TP1, TP2, TP5, TP6
	Thumbnail previews	TP1, TP3, TP4
	Availability of data attributes	TP1, TP2, TP3, TP5, TP6, TP7
	Metadata	TP1, TP2, TP3, TP4, TP5,TP7
	Validation with other datasets and people	TP6

The graph shown in Figure 24 shows the total time taken for each participant to complete the whole test.

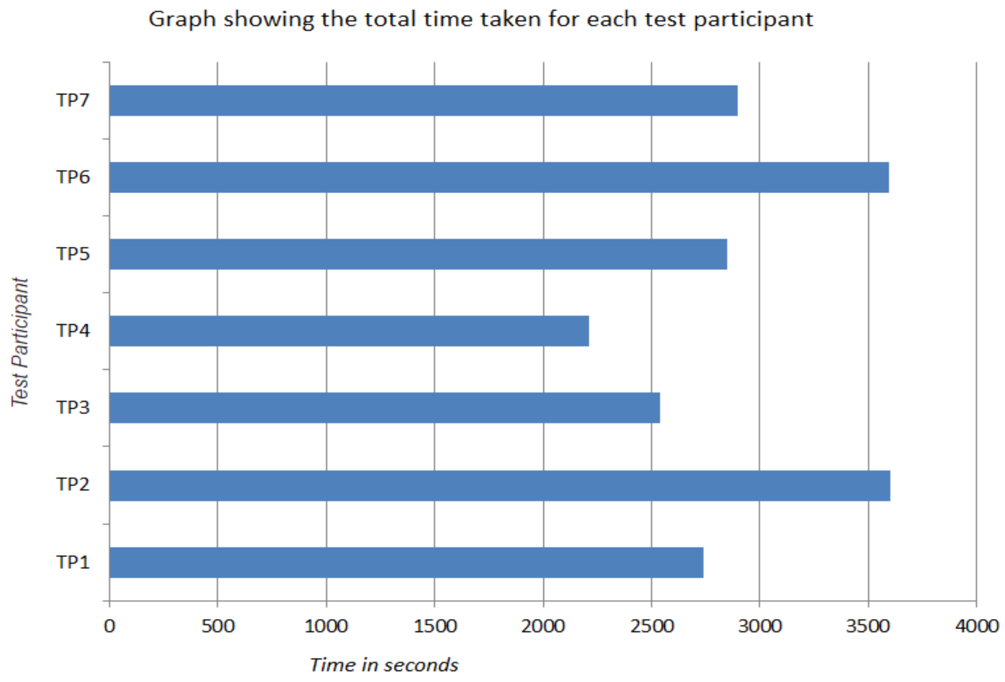


Figure 24: Graph showing the total time taken by each of the participants to complete the test

The graph shown in Figure 25 represents the time taken to answer some of the questions for the usability testing. The time shown in the graph only considers time taken to answer the questions mentioned. It does not consider the time taken to fill in the forms, explore the applications and give recommendations for improvement. The questions represented in the graph required the test persons to complete some tasks. The questions related to tasks completed with the Inspire geoportal are labelled Q1-Q3. The questions related to tasks completed with the prototype are labelled Q5-Q11. The test persons (TP) each have a time line that represents the time taken to complete those tasks. Bars of the same colour on a timeline show similar questions that were carried out using two different applications, i.e. Inspire geoportal and the prototype.

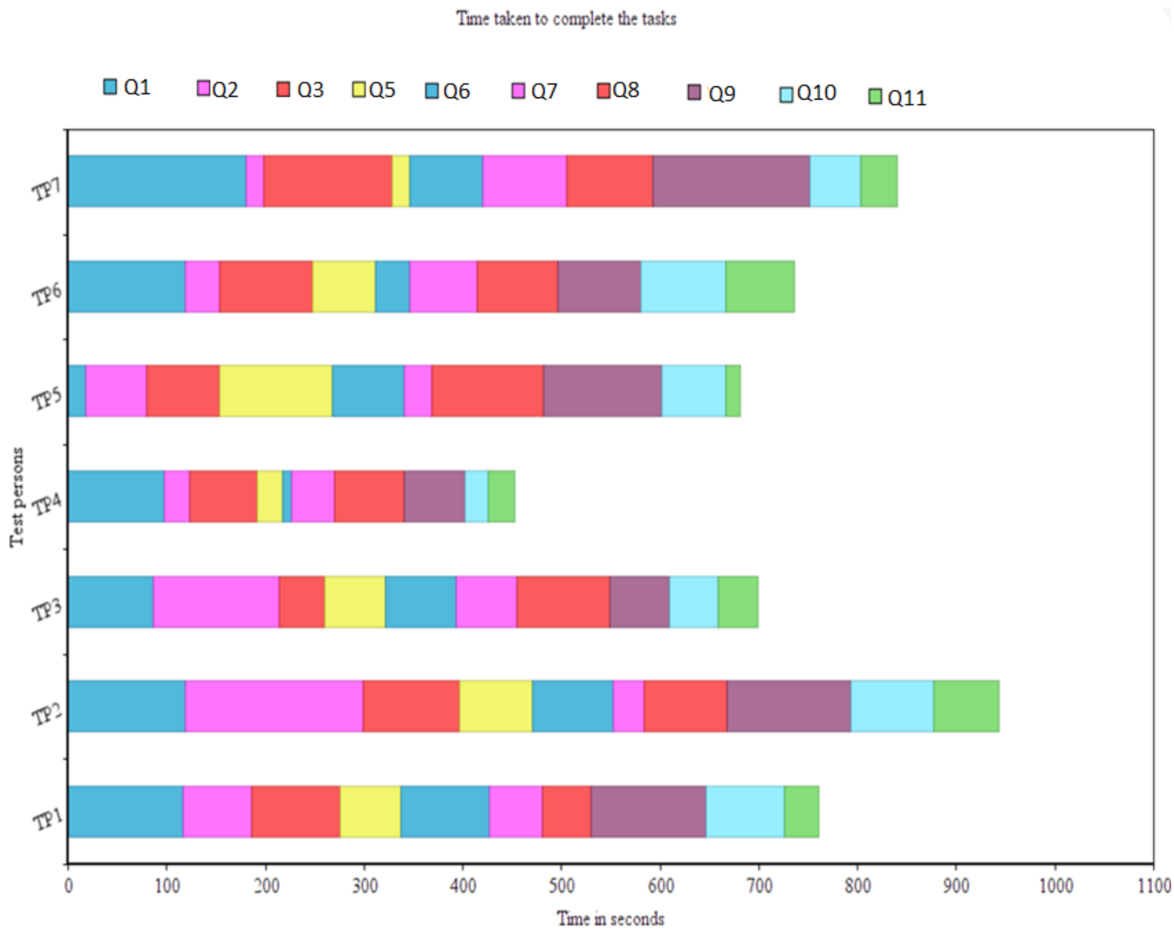


Figure 25: Figure showing the time taken to complete each of the tasks.

The test persons were asked to assume the role of a researcher in the field of road safety. Different questions that required them to use both the Inspire geoportal and the prototype were then asked. For full description, see appendix. Below are the questions;

Inspire geoportal

Q1: By looking at the information provided about the dataset, can you identify attributes within this dataset?
Q2: If yes, which attributes can you identify and how were you able to identify them?
Q3: Do you think this dataset could be useful to you for the given task? Justify your reason.

Prototype

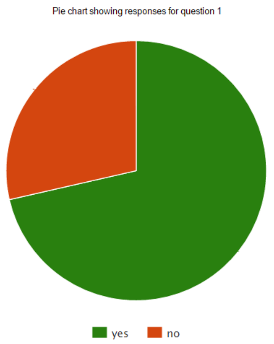
Q5: Does the workflow provide any useful information for you? Justify your answer.
Q6: Can you find any data that can be useful in this scenario?
Q7: If yes, which data can you obtain with this application and how were you able to identify it?
Q8: Judging from the information provided in the different previews about the data layers, do you think the data is appropriate for use in this scenario? Justify your answer. (Indicate which is appropriate for use and which is not)
Q9: Looking at the track road and the secondary roads, which do you think is more likely to have more accidents on it? Why?
Q10: Do you think all the previews provide enough information that help in understanding the data? Explain.

Q11: Current the previews provided are static images of the real-world representation; do you think that dynamic data would be more useful? Explain.

5.7.2. Part 2

This section gives the results of the first task which was completed using the Inspire geoportal.

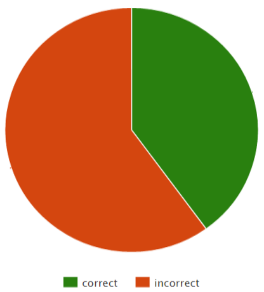
Results for question 1



By looking at the information provided about the dataset, can you identify attributes within this dataset?

Five of the 7 participants answered with yes while two of them (TP4 and TP7) said they could not identify the attributes within the dataset. The two test persons who answered 'no' assumed the attributes were the ones mentioned in the abstract.

Results for question 2

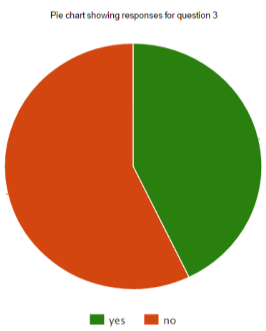


If yes, which attributes can you identify and how were you able to identify them?

For question 2, two of the participants (TP3 and TP6) gave the answers that were expected for this question, the attributes that were mentioned in the abstract. However three of them gave the metadata attributes instead for example spatial extent and date of publication.

Results for question 3

.



Do you think this dataset could be useful to you for the given task? Justify your reason

Three of the participants thought the dataset could be useful and two of the three continued to explain that it would be useful but with limitations. Four of the test participants thought the dataset would not be useful because they could not find the attributes that they were interested in from this dataset.

Results for question 4

What areas of improvement can you suggest to help you understand the contents of the dataset better?
Four of the participants suggested that the data attributes should be made more explicit to the user. The suggestions that were given included; adding some of the attribute data as a table, listing the attributes in the metadata and using visualization to make the attributes stand out. Three suggested more interactivity with the map layer to allow for querying of the features. One suggested more information to for validation of the dataset.

5.7.3. Part 3

This section provides results for part 3 which involved the use of the prototype developed to accomplish the tasks assigned.

Results for question 5

Does the workflow provide any useful information for you? Justify your answer.

All the participants agreed that the workflows were useful and provided information about how the datasets could be used. One of them suggested that including the sources of the workflows would be helpful.

Results for question 6

Can you find any data that can be useful in this scenario?

For this question, all the participants answered yes, i.e. they could find data that was useful for the given scenario.

Results for question 7

If yes, which data can you obtain with this application and how were you able to identify it?

Three of the participants used the tree data structure to identify the relevant data, two used the legend, one used the workflow diagram and one did not answer the question appropriately.

Results for question 8

Judging from the information provided in the different previews about the data layers, do you think the data is appropriate for use in this scenario? Justify your answer. (Indicate which is appropriate for use and which is not)

Six of the participants thought that all the data was useful though one of them stated that more data would be needed. Only one participant thought some of it was not useful like maximum speed limit.

Results for question 9

Looking at the track road and the secondary roads, which do you think is more likely to have more accidents on it? Why?

Five of the participants thought that secondary roads would be more prone to accidents by looking at the "real-world representation" and two of the participants did not understand what the question required and therefore did not give applicable answers.

Results for question 10

Do you think all the previews provide enough information that help in understanding the data? Explain.

Six of the participants found that all the previews were relevant because they provided different information especially the real-world representation. One of them suggested that more text was required to explain the images. One of the participants however thought did not think that they were all relevant because they gave similar information.

<u>Results for question 11</u>

Currently the previews provided are static images of the real-world representation, do you think that dynamic data would be more useful? Explain.

Five participants thought that more dynamic data that shows the real situation would be helpful. Two of the participants however did think it would add much value for this scenario.

<u>Results for question 12</u>

What suggestions would you give to improve the prototype?

Two participants suggested an improvement in the way the legend was presented. It would be better if it was displayed after displaying a particular layer. One of the participants suggested an improvement in the workflows by making it more specific and relating it to the actual data required directly. Other suggestions included making the font bigger, providing a legend for the completeness of data in the prototype, linking the data to metadata, specifying the target users and creating more hyperlinks. Including spatial coverage for the datasets was also suggested.

<u>Results for question 13</u>

What did you think of the test in general? (You can just express your thoughts)

The participants found it quite interesting and very interactive.

5.8. Conclusion

The evaluation of the prototype involved doing usability testing. Seven (7) test parsons volunteered. They were required to provide some background information. The test had two parts: one involved using the Inspire geoportal and the second one involved using the prototype. The test persons were also required to think aloud as they did the test. Test was recorded using Camtasia which was running on a laptop. The result was a video recording which included screen recording of the actions of each participant and a filled out form. The results of the evaluation are analysed and discussed in the Chapter 6.

6. DISCUSSION

This Chapter explains the results from the previous chapters which included; evaluation of geoportals, prototyping and usability testing.

6.1. Evaluation of geoportals

This section explains the current state of geoportals on the basis of the examples that were identified for this research. The areas of interest include the use of keywords, ranking of search results and the contents of metadata that are useful in identifying relevant datasets for a given purpose.

The use of keywords for searching for relevant datasets helps the users in narrowing down the number of datasets that have to be evaluated for the fitness for use for a particular purpose. The geoportals that were evaluated all provided this option. They all used predefined keywords and some even allowed for user-defined keywords. This functionality was helpful and could have been re-used as an initial step in this research to help users in finding relevant datasets.

Some of the geoportals also provided the option of ranking the search results by relevance and in ascending and descending order. However, none of them used ranking of the results in terms of popularity. This would be helpful in quickly identifying and re-using of datasets that other people found useful.

The access to full metadata is important for geoportals because it is what is commonly used in assessing the fitness for use of a dataset. The geoportals that were evaluated all provided full access to metadata. The contents of metadata that were of interest for this research were the use of previews, explicit definition of data attributes and availability of lineage and information about intended use of the datasets. All the geoportals provided an overview map of datasets where applicable however there was limited ability for the users to interact with the map. The problem with the overview maps displayed with the geoportals is that they do not enable the users to understand the features within the dataset.

In addition, the metadata in these geoportals do not clearly state the attributes the potential user of the dataset should expect. This would be helpful in quickly identifying possibly useful datasets and this could be done using text, a preview of the table of the dataset or images. The information displayed in metadata is in form text format which is usually more difficult to read and understand compared to images and other graphical representations. The Inspire geoportal for example provided information about the lineage in a text format which could be displayed as a workflow diagram. Workflow diagrams could give suggestions of how a dataset could be used and also on how it was created. None of the geoportals used workflow diagrams and the only previews shown in some of them were thumbnail previews that do not provide enough information about the data attributes. The use of previews that provide information about the dataset would be helpful in determining the suitability of a dataset for a particular purpose.

6.2. Development of the prototype

The design of the prototype was based on the questions a researcher in the field of road safety would have. A more reliable application could have been achieved if people who fit this user profile were actually presented with mock-ups during the development stage and asked for requirements they would want the system to provide. Because of the limitation in time however, this was not done.

The workflows included were from literature and each component was included to present a different view of the datasets used. The limitation of the prototype was that there were only three datasets presented and therefore the images presented were assigned to the attributes manually. In a real scenario, this would have to be done automatically since there is always a lot of geospatial data made available to users. The components would need to be more interactive and the required data for each component generated atomically so that the data providers do not have to do a lot of work to make their data accessible to other users.

6.3. Usability testing

The testing was done with researchers at ITC, however, they were not experts in the field for road safety. From literature in Chapter 2, it was discovered that there are two main users of geospatial data; experts and non-experts. Of all the participants, only one had more than five years of experience working with geospatial data. Four participants had 3-5 years' experience and two had less than three years of experience. This could have had an impact of the results especially since they were not experts at road safety. The testing was for the concept of the using previews so even though the results may not have been perfect; they give relevant insight into the potential for using previews with geospatial data.

From the usability testing, the first part involved using the Inspire geoportal to find a dataset and assess it for fitness of use. One of the questions involved identifying the attributes of the dataset (contents within the dataset). Three of the participants identified metadata attributes which was not what the question was asking. Two of these participants had a background in natural resource management and one from computer science. Their background could have been the reason they misunderstood the question. Probably attributes in their field of study usually refer to metadata data characteristics. Two of the participants were able to identify the contents within the dataset. Although the abstract mentions some of them, it does not make it obvious to the user because two other participants were not able to identify these contents even though they saw the abstract. In addition, it takes time to read the abstract. This shows that the attributes need to be made more explicit to the use using visualization so that the user does not spend a lot of time trying to identify them. In question 3 of the first part, participants were required to determine whether the dataset could be useful for the given tasked. Many of them expressed that they would need more details about the dataset to be able to answer that. This shows that more details about the dataset need to be made available to the users to enable them to answer this question about the datasets.

The timeline graph shown in Figure 24 showed the time it took to complete the tasks for each user. There were two tasks that were performed with both the prototype and the Inspire geoportal. The first one required the test persons to identify the attributes of the datasets. Most of the test persons took a longer time using the geoportal than using the prototype. This shows that making the attributes more obvious accessible saves the user's time in finding relevant datasets. The second task required them to determine the suitability of the datasets for use. For this question, some of the participants took longer using the geoportal and others spent more time while using the prototype. This shows that a more.

The second part of the usability test involved the use of the prototype that was implemented for this research. The purpose was to test the relevance of the different components that were implemented. The use of workflows seemed to be useful to the participants since most of all of them stated that it showed how the dataset could be used. The workflows presented in the prototype showed processes where the data could be used. It would have been more prudent to compare lineage form the Inspire geoportal with workflow diagrams that show the processes the dataset has undergone in the prototype. This was not done because there was no detailed information about these processes. Including workflows in geoportals that suggest how a dataset could be used could be helpful to potential users looking to find relevant datasets. This would however be more meaningful if it is applied for datasets that are dedicated for a particular field, or else, the sources of the workflows should be indicated so that they can be verified by the user. The trusted source of the workflow would probably be the producer of the dataset to show how he intended the data to be used. The use of workflows also provides information faster that having to read a lot of text. The possibility uploading workflows by people who have made use of the datasets would also help those looking for datasets to identify faster datasets relevant to them. Workflows were only used at the initial stage while trying to find a relevant dataset; this means that looking at the workflow at the beginning helps people weed out irrelevant datasets. A link between workflows and datasets are also helpful.

The use of real-world images assigned to the datasets also seemed to give more insight to the participants about the dataset. This is an option that could be explored in making a dataset more understandable. This was useful to the participants because they were not familiar with the data or the location from which the data was taken. In addition, they were not experts in this field. However, this also goes to show the users of spatial data in the world today where the producers are not necessarily the users of the data. The presentation of the context in which the data is meant is therefore useful in helping a potential user of a dataset understand and give better judgement as to whether it is useful or not. The use of images to present the actual situation the data is presenting seems to be helpful. The use of images seems to be better received that text, reading text about a dataset. This more tests would have to be done to provide a more solid argument.

Many of the test participants used the data tree structure to try and understand the contents of a dataset. The use of search a system provided the participants with a quick understanding of the contents with the datasets. The prototype only showed three datasets so the use of a data tree was convenient. However, in real-world situation where many datasets covering different fields are involved, a way of querying the structure would probably be more helpful. For example an ontology diagram that represents the different datasets could be created where if a user searches for a particular dataset by typing "roads", only related datasets and concepts are highlighted.

One of the items that people looking for datasets consider is how complete that data is. In the prototype, this was presented in the prototype by using circles with red colour to show incomplete data and blue colour for complete data. Only two of the participants used this to assess the completeness of the data attributes. This seems to suggest that the completeness of the data attributes need to be presented in a more obvious way to the user. However, it could also suggest that people consider the completeness of a dataset at later stages.

6.4. Metadata handling and standards

The OGC (Open Geospatial Consortium) sets standards for catalogue interfaces (CSW) that help in regulating them (OGC, 2016). The OGC Catalogue service is associated with three main classes; OGC service class, Discovery class and Manager Class. Each of these classes provides a different functionality in the catalogue service. Based on this research, this can be related to a dashboard that was used to develop the prototype for this research. Just like the different components of the dashboard, more provision can be made on the catalogue service interface to accommodate data from other sources to help enrich the datasets. These catalogue service interfaces can be related to data sources for example Mapillary (for real-world images) and OSM (for ontologies) which can be queried to help in the discovery of datasets.

For the standards to be effective, producers or publishers of datasets would have to include data attributes in the metadata they provide and these would have to be understood for the right previews to be generated. This would probably require the some of the common terms of data attributes to be agreed upon or else a forum would have to be incorporated that allows producers to specify that previews of their data as is the case with OSM. Because otherwise, different people name data attributes differently and some of the names may be understood by other users and in addition, with many names, the automation of generating dataset previews would be more difficult. This therefore would require the standardization of attribute names. This is recommended to ensure that data is correctly assigned to the semantics it represents.

6.5. Conclusions

In this Chapter the results at the different stages of the research were discussed. Chapter 7 contains the conclusions derived from the research.

7. CONCLUSIONS AND RECOMMENDATIONS

7.1. Overview

Current geoportals make use of metadata to create previews of geospatial data; however, the metadata does not provide enough information about the datasets to enable the users to explicitly identify their attributes. This research explored the use of semantic previews to make these data attributes explicit to the potential users of the datasets. A prototype was designed that incorporated data structure, real-world representation, completeness of data attributes, map rendering and a link to workflows on its interface. The prototype was then evaluated to determine whether semantic previews provide more information about datasets and if they are useful in determining their fitness for use. From the usability testing, it was discovered that semantic previews are useful in understanding the contents of a dataset. The availability of data attributes also helps in assessing the fitness for use of the dataset. The completeness of the data attributes however was not used by the many test persons probably because it was not very clear to them. Therefore a more obvious solution needs to be implemented. One of the challenges that can be foreseen is finding the relevant previews to present for the different fields. This is because different people understand things differently and therefore would require a platform that takes all this into consideration. This would require the providers of the data to provide more information about a datasets since they are the most knowledgeable about it, to enable other users to understand it better.

7.2. Answers to the research questions

Questions for objective 1

1. **What are the strengths and weaknesses of existing geoportals in terms of discovery of relevant data?**

The strength of the current geoportals lies in the provision of metadata that gives the characteristics of the datasets, for example the producer of the datasets, the spatial extent of the dataset among others. Some of them even follow the metadata standards that were set to allow for discovery of the datasets. The weakness of these geoportals is that most of the information is provided in text format so the user needs to read all of it to have an idea of what the dataset is about. In addition, the contents of the datasets are usually not clearly stated in the metadata which means the user has to just assume and therefore spend more time trying to identify useful datasets for his particular purpose.

2. **How can these geoportals be made more efficient and effective for discovery of relevant data?**

The attributes contents of the datasets can be made more obvious to the user by using graphical display to enable users to save time they would have spent reading all the text that is available about the resources in the metadata.

3. **What can be the role of (semantic) previews?**

Semantic previews can be used to display the most important features of a dataset that help the potential user in assessing the suitability of a dataset for a particular purpose. Data structure and images that are interactive can be used to make the data attributes within the datasets more obvious to the user. Providing the user with graphical representation of the elements they consider helpful and important in assessing the fitness for use of a dataset.

Questions for objective 2

1. **Who are the users that the research is targeting and what are their requirements?**

This research narrowed down the user group to researchers in the field of road safety. This is covered in Section 4.3. However, the potential users of this research in general include everyone who uses geospatial data. Based on literature in Chapter 2, users of geospatial include both experts and non-experts. These people are interested in the attributes of the dataset that enable them to understand the dataset and therefore determine its suitability for a particular purpose. While searching for datasets, the users consider mainly the attributes of the dataset, the metadata and the data source. This research would be able to provide them with previews that would enable them to understand these datasets.

2. What kind of geo-datasets are of interest to these users?

The researchers are interested in datasets that provide information used in analysis of road accidents. The attributes of interest include; road type, road length, road curvature, road surface, maximum speed limit, utilities for example traffic signals. In a wider context however, geospatial datasets would include those from VGI sources and satellite images.

3. How do users make use of previews?

This question was answered during the usability testing. Previews are used to give insight about the data without the access to the data as a kind of overview. All the test persons looked at previews as a link to more information about the dataset. Many of them expected hyperlink to more detailed information about the data.

Questions for objective 3

1. How can the previews be semantically enriched, e.g., by using Semantic Web technology?

Based on literature in Chapter 2, ontologies are supported by the current Semantic web. These ontologies can be created for geospatial data and incorporated into existing systems as ontology diagram or data structure to help in the discovery of more information about the dataset. In the prototype where the data tree structure was used, many of the test persons utilised it for determining the contents of the dataset. Most of the test persons also found the real-world representation of the features helpful in understanding the dataset better. The use of images can therefore be explored to give more information about data attributes to people who are not familiar with the data.

2. What tools are suitable for implementing comprehensive previews?

Openlayers is very easy to learn and use when it comes to rendering vector and raster data. In addition, OpenLayers can be integrated with D3 libraries. D3 is very dynamic and therefore very useful in implementing new systems that support visualizations. The images used in the implementation of the prototype were from Mapillary. Mapillary API was used to retrieve real-world images for the area of coverage of the data. The images were free and therefore convenient to use for area which are represented. OSM feature page also provided some of the images for features that were not covered by Mapillary. For this research, there was no detailed study about the available tools for implementing previews. The tools selected were based on popularity, technical limitation and knowledge of the researcher. This means that there are probably other tools that can be used to perform the same functions. Because the prototype was a web application, JavaScript, HTML and CSS which are very popular tools for designing web applications were also used.

3. What are the strengths and weaknesses of the prototype?

Strengths: the prototype makes the data attributes more obvious to the user and also enables them to interact with the dataset and therefore be able to understand it better. Images of the real-world are used which give insight as to what the situation is and also the use of workflows that suggests potential use of the dataset.

Weaknesses: Lack of interactivity between some components of the prototype makes it a bit confusing for the user. In addition, the user has no way of querying the data and where for the case of many datasets it would be more difficult to use. The source of the workflows was not indicated which makes it difficult to rely on and also interface needs to be re-arranged so that some things are more intuitive to the user like in the completeness of data attributes so that they can make use of it assessing datasets.

4. What can be done to improve the prototype?

The prototype can be improved by including more datasets and making the components more interactive with each other for example the data tree and the map such that when an element is selected in the data structure, it is reflected in the map and vice versa. Links to metadata would also be helpful so that the user is provided with all the information about the dataset in one place.

Questions for objective 4

1. **How can the prototype be integrated into a catalogue service?**

The OGC standard for catalogue services defines three classes which are discussed in Section 6.4. These classes can be increased to accommodate images from other geospatial data sources like Mapillary and OSM to enrich the catalogue service and provide alternative means of querying data from sources people are familiar with. By integrating different data sources, uses are provided with more options that they can utilise in understanding the datasets and therefore they can choose to use the ones they are familiar with. The CSW could also provide links between datasets, by comparing their preview elements (in data hierarchy, real-world representations, or the data attributes).

2. **How can metadata standards be improved, based on this research?**

This has not been covered by this research. However, from the usability testing, it was suggested that the geoportals needed to make the attributes more explicit. The prototype explicitly provided this. For the previews to be incorporated into geoportals, some standards have to be agreed upon. OSM for example has established standards for the tags. A similar approach can be adopted or even re-used where applicable to facilitate uniformity in data. Enabling of the users to upload their own data or enrich the images and terms in a controlled way.

7.3. Recommendations

7.3.1. Recommendations for the prototype

It is recommended that the elements of the prototype be made more interactive with each other so that with just a click, the used is instantly availed various information about the dataset of interest. For example by displaying a layer, the legend for that layer is displayed instantly.

Another recommendation for the improvement of the prototype is to incorporate metadata of the datasets so that users have not just information about the contents of the data but the characteristics of the dataset as well.

The possibility of linking the workflows directly to the actual data that is required for the use cases that are represented is also recommended. The users can also be provided with the option to upload their own workflows showing how they used the dataset. this could be helpful to other users of the dataset.

7.3.2. Recommendations for research

Based this research, increasing the CSW classes is recommended so that more information about datasets can be presented to users and therefore eliminate ambiguity about a dataset. In addition, standards could be considered in the naming of attribute data so that it is names that can be easily identified are incorporated and used with ontologies to enable people have more access to relevant data for particular purposes.

One of the challenges in using previews is identifying the right previews to show different types of data to different classes of people. This could be an interesting area of research to determine whether all classes of people understand the previews in the same way. This would be helpful in future integration to the existing systems.

REFERENCES

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-Centered Design. *Encyclopedia of Human-Computer Interaction*.

Aditya, T., & Kraak, M.-J. (2006). Geospatial Data Infrastructure Portals: Using the National Atlas as a Metaphor. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *41*(2), 115–134. http://doi.org/10.3138/P3P1-5327-X2QK-616M

Anandhan, A., Dhandapani, S., Reza, H., & Namasivayam, K. (2006). Web Usability Testing ― CARE Methodology. In *Third International Conference on Information Technology: New Generations (ITNG'06)* (pp. 495–500). IEEE. http://doi.org/10.1109/ITNG.2006.141

Angeletou, S. (2008). *Semantic Enrichment of Folksonomies*.

Athanasis, N., Kalabokidis, K., Vaitis, M., & Soulakellis, N. (2009). Towards a Semantics-Based Approach in the Development of Geographic Portals. *Computers & Geosciences*, *35*(2), 301–308. http://doi.org/10.1016/j.cageo.2008.01.014

Begelman, G., Keller, P., & Smadja, F. (2006). Automated Tag Clustering: Improving search and exploration in the tag space. *Collaborative Web Tagging Workshop at WWW2006 Edinburgh Scotland*, *22*(2), 15–33. http://doi.org/10.1.1.120.5736

Bevan, N., Carter, J., & Harker, S. (2015). ISO 9241-11 Revised: What Have We Learnt About Usability Since 1998? In *LNCS* (Vol. 9169, pp. 143–151). http://doi.org/10.1007/978-3-319-20901-2_13

Bogdanović, M., Stanimirović, A., & Stoimenov, L. (2015). Methodology for geospatial data source discovery in ontology-driven geo-information integration architectures. *Web Semantics: Science, Services and Agents on the World Wide Web*, *32*, 1–15. http://doi.org/10.1016/j.websem.2015.01.002

Braggaar, R. (2016). Dutch national geoportals: fit for use or work in progress? An assessment of the geoportals and suggestions for future improvements. *ReasearchGate*. Retrieved from https://www.researchgate.net/publication/303720327

Bruin, S. de, Bregt, A., & Ven, M. van de. (2001). Assessing fitness for use: the expected value of spatial data sets. *International Journal of Geographical Information Science*, *15*(5), 457–471. http://doi.org/10.1080/13658810110053116

Carline, A., Pinar, B., Evelien, B., & Deniz, K. (2016). *Sharing Workflows for Volunteered Geo-Information*.

Clarke, M., & Harley, P. (2014). How Smart Is Your Content? Using Semantic Enrichment to Improve Your User Experience and Your Bottom Line. *Science Editor*, *37*(2), 5.

Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., & Rau, R. (2011). OSMonto - An Ontology of OpenStreetMap Tags. In *State of the Map Europe*. Retrieved from https://pdfs.semanticscholar.org/8466/8a0ada3eb1cb08afdaddd5cd60ac0ba5cda9.pdf

Delikostidis, I. (2011). *Improving the Usability of Pedestrian Navigation Systems*. University of Twente.

Devillers, R., Bédard, Y., & Jeansoulin, R. (2005). Multidimensional Management of Geospatial Data Quality Information for its Dynamic Use Within GIS. *Photogrammetric Engineering & Remote Sensing*, *71*(2), 205–215. http://doi.org/10.14358/PERS.71.2.205

Devillers, R., Gervais, M., Bédard, Y., & Jeansoulin, R. (2002). Spatial Data Quality : From Metadata to Quality Indicators and Contextual End - User Manual. In *OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management* (pp. 45–55). IEEE. http://doi.org/10.1.1.460.4972

Dictionary.com. (2016). Retrieved November 8, 2016, from http://www.dictionary.com/browse/preview

Elzakker, C. P. J. M. (2004). *The use of maps in the exploration of geographic data Nederlandse Geografische Studies / Netherlands Geographical Studies Plaatselijke Redacteuren / Local Editors Redactie-Adviseurs / Editorial Advisory Board*. Utrecht University. Retrieved from https://www.itc.nl/library/Papers_2004/phd/vanelzakker.pdf

Elzakker, C. P. J. M., & Wealands, K. (2007). 34 Use and Users of Multimedia Cartography.

Few, S. (2006). Clarifying the vision. In *Information Dashboard Design The Effective Visual Communication of Data* (p. 223).

Fonseca, F. (2013). Geo-Ontologies, (September). Retrieved from https://www.researchgate.net/publication/40741418

Fonseca, F. T., & Egenhofer, M. J. (1999). Ontology-driven geographic information systems. In *Proceedings of the seventh ACM international symposium on Advances in geographic information systems - GIS '99* (pp. 14–19). New York, New York, USA: ACM Press. http://doi.org/10.1145/320134.320137

GeoViQua. (2014a). GeoLabel. Retrieved August 23, 2016, from http://www.geoviqua.org/GeoLabel.htm

GeoViQua. (2014b). Overview. Retrieved August 23, 2016, from http://www.geoviqua.org/Overview.htm

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211–221. http://doi.org/10.1007/s10708-007-9111-y

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, *5*(2), 199–220. http://doi.org/10.1006/knac.1993.1008

Gulliksen, J., Ransson, B. G., Boivie, I., Blomkvist, S., Persson, J., & Sa Cajander, Å. (2003). Key principles for user-centred systems design. *BEHAVIOUR & INFORMATION TECHNOLOGY*, *22*(6), 397–409. http://doi.org/10.1080/01449290310001624329

Haklay, M., & Weber, P. (2008). OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*, *7*(4), 12–18. http://doi.org/10.1109/MPRV.2008.80

Haworth, B. (2016). Emergency management perspectives on volunteered geographic information: Opportunities, challenges and change. *Computers, Environment and Urban Systems*, *57*, 189–198. http://doi.org/10.1016/j.compenvurbsys.2016.02.009

http://www.nationaalgeoregister.nl. (2017). Nationaal georegister. Retrieved February 12, 2017, from http://www.nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/home

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the Information Age Solving Problems with Visual Analytics*. Eurographics Association. Retrieved from http://diglib.eg.org/handle/10.2312/14803

Kitchin, R., Lauriault, T. P., & McArdle, G. (2015). Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, *2*(1), 6–28. http://doi.org/10.1080/21681376.2014.983149

Klien, E. (2007). A Rule-Based Strategy for the Semantic Annotation of Geodata. *Transactions in GIS*, *11*(3), 437–452. http://doi.org/10.1111/j.1467-9671.2007.01054.x

Kraak, M.-J. (2003). Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing*, *57*(5), 390–399. http://doi.org/10.1016/S0924-2716(02)00167-3

Kuhn, W. (2005). LNCS 3534 - Geospatial Semantics: Why, of What, and How? *Journal on Data Semantics III LNCS*, *3534*, 1–24. http://doi.org/10.1007/11496168_1

Lanter, D. P. (1990). Lineage in GIS: The Problem and a Solution. *National Center for Geographic Information and Analysis*.

Maguire, D. J., & Longley, P. A. (2005). The Emergence of Geoportals and their Role in Spatial Data Infrastructures. *Computers, Environment and Urban Systems*, *29*(1), 3–14. http://doi.org/10.1016/j.compenvurbsys.2004.05.012

Mapillary. (2016). Retrieved January 3, 2017, from https://www.mapillary.com/

McArdle, G., & Kitchin, R. (2016). THE Dublin Dashboard: Design and Development of a Real-Time Analytical Urban Dashboard. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *IV-4/W1*, 19–25. http://doi.org/10.5194/isprs-annals-IV-4-W1-19-2016

Mendoza, G. A., Song, B., & Mladenoff, D. J. (2006). Visualization with Spatial Data. *Computer Applications in Sustainable Forest Management*, 127–142.

Merriam-Webster. (2015). Retrieved August 8, 2016, from http://www.merriam-webster.com/dictionary/preview

Merriam-Webster. (2016). Retrieved November 8, 2016, from http://www.merriam-webster.com/dictionary/semantics

Mishra, A., & Mishra, R. P. (2012). *Global Trends in Computing and Communication Systems*. (P. V. Krishna, M. R. Babu, & E. Ariwa, Eds.)*Communications in Computer and Information Science* (Vol. 269). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-29219-4

Mobasheri, A., Van Oosterom, P., Zlatanova, S., & Bakillah, M. (2013). Semantic Annotation of existing Geo-Datasets: A Case Study of Disaster Response in Netherlands.

Nöllenburg, M. (2006). Geographic Visualization. In *Human-Centered Visualization Environments* (pp. 257–294). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-540-71949-6_6

OGC. (2016). Open Geospatial Consortium OGC ® Catalogue Services 3 . 0 - General Model, 1–91.

Paulolyslager.com. (2017). Paul Olyslager. Retrieved February 6, 2017, from https://www.paulolyslager.com/tagclouds-useful-keyword-spam/

Pauwels, K., Ambler, T., Clark, B. H., Lapointe, P., Reibstein, M. D., Skiera, B., … Wiesel, T. (2009). Dashboards as a Service Why, What, How, and What Research Is Needed? http://doi.org/10.1177/1094670509344213

Pôças, I., Gonçalves, J., Marcos, B., Alonso, J., Castro, P., & Honrado, J. P. (2014). Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring. *International Journal of Geographical Information Science*, *28*(11), 2356–2371. http://doi.org/10.1080/13658816.2014.924627

Rasheed, Z., & Shah, M. (2002). Movie Genre Classification By Exploiting Audio-Visual Features Of Previews. In *16th International Conference on Pattern Recognition (ICPR)* (pp. 1086–1089). Quebec City, Canada: IEEE Computer SOC.

Ristoski, P., & Paulheim, H. (2016). Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey. *Web Semantics: Science, Services and Agents on the World Wide Web*, *36*, 1–22. http://doi.org/10.1016/j.websem.2016.01.001

Rivard, K., & Cogswell, D. (2004). Are You Drowning in BI Reports? Using Analytical Dashboards to Cut Through the Clutter. *In- Formation Management*, *14*(4), 26. Retrieved from www.dmreview.com

Semantics | Definition of Semantics by Merriam-Webster. (2017). Retrieved January 26, 2017, from https://www.merriam-webster.com/dictionary/semantics

Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *IEEE Sympsium on Visual Languages* (p. 336-). Washington, DC, USA. Retrieved from http://www/cs.umd.edu/projects/hcil/

Shneiderman, B., Byrd, D., & Croft, W. B. (1997). Clarifying Search: A User-Interface Framework for Text Searches. Retrieved from http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html

Sinclair, J., & Cardew-Hall, M. (2007). The folksonomy tag cloud: when is it useful? *Journal of Information Science*, *34*(1), 15–29. http://doi.org/10.1177/0165551506078083

surveymonkey.com. (2017). SurveyMonkey. Retrieved January 30, 2017, from https://nl.surveymonkey.com/

TechSmith. (2016). Best Practices for App Previews. Retrieved August 23, 2016, from https://www.techsmith.com/techsmith-appshow-best-practices.html

techtarget.com. (2017). TechTarget. Retrieved January 30, 2017, from http://searchsoftwarequality.techtarget.com/definition/requirements-analysis

UNESCO-IHE. (2016). Africa and Europe to strengthen links on Water and Climate: AfriAlliance.

Retrieved August 23, 2016, from https://www.unesco-ihe.org/news/africa-and-europe-strengthen-links-water-and-climate-afrialliance

Usabilityfirst.com. (2015). Usability First - Methods - Usability Testing | Usability First. Retrieved January 30, 2017, from http://www.usabilityfirst.com/usability-methods/usability-testing/

Vandenbroucke, D., Janssen, K., & Orshoven, J. Van. (2008). INSPIRE State of Play: Generic approach to assess the status of NSDIs. In J. Crompvoets, A. Rajabifard, B. Van Loenen, & T. D. Fernández (Eds.), *A Multi-View Framework to Assess SDIs*. Australia.: Digital Print Centre, The University of Melbourne.

Vockner, B., & Mittlböck, M. (2014). Geo-Enrichment and Semantic Enhancement of Metadata Sets to Augment Discovery in Geoportals. *ISPRS International Journal of Geo-Information*, *3*(1), 345–367. http://doi.org/10.3390/ijgi3010345

www.nchsoftware.com. (2017). ClickCharts Charting, Mapping & Flowchart Software. Retrieved February 9, 2017, from http://www.nchsoftware.com/chart/

www.packtpub.com. (2017). Advantages of using OpenLayers. Retrieved February 12, 2017, from https://www.packtpub.com/mapt/book/web-development/9781782162360/1/ch01lvl1sec11/advantages-of-using-openlayers

www.techtarget.com. (2017). What is workflow? - Definition from WhatIs.com. Retrieved February 9, 2017, from http://searchcio.techtarget.com/definition/workflow

www.wiki.openstreetmap.org. (n.d.). LinkedGeoData - OpenStreetMap Wiki. Retrieved February 9, 2017, from http://wiki.openstreetmap.org/wiki/LinkedGeoData

www.wiki.openstreetmap.org. (2017). Map Features - OpenStreetMap Wiki. Retrieved February 9, 2017, from http://wiki.openstreetmap.org/wiki/Map_Features#Roads

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41–59. http://doi.org/10.1016/j.accinf.2011.08.002

APPENDICES

Appendix 1

Before the test

- Arrive at the room for testing 30 minutes prior to the arrival of the test person
- Ensure that it is clean and comfortable for the test persons
- Switch on the lap
- Go to the Inspire geoportal
- Open the application of the prototype
- Check whether the application is working fine
- Check Camtasia by making a sample recording
- Place the hard copy of the tasks and instructions of the users on the table where they will be seated
- Place a pen and ensure that it works well

Introduction of the test person

- Welcome the test person by name
- Invite them to take a seat
- Ask them to ensure that they are comfortable
- Present the hard copy of the tasks and instructions
- Ask the test person to speak their thoughts as they do the test and use mouse pointer to refer to something on the screen during tasks
- Ask whether the test person has any questions before we can begin the test
- Ask the test person to inform you when they are through with a particular task
- Present the first task for the user to fill in.

During the test

- Start Camtasia
- When the test person is through with the pre-test, give them the first part of the test and show them where the Inspire geoportal is.
- Give the user the second part of the task when first task is completed and show them the application prototype to use for the second part of the tasks
- After they are through with the test, ask the user what their thoughts are on the test
- After the test, press F10 to stop recording
- Ensure that the recording is saved

After the test

- Ask the test person what they thought of the test
- Thank them for their participation and wish them well

Appendix 2

Usability testing

Introduction

The test will be split into two parts; the first part of the test is using Inspire geoportal and the second part will be testing an application prototype that I have designed that uses previews.

Objective of the testing

- To determine whether the use of previews is helpful in providing more information about a dataset
- To determine whether the previews used are useful in determining the fitness for use of a dataset
- To find out how the prototype can be improved where applicable

Pre-test questions

1. How old are you?
   - 18-24
   - 25-34
   - 35-44
   - 45- 54
   - Above 55

2. What is your country of origin? Please indicate.
   …………………………………………………………………………...

3. What is your field of study?
   ……………………………………………………………………………………………………
   ……………………………………………..

4. What is your highest educational degree?
   - Bachelor
   - Master
   - PhD
   - Other (please specify)……………………………………………….

5. How much experience do you have working with geospatial data in your studies and/or in your professional life? (Only one answer possible)
   - Less than 3 years
   - 3-5 years
   - 5-10 years
   - More than 10 years

6. Which sources do you use when you are looking for a dataset to use? (More than one answer possible)
   - Geoportal
   - Database
   - Web browser
   - Others
     (Specify)………………………………………………………………………………………
     ……………………………………………………………………………………………………

7. If you have used geoportals before, which geoportal have you used?(More than one answer possible)
- Inspire Geoportal
- GEOSS
- Nationaal Georegister
- National geoportal (please specify)………………………………… ………………………………………………………………………

8. How often do you use geoportals?(Only one answer possible)
- Weekly
- Fortnightly (every 2 weeks)
- Monthly
- Less than once per month
- 2-4 times per year
- Once per year or less
- Never

9. How do you normally determine the fitness for use of a dataset for your specific purpose? By looking at : (More than one answer possible)
- The data source
- Thumbnail previews of that dataset
- Availability of required data attributes
- Metadata of the dataset
- Other (specify)……………………………………………………

Part 1

Inspire geoportal

A. Explore the Inspire geoportal (This should take utmost 5mins)

From the Inspire geoportal, try to find a dataset of your choice following the instructions below;

- In the search box on the top right window, type in a keyword word of your choice to help you find a dataset of your interest.
- From the search results, select one and try to understand what the dataset is about.

B. Task

Imagine you are a foreign professional in Uganda in the field of road safety. You have been hired by the local government to carry out some research on the roads in Kampala. The purpose for the research is to find out the roads with high accident rates and determine possible reasons for these accident rates to find countermeasures to reduce them. To do this, different data will be needed;

- Crash data which can be obtained from the police
- Road utilities like lighting on the roads, traffic signs and signals
- Road characteristics data, maximum speed, road type, road length, curvature, road surface

Assume that you have already obtained the crash data. Try to find the other data that you require for your task.

- Uncheck all the checked boxes
- In the search box on the right, type in the keyword "roads"
- Click "Advanced Search" in the right window
- Refine Search by "Metadata Language"
- Select "English"
- Type the keyword "transportation"
- In the search results, look for "Top10Vector- Road network" and open it.(Assume the dataset is for your area of interest)

Questions (Please think aloud as you write down the answers in the space provided)

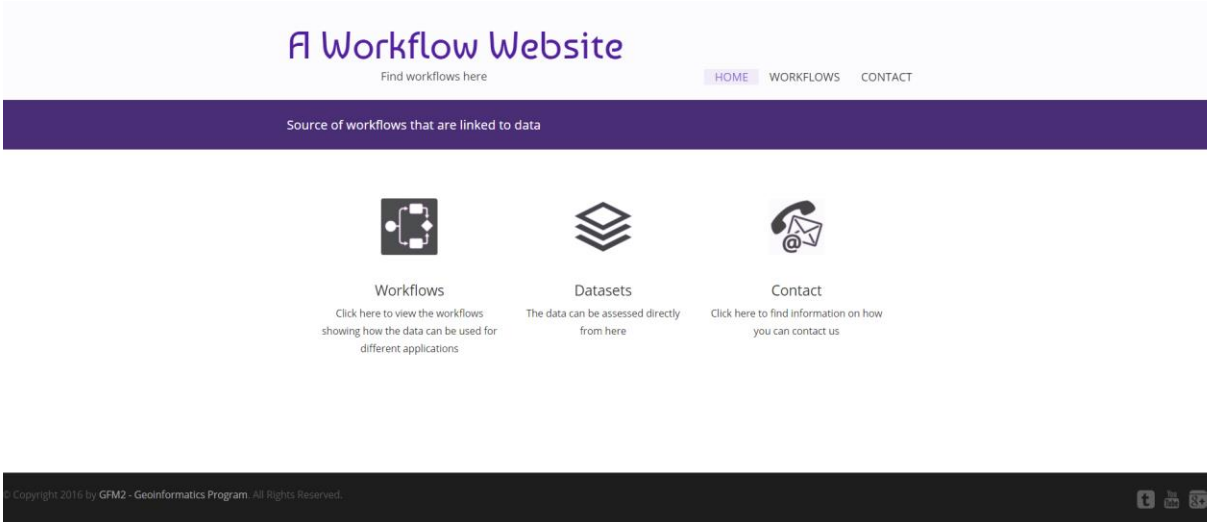
1. By looking at the information provided about the dataset, can you identify attributes within this dataset?
………
2. If yes, which attributes can you identify and how were you able to identify them?
………
3. Do you think this dataset could be useful to you for the given task? Justify your reason.
………
4. What areas of improvement can you suggest to help you understand the contents of the dataset better?
………

Part 2

Prototype

A. Explore the prototype (This should take about 7 mins)

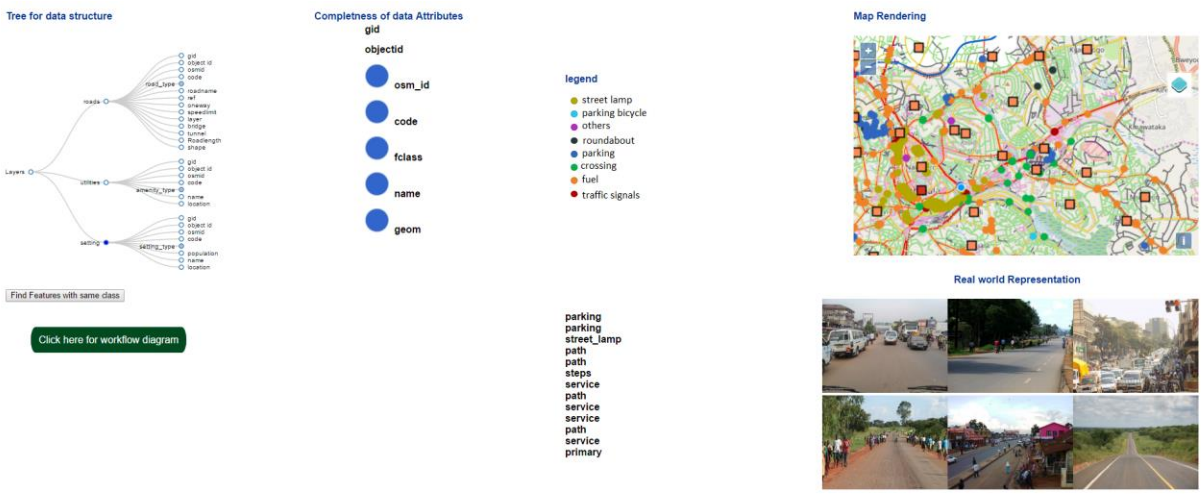
The prototype is a web application that shows previews of spatial datasets. The web page has three icons; Workflows, datasets and contact.



Go to the application (prototype). The web page has three icons; Workflows, datasets and contact

The workflows show how data in the application can be used to accomplish different purposes. In this prototype, only two workflow diagrams are presented.

- Click on the workflow icon, it will redirect to a page with the workflows
- Select the workflows and have a look at them. The boxes with red text are the data that can be obtained with the prototype and clicking inside them will redirect you to the data page.
- Alternatively, you can go directly to the data page by selecting the "Datasets" icon.
- In the data page, there are four windows, the data structure, the completeness of the attributes, map rendering and the real-world representation and a button for the workflows as shown below;



Data structure
This shows the data structure within the application. It shows the attributes within the datasets. Click on the nodes to observe how they behave and interact with the window entitled "completeness of the data attributes". Hovering over the nodes shows the data type of the node if applicable.

Completeness of the data attributes
Blue colour shows attributes that are complete
Red colour shows attributes that are incomplete missing values

Map rendering

This icon  in the map window shows the data layers available in this application. Checking the layer renders the features on the map.
Selecting a feature that is displayed in the map shows the real-world representation of the feature and the completeness of the data attributes (Zoom in to select only one feature otherwise many will be displayed at the same time).

Real-world representation
Click on the feature that is displayed in the map first to show a real-world representation of it.

The "Click here for workflow" button shows the workflows that available in the application for which the datasets can be used.

B. Task
(This task is the same as the previous task)
Imagine again you are a foreign professional in Uganda in the field of road safety. You have been hired by the local government to carry out some research on the roads in Kampala. The purpose for the research is to find out the roads with high accident rates and determine possible reasons for these accident rates to find countermeasures to reduce them. To do this, different data will be needed;

- Crash data which can be obtained from the police
- Road utilities like lighting on the roads, traffic signs and signals
- Road characteristics data, maximum speed, road type, road length, curvature, road surface

Assume again that you have already obtained the crash data. Try to find the other data that you require for your task.

- ➢ In the workflow page, choose the appropriate workflow for your task.
- ➢ From the workflow click on the box with red text and you will be directed to the data page

Questions (Please think aloud as you write down the answers in the space provided)

5. Does the workflow provide any useful information for you? Justify your answer.
…………………………………………………………………………………………………
6. Can you find any data that can be useful in this scenario?
……………………………………………………………………………………….
7. If yes, which data can you obtain with this application and how were you able to identify it?

………
………
…………………………………………………………………………………….

8. Judging from the information provided in the different previews about the data layers, do you think the data is appropriate for use in this scenario? Justify your answer. (Indicate which is appropriate for use and which is not)
………
………
………
………
………….

9. Looking at the track road and the secondary roads, which do you think is more likely to have more accidents on it? Why?
………
………
…………………………………………………………………………………….

10. Do you think all the previews provide enough information that help in understanding the data? Explain.
………
………………………………………
………
………
…………………………………………………………………………………..

11. Current the previews provided are static images of the real-world representation, do you think that dynamic data would be more useful? Explain.
………
………
………
…………………………………………………

12. What suggestions would you give to improve the prototype?
………
………
………
………
…………

13. What did you think of the test in general? (You can just speak your thoughts)