Geographical Information Management and Applications

A method for creating web GIS assessments from the user perspective

Master's thesis

Danny van Hienen 4055675

d.vanhienen@students.uu.nl

Supervisor: prof. dr. S.C.M. Geertman

Responsible professor: prof. dr. D.J. Karssenberg

Universiteit Utrecht

TU Delft

University of Twente.

Wageningen UR

A method for creating web GIS assessments from the user perspective

Master's thesis

Danny van Hienen

Summary

Web GIS assessments differ, because of the varying product characteristics, user types and user goals. Assessment developers have to invent the wheel repeatedly, because a standardized method is lacking for web GIS assessments from the user perspective. The following research question is formulated to develop a method that guides assessment developers through the process: *what is an appropriate standardized method for developing web GIS assessments from the perspective of the user?*

Within the field of software quality assessments, a standardized method involves an approach to define quality in small parts of the product and use their definitions to create an assessment. Hence, assessment developers define what needs to be evaluated from a pool of product characteristics. Developers can translate the selected web GIS characteristics into statements, which users can rate on the level to which they agree. The resulting statement ratings can be analysed and compared.

The assessment method consists of six steps and is developed for two case studies, Floodlabel (n = 109) and GeoWeb (n = 37). The first is product of a research project, namely a prototype web site for informing Dutch homeowners about flood risks and mitigation. The second is an application of Sweco, which is used for advanced GIS-implementations. For both case studies, a questionnaire is created and analysed by following an assessment method. That method includes a step-by-step approach to define the research set up, define the questionnaire content, create an online questionnaire, and generate and interpret the results. All steps are based on the literature background section, which discusses what is understood as a web GIS, a standardized assessment and how the user perspective can be included.

This study's conclusion is that the method is appropriate for creating web GIS assessments from the user perspective. Using the rating scale set up ensures a consistent questionnaire, which can be easily understood by respondents. With the four sub steps to write out the questionnaire statements relevant web GIS characteristics can be assessed. The final ratings can be generated easily and used for comparing user groups and product characteristics. However, it must be acknowledged, that a successful application depends on the developer's ability to select the most appropriate web GIS characteristics, write the statements out clearly and understandable, provide sufficient user guidance for participating to the research, avoid GIS jargon and provide definitions.

After all, this study has taken a first step for developing a standardized web GIS assessment method. Yet, the method should be examined by more researchers and applied for non-scientific purposes as well. That would lead to more insights in what an appropriate method is, how it can be applied, and which challenges occur. The method is ready for application on a much wider scale and, hence, developing to a standardized guidance for anyone who wants to assess a web GIS.

Contents

List of figures

List of tables

1. Introduction

Within the rapidly developing world of Geographical Information Systems (GIS), products are changing fast and are more and more used by organizations. An important development is the upcoming presence of web GIS. In contrary to installing all GIS components on every computer, web GIS products *“take advantage of web technology for communicating among its components”* (Veenendaal, Brovelli & Li, 2017, p. 3). This leads to a shift from traditional desktop GIS to web GIS applications, with a broader range of users that can be reached. In this way, organizations identify opportunities to work more efficient, effective and satisfied with the support of a web GIS product.

Essentially, the purpose of a web GIS is to fulfil the user needs. Consider a municipality employee responsible for traffic sign maintenance. Once a week the user visits a web GIS that visualizes whether traffic sign maintenance action is required. Since the user only needs to know whether action is required, other information or functionalities are unnecessary. That means a functionality of generating buffers around objects is rather redundant than useful. In this way, users experience a product on the degree to which their user needs are supported.

Experiencing a web GIS involves how the user interacts with the system, which can be defined as the human-computer interaction. The user wants to reach certain goals and web GIS products should be configured in a way the user is able to reach these goals. An example of human-computer interaction is a mouse click on the screen triggering a pop-up on the screen. The user retrieves the information and thus, experiences using the product. A simple illustration of this process is visualised in figure 1.1.

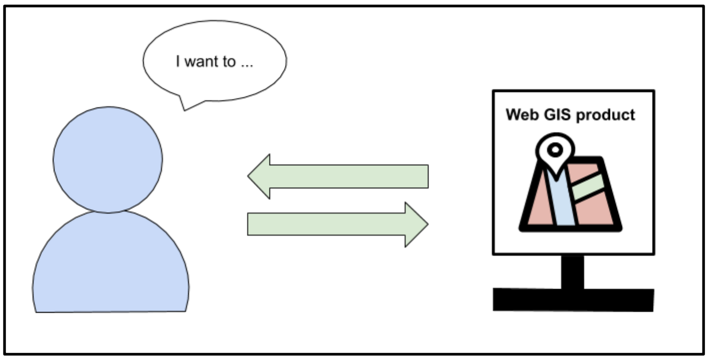

Figure 1.1 Human-computer interaction of a web GIS

Shaping web GIS products to the user needs is a common challenge for product owners, managers or anyone who provides a web GIS solution to one or more users. Hence, user needs are commonly involved in early development stages to avoid situations were the product needs to be revised after completion (Kujala, 2008; Sluter, van Elzakker & Ivánová, 2017). However, this does not certainly mean the final product meets the user needs, because they differ over time, might be interpreted incorrectly or are unreachable for the product. For that reason, product owners need to assess the product from the user perspective to be aware of its quality continuously. In this way, the user can share an opinion on the degree to which the product has added value (Lewis, 2014; Wagner, 2013).

1.1 Problem statement and research objectives

Quickly evaluating the product from the user perspective can be performed with an assessment. Since web GIS products have different product characteristics, user types and user goals, the assessments differ as well. Hence, assessment developers have to develop their own assessment set up, content, and approach of analysis on their own. Within the conceptual model figure 1.3, this problem is visualised. The figure shows that assessment developers must create different assessments for different products with varying users. How they can create this is not standardized, so they have to invent the wheel repeatedly.

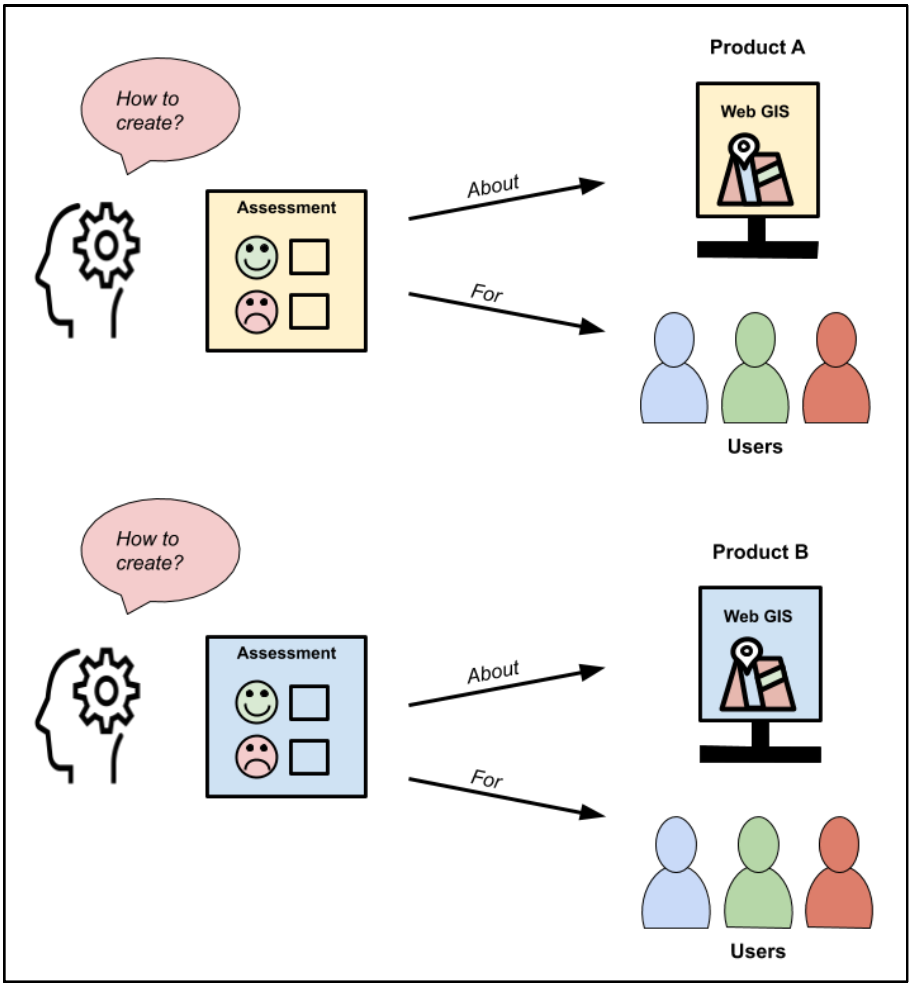

Figure 1.2 Visualisation of the problem statement

Therefore, a standardized method should be developed to guide assessments developers in the process of creating assessment. However, there currently is no standardized method for developing web GIS assessments from the user perspective. So, this study's research objective is to examine what an appropriate method is. Hence, the following research question is formulated:

What is an appropriate standardized method for developing web GIS assessments from the perspective of the user?

The main goal of the method is to provide support by developing web GIS assessments. Reaching this study objective requires a theoretical foundation where the method is based on. Hence, three sub research questions lay the foundation for designing the method.

The first step in the process of developing a web GIS assessment is to define what needs to be assessed. Therefore, developers need to be aware of what is understood as a web GIS application and which potentially relevant product characteristics exist. Having that information ensures that they know what needs to be assessed. For that purpose, the following sub research question is formulated:

What is generally understood as a web GIS application? (SQ 1)

Since a standardized method for developing web GIS assessments is lacking, it is necessary to formulate what is generally understood as a standardize method. Based on literature from general software studies, the best-practices can be used to develop a method for web GIS. For that reason, the second sub research question is formulated as follows:

What is generally understood as a standardized assessment method? (SQ 2)

The assessments must evaluate the products from the user perspective, which requires special attention, because users can experience the same product differently. In other words, assessments must acquire the user's background and subjective measurements. How that can be performed is examined by formulating the third sub research question:

How can web GIS applications be assessed from the user perspective? (SQ 3)

The three sub research questions are discussed in the theoretical background section and lay the foundation for developing the method that guides assessment developers.

1.2 Research limitations, scope and case studies

As mentioned earlier, web GIS has evolved rapidly the last decades, which has led to implementation in the businesses of companies, governments, educational institutions and other organizations. As a result, web GIS users vary from experienced to inexperienced, low to high educated and young to old users and thus, users experience a product differently. For that reason, this study's method is applied to case studies with differing user types, being Floodlabel and GeoWeb.

The first concerns the prototype web site *floodlabel.net*, which is designed within the bigger FLOODLABEL research project. One of the goals of this research project is to explore how Dutch homeowners can be informed about flood risks and activated to flood risk mitigation. For that purpose, the prototype web site assigns *Floodlabels* to each premise in the Netherlands to show homeowners risks for different flooding types. These labels are generated by using geographical data of rainfall, river floods and emerging groundwater (Urban Europe, 2017).

The second case study regards to the software product GeoWeb, which is a collaborative product of Sweco and Esri Netherlands. For each customer organization, developers and administrators create web GIS solutions using GeoWeb, resulting in diverse solutions with varying geographical information, functionalities, designs and so forth. Therefore, Sweco supplies the product to organizations differing from municipalities to grid companies, with different types of end users, varying from inexperienced to experienced GIS users (Sweco Nederland B.V., 2020).

For both case studies an assessment is developed and disseminated among the case study respondents, resulting in quality perceptions on the products from the user perspective. These quality perceptions are transformed into ratings and scores to make them interpretable. Additionally, the assessment itself is assessed on its reliability and validity. Using these results, the appropriateness of the method can be discussed to answer the main research question (Scheepers, Tobi & Boeije, 2016). An overview of the research steps is visualised in figure 1.3.

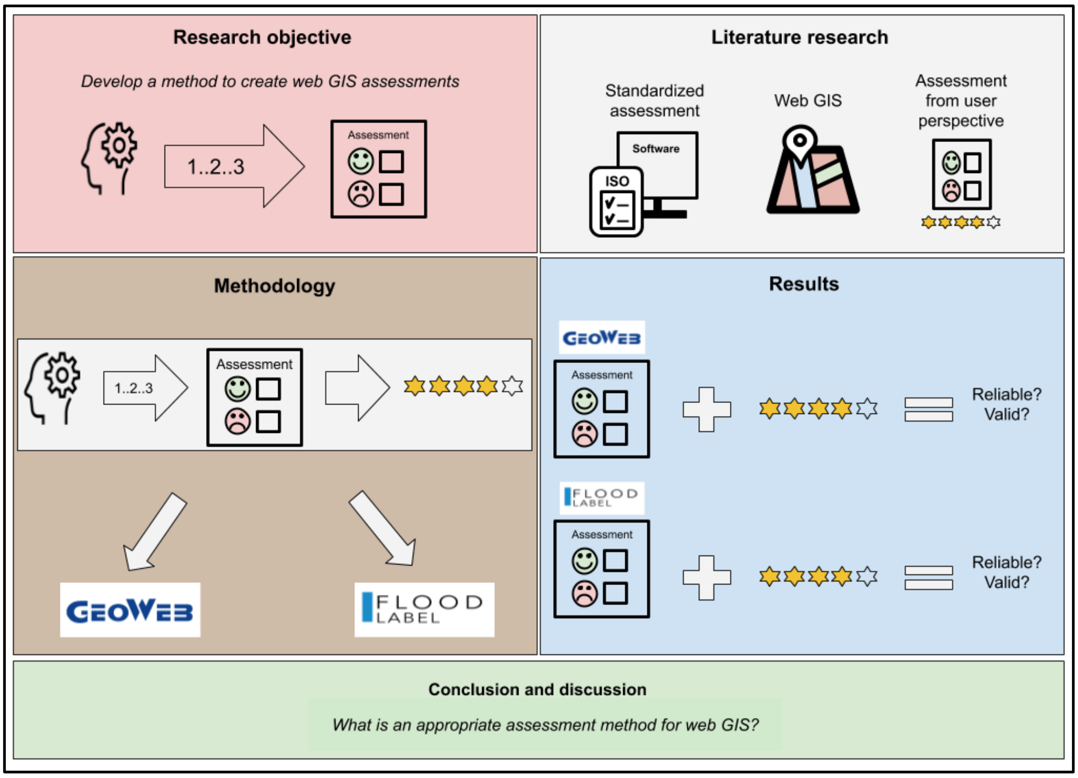

Figure 1.3 Overview of the research steps

Within the theoretical background section is discussed what should be included in this method, by elaborating on standardized assessment methods, what is understood as a web GIS and how the user perspective can be included (section 2). That information is used to develop the method, apply it on the Floodlabel and GeoWeb case studies and evaluate its appropriateness. How that is performed is discussed in the methodology (section 3). The study results are three-fold, namely applying the method to create questionnaires, generate and interpret assessment results, and discuss the method’s reliability and validity (section 4). The findings provide input for answering the research questions in the conclusion (section 5). Lastly, a discussion section defines the added value of this study and suggestions for further research (section 6).

1.3 Relevance

The research gap to fill for this study involves the lacking standardized method for developing a web GIS assessment. Such a method should be applicable to web GIS with widely varying types of purposes, differing software products and diverse user types. The assessments are targeting the process of figure 1.1, hence, this study combines scientific literature to cover definitions of the human-computer interaction between web GIS and its users. Based on that literature background, this study develops and tests a step-by-step process as a guidance of developing an assessment (Sibisi & van Waveren, 2007; Yan et al., 2019). In this way, the research gap is filled of a method for developing a web GIS assessment from the user perspective.

Besides that, this study's societal relevance regards the benefits of using the method in practice. Ideally, the method suits for developing an assessment for all web GIS products. Thus, product owners within commercial, governmental or educational organizations can make use of the standardized assessment method. The benefits of the method are the guidance as a step-by-step process and reduced time to assess a web GIS. Following the method facilitates developers to easily create an appropriate assessment, generate results and interpret the scores. In this way, developers quickly acquire the user perspective on a web GIS.

2. Theoretical Background

For developing a web GIS assessment method from the user perspective, this chapter provides the required background information. The first section 2.1 introduces the field of web GIS products and discusses relevant characteristics to consider in an assessment. After that, section 2.2 elaborates on standardized assessment methods for software products and how these can be specified to web GIS. Lastly, section 2.3 elaborates on how the user perspective can be included in an assessment.

2.1 Web GIS products

Within figure 1.1 is visualized that a user interacts with a web GIS product. In a web GIS assessment, the user shares experiences regarding this human-computer interaction. To be able to assess a web GIS, the user needs to evaluate different product characteristics. Hence, this section introduces the concept of web GIS and discusses characteristics relevant for assessments.

Defining web GIS starts with defining the bigger concept of GIS. Although varying GIS-definitions have been provided the last decades, most of them involve a description of how computer-based systems handle with georeferenced data. As an illustration, Huisman and de By (2009, p. 32) emphasize that *georeferenced data* is key to all involved handlings and define GIS as *"a computer-based system that provides the following four sets of capabilities to handle georeferenced data: data capture and preparation (1), data management, including storage and maintenance (2), data manipulation and analysis (3), and data presentation (4)."*

Recently, web technology developments changed the field of GIS exceptionally. Nowadays, traditional offline software applications are being replaced by online cloud-based solutions, which more and more support the four sets of capabilities to handle georeferenced data identified by Huisman and de By (2009) in connected and cooperative environments (Heywood, Cornelius & Carver, 2006; Tolpekin & Stein, 2013).

Because the field of web GIS is relatively new and integrated in differing fields of science and practice, numerous associable concepts are being used. Nevertheless, the following definition covers most of these associable concepts and defines web GIS as: *"a fully-fledged GIS that takes advantage of web technology for communicating among its components (data, functionality, and interface)*" (Veenendaal et al., 2017). So, a web GIS should have the same capabilities as a traditional GIS, but components can be stored at different physical locations, where communication between components is supported by web technology.

As an illustration, figure 2.1 visualizes the communication between web GIS components facilitated by web technology. Within this example, webservices function as communicator between geo-resources and web clients. The communication between client and server requires input, which can be communicated in different formats and languages such as the eXtensive Markup Language (XML). In short, XML supports transferring data in an interoperable format. The web viewer in an internet browser (PDOK Viewer) reads XML files published by a webservice, with data from geo-resources (National parks layer) (Agrawal & Gupta, 2017;

Castronova, Goodall & Elag, 2013; Veenendaal et al., 2017). Elaborating in more detail about web GIS architecture is outside this study's scope but section 8.5 of Tolpekin and Stein (2013) provides a clear and detailed description.

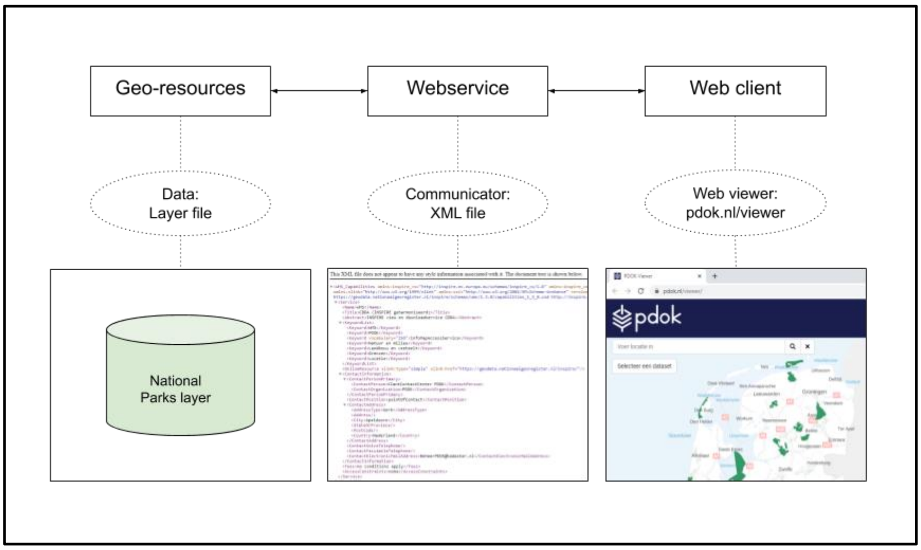

Figure 2.1 Basic representation of a Web GIS architecture. Source: Author

2.1.1 Web GIS product characteristics

This section narratively describes web GIS product characteristics. Several characteristics are discussed from the user perspective. Within table 2.1, a non-exhaustive list covers relevant product characteristics from the user perspective.

Table 2.1 Overview of web GIS product characteristics. Source: author

Product characteristic	***Description***	***Examples***
Basic (GIS) functionalities	Functionalities related to using the map	Zooming, panning, scrolling or turning layers on and off
Advanced (GIS) functionalities	Functionalities handling georeferenced data	Overlay functions or network analysis
Georeferenced data	Data with a location component	Added value or up-to-datedness of data
Import and export of georeferenced data	Any method for adding data or extracting data to or from the application	Upload or extract a shapefile
Georeferenced data descriptions	Any way of providing descriptive information about the georeferenced data	Titles, legends or references to external metadata
Visualization of georeferenced data	Any representation of georeferenced data	Colours, symbols and classification of data
Map interface	The component of a web application that includes a map	Web viewer with integrated functionalities or a map as relatively small component
User guidance	Anything that supports using the application	Documentation guides or support functions

Basic and advanced GIS functionalities

Starting with a description of functionalities, which facilitate activities in a product. In the beginning of web GIS applications, interactive maps enabled the user to communicate with map representations, but users were limited to browsing, zooming, panning, and turning layers on and off (Unrau & Kray, 2019). This study assigns these functionalities as basic functionalities. Nowadays, more advanced traditional GIS functionalities are included in web GIS products as well. Examples are overlay analysis, measurement techniques, attribute queries, neighbourhood functions, proximity analysis and network analysis. Therefore, advanced functionalities are defined as functionalities handling georeferenced data (Heywood et al., 2006; Poplin, 2015; Tolpekin & Stein, 2013; Unrau & Kray, 2019; Veenendaal et al., 2017).

Examples of web GIS products

The first example is a web site showing the weather for the United Kingdom (UK) per three hours. Assumingly, the product goal is to inform users about the weather at a specific location on a specific moment in time. Therefore, this web site includes a web map with weather descriptions for eight moments a day, which can be visualised by using a timeline function. Besides that, the user is able to view a location within the UK on various levels of scale, either with a search function or with the zooming and panning functions. When the user has managed to find the location of interest, a pop-up function provides the location with the weather type for the selected moment in time. The weather attribute values are visualised in the map by either a sun or a cloud and a sun (figure 2.2), which should be easily recognized and hence support communicating the information to the user.

Reflecting on this web GIS product it includes the following characteristics influencing the user: zooming, panning, searching for data, a timeline, pop ups, visualizations and data descriptions. These characteristics support achieving the user goal of being informed about the weather. For that reason, they can be included in an assessment from the user perspective, where the user shares experiences of how the product characteristics facilitate reaching the user goals.

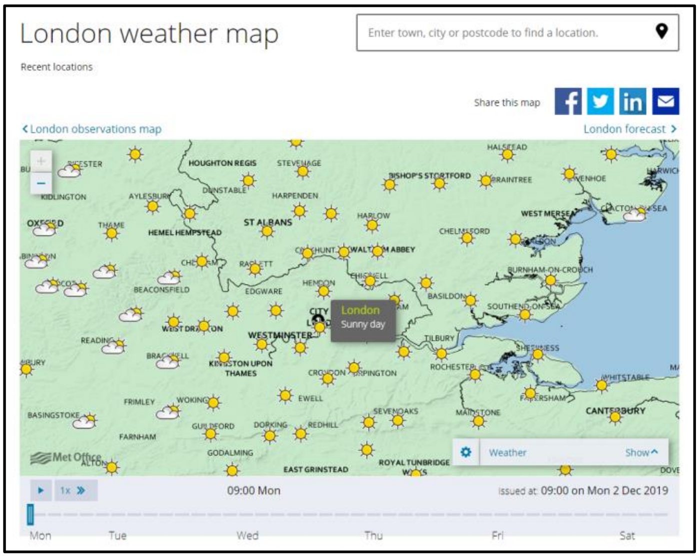

Figure 2.2 Screen picture of the UK weather map for the city of London. Source: Met Office (2019)

The second example is about a web GIS application of a Belgium company called Siggis. As an introduction the following product background information for *De Watergroep* is provided on their web site:

> *"the goal is to expand the use of geographical information, both within and outside the office environment. The assignment involves implementing an online GIS viewer for the drinking water network with editing, spatial analysis, tracing, printing and exporting functionalities, including an intuitive management application featuring authentication and a mobile offline viewer."* - (Siggis, 2019)

Multiple web GIS product characteristics are mentioned in this product description. For example, their targeted users are within and outside the office environment, which indicates that applications must be used on different types of devices, such as desktop computers, tablets or mobile phones. Using web GIS on different types of devices, requires varying interfaces and might lead to different experiences. Additionally, users must be able to perform advanced functionalities such as editing, spatial analysis, tracing, printing and exporting data. Commonly, in a web viewer these functionalities can be executed by clicking on the right buttons, located in a toolbar. Hence, an application toolbar can consist of multiple tabs and underlying buttons, which can be seen in figure 2.3, where the viewer consists of seven tabs with underlying functionality buttons. Assumingly, inexperienced users might have trouble by finding the right tools for their analysis. So, applications can provide user guidance such as documentation guides, descriptive texts or communities for support.

To reflect, this example shows different product characteristics in comparison to the UK weather application, because of the more advanced user goals regarding handlings with georeferenced data.

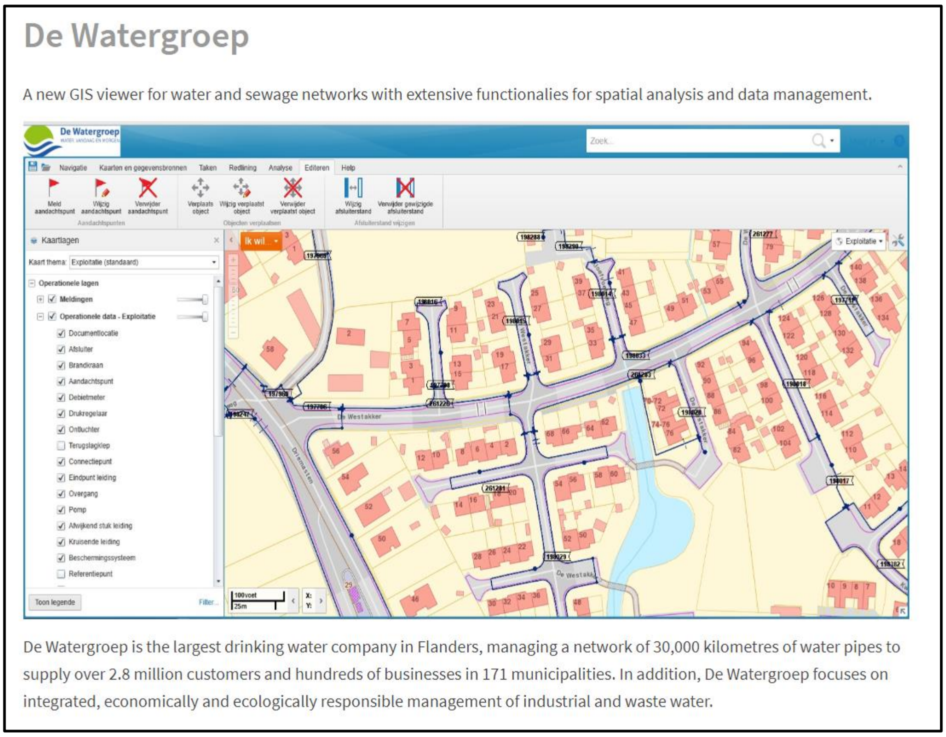

Figure 2.3 Screen picture of a Siggis (Belgium) project description involving a web GIS viewer. Source: Siggis (2019)

Georeferenced data

An important characteristic of georeferenced data is its quality. Assessing the data quality can be a time-consuming process, which examines the accuracy, precision, completeness and whether the data is up to date or not (Heywood et al., 2006). Next to that, georeferenced data can be assessed on how it is used. For example, by assessing which data is included, how this contributes to reaching the user goals and how the product explains the meaning of the data. This can be supported by the following product characteristics of the selection georeferenced data, metadata, a legend and data visualization.

Data security and privacy

Data security and privacy has become an important topic in computer sciences and public debate. Since web GIS products might disseminate privacy sensitive data, developers should deal with related questions properly. Currently, the European Union provides detailed regulations within the General Data Protection Regulation (GDPR), *"with the purpose of protection natural persons with regard to the processing of personal data and on the free movement of such data"* (Eur-Lex, 2016). For this reason, web GIS developers must secure that products do not violate these regulations, by for example, creating personal accounts with certain rights (view, edit or create) to use applications or data.

In summary, several product characteristics influence the user experiences, meaning that a user assesses a web GIS product on different characteristics. Hence, the non-exhaustive list within table 2.1 functions as starting point for defining web GIS products.

2.2 Standardized Assessment Method

This section elaborates on assessment methods for software products and how they can be used for web GIS. Developing an assessment requires defining the research set up and the assessment content. For this study, the assessments aim at how the user is able to reach the user goals by using the product. Hence, the user perspective, human-computer interaction and web GIS product must be included in an assessment (figure 2.4).

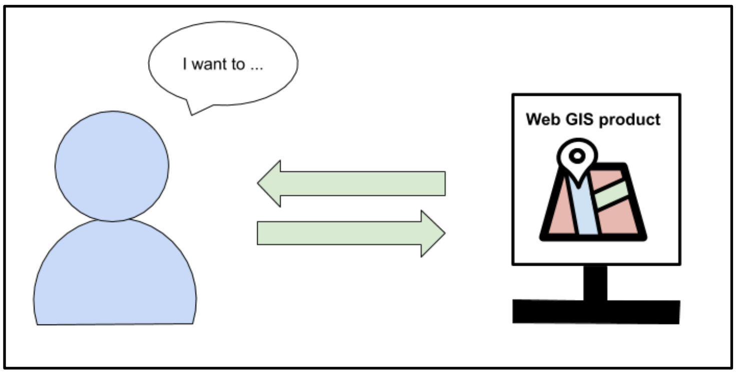

Figure 2.4 Human-computer interaction of a web GIS

Both the human-computer interaction (green arrows in figure 2.4) and the web GIS product can be assessed in various ways. Hence, this section describes two ISO quality models to cover the *quality in use* (human-computer interaction) and *product quality* (web GIS product).

Product Quality Model

Internationally accepted standards provide quality models for defining software quality as a set of definitions (ISO/IEC 25010, 2011). The International Organization of Standardization (ISO) designs, among other things, evaluation models for software products and software-intensive computer systems. Within *Systems and software Quality Requirements and Evaluation: SQuaRE* (2011), ISO provides quality models with well-known characteristics and sub characteristics for types of software quality. These models transform the initially vague concept of quality into smaller and more manageable parts, which can be measured by assessments (ISO/IEC 25010, 2011; Wagner, 2013).

The *product quality model* defines software products as a set of characteristics, with sub characteristics. All of these (sub) characteristics are assigned with definitions, which can be used to include in an assessment (definitions are included in appendices 8.1 and 8.2). Regularly, the complete list of quality (sub) characteristics is too long to include in an assessment. Hence, the quality models function as pools of characteristics from which assessment developers can select the most appropriate characteristics (P. Miguel, Mauricio & Rodríguez, 2014; Sibisi & van Waveren, 2007).

The *product quality model* categorizes software product quality into eight characteristics. From these characteristics the following five are relevant for web GIS assessments from the user perspective: functional suitability, performance efficiency, usability, reliability and security (ISO/IEC 25010, 2011). According to ISO/IEC 25010 (2011, p. 7), these product characteristics influence the human-computer interaction of primary users.

Both ISO product characteristics and the characteristics described in section 2.1.1 can be used to define web GIS products. Building on figure 2.4, these lists of product characteristics are attached to the web GIS product in figure 2.5. For each assessment individually, developers have to select the most relevant ones.

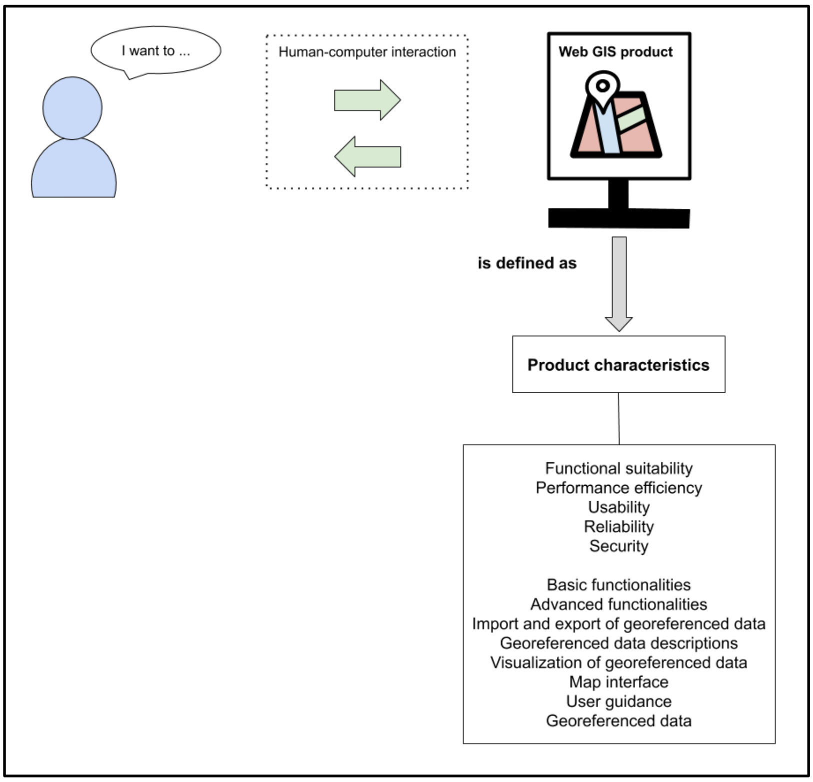

Figure 2.5 Web GIS product defined by characteristics

Quality in use model

Comparable to the product quality, the human-computer interaction can be covered by characteristics from the *quality in use model*. This model examines the *"degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, freedom from risk and satisfaction in specific contexts of use"* (ISO/IEC 25010, 2011, p. 8). To measure this, quality in use is composed of the following characteristics: effectiveness, efficiency, satisfaction, freedom from risk and context coverage. That means that users evaluate how (e.g. with effectiveness or efficiency) a product facilitates reaching the user goals. Building on figure 2.4, these five characteristics can be used to define the human-computer interaction (figure 2.6).

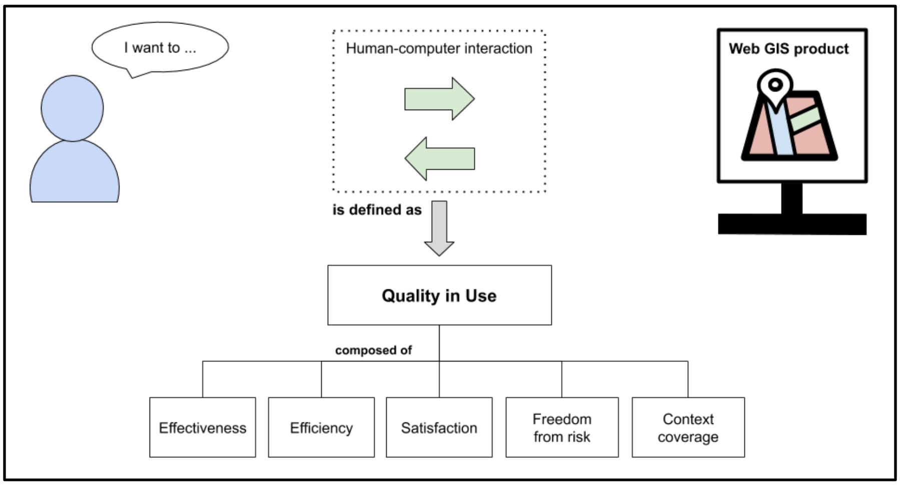

Figure 2.6 Human-computer interaction defined as quality in use

In sum, section 2.1 elaborates on *what* can be assessed, namely the characteristics of a web GIS product. Section 2.2 discusses *how* these products can be assessed by using quality definitions and the next section 2.3 is related to the users *who* assess a web GIS.

2.3 User perspective web GIS assessment

Nowadays, web GIS products have differing types of users with varying backgrounds and user goals. Hence, users experience a web GIS in their own way and perceive its quality differently. Quality is not a fixed and universal property of a web GIS (Wagner, 2013).

Because quality is perceived differently by web GIS users, the user characteristics must be identified for an assessment to put the results into a meaningful context. Section 2.3.1 elaborates on examples of these user characteristics and section 2.3.2 elaborates on considering them in a web GIS assessment.

2.3.1 User characteristics

This section identifies user characteristics to acquire in an assessment. For each assessment individually, developers must select the most relevant characteristics. To build further on figure 2.4, figure 2.7 lists several user characteristics to define a web GIS user. This section describes these characteristics.

Commonly, web GIS products are developed to create added value for the users. Each user has a certain goal, which can be reached by executing tasks on a product. Those tasks need to be facilitated by the product in an acceptable way. Hence, web GIS products are assessed on how they facilitate the user to reach specified goals (ISO/IEC 25010, 2011).

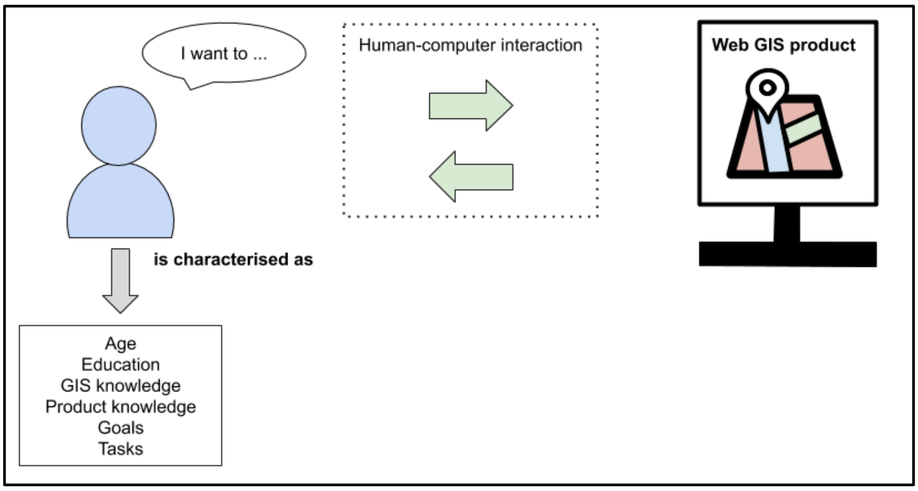

Figure 2.7 User characteristics in the human-computer interaction of web GIS products

Next to that, users can be characterised by different demographical and professional characteristics. These different characteristics as age, educational background, GIS knowledge, or product knowledge, influence the way a product is used and is experienced (Haklay, 2010). As an illustration, figure 2.8 visualizes two user quality perceptions on *Nationaal Georegister's* open source web viewer (PDOK, 2020).

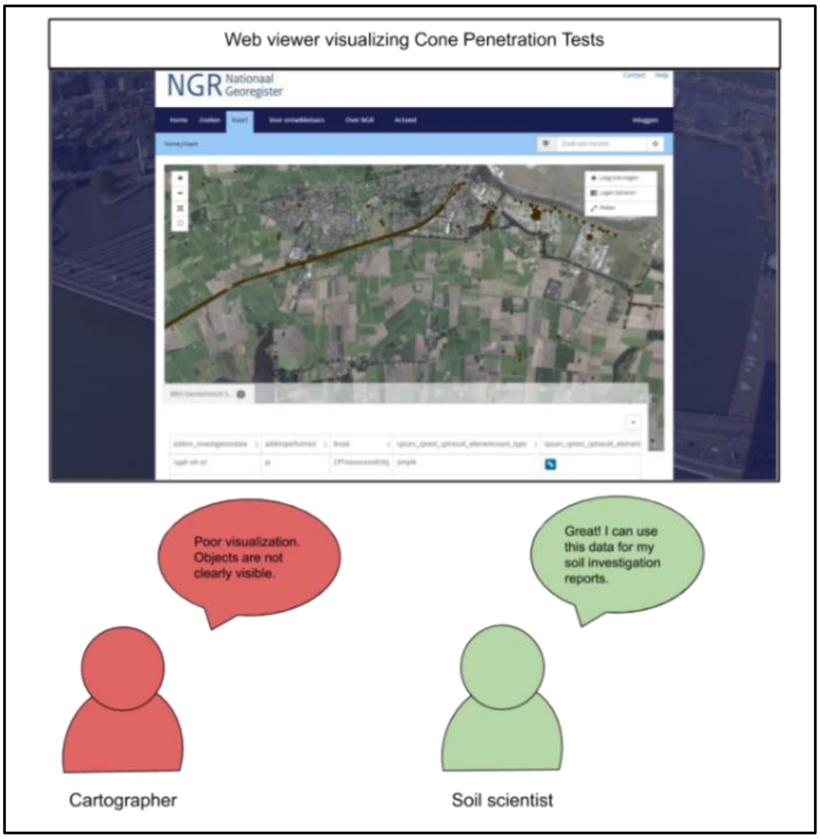

Figure 2.8 Fictive user experiences of Nationaal Georegister's web viewer. Source: author and (PDOK, 2020)

The web viewer visualizes Cone Penetration Tests (CPT) in the north-east of the Dutch province of Groningen. One of the users is a cartographer and the other one is a soil scientist. The latter is enthusiast about this web viewer, because the visualised CPT data can be used for further soil investigation. So, the viewer facilitates this soil scientist's user tasks to reach the goal of acquiring CPT data for further soil investigation. In contrary, the cartographer has a different perspective on the web viewer and disapproves the map visualization, because of the brown coloured icons and an aerial image as background map. The cartographer's goal is to improve map readability and data visualizations of open source web viewers. Hence, the cartographer experiences using the application negatively. Reflecting on these two user experiences, both users have different goals for the web viewer and hence focus on different web GIS product characteristics (georeferenced data and data visualisation). As a result, the users assess a singular web GIS product differently.

Furthermore, using advanced web GIS products requires GIS knowledge and skills. For that reason, web GIS developers consider their users and context of use for selecting appropriate functionalities. To illustrate that, figure 2.9 shows the conceptual relations between users, functionalities and data context. One interpretation is that laymen do not need high functionalities if all data context is known, the product's goal is then to present data. However, when data context is unknown, professional users should use high-end functionalities to explore and analyse the data (Haklay, 2010; Veenendaal et al., 2017). These relations show the influence of users, data context and functionalities on the product's type of use.

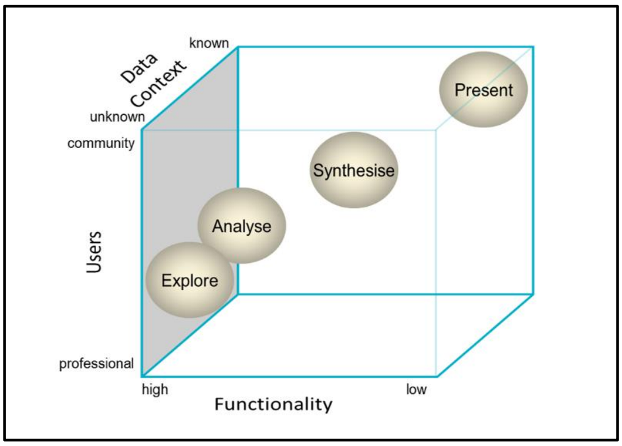

Figure 2.9 Component dimensions influencing web map use.
Adapted from Maceachren and Kraak (1997) by (Veenendaal, 2015)

2.3.2 Assessments methods

Frequently, questionnaire methods are used for a product assessment from the user perspective. Especially, within the usability field of human-computer interaction (HCI) (Hornbæk, 2006). These questionnaire methods can include open-ended questions or closed questions. Examples are the System Usability Scale (SUS) and Post-study System Usability Questionnaire (PSSUQ), which target experiences of a product's usability. These methods include intercorrelated rating scale statements, where respondents have to select a level to which they agree to the statements (Brooke, 1996; Lewis, 2002)(figure 2.10). An advantage of these methods is the possibility to compare the rankings among different users, user types or products (Haklay, 2010; Lewis, 2014).

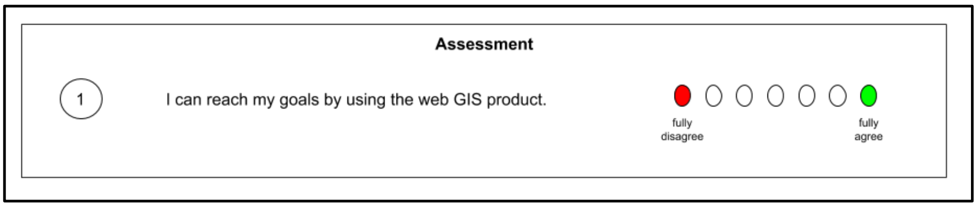

Figure 2.10 Rating scale statement assessment set up. Source: Author

2.4 Summary

After all, this chapter combines theoretical background knowledge for developing web GIS assessments from the user perspective. Such an assessment targets the process of the human-computer interaction between a web GIS product and its user (figure 1.1 and figure 2.4). Therefore, this chapter discusses how the user perspective, the interaction and the web GIS product can be included in an assessment. Overall, this results in a model visualised in figure 2.11, where the input of the sections is combined and related to each other.

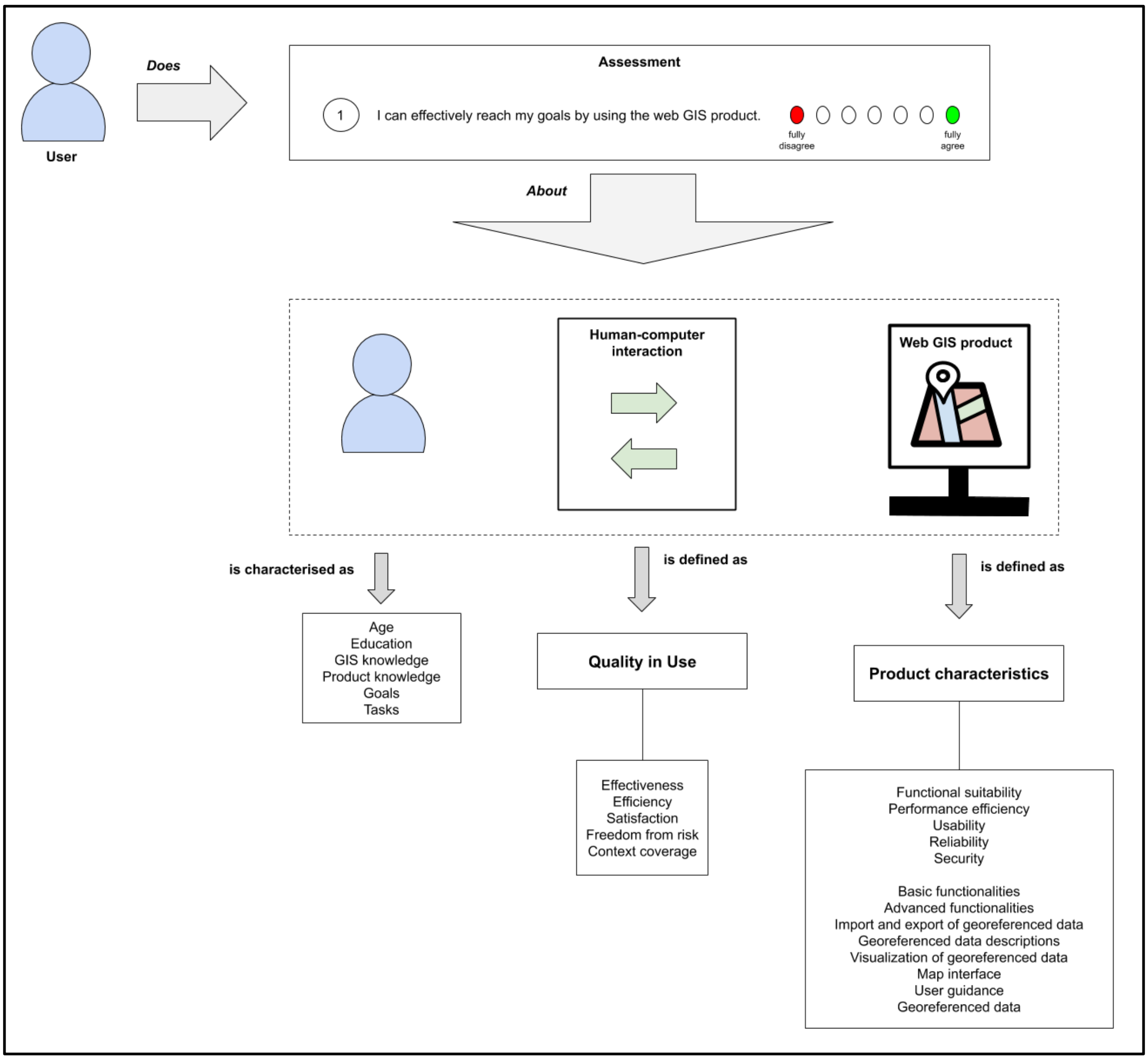

Figure 2.11 Overview web GIS assessment from the user perspective

Assessing a web GIS from the user perspective can be performed by developing a questionnaire with rating scale statements. The content of the statements can be based on figure 2.5, 2.6 and 2.7. Within section 2.1 Web GIS is discussed on which product characteristics a user can assess a web GIS and section 2.2 states that a web GIS can be defined by a list of product characteristics (figure 2.5). So, for each web GIS the most relevant product characteristics need to be selected for an assessment. Subsequently, that set of product characteristics need to be assessed on how they facilitate the human-computer interaction. That interaction can be defined by the quality in use characteristics (figure 2.6). Lastly, the user assesses web GIS on how it facilitates reaching the user goals, besides, the user perception is influenced by demographical and professional characteristics (figure 2.7).

3. Methodology

This chapter discusses how the method is created, applied on the case studies and how the results are analysed. Decisions and considerations are discussed for the case studies of Floodlabel and GeoWeb (section 3.1 Case studies) and how questionnaires are shared among respondents (section 3.2 Case study respondents). Within section 3.3 (Methods) is explained how the method is developed and applied to the case studies. Developing the method is based on figure 2.11, which explains how the human-computer interaction can be targeted in an assessment.

3.1 Case studies

Both case studies have different types of product developers, product owners, user groups and application purposes. So, varying insights are generated for the applicability of this study's assessment method.

The first case study involves the FLOODLABEL research project. Currently, flooding and other types of inundations are a threat for certain European urban areas. For that reason, governments need to perform flood protection measures. Additionally, homeowners can take measures as well, to ensure that urban areas are well-protected. Supporting and guiding that process is the primary goal of the FLOODLABEL research project (Urban Europe, 2017). Among other things, the research project developed a web GIS application, namely a combined Flood Information System (FIS) and Decision Support System (DSS). The prototype of the application can be accessed via the following url: https://floodlabel.net.

Commissioned by the University of Utrecht the application is developed by the company *Nelen & Schuurmans*, which creates information products related to water management (Nelen & Schuurmans, 2019). The prototype web site generates a so-called Floodlabel for Dutch premises based on georeferenced data sets. Users have to fill in their residential building address and a resulting web page visualizes the generated Floodlabel, a map with three georeferenced data sets and three buttons with additional information. Most relevant for this study is the integrated map showing precipitation, river flooding and ground water information. On this map, the user can switch between different layers, zoom in and out and navigate through the data and hence, is assigned as a basic application.

The second case study is about Sweco's web GIS software GeoWeb. Sweco is an internationally oriented architecture and engineering consultancy company, with their origins from Sweden. The company took over the former Dutch company Grontmij in 2015, including its Dutch head office in *De Bilt* (Sweco AB, 2020). Within the Transport and Mobility division, Team GIS and IT owns GeoWeb and is developed in collaboration with Esri Netherlands. Currently, GeoWeb solutions are implemented in more than hundred customer organizations, both governmental and commercial. All of these organizations have their own user needs, product goals and hence different ways of using GeoWeb.

Most of the organizations are using an HTML5 viewer as application for the end users. Within this viewer, GeoWeb administrators are able to include appropriate functionalities, geographical information, user guidance, extension tools and most of the traditional desktop GIS characteristics. Additionally, GeoWeb software includes modules for workflows, printing and reporting. That all enables Sweco to offer advanced web GIS solutions shaped to the user needs (Sweco Nederland B.V., 2020).

3.2 Case study respondents

This section elaborates on the selection process of respondents. Essentially, this is based on relevant user characteristics for a web GIS assessment (section 2.3). For both case studies the selection approach and final respondent group is described. Important to mention, the focus lies on a web GIS assessment research population, rather than the Floodlabel and GeoWeb populations. So, the selection process is based on an appropriate quantity and diversity of web GIS users.

3.2.1 Floodlabel respondent selection approach

Dutch homeowners are the targeted user group of the Floodlabel project. To enhance the chance that people will participate to the research, the research area is based on the relevance of communicating flood risks to homeowners. The Floodlabel related study of Snel et al. (2019) includes Dordrecht, Venlo and Zwolle in its research area.

Venlo was selected as research area based on the higher flood probability in adjacent river Meuse than the flood probabilities for adjacent rivers of Dordrecht and Zwolle. However, the field work in Venlo did not resulted in an appropriate number of respondents, so Zwolle was incorporated in the research area as well.

Defining the research population was based on the distribution of educational groups within neighbourhoods of Venlo and Zwolle. Assumingly, perspectives on flood risks and mitigation measures differ among the educational groups with the population. The Dutch Central Bureau of Statistics (CBS) uses a standard educational division (SOI in Dutch) and one classification includes three educational levels based on a hierarchical division of *high*, *middle* and *low* (CBS, 2016). Neighbourhoods were selected by aiming at acquiring thirty respondents per educational group.

As a result, 671 residential buildings were visited in Venlo's neighbourhoods Hogekamp, Hagerhof-Oost and Hagerhof-West, leading to twenty-nine participating respondents. Regarding Zwolle Stadshagen, 181 residential buildings were visited, leading to eleven participating respondents. Later on, the questionnaire was shared on social media channels, with the purpose of aiming to have thirty respondents per educational group and hundred respondents in total. The following section elaborates in more detail about the respondent characteristics.

3.2.2 Floodlabel respondent characteristics

This section describes the research population of the Floodlabel case, including 109 respondents with valid results. Each respondent was asked to share his or her level of education, age, user goal and used type of device.

Level of education
According to the division on education level, respondents of the lowest educational group are represented by nine respondents with valid results (8.3 percent of total). These people were the hardest to reach on the field and via social media. Despite the selection of neighbourhoods with relatively high amount of people in this lower educational group, they lacked motivation or did not succeed to complete the questionnaire. Respondents in the middle educational group are represented by twenty-two valid results (20.2 percent of total). Hence, the desired minimum of thirty respondents is not reached. The higher educational group represents the majority of the total research population by a number of seventy-eight respondents with valid results (71.6 percent of total).

Age
The mean age of the respondents is thirty-nine and the median thirty-two. The youngest respondent of the Floodlabel case is eighteen years old and the oldest seventy-three years. Notable, the group of respondents from twenty to thirty years old is represented the most in this research population (n = 49) and people from thirty to fifty years old are less represented (n = 20). Besides, for the purpose of analysing questionnaire results between groups of age an additional variable divides the respondents in three equal groups. These groups represent people below twenty-five years old (n = 33), people from twenty-five to fifty years old (n = 38) and people from fifty years or older (n = 38).

User goal
Respondents were asked about their goal for the floodlabel.net web site. Prior to assessing the web site, respondents had to visit and experience using the web site by filling in their residential address and explore all web site components. Subsequently, respondents were asked to think about why they would use the Floodlabel web site. That resulted in answers with meanings differing from *'no goal'*, *'flood risks of house'* to *'requested by student'*. To interpret these user goals, they are coded and assigned a label (section 8.4). Since, multiple labels are overlapping, another variable is created to divide the respondents in groups whether they were motivated to participate by either this research or the Floodlabel web site. As a result, the Floodlabel-motivated group consists of fifty-five respondents and the group interested in this research of fifty-four. Using this information enables comparing questionnaire results between and within groups of user goals.

Device type
The majority of respondents with valid results (n = 63, 57.8 percent) used a desktop computer or laptop to visit the Floodlabel web site. Less respondents used a mobile phone (n = 39, 35.8 percent), and a few respondents a tablet for visiting the web site (n = 7, 6.4 percent).

Invalid results
Due to research errors, respondents 14, 32, 64 and 66 are excluded from analysis. Both respondents 14 and 64 experienced problems with using the floodlabel.net web site. Respondent 14 mentions the occurrence of a blank screen while trying to reach the web site and respondent 64's address is not recognized by floodlabel.net. For these reasons the respondents rate all questionnaire statements with the most negative score. Although the error of not finding an address is relevant for the case study, other statements are negatively

influenced. Furthermore, respondents 63 and 66 are representing the answers of a singular person, because of the identical results and mail address. Hence, respondent 66 is excluded from analysis. Furthermore, respondent 32 misinterpreted the rating scales. Four statements are assigned with a rating of 2 and expanded with comments in the direction of *"everything is clear"* or *"everything works fine"*. These comments obviously not reflect the negative ratings. Besides that, the respondent did not comment on all individual statements, hence cannot be concluded whether the user consistently interpreted the rating scales in the opposite direction, so, respondent 32 is excluded from analysis.

3.2.3 GeoWeb respondent selection approach

Gathering as much as possible insights from respondents with different GIS knowledge and experience is performed in collaboration with a contact person at Sweco. With this support, GeoWeb administrators of customer organizations were contacted. These administrators were asked to distribute the online survey to their GeoWeb users. Enhancing reliability of statistical analysis, the minimum respondents was set to thirty, but the aim was reaching seventy-five respondents with appropriate groups of people in different GIS knowledge groups.

3.2.4 GeoWeb respondent characteristics

In total, a number of thirty-seven respondents participated to the GeoWeb questionnaire. This section elaborates on the respondent characteristics of GIS knowledge and skills, age, educational level, user goal, frequency of use, and type of device.

GIS knowledge and skills

Two questions were included in the questionnaire regarding the GIS knowledge and skills. One question in the introduction section about the respondent's current level of GIS knowledge and skills. Another question in the final section involved the respondents' perception of whether they have sufficient GIS knowledge and skills for using GeoWeb, to which they had to rate this perception on a scale from one to seven.

As a result, sixteen people have working and educational experience, twelve people only have working experience and nine people have basic skills. For the group with basic skills the mean rating of perceived GIS knowledge (5) is lower than both other groups (6), hence this slight difference can be considered in the analysis by creating a user type with basic GIS knowledge.

Age

Respondents at least twenty-seven and maximum sixty-three years old. On average, they are nearly forty-six years old and the quartile boundaries are respectively 37.5, 46 and 56. Additionally, two classes are defined with a class boundary of forty-five. Eighteen respondents are included in the group younger than 45, nineteen to the other group.

Education

Most of the respondents are assigned to highest education level group, namely 78.4 percent. The other eight respondents are assigned to the middle education level group.

User goal

GeoWeb users from eight organizations, with different user goals were selected. Hence, they were asked to share their goals for using GeoWeb. These organizations and user goals are visualised in appendix 8.5 and are divided into two groups. The criteria for classification is based on the whether the respondent assumingly manipulates georeferenced data or not (table 2.1). Viewing or acquiring information in a GeoWeb environment is assigned as basic use. Other user goals as managing GeoWeb environments, performing analysis or producing output are assigned as advanced use.

The basic group consists of twenty-three respondents and the advanced group of fourteen respondents. It should be noted that these classes have arbitrary boundaries and might be incorrect in one or more individual cases. However, to provide support for this classification, relatively more respondents assigned with basic user goals have basic GIS experience and knowledge (35 percent) than respondents with advanced user goals (7 percent). Additionally, this is supported by a significant difference in mean ranks of GIS experience and knowledge between the two groups (Mann-Whitney U test: $p < 0.05$).

Frequency of use
None of the respondents uses GeoWeb on a yearly basis. Out of the total thirty-seven respondents, twenty people use GeoWeb on a daily basis (54 percent), thirteen on a weekly basis (35 percent) and four on a monthly basis (11 percent).

Device type
All respondents except one are using GeoWeb on a desktop computer or laptop. That one respondent with a divergent answer identifies using each device type of a mobile phone, tablet and desktop computer or laptop.

Invalid results
Lastly, none of the respondents identified problems while participating to the questionnaire, so all respondents are included in the analysis.

3.3 The method

This section aims at developing the method for creating a web GIS assessment from the user perspective. The method consists of six steps, defining the research set up (step 1), defining the assessment content (step 2), creating the assessment (step 3), and generating and interpreting the results (step 4). Additionally, step 5 and 6 examine the reliability and validity of the applied method. How the study has come to design these steps is discussed in this section.

The method is based on figure 2.11, which visualizes how web GIS can be assessed from the user perspective. For an appropriate method a standardized research set up is needed, hence, section 3.3.1 elaborates on how rating scale statements can be used for an assessment (step 1). Subsequently, the content of the questionnaire statements must target the human-computer interaction, so, section 3.3.2 provides a four-step approach to start with a pool of web GIS product characteristics and end with a set of written out statements (step 2). Thereafter, the assessment can be created and disseminated among the respondents, which is discussed in section 3.3.3 (step 3). How the questionnaire results are generated and analysed is discussed in section 3.3.4 (step 4). To gain insights in the methods' reliability and validity, sections 3.3.5 and 3.3.6 elaborate on how respondent comments are used to evaluate the application of the method (step 5 and 6).

3.3.1 Assessment research set up

The four sub steps of figure 3.1 provide guidance by defining the research set up. According to the theoretical background section, rating scale statements are a sufficient research set up for web GIS assessments from the user perspective. To enhance repeatability and reliability of the questionnaires, they have to be designed in an online environment. Next to that, open-ended questions can be used to acquire information about the user and textual descriptions need to be added for guiding the user to fill in the questionnaire. Hence, this section elaborates on how these components of the research set up are applied to the case studies.

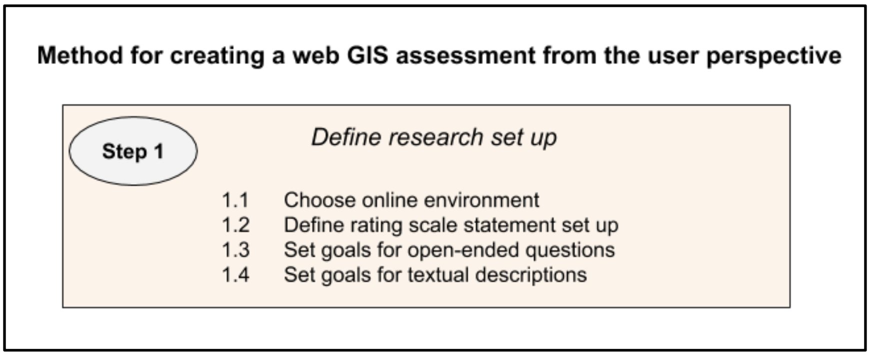

Figure 3.1 Step 1 of the method, defining the research set up

Google Forms

The online environment Google Forms can be used without licenses, does not limit the number of questions and respondents, and generates free output data in the interoperable Google Spreadsheet format. Furthermore, Google Forms supports the following question formats: *short answer*, *multiple choice*, *linear scale* and *checkbox grid*. Next to that, sections with textual descriptions can be added to a questionnaire. For the repeatability of the research, Google Forms enables copying complete online documents and objects within documents (Google, 2019). That all makes Google Forms suitable for this study's questionnaires.

Rating scale statements

Choosing for rating scale statements is based on two standardized methods in usability studies, the System Usability Scale (SUS) and Post-study System Usability Questionnaire (PSSUQ) (Brooke, 1996; Lewis, 2002). Both methods involve rating scale statements about usability characteristics, which means how people experience using a computer system. Besides, the methods are numerously used and analysed the last decades, which provided input for defining this study's research set up (Bangor, Kortum & Miller, 2008; Brooke, 2013; Sauro & Lewis, 2011, 2012). Questionnaire examples of the SUS and PSSUQ are included in appendix 8.3.

Rating scale statements support generating scores for a singular statement and groups of statements. These scores enable comparing scores between and within user groups (e.g. age classes) or product components (e.g. functionality). The mean statement scores indicate which statements are ranked higher or lower than others (Bangor et al., 2008; Finstad, 2010; Sauro & Lewis, 2012). An example of a rating scale is visualised in figure 3.2.

Figure 3.2 Screen dump of rating scale statement and explanation question in Google Forms. Source: author

Based on the SUS and PSSUQ literature, the research set up components of this study's assessment are defined. These components are briefly discussed in this section and an overview is provided in table 3.1.

Table 3.1 Overview research set up components

Component	***Implementation***
Number of statements	10 - 16
Wording of statements	All positive
Number of rating scales	7
Wording of rating scales	Fully disagree (1) – fully agree (7)
No answer option	Excluded
Final product score	Multiply 0 – 100
Interpretation of final product score	Adjective rating scale for overall product quality
Final scores for groups	For each product characteristic theme
Open-ended questions	Rating explanations for negative ratings and user tasks
Text	Providing definitions for GIS jargon
Text	Guidance for assessment

Number of statements

One of the most important characteristics is the questionnaire length. At one hand, developers want to acquire as much relevant information as possible, but at the other hand users want to spend at least as possible time on a questionnaire. So, developers need to find a balance between including as much as possible statements and diminishing the questionnaire length. With that in mind, both SUS and the 3rd version of PSSUQ are relatively short questionnaires, including respectively ten and sixteen statements dedicated to the usability of products (Brooke, 1996; Sauro & Lewis, 2012). Hence, this study aimed at including ten to sixteen statements.

Wording of statements

Statements can either have positive or negative wordings, meaning that they state something is good or bad. The SUS uses statements with both positive and negative wordings, which requires attention of the respondent because a high rating scale does not reflect a positive rating for all statements. According to Sauro and Lewis (2011), inconsistency in wording can better be avoided, because probabilities of measurement errors by researchers and respondents have to be diminished. Hence, this study uses positive worded statements only.

Number and wording of rating scales
Rating scales can have differing wordings and numbers of scales. Considering consistency, it is desirable to have a similar set up for all statements. The PSSUQ includes seven rating scales, because reliability of individual statements increases with more scales and tends to level off around the number of seven (Lewis, 2002). Furthermore, the SUS assigns higher rating scales with the positively worded *fully agree*, in contrary to the PSSUQ. Assuming that people tend to assign a higher rating scale more positive, this study uses seven rating scales where the first is represented by *fully disagree* and the seventh by *fully agree*.

No answer
Closed-ended questions force the respondents to select a predefined answer. Therefore, the PSSUQ includes the option of not choosing a rating (Lewis, 2002). At one hand, adding such an option provides the respondent to reflect a weak or no opinion about a statement in the results. At the other hand, the respondent is forced to have an opinion about the statement. The latter is not a problem if the assessment considers relevant statements only. Hence this study excludes a no answer option.

Relative importance product characteristic themes
Respondents were asked to fill in the relative importance of each product characteristic theme. The question's purpose is to compare the themes to each other and assign them with one of the following scales *not important*, *less important*, *neutral*, *important* or *very important*.

Open-ended questions
Next to the closed-ended rating scale statements, open-ended questions are included in the assessments as well. These questions provide added value in multiply ways, because they provide room for answers that might be unknown prior to data collection. For example, adding open-ended questions for an explanation of statement ratings provide insights in the respondent's reasoning behind the rating. After data collection, these reasonings are used to list the comments and examine how the respondents interpreted the statements. Furthermore, open-ended questions are used to let the user define their user tasks and goals, since these were unclear in advance of the data collection.

Textual descriptions
Additionally, textual descriptions in the questionnaire introduce the topic, web GIS product, research goals and how to do the assessment. Most importantly, emphasis lays on the fact that respondents had to assess the product from the perspective of reaching the user goals.

Besides that, the field of GIS is for many people undiscovered. Accordingly, GIS jargon can be misinterpreted or misunderstood very easily. Hence it is important to provide definitions of theoretical concepts related to GIS or computer systems in general. Based on estimating the GIS and computer system knowledge of the case study users, all GIS related definitions and advanced computer system definitions are explained within the Floodlabel assessment. These definitions are the product quality characteristics. For the GeoWeb case study only the difference between *basic* and *advanced functionalities* is explained, since it was expected that all users know what a web GIS product is. Additionally, small chances in wording improve the understandability of theoretical definitions, such as replacing *web GIS application* by a more understandable term as *product* (GeoWeb) and *web site* (Floodlabel).

Questionnaire validity

Lastly, examining questionnaire validity after data collection requires specific types of questions. These questions need to be configured in advance of the data collection and have the purpose of examining how respondents interpret questionnaire content in comparison to the intended meaning. So, for this study's assessments, a control question is included to examine the degree to which the group of *usability* statements cover its theoretical concept and another question related to the overall web GIS product regarding reaching their user goals.

3.3.2 Assessment content

The four sub steps of figure 3.3 guide developers at defining the assessment content. Within the theoretical background section is explained that the human-computer interaction and a web GIS product can be covered by characteristics (figure 2.11). Hence, the four sub steps select and combine the most relevant product characteristics and assign a quality in use characteristic to each combination. That enables to assess the user experience of all relevant product characteristics in a questionnaire. For each combination a statement is written out. Explanations of the four sub steps are provided in this section.

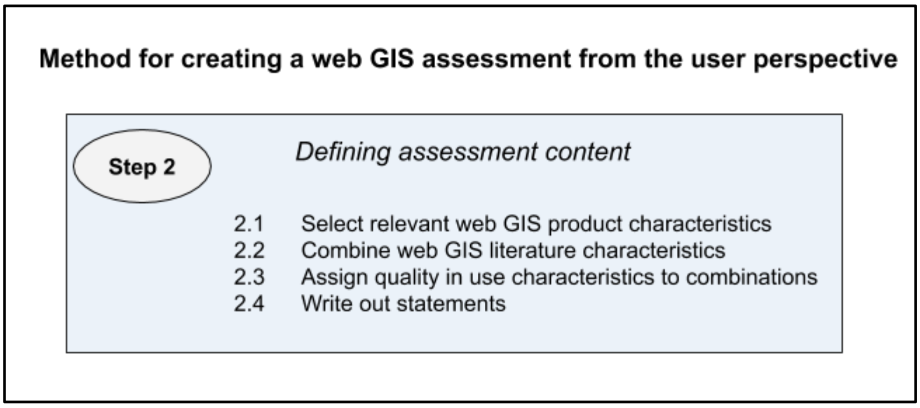

Figure 3.3 Step 2 of the method, defining assessment content

Step 2.1 – Relevance of web GIS product characteristics

The first sub step concerns defining which web GIS product characteristics are relevant. The pool of product characteristics of figure 2.11 can be used as a starting point. Since the product characteristics have a high abstract level, their sub characteristics are considered as well. An overview of these sub characteristics is provided in appendix 8.2. Figure 3.4 shows the starting pool that can be used for selecting the relevant product characteristics. Within this figure, six classes, or product characteristic themes, are created to structure the list of sub characteristics.

The method's step 2.1 concerns defining which product characteristics are relevant for the assessment. Hence, the key in this process is defining what needs to be evaluated by the users. So, for all sub characteristics of figure 3.4 can, iteratively, be determined whether the targeted web GIS product includes the product characteristic. As an example, can be examined whether the product includes advanced functionalities. If so, the product characteristic *advanced functionalities* can be included in the assessment, if not, that characteristic must be excluded.

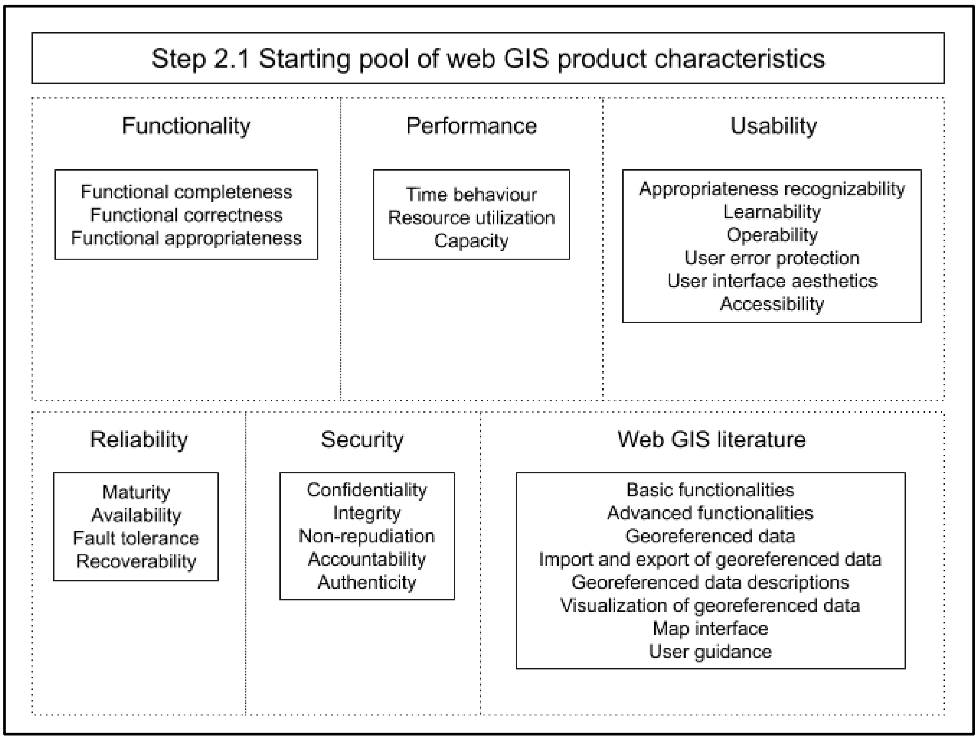

Figure 3.4 Step 2.1 Starting pool of web GIS product characteristics

Step 2.2 – Combine web GIS product characteristics

The next sub step aims at diminishing the number of selected characteristics. For that purpose, two questions can be asked. The first aims at removing redundant characteristics when looking inside every product theme, except the *web GIS literature* theme. The second aims at integrating the *web GIS literature* characteristics into the others. So, for each *web GIS literature* characteristic is defined whether it overlaps with other characteristics and can be combined or excluded from the assessment. After that, the remaining list of combinations can be assigned with a statement number.

Step 2.3 – Assign quality in use characteristic

Steps 2.1 and 2.2 focus on including the web GIS product characteristics. Step 2.3 includes how users interact with the product (figure 2.11). Hence, for each statement is defined which quality in use (sub) characteristic is most suitable. The complete list of (sub) characteristics is provided in appendices 8.1 and 8.2. A singular (sub) characteristic can be assigned to a combination, considering how the user might experience the product characteristic. That can be defined by considering how the product characteristic supports reaching the user goal.

Step 2.4 – Writing out statements

For writing out the statements it should be considered that GIS or IT jargon is avoided as much as possible, the statements are positively worded and are written out in a way that they can be understood by all respondents.

3.3.3 Creating assessment

After writing out all statements, the online assessments can be created by following step three of the method (figure 3.5). Within the assessment, the statements, open-ended questions and textual descriptions need to be included. Subsequently, it is necessary to test the assessment multiple times, by varying users. That enables to make improvements and decide when the assessment is ready to be disseminated.

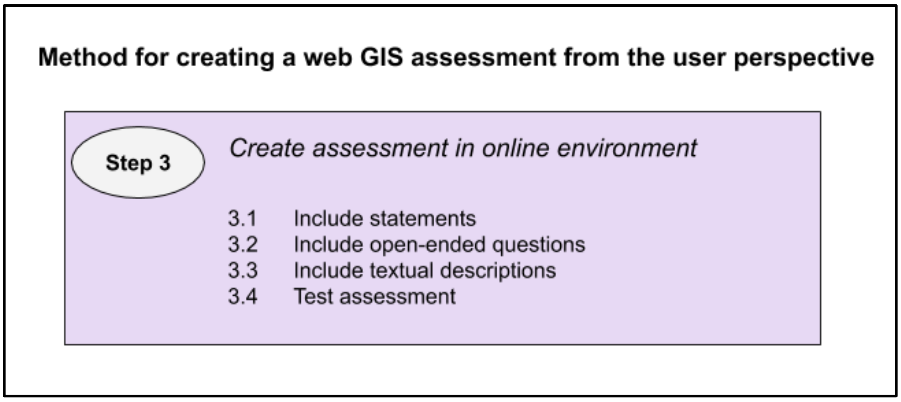

Figure 3.5 Step 3 of the method, creating online assessment

3.3.4 Case study scores

The fourth step of the method regards the generation and interpretation of the questionnaire results (figure 3.6). Respondents need to rate the statements from one to seven, reflecting the level to which they agree to the statement. Hence the resulting scores indicate how respondents assess the case studies' products. For both products separately, scores are generated for individual statements, product characteristic, the user groups and the overall products. Besides, with the support of the adjective rating scale and user comments the resulting scores can be interpreted comprehensively.

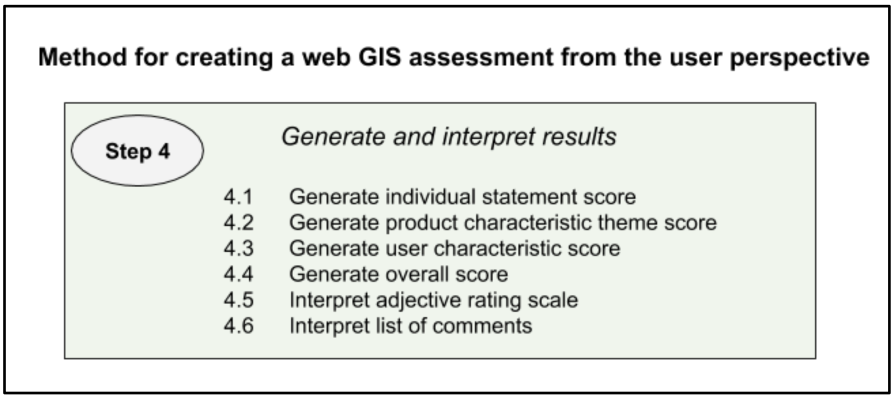

Figure 3.6 Step 4 of the method, generating and interpreting results

Statement score

For each statement the average score is generated to rank and compare the statements. Next to that, the rating scales of one, two and three are lower than the middle rating scale of four. Thus, counting the lower rating scales for each statement provide insights in the number of respondents rating a statement negatively.

Theme score
For all product characteristic themes (e.g. *functionalities* and *performance),* statements scores are generated as well. These scores can be used to rank and compare the themes, following a similar approach of the PSSUQ, which generate scores for *System Quality, Information Quality* and *Interface Quality* (Lewis, 2002).

User characteristic score
The overall rating, theme ratings and statement ratings of different user groups are analysed by different statistical tests. Selecting an appropriate statistical tests involves considering the measurement scale, number of population groups and meeting normality and homogeneity of variances assumptions (de Vocht, 2016). To illustrate how different statistical tests can be performed on the assessment results, the statistical tests listed in table 3.2 are executed.

Table 3.2 Statistical tests for user group analysis

User characteristic	***Case study***	***Measurement scale***	***Number of groups***	***Method***	***Score***
Organization	GeoWeb	Nominal	8	Scaled score	Overall
Organization	GeoWeb	Ordinal	5	Ranking	Themes
Adv/Bsc user goal	GeoWeb	Ordinal	2	Mann-Whitney U	Functionality
Education	Floodlabel	Ordinal	2	t-test	Overall
Device	Floodlabel	Nominal	2	t-test	Overall
Device	Floodlabel	Nominal	2	Histogram	Functionality
Device	Floodlabel	Nominal	2	t-test	Functionality

Overall score
For the purpose of interpreting the questionnaire results, the SUS multiplies the sum of scores to generate a score between zero to hundred. Noteworthy, SUS-developer Brooke emphasizes that this was rather a marketing strategy, than anything scientific. Since, product owners were more likely to understand rating scales from zero to hundred than ten to fifty. Hence, including a scaled summed score improves readability of the results, but must be attached with a remark to be careful with its interpretation, because, scores cannot be seen as percentages where 80 is two times better than 40 (Brooke, 2013).

To calculate the overall score, the average ratings for singular statements are generated and multiplied by the number of statements. That result is scaled up to a rating between zero and hundred, with the following formula where M is the mean rating per statement and n is the number of statements:

Equation 1 Overall score formula

$$Overall\ score\ =\ \big((M-1)*n\big)*\frac{100}{n\ *6}$$

3.3.5 Reliability

Web GIS assessments need to meet scientific requirements of questionnaire reliability and validity. Reliability concerns the research process and set up, i.e. how the questionnaire measures and is executed. Enhancing reliability must be performed before and after the research execution on how potential external factors influence the questionnaire content, meaning that results must be accurate, stable and not random. Therefore, the questionnaires are assessed on the way they generate similar results if the research is repeated and on the consistency of the respondents' answers within a singular assessment (ISO/IEC 25020, 2007; Sauro & Lewis, 2012; Scheepers et al., 2016).

Type of device

It is undesirable for the questionnaire reliability if different research set ups influence the process of filling in the questionnaire. Since the questionnaire is designed in an online Google Forms format, using different devices might influence the amount of time spent on the questionnaire by the respondents. Hence, respondents had to mention on which device type they participated and the minutes they spent on filling in the questionnaire.

Table 3.3 shows the descriptive statistics of the device groups for the *duration* variable values. For both *Desktop* and *Mobile* is n > 30, in contrary to the *Tablet* group. Four respondents are outliers of this variable, namely respondents 13, 22, 51 (20 minutes, twice mobile and once desktop) and 36 (35 minutes, desktop). Performing a Kruskal-Wallis H test shows that there is no indication of any significant effect from device on the duration of the questionnaire, $\chi^2(2) = 2.122$, $p = 0.346$. So, there is no indication of a violation to reliability.

Table 3.3 Descriptive statistics device type groups Floodlabel

Device type	***n***	***M***	***SD***	***Min***	***Max***
Desktop	59	9.29	5.014	2	35
Tablet	7	8.00	4.865	4	15
Mobile	36	9.33	3.538	5	20

Regarding the GeoWeb case, all but one respondent participated to the research on a desktop computer or laptop. So, there is no indication of any violation to reliability.

No answer option and suggestions for improvement

The research set up can be criticized by the respondents in the comment areas. Therefore, the comments are listed and checked on any remarks about the absent *no answer option* and any other suggestions for improvement on the research set up.

3.3.6 Validity

The assessment is evaluated on its validity, which concerns how the content is based on a theoretical background. Questionnaires must ask what they claim to ask and not systematically generate misleading results, so the content needs to be similar to the corresponding theoretical foundation. Furthermore, validity concerns the degree to which the questionnaire covers theoretical knowledge, so the comments on the rating scales are analysed on their relatedness to the statements (Haklay, 2010; Scheepers et al., 2016).

Mainly, this study's method involves a guidance for selecting the most appropriate web GIS characteristics for an assessment (section 3.3.2). Hence, applying the method concerns including and excluding the characteristics. So, analysing the method's validity includes evaluating whether a sufficient set of characteristics is selected. That can be performed by analysing respondent comments on remarks about the absence or presence of certain web GIS characteristics.

Adjective rating scale

To interpret the final overall product scores, Bangor, Kortum and Miller (2008) suggest adding an adjective rating scale. Basically, it consists of three major components, an overall quality modifier, a synonym for usability that is likely to be quickly understood and a reference to the product. That results in a question which intends to add a qualitative answer for explaining overall usability scores. In specific, these adjectives are *worst imaginable*, *awful*, *poor*, *ok*, *good*, *excellent* and *best imaginable*. After data collection, mapping the sum of scores against the adjective rating scales provides insights in which overall scores are perceived as *worst imaginable*, *awful* and so forth (Bangor et al., 2008).

3.4 Summary

Two case studies are selected to assess the products of Floodlabel and GeoWeb. The Floodlabel respondents (n = 109) can be divided in user characteristics groups of educational background, user goal, device type and age. The GeoWeb respondents (n = 37) can be divided in groups based on GIS knowledge, age, educational level, user goal, frequency of use and organization.

Both case studies are used to apply this study's method on. Within figure 3.7, an overview of the method is provided. It contains of six steps, regarding the research set up (step 1), assessment content (step 2), creation of assessments (step 3), the results (step 4), the method's reliability (step 5) and validity (step 6). This section elaborated on why these steps and sub steps are designed, the results section (section 4), discusses the application on the case studies.

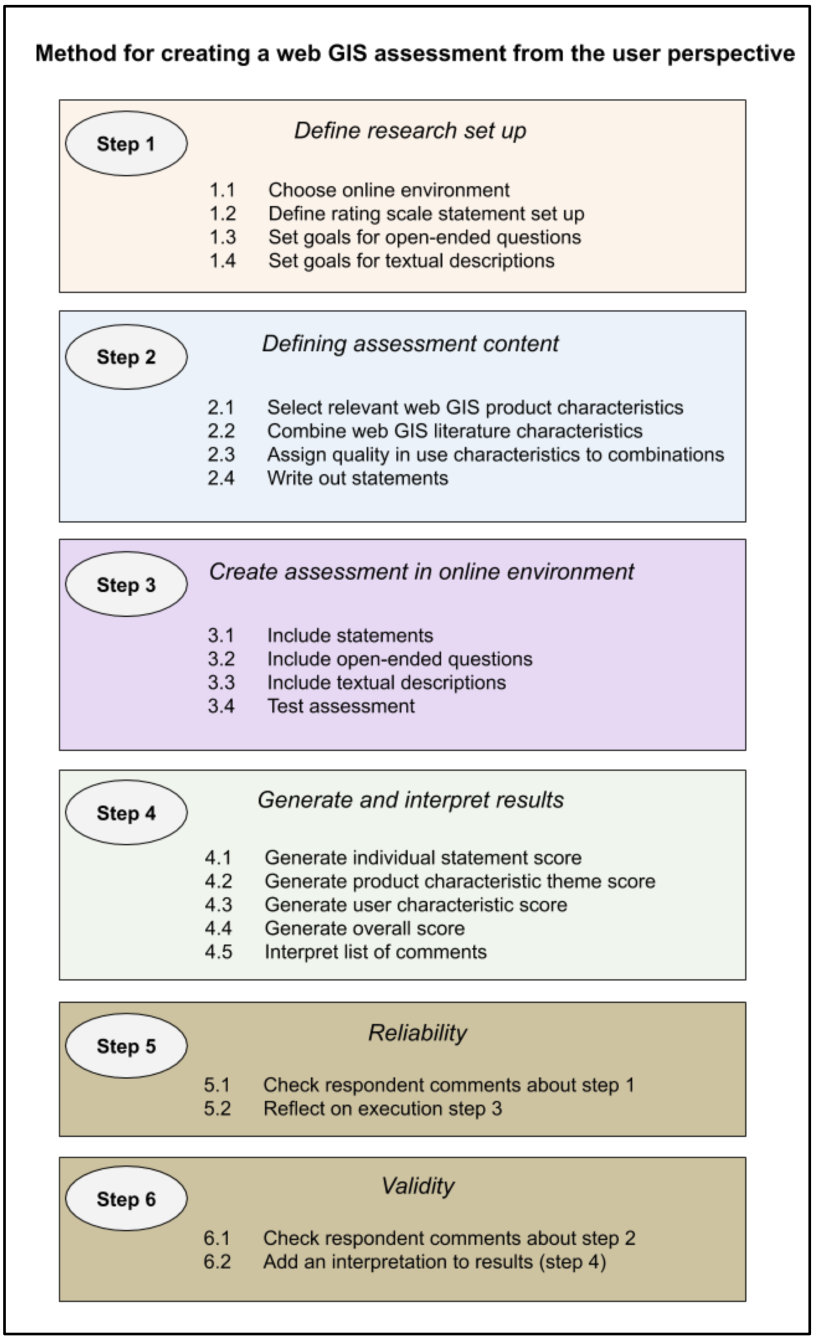

Figure 3.7 Method for creating web GIS assessments from the user perspective

4. Results

The research objective is to create a method for assessing web GIS products from the user perspective. Hence, the methodology section elaborated on the four-step approach to define the research set up (step 1), define questionnaire content (step 2), create the assessment (step 3), and generate and interpret the results (step 4). This section shows how the method is applied to the case studies. Section 4.1 elaborates on how the questionnaires are created. Section 4.2 on the resulting scores and how they can be interpreted. Section 4.3 and 4.4 respectively discuss the questionnaires' reliability and validity (step 5 and 6).

4.1 Case study questionnaires

Creating the questionnaire statements is based on following step 2 of the method (figure 3.3). In short, these steps filter the web GIS product characteristics on relevance (step 2.1), remove redundant characteristics (step 2.2), assign a quality in use characteristic (step 2.3) and written out clearly and understandable (step 2.4). Each of the steps is discussed in this section.

Step 2.1 – Relevance of web GIS product characteristics

The pool of potential web GIS product characteristics is used as starting point (figure 3.4). All characteristics are assessed on relevance for the case studies and visualised per product characteristic theme. The GeoWeb and Floodlabel results are showed in figure 4.2. Since the Floodlabel product does not include advanced handlings with georeferenced data, the blue coloured characteristics are not assigned as relevant for Floodlabel.

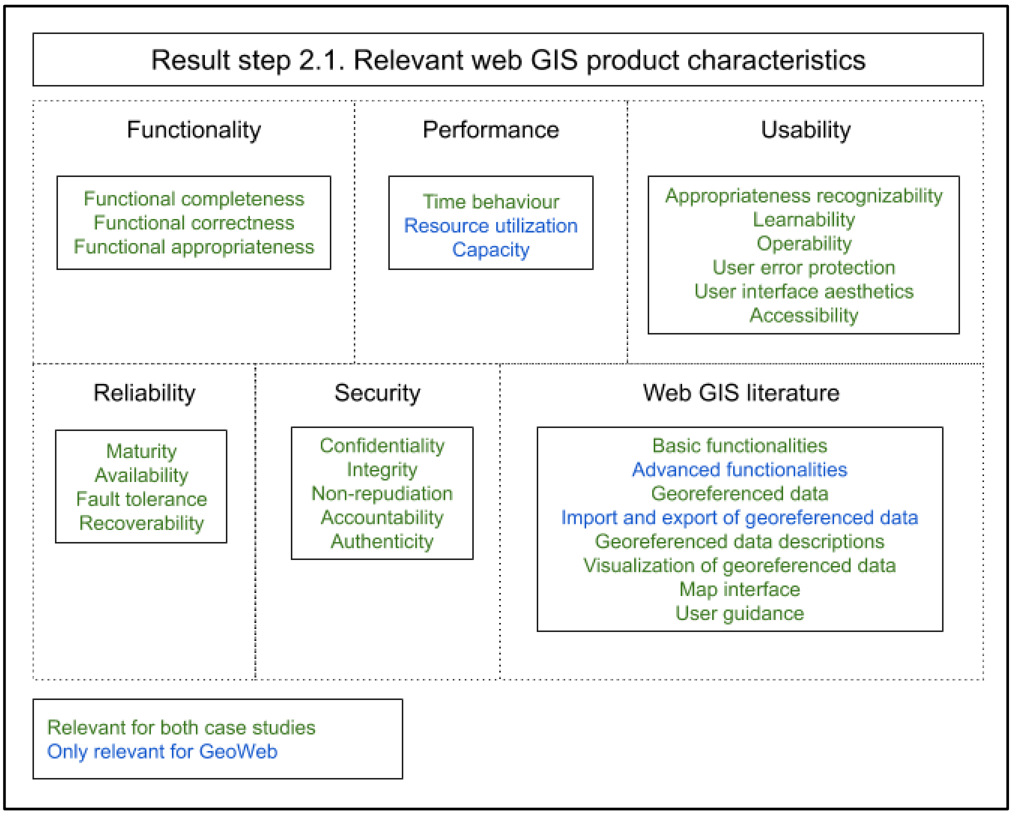

Figure 4.1 Assessment content relevance web GIS product characteristics

Step 2.2 – Combine web GIS product characteristics

With the purpose of limiting the number of characteristics to include in the assessment, step 2.2 is executed. For each theme except *web GIS literature*, the relative importance of its sub characteristics is considered. Multiple characteristics are excluded if they seem to be redundant in the questionnaire. As a result, figure 4.2 shows the list of the remaining characteristics.

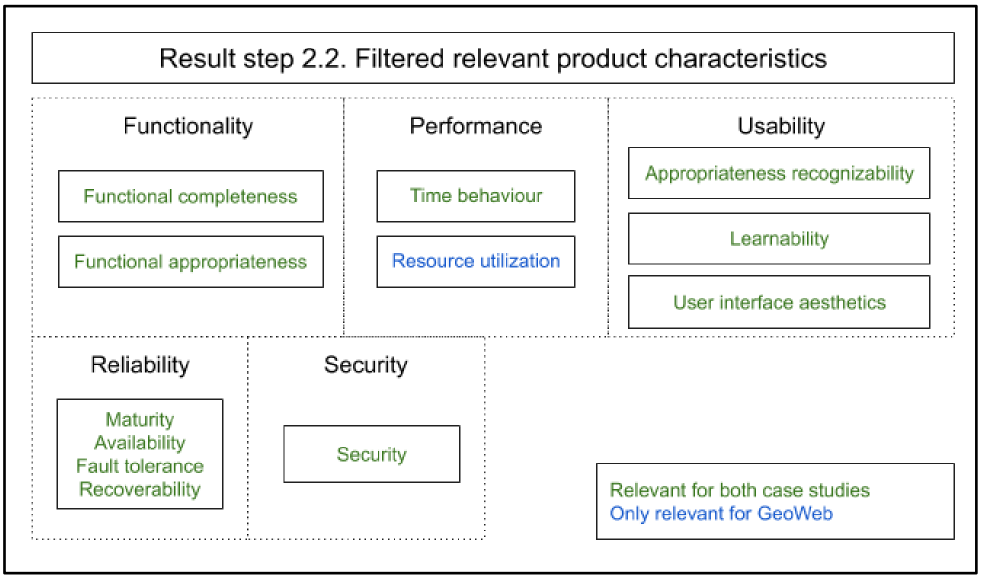

Figure 4.2 Filtered relevant characteristics

After that, *web GIS literature* characteristics are integrated in the others, using the set-up of table 4.1 where all combinations are identified with a green check mark. These combinations show which characteristics can be assessed together in a singular statement. For example, the *basic functionality* is combined with the *functional completeness* by examining the *completeness of basic functionalities*.

Table 4.1 Integrate web GIS literature characteristics into others

Characteristic	*Basic functionalities*	*Advanced functionalities*	*Georef. data*	*Import and export georef. data*	*Georefer. data descriptions*	*Visualization of georefer. data*	*Map interface*	*User guidance*
Functional completeness	✓							
Functional appropriateness		✓						
Time behaviour	✓	✓						
Resource utilization				✓				
Learnability								✓
Userface aesthetics							✓	
Appropriateness recognizability			✓		✓	✓		
Reliability								
Security								

As a result of the selection process, thirteen statements for GeoWeb and eleven for Floodlabel can be created. Within figure 4.3, the combinations are visualised in an overview for creating the statements.

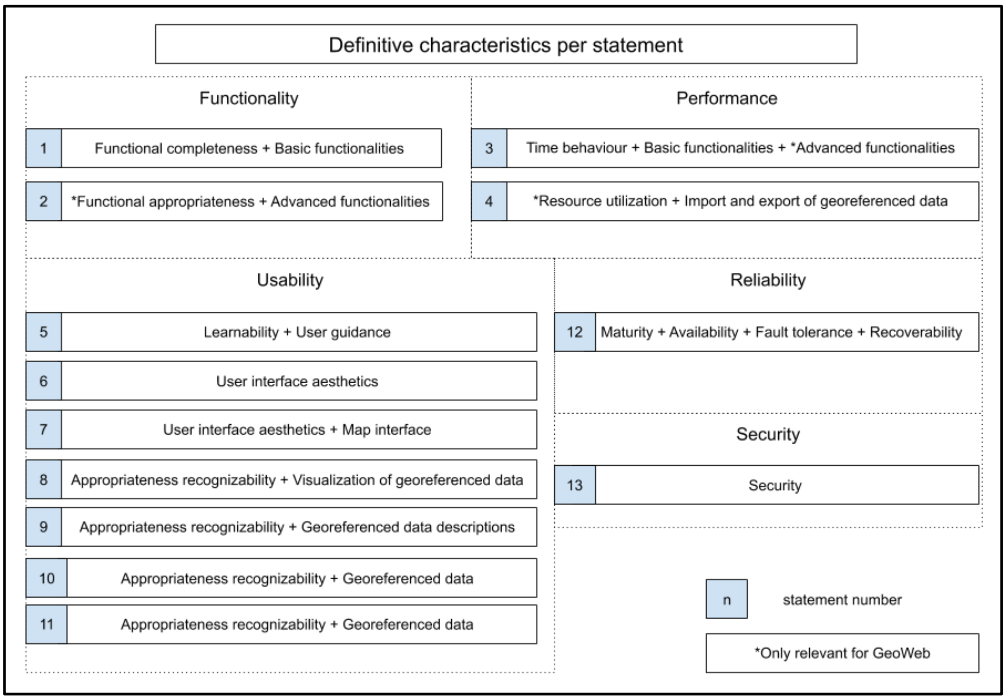

Figure 4.3 Product characteristic combinations for creating statements

Step 2.3 – Assign quality in use characteristic

The resulting statements listed in figure 4.3 represent the relevant components of the case study products. These characteristics must be assessed on experiences by using them. So, for each statement the most suitable *quality in use* characteristic is defined and listed in table 4.2. Considered is *how* the product characteristic supports reaching the user goal. As an example, the first statement aims at the *completeness of basic functionalities* and the assigned *quality in use* characteristic views this from the perspective of how *effective* the complete set of basic functionalities is for reaching the user goals.

Table 4.2 Quality in use characteristic per statement

Statement	***Quality in use characteristic***
1	*effectiveness*
2*	*efficiency*
3	*efficiency*
4*	*effectiveness*
5	*effectiveness*
6	*satisfaction*
7	*satisfaction*
8	*usefulness*
9	*effectiveness*
10	*usefulness*
11	*usefulness*
12	*freedom from risk*
13	*freedom from risk*

*for GeoWeb assessment only

Step 2.4 – Definitive statements

Both case studies have Dutch speaking respondents, so the statements are formulated in Dutch (an English written version is included in appendix 8.6).

Next to the thirteen statements of figure 4.3, three statements were added with the purpose of an adjective rating scale (16) and control variables for validating the operationalisation of *usability* (14) and *overall web GIS quality* (15). The results are visualised in table 4.3.

Table 4.3 Dutch written assessment statements with in-text reference

No.	***Text***
1	Ik kan goed gebruik maken van de kaart door te wisselen van thema en te scrollen, zoomen en bewegen op de kaart.
2	Er zijn niet meer of minder dan benodigde functionaliteiten beschikbaar die handelingen verrichten met geografische informatie (zoals (ruimtelijke-)zoekopdrachten of functionaliteiten voor overlay-, neighbourhood- en netwerkanalyses).
3	Ik ben tevreden over de snelheid van het product tijdens het uitvoeren van mijn handelingen.
4	Ik ben tevreden over de mogelijke dataformaten die het product kan verwerken als ik geografische data importeer of exporteer.
5	Er zijn genoeg beschikbare helpfuncties om het product (te leren) te gebruiken.
6	De interface van het complete product draagt positief bij aan wat ik met het product wil bereiken.
7	De interface van de interactieve kaart draagt positief bij aan wat ik met het product wil bereiken.

8	De geografische informatie op de kaart is duidelijk en valt goed te begrijpen.
9	Met behulp van bijvoorbeeld een legenda, titel, tekst of andere beschrijvingen van de geografische informatie op de kaart kan ik begrijpen waar het over gaat.
10	Ik ben van mening dat de geografische informatie voldoende up-to-date is voor mijn gebruikersdoelen.
11	De gekozen geografische informatie op de kaart draagt bij aan wat ik wil bereiken met het product.
12	Het product is betrouwbaar, werkend, beschikbaar en hersteld zich bij uitschakelen altijd als ik mijn gebruikerstaken uitvoer.
13	Ik ben van mening dat het product voldoende rekening houdt met data-veiligheid en privacy.
14	Ik ervaar het product als gebruiksvriendelijk.
15	Op basis van mijn ervaring met het product denk ik dat het geschikt is voor wat ik ermee wil bereiken.
16	Over het algemeen vind ik het niveau van de product

To summarize, figure 4.4 visualizes the performed steps related to the method's step 1 and 2. For each sub step the most important decisions, outcomes or considerations are highlighted.

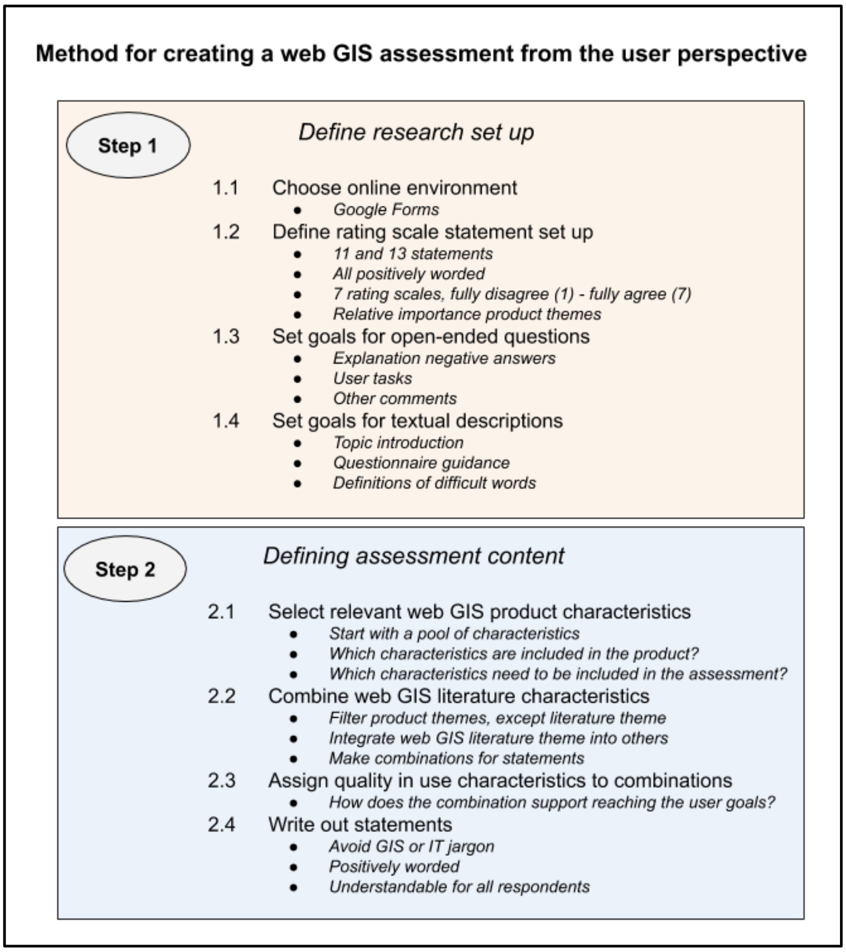

Figure 4.4 Application of the method's step 1 and 2

Step 3 – Creating and testing the questionnaires

The questionnaires are created in Google Forms and were tested by eighteen people with differing ages, educational backgrounds and affiliations with GIS. The feedback involved detailed suggestions for improvements of the questionnaire, e.g. *“avoid difficult words”*, *“increase limit of characteristics for open-ended questions”* and several lay-out related comments (figure 4.5). Furthermore, the average duration provided an indication of the questionnaire length and using this information the questionnaire was revised multiple times.

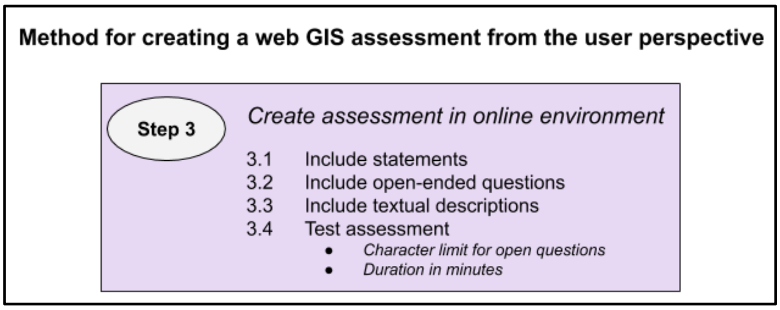

Figure 4.5 Application of the method's step 3

4.2 Case study scores

Applying step 4 of the method concerns generating and interpreting the questionnaire results (figure 3.6). After data collection, the statement ratings are transformed into scores for the overall product quality, five product quality themes and for each statement individually. Noteworthy, the scores for both case studies cannot be compared to each other, because of the differing assessments, user groups and web GIS products. The following scores are generated and identified in this section: *overall score, theme score, statement score, user characteristic score and control variables’ scores.*

Individual statement scores

The mean scores and percentages of negatively rated results of the individual thirteen statements are visualised in table 4.4.

Table 4.4 GeoWeb (left) and Floodlabel (right) mean scores and frequencies of low ratings per statement

Statement	***Mean***	***3 or less (%)***
1	5.24	14
2	4.00*	22
3	4.24*	30
4	4.19*	14
5	4.59*	16
6	4.59*	24
7	4.70*	14
8	5.32	8
9	5.30	8
10	4.95	22
11	5.38	5
12	4.73*	16
13	5.05	5
Mean	**4.78**	**16**

Statement	***Mean***	***3 or less (%)***
1	5.41	11
2	-	-
3	6.20	3
4	-	-
5	4.86*	21
6	5.05*	17
7	5.12*	16
8	5.63	6
9	5.35	13
10	5.39	7
11	5.31*	7
12	5.44	10
13	4.61*	19
Mean	**5.31**	**12**

*means below average mean

For both case studies the *usability* statements 5, 6 and 7 score below average, where *usability* statements 8, 9, 10 and 11 score on average or higher. These first three *usability* statements concern the *user guidance* for *learnability* and *interface*s of the product and others are related to *georeferenced data*.

Furthermore, almost one third of the GeoWeb respondents negatively rate the first *performance* statement (30 percent). Hence it is interesting for the GeoWeb product owners to list the explanation comments for this statement.

The majority of ratings is either neutral or higher for both case studies. As visible in the total row in table 4.4, sixteen and twelve percent of the answers are three or lower for respectively GeoWeb and Floodlabel.

Theme score
The final score is also generated for the five product characteristic themes which are visualised in table 4.5.

Table 4.5 Floodlabel (left) and GeoWeb (right) means per theme

Theme	***Mean rating***	***Mean importance***
Performance	6.20	2.90
Reliability	5.44	3.24
Functionalities	5.41	3.30
Usability	5.27	3.41
Security	4.61	3.06
Total	**5.32**	

Theme	***Mean rating***	***Mean importance***
Security	5.05	2.59
Usability	4.92	3.54
Reliability	4.73	3.54
Functionalities	4.62	3.24
Performance	4.22	3.32
Total	**4.77**	

Using these scores, themes can be ranked and compared. For example, *performance* scores 6.2 out of 7.0 for the Floodlabel case, which is a very positive result. In contrary, the security characteristic is rated the most negatively with 4.61. However, that score still is more positive than negative. Using the average ratings, themes can be ranked accordingly from high to low: *performance*, *reliability*, *functionalities*, *usability* and *security*.

According the GeoWeb ratings per theme, *performance* scores the least positive (4.22) and *security* the most positive (5.05). In comparison to the Floodlabel these extreme values are closer to the mean ratings of the total group of statements. For this group of GeoWeb users the following ranking can be made from high to low: *security*, *usability*, *reliability*, *functionalities* and *performance*.

Within both case studies the characteristics with the lowest perceived relative importance have the highest mean rating on the statements. Nevertheless, the mean importance ratings do not show noteworthy differences

User characteristic score
Table 4.6 visualizes the number of respondents, average rating and scaled score for two GeoWeb organizations. Although these scores provide insights in the overall scores of the organizations' products, comparing them requires more respondents per organization.

Table 4.6 GeoWeb overall score for two organizations

Organization	***N***	***Average statement rating***	***Overall score (0 -100)***
Water board of Limburg	12	5.12	69
Municipality of 's-Hertogenbosch	17	4.37	56

Regarding the Floodlabel education groups, the *low* (n = 9) and *middle* (n = 22) can be combined to assess the mean rating differences with the *high* education group (table 4.7). Both assumptions are met for performing a t-test ($n > 30$, $F(3.077)$, $p > .05$) and a t-test shows a significance difference in population means between these two groups: $t(107) = -3.710$, $p <$ 0.001. Which indicates that higher educated people tend to rate the product lower.

Table 4.7 Floodlabel education group descriptive statistics

Education group	***n***	***M***	***SD***
Low & middle	31	5.85	0.69
High	78	5.11	1.03

Furthermore, Floodlabel respondents using a desktop computer or laptop have slightly higher ratings but significant ($t(100) = 2.031$, $p < .05$). That difference is partly explained by the *functionality* statement, where a bigger group of respondents in the *desktop* group rated the statement with a 7 (28.57 percent) in comparison to a single 7 within the *mobile* population (2.56 percent). Performing a t-test shows a significant difference in population means ($F(.003)$, $p = .958$, $t(100) = 2.294$, $p = < 0.25$). So, people using a mobile device tend to rate the product functionalities lower than people on a desktop or laptop.

Table 4.8 Floodlabel device type descriptive statistics

Device	***n***	***Overall mean***	***Overall SD***	***Funct. mean***	***Funct. SD***
Desktop or laptop	63	5.46	1.35	5.63	1.35
Mobile	39	5.04	1.38	5.00	1.38

Within the questionnaire the user goals of the respondents were acquired. Section 3.2.4 explains how GeoWeb users can be separated into groups of basic and advanced use. For each web GIS characteristic theme, scores for both groups are generated (table 4.9).

Table 4.9 GeoWeb basic and advanced user product theme average ratings

User group	***N***	***Total***	***Functio -nalities***	***Perform-ance***	***Usability***	***Reliability***	***Security***
Basic	23	4.67	4.37	4.20	4.80	4.57	5.17
Advanced	14	4.95	5.04	4.25	5.10	5.00	4.86

Regarding these results, the functionality statements are on average rated with a 4 by users with basic goals and a 5 by users with advanced goals. These findings might indicate that users with basic knowledge and user goals tend to rate the functionality statements lower, which is a motivation to perform a Mann-Whitney U test on the mean ratings. This test shows a significant difference between these two respondent groups on the mean functionality rating ($U = 98$, $p < .05$). Nevertheless, should be noted that this regards a slight difference and the numbers of respondents per group are twenty-three and fourteen.

Overall score

The overall score covers the statements regarding the first thirteen statements of figure 4.3.

Table 4.10 Case study final scores

Case study	***Number of valid results***	***Mean rating per statement***	***Number of statements***	***Overall score (0 -100)***
Floodlabel	109	5.31	11	72
GeoWeb	37	4.78	13	63

For both case studies the overall score frequencies can be visualised in histograms figures 4.6 and 4.7. These histograms show the division of overall ratings and provide insights in how the respondent experiences using the product in general.

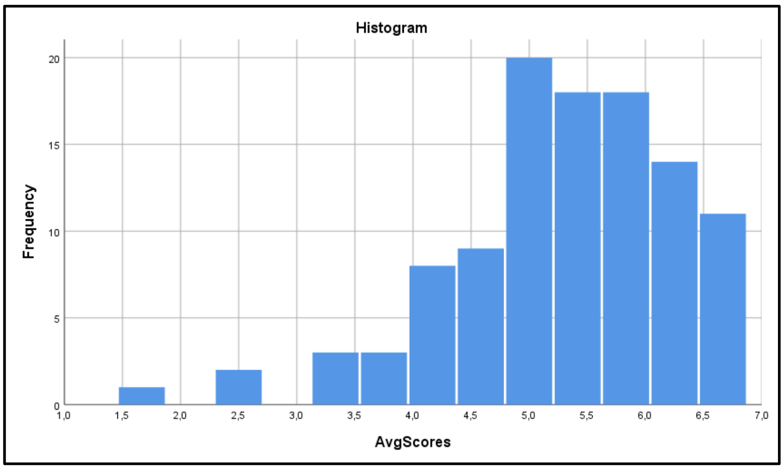

Figure 4.6 Histogram of the frequency average scores of Floodlabel

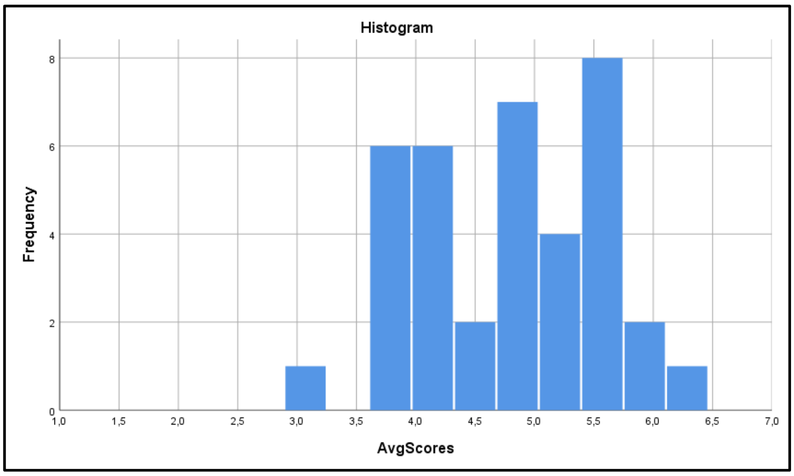

Figure 4.7 Histogram of the frequency average scores of GeoWeb

To summarize, scores are created for different purposes and can be interpreted in various ways. An overview of the performed actions regarding step 4 of the method is provided in figure 4.8.

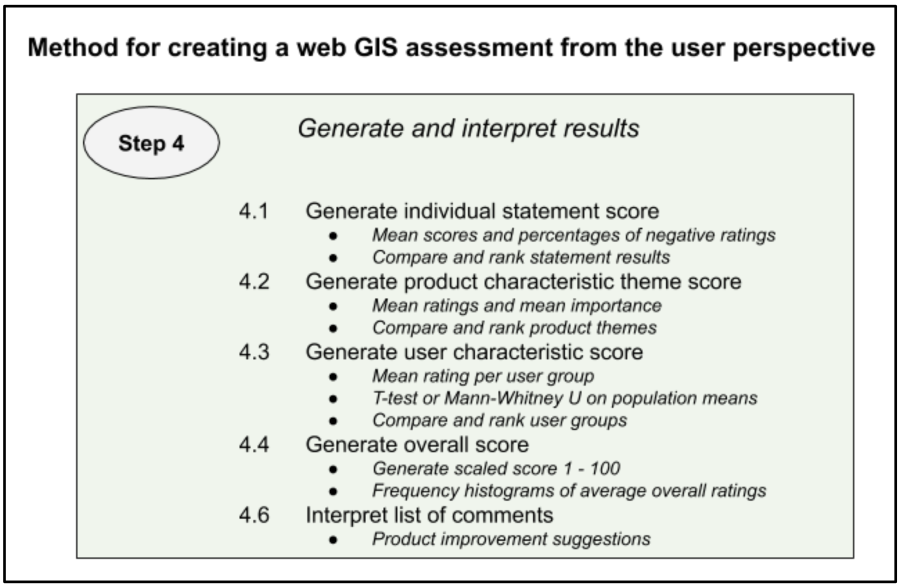

Figure 4.8 Application of the method's step 4

4.3 Assessment method reliability

This section elaborates on the reliability of the assessments, step 5 of the method. Based on the provided comments by the respondent the absent *no answer rating* and other suggestions are discussed in this section. Additionally, two small remarks are made on the research set up.

No answer

Section 3.3.1 discusses adding an option of *'no answer'* to the rating scales. For both case studies, respondents mentioned any inability to answer a question. Regarding Floodlabel, an overview of these comments is visualized in appendix 8.7. For example, *'no opinion'*, *'no idea'* or *'difficult to rate, again'* indicate lacking knowledge to rate a statement. However, the impact on the final results seems to be low because of the total number of 109 respondents.

Within the GeoWeb group, respondent 20 made the following statement: *"Not all questions can be answered sufficiently. Since the presence of GeoWeb components that I have never seen before, I cannot have an opinion about them. However, it is not possible to answer 'not applicable' or 'no opinion'."* Besides that, other relevant comments are listed in appendix table 8.9 and similar to the Floodlabel case, these numbers represent a small part of the total numbers of ratings per statement.

Suggestions for improvements research set up

For both case studies, a sum of five respondents made remarks about the maximum of hundred characters for the text fields, which was perceived to be low for a few questions.

According to the explanation comments related to the statements, Floodlabel's respondent 72 identifies the possibility to give an explanation if the rating is not 3 or less. That is because of the lacking validation based on the previous statement answer, which enables providing answers in all situations. As a result, not all respondents have filled in a comment when they rated a statement with 3 or less and some respondents filled in a comment when they rated a statement 4 or higher. Because the number of comments is not assessed in the analysis section, there is no indication of a violation to reliability.

Regarding the Floodlabel case, respondent 36 mentioned the overload of difficult and theoretical concepts, *"which negatively influences the understandability and speed of filling in the questionnaire."* Expanded with the remark of *"it took way more time than promised."* This respondent has the age of 70 years, is highly educated and is an outlier of the minutes spent on the questionnaire (35 minutes). One respondent (ID = 32) misinterpreted the way of rating the statements either positively or negatively.

Other improvements

Acquiring the user types can be improved by including a closed question with a few in advance defined user tasks. Defining specific user profiles is based on the characteristics of age, device type, GIS knowledge, educational level and user tasks. The latter was acquired with an open question in the assessments. For the GeoWeb case, the user tasks are divided into groups of basic and advanced use. However, for the Floodlabel case the results are too varying and a lot of them are irrelevant (appendix 8.4). In hindsight, a closed question with a few in advance defined user tasks would be more sufficient.

Besides that, the question to explain low rated answers could be changed to a question focusing on all ratings instead. This would stimulate an increase in number of comments and would contribute to more input for reliability and validity analysis or comments about the product itself.

Overall, there are no strong indications of violations to reliability. Several improvements of assumingly minor influence on the reliability are identified by the respondents but the research set up is appropriate for the purpose of the assessment. Figure 4.9 shows the performed actions regarding the method's step 5.

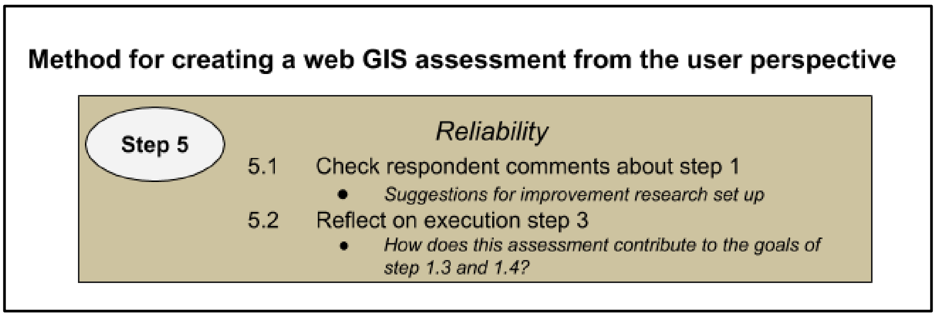

Figure 4.9 Application of the method's step 5

4.4 Assessment method validity

This section discusses the method's validity, step 6 of the method. That is performed by evaluating the respondent comments and interpreting the overall scores with the adjective rating scale results. Respondent comments are used to discuss the questionnaire content on whether it is selected appropriately and whether the statements are correctly interpreted. Next to that, the adjective rating scale provides an interpretation of the statement ratings.

Validity of defining questionnaire content (step 2)
Several respondents make comments on the absence or presence of certain web GIS characteristics, these are highlighted in this section to discuss the method's validity (section 3.3.6).

Firstly, *functional correctness* is excluded from the Floodlabel assessment, but respondent 71 mentioned the functional incorrectness of the theme switch arrows. That supports the decision of covering *functional correctness* by the *functional completeness* (figure 4.2).

Secondly, GeoWeb respondent 24 mentioned that the capacity limits are rapidly exceeded when importing shapefiles. That comment supports integrating *capacity* into the *resource utilization* statement (figure 4.2).

Thirdly, another example is related to the overall usability statement, where Floodlabel respondent 29 commented that the product is lacking any audio function for reading aloud. That comment supports the decision of integrating *accessibility* in the other *usability* statements (figure 4.2).

Fourthly, based on Floodlabel comments at multiple statements, an additional statement regarding the completeness of *georeferenced data* might have been sufficient, because five respondents (10, 61, 73, 80 and 93) mentioned the absent information about their residence.

Understandability of statements
For relevant comments is assessed whether it is interpreted *incorrect, partially correct* or *correct* and the results are provided in appendix 8.8. Notable comments related to statements 5, 6, 7 and 13 are highlighted in this section.

Firstly, statement 5 concerns the degree to which the product provides sufficient user guidance for (learning) to use the product. Essentially, this statement should be assessed *by the degree to which the user is able (to learn) to use the system and if necessary, with the support of user guidance* (appendix 8.2). However, this statement might be too complex, since eight respondents only mentioned the absent user guidance. Hence, these respondents are not mentioning their ability (to learn) to use the system and might have interpreted the statements as whether there is any user guidance.

Secondly, statements 6 and 7 include the concept of an interface. In contrary to the Floodlabel assessment, an interface definition is not provided in the GeoWeb questionnaire. Accordingly, several comments are: *"Excuse me? Do you mean that it is easy to use? No, not always."*, *"I don't understand what you mean"* and *What is an interface?"* These comments indicate that the definition of *interface* should have been included in the GeoWeb questionnaire to enhance the understandability of the statement.

Thirdly, the reliability statement 13 caused a misunderstanding of the concept for Floodlabel respondent 10: *"Reliable: that needs to be proved in a real emergency situation!"* Which clearly is another meaning of reliability.

Interpretation statement results
The adjective rating scale provides an interpretation of the overall ratings (section 3.3.6). Regarding both case studies the average adjective rating scale is rounded off to *fair* (Floodlabel 4.88 and GeoWeb 4.59) and the modus of GeoWeb is *Adequate*.

Within figure 4.10 the box plots for the average ratings per adjective rating scale the Floodlabel case study are visualised. In general, this figure shows that the higher adjective rating scales have higher minimum *average ratings*, mean *average ratings* and interquartile range (IQR).

Analysing the adjective rating scale and the average ratings is performed with a Kruskal-Wallis H test on the rankings. Both cases indicate significant effects of choosing an adjective rating scale on the average rating scale statements, Floodlabel χ^2 (6) = 67.170, $p = 0.000$ and GeoWeb χ^2 (5) = 17.11, $p = 0.000$. This indicates that people tend to give a higher adjective rating scale when they rate all statements on average higher. However, should be noted that is desirable to have bigger sample sizes to enhance the reliability of the tests.

More specific, all respondents rating the product as *excellent* have an average rating between 6 and 7. Noteworthy, at least a quarter of the respondents rating Floodlabel with negatively worded ratings have an average rating of 4 or higher. So, *inadequate* or *marginal* correlates for several respondents with a neutral score of 4 or a positive score of 5.

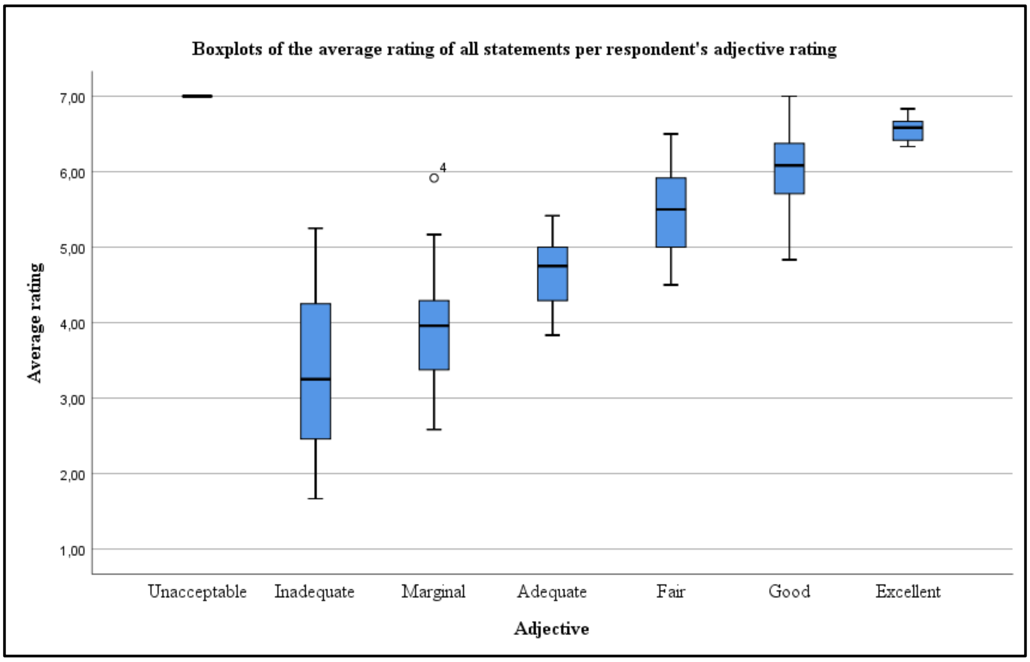

Figure 4.10 Boxplots Floodlabel adjective rating scale and average rating

To summarize, figure 4.11 visualizes the performed actions regarding the method's step 6.

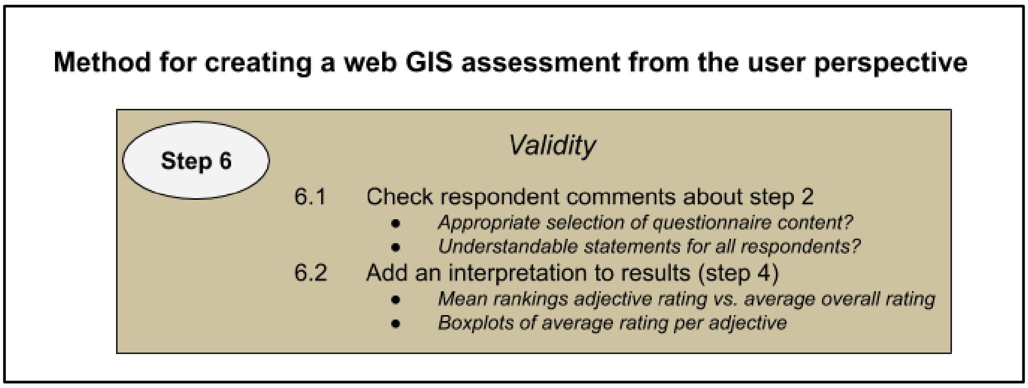

Figure 4.11 Application of the method's step 6

4.5 Summary

The method for developing a web GIS assessment from the user perspective is applied to the Floodlabel and GeoWeb cases. An overview of the most important results is provided in figure 4.12, including comments, performed actions and considerations for applying the study's method. Step 1 is mostly applied in the methodology section 3.3.1.

The results of applying step 2 are described in section 4.1. Starting with a pool of characteristics, the most relevant product characteristics are selected for the case studies. Subsequently, these characteristics are filtered and the *web GIS literature* characteristics are integrated into the other themes. Within figure 4.3, the relevant combinations of web GIS product characteristics are visualised. To each of these combinations the most appropriate *quality in use* characteristic is assigned. These combinations are used to create the definitive statements, which are visualised in table 4.3.

After applying step 3 to create the assessment and disseminate it among the respondents, step 4 is applied to generate and interpret the results. Section 4.2 elaborates on analysing questionnaire results, ratings are generated for the individual statements, groups of statements, groups of users and the overall product. This section shows the added value of using mean ratings, frequencies, rankings, histograms and t-tests to analyse the assessment results.

Within section 4.3 the method's reliability is discussed using respondent feedback (step 5). Several respondents suggest raising the hundred-character text field limit and mention the inability to answer a statement. The latter can be solved by adding a *no answer* option, but for this study the absence of a *no answer* option seems sufficient. Overall, the comments do not show strong violations to reliability, so, the research set up is appropriate for the purpose of the assessment.

The final section 4.4 discusses the method's validity (step 6). Multiple respondents mention excluded web GIS characteristics such as the *functional correctness* and *capacity*. These comments support the decisions to cover the concepts by other characteristics. Based on comments of five respondents the Floodlabel questionnaire can be improved by including a statement about *georeferenced data completeness*. Next to that, GeoWeb respondent comments emphasize the need for providing definitions of technical concepts, such as an interface. Lastly, the adjective rating scale results provide an interpretation of the overall statement ratings.

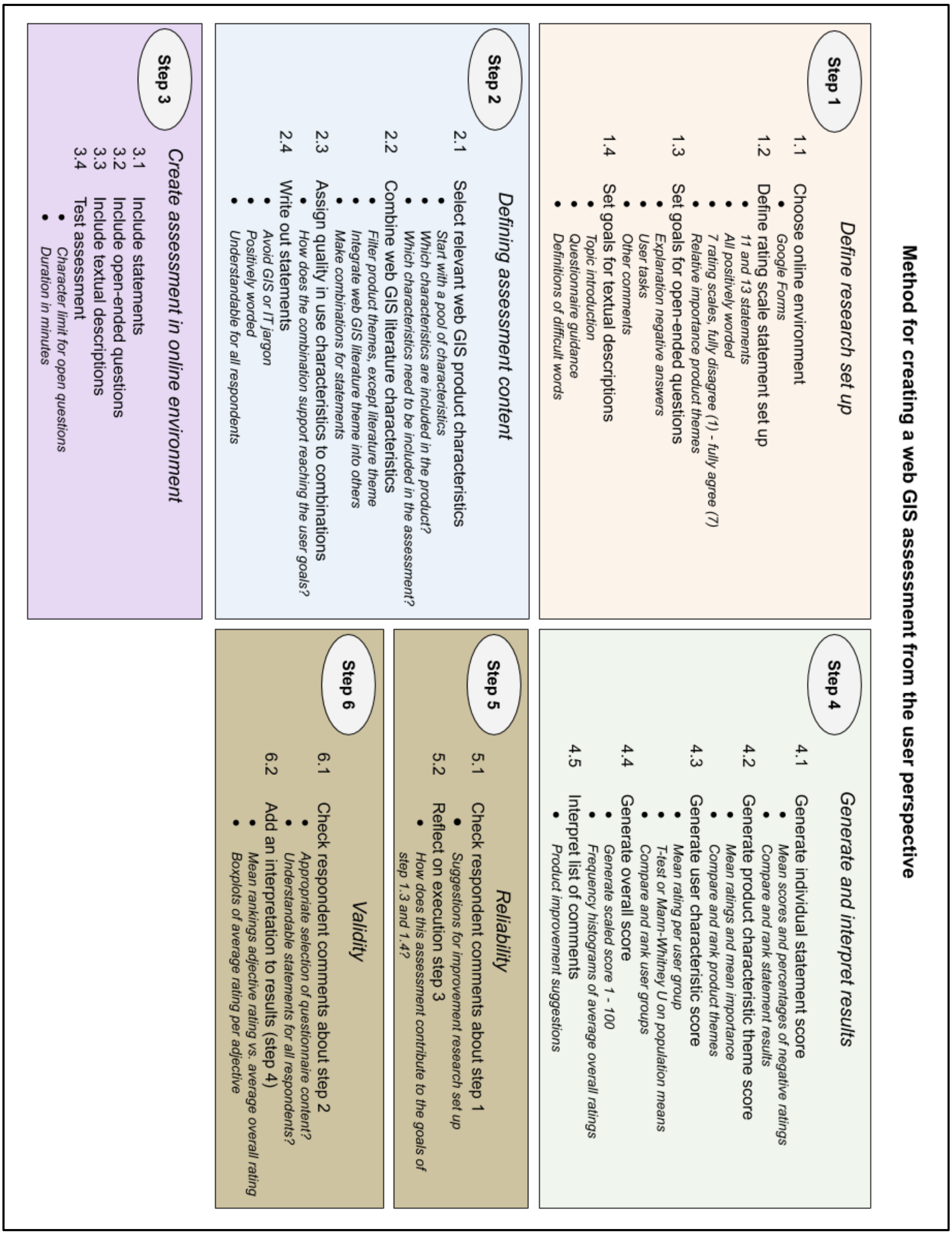

Figure 4.12 The complete method for creating a web GIS assessment from the user perspective

5. Conclusion

This study developed a method to for creating web GIS assessments from the user perspective. To achieve this, the literature background section provided information about web GIS products, standardized assessment methods and assessments from the user perspective. Using that information, a research set up for the assessments is suggested, an approach to define the content is provided and the questionnaire results are generated and analysed. Within this section the research questions are answered to summarize and conclude the findings for reaching the research objective. To begin with briefly answering the three sub research questions.

What is generally understood as a web GIS application?

Generally understood as a web GIS is *"a fully-fledged GIS that takes advantage of web technology for communicating among its components (data, functionality, and interface)*" (Veenendaal et al., 2017). Additionally, *georeferenced data* is key to all involved handlings within a GIS (Huisman & de By, 2009, p. 32). In comparison to general software products, GIS have distinctive characteristics, such as *basic and advanced GIS functionalities*, *the map interface*, *user guidance* and the *import*, *export*, *description*s and *visualization of georeferenced data* (Heywood et al., 2006; Poplin, 2015; Tolpekin & Stein, 2013; Unrau & Kray, 2019; Veenendaal et al., 2017). These product characteristics can be used to divide a web GIS in smaller and measurable parts.

What is generally understood as a standardized assessment method?

Within the field of software quality assessments, a standardized assessment method involves an approach to define quality in small parts of the product and use their definitions to create an assessment. This study focuses on the quality models of ISO, which provides internationally accepted quality models with definitions of software characteristics (ISO/IEC 25010, 2011). These models support assessing the human-computer interaction of a web GIS. Within figure 2.11, an overview is provided how the human-computer interaction components relate to each other and can be divided into smaller manageable parts. By providing definitions of quality, the quality models support the process of defining what is relevant for an assessment. So, web GIS assessment developers define what needs to be evaluated from a pool of web GIS characteristics. In addition to the list of web GIS characteristics from sub question one, the following characteristics are relevant as well: *functionality*, *performance*, *usability*, *reliability* and *security*. So, the assessment method includes defining the questionnaire with the support of these web GIS product characteristics.

How can web GIS applications be assessed from the user perspective?

To assess web GIS products from the user perspective, rating scale statements are an appropriate research set up. Developers can translate the selected web GIS characteristics into statements, which users can rate on the level to which they agree. The resulting statement ratings can be analysed and compared. In this way, the user is able to assess the product on how it facilitates reaching the user goals and developers can compare ratings between different users based on user tasks or characteristics such as GIS knowledge or age.

With this literature background information, the assessment method is developed for two case studies, Floodlabel and GeoWeb. The Floodlabel case study targets a prototype tool for informing Dutch homeowners on flood risks and mitigation measures (Urban Europe, 2017). Within this application an interactive map is integrated to visualize georeferenced data. A total of 109 respondents have participated to the research. The GeoWeb case study targets a web GIS application of Sweco, which is used by numerous customer organizations (Sweco Nederland B.V., 2020). That application facilitates advanced functionalities handling georeferenced data and is used by relatively more experienced GIS users. For that case study a total of 37 respondents participated to the research. Both case studies are used to apply the created method with the purpose of answering the main research question.

What is an appropriate standardized method for developing web GIS assessments from the perspective of the user?

In short, the assessment method includes six steps to create an assessment, which are described in figure 4.12. The first step regards the research set up, the second concerns defining the assessment content, the third guides by creating the assessments and the fourth step is about generating and interpreting the results. After applying the method, the final steps examine the methods reliability and validity.

The study results are four-fold, namely, applying the method to create the questionnaires (section 4.1), generating assessment ratings and interpret the results (section 4.2) and discussing the method's reliability (section 4.3) and validity (section 4.4). In this way, all six steps of the method are examined on their application to the case studies. The next paragraphs elaborate on the results, following the results section order.

Defining questionnaire statements
Step 2 of the method includes four sub steps to define what needs to be assessed (section 3.3.2). The pool of web GIS product characteristics as starting point (figure 2.11). The first sub step concerns selecting all potentially relevant web GIS product characteristics. Since the number of statements must be limited as much as possible, characteristics are removed and combined in the second sub step. Thirdly, the remaining set is assigned with a quality in use characteristic and finally, the statements are written out using the definitions of the characteristics, only positive wording statements and as less as possible GIS jargon (table 4.3).

Questionnaire results
The method's step 3 concerns generating and interpreting the questionnaire results. Final rating scores are generated per statement, groups of statements, groups of users and the overall product. These scores are analysed and compared by rankings, t-tests and histograms. These interpretations provide useful insights in how users experience the product. However, developers must be careful with drawing conclusions when interpreting the questionnaire results. Since, the research set up is designed to include perceptions of quality, the results are of an ordinal measurement scale. Hence, they only identify differences in ratings and are not suitable for arithmetic computations. The adjective rating scale is a useful tool to provide

additional insights for interpreting the results. However, it is always important to be careful with interpretations of the assessment ratings and emphasize the ordinal measurement scale.

The overall score interpretation is supported by the adjective rating scale. Using the boxplots of figure 4.8 supports estimating a correlated adjective for an average final rating. However, the ranges of the average ratings per adjective are too big to assign an adjective to an average rating.

Method's reliability
To assess step 1 of the method, respondents were asked to provide feedback on the research set up. Accordingly, they mentioned several small improvements, as raising the field limit of hundred characters for the open answers and considering an option to not rating a statement. Improvements can be made to acquire the user tasks and comments on statement ratings. Nevertheless, the suggested improvements have small influence on the case study results and reliability requirements are met. Thus, can be stated that the rating scale statement in an online questionnaire is experienced as an appropriate research set up.

Method's validity
The validity section assesses step 2 and 3 of the method. Respondent comments are examined to evaluate the selection process of the questionnaire content (step 2). Several respondents support the exclusion of product characteristics such as the *capacity*. However, multiple respondents mention the incompleteness of the *georeferenced data* for Floodlabel and an additional statement targeting the completeness of the *georeferenced data* would have been sufficient. Overall, most of the chosen product characteristics seem to be appropriate according to the respondent comments.

Furthermore, additional attention should have been paid to providing definitions of IT jargon and formulation of statements. The first *user guidance* statement is interpreted incorrectly by eight Floodlabel respondents, because the statement was not simple enough. Besides, several GeoWeb respondents struggled with the definition of an interface, which was lacking in the questionnaire description. It can be concluded that sufficient attention is needed for creating easy to understand statements (step 3).

Respondents were asked to explain, if they wanted, why they rated to statements negatively. In hindsight, if that question was obliged, the time spent on the questionnaire was raised significantly, but more meaningful results were generated. Those results can be used for the reliability and validity of the questionnaire, because they provide insights in how the respondents interpret the statements.

To conclude, the designed method is appropriate for designing web GIS assessments from the user perspective. With the support of the six steps and their sub steps, developers are able to assess web GIS products appropriately. That is, because the rating scale statements set up ensures a consistent questionnaire, which can be easily understood by respondents (step 1); statements can be created for relevant product characteristics (step 2); replicable and understandable questionnaires can be created online (step 3); and results can be generated and interpreted to acquire meaningful insights in the product (step 4). Finally, examining the method's reliability and validity puts the assessment into right perspective (step 5 and 6).

However, a successful application of the method depends on the developer's ability to select the most appropriate web GIS characteristics, write the statements out clearly and understandable, provide sufficient user guidance for participating to the research, avoid GIS jargon and provide definitions. Most importantly, developers must consider to which degree the targeted user group is able to understand the questionnaire.

6. Discussion

This section puts the study in perspective of the current literature work, elaborates on the study's scope and identifies challenges for further research.

Assessing a web GIS from the user perspective concerns how a user experiences a product. Targeting that experience requires focusing on three main components: the user, the computer and the interaction. These components must be included in an assessment.

Since assessment developers know their products, they have to define which product characteristics need to be assessed. That ensures the user is able to share experiences of interacting with the relevant product characteristics. To put these experiences into the right perspective, the user characteristics must be included in the assessment as well.

The research objective is to develop a method for creating web GIS assessments from the user perspective. This method includes an appropriate research set up, a four-step approach of writing out questionnaire statements, and analysis for generating and interpreting the results. These components are based on findings from scientific literature, which is discussed in the following section.

Literature and scope

Firstly, to define relevant web GIS characteristics for a user assessment, the following GIS introductory books and GIS usability studies are explored: Agrawal and Gupta (2017), Heywood et al. (2006), Poplin (2015), Tolpekin and Stein (2013), Unrau and Kray (2019) and Veenendaal et al. (2017). Based on that knowledge, a list of web GIS product characteristics is selected to function as a starting point of defining what is important (table 2.1). That list can be used in further research, saving time to explore potential web GIS characteristics. However, one should consider other web GIS characteristics as well, because of the widely varying and changing types of web GIS products. Hence, less general or new characteristics might be relevant for an assessment.

Secondly, the method's framework is based on the quality models of the International Organization of Standardization. These models operationalize the abstract and vague concept of quality into smaller and more manageable parts (ISO/IEC 25010, 2011; Wagner, 2013). The models are widely used and internationally accepted. However, the models consist of arbitrary characteristics that must be evaluated by assessments developers if they use them for their products.

Thirdly, the research set up for assessing web GIS from the user perspective is based on usability studies about SUS and PSSUQ. Both methods are using intercorrelated rating scale statements, where respondents rate a level to which they agree to the statements (Brooke, 1996; Lewis, 2002). To acquire the user perception such a subjective measurement is sufficient. However, an objective measurement tool might increase the added value of the assessment. As an example, loading time can be monitored and complemented to subjective performance experiences of the user. That would give assessment developers a better understanding of desirable performance goals.

Further research
To acquire more insights of the appropriateness of the method, the method should be applied to more case studies, executed by persons who are not yet involved in this study and be analysed in further research. That research must explore and solve occurring challenges, such as selecting relevant web GIS characteristics, write the statements out clearly and understandable, generate interpretable results and include objective measurement methods to strengthen the results. The only way of improving the method is using and testing it.

From now on, the method can be used by any individual or organization to assess a web GIS. The step-by-step approach provides guidance by quickly and easily creating an assessment and generate results. In this way, people can identify how users think about a web GIS, how it supports reaching the user goals and what can be improved. These insights provide input for discussions between product owners, product developers and users.

If the method is used more frequently by assessment developers, additional research can be done by interviewing the developers. That research can assess the method on how the developers experience applying it. Subsequently, results from these studies can be used to make improvements for using the method in practice.

In the future the method can be widely used by every organization. For example, organizations with web GIS products installed at numerous customer organizations and with multiple product versions. Comparisons can be made between different product versions or product-implementations. Questions can raise as: why is the product at organization A better experienced than at organization B? Are all product characteristics better experienced at newer product versions? Which users struggle with the web GIS? How can we improve our web GIS to support the users?

Within ten years, the method can be as widely used as the System Usability Scale, if organizations and researchers pick it up, analyse it and make improvements. The method's applicability, replicability and simplicity are strong pros for using it as a guidance.

After all, this study has taken a first step for developing a standardized web GIS assessment method. Yet, the method should be examined by more researchers and applied for non-scientific purposes as well. That would lead to more insights in what an appropriate method is, how it can be applied, and which challenges occur. The method is ready for application on a much wider scale and, hence, developing to a standardized guidance for anyone who wants to assess a web GIS.

7. References

Agrawal, S. & Gupta, R. D. (2017). Web GIS and its architecture: a review. *Arabian Journal of Geosciences*, *10*(23), 1–13.

Bangor, A., Kortum, P. T. & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction*, *24*(6), 574–594.

Brooke, J. (1996). SUS - A quick and dirty usability scale. In *Usability Evaluation In Industry*, 189–194. London: Taylor & Francis.

Brooke, J. (2013). SUS: A Retrospective. *Journal of Usability Studies*, *08*(2), 29–40.

Castronova, A. M., Goodall, J. L. & Elag, M. M. (2013). Models as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) standard. *Environmental Modelling and Software*, *41*, 72–83.

CBS. (2016). Standaard Onderwijsindeling 2016. Retrieved October 8, 2019, from https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/onderwijs-en-beroepen/standaard-onderwijsindeling--soi--/standaard-onderwijsindeling-2016

de Vocht, A. (2016). *Basishandboek SPSS 23: IBM SPSS Statistics* (2nd ed.). Utrecht: Bijleveld Press.

Eur-Lex. (2016). Document 32016R0679 (General Data Protection Regulation). Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Finstad, K. (2010). The usability metric for user experience. *Interacting with Computers*, *22*(5), 323–327.

Google. (2019). Get started with Forms. Retrieved December 4, 2019, from G Suite Learning Center website: https://support.google.com/a/users/answer/9302965?hl=en

Haklay, M. M. (2010). *Interacting with Geospatial Technologies* (1st ed.). Chichester, West Sussex, United Kingdom: John Wiley & Sons, Inc.

Heywood, I., Cornelius, S. & Carver, S. (2006). *An introduction to geograhical information systems* (3rd ed.). Harlow, England: Pearson Education Limited.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human Computer Studies*, *64*(2), 79–102.

Huisman, O. & de By, R. A. (2009). *Principles of geographic information systems: an introductory textbook.* Enschede, The Netherlands: The International Institute for Geo-Information Science and Earth Observation (ITC).

ISO/IEC 25010. (2011). Systems and software Quality Requirements and Evaluation (SQuaRE). In *BSI Standards Publication*. Retrieved from https://pdfs.semanticscholar.org/57a5/b99eceff9da205e244337c9f4678b5b23d25.pdf

ISO/IEC 25020. (2007). *INTERNATIONAL STANDARD. ISO / IEC 25020*. Retrieved from https://www.academia.edu/6692668/ISO_IEC_25020_2007_E_Character_PDF_document

Kujala, S. (2008). Effective user involvement in product development by improving the analysis of user needs. *Behaviour and Information Technology*, *27*(6), 457–473.

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human-Computer Interaction, 14*(3–4), 463–488.

Lewis, J. R. (2014). Usability: Lessons Learned. and Yet to Be Learned. *International Journal of Human-Computer Interaction*, *30*(9), 663–684.

Met Office. (2019). UK weather map. Retrieved December 2, 2019, from https://www.metoffice.gov.uk/public/weather/forecast/map/gcpvj0v07#?zoom=9&lat=51.56&lon=-0.00&map=SignificantWeather&fcTime=1575230400

Nelen & Schuurmans. (2019). Producten. Retrieved December 5, 2019, from https://www.nelen-schuurmans.nl/

P. Miguel, J., Mauricio, D. & Rodríguez, G. (2014). A Review of Software Quality Models for the Evaluation of Software Products. *International Journal of Software Engineering & Applications*, *5*(6), 31–53.

PDOK. (2020). Nationaal Georegister. Kaart. Retrieved April 3, 2020, from http://nationaalgeoregister.nl/geonetwork/srv/dut/catalog.search#/map

Poplin, A. (2015). How user-friendly are online interactive maps? Survey based on experiments with heterogeneous users. *Cartography and Geographic Information Science*, *42*(4), 358–376.

Sauro, J. & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? *Conference on Human Factors in Computing Systems - Proceedings*, 2215–2223.

Sauro, J. & Lewis, J. R. (2012). Quantifying the User Experience. In *Practical statistics for user research*. Waltham, USA: Elsevier.

Scheepers, P., Tobi, H. & Boeije, H. (2016). *Onderzoeksmethoden* (9th ed.). Amsterdam: Boom uitgevers.

Sibisi, M. & van Waveren, C. C. (2007). A process framework for customising software quality models. *IEEE AFRICON Conference*, 1–8. https://doi.org/10.1109/AFRCON.2007.4401495

Siggis. (2019). De Watergroep. Retrieved November 30, 2019, from https://www.siggis.be/en/portfolio/de-watergroep/

Sluter, C. R., van Elzakker, C. P. J. M. & Ivánová, I. (2017). Requirements Elicitation for Geo-information Solutions. *Cartographic Journal*, *54*(1), 77–90.

Snel, K. A. W., Witte, P. A., Hartmann, T., Geertman, S. C. M., Snel, K. A. W., Witte, P. A., … Geertman, S. C. M. (2019). More than a one-size-fits-all approach – tailoring flood risk communication to plural residents' perspectives. *Water International*, *44*(5), 554–570.

Sweco AB. (2020). The most approachable and committed partner with recognised expertise. Retrieved May 10, 2020, from https://www.sweco.nl/en/about-us/

Sweco Nederland B.V. (2020). GeoWeb. Alle data over onze leef- en werkomgeving overzichtelijk in kaart. Retrieved May 10, 2020, from GeoWeb website: https://geoweb.software/over-geoweb/

Tolpekin, V. & Stein, A. (2013). *The core of GIScience: a process-based approach*. Enschede, The Netherlands: Faculty of Geo-Information Science and Earth Observation (ITC).

Unrau, R. & Kray, C. (2019). Usability evaluation for geographic information systems: a systematic literature review. *International Journal of Geographical Information Science*, *33*(4), 645–665.

Urban Europe. (2017). FLOODLABEL project description. Retrieved September 4, 2019, from https://jpi-urbaneurope.eu/project/floodlabel/

Veenendaal, B., Brovelli, M. A. & Li, S. (2017). Review of web mapping: Eras, trends and directions. *ISPRS International Journal of Geo-Information*, *6*(10), 1–31.

Wagner, S. (2013). *Software product quality control*. Berlin, Heidelberg: Springer.

Yan, M., Xia, X., Zhang, X., Xu, L., Yang, D. & Li, S. (2019). Software quality assessment model: a systematic mapping study. *Science China Information Sciences*, *62*(9), 1–18.

8. Appendices

8.1 ISO product quality and quality in use model characteristic definitions

Table 8.1 Product Quality model characteristics for the primary user. Source: ISO/IEC 25010 (2011)

Product quality model characteristic	***Definition ISO/IEC 25010 (2011)***
Functional suitability	degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions
Performance efficiency	performance relative to the amount of resources used under stated conditions
Usability	degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in
Reliability	degree to which a system, product or component performs specified functions under specified conditions for a specified period of time
Security	degree to which a product or system protects information and data so that persons or other products or systems have the degree of data access

Table 8.2 Relevant characteristics of the Quality in use model. Source: ISO/IEC 25010 (2011)

Quality in use characteristic	***Definition ISO/IEC 25010 (2011)***
Effectiveness	accuracy and completeness with which users achieve specified goals
Efficiency	resources expended in relation to the accuracy and completeness with which users achieve goals
Satisfaction	degree to which user needs are satisfied when a product or system is used in a specified context of use
Freedom from risk	degree to which a product or system mitigates the potential risk to economic status, human life, health, or the environment
Context coverage	degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in both specified contexts of use and in contexts beyond those initially explicitly identified

8.2 ISO product quality and quality in use model sub characteristics and definitions

Table 8.3 Relevant characteristics and their sub characteristics. Source: ISO/IEC 25010 (2011).

Characteristic	***Sub characteristics***
Functional suitability	Functional completeness; Functional correctness; Functional appropriateness
Performance efficiency	Time behaviour; Resource utilization; Capacity
Usability	Appropriateness recognizability; Learnability; Operability; User error protection; User interface aesthetics; Accessibility
Reliability	Maturity; Availability; Fault tolerance; Recoverability
Security	Confidentiality; Integrity; Non-repudiation; Accountability; Authenticity
Effectiveness	-
Efficiency	-
Satisfaction	Usefulness; Trust; Pleasure; Comfort
Freedom from risk	Economic risk mitigation; Healthy and safety risk mitigation; Environmental risk mitigation
Context coverage	Context completeness; Flexibility

Table 8.4 Quality model sub characteristic definitions. Source: ISO/IEC 25010 (2011)

Sub characteristic	***Definition***
Functional completeness	degree to which the set of functions covers all the specified tasks and user objectives
Functional correctness	degree to which a product or system provides the correct results with the needed degree of precision
Functional appropriateness	degree to which the functions facilitate the accomplishment of specified tasks and objectives
Time behaviour	degree to which the response and processing times and throughput rates of a product or system, when performing its functions, meet requirements
Resource utilization	degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements
Capacity	degree to which the maximum limits of a product or system parameter meet requirements
Appropriateness recognizability	degree to which users can recognize whether a product or system is appropriate for their needs
Learnability	degree to which a product or system can be used by specified users to achieve specified goals of learning to use the product or system with effectiveness, efficiency, freedom from risk and satisfaction in a specified context of use
Operability	degree to which a product or system has attributes that make it easy to operate and control
User error protection	degree to which a system protects users against making errors
User interface aesthetics	degree to which a user interface enables pleasing and satisfying interaction for the user

Accessibility	degree to which a product or system can be used by people with the widest range of characteristics and capabilities to achieve a specified goal in a specified context of use
Maturity	degree to which a system, product or component meets needs for reliability under normal operation
Availability	degree to which a system, product or component is operational and accessible when required for use
Fault tolerance	degree to which a system, product or component operates as intended despite the presence of hardware or software faults
Recoverability	degree to which, in the event of an interruption or a failure, a product or system can recover the data directly affected and re-establish the desired state of the system
Confidentiality	degree to which a product or system ensures that data are accessible only to those authorized to have access
Integrity	degree to which a system, product or component prevents unauthorized access to, or modification of, computer programs or data
Non-repudiation	degree to which actions or events can be proven to have taken place, so that the events or actions cannot be repudiated late
Accountability	degree to which the actions of an entity can be traced uniquely to the entity
Authenticity	degree to which the identity of a subject or resource can be proved to be the one claimed
Usefulness	degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use
Trust	degree to which a user or other stakeholder has confidence that a product or system will behave as intended
Pleasure	degree to which a user obtains pleasure from fulfilling their personal needs
Comfort	degree to which the user is satisfied with physical comfort
Economic risk mitigation	degree to which a product or system mitigates the potential risk to financial status, efficient operation, commercial property, reputation or other resources in the intended contexts of use
Healthy and safety risk mitigation	degree to which a product or system mitigates the potential risk to people in the intended contexts of use
Environmental risk mitigation	degree to which a product or system mitigates the potential risk to property or the environment in the intended contexts of use
Context completeness	degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in all the specified contexts of use
Flexibility	degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in contexts beyond those initially specified in the requirements

8.3 Examples of SUS and PSSUQ

System Usability Scale

© Digital Equipment Corporation, 1986.

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Figure 8.1 System Usability Scale. Source: Brooke (1996)

The Post-Study System Usability Questionnaire Items

The first item illustrates the item format. The remaining items show only the item text to conserve space. Each item also has an area for comments (not shown).

1. Overall, I am satisfied with how easy it is to use this system.

STRONGLY AGREE						STRONGLY DISAGREE	
1	2	3	4	5	6	7	N/A

2. It was simple to use this system.
3. I could effectively complete the tasks and scenarios using this system.
4. I was able to complete the tasks and scenarios quickly using this system.
5. I was able to efficiently complete the tasks and scenarios using this system.
6. I felt comfortable using this system.
7. It was easy to learn to use this system.
8. I believe I could become productive quickly using this system.
9. The system gave error messages that clearly told me how to fix problems.
10. Whenever I made a mistake using the system, I could recover easily and quickly.
11. The information (such as on-line help, on-screen messages and other documentation) provided with this system was clear.
12. It was easy to find the information I needed.
13. The information provided for the system was easy to understand.
14. The information was effective in helping me complete the tasks and scenarios.
15. The organization of information on the system screens was clear.

Note: The “interface” includes those items that you use to interact with the system. For example, some components of the interface are the keyboard, the mouse, the microphone, and the screens (including their use of graphics and language).

16. The interface of this system was pleasant.
17. I liked using the interface of this system.
18. This system has all the functions and capabilities I expect it to have.
19. Overall, I am satisfied with this system.

Figure 8.2 PSSUQ. Source: Lewis (2002)

8.4 Floodlabel user goal labels

Table 8.5 Descriptive statistics of Floodlabel user goal labels

Floodlabel user goal label	***Frequency***	***Percentage***
Requested in questionnaire	38	32.76
Flood risks of house	19	16.38
Interesting	15	12.93
Requested by student	14	12.07
To be informed	10	8.62
Advice on damage prevention & steps of action by flooding	7	6.03
Potential damage of house by flooding	6	5.17
No goal	3	2.59
Other	3	2.59
Website does not have added value for me	1	0.86
Total	***116***	***100.00***

8.5 Label values GeoWeb questionnaire results

Table 8.6 Frequency table of GeoWeb respondents per organization

Organization	***Frequency***
Sweco	2
Municipality of Vught	1
Water board of Limburg	12
Environmental service of Midden-Holland	2
Municipalities of Tilburg and Breda	1
Municipality of Breda	1
Municipality of ‘s-Hertogenbosch	17
Province of Zuid-Holland	1
Total	**37**

User goals GeoWeb Respondents labels:

Type gebruik

1 = Databeheer en distributie

2 = Locatie informatie inwinnen

3 = Basisregistratie, combineren van geografische gegevens

4 = Controle ingevoerde gegevens

5 = Visualiseren bestaande basisdata

6 = Analyse en onderzoek

7 = Streetsmart extensie

8 = Foto’s bekijken

Inhoudelijk

A = Monitoringsnetwerk

B = Omgevingsmeldingen verwerken

C = Inventarisatie omgevingsfactoren beekherstelprojecten.

D = Belastingheffing

E = Kadastrale gegevens en adressen

F = Gegevens zoeken voor bouwplannen

G = Bestemmingsplannen

H = Cyclorama foto’s bekijken om perceeleigenaar te achterhalen

8.6 Written out statements translated to English

Table 8.7 Written out statements translated to English

Number	***Statement***
1	I am able to use the map by changing themes and scrolling, zooming and panning.
2	There are no less or more than necessary functionalities available, to perform handlings with georeferenced data (e.g. (spatial) queries or overlay, neighbourhood and network analysis.)
3	I am satisfied by the product's response rate while executing my actions.
4	I am satisfied by the supported data types for importing and exporting georeferenced data.
5	The product offers sufficient user guidance for (learning) to use the product.
6	The product's interface positively contributes to reaching my user goals.
7	The interactive map's interface positively contributes to reaching my user goals.
8	The georeferenced data on the map is clear and interpretable.
9	Using for example a legend, title, text or other georeferenced data descriptions, I am able to understand its meaning.
10	In my opinion the georeferenced data is sufficiently up to date for my user goals.
11	The georeferenced data on the map contributes to reaching my user goals.
12	I am experiencing this product as user friendly.
13	The product is always reliable, working, available and recovers when I perform my user tasks.
14	In my opinion the product sufficiently considers data security and privacy.
15	Based on my experiences with the product, I think it is sufficient for my user goals.
16	Overall, I think that the quality of the product is:

8.7 Explanation comments for reliability of questionnaires

Table 8.8 Floodlabel comments on a no answer option

Statement	***ID***	***Rating***	***Comment***
5	37	4	Not investigated
10	103	4	No opinion
12	13	3	I cannot rate this. Restarting the web site leads to the need of entering my information again.
12	44	4	Did not use it
12	53	1	I am not able to rate this
13	13	3	Difficult to rate, again
13	44	4	No opinion
13	93	2	I do not know this
15	38	1	No idea

Table 8.9 GeoWeb comments on a no answer option

Statement	***ID***	***Rating***	***Comment***
2	20	1	No idea. I am far from familiar with all functionalities …
4	5	4	Rated ‘neutral’, since I do not use these functions
4	20	1	N.a.
5	20	1	N.a., never had a look at it
13	14	3	I cannot rate the security components behind the system
13	19	5	No idea
13	26	4	No opinion

8.8 Explanation comments validity of questionnaires

Table 8.10 Statement interpretation within explanation comments Floodlabel

Statement	Comments (n)	Number of comments that are interpreted…		
		incorrectly	partially correct	correctly
1	16	2	2	12
3	5	2	0	3
5	23	2	8	13
6	20	1	0	19
7	15	5	0	10
8	5	1	0	4
9	14	0	0	14
10	7	4	2	1
11	7	1	1	5
12	14	1	0	10
13	24	8	2	4
14	11	0	2	22
15	9	1	0	8
16	7	3	0	4

Table 8.11 Statement interpretation within explanation comments GeoWeb

Statement	Comments (n)	Number of comments that are interpreted…		
		incorrectly	partially correct	correctly
1	6	2	1	3
2	9	2	2	5
3	12	0	0	12
4	6	1	1	4
5	7	0	0	7
6	10	2	1	7
7	6	1	1	4
8	4	0	1	3
9	3	0	0	3
10	9	0	2	7
11	2	1	0	1
12	7	0	0	9
13	8	0	2	5
14	9	0	6	2
15	4	0	0	4
16	0	2	1	3