Communicating uncertainty

in spatial data

Exploring the usefulness in municipal information chains

DISSERTATION

Nico de Graaff

Submitted in part fulfilment of the requirements for the degree of
Master of Science in Geographical Information Systems (UNIGIS)
Faculty of Earth and Life Sciences
VU University Amsterdam
The Netherlands
June 2013

This page intentionally left blank

“ As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality. ”

- Albert Einstein -

This page intentionally left blank

Abstract

In spatially managed processes, uncertainty is an inherent property of spatial data. Uncertainty in spatial data might lead to misunderstandings, disagreements, or poorly grounded decisions. In municipal organisations uncertainty in spatial data not always seem to be recognized, information about uncertainty seem to be supply driven, and there are probably gaps in communication about uncertainty. This research aims to investigate the usefulness of information about uncertainty in spatial data of key Geo-registrations in municipal chain processes. A literature study, a survey, and in-depth interviews are conducted to explore which factors influence usefulness of information about uncertainty in spatial data in municipal information chains. The technology acceptance model (TAM) is used to model and measure the relationships between possible factors and usefulness of information about uncertainty. The results of the survey are analysed using rank order correlation and multi regression analysis. The results of this research revealed that more than 74% of the respondents perceived information about uncertainty as useful. Information needs, regular tasks, and type of visualization seem to be the most influential factors which affect usefulness of information about uncertainty. Experience in spatial data analysis and data creation seem to influence the perceived ease of use. The in-depth interviews revealed that in a municipal information chain uncertainty awareness related to problem recognition are influential factors. The way how users handle uncertain data is related to the gap between required information about uncertainty and provided information about uncertainty. To reduce the gap in communicating about uncertainty in spatial data, data providers and data users should increase their knowledge of concepts of uncertainty and work towards a shared perception of spatial data and its inherent uncertainty.

Keywords: Uncertainty, Spatial data quality, key Geo-registrations

This page intentionally left blank

Samenvatting

In ruimtelijk beheerde processen is onzekerheid een inherente eigenschap van ruimtelijke gegevens. Onzekerheid in ruimtelijke gegevens kan leiden tot misinterpretaties, verschillen van inzicht of slecht onderbouwde beslissingen. In gemeentelijke organisaties lijkt onzekerheid in ruimtelijke data niet altijd herkend te worden, informatie over onzekerheid lijkt aanbod gedreven, en er zijn mogelijk hiaten in de communicatie over onzekerheid in ruimtelijke gegevens. Dit onderzoek heeft als doel de bruikbaarheid van de informatie over de onzekerheid in ruimtelijke gegevens van ruimtelijke Basisregistraties in gemeentelijke ketenprocessen te onderzoeken. Een literatuurstudie, een enquête en diepte-interviews zijn uitgevoerd om te onderzoeken welke factoren van invloed zijn op bruikbaarheid van informatie over de onzekerheid in ruimtelijke gegevens binnen de gemeentelijke ketenprocessen.
Het technologie acceptatie model (TAM) wordt gebruikt voor het modelleren en het meten van mogelijke relaties tussen invloedrijke factoren en bruikbaarheid van informatie over onzekerheid. De resultaten van de enquête zijn geanalyseerd met behulp rangordecorrelatie en meervoudige regressieanalyse. Uit de resultaten van dit onderzoek bleek dat meer dan 74% van de respondenten informatie over onzekerheid als nuttig ervaren. De verschillende informatiebehoefte over onzekerheid in ruimtelijke gegevens, reguliere taken, en het type van visualisatie lijken de meest invloedrijke factoren van de bruikbaarheid van informatie over de onzekerheid in ruimtelijke gegevens. Ervaring in ruimtelijke analyse van gegevens en het opbouwen van ruimtelijke gegevens lijkt het ervaren gebruiksgemak te beïnvloeden. Uit de diepte-interviews is gebleken dat in een gemeentelijke informatieketen bewustwording van onzekerheid in ruimtelijke gegevens in relatie tot probleem herkenning invloedrijke factoren kunnen zijn. De manier waarop gebruikers omgaan met onzekere informatie heeft betrekking op de kloof tussen benodigde informatie over onzekerheid en verstrekte informatie over onzekerheid. Om de kloof in de communicatie over onzekerheid in ruimtelijke gegevens tussen data leveranciers en gebruikers van die gegevens te verkleinen zou hun kennis van concepten van onzekerheid vergroot moeten worden om te streven naar een gedeelde perceptie van gegevens en de inherente onzekerheid in deze gegevens.

Trefwoorden: Onzekerheid, gegevenskwaliteit, Basisregistraties

This page intentionally left blank

Table of contents

List of figures

List of tables

This page intentionally left blank

Disclaimer

The results presented in this thesis are based on my own research at the Faculty of Earth and Life Sciences of the VU University Amsterdam. All assistance received from other individuals and organisations has been acknowledged and full reference is made to all published and unpublished sources. This thesis has not been submitted previously for a degree at any institution.

Signed:
Amsterdam, 18-10-2013

ing. N.J. De Graaff

This page intentionally left blank

Acknowledgements

First and foremost I like to express my gratitude and love to my wife Saskia and our children Renée and Robin, for the support fulfilling my ambition to increase my knowledge in the sciences of geographic information science.

My appreciation and gratitude are extended to my advisor Drs. Ronnie Lassche who supervised me through this research process. I also wish to acknowledge and thank the UNIGIS program director, Professor Henk Scholten, and Dr. Jasper Dekkers who guided me through the process of proposing my research.

I also like to express my gratitude to my managers Ad van der Meer and Dick Krijtenburg of the Municipality of Amsterdam who gave me the opportunity to fulfilling my ambition to conduct this study, and my co-workers for their inexhaustible feedback.

Finally, I would thank all the respondents who have invested their sparse time to participate in this research. None of this research would have been possible without the efforts of the respondents who contributed to the survey and the interviews.

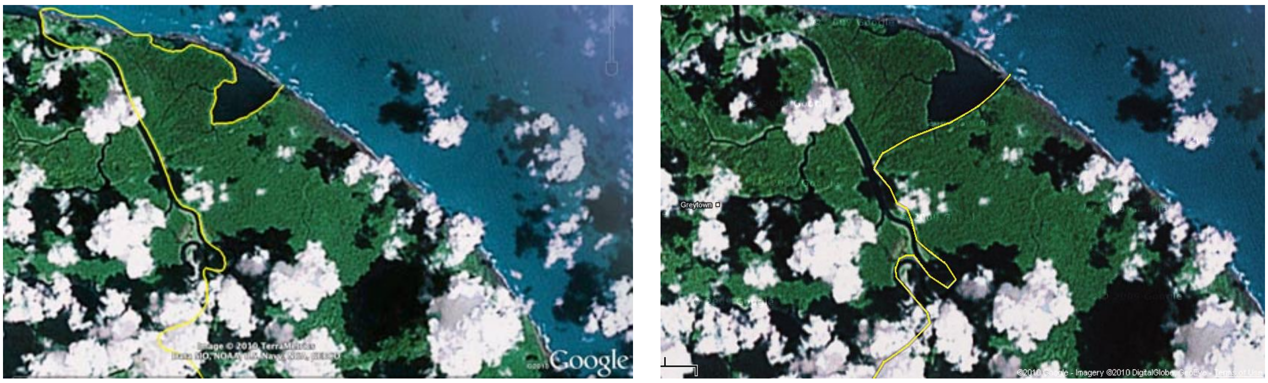
This page intentionally left blank

1. Introduction

The last decade there is an increasing growth of supply and application of geospatial data (Scholten et al. 2009). This growth implies the intensifying of the interaction between suppliers and users of geospatial data. Uncertainty in spatial data is inherent caused by a complex world, imperfect human perception, and imperfection in measurements and representation (Longley et al. 2011b). These different appearances of uncertainty affect the use of the data, which ranges from visualisation to simple queries or complex analyses (Rodolphe Devillers et al. 2010). From the scientific domain much research is performed concerning principles of uncertainty, propagation, and visualizing uncertainty in spatial data (MacEachren et al. 2005). In practice, there is limited attention from users for the consequences of use of spatial data with uncertainty. The challenge is to draw the attention of users of spatial data and raise their awareness that there are potential consequences of using data with uncertainty (Rodolphe Devillers et al. 2010). These potential consequences are that data sets do not fit their needs or requirements. This study aims to contribute to reduce the apparent gap between science and practice of the perception of usability of information about uncertainty in spatial data.

1.1. Exploring the problem field

Inaccurate, poorly defined, or even uncertain reliability of spatial data might lead to major disagreements. In the beginning of November 2010, a border dispute between Nicaragua and Costa Rica lead to an international conflict. The cause of the dispute was the erroneous depiction of the border located at the San Juan River near the Caribbean coast which was published in Google Maps (Moses 2010). Google promised to correct the 'inaccuracy' as soon as possible, and reconstructed the borderline a few days after the error was reported.
In Figure 1 the erroneous location, and the corrected location are illustrated.



A. Erroneous borderline **B. Corrected borderline**

Figure 1.01. The borderline dispute between Nicaragua and Costa Rica.

This event raises interesting questions concerning uncertainty in spatial data, such as; how accurate is this border defined, what is the basis for its demarcation, what is the primary source of this geographic data?
Google Maps left all these questions unanswered, but corrected the data immediately after this incident. This example illustrates the importance of providing information about uncertainty in spatial data. This research focuses on the use of spatial data in municipal processes. To place this research in the intended context the problem field will be further explored in this context.

1.2. Problem field

Two cases are used to illustrate the impact of uncertainty in spatial data in a municipal context. Case 1 illustrates uncertainty in the meaning of data, Case 2 illustrates the effects of positional uncertainty.

Case 1. The National Museum in Amsterdam

On Thursday November the 3th in 2011 an article appeared on the local media AT5.
The article reported that according the formal key registration the size of the internal area of the National Museum in Amsterdam was 1 square meter. A few directors and officials went to get started with this information and calculated how much incoming tax the municipal will miss due to this 'error'. Further consequences were a political agitation and a bad turn to the citizens.

ZAKELIJK

donderdag 03 november 2011 09:05

'Rijksmuseum is één vierkante meter'

Volgens het basisgebouwenregister van de gemeente Amsterdam is het Rijksmuseum niet groter dan één vierkante meter. Een foutje, maar wel een dure.

Foto: rogier righART

Figure 1.02. The AT5 article of National Museum in Amsterdam.

Two issues of this case are interesting:
First: the users of the data (the journalist and some municipal officials) were not aware of the fact that the value of 1 square meter is not a factual value, but a default value.
When this value 1 is assigned to a building according to the accompanying national data catalogue it means “the real value is unknown”.
Second: the default value did not refer to the museum itself, but to an auxiliary building.
The assigned value for the area of the main building was 38149 square meters.
What possibly went wrong? The users were not properly informed about the meaning of the data. The national data portal for information about addresses and buildings did not provide sufficient information about the meaning of the data (Kadaster 2012).

Case 2. Cadastral parcels and buildings
Widely used data in a municipal context are the relations between cadastral parcels and dwellings. Those relations provide insight in the legal status of the involved legal entities and the real estate. Purposes of use are legal transactions or administration of real estate.
One of the methods to determine the relationship between cadastral parcels and dwellings is to use the spatial relationship (e.g. spatial intersection) between the geometry of the cadastral parcel and the building. The data quality standards of both entities differ[1] which results in remarkable spatial differences. Figure 1.03 illustrates in blue the boundaries of the cadastral parcels, and in red the buildings. The picture shows that there are differences of 1 meter or more between the boundaries of both entities, despite the fact that those boundaries in most case should be equal.

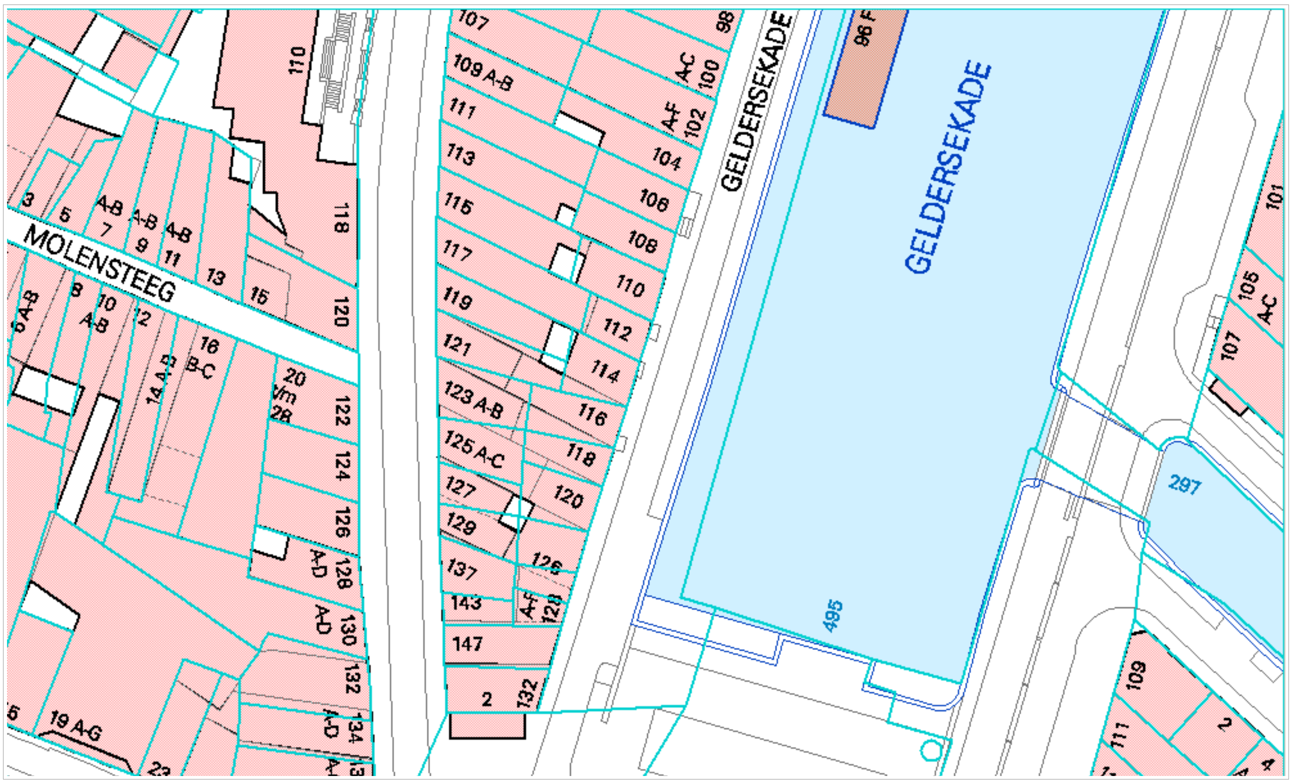


Figure 1.03. Cadastral parcels and buildings.

The different positional inaccuracies lead to topological inconsistency, which affects the quality of the spatially derived relationships. It is likely that as this is a basis for legal actions it also affects the quality of the legal act itself.

The two cases showed problems from a different perspective and with different consequences. What these two cases have in common is that in both case the information about uncertainty in spatial data is likely supply-oriented or inadequate. According to Devillers et al. (2010) raising the awareness of users for spatial data characteristics could increase the interest for spatial data quality. The implementation of the system of Key registrations (in Dutch: 'Basisregistraties') in the Netherlands performed by the national government is predominantly supply driven (BZK 2011). Due to the nature of this development, cognitive, normative, or technological gaps might be arise in the communication about uncertainty in spatial data (Grira et al. 2010). As a result of this possible communication gap limited user awareness and limited knowledge of uncertainty in spatial data might lead to misinterpretation. The link with this research is to explore what factors affect the usefulness of information about uncertainty in spatial data.

1 In this use case data of the municipality of Amsterdam is used. The data quality standard for positional accuracy of buildings has a maximum standard deviation of 5 cm. For cadastral data this varies between a few centimetres to meters.

1.3. Scope of the research

The scope of this research is framed by the use of key Geo-registrations in a municipal information chain. In a municipal organisation several processes use key Geo-registrations which are centrally administered by a supportive municipal division and used by many different users for different purposes[2]. These predominantly spatially managed data sets are maintained and used in processes which are cross organizational, which may be defined as 'chain processes'. As illustrated in Figure 1.04 the flow of data passes through various links in the chain. The green framed area in Figure 1.04 demarcates the focus of this research.

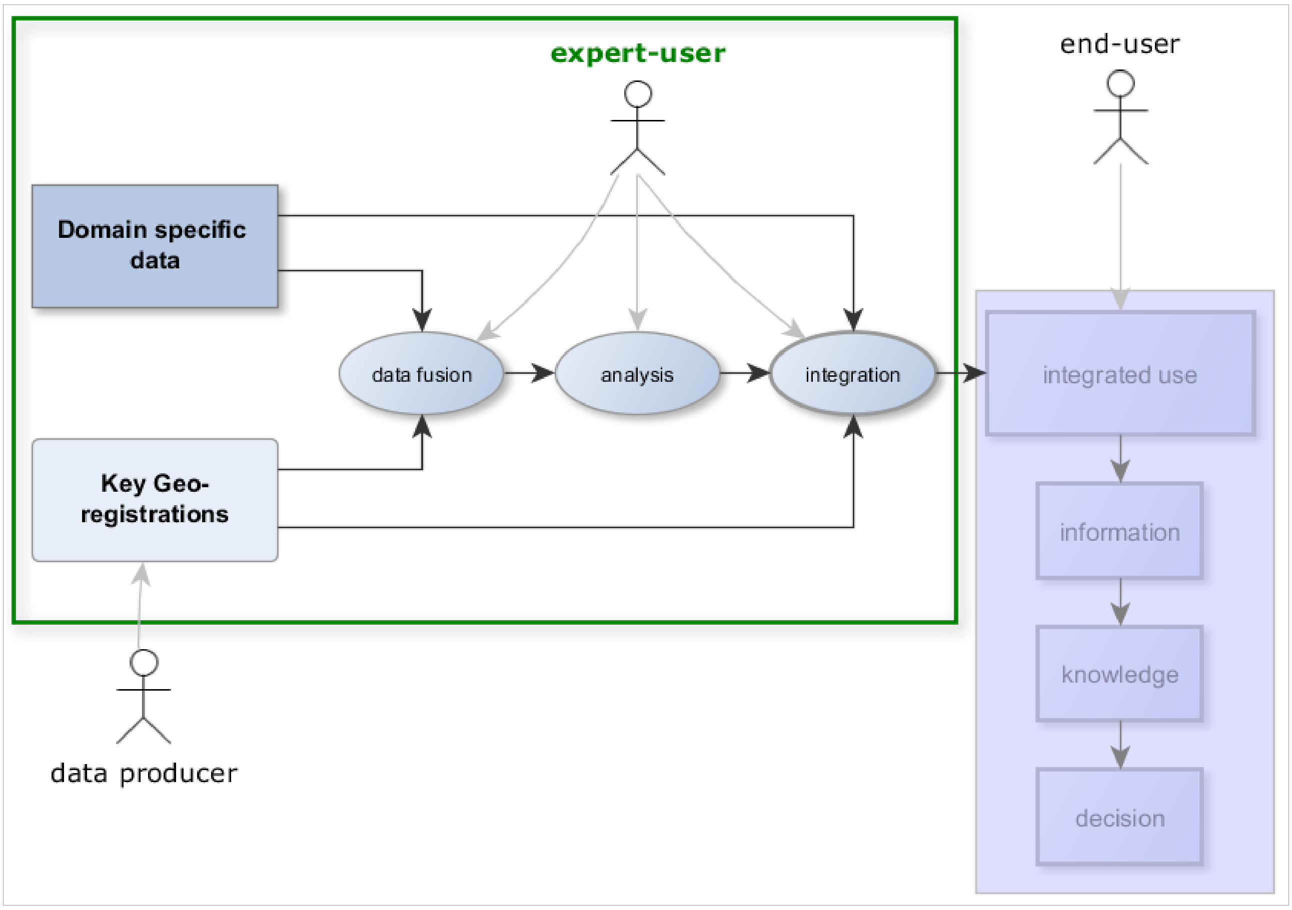


Figure 1.04. Municipal information chain.

In many municipal processes key Geo-registrations are used in combination with domain specific data applicable for a specific task. Expert users fulfil a key role to provide end users with the appropriate data or information in the form of maps, applications or reports. They combine different data sources, perform analysis, and process the data into a application or report. One of the tasks of an expert user is to assess the characteristics of the base data and provide the end user information about the usability and limitations of the integrated data.
To support this task the expert user might want to know how uncertainty is present in the key Geo-registrations. The demarcation of this research focuses around this part of the task of the expert user.

2 Based on the principle "acquire once, use manifold" (VROM 2008)

1.4. Thesis structure

The previous section of this thesis provided an introductory text and described the problem field. The next section defines the research objective, chapter two describes the research methodology. To establish the theoretical framework the literature study is discussed in the chapters three to five. Chapter three discusses different taxonomies and conceptualization of uncertainty in spatial data. How uncertainty in spatial data might be communicated is discussed in chapter four. Chapter five describes the Dutch system of key registrations and its use in a municipal information chain. The shift from the theoretical perspective to the empirical perspective is described in chapter six. The research results are broadly reported and explained in chapter seven. The synthesis and conclusions are stated in chapter eight.

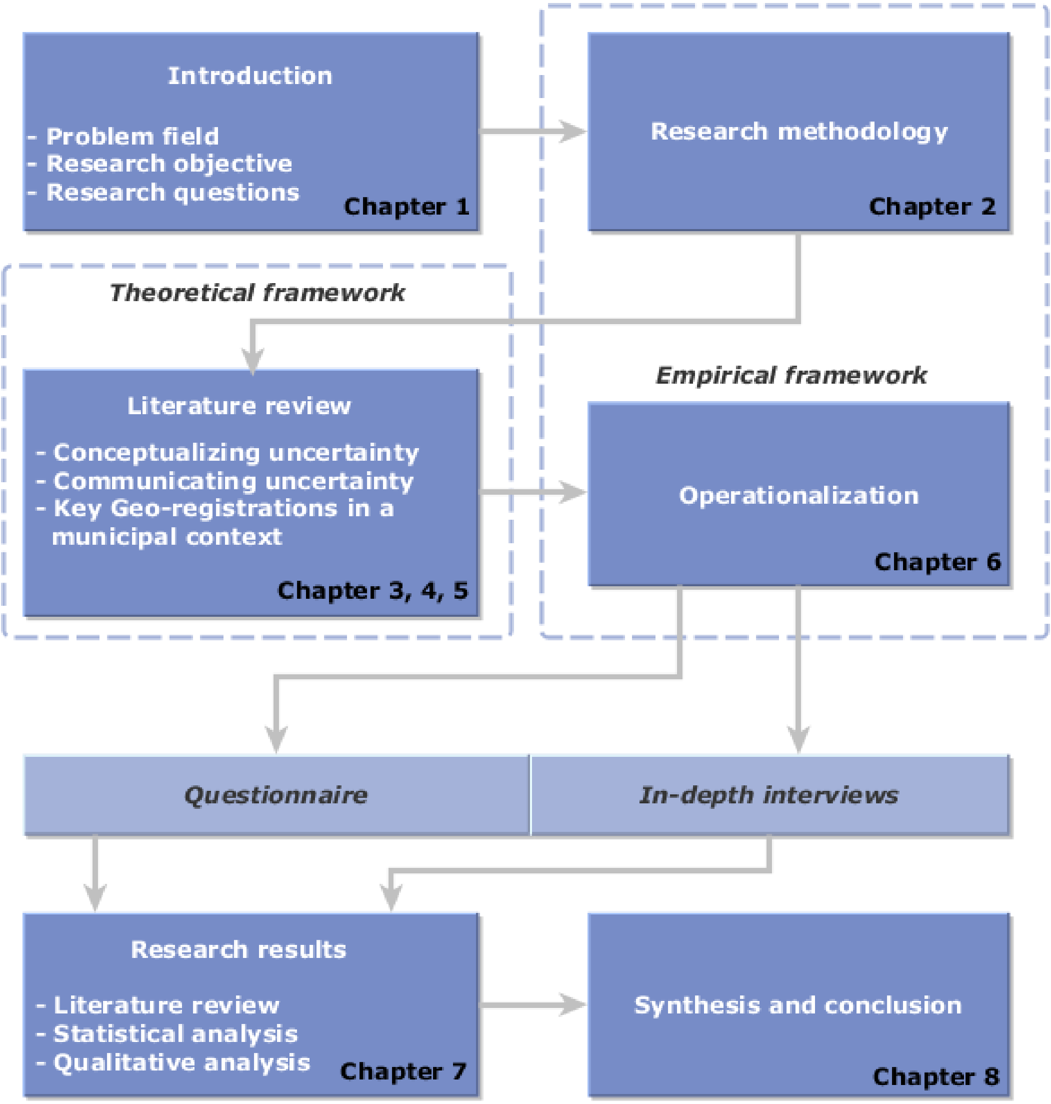


Figure 1.05. Thesis structure.

1.5. Research objective

This research focuses on investigating the usefulness of communication about uncertainty in key Geo-registrations in a municipal information chain.
Communication about uncertainty in key Geo-registrations in the context of this research is intended as 'providing information about uncertainty in spatial data as a part of the key Geo-registrations' used in a municipal context.

Therefore the objective of this research is:

"Explore how uncertainty appears in spatial data and the usefulness of communicating information about uncertainty in key Geo-registrations used in municipal chain processes."

The specific research aims are:

1. Explore and define concepts of uncertainty in spatial data.
2. Explore how uncertainty in spatial data appears in a municipal information chain.
3. Explore how uncertainty in spatial data can be communicated.
4. Explore the information needs about uncertainty in spatial data.
5. Explore the gap between information needs and information provided about uncertainty in spatial data.
6. Investigate the usefulness of information about uncertainty in spatial data.

To achieve these research aims a main research question is formulated, and supporting research questions are formulated.

The main research question is formulated as follows:

> Which factors influence the usefulness of information about uncertainty in spatial data of key Geo-registrations in municipal chain processes?

To work towards the answer of the research question the following sub questions are formulated:

1. How is uncertainty in spatial data defined?
2. How are uncertainty and spatial data quality related?
3. How does uncertainty in spatial data appear in a municipal information chain?
4. How can uncertainty in spatial data be communicated?
5. What are the information needs about uncertainty in spatial data?
6. How do users perceive usefulness of information about uncertainty in spatial data?
7. Which factors affect the usefulness of information about uncertainty in spatial data?

The purpose of sub questions 1 and 2 is to explore definitions which will be used in this research to achieve an unambiguous understanding of uncertainty and spatial data quality. The research problem is explored in a municipal context by answering sub question 3. Sub question 4 extends the theoretical foundation by investigating the different methods of visualizing uncertainty in spatial data. Sub questions 5, 6 and 7 are asked to determine *what* information needs are, *why* the users consider this information as useful, and what factors affect the usefulness of information about uncertainty in spatial data.

2. Research methodology

To explore the usefulness of information about uncertainty in spatial data a thorough understanding of the concepts of uncertainty, visualization methods, and user perceptions is required. To obtain answers on the research questions and achieve the research aims the mixed method approach is chosen. Both quantitative and qualitative methodological approaches are required to make generalizations and achieve a deeper understanding of the research problem (O'Leary 2010). The methods applied in this research consist of literature review, a survey using a questionnaire, and in-depth interviews (see Figure 2.01). The literature review explores concepts of uncertainty in spatial data, conducted research of communication of uncertainty in spatial data and contextual concepts of the use of key Geo-registrations in a municipal context. The output of the literature review is used as input for the survey and in-depth interviews. Preliminary results of the survey are used as input for the in-depth interviews to put more focus on the questions to be asked. The results of both studies are analysed and serve as a basis for discussion and conclusion. The Three aforementioned methods explained more in detail in the next sections of this chapter.

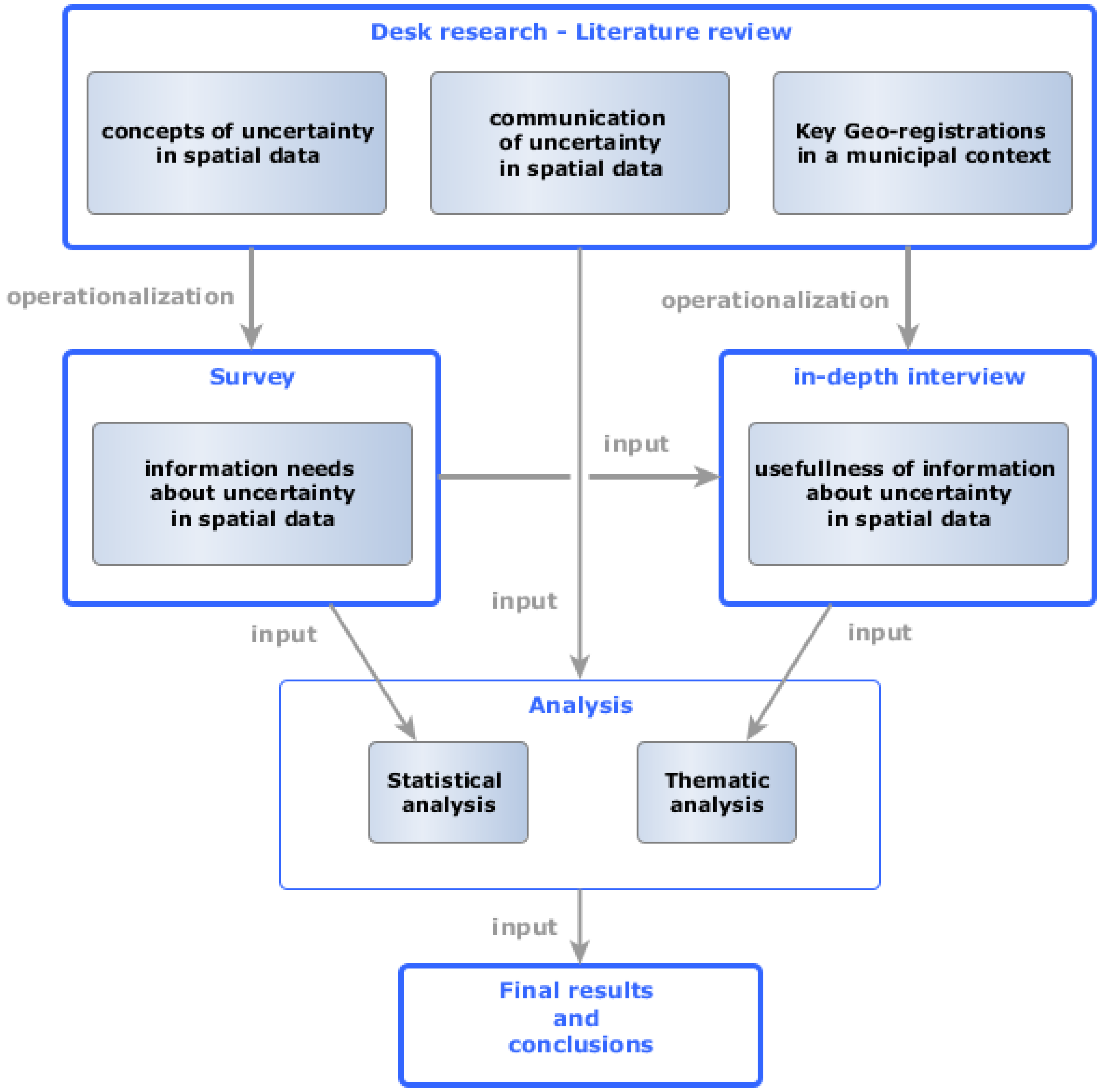


Figure 2.01. Research methodology.

2.1. Coherence of the research aims, questions, and methods

The methods applied in this research consist of desk research, a survey using a questionnaire, and in-depth interviews. Figure 2.02 illustrates the relationship between the research aims, research questions, and the research methods. Each research aim is related to one or more research questions, and each research question is related to a research method.

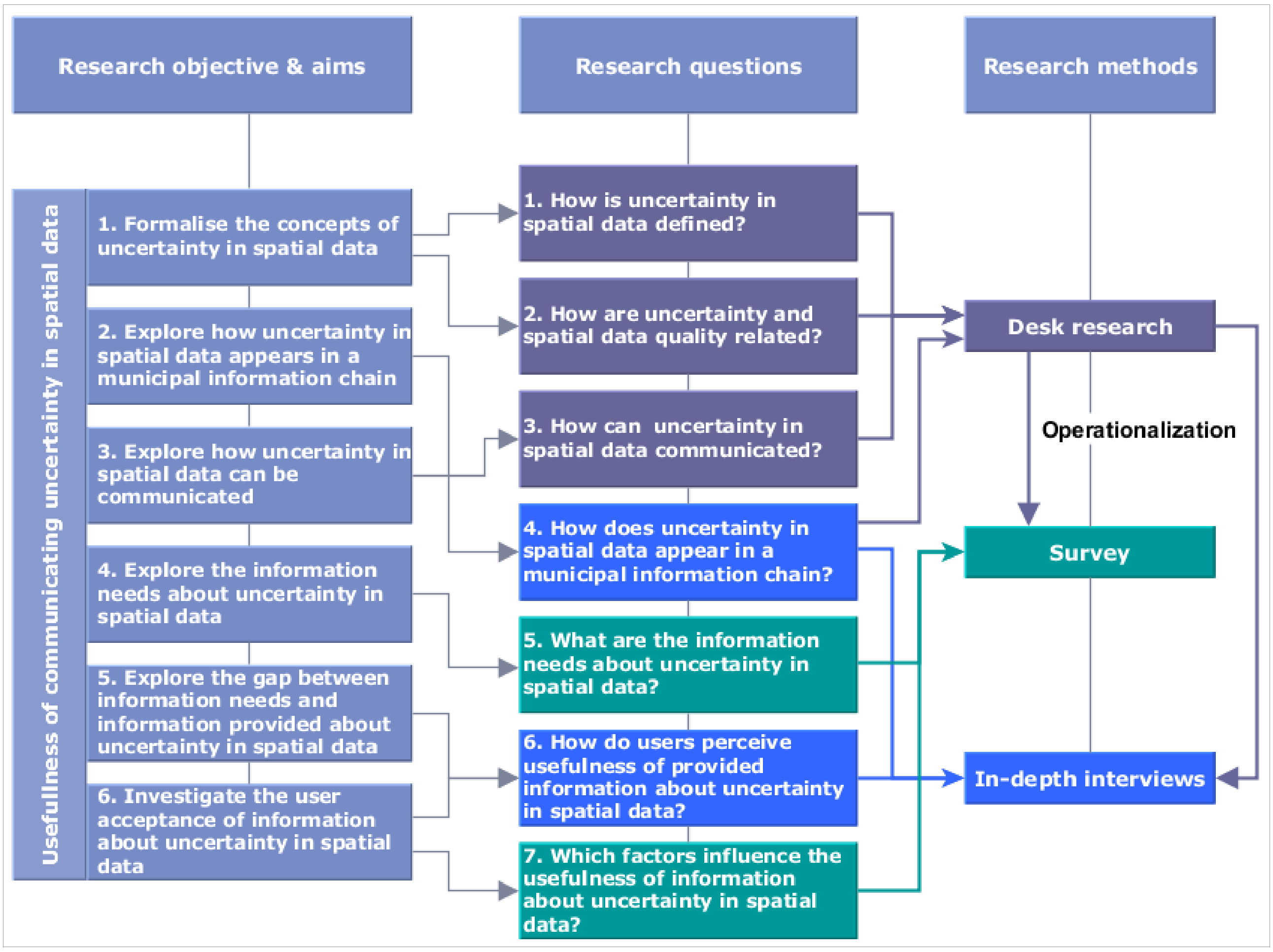


Figure 2.02. Research methodology.

As illustrated in Figure 2.02 three methods to obtain the answers on the research questions. The desk research is used to establish a theoretical framework. To obtain quantitative data and perform quantitative analysis the survey is applied, The in-depth interviews are used to gain insight into qualitative aspects of the usefulness of information about uncertainty. The outcomes of each method are analysed separately using suitable analysis methods. Results of these analysis are discussed in the chapter six which establishes the discussion and conclusion.

2.2. Desk research

The desk research aims to establish a theoretical framework regarding definitions and concepts, and discuss previously conducted research using a literature review.
The supporting research questions must be answered through this desk research:

1. How is uncertainty in spatial data defined?
2. How are uncertainty and spatial data quality related?
3. How can uncertainty in spatial data be communicated?
4. Explore the use of key Geo-registrations in a municipal context

The first research question concerns the definition of uncertainty in spatial data, this part of the desk research establishes definitions and concepts of uncertainty in spatial data. Research question 2 concerns the relationship between uncertainty in spatial data and spatial data quality. The approach of this part of the desk research is to investigate theoretical relationships between uncertainty in spatial data and spatial data quality by conducting a literature review. The third research question is about communication of uncertainty in spatial data. The aim of this part of the desk research is to explore previously conducted research with a focus on topical aspects and methodological aspects. The elaboration of the desk research and literature review is established in chapter three, four, and five.

2.3. Survey

The objective of the questionnaire is to explore the information needs about uncertainty in spatial data, and to investigate which factors influence the usefulness information about uncertianty. The survey is implemented through questionnaire using computer-assisted web interviewing (CAWI). A survey using an electronic questionnaire has the ability to reach many people in a relatively short period (Babbie 2001; Bethlehem 2009), and therefore contributes to the validity of the method. The survey process and survey design are explained in this section.

2.3.1. The survey process

The survey process performed in this research consists of various steps which are illustrated in Figure 2.03. This paragraph discusses these steps in general, the following sections describe those steps more in detail. The first step in the process is the survey design, which also specifies the objectives of the survey. The second step is data collection, the applied method in this survey is Computer-assisted web interviewing (CAWI). The third step is the data editing with the main purpose to transform the data into a suitable form for perform statistical operations. The fourth step in the survey process is the non response correction.
Non response might occur when elements or questions of the questionnaire do not provide sufficient information to meet the objective of the survey (Bethlehem 2009).
Step five of the survey contains an deductive analysis of the results by testing hypotheses. Finally, the last step of the survey process is publish the results including the outcomes of the hypothesis tests.

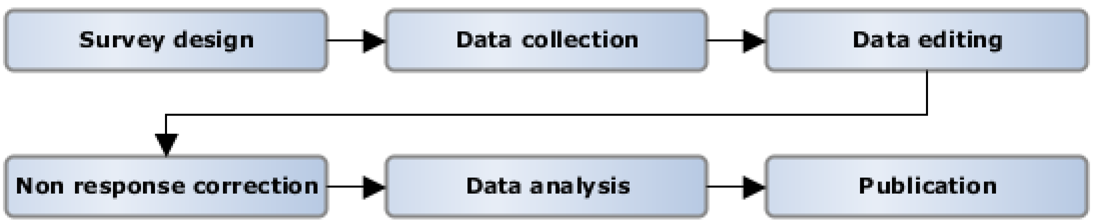


Figure 2.03. The survey process (Bethlehem 2009).

2.3.2. Survey design

To achieve the survey design several design decisions have been made. This paragraph describes of which steps the survey design consists and explains every step.
The steps applied in the survey design are illustrated in Figure 2.04.

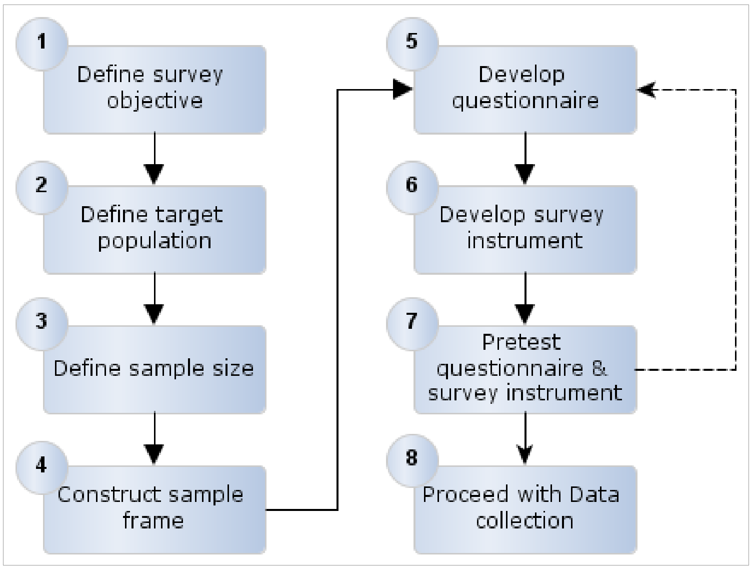


Figure 2.04. Steps in the survey design.

Step 1. Define the survey objective
The objective of the survey is to answer the supporting research questions.
A survey is used to provide answers on the supporting research questions five and seven:
5. What are the information needs about uncertainty in spatial data?
7. Which factors affect the usefulness of information about uncertainty in spatial data?
The goal of the questionnaire is to measure variables as part of the construct information acceptance. Those variables consists of user characteristics, task characteristics, communicated uncertainty, and type of visualization which is provided to use information about uncertainty.

Step 2&3. Define target population & sample size
The survey focuses on users of key Geo-registrations in 25 large municipalities which are defined with more than 100,000 inhabitants (CBS n.d.). Particular in large municipalities the use of key Geo-registrations might be chained across the organisation. According to the survey carried out by Geobusiness Netherlands, the number of municipal professionals who work in the field of geographic services is estimated on *390* (Geobusiness Nederland 2010). With an estimated population size of 390, the strived size of the sample is approximately 100. Therefore a survey using a questionnaire is a suitable method to gather the data (Babbie 2001). In this survey a distinction will be made between two different types of users according to their task, education, skills, and applicability of key Geo-registrations which are listed in Table 2.01.

type of user	**information demand**	**user demand**
Information specialist	raw data	analysis,flexibility
Policy planner	raw and pretreated data	analysis, good accessibility

Table 2.01. Users and demands, adapted and modified from Nijkamp & Scholten (1993)

In order to be able to construct a sample frame and perform a suitable selection of the sample the profile of the participant is extended. The profile of a potential participant contains the following characteristics:

The participant which attends to the survey must specify to a certain profile:

- The participant works in a municipal context, and participates in a municipal data chain
- The participant works frequently with Geographic data and combines base data (Topography, Cadastral, addresses & buildings) with domain specific data.

Within the domain where the participant operates, one of the following tasks are performed:

- Usage base data as a reference and support for domain specific tasks.
- Visualization of topography or cadastral data combined with domain specific data
- Analysis of key Geo-registrations, derivation of qualitative or quantitative data
- combine base data with domain specific data, i.e. derivation, integration

Step 4. Construct sample frame
There is no list available of all professionals which use key Geo-registrations.
To construct the list of participants, key persons of municipalities were approached to provide a list of names with potential participant. The selection performed by the key person was based on a profile that the potential participant should fit.

Step 5. Development of the questionnaire
To provide answers on the research questions, measurable variables are developed based on the technology acceptance model (Venkatesh & Davis 2000). These variables are further specialized to implement a psychometric scale used in psychology (Trochim 2006). The scale applied is the Likert Scale. In the questionnaire, several well known criteria for formulating question text (Bethlehem 2009) are implemented to reduce observation errors. The questions for the questionnaire are developed to fit in the construct which is based on the technology acceptance model . Further details of the application of technology acceptance model in this research is discussed in chapter six.

Step 6. Development of the survey instrument
The survey chosen instrument which facilitates the questionnaire is computer-assisted web interviewing (CAWI). Conducting a survey must be feasible within a reasonable time and still have a good response. A survey instrument should be easily accessible and easy to operate for the respondents, and for the researcher the data must be well organized and structured to perform analytical operations. Taking these requirements into account, the appropriate instrument will be an electronic form accessible via a website (Bethlehem 2009; O'Leary 2010). For this survey a website is developed (Figure 2.06), containing an introductory page followed by the page with the questionnaire which are provided in Appendix 1 and 2.

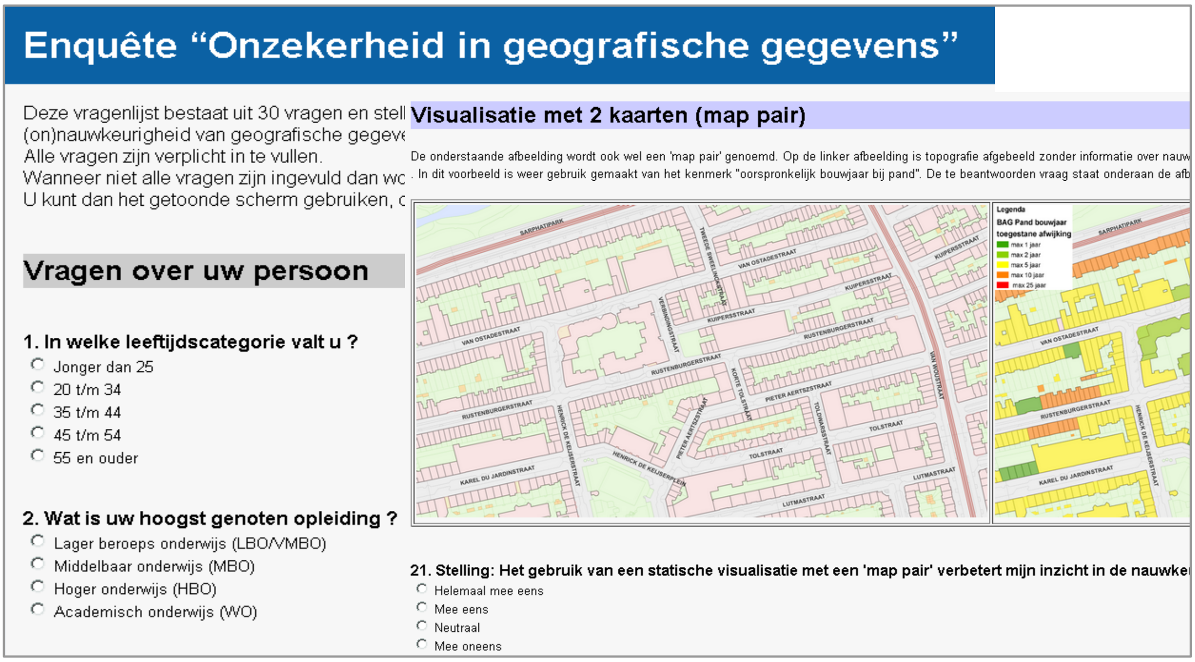


Figure 2.05 Impression of the survey instrument.

Step 7. Pre-testing the questionnaire and the survey instrument
Both the survey instrument and the questionnaire are pre-tested to reduce design bias and optimize the measurement scale. Experts who are familiar with research and key Geo-registrations performed this test.The performance of a pre-test of the questionnaire has several goals.Those goals can be divided into two groups; the content of the questionnaire and the survey instrument.

The content of the questionnaire
The content of the questionnaire is tested on several aspects. An important aspect is the validity of the questions. The questions should measure the target variables which are defined for the purpose of the questionnaire. To improve validity, questions are tested on question formulation, completeness of questions (missing relevant questions), duration of questionnaire, understandability, and the structure of questionnaire.

The survey instrument
The survey instrument is also a key factor to gather the data to obtain the answers on the research question. It is important for the participants to understand the purpose and know the duration of the questionnaire. The tool which is used must be easy to use and accessible. Three experienced users including a professional tester carried out the pre-testing.
The performed pre-test led to adjustments in the cover letter, the formulation of the questions and the supporting text of the questions.

Step 8. Data collection
The data collection is performed using computer-assisted web interviewing (CAWI). This implies that the respondents respond without direct intervention and independent of the researcher. The instrument used is self administered and stores the gathered data in a tabular form which is directly imported into the statistical package "R" for statistical analysis.

2.3.3. Data analysis

The approach for analysis chosen must be appropriate with the scale of the collected data and the purpose for the analysis which are related to the research question. The collected data from the survey is of an ordinal scale. The main purpose is to investigate distributions and relationships. The used statistical methods should be appropriate for ordinal data and how they are related to the research question. Regression analysis is applied to develop a statistical model that can be used to investigate relationships between influential factors and usefulness. Rank order correlation is applied to measure the strength of the association between quantitative variables.

2.4. In-depth interviews

One of the key methods of qualitative research are in-depth interviews.
In-depth interviews are open-ended and have exploratory nature (O'Leary 2010; Cook 2012). The objective of the conducted interviews is to understand the respondent's perception about the appearance of uncertainty of spatial data in a municipal information chain, and the usefulness of communicating about uncertainty in spatial data. The interview is semi-structured based on a predefined questioning plan, but can shift in order to adapt the flow of the conversation. The advantage of this approach is the ability to respond to new developments in the conversation and to study the perception of the participant in depth.
Uncertainty in spatial data might be an ambiguous concept for participants. By introducing the concepts of uncertainty in the interview by an introductory letter as provided in Appendix 3, the interviewer can set the same 'frame of mind' with the interviewee to achieve a common understanding of terminology and definitions. The target population is the same as the population defined for the survey and consists of information specialists with expert knowledge of (their) municipal processes and GIS. The participants of in-depth interviews consists of key informants. Those key informants are non random selected and are probably the most knowledgeable about spatial data and the business process of the municipal organisation which they represent.

According to Kvale (1996) in-depth interviews consists of the following stages as depicted in Figure 2.11; The starting point of the process of collecting data through in-depth interviews is to thematise the research aims (1), followed by designing the process of interviewing (2).
The introductory letter including preparation for the interviewee is included in Appendix 5.
Stage 3, the interviewing is supported by material consisting of examples of uncertainty in spatial data and visualization methods which are included in Appendix 6.

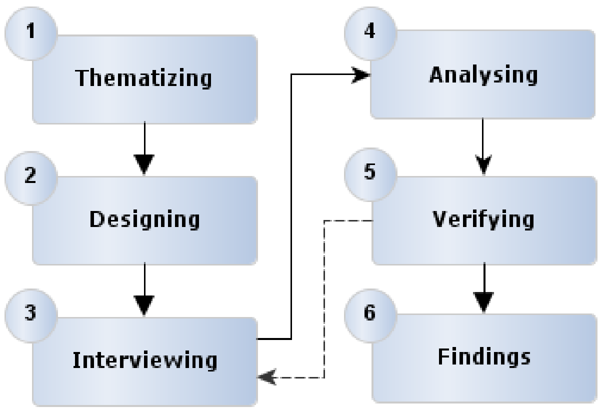


Figure 2.06. The stages of the conducted in-depth interviews.

The results of the interviews are transcribed in an unambiguous and structured way to support the thematic data analysis (4). The preliminary findings are verified against logic and theoretical grounds(5). If necessary, illogical or questionable answers are verified with the interviewee. Finally the findings serve as input for the integrated analysis.

Population & sample
The population characteristics must match the scope of this research, the use of key Geo-registration in a municipal context. Therefore the sample must be representative for municipal tasks which use key Geo-registrations. This implies that variety and in-depth understanding is of greater importance than sample size. The target sample size is 10 in-depth interviews which cover municipal tasks with the most intensive use of key Geo-registrations.

Analysis
The qualitative analysis is performed on the transcriptions structured by themes which will be explained in chapter six, and compared with the results of the literature review.

3. Conceptualizing uncertainty in spatial data

Uncertainty is a concept which might be interpreted in different ways for different goals. Therefore the goal of this chapter is to explore concepts of uncertainty in spatial data and how it is incorporated in this research. The first section explores the general concept of uncertainty. Section 3.2 discusses different taxonomies of uncertainty in spatial data, and Section 3.3 focuses on a more detailed level of uncertainty in spatial data and how the concept is incorporated in this research. Due to the possible relationship between uncertainty in spatial data and spatial data quality this relationship is discussed in Section 3.4 of this chapter.

3.1. Concepts of uncertainty

According to Websters dictionary (1986) uncertainty is defined as:

“*The quality of state being uncertain: lack of certainty, Something that is not certain: something doubtful or unknown*”.

The online dictionary Oxford dictionary (2010b) gives a similar definition of uncertainty:

“*The state of being uncertain*”.

Both definitions refer to the concept 'uncertain', the meaning of this concept according to Websters(1986) is given as “*Not fixed in time, not certain to occur, not known, Not fixed in place,direction, or course , doubtful, changeable, fickle, variable*”.

Skeels et al. (2010) states that uncertainty is a psychological awareness which is subdivided into three levels of uncertainty awareness:

1. Known knowns; what we actually know
2. Unknown knowns; we know what we should know, and are aware of its existence
3. Unidentified unknowns; what we should know, but are not aware of

Figure 3.01. Levels of uncertainty awareness , adapted from Skeels et al. (2010).

Uncertainty level 1, “Known knowns” indicates that uncertainty about phenomena is minimal. When deviations in characteristics of spatial phenomena objectively are definable and quantifiable, it may be defined as error (MacEachren et al. 2005), otherwise those measures remain uncertain.

When sources of uncertainty are known and measures are not (yet) quantified uncertainty level 2 appears, the "Unknown knowns". Users might know there are uncertainties, but are not quantified to assess the fitness for their use. Unidentified unknowns (uncertainty level 3) appear when users are not aware of uncertainty in spatial data. This level of uncertainty might cause a gap in recognized information needs. The link between the unknown knowns and the unidentified unknowns with this research is the investigation of the usefulness of providing information about uncertainty appearing in those two uncertainty levels.

3.2. Uncertainty in spatial data

The purpose of this section is to formalize the concepts of uncertainty in spatial data in the context of this research. The different established concepts and taxonomies will be discussed.

Uncertainty in spatial data is endemic (Longley et al. 2011a) and appears in different forms. To understand the nature of uncertainty it is important to realise that our world is complex and that geographic representations of our world might be incomplete due to cognitive limitations, generalizations of reality, or measurement errors (Longley et al. 2011b). Uncertainty arises in our perception of reality and how clear geographic entities are able to be translated from our perception of the physical reality to a model. In the process of modelling itself discordance and non-specificity appear when entities are ambiguously defined to formulate a model. Measurements are performed to represent those models and the corresponding rules for data collection to store the results in spatio-temporal databases for further use. Due to imprecision of measurements uncertainty occurs in the accuracy of the measured objects.
Figure 3.02 illustrates this process discussed above.

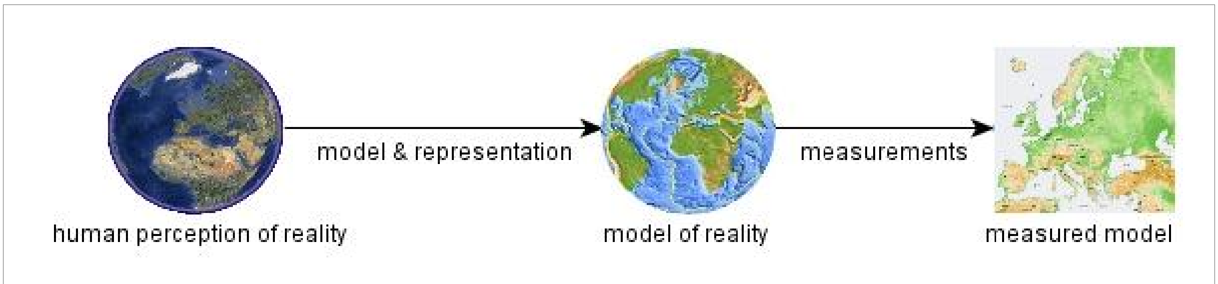


Figure 3.02. From reality to measured model.

Measured models are used for analysis. Uncertainty in measured models will propagate in results of performed analytic actions and perceptions (Longley et al. 2011b).

In this research uncertainty in spatial data is used as an umbrella term which describes different appearances of uncertainty. Different appearances of uncertainty are described in taxonomies from Shi (2010), Fisher, (1999) and Shu (2003).
These taxonomies are explained, compared and discussed in the subsequent paragraphs.

The taxonomy of Shi

The taxonomy of Shi (2010), divides uncertainty into imprecision, vagueness and ambiguity, which is illustrated in Figure 3.03. According to Shi, imprecision refers to the variation and repeatability of measurements. Imprecision occurs in geometry, thematic attributes, and temporal characteristics of an object instance. In statistical theory precision is a degree of dispersion and measured as standard deviation and can be modeled using confidence region models grounded with solid statistical theory and probability. Vagueness occurs when the demarcation of objects can not be sharply defined, or when the membership of classes which an object might belong to is not clear. Vagueness can be modeled using fuzzy set theory (Zadeh 1965) to obtain measures and incorporating in analysis. Ambiguity refers to the lack of a clear definition. An object class might be discord when an object can be assigned to more than one class. Non-specificity arises when an object class and the instance of an object are open for interpretation, or candidate classes are not specific enough. Measures for ambiguity like discord measure an non-specificity measures are based on evidence theory founded by Dempster and Shafer (1976).

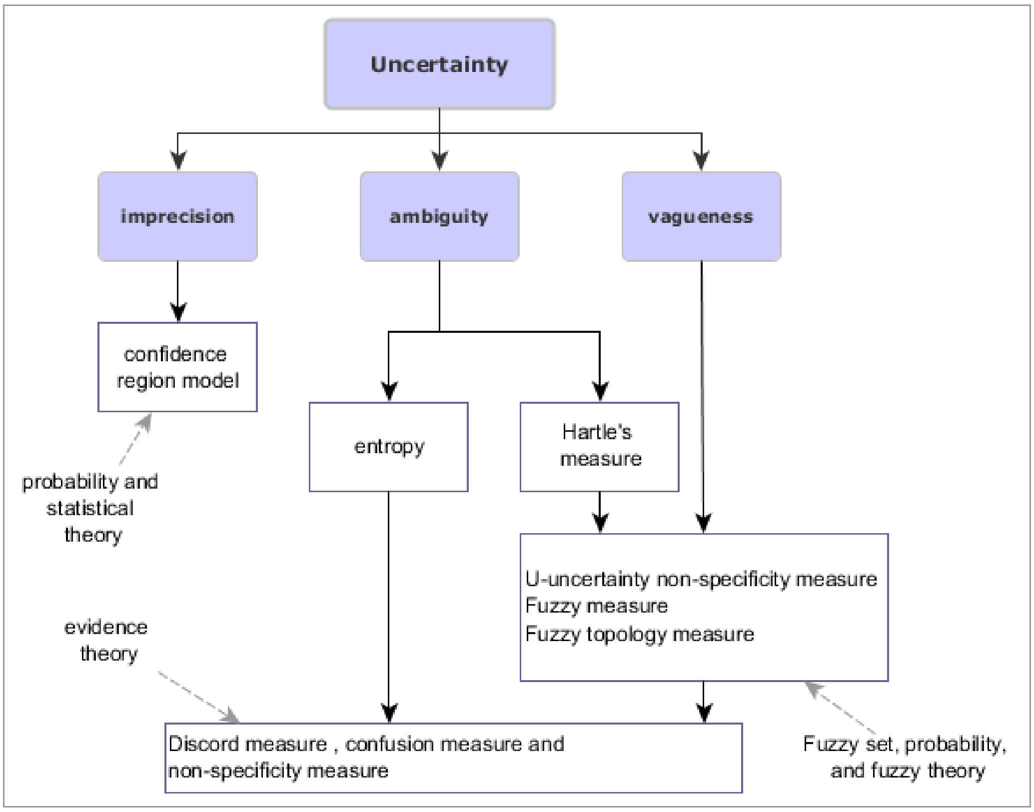


Figure 3.03. The taxonomy of Shi.

The taxonomy of Fisher.
The taxonomy of Fisher (1999) which is partly adapted from Klir & Yuan (1995) approaches the concept of uncertainty from its ability to define object classes or individual objects. Figure 3.04 illustrates the taxonomy of Fisher.

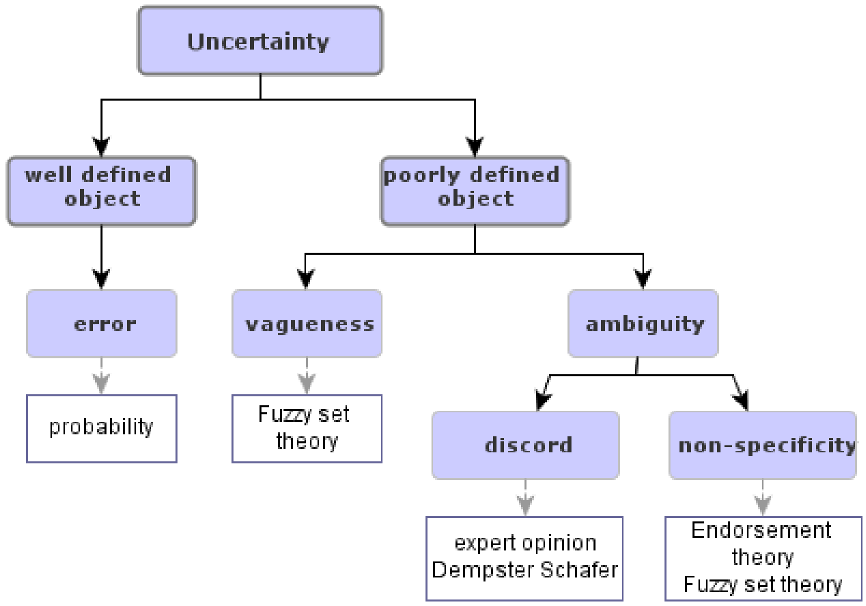


Figure 3.04. Taxonomy of Fisher.

When an object objectively can be well defined the appeared uncertainty might be caused by errors in measurements and can be handled using probabilistic methods.
When the object is poorly defined, then vagueness and ambiguity arises. The terms vagueness and ambiguity are equal with the corresponding terms discusses in the taxonomy of Shi in the previous paragraph.

The difference between the taxonomy of Shi and the taxonomy of Fisher is the use of the term *imprecision* versus the term *error*. Despite this difference, in the referring literature (P. F. Fisher 1999; WenZhong Shi 2010) both terms are defined as probabilistic in nature. According to several dictionaries, the linguistic definition of 'Error' varies. The first meaning of the noun 'error' which were given is 'mistake'. This might be conceived as negative, or wrong. According to the Oxford online dictionary (Oxford dictionary 2010a), the meaning of 'Error' is defined as: *"a measure of the estimated difference between the observed or calculated value of a quantity and its true value"*, which not always is perceived as 'wrong'.
If either the taxonomy of Shi or the taxonomy of Fisher is used, in both cases the perceived meaning of 'imprecision' or 'error' has to be taken into account.
When measured values are inaccurate within the range of conformance levels they assumed to be acceptable. When measured values are outside conformance levels they are rejected and might be perceived as errors.

The taxonomy of Shu et al.
The taxonomy of Shu et al. (2003) has a quite different approach compared to the previously discussed taxonomies of Fisher and Shi. The approach of Shu et al. puts the emphasis on the appearance of uncertainty. Uncertainty appears from a human, earth, and the computing machine perspective. From a human perspective cognition uncertainty appears, computational uncertainty appears from a computing machine perspective, and geographic uncertainty appears from an earth perspective. The relations between those perspectives are illustrated in Figure 3.05.

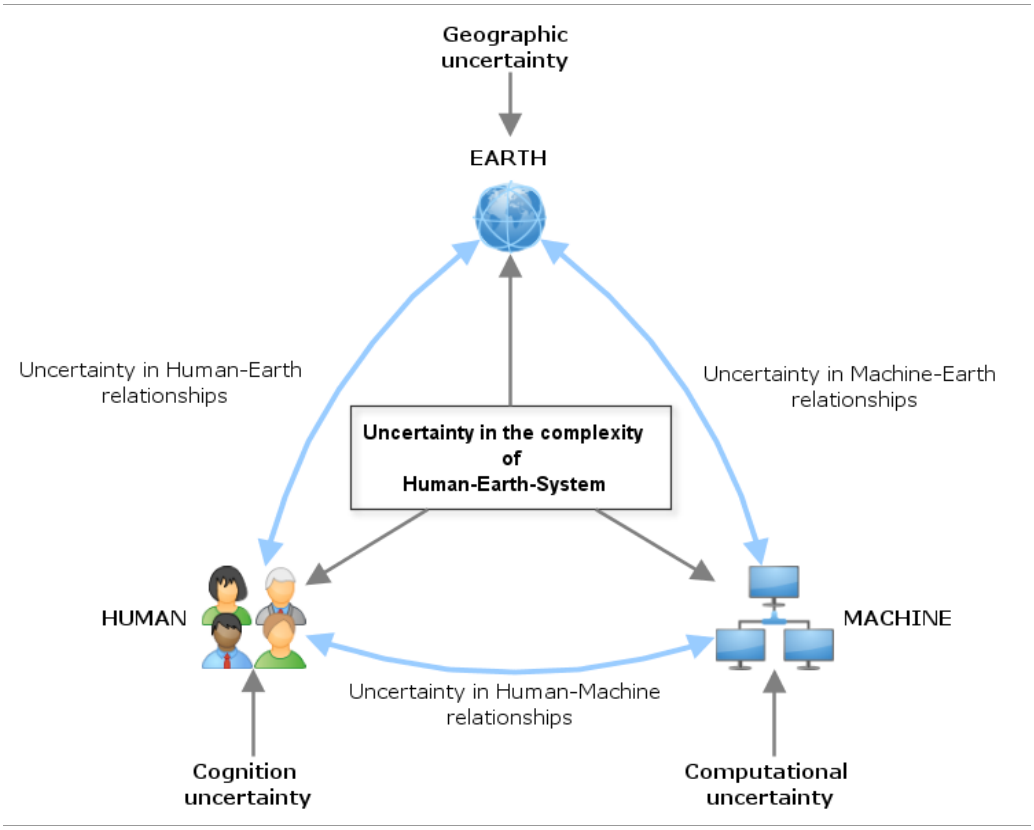


Figure 3.05. The nature of uncertainty according to Shu et al.

This taxonomy is divided into uncertainty of entities and uncertainty of 'human-machine-earth' relations. Uncertainty in geographic entities arise due to the complexity of natural and anthropogenic features which might difficult to define through their possible fuzzy nature. Cognitive uncertainty is divided into by human perception, limitations in knowledge (memory), and logical reasoning (thinking). Uncertainty of human-machine-earth relations is divided into inaccuracy, incompleteness, inconsistency, and imprecision. According to Shu et al. (2003, p.2), inaccuracy refers to " to some deviation of measurement value from the true value". A deviation might be interpreted as an approximate difference or as an incorrect difference when the deviation is not acceptable by the user. Incompleteness is explained as the missing of values due to partial computational and cognitive description of geographic phenomena. Representational diversity(conflicting), semantic mismatch(incoherency), and semantic contradiction (invalidity) are appointed as the components of inconsistency. According to Shu et al. imprecision refers to the degree of exactness of computational and cognitive values closely related to resolution of possible values. The concept imprecision how Shu et al. this explains, is subdivided into non-specificity, ambiguity or confusion, and vagueness or fuzziness. Compared to previously discussed definitions of the concept imprecision, Shu et al. groups the components ambiguity and vagueness under the concept imprecision, where Shi and Fisher. explicitly separate the concept imprecision from vagueness and ambiguity. The difference in approach between Shu et al. and the aforementioned taxonomies is a consequence of a difference in the definition of the concept imprecision.

3.3. The implemented taxonomy of uncertainty

The previous paragraphs discussed taxonomies which showed agreements and differences of the concept of uncertainty in spatial data. To provide a consistent theoretical foundation for this research several aspects of the taxonomies of Shi and Fisher are adopted, modified, and expanded. The motivation for this choice is that the taxonomy of Shu has a more holistic approach which less fits within the scope and focus of this research. The following paragraphs describes a taxonomy of uncertainty which shall be used in this research and discusses the definitions of the terms used in this taxonomy. Ambiguity, vagueness, and inaccuracy are the key elements of the taxonomy of uncertainty in spatial data which will be used in this research, and are illustrated in Figure 3.06.

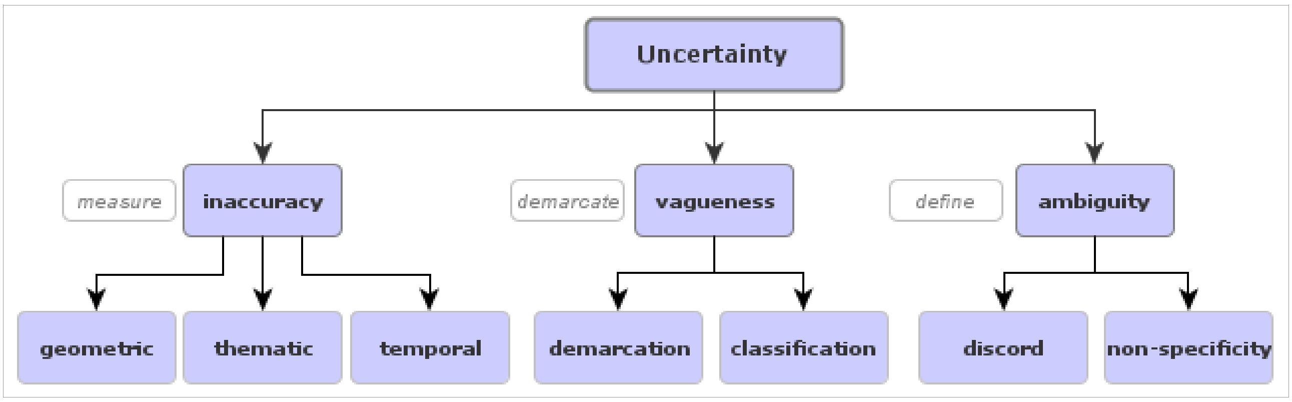


Figure 3.06. Adapted and modified taxonomy of uncertainty.

3.3.1. Ambiguity

Ambiguity occurs when there is uncertainty about the classification of a geographic phenomena mainly caused by different perceptions (P. F. Fisher 1999). Ambiguity is divided into discordance and non-specificity. Discordance applies when the definition of a phenomena allows to assign the phenomena to more than one class. Examples of discordance are the borderline dispute between Nicaragua and Costa Rica (discordance in demarcation), and when the same area is classified differently which is perceived and mapped by different persons. Non-specificity occurs when it is not possible to assign a phenomena to a class because the definition is not specific enough and open for interpretation (Devillers & Jeansoulin 2006). An example of non-specificity is the assignment of land covered with forest.

3.3.2. Vagueness

Vagueness appears in the form of fuzzy boundaries and inhomogeneous areal features. Boundaries of geographic objects are not always crisp due to the fuzzy transition between to different classified features. Areal features of geographic objects are not always homogeneous due to mixed types of land cover. Most buildings are accurate to demarcate due to their crispness. Boundaries of more natural entities like woodland or natural water bodies are not always crisp with the consequence that the demarcation of such entities is uncertain. According to specifications of the key registration large scale topography the transition between water body and other type of terrain must be acquired, but the vegetation like reeds are not separately distinguished. Formal models to describe fuzzy boundaries and fuzzy classification are established by Zadeh (1965) by the introduction of fuzzy logic and fuzzy set theory. Fuzzy logic extends boolean logic by introducing the concept that phenomena can be partial member of a class, instead of true [1] or false [2]. This partial membership is expresses in a continuous value between 0 and 1 using fuzzy membership functions (Burrough & McDonell 1998, pp.268–273; V. B. Robinson 2003).

3.3.3. Inaccuracy

Where Shi applied imprecision, in this taxonomy the term inaccuracy has been chosen. Precision differs from (in)accuracy, and describes the repeatability of *measurements* as a quality measure (Shi 2010). Accuracy is the degree to which stored values matches a true or accepted value (Foote & Huebner 1999; Longley et al. 2011b; WenZhong Shi 2010) . Inaccuracy may be defined as the discrepancy between a stored value and its 'true' given value. Heuvelink (cited in Longley et. al. 2011b) defined accuracy as "the difference between reality and our representation of reality". The use of the phrase 'our representation' implies that there are a variety of possible views of a complex world we want to model. The motivation to adapt the concept inaccuracy instead of imprecision is based on the previous explanation of this concepts and fits better in the frame of concepts within this research.

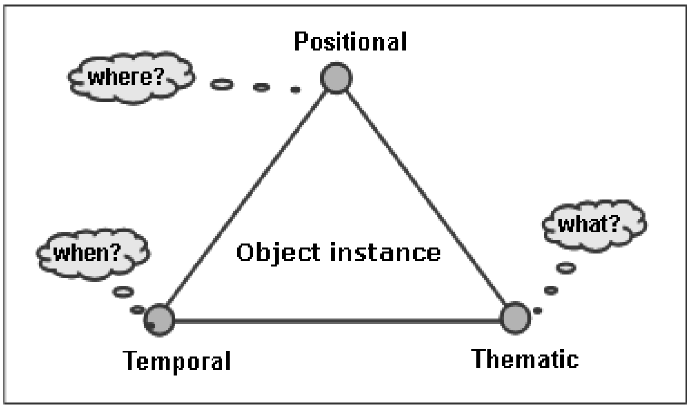


Figure 3.07. Spatio-temporal theme triangle (P. van Oosterom 2010).

Inaccuracy can be divided into three elements; positional inaccuracy, thematic inaccuracy, and temporal inaccuracy. In spatio-temporal data every object instance has positional, temporal and thematic characteristics which inherently contain inaccuracies which are depicted in Figure 3.07.

3.3.3.1. Positional Inaccuracy

When inaccuracy occurs in geometry of individual objects (shape and location) then the discrepancy between the 'real' value and the stored value is uncertain to the user.
The term 'error' is frequently used as a synonym for discrepancy between stored and 'true' value. In the context of inaccuracy, errors are not with an unlimited precision quantifiable, but with in a certain range errors are quantifiable.

3.3.3.2. Thematic inaccuracy

Thematic inaccuracy may occur in attributes of an individual object, and can be defined as the discrepancy of an attribute value to its true value (Foote & Huebner 1999). In terms of uncertainty, the user is uncertain about the magnitude of this discrepancy. Thematic inaccuracy of quantitative variables can be investigated using probabilistic functions, and qualitative or categorical variables can be assessed using error matrices, or the incidence of Defect Method (WenZhong Shi 2010).

3.3.3.3. Temporal inaccuracy

At the most fundamental level, time can be interpreted as a relative view of time and as an absolute view of time (Peuquet 2000). In the *relative* view time is expressed as a time interval (duration between two time instants) or as time series which consists of multiple time intervals (Peter Van Oosterom et al. 2006). When time is considered as an *absolute* entity with respect to space, the view focuses on time as the subject matter, the objects which are considered are with respect to geometry and and theme unchanged. Time is a continuous phenomena which is often represented as a discrete numerical scale, with a chosen time scale and granularity (Bordogna et al. 2009). Due tot the extendible character of the time indications uncertainty in time can be modeled using fuzzy instant, fuzzy intervals, and fuzzy time series (Campaña et al. 2007; Bordogna et al. 2009).

3.4. Uncertainty and spatial data quality

Uncertainty is related to data quality, this section discusses the nature of this relationship. Prior to discussing the relationship between uncertainty and spatial data quality, first the concept of data quality is discussed.

The concept of spatial data quality

The concept of spatial data quality is often interpreted in different ways, when GI-specialists in the 'field' are asked to give a description of spatial data quality, different answers are given. Some define spatial data quality “data which is exempt from errors”, or “data which is very accurate”. Others define spatial data quality when the data fit their purpose of use. According to the ISO 19011: 2002 standard is Quality is defined as “totality of characteristics of a product that bear on its ability to satisfy stated and implied needs” (Van Oort 2005, p.13). The elaboration of standardization in spatial data quality is established in the ISO standard *19113:2002 Geographicinformation – Quality principles*. Devillers & Jeansoulin (2006) define two groups for data quality; *internal* data quality and *external* data quality. Internal data quality is the agreement between the data produced and what should have been produced (specifications), and is mainly the concern of the data producers. External data quality is the level of concordance between the data product and the users needs in a specific context. Figure 3.08 illustrates relationships between internal en external data quality in the context of key registrations.

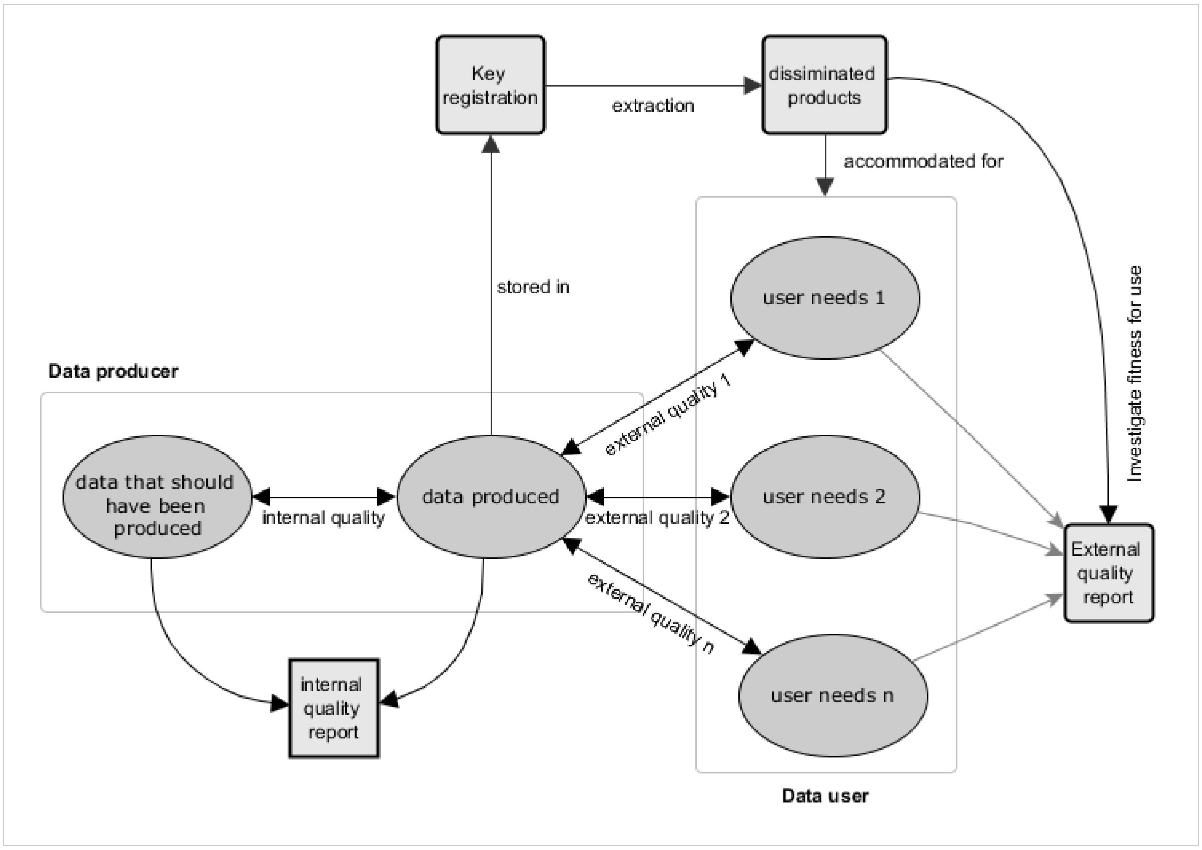


Figure 3.08. Spatial data quality , Adapted and modified from Devillers and Jeansoulin (2006).

Also in Figure 3.08 the distinction between production of data to maintain a key registration, and the dissemination to products to provide users of key Geo-registrations is illustrated.
In most cases key registrations like Cadastre, topography (TOP10NL), or Addresses & buildings are stored in a single database, and are extracted to products in different forms and formats to meet the users needs. This implies that evaluation of external data quality might not be based on the same data quality standards and data quality measures as in internal data quality procedures.

The standards concerning spatial data quality are listed in Appendix 4.A (Jakobsson & Giversen 2008). When datasets are assessed from an usability perspective (ISO 1998; Folmer & Bosch 2004; Blake & Mangiameli 2011) the data quality elements described in Appendix 4B. are considered. External data quality elements complemented with internal data quality elements might be useful to test the usability of a dataset. To carry out an investigation to asses the fitness for use and to measure the conformance of the data, the data users should have their requirements established, and the data should provide information about the definitions and the accuracy of the data. And how does uncertainty fit in this argument?
Figure 3.09 illustrates relationships between spatial data quality and uncertainty.

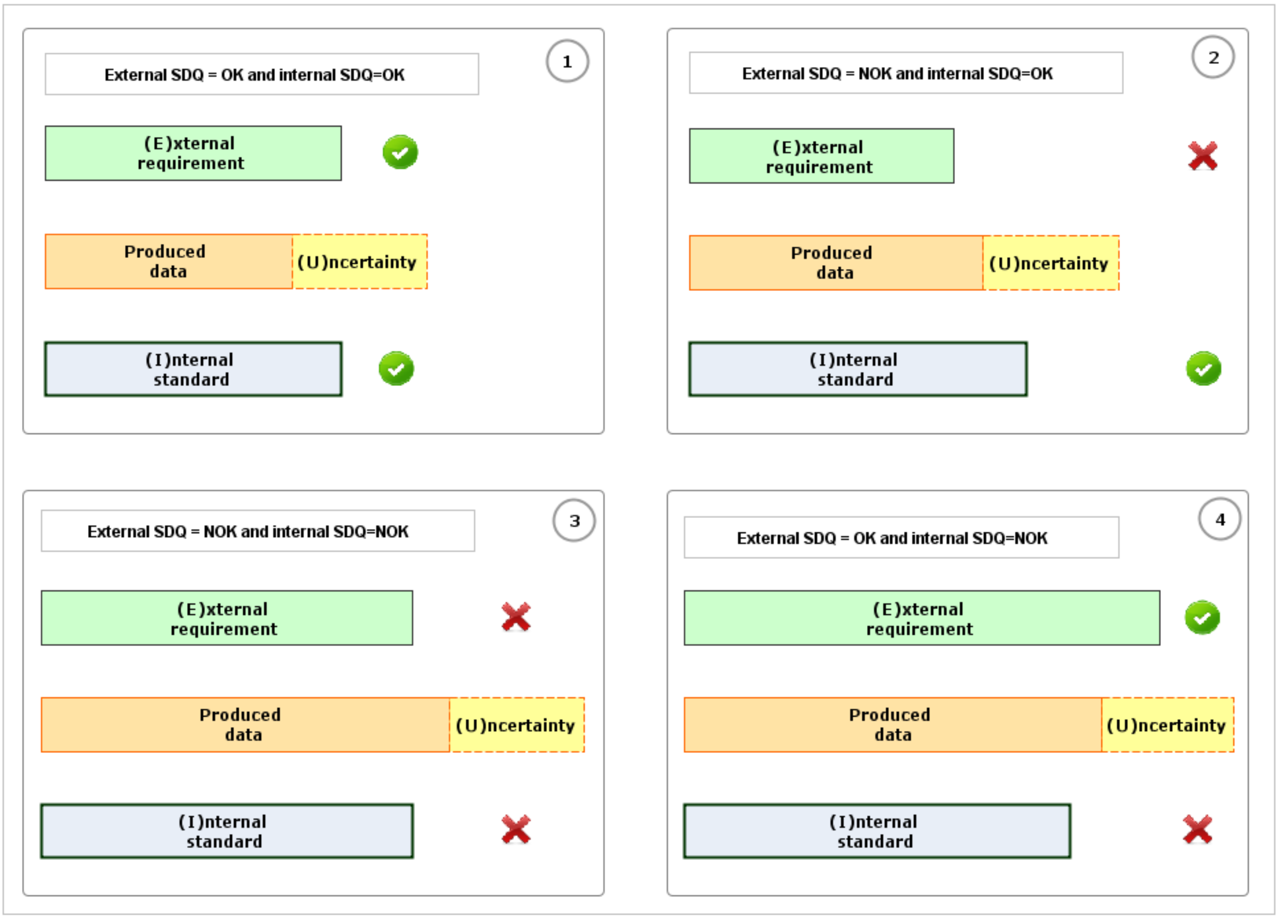


Figure 3.09. Relationships between uncertainty and data quality.

Three types of bars are depicted in four boxes which represent:

- A quantification of a requirement, external data quality (**E**),
- A quantification of a value of uncertainty (**U**),
- A quantification of an internal data quality standard (**I**)

An abstraction of this statement for each box is given:

Box 1: $E \leq U$ and $I \leq U$; **Box 2**: $E \leq U$ and $I > U$;
Box 3: $E > U$ and $I > U$; **Box 4**: $E > U$ and $I \leq U$;

When the value of a requirement E or an internal standard I equals or smaller than the quantified uncertainty then there is conformance. When this is not the case the produced data does not fit the users requirements or it does not conforms the data quality standard. Hence I would state that information about uncertainty in spatial data probably supports the investigation of external spatial data quality from a data user perspective, and supports the data producers to test the conformance of the data against the data quality standards.

4. Communicating uncertainty

The purpose of this chapter is to explore different aspects of how uncertainty in spatial data can be communicated, and discusses previously conducted research concerning usability of visualizing uncertainty. Different methods of communicating of uncertainty in spatial data are briefly discussed, and usability studies based on a literature study of previously conducted research are discussed.

4.1. Methodological approaches

Communicating uncertainty in spatial data can be interpreted as a part of information visualization. According to Mazza (2009) visual representations should be designed with the consideration of a collection of variables. Those variables are divided in five groups which are shown in Table 4.01.

Purpose	Level of measurement	Dimensions	Data structure	type of interaction
Communicative	Nominal	Univariate	Linear	Static
Exploratory	Ordinal	Bivariate	Temporal	Transformable
Confirmatory	Interval	Trivariate	Spatial	Manipulable
	Ratio	Multivariate	Hierarchical	
	Cyclic		Network	

Table 4.01. Variables to consider for designing visual representations, based on Mazza (2009)

Values of uncertainty measures which are implemented into variables of a defined data set can be mapped to visual variables like colour hue, colour saturation, symbol size, transparency and so on (Slocum et al. 2009). Representation of uncertainty in spatial data can either be applied as a communicative purpose, or as an exploratory purpose. A visual representation with communicative purpose supports a particular message or an idea. A representation with an exploratory purpose is intended to give the data user the ability to 'explore' the data to discover information and relationships in the data. Reinke & Hunter (2002) developed a framework for communicating uncertainty in spatial databases. The approach is a transformation from data quality measures into an interactive visualization of spatial data quality. The factors which might affect the display design are based on personal, social, and cultural characteristics. The interaction with a visualisation can be based on different communication goals; notification, identification, quantification and evaluation. The framework of this process is illustrated in Figure 4.01.

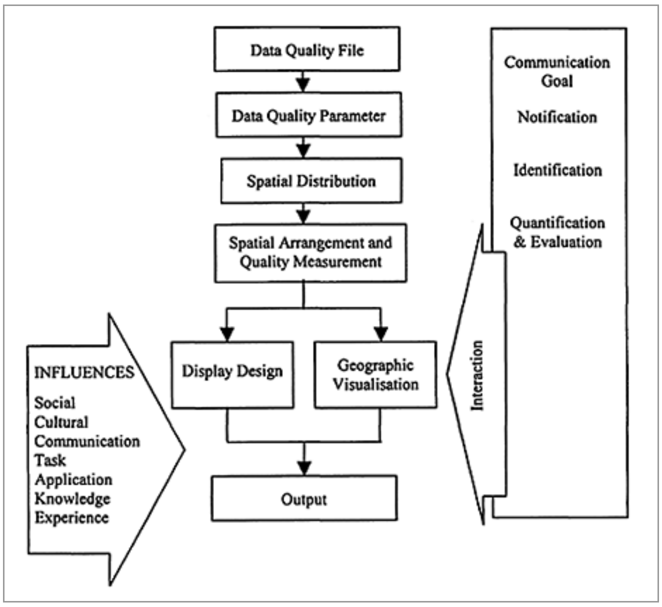


Figure 4.01. Reinke & Hunter's framework for communicating uncertainty.

4.2. Methods of communicating uncertainty in spatial data.

To communicate uncertainty in spatial data several methods are developed. Those methods can be grouped into visualization of uncertainty, metadata, and data quality reports (Pham et al. 2009; WenZhong Shi 2010). This section explores previously conducted research regarding these methods.

4.2.1. Visualization of uncertainty

Several studies concerning visualization of uncertainty in spatial data are conducted by Brewer, MacEachren, Goodchild, Ware, Shi, Zuk, and many others. Most studies conducted concerned the investigation of visual variables which are grounded with solid theories established by Bertin (1983), Tufte (2001), and based on Gstalt principles (Zuk & Carpendale 2006). Drecki (2009) gives an overview of cartographic solutions which consist of visual variables, static representations, dynamic representations, and non-visual representations (sound maps). Important contributions to visualizing uncertainty are made by MacEachren (1992; 2005) and Pang (2001; 2008). Recently conducted research concerning to visualisation of uncertainty is carried out in the domain of urban planning and performed by Zhang (2008), Keijzer (2010), and Vullings et al. (2009) which are related to the GEO3 research project 'Dealing with uncertainty in spatial planning'. One of the aims of this research project was the improvement of the visualisation of uncertain spatial planning objects, regarding incomplete defined objects, and discrete defined field objects. The mapping of improved classification of uncertainty on visual variables contributed to achieve this specific research aim.

Table 4.2 gives a summary of the relevant methods of visualizing uncertainty in spatial data.

Name of the Method	Description
Map pairs	one map with data of the phenomena, and one map illustrating uncertainty using visual variables (Slocum et al. 2009).
Bivariate maps with *Intrinsic* visual variables	Intrinsic variables are intrinsic to the data, the representation of the phenomena is modified using visual variables (Slocum et al. 2009).
Bivariate maps with *Extrinsic* visual variables	Extrinsic variables will be added to the data using visual variables and are distinguisable from the phenomena in one visualization (Slocum et al. 2009).
Animation	*(1) Animation* of different realisations of the uncertain attribute to emphasise the uncertainty (Ehlschlaeger et al. 1997). *(2) Blinking Regions*, where two images of data and uncertainty are overlaid on top of another and alternately displayed (Kardos et al. 2006). *(3) Blinking Pixels* where the displayed data is manipulated to blink through constantly changing colour of the pixels of more uncertain data with a rate of change proportional to the uncertainty (Zhang 2008).
Hierarchical Spatial Data structures	These are used as a transparent tessellated layer on top of the data. A finer tessellation indicates less uncertain areas, a coarser tessellation indicates more uncertain regions (Kardos et al. 2007).
Colour Models	*(1) RGB Colour Scheme*, red, green, blue represents the variables and colour intensity represents the uncertainty with higher intensity depicting lower uncertainty (Cliburn et al. 2002, MacEachren 2005). *(2) Whitening*, where the colour hue is used to represent the data and the saturation-intensity (whiteness) is used to represent the uncertainty. (Hengl 2003). A similar result is obtained by the technique of pixel mixing (Longley et al. 2011a).
Glyphs	Uncertainty and the data is represented in a bivariate depiction through pictorial symbols, known as glyphs (Pang 2001).
Contouring	Contour lines of different colours are used to distinguish between different variables and their uncertainties with the intensity of colour. Positional uncertainty is depicted through the gap widths in the dotted contour lines where higher uncertainty leads to wider gaps (Dutton 1992, Pang 2001). The concept of contouring can be used in an animated environment as animated isolines (Fauerbach et al. 1996).
Focus Metaphors	Uncertain data is depicted out of focus, e.g. foggy and more certain in focus, e.g. crisp boundaries. Another metaphor of this method is the Opacity method where less uncertain data is seen less opaque and more uncertain data is more opaque (MacEachren 2005). This concept can also be used in reverse where uncertain data is shown more transparently (MacEachren et al. 2005, Vullings et al. 2009)

Table 4.02. Visualisation methods of uncertainty (Gerharz & Pebesma 2009)

4.2.2. Metadata

The accessibility of geographic data improved with the growing maturity of geoportals and the use of standardized metadata. Geographic data becomes available on international level through the Inspire Portal, and on national level the Dutch National Georegister. Many other initiatives are still being developed to create domain specific portals for agriculture, archaeology, marine biology, and many more. In order to enable users to investigate the fitness for their use and provide information about data quality, producers of data sets are obligated to provide metadata. Several aspects of the providing information about data spatial data might be applied in different levels of the data (Devillers, Bédard, & Jeansoulin, 2005). Those levels are data series, data set, object class, and object instance. Metadata provides information about data sets, data series and are drafted according to formal standards to ensure interoperability (Jakobsson & Giversen 2008; JRC 2008; WenZhong Shi 2010). Several standards are developed and extended in different countries. The most comprehensive standard is the ISO metadata standard ISO 19115. This standard is developed by the technical committee for Geo-information and Geo-informatics (ISO/TC 211 2011). The ISO metadata standard ISO 19115 contains metadata elements which facilitates to capture elements of spatial data quality. This offers possibilities to provide information about uncertainty in spatial data. According to Comber et al. (2006) metadata reporting possibilities had also its limitations and are insufficient to explicitly link them, and therefore Comber et al. proposed expansion of metadata elements to facilitate free text descriptions. In addition to these developments this research proposes as depicted in Figure 4.02 a method how elements of uncertainty can be incorporated in metadata elements of the ISO 19115 standard. The diagram in Figure 4.02 shows how the elements of uncertainty can be mapped on elements of spatial data quality. Chapter 3 showed that uncertainty and spatial data quality are closely related. Therefore its recommended to perform further research if, or how to incorporate uncertainty into standards.

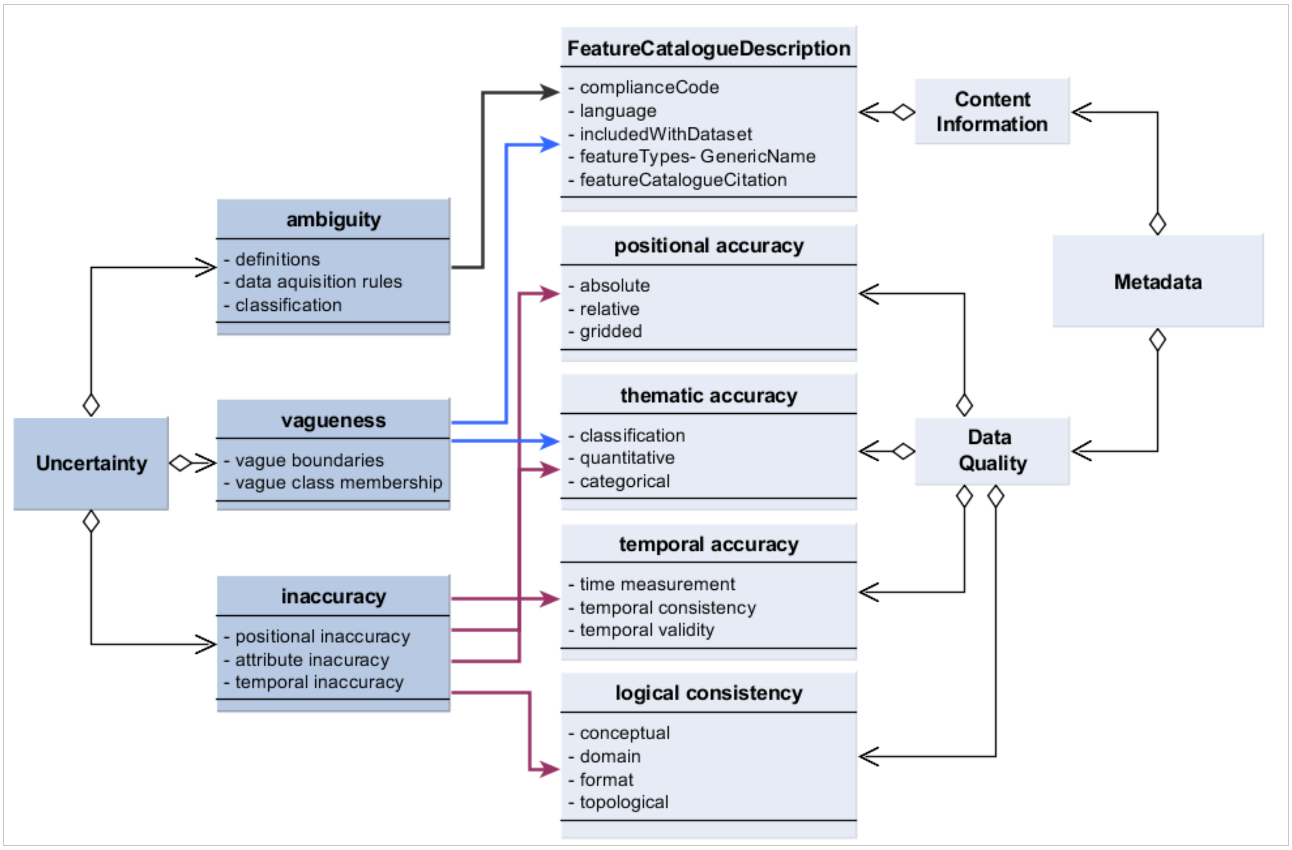


Figure 4.02. Uncertainty, data quality, and metadata.

4.3. Usability of visualizing uncertainty in spatial data

From previously conducted studies concerning usability of visualizing uncertainty in spatial data one may deduce that investigating usability of visualizing uncertainty in spatial data concerns various areas of interest. These are discussed separately in the subsequent paragraphs.

4.3.1. Usability of visualization methods

The studies concerning the usability of visualization methods are conducted by many researchers for a variety of application areas. Aerts, Clarke, & Keuper (2003) tested visualization techniques to represent model uncertainty of an urban growth model called SLEUTH. The objective of this research was to compare two visualization techniques; static comparison and toggling. To investigate the usefulness, a web based survey was developed containing the uncertainty visualizations, and was complemented with a questionnaire. The result of this research was that the static comparison proved to be slightly more favourable than the toggling.

More recent research concerning visualization methods of uncertainty in spatial data is conducted by Kardos, Benwell & Moore (2007) and by Viard, Caumon, & Lévy (2011). Kardos, Benwell & Moore investigated alternative ways of visualizing attribute uncertainty in census data using the Hexagonal or Rhombus (HoR) trustree tessellation.
An internet survey was conducted to assess the usability and effectiveness of six different trustree methods applied to New Zealand 2001 census data. The results of this research showed that a transparent Hor trustree visualization containing uncertainty measures overlaying a choropleth layer with census data shown adjacent with the original choropleth is the most usable and effective way to express spatial and attribute uncertainty. Figure 4.03 shows a map pair with left: the choropleth map with census data, and right: the map containing census data overlayed with the transparent Hor trustree visualization.

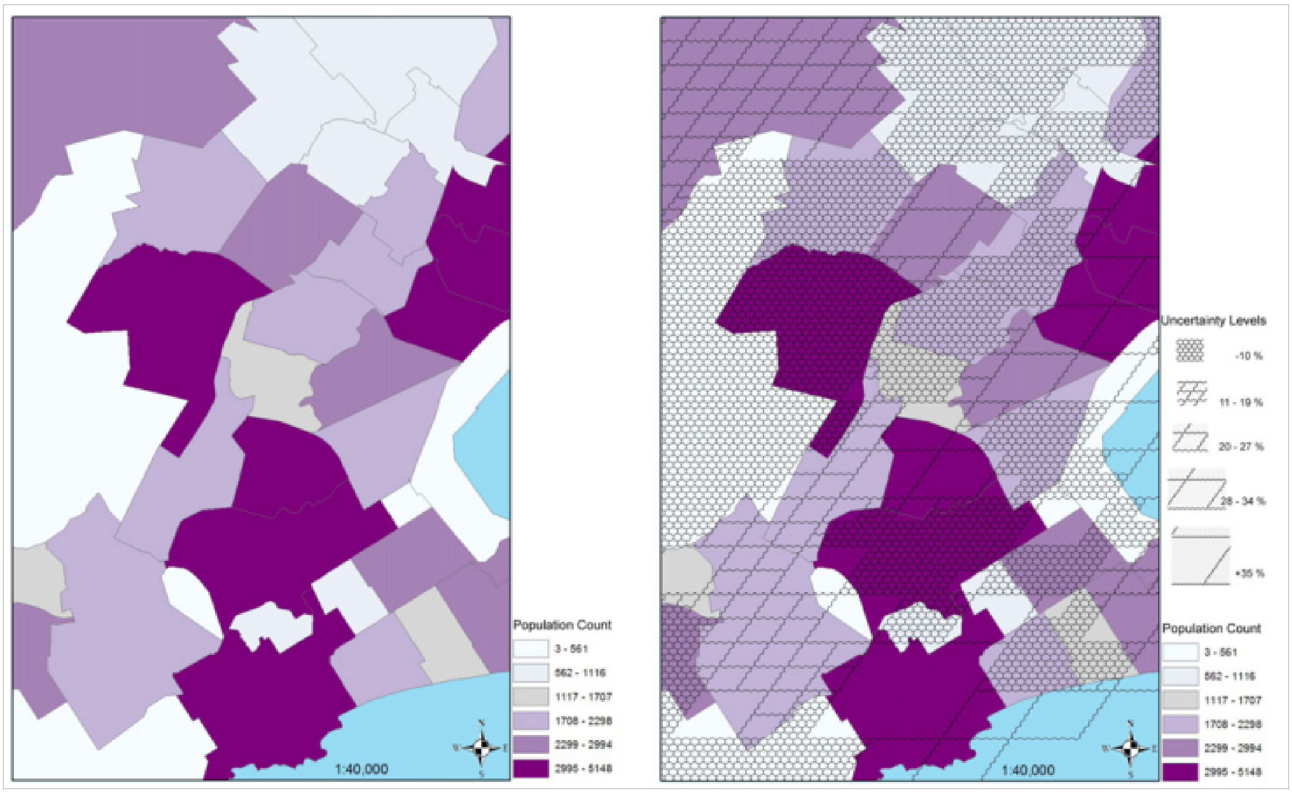


Figure 4.03. Map pairs using transparent Hor trustree visualization.

Viard, Caumon, & Lévy (2011) studied the quality of decision-making using adjacent versus coincident representations of spatial uncertainty as compared to the presentation of data without uncertainty. Their research was conducted in the context of a decision process. The used method was based on an illustrated questionnaire with static maps which visualized geological data combined with visualisation of uncertainty. The results of this research showed contradictory results. When visualisations were increased in complexity the significance of the difference between the two used methods decreased. This was interpreted as when the complexity of a visualization increases, this might perceptual and cognitive overload depending on the skills of the user. The researchers stated that when designing uncertainty visualizations the skills of user must be taken into account to actually add value to decision making.

4.3.2. Research concerning usability of information about uncertainty

Several studies are conducted to investigate the usability of information about uncertainty in spatial data. Those studies focused on decision-making and the use of information about uncertainty in domain specific applications.The previously conducted studies approached their research objective differently. This literature review covers two approaches. Several studies investigated the influence of uncertainty representations on decision-making, other studies investigated the influence of human factors on using information about uncertainty in spatial data.

Influence of representations on decision-making
Hope (2005) investigated the effects of different uncertainty representations on the decisions being made. Hope designed two case studies; the first case study incorporated information about thematic uncertainty to test the effects on decision-making, and the second study incorporated information about positional uncertainty to test the effects on decision-making. The findings of her research were significant regarding the influence of how information about uncertainty is portrayed. Hope (2005) concluded that:

"Decision-makers do not necessarily respond in a rational manner to the inclusion of uncertainty information and can make significantly different decisions dependent upon the form by which the information is portrayed".

The earlier mentioned research of Viard, Caumon, & Lévy (2011) also pointed out that uncertainty visualizations adds value to decisions which must be taken, but that the design of visualisations and the perceptual and cognitive overload of the user must be taken into account.

Human factors
Conducted research with the focus on human factors related to information about uncertainty in spatial data is limited. Roth (2009) investigated the impact of the user's domain expertise on the use of uncertain geographic information in a risk assessment scenario. The results of this research indicated that decisions of users might be affected by both domain and map-use expertise, and that domain expertise is most influential in risk assessment and map-use expertise had the biggest impact on the perceived difficulty of the risk assessment.
Roth (2009) also concluded that:

"The key to designing useful and usable representations and visualizations of uncertainty (or of anything else) is to know the end user".

Boukhelifa & Duke (2009) investigated why visualization of uncertainty in spatial data is rarely adopted in the practical application of visualization. They state that the application of uncertainty visualization is limited due to the fact that the relationship between accuracy and uncertainty is debated by users, and in the field of cartography the terms uncertainty and data quality have a close relationship but these two concepts are sometimes used synonymously. The application of uncertainty visualization can be limited through the quality and scope of uncertainty data, the limited confidence in spatial data quality, and by the perceptual and cognitive confusion that the depiction of this data can generate (Boukhelifa & Duke 2009).

5. Key Geo-registrations in a municipal context

As stated in the introduction and the research objective, this research focuses on uncertainty in spatial data with an emphasis on the use of key Geo-registrations in a municipal context. This chapter puts the previously discussed theories in context, and describes briefly the Dutch system of key registrations and its use in municipal processes. The description of this topic is narrowed by the key Geo-registrations which are used in municipal spatially managed processes. The first section describes the Dutch system of key registrations, the second section describes the most commonly used registrations used in municipal processes, and links this subject with the research aims.

5.1. The Dutch system of key registrations

The Dutch system of key registrations is one of the building blocks of the Dutch national execution program 'NUP' which aims to improve public services to citizens and businesses (BZ 2008). To achieve the improvement of interoperability and cooperation between public organizations, citizens an businesses a set of strategic and tactical principles are established in a national architecture called NORA (Dutch governmental reference architecture).
A specialization of those architecture principles which focuses on municipal information needs are established in the GEMMA architecture. The GEMMA architecture covers architecture principles, information models, adapted standards, and practical guidelines for implementation of the the afore mentioned NORA (KING 2011). The GEMMA information architecture is implemented throughout the reference model of municipal base data (RSGB), which is related to the national system of key registrations. Figure 5.01 illustrates the relations between programmes as described above. Beside the system of key registrations domain specific information models for Geo-information are developed and maintained under the supervision of Geonovum. Geonovum is the National Spatial Data Infrastructure executive committee in the Netherlands and develops and manages the national Geo standards.

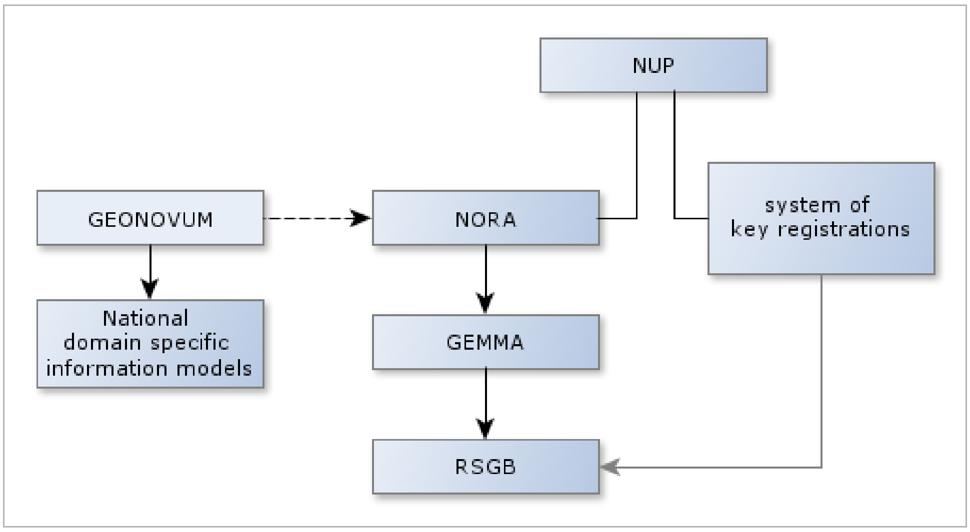


Figure 5.01. Relations between national programmes and system of key registrations.

The Dutch system of key registrations consists of 13 registrations, including 7 Geo-registrations. An important characteristic of key registrations is that they are grounded on legislation. Every key registration has its own legislation with accompanying laws.

The items 7 to 13 belong to the group key Geo-registrations. Not all the registrations which are mentioned in this list are not yet fully implemented. The planning of the implementation was started in April 2007 and the estimated finalization is planned in the year 2016 (MIM 2011).

1. Key registration of Citizens (GBA/BRP)
2. Key registration of non-residents (RNI)
3. New trade registry (NHR)
4. Key registration of income (BRI)
5. Key registration of salary, benefit, and employment relationships (BLAU)
6. Key registration vehicles (BRV)
7. Key registration addresses (BRA)
8. Key registration buildings (BGR)
9. Key registration land registration (BRK)
10. Key registration of properties and real estate value (BRWOZ)
11. Key registration topography (BRT)
12. Key registration large scale topography (BGT)
13. Key registration of the subsurface (BRO)

Figure 5.02 illustrates the system of key registrations, the relationships between key Geo-registrations, and the most relevant data elements which contain geometry. The registrations which are part of the key Geo-registrations are grouped within the green outline.

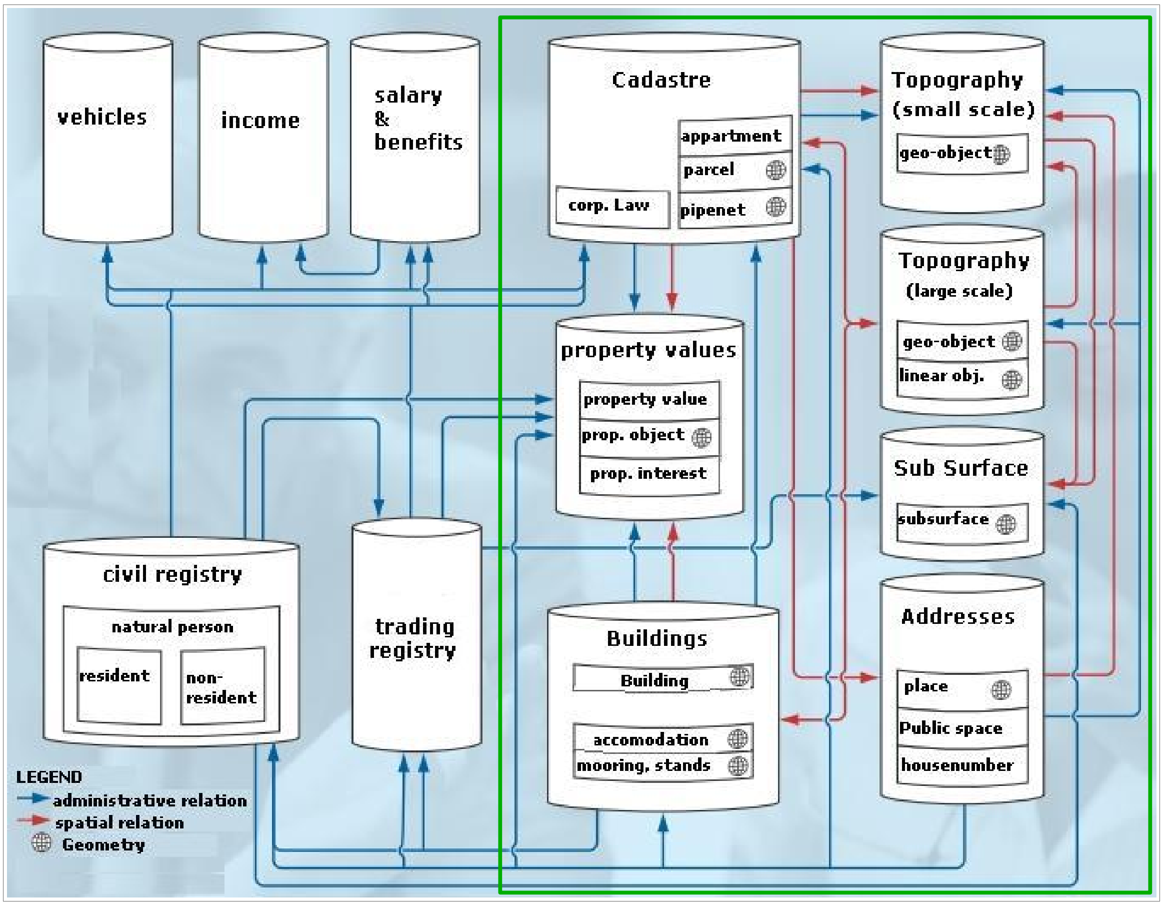


Figure 5.02. The Dutch system of key registrations after (MIM 2011).

5.2. Key Geo-registrations in municipal processes

Key Geo-registrations are widely used in spatially managed municipal processes, they form the foundation for the creation of derived data, the basis for added value, or as reference data for visualisation purposes. Beside the system of key registrations, national established information standards and municipal established registrations are used to support spatially managed municipal processes to carry out public tasks. These tasks are domain specific and require domain specific information models which are related to key Geo-registrations. Figure 5.03 depicts how domain specific data sets build on top of the system of key registrations and supports municipal processes.

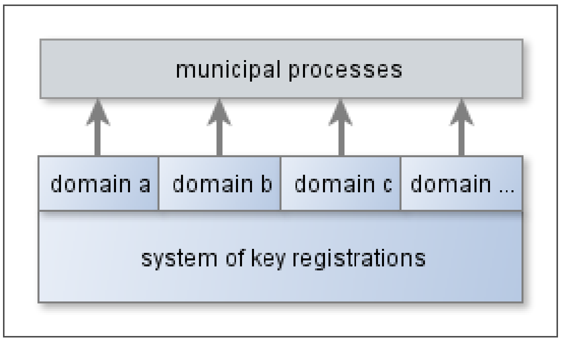


Figure 5.03. Key registrations and municipal processes.

The utilization of key Geo-registrations in the context of all municipal processes is too diverse to discuss in this research. How key Geo-registrations are used in municipal processes is explained from a data oriented perspective and from a process perspective.

The data oriented perspective.
As mentioned in the previous paragraph usage of key Geo-registrations might appear in different forms. The mind map in Figure 5.04 illustrates a taxonomy with three main groups how usage can be classified. Input for this mind map is based on de Smith et al (2007) and Longley et al. (2011a).

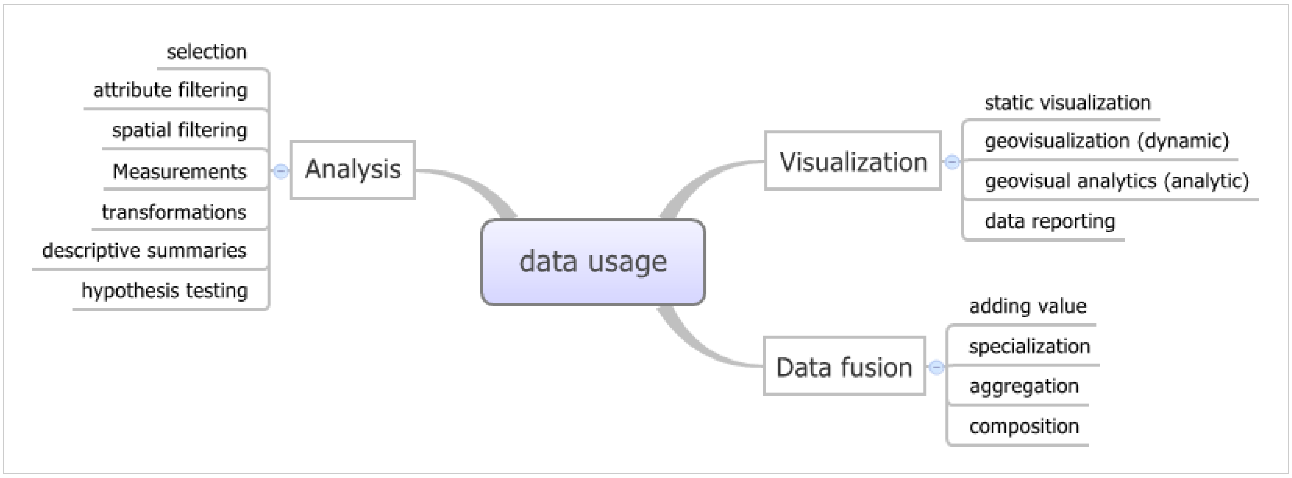


Figure 5.04. Municipal data usage.

The three groups of data usage are data fusion, visualisation, and analysis. Visualisation and analysis cover well-known methods used in Geographic information science. Data fusion in a municipal context will be further explained. When expert users "*adding value*" to a dataset (i.e. A key Geo-registration), domain specific data can be added to the dataset, or objects from both datasets can be explicitly linked. Examples of this kind of usage are linking cadastral parcels with ground lease records, linking buildings with maintenance plans, linking objects of small scale topography with maintainable objects to administer public space.

Specialization is applied when topographic objects from a thematic perspective are too generalized for the intended use or a specific task and require further division into more specific features. An example for specialization of objects is the maintenance of public green areas. In national large scale topographic data, green areas are acquired and registered only as 'green areas'. For the maintenance of public green more specific features are needed, for example groups of trees, grasses, bushes, and many more. *Aggregation* and *composition* are methods to create objects which are composed of multiple objects. When an object is composed out of objects with the same characteristics the aggregation operation is applied. Objects composed of objects with different characteristics the composition is applied.

The process oriented perspective.
Spatially managed municipal processes using both domain specific data, and data sets extracted from key Geo-registrations (base data). Depending on the purpose of use and information needs the base data and thematic data might be integrated an manipulated by data fusion, or combined as-is using visual overlay techniques. Details about those methods were discussed in the previous paragraph. Figure 5.05 illustrates an example of a municipal information chain.

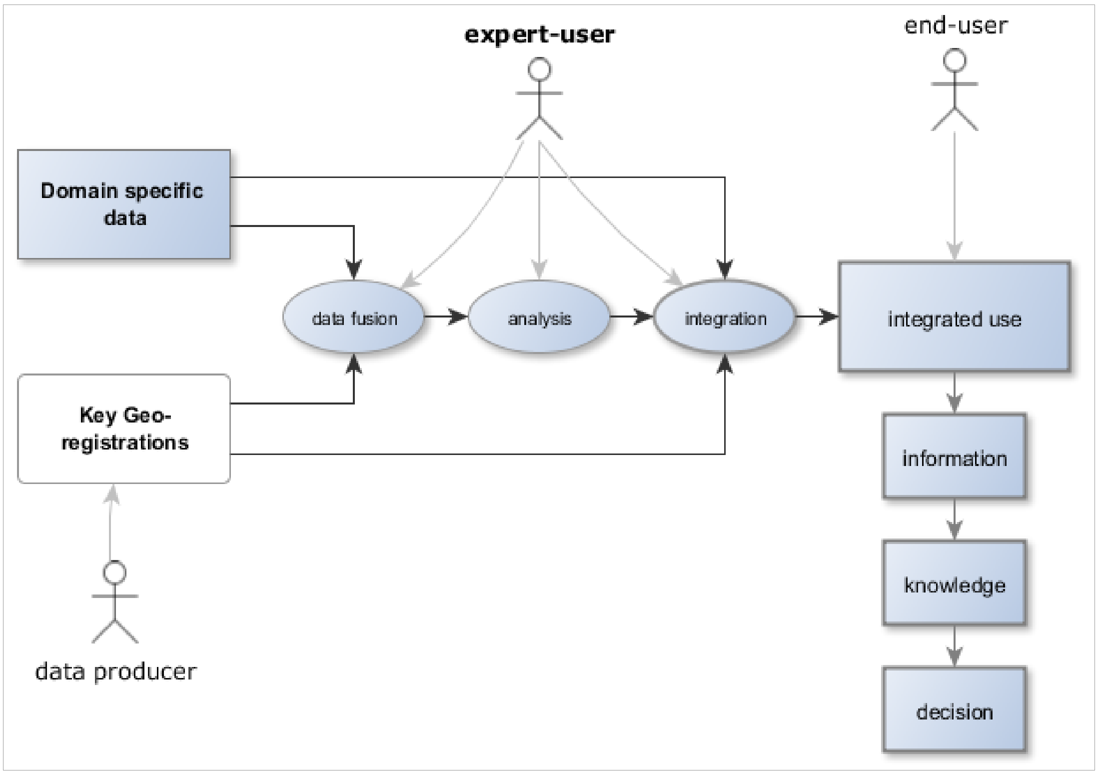


Figure 5.05. Municipal data chain.

The process of combining base data with thematic data and performing analysis is usually performed by expert users to prepare data for end users. The actual use, gathering information and translation to knowledge which are required to take strategic, tactical, or operational decisions is usually performed by end-users.
To exemplify how uncertainty might appear in a municipal information chain a concrete example of a municipal process will be discussed with the focus on data fusion.
The example use case discussed is the management of urban parks and green space.
The outline of the case is as follows; first the used data is briefly described, then flow of the used data related to the process is described. In this use case there will be referred how uncertainty might appear.

The initial situation in this example is an existing database or dataset which supports the process of managing urban parks and green space. The type of data usage is adding value and specialization. Figure 5.06 illustrates the process flow.

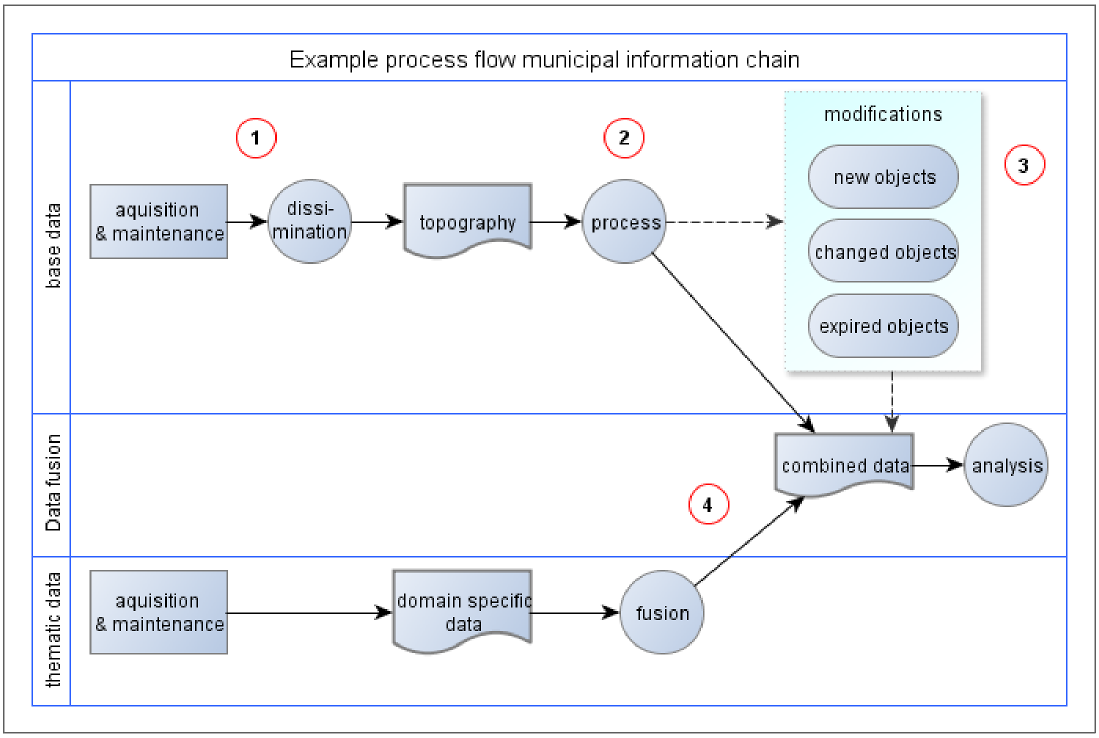


Figure 5.06. Example process flow municipal information chain.

Large-scale topography is maintained by another organisation which extracts and distributes products from the key Geo-registration large scale topography *(1)*.
The organisation which maintains the urban parks and green space, receives the updated topographic data and integrates the base data with domain-specific data in their own back-office system *(2)*. An important step in the process is determining the impact of the updated topography. The updated topography may consist of new objects, changed objects, or expired objects *(3)*. At mark *3* ambiguity may arise, the question might be asked is: when is an object perceived as new, changed or expired?
There might be a difference of perception between the data creator of the topographic data and the user. When data acquisition rules are ambiguously defined or not known by the user, the user might have a different interpretation of the modifications. This might imply omission or commission in the processed dataset. When the impact analysis with respect to the domain specific content is performed, the topographic modifications are implemented into the thematic dataset through data fusion *(4)*. This implementation consists of two types of operations; adding value through linking maintenance data with topographic(green) objects, and further specialization of topographic objects. At mark *4* where new 'green' objects are formed all main aspects of uncertainty might occur:

- Ambiguity: Is this topographic object a green object, or just unpaved terrain?
- Vagueness: How accurate is the boundary of this green area determined, and is it all green or is it a mixture of different types of land cover?
- Inaccuracy: What is the positional accuracy of the boundaries, how accurate are the topographic green objects classified?

After the data fusion procedure the expert-user probably wants to make a statement of the result of the data fusion, therefore understanding of uncertainty in spatial data might improve the insight of the limitations of the data, and the conformance of the data to its data quality specifications drafted by the user.

6. From theory to empiricism

The preceding chapters three, four, and five covered the literature review of this research. This chapter aims to evaluate the literature review, refines these theories to specific concepts, and operationalizes these concepts to measurable indicators. These indicators form the base for the design of the survey and the in-depth interviews.

6.1. Evaluation literature review

The literature review discussed four topics:

1. Concepts of uncertainty in spatial data.
2. Explore how information about uncertainty in spatial data can be communicated.
3. Investigate previous conducted research about the usefulness of visualizing uncertainty.
4. Explore the use of Key-Georegistrations in a municipal context

The following paragraphs provide a summary of the findings of the literature review ordered by topic.

6.1.1. Concepts of uncertainty in spatial data

Available literature discussed different definitions and different taxonomies of uncertainty which are discussed in chapter three. A notifiable finding is the co-existence of different taxonomies on uncertainty in spatial data established by Shi (2010), Fisher (1999), and Shu et al (2003). Remarkable is the limited available research on the relationship between uncertainty and spatial data quality. Section 3.3 discussed the implemented taxonomy of uncertainty in spatial data. The concepts discussed in this taxonomy are operationalized the survey and the in-depth interviews.

6.1.2. Communication of uncertainty in spatial data.

Communication about uncertainty and visualizing uncertainty in particular has been researched since 1988. Chapter four discussed conducted research which are relevant for this research. Despite the extensive methods and theories that have been are developed the in the last years, it is remarkable that these theories and methods are applied in a limited extent in the different fields of Geo-information in a municipal context. Concepts of importance are the type of user interaction with the visualization, and the type of visualization of information about uncertainty in spatial data.

6.1.3. Previously conducted research: usefulness of visualizing uncertainty.

The study of earlier studies yielded some interesting concepts regarding communicating information about uncertainty in spatial data. The concepts concerned cognitive aspects, different aspects of tasks, and perception.

Cognitive overload
Various studies indicate that the complexity of data and how uncertainty is visualized might affect the perceived ease of use of information about uncertainty (Sanyal et al. 2009). The inherent complexity of data and depiction of uncertainty might cause perceptual and cognitive confusion (Boukhelifa & Duke 2009; Viard, Caumon & Lévy 2011). But when this complexity is taken into account when designing visualizations, the user is exempt from cognitive overload it might add value.

Context and task dependent
According to Sanyal et al. (2009) and Boin (2008), the chosen method of a visualization method should depend on the context. The effectiveness of the chosen visualization methods also depends on the task performed by the user.

Task and domain experience
Roth (2009) investigated the impact of expertise on the use of uncertain geographic information using a case study about risk assessment. Roth found that the confidence of the user might be affected by domain experience and the users skills of map-use. The level of domain experience influenced the risk assessment itself. The level of map-use skills dominated the impact on the perceived difficulty of the uncertainty visualizations.

Irrational perception
The study of Hope (2005) provided strong evidence that the form and type of representation of uncertainty in spatial data affects decision-making. According to Hope "decision-makers do not necessarily respond in a rational manner to the inclusion of uncertainty information and can make significantly different decisions dependent upon the form by which the information is portrayed".

6.1.4. The use of key Geo-registrations in a municipal context

Information about uncertainty in spatial data of key Geo-registrations in municipal chain processes is an important factor in the scope of the study. In order to answer the main research question it is essential to understand the nature of use of key Geo-registrations in a municipal context. Key Geo-registrations are related in different ways to domain specific data. The desk research showed that key Geo-registrations are the foundation for spatially managed processes.Uncertainty is inherent in spatial data (Longley et al. 2011b), and therefore occurs in various ways and on various steps in these chain processes. The different types of users with different backgrounds and perceptions, with different tasks, needs, concerns, and interests can be of importance to make a statement about the usefulness of information about uncertainty in spatial data. As a significant link in the municipal information chain the prementioned (expert) users are topic if investigation how they notify, classify, quantify (Reinke & Hunter 2002), and handle uncertainty in spatial data.

6.2. Operationalization

Three important concepts are extracted from theory. The definitions and adapted taxonomy of uncertainty are organized in concept (**1**). This concept provides uniformity and consistency of used terminology and definitions throughout the research. Elements of the adapted taxonomy of uncertainty are operationalized into measurable variables regarding information needs about uncertainty in spatial data. Usefulness of information about uncertainty is organized into concepts (**2**). These concepts are conceptualized in personal characteristics, task characteristics, information needs, and type of visualization. Due to the general nature of this concept the operationalized variables shall be measured using a survey. Perceived usefulness in a municipal context is organized in concept (**3**). This concept is operationalized into qualitative indicators, namely notification(awareness), perception (experience as a problem), quantification (how 'big' is the problem) and its possible propagation), and handling methods (how to deal with uncertainty), and which information is preferred and which information about uncertainty is available. Due to the qualitative nature the operationalized variables are measured using in-depth interviews. Figure 6.01 illustrates the relationships between theory, concepts, and variables. The elaboration of the operationalization from concepts to measurable variables is discussed in the following paragraphs organized by research method, survey and in-depth interview.

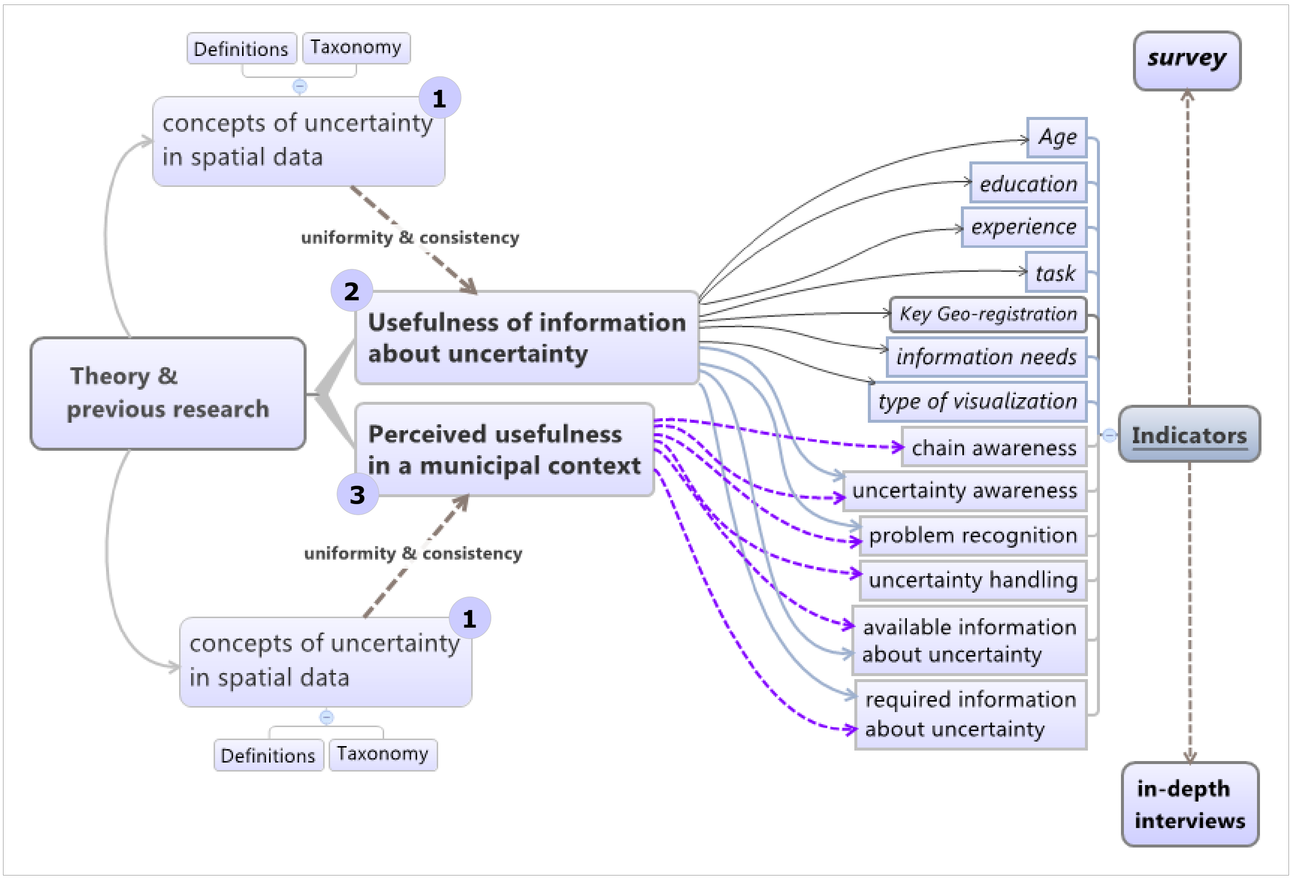


Figure 6.01. Conceptualized theory.

6.2.1. Usefulness of information about uncertainty

To measure usefulness of information about uncertainty in spatial data measurable variables are defined using the technology acceptance model (TAM). These variables are further specialized to implement the psychometric Likert scale used in psychology (Trochim 2006). In the questionnaire, several well known criteria for formulating question text (Bethlehem 2009) are implemented to reduce observation errors. The questions for the questionnaire are developed to fit in the construct which is based on the technology acceptance model (Venkatesh & Davis 2000). The technology acceptance model (TAM) which is established by Fred Davis (1989) is an extension of the theory of reasoned action. The theory of reasoned action (TRA) is a fundamental theory of human behavioural studies (Davis et al. 2003). The underlying concept of user acceptance models is based on three core constructs:

- Individual reactions to using information technology
- Intentions to use information technology
- Actual use of information technology

Davis adapted TRA and extended the model with the constructs *perceived usefulness* (PU), *perceived ease of use* (PEU), which affect the behavioural intention of the user to the actual system use. PU and PEU are hypothesized to be directly affect by external variables.
These external variables are constructed from system characteristics. Figure 6.02 illustrates the final version of the technology acceptance model as defined by Venkatesh and Davis (1996).

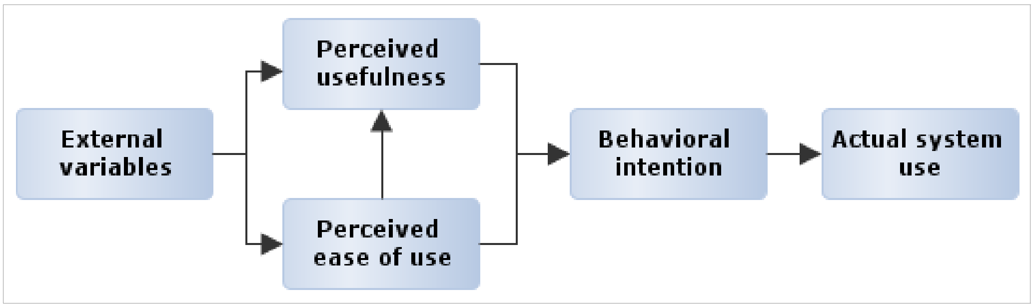


Figure 6.02. Final version of TAM ,(Venkatesh and Davis 1996).

The Technology Acceptance Model is applied to measure the usefulness of communicating about uncertainty in spatial data. The external variables are constructed from system characteristics which are composed of variables divided into four groups:

a) Personal characteristics; age, education,experience.

b) Task characteristics; the tasks which are carried out frequently by the respondents, and the registrations used to carry out their tasks.

c) Uncertainty information needs; required information needs of the respondents about uncertainty is spatial data.

d) Type of visualization; the form of how information about uncertainty is visualized

Figure 6.03 illustrates the implementation of the external variables in the technology acceptance model.

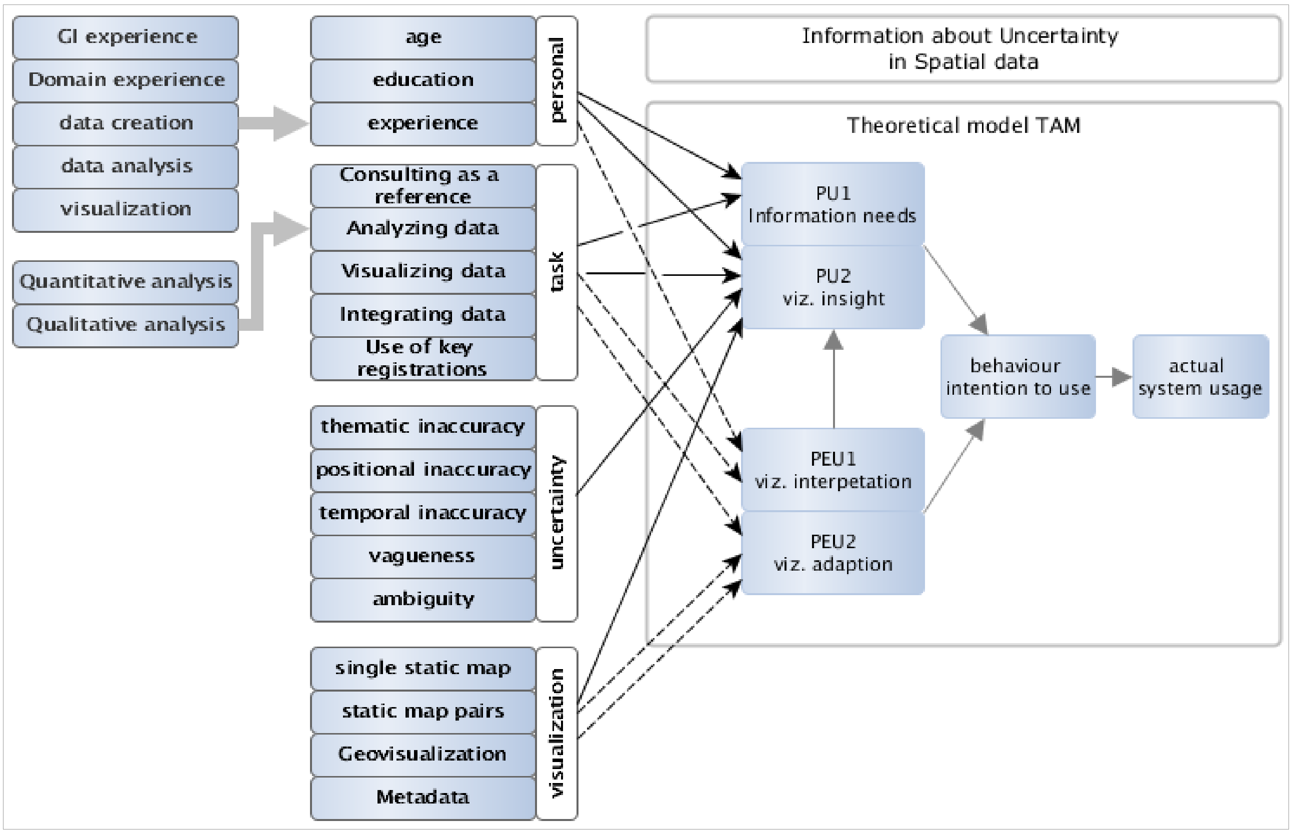


Figure 6.03. Measurable independent variables and TAM.

To explore perceived usefulness and perceived ease of use more in detail PU is divided into PU1 and PU2, and PEU is divided into PEU1 and PEU2.

PU1 : perceived information needs
Constructed psychometric scale items based on the adapted taxonomy of uncertainty in a municipal context.

PU2 : perceived insight in map accuracy through visualization
Constructed psychometric scale items which indicates the perceived informational value of visualizing map accuracy.

PEU1 : perceived Interpretation of visualization
Constructed psychometric scale items which indicates the PEU of visualization of uncertainty. This indicates how well the respondents can interpret the visualizations demonstrated or described in the questionnaire.

PEU2 : Perceived adaption
Constructed psychometric scale items which indicates the perceived adaption of visualization of uncertainty into the tasks of the respondents. This indicates how well the respondents can adapt the visualizations demonstrated or described in the questionnaire into their own working procedures.

Each explanatory variable of this construct is based on the median of a predefined group of response (measured) variables. Table 6.01 gives the composition of the explanatory variables.

response variable	description of response variable	explanatory variable
inf13_ambiq	The level of information needs about ambiguity	**PU1**
inf14_vague	The level of information needs about vagueness	**PU1**
inf15_posacc	The level of information needs about positional accuracy	**PU1**
inf16_attracc	The level of information needs about attribute accuracy	**PU1**
inf17_time	The level of information needs about temporal consistency	**PU1**
viz18_single_ins	The level of insight in map accuracy using a single static map	**PU2**
viz21_pair_ins	The level of insight in map accuracy using a static map pair	**PU2**
viz24_dyn_ins	The level of insight in map accuracy using dynamic visualization	**PU2**
viz27_mta_ins	The level of insight in map accuracy using metadata	**PU2**
viz28_mta_defi	Level of insight in ambiguity using metadata	**PU2**
viz19_single_interp	The ease of use interpreting uncertainty on a single static map	**PEU1**
viz22_pair_interp	The ease of use interpreting uncertainty on a static map pair	**PEU1**
viz25_dyn_interp	Dynamic visualization is preferred over static visualization	**PEU1**
viz29_mta_compli	The ease of use interpreting uncertainty using metadata	**PEU1**
viz20_single_adapt	The level of adaption of a single static map into current tasks	**PEU2**
viz23_pair_adapt	The level of adaption of a static map pair into current tasks	**PEU2**
viz26_dyn_adapt	The level of adaption of dynamic visualization into current tasks	**PEU2**
viz30_mta_adapt	The level of adaption of metadata into current tasks	**PEU2**

Table 6.01 . The composition of the explanatory variables

The variables of the construct are operationalized into the survey questionnaire. The questionnaire consists of 30 questions and are measured on an ordinal scale from 0 - 5. Each question represents a measured variable which are used to for further analysis. The construct contains connectors which represent possible relationships. These possible relationships are operationalized into the following hypotheses to test:

1. Personal characteristics affects perceived usefulness
2. Personal characteristics affects perceived ease of use
3. Task characteristics affects perceived usefulness
4. Task characteristics affects perceived ease of use
5. Information needs affects perceived usefulness
6. Information needs affects perceived ease of use
7. The type of visualization affects perceived usefulness
8. The type of visualization affects perceived ease of use
9. The registration used affects perceived usefulness
10. The registration used affects perceived usefulness

The assumed relationships are examined for strength by using correlation analysis, the effect of the response variables on the explanatory variables are analysed by regression analysis.

6.2.2. Perceived usefulness of information about uncertainty

The emphasis of the concept of perceived usefulness of information about uncertainty is on the perception of expert-users. This part of the research has the goal to investigate *why* the expert-users formed their opinion about usefulness of information about uncertainty in spatial data. To achieve the previous mentioned goal, the concept is operationalized into a thematized in-depth interview with a qualitative approach. The themes are based on the desk research discussed in chapter four (Research concerning usability of information about uncertainty), & chapter five (key Geo-registrations in a municipal context). Hence, these themes are illustrated in Figure 6.04 and are explained in the next paragraph.

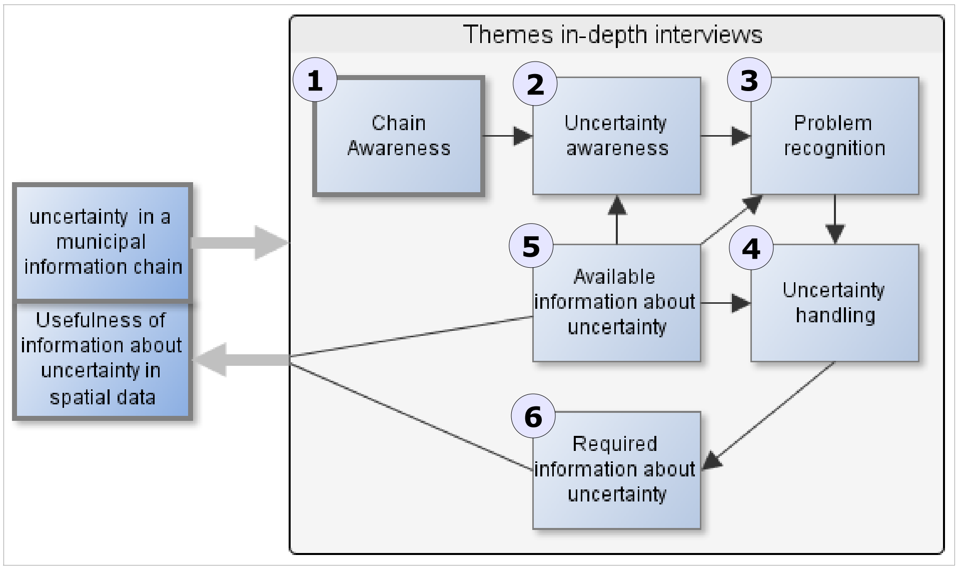


Figure 6.04. Interrelated themes in-depth interviews.

1. Chain awareness
Users should recognize that their activities are part of a municipal information chain, and that their activities constitute a junction in the municipal information system to facilitate primary processes.

2. Uncertainty awareness
Users should understand concept of uncertainty in spatial data and recognize uncertainty as a problem or an issue which should be taken into account.

3. Problem recognition
Based on the work of Reinke & Hunter(2002) users should notify, classify, visualize, and evaluate uncertainty in the context of their tasks.

4. Uncertainty handling
Depending of the context, business demands, and available resources uncertainty can be handled as followed (Vullings et al. 2009; Hope 2005) :

a. Ignore uncertainty in spatial data
b. Incorporate expert knowledge to understand the limitations
c. Make adjustments in the working procedures to reduce the effects of uncertainty
d. Reducing uncertainty through investigation, or inquiry by the data producer
e. Self-improving of the data; acquire missing data, or improve accuracy

5. Available information about uncertainty
The ability to become aware of uncertainty in spatial data, recognize possible problems, and handling strategy, information about uncertainty should be available. The purpose of this theme is investigate how the current availability of information about uncertainty is experienced.

6. Required information about uncertainty

To gain new insights in required information about uncertainty in spatial data is investigated. The focus on this theme is to discover possible relationships between required information about uncertainty and the context of the use of spatial data.

Relationships between the themes
The themes are interrelated as illustrated in Figure 6.04. The statement in this research is that expert-users should be aware of a municipal information chain, and that data 'flowing' through this chain contains uncertainties. To recognize possible problems the users should have be aware of uncertainty in spatial data to recognize this phenomena. When uncertainty in spatial data is identified and classified, it can be quantified. To quantify uncertainty, quantitative information about uncertainty is required. With available information about uncertainty the user should be able to handle uncertainty. To test theory and practice, the available information about uncertainty is investigated, and required information about uncertainty is investigated

7. Research results & discussion

This chapter reports and discusses the results of the conducted research. Section 7.1 covers the findings of the survey, Section 7.2 covers the findings of the in-depth interviews.

7.1. Results survey

The questionnaire of the survey aimed to obtain data to answer two research sub-questions:

(5) What are the information needs about uncertainty in spatial data?

(7) Which factors affect the usefulness of information about uncertainty in spatial data?

Section 7.1.1 describes the response. The descriptive statistics which are used for sub-question 5 are reported in Section 7.1.3. The results of investigating possible relationships is described in Section 7.1.4.

7.1.1. Survey response

The intended population are GIS specialists which are employed by large municipalities. The size of the population is estimated on 390 which is based on the report of the Dutch business association of Geo-information (Geobusiness Nederland 2010).

Target population (large municipalities) : **390**
Sample : ***100***
Response : ***44***

The response rate is computed as : $\frac{nR}{nR+nNR}$ (Bethlehem 2009)

Where *nR* is the number of responses, and *nNR* the number of non-reponses.

Hence, the computed response rate is : $\frac{44}{44+56}$ = **44**%

To improve the validity of the outcomes of the survey, the respondents should be a reliable representation of the population. The questionnaire contains elements which covers personal characteristics of the respondents as stated in Section 2.3.2. The sample size is computed with the a given error probability of 5% ,a preferred effect size of 0.5, and the number of predictors (Erdfelder et al. 1996).

The sample is assumed to be representative when there is an equal distribution of personal characteristics of the respondents. The quantification of the variability of the responses is expressed in Simpson's diversity index (Mcdonald & Dimmick 2003). A perfect homogeneous sample has a diversity index of 1, and a perfect heterogeneous sample has a diversity index of 0. Simpson's index takes the number of categories into account, the number of absent categories, and depends on the sample size. Due to the sample size of this survey a perfect heterogeneous sample has a diversity index of 0.8.
Simpson's diversity index is calculated as:

$$\boldsymbol{D}=1-\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{p}^2 \boldsymbol{i}$$

Where ***n*** is the number of categories, ***p*** is the proportion of responses in a category, and ***i*** is the index.

A detailed summary of the results of the questionnaire is provided in Appendix 5. The summary in the subsequent section provides an overview of the distribution and the central tendency of the response.

7.1.2. Summary of the personal characteristics of the respondents

The purpose of this section is to provide an overview of the response. Figure 6.01 provides the overview the distribution and variation of the respondents using a pie chart for each variable. According the the outcomes the majority of the respondents is middle aged or above, well-educated and has more than 5 years experience working with geographic data. The level of experience of the respondents if investigated for creating geographic data, performing spatial analysis, interactive visualization if geographic data, and the level of domain experience. From the outcomes can be deduced that the majority of the respondents is above the level of novice and more than half of the respondents has the level of very experienced or expert. These findings enhance the representativeness of the survey because they match the profile of the expert users as described in Figure 1.04.

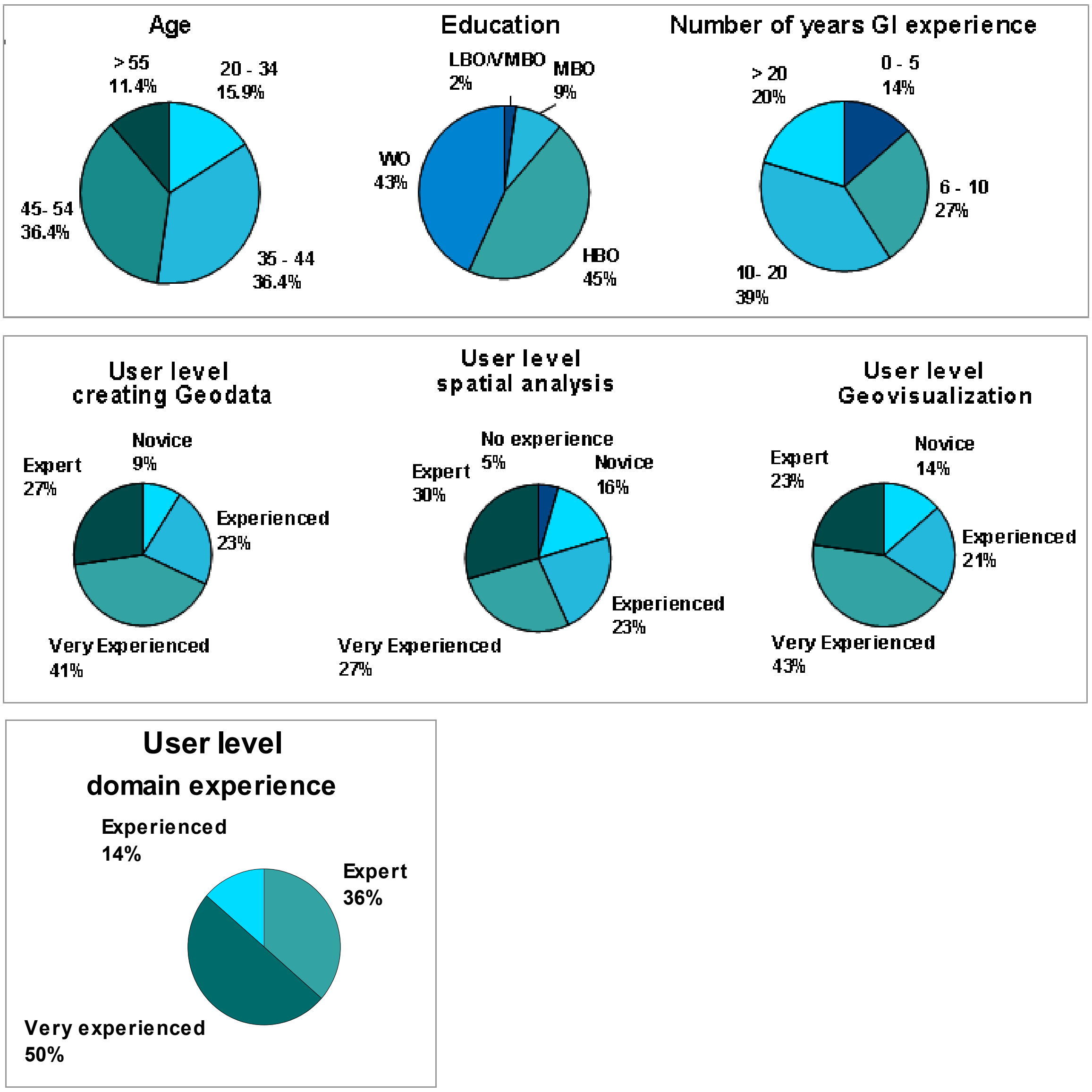


Figure 7.01. Distribution of personal characteristics.

7.1.3. Information needs about uncertainty in spatial data

The outcomes of the questionnaire showed that respondents appeared to be positive about information about uncertainty in spatial data. Figure 7.02 represents the results of the questionnaire to what extent the respondents agree about the usefulness of information needs about uncertainty in spatial data. The majority of the respondents rated the questions "agree" or "strongly agree". It is clear that Figure 7.02 shows vagueness has the largest response, and temporal consistency the smallest response. Despite the lowest rank, the smallest response represents 75% of the respondents.

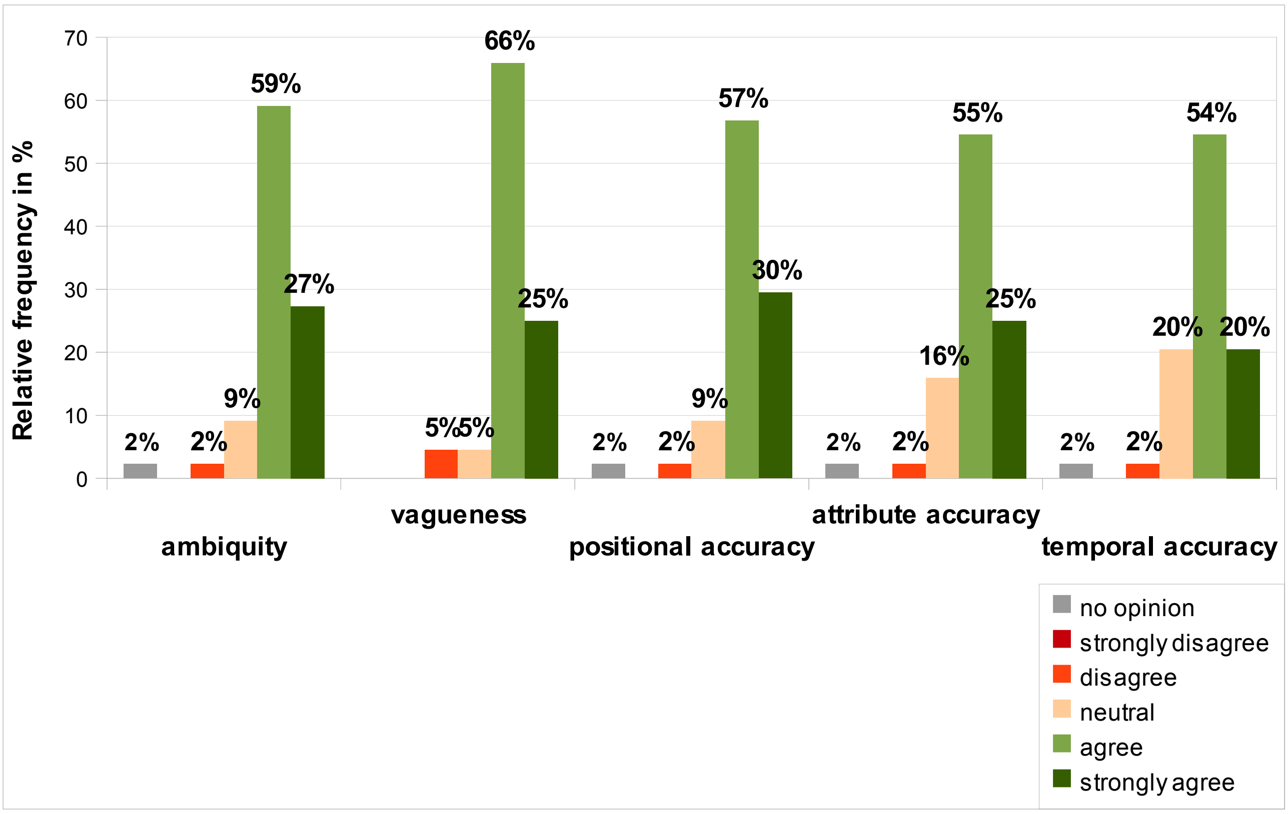


Figure 7.02. Survey response of information needs about uncertainty.

The proper centrality measures for ordinal data are the median and mode (Berenson & Levine 1992). For each element the calculated median is equals 4 (agree). The mode slightly differs. The outcomes of the central tendency is also provided in Table 7.01.

Category	min	max	median	mode	IQR	diversity index
Ambiguity	0	5	4	4	1,00	0,567
Vagueness	2	5	4	3	0,25	0,499
Positional accuracy	0	5	4	4	1,00	0,581
Attribute accuracy	0	5	4	4	0,25	0,614
Temporal accuracy	0	5	4	4	0,25	0,618

Table 7.01. Centrality measures for information needs about uncertainty in spatial data

From the graph and tables can also be deduced that the distribution and diversity of the response differ. The diversity is calculated according to Simpson's diversity index (Mcdonald & Dimmick 2003). Information needs about temporal inconsistency has the highest rate of diversity, information needs about vagueness has the lowest rate of diversity.

7.1.4. Analysis of possible relationships

Approach
To investigate which factors affect the usefulness of information about uncertainty in spatial data several statistical tests are performed to investigate possible relationships between dependent and independent variables as illustrated in Figure 6.03 in chapter six. The analysis performed uses multi-level constructs which are operationalized in statistical models. The external variables in the TAM are the first level and are generalizations of level two which represent the measured variables in the questionnaire. An overview of the performed analysis and statistical tests is given in Appendix 7. This section discusses the statistical tests which are significant. Rank order correlation is used to to investigate the strength of relationships, multi regression analysis and ordinal regression are used to investigate relationships between dependent and independent variables. The starting point to use different regression methods is based on the statistical principle that a statistical model should fit the data and not vice versa (Berenson & Levine 1992). Depending of the characteristics of the data and assumptions of the statistical methods the most optimal combination of data and fitting model is chosen to perform the analysis. The statistical models are implemented using the software package R (R Core Team 2012b).

Structure of this section
This section is organised from general to more specific. First a summary of the general results is described to provide a holistic overview of the results. This summary is followed by a more detailed analysis of the relationships in the Sections 7.1.4.2 to 7.1.4.5.
The final part of this section consists of the descriptions of the results of the rank-order correlation, the regression analysis, and the corresponding statistical models.

7.1.4.1. General results of the analysis of the relationships

This section attempts to provide insight in the distribution and partition of the independent variables in respect to the dependent variables. In order to achieve this, the significant results of the rank-order correlation analysis and the regression analysis are visualized in Figure 7.03. The independent variables are arranged into 5 groups; personal characteristics, user tasks, the use of registrations, information needs about uncertainty, and type of visualization. The dependent variables PU1, PU2, PEU1, and PEU2 are depicted in blue. The coloured connectors indicate at a general level which relationships are significant. The coloured bars attached indicate which dependent variables are affected by which independent variables.

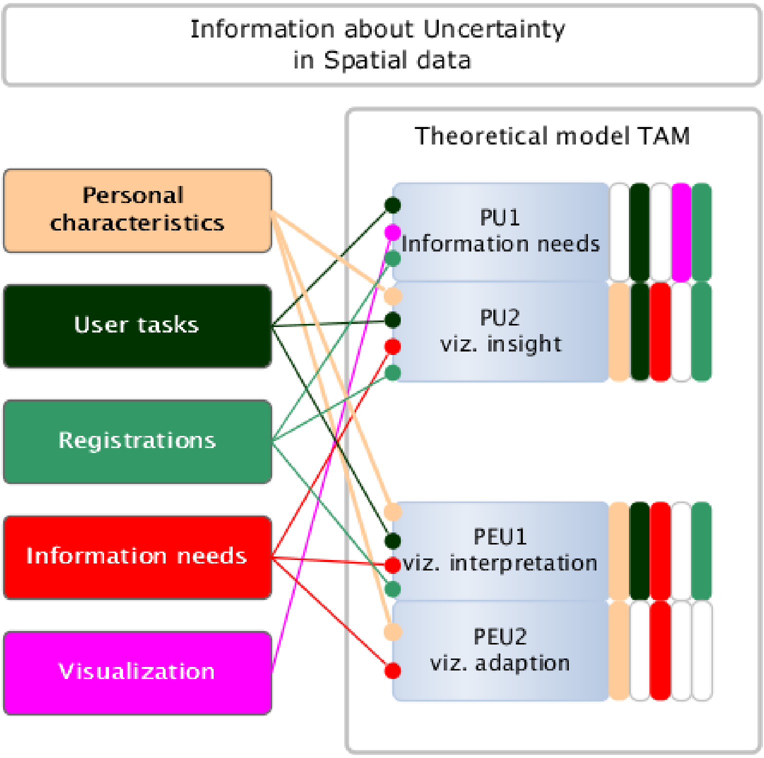


Figure 7.03. Generalized results analysis of relationships.

On a general level, personal characteristics, user tasks, the use of registrations, information needs about uncertainty have relationships with three dependent variables. The type of visualization has a relationship with PU1.

From statistical analysis is derived that a number of variables affect the dependent variables. Figures 7.04 and Figure 7.05 give an overview of the most influential variables.
The variables are arranged into groups; information needs, personal characteristics, tasks, and type of visualization. An additional distinctive element is 'Registrations'. This variable represents the the group 'Used Registrations' which may affect PU or PEU.

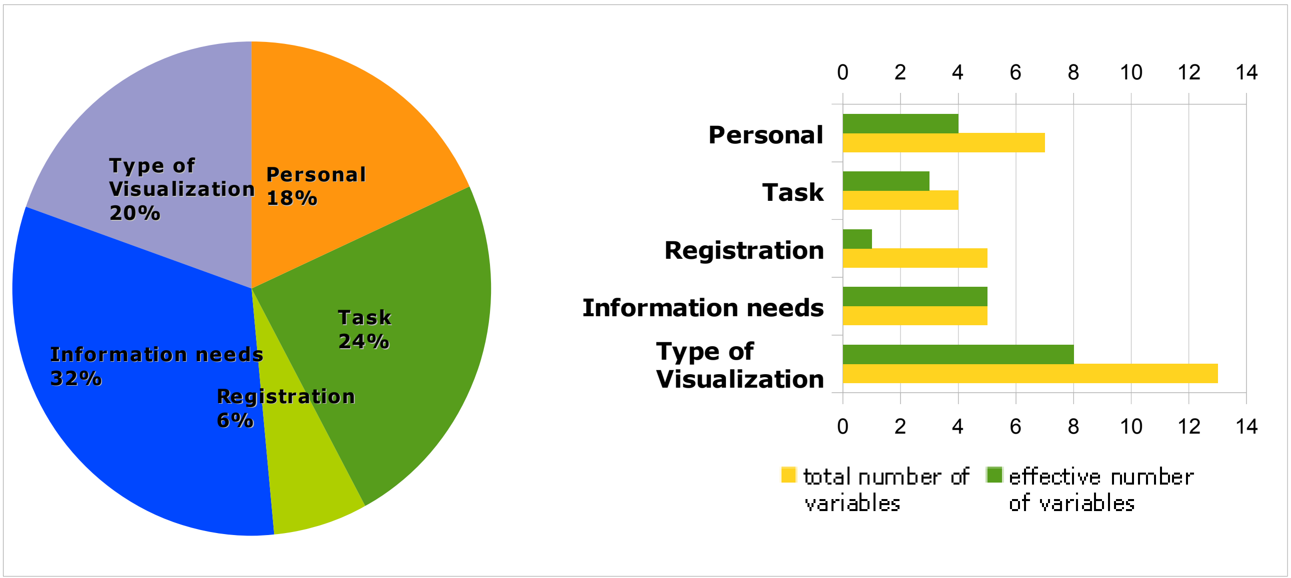


Figure 7.04. Generalized results of the efficacy of variables.

From the charts in Figure 7.04 can be deduced that variables from the group information needs are the most influential. The bar chart on the right side in figure 7.04 provides insight in the ration of the number of variables measures and effective variables. To investigate the differences in influence between the independent variables the number of affected variables are presented in the bar chart in Figure 7.05. The visualization type metadata seems affect the largest number of variables.

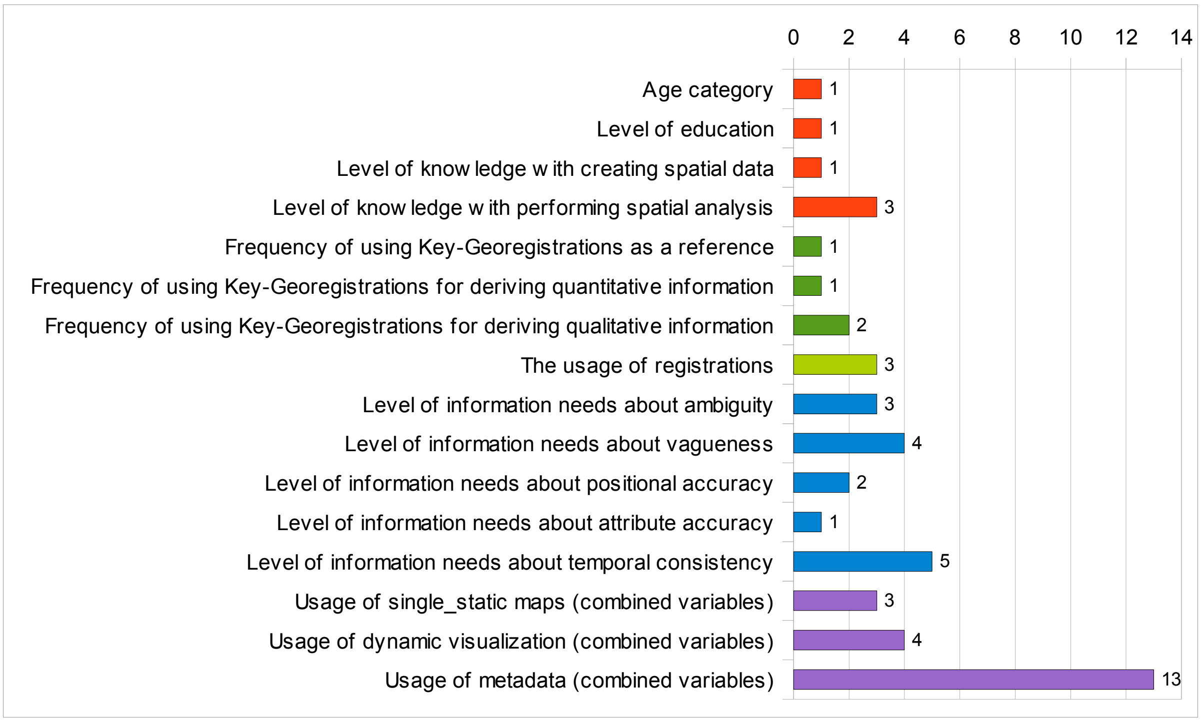


Figure 7.05.Overview of the effectiveness of independent variable.

7.1.4.2. Effect on perceived Usefulness (PU1)

This section describes the interpretation of the results of the analysis on the effect on PU1. The diagram in Figure 7.06 gives and overview of the distribution of the variables ordered by category. Each arrow indicates if PU1 is affected by on or more variables. The typology of the arrow indicates if the predicted relationship is derived from correlation or regression. The predicted relationships are listed in Table 7.02.

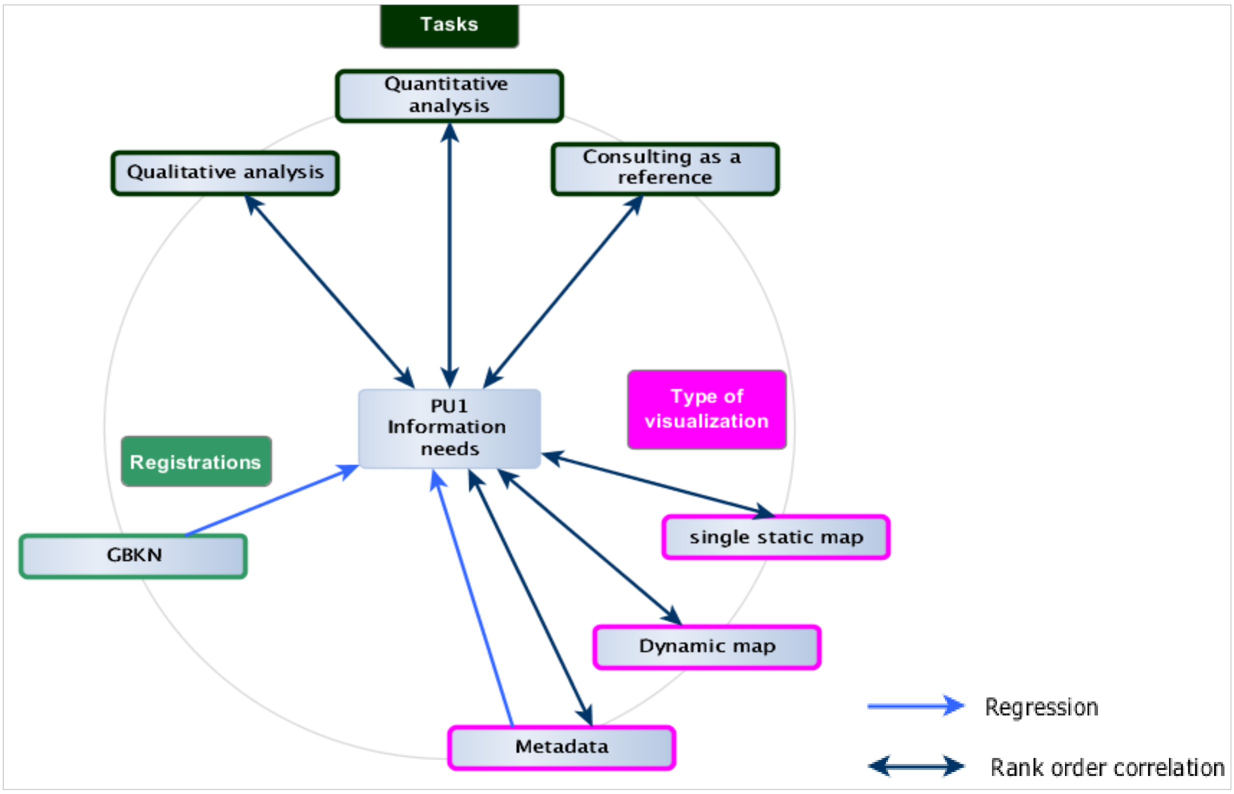


Figure 7.06. Effect on perceived usefulness PU1.

independent	dependent	Ro	P-value	method
Task_analysis-qualitative	PU1_positional-accuracy	0,382	0,011	Rank-order correlation
Task_analysis-quantitative	PU1_positional-accuracy	0,331	0,028	Rank-order correlation
Task_reference	PU1_positional-accuracy	0,374	0,012	Rank-order correlation
viz_dyn_ins	PU1_ambiguity	0,325	0,031	Rank-order correlation
viz_mta_adapt	PU1_temporal-accuracy	0,359	0,017	Rank-order correlation
viz_mta_adapt	PU1_vagueness	0,446	0,002	Rank-order correlation
viz_mta_definitions	PU1_temporal-accuracy	0,348	0,021	Rank-order correlation
viz_mta_ins	PU1_ambiguity	0,335	0,026	Rank-order correlation
viz_mta_ins	PU1_attribute-accuracy	0,370	0,013	Rank-order correlation
viz_mta_ins	PU1_vagueness	0,415	0,005	Rank-order correlation
viz_mta_interp	PU1_positional-accuracy	0,343	0,023	Rank-order correlation
viz_mta_interp	PU1_temporal-accuracy	0,360	0,016	Rank-order correlation
viz_mta_interp	PU1_vagueness	0,329	0,029	Rank-order correlation
viz_single_adapt	PU1_positional-accuracy	-0,335	0,026	Rank-order correlation
viz_single_adapt	PU1_temporal-accuracy	-0,306	0,043	Rank-order correlation
viz_single_ins	PU1_vagueness	0,322	0,033	Rank-order correlation
v_mta	PU1_all	-	0,021	Ordinal regression
registration_gbkn	PU1_all	-	0,029	Ordinal regression

Table 7.02. Effect on perceived usefulness PU1

The Spearman's Rank order tests indicate a moderate association between tasks performed by respondents, Type of visualization, used registration, and perceived usefulness. A notifiable finding is the fact the metadata has the highest number of relationships with PU1, and the strongest relationship discovered is the relationship between perceived adaptability of metadata and PU1. Users which perform the tasks quantitative and qualitative analysis and use large scale topography as a reference and perceive information about uncertainty useful when this information is visualized using metadata, single static map, and a dynamic map. Thematic accuracy does probably not influence any dependent variables.

7.1.4.3. Effect on perceived usefulness, insight (PU2)

This section describes the interpretation of the results of the analysis on the effect on PU2. The diagram in Figure 7.07 gives and overview of the distribution of the variables ordered by category. Each arrow indicates if PU2 is affected by on or more variables. The typology of the arrow indicates if the predicted relationship is derived from correlation or regression. The predicted relationships are listed in Table 7.03.

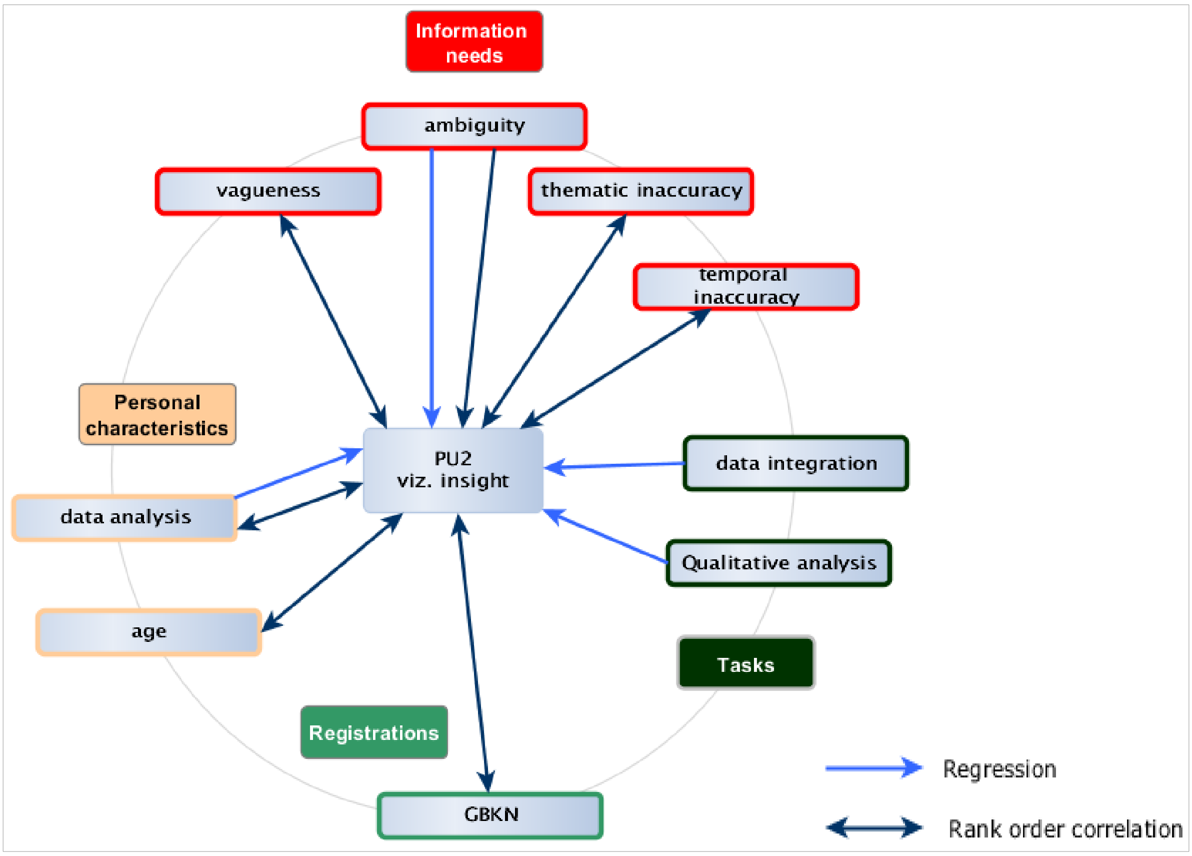


Figure 7.07. Effect on perceived usefulness PU2.

	independent	dependent	Ro	P-value	method
	Age	PU2_viz_mta_definitions	0,309	0,041	Rank-order correlation
	Experience_data-analysis	PU2_viz_mta_ins	0,353	0,019	Rank-order correlation
	Info_ambiguity	PU2_viz_dyn_ins	0,325	0,031	Rank-order correlation
	Info_ambiguity	PU2_viz_mta_ins	0,335	0,026	Rank-order correlation
	Info_vagueness	PU2_viz_single_ins	0,322	0,033	Rank-order correlation
	Info_vagueness	PU2_viz_mta_ins	**0,415**	0,005	Rank-order correlation
	Info_attribute-accuracy	PU2_viz_mta_ins	0,370	0,013	Rank-order correlation
	Info_attribute-accuracy	PU2_viz_mta_ins	**0,419**	0,005	Rank-order correlation
	Info_temporal-accuracy	PU2_viz_mta_definitions	0,348	0,021	Rank-order correlation
	registration_gbkn	PU2_viz_dyn_ins	0,361	0,016	Rank-order correlation
	task11_freq_qualit, task12_freq_integr	viz21_pair_ins,viz24_dyn_ins	-	0,020	Ordinal regression
	pers06_ana_exp	viz27_mta_ins , viz28_mta_defi	-	0,011	Ordinal regression
	Info_ambiguity	viz21_pair_ins,viz24_dyn_ins, viz27_mta_ins	-	0,036	Multi regression analysis

Table 7.03. Effect on perceived usefulness PU2

Respondents which are experienced with data analysis and which perform qualitative data analysis and perform data integration with large scale topography, with information needs about ambiguity, vagueness, thematic accuracy and temporal accuracy might find information about uncertainty useful. Especially in metadata. There is a correlation between age and insight in definitions using metadata. When age increases, the perceived usefulness increases.

7.1.4.4. Effect on perceived ease of use, interpretation (PEU1)

This section describes the interpretation of the results of the analysis on the effect on PEU1. The diagram in Figure 7.08 gives and overview of the distribution of the variables ordered by category. Each arrow indicates if PEU1 is affected by on or more variables. The typology of the arrow indicates if the predicted relationship is derived from correlation or regression. The predicted relationships are listed in Table 7.04.

According to the rank-order correlation analysis the relationship between use of large scale topography, cadastral data ,and interpretation of dynamic visualizations is the strongest. The level of experience with data analysis and data creation affect PEU1, as well as the level of qualitative analysis. The relations between information needs and PEU2 may indicate which information needs are easier to interpret when the values of the previously mentioned independent variables change.

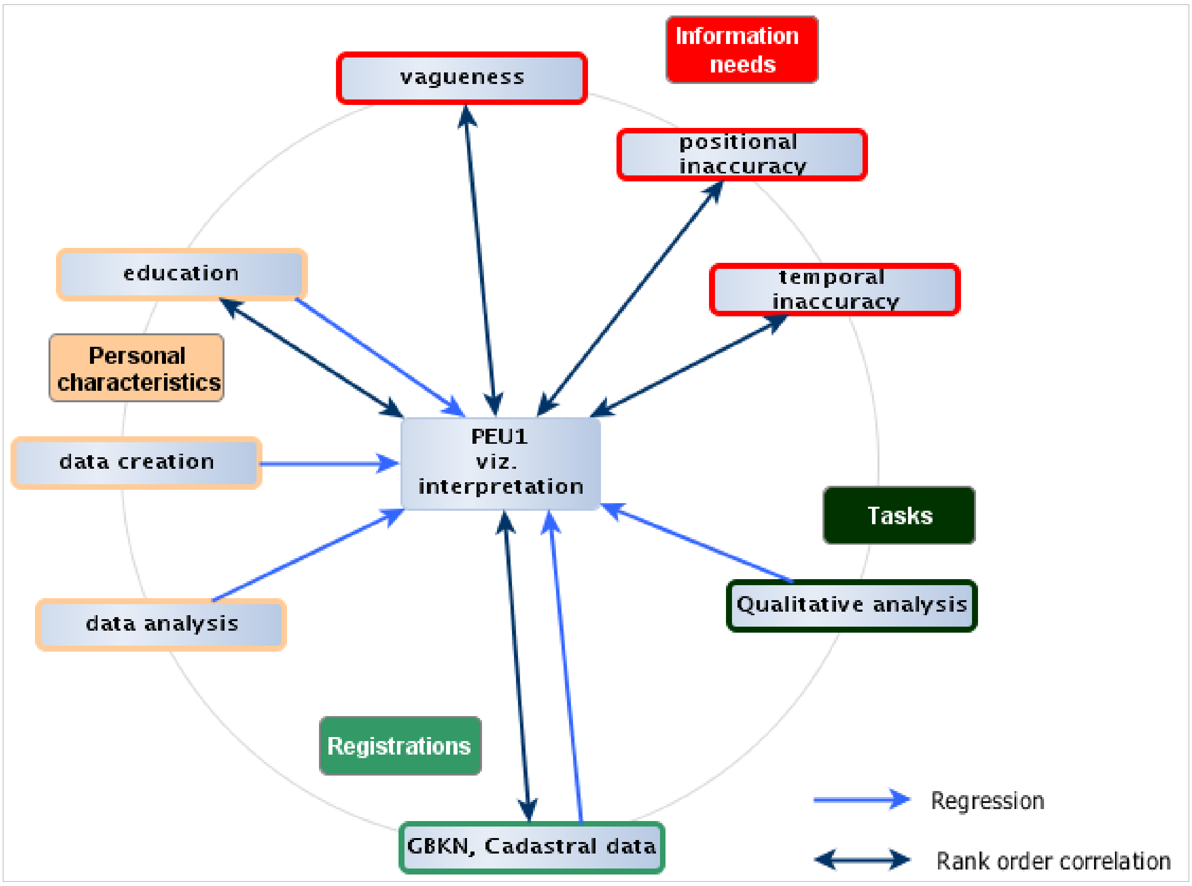


Figure 7.08. Effect on perceived usefulness PEU1.

	independent	dependent	Ro	P-value	method
	Education	PEU1_viz_dyn_interp	-0,333	0,027	Rank-order correlation
	Info_vagueness	PEU1_viz_mta_interp	0,329	0,029	Rank-order correlation
	Info_positional-accuracy	PEU1_viz_mta_interp	0,343	0,023	Rank-order correlation
	Info_temporal-accuracy	PEU1_viz_mta_interp	0,360	0,016	Rank-order correlation
	registration_gbkn	PEU1_viz_dyn_interp	0,389	0,009	Rank-order correlation
	registration_lki	PEU1_all	-	0,033	Ordinal regression
	registration_akr	PEU1_all	-	0,042	Ordinal regression
	Task_analysis-qualitative	PEU1_all	-	0,042	Multi regression analysis
	Experience_data-analysis	PEU1_all	-	0,030	Ordinal regression
	Experience_creating_GI_data	PEU1_all	-	0,012	Ordinal regression
	Education	PEU1_all	-	0,010	Multi regression analysis

Table 7.04. Effect on perceived usefulness PEU1

7.1.4.5. Effect on perceived ease of use, adaption (PEU2)

This section describes the interpretation of the results of the analysis on the effect on PEU2. The diagram in Figure 7.09 gives and overview of the distribution of the variables ordered by category. Each arrow indicates if PEU2 is affected by on or more variables. The typology of the arrow indicates if the predicted relationship is derived from correlation or regression. The predicted relationships are listed in Table 7.05.

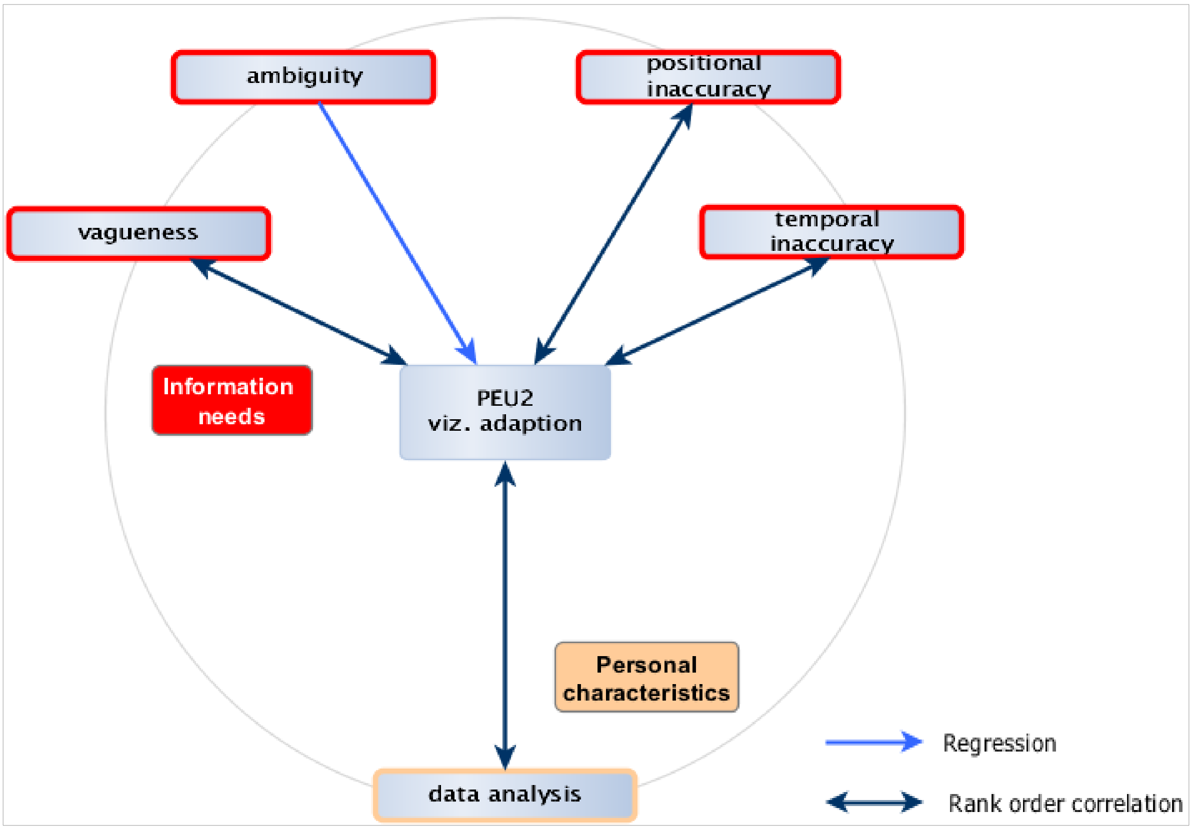


Figure 7.09. Effect on perceived usefulness PEU2.

The perceived ease of use regarding adaption is affected by information needs and probably by when respondents are experienced with spatial data analysis. The discovered correlation is between this experience with spatial data analysis and visualization using metadata. According to the outcomes the most probable adaptive types of visualization are the use of metadata and single static maps when the information needs are ambiguity, positional accuracy, and temporal accuracy.

	independent	dependent	Ro	P-value	method
	Experience_data-analysis	PEU2_viz_mta_adapt	0,298	0,050	Rank-order correlation
	Info_vagueness	PEU2_viz_mta_adapt	0,446	0,002	Rank-order correlation
	Info_positional-accuracy	PEU2_viz_single_adapt	-0,335	0,026	Rank-order correlation
	Info_temporal-accuracy	PEU2_viz_single_adapt	-0,306	0,043	Rank-order correlation
	Info_temporal-accuracy	PEU2_viz_mta_adapt	0,359	0,017	Rank-order correlation
	Info_ambiguity	PEU2_all	-	0,0005	Multi regression analysis

Table 7.05. Effect on perceived usefulness PEU2

7.1.4.6. Results Rank-order correlation analysis

To determine the strength of the association between ordinal variables Spearman's Rank Order Correlation or Kendall's Tau can be used. The choice is arbitrary, there still is a discussion between statisticians what causes differences between those two methods (Fredricks & Nelsen 2007). To determine the level of significance the Table of Zar (1972) with critical values of the Spearman's Rank Order Correlation Coefficients is used. For a two-tailed test with a 5% significance level and with a degree of freedom of 44 observations the critical value of **0.298** is applied. Table 7.06 shows the 23 significant results where p-value < 0.05 of the correlation analysis. A complete overview is given in Appendix 7.

Variable 1		Variable 2	Ro	p_val
Age category	–	The level of insight in ambiguity using metadata	0,31	0,041
Education	–	Dynamic visualization is preferred over static visualization of uncertainty	-0,33	0,027
The level of knowledge of performing spatial analysis	–	The level of insight in ambiguity using metadata	0,35	0,019
The level of knowledge of performing spatial analysis	–	The level of adaption of metadata into current tasks	0,30	0,050
Frequency of using Key Georegistrations as a reference	–	The level of information needs about positional accuracy	0,37	0,012
Frequency of using Key Georegistrations to derive quantitative information	–	The level of information needs about positional accuracy	0,33	0,028
Frequency of using Key Georegistrations to derive qualitative information	–	The level of information needs about positional accuracy	0,38	0,011
The level of information needs about ambiguity	–	The level of insight in map accuracy using dy-namic visualization	0,32	0,031
The level of information needs about ambiguity	–	The level of insight in map accuracy using metadata	0,33	0,026
The level of information needs about vagueness	–	The level of insight in map accuracy using a single static map	0,32	0,033
The level of information needs about vagueness	–	The level of insight in map accuracy using metadata	0,42	0,005
The level of information needs about vagueness	–	The ease of use interpreting uncertainty using metadata	0,33	0,029
The level of information needs about vagueness	–	The level of adaption of metadata into current-Tasks	0,45	0,002
The level of information needs about positional accuracy	–	The level of adaption of a single static map into current Tasks	-0,33	0,026
The level of information needs about positional accuracy	–	The ease of use interpreting uncertainty using metadata	0,34	0,023
The level of information needs about attribute accuracy	–	The level of insight in map accuracy using metadata	0,37	0,013
The level of information needs about temporal consistency	–	The level of adaption of a single static map into current Tasks	-0,31	0,043
The level of information needs about temporal consistency	–	The level of insight in map accuracy using metadata	0,42	0,005
The level of information needs about temporal consistency	–	The level of insight in ambiguity using metadata	0,35	0,021
The level of information needs about temporal consistency	–	The ease of use interpreting uncertainty using metadata	0,36	0,016
The level of information needs about temporal consistency	–	The level of adaption of metadata into current-Tasks	0,36	0,017
The use of large scale topography	–	The level of insight in map accuracy using dy-namic visualization	0,36	0,016
The use of large scale topography	–	Dynamic visualization is preferred over static visualization of uncertainty	0,39	0,009

Table 7.06. Significant results Spearman's Rank-order correlation

The table with the results shows four items; the two variables, the calculated rank-order correlation, and the calculated p-value. In order to assess the strength of the correlation the following interpretation guidelines are applied:

ρ	interpretation of strength of correlation
< 0.15	very weak
0.15 – 0.25	weak
0.25 – 0.40	moderate
0.40 – 0.75	strong
> 0.75	very strong

Table 7.07. Interpretation of rank-order correlation (University of Groningen 2012)

Assumptions of Spearman's rank-order correlation

- Assumption 1: The two variables should be measured on an ordinal, interval or ratio scale.
- Assumption 2: There needs to be a monotonic relationship between the two variables. A monotonic relationship exists when either the variables increase in value together, or as one variable value increases, the other variable value decreases.

Definition of the Spearman's Rank-order correlation *ρ:*

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

where d_i is the difference in paired ranks and n = the number of cases.

For Spearman's test, p-values are computed using algorithm AS 89 for n < 1290 and exact = TRUE. Because for the number of observations > 10, the applied software R uses an Edgeworth series approximation for larger sample sizes (R Core Team 2012a).

7.1.4.7. Results regression analysis

Regression analysis is used to investigate the effect of independent variables on dependent variables (Berenson & Levine 1992). In this research the PU1, PU2, PEU1, and PEU2 of TAM are operationalized into dependent variables, and the external variables are operationalized into independent variables. In order to make statements and inferences about regression the observations must fit a statistical model (Bethlehem 2009). Due to the nature of the scale of the variables (ordinal) and distribution of the observations the Multi regression model and the ordinal regression model is used. First the results of the regression analysis are presented. The subsequent paragraphs explain the used models. Table 7.08 shows the results of the hypotheses which are significant or partial significant. An overview of the implemented models is given in Appedix 6 and an overview of the results of the regression analysis is given in Appendix 7.

ID	Hypothesis	Multi Linear Regression	Ordinal regression logit link	Ordinal regression clog-log link
HYP02C	Education affects PEU1	Significant	Significant	Significant
HYP06A	Task affects PEU1	Significant	Partial significant	NOT significant
HYP07_2	Information needs affects PU2	Significant	NOT significant	NOT significant
HYP07A	Information needs affects PEU2	Significant	NOT significant	NOT significant
HYP09B	Use of GBKN affects PU1	NOT significant	Significant	NOT significant
HYP08D	Metadata affects PU1	NOT significant	NOT significant	Significant
HYP03D	Experience affects PARTIAL PU2	Assumptions not met	Partial significant	NOT significant
HYP04A	Experience affects PEU1	Assumptions not met	Partial significant	NOT significant
HYP05B_2	Task affects PU2	NOT significant	NOT significant	Partial significant
HYP08	Type of visualization affects PU1	NOT significant	NOT significant	Partial significant
HYP09	Use of registrations affect PU1	NOT significant	Partial significant	NOT significant
HYP13C	Registration affects dynamic map use	Partial significant	NOT significant	NOT significant

Table 7.08 Results regression analysis

The results presented in Table 7.04 shows by each hypothesis the results of the performed regression method. The independent variable may consist out of more than one dummy variables. When there not all the independent variables have a significant effect on the dependent variable, the result in the table is classified as 'Partial significant'.

Multi Linear regression model

To apply the multi regression model four assumptions must be met (Berenson & Levine 1992):

1. Variables are normally distributed
2. The assumption of a linear relationship between the independent and dependent variable(s)
3. Variables are measured without error (reliably)
4. The assumption of homoscedasticity

These assumptions are implemented in the used software "R" and calculates the degree of conformance which are reported in Appendix 7.

The formal model for multiple regression given *n* observations is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_n x_{ni} + \epsilon_i$$

Where:

β_0 is the intercept,

$\beta_1 x_{1i}$ is the slope of Y with variable x_1 holding variables $x_2 \ldots x_n$ constant,

$\beta_2 x_{2i}$ is the slope of Y with variable x_2 holding variables $x_3 \ldots x_n$ constant,

and ϵ_i represents the model deviation.

The F-test is used for overall model test and the t-tests are used for testing the individual regression coefficients. Multi linear regression performed using one predictor requires 89 respondents by medium effect size of 0.15 , a required power of 0.95, and an error probability of 0.05 (Dattalo 2008).

Multi linear regression			Signifance Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05	
Hypothesis	**link**	**Independent variable**	**estimate**	**prob. t-test**
Education affects PEU1	-	pers02_edu	-0,250	0,010 **
Task affects PEU1	-	task11_freq_qualit	0,166	0,042 *
Info needs affects PU2	-	inf13_ambiq	0,245	0,037 *
		inf13_ambiq	0,305	0,008 ***
Registration affects dynamic map use	-	data_ind_gbkn	1,141	0,004 ***

Table 7.09 Results multi linear regression analysis

The Ordinal regression model

Ordinal regression, also known as the ordered logit or ordered logistic regression model is a regression model for ordinal dependent variables (Rodríguez 2007).
The ordinal regression model, or cumulative link model (Agresti 2002) can also be applied for ordinal response variables or ordered categorical data where the distances between categories are unknown. The ordinal regression model is applied using two different link functions:

Model Ordinal regression with the logit link:

$$\mathrm{logit}(P(Y_i)) = \theta_j + \beta_1(X_{i1} + \beta_2(X_{i2}) \ \ldots + \beta_k(X_{ik}))$$

Where:
$\mathrm{logit}(P(Y_i))$ is the cumulative probability for the j^{th} category, for the i^{th} case,
Θ_j is the threshold for the j^{th} category,
β_1, β_2, ...β_k, are the regression coefficients,
and X_{i1}, X_{i2}, X_{ik} are the predictors (independent variables) for the i^{th} case.

Model with the complementary log-log link

$$\log\{-\log[1-P(Y_i \leq j)]\} = \theta_j + \beta_1(X_{i1} + \beta_2(X_{i2}) \ \ldots + \beta_k(X_{ik}))$$

Where:
$\log\{ - \log [1 - P(Y_i \leq j)] \}$ is the cumulative probability for the j^{th} category, for the i^{th} case,
Θ_j is the threshold for the j^{th} category,
β_1, β_2, ...β_k, are the regression coefficients,
and X_{i1}, X_{i2}, X_{ik} are the predictors (independent variables) for the i^{th} case.

The logit link function is probably the optimal link for evenly distributed categories, and the complementary log-log link function is probably the most optimal link function when higher categories are more probable (Andric et al. 2010). The Akaike information criterion (AIC) is used to compare models to investigate which model would best approximate the data. Smaller values indicate a better fit.

Ordinal regression				
Hypothesis	**link**	**Independent variable**	**estimate**	**prob. t-test**
Use of GBKN affects PU1	logit	data_ind_gbkn	3,743	0,029 *
Metadata affects PU1	cloglog	v_mta	0,602	0,021 *
Experience affects PEU1	logit	pers05_gi_b_exp	1,631	0,001 **
Experience affects PEU1	logit	pers06_ana_exp	-0,772	0,035 *
The use of registrations affect PEU1	cloglog	data_ind_lki	1,312	0,033 *
		data_ind_akr	-1,259	0,042 *
Experience affects PARTIAL PU2	logit	pers06_ana_exp	0,733	0,012 *
Experience affects PEU1	logit	pers05_gi_b_exp	1,631	0,001 ***
		pers06_ana_exp	-0,772	0,035 *
Task affects PU2	cloglog	task11_freq_qualit	0,396	0,038 *
		task12_freq_integr	-0,360	0,020 *
Type of visualization affects PU1	cloglog	v_mta	0,602	0,021 *
Use of registrations affect PU1	logit	data_ind_gbkn	3,108	0,033 *

Table 7.10 Results ordinal regression analysis

7.2. Results in-depth interviews

The findings of the in-depth interviews are discussed in this section of the chapter. The subsequent paragraphs discusses the findings and are organized by the established themes discussed in Section 6.2.2. To structure the analysis the findings are ordered by the definition of uncertainty and the adapted taxonomy discussed in chapter three.

Theme 1. Chain awareness
All the respondents recognized the concept of an information chain and their role in a municipal information chain, as well on a organisational level as on the level of their own role in several working processes. Their role as they perceive is an intermediary role in a chain process between the data supplier and the end-users. The respondents use several key registrations (in the form of data products) which are used for multiple purposes using a variety of methods for of processing, analysis, and visualization. The interviewees emphasize on their role to provide the end user suitable information for decision-making.

Theme 2. Uncertainty awareness
The concept of 'uncertainty awareness' in this research is defined as the ability of the respondent to notify, identify, quantify, and evaluate uncertainty in spatial data (Reinke & Hunter 2002). To investigate the appearance of uncertainty in a municipal information chain the understanding of the respondents awareness of uncertainty in spatial data is essential.

Awareness of uncertainty in spatial data can be divided into different categories which are mentioned in Section 4.1. Ignorance of uncertainty was not mentioned in the model of Skeels et al. (2010) and is added as an element of the theme. The findings of the interview are classified in these categories and described in the subsequent paragraphs.

Ignorance
The interviews revealed that uncertainty was ignored due to the compulsion of the demanding organisation to keep provided information free from uncertainties for the purpose decision-making.

Known knowns
Uncertainty in spatial data was by the interviewees notified and identified, and in some cases (i.e. positional inaccuracy) quantifiable. Uncertainty was notified during the use of the data by expert knowledge, the usage of a combination of different data sources, or by spatial transformation of the data like buffering or aggregation.

Unknown knowns
In the case of unknown knowns there is uncertainty, but it is not definable and quantifiable. For the interviewees it is not always clear (uncertain) how the information model is implemented in the derived data products. Therefore they experience that the meaning of the data is mostly unclear, and there is a lack of clear definitions.
Related to the previous statement is the experience that the formulation of the survey rules of the data producer are based on certain interpretations which may differ from the perspective of the data users. Other unknown knowns are of information about completeness and currency and limited knowledge about the responsible data producer of a data source which refer to registrations which are not key Registrations.

Unidentified unknowns
Unidentified unknowns appear when users are not aware of uncertainty in spatial data. The in-depth interview revealed that 80% of the interviewees was not aware of the inhomogeneous positional inaccuracy of the cadastral map and the limitations of its use. It is notifiable that the this respondents the key Geo-registrations consider as the 'truth', a source were they completely rely on.

Theme 3. Uncertainty and problem recognition
Problem recognition caused by uncertainty in this research is interpreted as how the respondents experience uncertainty in spatial data as a problem.
The interviews showed that problem recognition regarding uncertainty is experienced differently by the respondents. The findings of the in-depth interviews are thematised according the adapted taxonomy of uncertainty which is introduced in chapter three. The in-depth interviews revealed that the respondents pointed out features of uncertainty which are not included in the taxonomy of uncertainty, but refer to elements of spatial data quality as adapted in the ISO 19113 standard for spatial data quality. The findings regarding this theme are organised by the taxonomy of uncertainty.

<u>Ambiguity (discordance, non-specificity)</u>
The results of interviews showed that ambiguity is an issue which is experienced by the interviewees as an unclear meaning of the data and lack of definitions. The translation from (a certain perspective) *from reality to the intended information model* was experienced as vague or indistinct. The definition of survey rules how to demarcate objects like buildings or urban green areas was experienced as non-specific. The demarcation of those objects were experienced as sometimes ambiguous, this ambiguity propagates in the demarcation of the areas.

<u>Vagueness</u>
Vagueness is not always recognized as a problem using key registrations and depends on the purpose of use. The boundaries of the urban ecological structure are derived form large scale topographic objects. The demarcation of those objects are sometimes ambiguous, those ambiguity propagates in the demarcation of the areas of the urban ecological structure. Vagueness in the demarcation may have impact on the verification of change requests against policy and affects the evaluation of permits against the compliance of municipal policies.

<u>Positional inaccuracy</u>
Positional inaccuracy is recognized as a problem when it shifts from acceptable inaccurate to erroneous (not acceptable inaccurate). Whether inaccuracy is acceptable depends of the requirements related to the requirements and purpose of use as discussed in Section 3.4.
Due to the unquantifiable positional inaccuracy in the large scale topography it occurs that users make extra costs to improve specific parts of the data set until it fit for their use.
A noteworthy finding during the in-depth interviews were statements of the respondents regarding '*The shift from inaccurate to erroneous*'. The respondents indicated that inaccuracies are acceptable, but when these inaccuracies exceed 'a certain' threshold value, inaccuracies become errors. These 'threshold values' were not explicitly formulated and quantified and are depending on the experience of the expert and the purpose of use.

<u>Attribute inaccuracy</u>
The degree of problem recognition of attribute inaccuracy differs among the interviewees. The importance of attribute accuracy depended on the task, and on the importance of the derived information. A prominent example is the attribute inaccuracy of year of construction (BAG), type of pavement of a road, or the size of the are of a dwelling (BAG).
These kind of inaccuracies lead to extra investigation (time and money) to determine the right or more accurate values.

<u>Temporal inaccuracy</u>
Key Geo-registrations are spatio-temporal. Inaccuracies in temporal aspects of the data are experienced in the inconsistencies of the temporal attributes of an object (correctness of ordered events or sequences). In municipal processes or events or 'cases' are dependent of temporal attributes of objects. The respondents experience when those temporal attributes are uncertain it may effect cases which are subject of an act of public law.
Temporal inaccuracy (uncertainty) is experienced when the actuality of used data sets is unknown. This is experienced when differences in actuality between used data sets occur.

Findings complementary to the adapted taxonomy of uncertainty
The in-depth interviews revealed elements of spatial data which are not included in the adapted taxonomy of uncertainty. These elements refer to elements of which are also elements of spatial data quality according to ISO 91113 and ISO 91115.

Completeness
Being uncertain about completeness of objects, relationships, or attributes occurred in different ways. Interviewees were uncertain about the completeness objects in a data set (*Data completeness*). For example the parking enforcement is highly depended of the degree of completeness of addresses. The addresses are used for determining the location of vehicles and extracting information for parking tickets. Another form of experiencing completeness is uncertainty about the completeness of the model. *Model completeness* refers to the agreement between the abstraction of reality and what is specified in the data set. Which objects are contained by the model and which are not?
Uncertainty about completeness of attributes is experienced for attributes of Key-registration Buildings. Especially uncertainty about the completeness of the original year of construction of a building is an issue in cities with historical city parts.
Also public spaces without legally established names might have effect on legal acts in municipal processes. Uncertainty about completeness in relationships is experienced by relations between registrations within the system of Key Registrations. Within the spatial domain the relationship between buildings and cadastral objects is experienced as a problem due to the effect on municipal processes.

Topological inconsistency
In domain specific data *Topological inconsistency is experienced in* spatial differences between areas, for example sliver polygons in features for parking areas.
These topological inconsistencies might propagate in derived information used for geocoding information.

Responsible data provider
The responsible data provider is in the context of this research defined as the responsible organization which produced the original (unmodified or enriched) data.
For Key registrations the data provider is undisputed, but for other used data sources the origins one respondent indicated that the responsible data provider is not always clear.

Insight in modifications
The interviews indicated that insight in the changes of the data might contribute to anticipate of handling data. From the perspective of of the interviewees is not only uncertainty of the data itself important, but also the changes which occur in a data set, and how those changes affect the domain specific data or triggers domain specific processes.

Theme 4. Handling uncertainty
When uncertainty in spatial data is recognized, users may choose different ways to handle those uncertainties. The discussions with the interviewees revealed 5 ways of handling uncertainty in spatial data:

1. Ignore uncertainty in spatial data
2. Incorporate expert knowledge to understand the limitations
3. Make adjustments in the working procedures to reduce the effects of uncertainty
4. Reducing uncertainty through investigation, or inquiry by the data producer
5. Self-improving of the data; acquire missing data, or improve accuracy

The way the respondents handling uncertainty in spatial data may indicate what kind information about uncertainty they might need. The next two themes focus on the current perception of information needs about uncertainty in spatial data.

Theme 5. The perception of the interviewees how they are informed about uncertainty
The questions asked regarding this theme were about the provided information about uncertainty in spatial data. The questions focused on the elements described in the adapted taxonomy of uncertainty and the additional elements mentioned in "Uncertainty & problem recognition". The findings regarding how interviewees were informed about uncertainty varies from "very limited" to "no usable information". According to the interviewees normative limited information about uncertainty was provided about positional accuracy and actuality.

Theme 6. How interviewees prefer to be informed about uncertainty
The interviewees indicated their preference how they wanted to be informed about uncertainty in spatial data. Those preferences are divided into the categories information content, the form of information, and added value.

Information content
The interviewees indicated to be interested in normative and actual information about uncertainty in spatial data. Normative information should contain the standards and threshold values, and the degree of conformance of the data according to these standards.
Actual information about uncertainty in spatial data stored in conjunction with the actual data, or on a data set level is also preferred by the interviewees.
Frequently mentioned elements are actuality, positional accuracy, and completeness.

The form how information about uncertainty in spatial data should be communicated
The preferences of the interviewees in what form information about uncertainty in spatial data should be communicated are diverse. On the dataset level reports of about spatial data quality (standards an conformance) are perceived as informative.
On the object level,the interviewees prefer information about uncertainty stored in conjunction with the actual data so the can perform analysis to fit their usefulness for a certain purpose. Additional requirements are that information about uncertainty must be easy adaptable in their processes, easy accessible and tuned on its users.

Perceived usefulness of information about uncertainty in spatial data
A part of the interviews was focused on the perceived usefulness of information about uncertainty in spatial data. The interviewees pointed out that information about uncertainty contribute tot efficiency by having detailed knowledge about the limitations and possibilities in advance. In practice there are examples of unnecessary efforts and investments to acquire additional data or explain uncertainties. During the use of data users observe discrepancies or inaccuracies without knowing the cause. If they are better informed about the cause the can process the data more efficient and take the imperfection of the data into account.

Improvement of knowledge
A notifiable finding from the in-depth interviews is a need of the users to expand their knowledge (for example by suitable education) about concepts of spatial data quality and uncertainty in spatial data. When users gain knowledge and have a shared perception about these concepts and the characteristics of the data, they believe that this improves the usability of the data and the information about uncertainty in this data.

"There is no truth. There is only perception."
Gustave Flaubert, novelist

8. Conclusions

To work towards the answer of the main research question first the sub-questions are answered. The main research question will be answered in Section 8.2 preceded by the synthesis of the results of the literature study, survey, and in-depth interviews. Section 8.3 discusses the limitations of this study and possibilities for further research.

8.1. Synthesis

Sub-question 1. How is uncertainty in spatial data defined?
The literature review discussed in chapter three showed a general definition of uncertainty and the existence of different taxonomies of uncertainty in the context of spatial data.
The different taxonomies showed overlapping and disjoint elements caused by different perspectives. The taxonomy of Shi was adapted with modifications for the purpose of the survey. Users can be uncertain about more characteristics of spatial data then described in the discussed taxonomies of uncertainty. Additional discovered characteristics in this research such as completeness, topological consistency, or responsible data provider can be classified according to the international standards for spatial data quality ISO19113, ISO19114 and the international standard for metadata ISO19115.

Sub-question 2. How are uncertainty and spatial data quality related?
Beside the difference in definition of uncertainty and spatial data quality there are similarities. These similarities were noted in Figure 4.02 of the corresponding elements of the taxonomies of uncertainty and spatial data quality. To achieve an unambiguous communication and use of information about uncertainty it is essential to emphasize the relationship between uncertainty and spatial data quality. To improve insight in the relationship between uncertainty and spatial data quality the following conclusion is formulated:

"*Information about uncertainty in spatial data provides input to investigate to what extent the characteristics of a data set fit the requirements specified by users or data producers*"

This relationship can be perceived from two perspectives, the perspective from data producers and the perspective from data users. The data producers can use information about uncertainty to assess the degree to what extent the produced data fit the formal requirements. Data users can use information about uncertainty to assess the fitness of use and how inherent imperfection of the data set affect the nature of their use and handling methods.

Sub-question 3. How can uncertainty in spatial data be communicated?
Chapter four discussed different methods and studies of communicating uncertainty.
Based on the literature study and the findings of the in-depth interviews it can be stated that communication about uncertainty in spatial data might be effective when it is designed within context of user tasks and easy adaptable. To prevent information about uncertainty to become unnecessary or excessive for various users information about uncertainty should be understandable and process-driven. This implies that uncertainty should be visualized on different ways adapted on the users needs, context, and abilities.

Sub-question 4. How does uncertainty in spatial data appear in a municipal information chain?
Based on the results of the in-depth interviews and desk-research the appearance of uncertainty in a municipal information chain explored. Section 5.2 described this phenomenon a hypothetical way. The results of the in-depth interviews illustrated examples how uncertainty appears in a municipal information chain.

Sub-question 5. What are the information needs about uncertainty in spatial data?
The results of the survey revealed the information needs of the respondents which are presented in chapter seven. As described in Section 7.2, additional information needs which were not mentioned in the adapted taxonomy of uncertainty were discovered at the in-depth interviews.

Sub-question 6. How do users perceive usefulness of information about uncertainty in spatial data?
The interviewed respondents experienced a gap between information needs and provided information about uncertainty. This confirms the experience from the field is that information about uncertainty or data quality provided by key Geo-registrations is limited.
This research showed that the majority of the expert-users information about uncertainty in spatial data considers as useful. These information needs not only concerns the elements of the taxonomy of uncertainty but also elements of the international standards of spatial data quality and metadata. Notifiable is the (re)discovery of the fact that users find it useful how the 'real world' is translated to an information model of a Key-Georegistration and extracted to data sets. This discovery emphasizes the fact that there are different perceptions of the establishment of data sets bases on Key-Georegistrations (Koláčný 1969; MacEachren 1995).
Visualization of uncertainty is explored by investigating the degree of insight in accuracy, the ease of use of different types of visualisations, and the adaptability of different types of visualisations in tasks of the users. Dynamic visualization and Metadata are perceived to provide most insight in map accuracy, are perceived as easy to use, and are the most optimal visualization methods to adapt.

Sub-question 7. Which factors affect the usefulness of information about uncertainty in spatial data?
To determine which factors influence the usefulness of information about uncertainty possible relationships a survey is conducted and results are statistically analysed. Section 7.15 provided a detailed summary of the determined relationships. The results derived from the literature review and the in-depth interview revealed factors which are complementary to the results of the survey. Factors derived from all methods are listed in Table 8.01.

nr.	Factor	Affected	explanation
S1	Information needs	PU2, PEU1, PEU2	PU1 : information needs PU2 : insight in map accuracy PEU1 : interpretation of visualizations PEU2 : adaptability of visualizations
S2	Task	PU1, PEU1	
S3	Type of visualization	PU1	
S4	Personal experience	PU2, PEU1, PEU2	
S5	Personal characteristics	PU2, PEU1	
S6	Type of registration	PU1, PU2, PEU1	
I1	Uncertainty awareness	General usefulness	
I2	Problem recognition	General usefulness	
I3	Uncertainty handling	General usefulness	
I4	Available information about uncertainty	General usefulness	
I5	Knowledge	General usefulness	

Table 8.01. Factors which influence the usefulness of information about uncertainty

The factors S1 – S6 are based on the results of the survey. From the determined relationships can generally be derived that perceived usefulness of information about uncertainty in spatial data is influenced by which task is performed, the information needs, and the type of visualization. The perceived ease of use is generally influenced by experience, tasks, information needs, and the level of interpretation of dynamic visualizations.

The factors I1 – I5 are based on the results of the literature review and the in-depth interviews. The coherence between uncertainty awareness, problem recognition, uncertainty handling, and available information about uncertainty is predefined and deductively investigated using the interviewees. Details of these factors were described in Section 7.2. The factor 'knowledge' is discovered during the interviews. Different areas of knowledge might affect the usefulness of information about uncertainty. *Conceptual knowledge* about spatial data and standards improves the understanding of Key-Georegistrations and expands the ability of the expert-user to improves the usability information about uncertainty. Knowledge of *how the translation from reality to an information model* is performed can provides the users of key-Georegistrations more insight in the meaning of the data, and therefore also in information about uncertainty of that data.

8.2. Conclusion

This research attempts to answer the research question "*Which factors influence the usefulness of information about uncertainty in spatial data of key Geo-registrations in municipal chain processes?*". The research question is answered to place the answers of the sub-questions in a municipal context. The scope of this research is the use of Key-Georegistrations in municipalities. Due to the nature of municipal organisations spatial data flows through different links of information chains, and are in some cases fused or enriched. Key-Georegistrations have a legal base which includes mandatory use. This implies that other organisations than the original data producer are not allowed to collect, change or improve the base data. These legal issues might limit the possibilities how to handle uncertainty in spatial data. In Municipalities are politics, public interest, responsibility of public finance, and a proper conduct important drivers. When expert-users are informed about uncertainty in spatial data and can estimate the uncertainty of the derived information for decision-makers,then political and organisational concerns might affect how uncertainty about spatial data is used. Rationality of decision-makers might be affected by the form of the portrayed information about uncertainty in spatial data (Hope 2005).

The left part of Figure 8.01 illustrates four key factors which affect usefulness of information about uncertainty in spatial data. Experience in spatial data analysis and creating data influence the perceived ease of use. The perceived usefulness is influenced by the interaction between task and registration use, information needs, and how information about uncertainty is visualized. The factors in the right part of Figure 8.01 are connected by the factor 'Information needs' which is the link between the factors which are derived from the survey and the in-depth interview. General usefulness of information about uncertainty is influenced by the extend how users are aware of uncertainty, if users recognize or ignore the uncertainty as a problem. When users experience uncertainty as a problem the handling methods depends on the balance between information about uncertainty they require and the availability of information about uncertainty in spatial data. Knowledge of concepts about uncertainty and a shared perception of data and uncertainty contributed to minimize the gap between supply and demand of information about uncertainty.

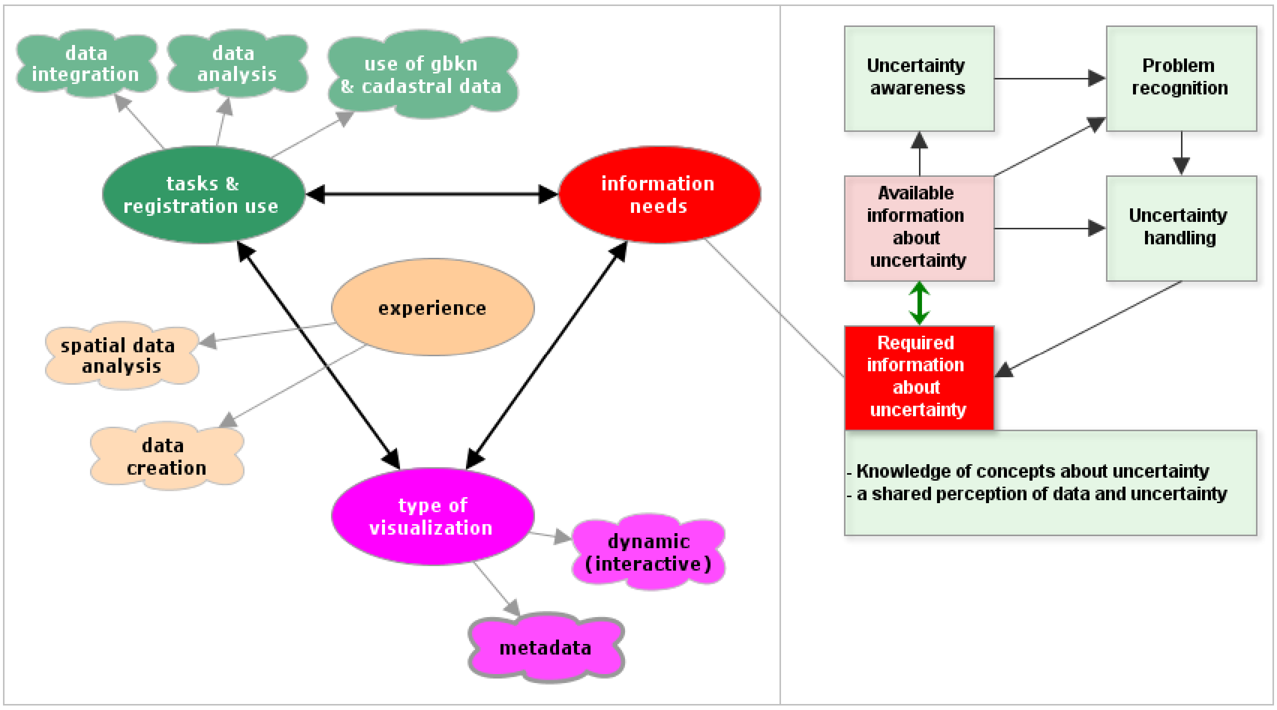


Figure 8.01. Factors which influence the usefulness of information about uncertainty.

8.3. Limitations of this study

Generalizability
The results of the questionnaire of this study are statistically significant and might be generalized on the targeted population. The generalizability can be limited because the extend of the investigated population is not exactly known. The approach of this study is based on behaviouralism with the focus on how people perceive and experience spatial phenomena in everyday life and/or made decisions (Gatrell et al. 2005). Therefore the transferability of the results of this research to other populations should be critically reviewed.

Sample size
The survey is based on a population were the size is not exactly known and is estimated based on a survey which was performed in 2010 (Geobusiness Nederland 2010). The sample size to determine correlation is sufficient, but for multi linear regression the size is limited due to the response of 44 compared to the required sample of 89.

The sample of the in-depth interview was 10, and response was 5. Therefore the generalizability of the results of the in-depth interviews are limited but transferable to comparable studies according to the study about sample size and saturation of Mason (2010). Guidelines mentioned by Creswell (1998) vary between 5 and 50 depending of the research field. Mason emphasizes that saturation determines the majority of qualitative sample size. Jette et al. (2003) suggested that when a researcher has expertise in the investigated topic the number of participants can be reduced.

8.4. Recommendations and further study

Unify the methods to communicate about uncertainty and spatial data quality
Considering the previous and current developments of the Dutch system of key registrations it can be argued that this are supply driven developments. These supply driven developments originates from different domains with their own historical backgrounds. Due to these different backgrounds concepts, used standards, and taxonomies for uncertainty and spatial data quality are different. Based on the findings of this study it is recommended to unify the methods to communicate about uncertainty and spatial data quality within the context of the use of the Dutch system of key registrations. Unification can lead to improved consistency in quality assessment of integrated use of key registrations.

Extend existing information models with elements of uncertainty
To assess data quality conformance or fitness for use, information about uncertainty in spatial data should be available at the level of a data set, object, or attribute. Further research to investigate feasibility and efficacy of how information about uncertainty in the information models of Key-Georegistrations can be remodelled is recommended. It is also recommended to adapt established standards for spatial data quality into these models and embed the concepts of uncertainty into the adaptation of these standards.

Close the knowledge gap
This study pointed out that experience and personal characteristics affects the usability of information about uncertainty in spatial data. It is recommended to identify which knowledge is required to improve the usability of information about uncertainty, and how this knowledge currently exists by users and embedded in their organisations. In addition to these research objectives is further to investigate how to close the gap between required knowledge and the existing knowledge.

References

Aerts, J.C.J.., Clarke, K.C. & Keuper, A.D., 2003. Testing popular visualization techniques for representing model uncertainty. *Cartography and Geographic Information Science*, 30(3), pp.249–261.

Agresti, A., 2002. *Categorical data analysis*, Wiley-interscience.

Andric, N. et al., 2010. Ordinal Response Models. Available at: http://isites.harvard.edu/fs/docs/icb.topic749412.files/OrdinalModels_v2.pdf [Accessed April 7, 2013].

Babbie, E., 2001. Modes of observation. In *The practice of social research*. Belmont,USA: Wadsworth Thomson learing inc., pp. 237–273.

Berenson, M.L. & Levine, D.M., 1992. *Basic business statistics , concepts and applications* 5th ed., New York, United States of America: Prentice Hall.

Bertin, J., 1983. *Semiology of Graphics: Diagrams, Networks, Maps*, Madison: University of Wisconsin Press.

Bethlehem, J., 2009. *Applied survey methods: a statistical perspective*, Hoboken , NJ, USA: John Wiley & Sons, Inc.

Blake, R. & Mangiameli, P., 2011. The Effects and Interactions of Data Quality and Problem Complexity on Classification. *Journal of Data and Information Quality (JDIQ)*, 2, pp.8:1–8:28.

Boin, A.T., 2008. *Exposing Uncertainty: Communicating spatial data quality via the Internet.* PhD THESIS. Australia: Department of Geomatics,The University of Melbourne.

Bordogna, G. et al., 2009. Managing imperfect temporal metadata in the Catalog services of Spatial Data Infrastructures compliant with INSPIRE. In *IFSA-EUSFLAT 2009 Proceedings*. The Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference. Lisbon, Portugal: European society for fuzzy logic and technology, pp. 915–920. Available at: http://www.eusflat.org/publications/proceedings/IFSA-EUSFLAT_2009/IFSA-EUSFLAT2009Proceedings.pdf [Accessed April 4, 2011].

Boukhelifa, N. & Duke, D.J., 2009. Uncertainty visualization: why might it fail? In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. CHI EA '09. Boston, MA, USA: ACM, pp. 4051–4056.

Burrough, P.A. & McDonell, R.A., 1998. *Principles of Geographic information systems*, Oxford University press.

BZ, D. ministry of internal affairs, 2008. Nationaal uitvoeringsprogramma dienstverlening en e-overheid. Available at: http://www.e-overheid.nl/images/stories/Publicaties/akkoordverklaring-1-december.pdf [Accessed June 14, 2011].

BZK, M. van B.Z. en K., 2011. Stelsel van Basisregistraties en implementatievisie.

Campaña, J. et al., 2007. A Fuzzy Set-Based Approach to Temporal Databases. In *Scalable*

Uncertainty Management. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, pp. 31–44. Available at: http://dx.doi.org/10.1007/978-3-540-75410-7_3.

CBS, CBS StatLine - Bevolking; ontwikkeling in gemeenten met 100 000 of meer inwoners. Available at: http://statline.cbs.nl/StatWeb/publication/default.aspx?DM=SLNL&PA=70748NED&D1=0-1%2c3%2c26%2c40%2c68%2c73%2c75%2c81%2c83&D2=0&D3=0&D4=a&D5=0%2c5-8&HDR=G2%2cT%2cG1&STB=G4%2cG3&VW=T [Accessed January 17, 2011].

Comber, A.J. et al., 2006. Using Metadata to Link Uncertainty and Data Quality Assessments. In *Progress in Spatial Data Handling*. Springer Berlin Heidelberg. Available at: http://www.springerlink.com/content/n34u6l5w6277875u/ [Accessed December 6, 2010].

Cook, K.E., 2012. In-Depth Interview. In *Qualitative research methods*. Thousand Oaks, USA: SAGE Publications. Available at: http://www.sage-ereference.com/research/n209.xml.

Creswell, J., 1998. *Qualitative inquiry and research design: Choosing among five traditions*, Thousand Oaks, CA: SAGE Publications.

Dattalo, P., 2008. Determining Sample Size : Balancing Power, Precision, and Practicality. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=220931&site=ehost-live.

Davis, F.D., 1989. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), p.319. Available at: http://www.jstor.org/discover/10.2307/249008?uid=3738736&uid=2&uid=4&sid=21101476216181 [Accessed December 10, 2012].

Davis, F.D. et al., 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3).

Devillers & Jeansoulin, 2006. *Fundamentals of Spatial Data Quality* 1st edition., London, Great Britain: ISTE Ltd.

Devillers, R., Bédard, Y. & Jeansoulin, R., 2005. Multidimensional management of geospatial data quality information for its dynamic use within GIS. *Photogrammetric engineering and remote sensing*, 71(2), pp.205–215.

Devillers, Rodolphe et al., 2010. Thirty Years of Research on Spatial Data Quality: Achievements, Failures, and Opportunities. *Transactions in GIS*, 14(4), pp.387–400. Available at: http://dx.doi.org/10.1111/j.1467-9671.2010.01212.x.

Drecki, I., 2009. Representing Geographical Information Uncertainty: Cartographic solutions and challenges. In International Cartographic Conference. Santiago de Chile.

Erdfelder, E., Faul, F. & Buchner, A., 1996. GPOWER: A general power analysis program. *Behavior Research Methods*, 28(1), pp.1–11. Available at: http://www.springerlink.com/content/q8x332u6j5650mk1/abstract/ [Accessed May 21, 2012].

Fisher, P.F., 1999. Models of uncertainty in spatial data. *Geographical information systems*, 1, pp.191–205.

Folmer, E. & Bosch, J., 2004. Architecting for usability: a survey. *Journal of Systems and Software*, 70(1-2), pp.61–78. Available at: http://www.sciencedirect.com/science/article/B6V0N-49N0GC5-

1/2/d557fb3a4aa5be98fe93312efd46e7f9 [Accessed December 5, 2010].

Foote, K.E. & Huebner, D.J., 1999. Error, Accuracy, and Precision. *The Geographer's Craft Project, Department of Geography, University of Texas*. Available at: http://www.colorado.edu/geography/gcraft/notes/error/error_ftoc.html [Accessed March 9, 2011].

Fredricks, G.A. & Nelsen, R.B., 2007. On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables. *Journal of Statistical Planning and Inference*, 137(7), pp.2143–2150. Available at: http://www.sciencedirect.com/science/article/pii/S0378375806002588 [Accessed December 17, 2012].

Gatrell, J.D., Bierly, G.D. & Jensen, R.R., 2005. *Research Design and Proposal Writing in Spatial Science*, USA: Springer.

Geobusiness Nederland, 2010. Marktmonitor Nederlandse Geo-informatiesector 2009/2010. Available at: www.geobusiness.nl/document152/geo+sector+in+kaart+20092010 [Accessed December 14, 2010].

Gerharz, L.E. & Pebesma, E.J., 2009. Usability of interactive and non-interactive visualisation of uncertain geospatial information. *Geoinformatik*, pp.223–230.

Grira, J., Bédard, Y. & Roche, S., 2010. Spatial Data Uncertainty in the VGI World: Going from Consumer to Producer. *Geomatica*, 64, pp.61–72.

Hope, S.J., 2005. *Decision-making under spatial uncertainty*. Masters Thesis. Melbourne, Australia: University of Melbourne.

ISO, I.O. for S., 1998. *International standard ISO 9241-11, Part 11 : Guidance on usability*, Geneva, Switzerland: International Organization for Standardization.

ISO/TC 211, 2011. ISO/TC 211 Geographic information/Geomatics. *ISO/TC 211 Geographic information/Geomatics*. Available at: http://www.isotc211.org/ [Accessed April 18, 2011].

Jakobsson, A. & Giversen, J., 2008. *Guidelines for Implementing the ISO 19100 Geographic Information Quality Standards in National Mapping and Cadastral Agencies*, Brussels, Belgium: Eurogeographics.

Jette, D.U., Grover, L. & Keck, C.P., 2003. A qualitative study of clinical decision making in recommending discharge placement from the acute care setting. *Physical therapy*, 83(3), pp.224–236.

JRC, D.T., 2008. INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119. Available at: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/metadata/MD_IR_and_ISO_20081219.pdf [Accessed May 28, 2011].

Kadaster, 2012. BAG Raadplegen. Available at: https://service1.kadaster.nl/bag/bag-burger/secured/selectie [Accessed December 16, 2012].

Kardos, J., Benwell, G.L. & Moore, A.B., 2007. Assessing different approaches to visualise spatial and attribute uncertainty in socioeconomic data using the hexagonal or rhombus (HoR) trustree. *Computers, Environment and Urban Systems*, 31(1), pp.91–106. Available at: http://www.sciencedirect.com/science/article/B6V9K-4KSSW34-1/2/484cc3d540863f5c470d9e95e333e8b5 [Accessed June 6, 2010].

Keijzer, S.C., 2010. *Visualizing scale related uncertainty in web maps for spatial planning*. Masters Thesis. Enschede, The Netherlands: GIMA; ITC Enschede, TU Delft, Utrecht University and Wageningen UR. Available at: http://igitur-archive.library.uu.nl/student-theses/2011-0228-200527/UUindex.html [Accessed May 15, 2011].

KING, K.N.G., 2011. GEMMA: oplossingen voor uw interne organisatie. Available at: http://www.kinggemeenten.nl/king-kwaliteitsinstituut-nederlandse-gemeenten/e-dienstverlening-verbeteren/gemma [Accessed June 14, 2011].

Klir, G.J. & Yuan, B., 1995. *Fuzzy sets and fuzzy logic: theory and applications*, Upper Saddle River, NJ, USA: Prentice Hall.

Koláčný, A., 1969. Cartographic Information—a Fundamental Concept and Term in Modern Cartography. *Cartographic Journal, The*, 6(1), pp.47–49.

Kvale, S., 1996. *InterViews: An Introduction to Qualitative Research Interviewing*, Sage Publications, Inc. Available at: http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/080395820X [Accessed February 13, 2012].

Longley, P.A. et al., 2011a. *Geographical information systems and Science* 3rd ed., John Wiley & Sons, Inc.

Longley, P.A. et al., 2011b. Uncertainty. In *Geographical information systems and Science*. John Wiley & Sons, Ltd, pp. 147–177.

MacEachren, A.M., 1995. *How maps work: representation, visualization, and design*, New York, USA: The Guilford Press.

MacEachren, A.M. et al., 2005. Visualizing Geospatial Information Uncertainty: What We Know and What We Need to Know. *Cartography and Geographic Information Science*, 32(3), pp.139–161.

MacEachren, A.M., 1992. Visualizing uncertain information. *Cartographic Perspective*, 13(3), pp.10–19.

Mason, M., 2010. Sample Size and Saturation in PhD Studies Using Qualitative Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 11(3). Available at: http://www.qualitative-research.net/index.php/fqs/article/view/1428 [Accessed December 29, 2012].

Mazza, R., 2009. *Introduction to Information Visualization*, London, UK: Springer-verlag.

Mcdonald, D.G. & Dimmick, J., 2003. The Conceptualization and Measurement of Diversity. *Communication Research*, 30(1), pp.60–79. Available at: http://crx.sagepub.com/content/30/1/60 [Accessed May 21, 2012].

MIM, M. of infrastructure and environment, 2011. Stelselposter,stelselhandboek Basisregistraties. *Stelselhandboek Basisregistraties*. Available at: https://wiki.noiv.nl/xwiki/bin/view/Stelselhandboek/Stelselposter [Accessed June 14, 2011].

Moses, A., 2010. Google Maps Mistake: Nicaragua Accused Of Invading Costa Rica. *The Sydney Morning Herald*. Available at: http://www.smh.com.au/technology/technology-news/google-maps-error-sparks-international-conflict-20101108-17jl3.html [Accessed November 22, 2010].

Nijkamp, P. & Scholten, H.J., 1993. Spatial information systems: design, modelling and use in planning. *International Journal of Geographical Information Systems*, 7, pp.85–96.

O'Leary, Z., 2010. *The essential guide to doing your research project* third., London, Great Brittain: Sage publications.

Van Oort, P., 2005. *Spatial data quality: from description to application*, Delft, The Netherlands: Netherlands Geodetic Commission.

Van Oosterom, P., 2010. Temporele aspecten in GIS-bestanden. *AGGN Newsletter*, 18(2), pp.6–12.

Van Oosterom, Peter et al., 2006. Aspects of a 4D Cadastre: A First Exploration. In 13th FIG Congress,Munich. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.132.5334 [Accessed March 19, 2011].

Oxford dictionary, 2010a. Error. *Oxford Dictionaries Online*. Available at: http://oxforddictionaries.com/view/entry/m_en_gb0897580#m_en_gb0897580 [Accessed February 21, 2011].

Oxford dictionary, 2010b. Uncertainty. *Oxford Dictionaries Online*. Available at: http://oxforddictionaries.com/view/entry/m_en_gb0897580#m_en_gb0897580 [Accessed February 21, 2011].

Pang, A., 2001. Visualizing Uncertainty in Geo-spatial Data. In *Proceedings of the Workshop on the Intersections between Geospatial Information and Information Technology*.

Pang, A., 2008. Visualizing Uncertainty in Natural Hazards. In A. Bostrom, S. French, & S. Gottlieb, eds. *Risk Assessment, Modeling and Decision Support*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 261–294. Available at: http://www.springerlink.com/content/u7312m483334gu03/ [Accessed July 25, 2011].

Peuquet, D.J., 2000. Space-time representation: An overview. In *Time in GIS: Issues in spatio-temporal modelling*. Delft, The Netherlands: Nederlandse Commissie voor Geodesie. Available at: http://www.ncg.knaw.nl/ [Accessed March 19, 2011].

Pham, B., Streit, A. & Brown, R., 2009. Visualisation of Information Uncertainty: Progress and Challenges. In *Trends in Interactive Visualization*. Advanced Information and Knowledge Processing. Springer London, pp. 19–48. Available at: http://dx.doi.org/10.1007/978-1-84800-269-2_2.

R Core Team, 2012a. R manual, Test for Association/Correlation Between Paired Samples. Available at: http://stat.ethz.ch/R-manual/R-devel/library/stats/html/cor.test.html [Accessed May 9, 2012].

R Core Team, 2012b. *R: A Language and Environment for Statistical Computing*, Vienna, Austria: Statistics Department of the University of Auckland. Available at: http://www.R-project.org.

Reinke, K. & Hunter, G.J., 2002. A Theory for Communicating Uncertainty in Spatial Databases. In *Spatial Data Quality*. London, UK: Taylor & Francis.

Robinson, V.B., 2003. A Perspective on the Fundamentals of Fuzzy Sets and their Use in Geographic Information Systems. *Transactions in GIS*, 7(1), pp.3–30. Available at: http://dx.doi.org/10.1111/1467-9671.00127 [Accessed April 25, 2011].

Rodríguez, G., 2007. Lecture Notes on Generalized Linear Models. Available at: http://data.princeton.edu/wws509/notes/ [Accessed February 26, 2013].

Roth, R.E., 2009. The Impact of User Expertise on Geographic Risk Assessment under Uncertain Conditions. *Cartography and Geographic Information Science*, 36(1), pp.29–43. Available at: http://www.ingentaconnect.com/content/acsm/cagis/2009/00000036/00000001/art00004.

Sanyal, J. et al., 2009. A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), pp.1209–1218.

Scholten, H.J., Van de Velde, R. & Van Manen, N., 2009. *Geospatial Technology and the Role of Location in Science*, Springer Netherlands. Available at: http://www.springer.com/earth+sciences+and+geography/geographical+information+systems/book/978-90-481-2619-4 [Accessed December 16, 2012].

Shafer, G., 1976. *A mathematical theory of evidence*, Princeton N.J.: Princeton University Press.

Shi, WenZhong, 2010. *Principles of modeling uncertainties in spatial data and spatial analyses*, Hong Kong: CRC press,Taylor and Francis Group, LLC.

Shu, H. et al., 2003. Uncertainty of Geographic Information and its Support in MADS. In International symposium on spatial data quality. Available at: http://infoscience.epfl.ch/record/99139/files/GIUncertainty_HK.pdf [Accessed October 16, 2010].

Skeels, M. et al., 2010. Revealing uncertainty for information visualization. *Inf Visualization*, 9(1), pp.70–81. Available at: http://dx.doi.org/10.1057/ivs.2009.1.

Slocum, T.A. et al., 2009. *Thematic Cartography and Geovisualisation* 3rd ed., Upper Saddle River, NJ, USA: Pearson Education International.

De Smith, M.J., Goodchild, M.F. & Longley, P.A., 2007. Conceptual frameworks for spatial analysis. In *Geospatial Analysis, a comprehensive guide to principles, techniques and software tools*. Leicester, United Kingdom: Troubador Publishing, pp. 37–38.

Trochim, W.M., 2006. The Research Methods Knowledge Base. *The Research Methods Knowledge Base*. Available at: http://www.socialresearchmethods.net/kb/ [Accessed December 17, 2010].

Tufte, E., 2001. *The Visual Display of Quantitative Information* 2nd ed., Graphics Press.

University of Groningen, 2012. The solution with Spearman's rho. *University of Groningen*. Available at: http://www.ppsw.rug.nl/~bve/stat/module8/3Spearman.html [Accessed September 1, 2012].

Venkatesh, V. & Davis, F.D., 2000. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *MANAGEMENT SCIENCE*, 46(2), pp.186–204. Available at: http://mansci.journal.informs.org/cgi/content/abstract/46/2/186.

Viard, T., Caumon, G. & Lévy, B., 2011. Adjacent versus coincident representations of geospatial uncertainty: Which promote better decisions? *Computers & Geosciences*, 37(4), pp.511–520. Available at: http://www.sciencedirect.com/science/article/pii/S0098300410003067 [Accessed July 25, 2011].

Vullings, L.A.E. et al., 2009. *Raamwerk voor omgaan met onzekerheid*, Wageningen: Alterra.

Websters Dictionary, 1986. Uncertainty. *Websters Third New International Dictionary*, volume 3, p.2484.

Zadeh, L., 1965. Application of fuzzy set theory. *Fuzzy sets, Information and control*, 8, pp.338–353.

Zar, J.H., 1972. Significance Testing of the Spearman Rank Correlation Coefficient. *Journal of the American Statistical Association*, 67(339), pp.578–580. Available at: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1972.10481251 [Accessed October 28, 2012].

Zhang, Q., 2008. *Animated representation of uncertainty and fuzzyness in spatial planning maps*. Enschede, The Netherlands: International institute for Geo-Information science and Earth observation.

Zuk, T. & Carpendale, S., 2006. Theoretical analysis of uncertainty visualizations. In *Proceedings of SPIE*. pp. 66–79.