GIMA
Geographical Information Management and Applications

Quality assessment of professional and VGI geo-data in The Netherlands

OpenStreetMap
LocatUs®
Google Maps

Author Stefan van den Berg
Supervisor drs. Maarten Zeylmans van Emmichoven
Professor prof. drs. Stan Geertman

Universiteit Utrecht
TU Delft
University of Twente.
Wageningen UR

This page is intentionally left blank

Acknowledgements

This master thesis titled ‘quality assessment of professional and VGI geo-data in The Netherlands’ is the result of an intensive six months of research in the field of Geographic Information Science. I would like to thank my supervisor Maarten Zeylmans for his assistance and support during this period. The motivation for doing this research came from him as he was genuinely curious of the geographic data quality of Locatus, Google Maps and Locatus in The Netherlands. I am glad to have systematically answered this curiosity with this research.

I would also like to thank Maartje Poelman from the Healthy Urban Living project at the University of Utrecht and Marco Groeneveld from Locatus. Maartje helped me find a relevant use-case and informed me on ongoing research within her research project which requires geographic data and Marco provided me with the necessary Locatus data which I could use for free.

I hope this thesis can be a contribution to the research field of geographic data quality.

Stefan van den Berg,
Utrecht, February 22nd 2018

List of abbreviations

AGI	Asserted Geographic Information
API	Application Programming Interface
BAG	Basic registration of addresses and buildings
CBS	Central Bureau for Statistics
CQ	Crowd Quality
GIMA	Geographic Information Management and Applications
ISIC	International Standard Industrial Classification of All Economic Activities
ISO	International Standardization Organisation
NACE	Statistical Classification of Economic Activities in the European Community
OSM	Openstreetmap
POI	Point of Interest
RMSE	Root Mean Square Error
SBI	Standard Industrial Classification
VGI	Volunteered Geographic Information

Abstract

The increasing popularity of VGI datasets like Openstreetmap makes it interesting to research what the quality of this type of geographic data is relative to professional geographic data. This research assesses the quality of two professional datasets (Google Maps and Locatus) and one VGI dataset (Openstreetmap) in the city of Utrecht, The Netherlands. Geographic data quality is classified in three main categories in order to systematically compare the quality of the test datasets. The first category is the intrinsic quality where the fundamental metrics are completeness, attribute accuracy and positional accuracy. The second category is the pragmatic quality which is determined by a use-case from the Healthy Urban Living project at the University of Utrecht. The third category is extrinsic quality which only applies to VGI data sources and is determined by the experience of the data producer. Overall, the quality of the Locatus dataset is relatively the highest in the study area where this dataset performs better in the analysis of the intrinsic and pragmatic quality compared to Google Maps and Openstreetmap. These datasets have a comparable intrinsic quality, but differ a lot in pragmatic quality which is moderate for Openstreetmap and poor for Google Maps. The extrinsic quality analysis of Openstreetmap results in high trustworthiness of this dataset in the study area, because almost all contributions in this area come from experienced mappers.

Keywords: geographic data quality, VGI, POI, Openstreetmap, Locatus, Google Places API

Table of contents

Acknowledgements 3

List of abbreviations 4

Abstract 5

1. Introduction 8

1.1 Motivation 8

1.2 Objectives 9

1.3 Research questions 10

1.4 Study-area 10

1.5 Report structure 12

2. Related work 12

3. Theoretic background 14

3.1 The development of Web 2.0 and the corresponding revolution in Geography 14

3.2 Volunteered Geographic Information 15

3.3 Quality assessment of geographic information 16

3.4 Summary 20

4. Ground truth 21

5. Methodology 24

5.1 Intrinsic quality metrics 24

5.3 Pragmatic quality metrics 27

5.4 Extrinsic quality 27

6. Data 29

6.1 Data description 29

6.2 Data inspection 31

6.3 Data processing 32

7. Results 33

7.1 Intrinsic quality 33

7.2 Pragmatic quality 35

7.3 Extrinsic quality 36

8. Discussion 38

8.1 Intrinsic data quality 38

8.2 Extrinsic data quality 39

8.3 Pragmatic data quality 39

8.4 Research value 39

9. Conclusion 40

10. Bibliography **42**

Appendices **46**

Appendix I: MySQL database creation 46

Appendix II: Google Place API Python script 47

Appendix III: Number of addresses per street in the city centre of Utrecht 48

Appendix IV: Random number generator 51

Appendix V: extended description of CBS classification 52

Appendix VI: Classification error matrices 56

Appendix VII: Positional accuracy tables 58

Appendix VIII: Pragmatic quality table 64

Appendix IX: OSM edits per user in Utrecht (2013 – 2017) 65

Appendix X: Results intrinsic quality analysis Nova shopping area 69

Appendix XI: Results intrinsic quality analysis Utrecht city centre 69

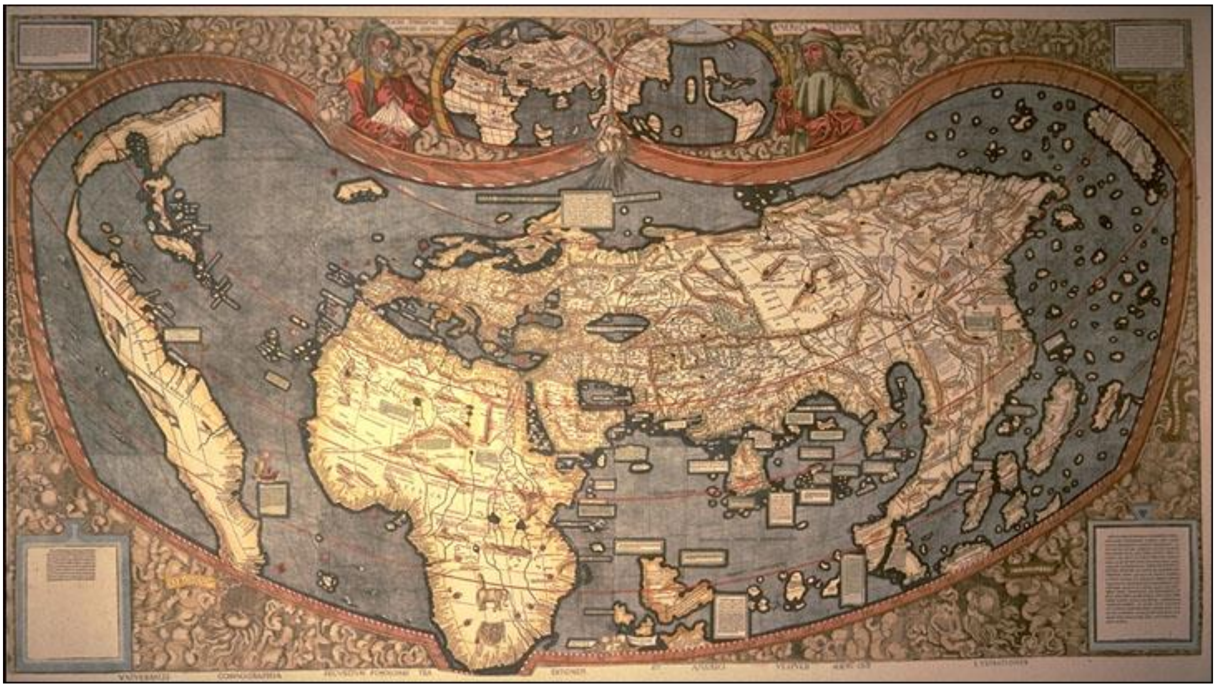
1. Introduction

1.1 Motivation

Geographic information has been of great importance throughout the history of mankind. Maps allowed sailors to travel around the world and helped people make sense of the wilderness. Where they used to be tangible pieces of paper, maps are nowadays digitally accessible and essential in the daily life for navigation and simple map queries. Easy-to-use mapping services of multinationals like *Google*, *Yahoo!* and *Microsoft* are booming, especially after the rise of mobile devices (Turner, 2006). Because of the importance of geographic information, a lot of different organisations strive to profit by offering it to the public. *Google* for example spends one billion dollar annually on maintaining their data on *Google Maps*.

Wroclawski (2014) argues in *The Guardian* that '*as more private companies offer us maps, we need an open-source, editable solution - a cartographical Wikipedia'*. This solution implies that the public, instead of professional organisations, is responsible for the collection and maintenance of its own geographic information. Goodchild (2007) discusses in his paper '*Citizens as sensors: the world of volunteered geography'* the first citizen who influenced in geographic information back in 1507. This was Martin Waldseemüller who labelled his own drawing of a new continent as 'America', even though he was not qualified to do so. Goodchild states that (p. 212) "*the events of 1507 provide an early echo of a remarkable phenomenon that has become evident in recent months: the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information, a function that for centuries has been reserved to official agencies*". The involvement of the public in geographic information is termed Volunteered Geographic Information (VGI). Platforms like *Openstreetmap*, *Wikimapia* and *Flickr* are examples of online places where the public can create, maintain and share their geographic information.

Figure 1: Composite image of the World Map by Martin Waldseemüller (1507)



Source: Hessler (2006)

The data collectors on these platforms are often termed ‘non-specialists’ or ‘amateur geographers’ (Mooney et al., 2010). *The Washington Post* (Hui, 2016) discusses groups of amateur geographers in the District Area of Washington that keep the *Openstreetmap* information of their local environment up-to-date. According to one of the amateurs “*it is up to the citizens to be the eyes and ears for local officials*”. Besides the easy-to-use and free online mapping services, there are a lot of commercial organisations selling more complex and extensive geographic information. Satellite images, high-resolution topography and soil information are examples of this information which is collected and maintained by professionals who follow a methodology specified by their organisation. This is the most significant difference with VGI which is collected by non-specialists who do not follow a certain methodology (Mocnik et al., 2017). This, however, does not mean that VGI cannot be a competitor of commercial datasets. *Openstreetmap* is currently the biggest VGI platform with more than 4.4 million users worldwide and has increasingly been implemented in professional projects (Dorn et al., 2015).

Due to the potential of VGI and the uncertainty of its quality, it is interesting to investigate the actual quality of VGI in comparison to professional geographic information. An example where the choice of data is relevant is the ‘Healthy Urban Living project’ at the *Utrecht University* in The Netherlands where research is conducted across three linked research pillars focussing on health challenges, healthy lifestyles and healthy environments. In this project, geographic information of ‘points of interest’ (POI) is required in order to research the urban environment. POI can be historical sites, landmarks, public services, shops, restaurants or bars and are in this research: points related to retail (De Tré et al., 2013). For the project there are several datasets that are plausible to use, but there is no real quality framework to determine which one is of the highest quality.

The increased potential and use of VGI data makes it important to be aware of the quality in comparison to professional geographic information (Senaratne et al., 2017). Goodchild & Li (2012) state that VGI-data for scientific research is often inadequate, but could be useful in the early stages. This report proposes an approach to research and compare the quality of VGI and professional geographic information.

1.2 Objectives

The main objective of this research is to assess the quality of professional and VGI points-of-interest (retail) datasets compared to each other in the city of Utrecht (The Netherlands). In order to reach this main objective, the first objective is to conduct a literature study. This results in an overview of existing theory regarding the assessment of geographic information in general and VGI.

(1) Determine quality metrics of geographic information in general and VGI for scientific research

The second objective is to collect geographic information of retail POI for the study-areas in order to research the quality of the test datasets. The professional test data sources will be *Locatus* and *Google Maps* and the VGI data source will be *Openstreetmap*. The biggest challenge is in the collection of the *Google Maps* data, because this data can only be accessed by using an Application Programming Interface (API) which means this data cannot be used within GIS software like *ArcGIS*, but has to be pre-processed with a Python script.

(2) Collect retail POI datasets of the test sources

The third objective is to collect a ground-truth dataset which can be used as reference dataset. A lot of research assumes that a professional dataset should be used as reference in order to assess the quality of VGI (Girres & Touya, 2010; Jonietz & Zipf, 2016). However, as Jonietz & Zipf (2016, p. 7) acknowledge: '*the traditional assumption of authoritative or commercial datasets being of a higher quality compared to VGI is no longer fully reliable*'. That is why ground truth has to be collected empirically by the researcher in order to be able to assess the quality of both VGI and professional datasets.

(3) Empirically collect a ground-truth dataset of the study areas which can serve as a reference dataset

The fourth objective is to determine the quality of retail POI information of the test datasets and compare the outcomes in the study-areas.

(4) Determine the quality of the test datasets and compare the outcomes in the study-area

1.3 Research questions

How can the quality of professional geographic information and VGI of Points-Of-Interest be assessed and to what extent does the quality of this information differ within the city of Utrecht?

- What quality metrics determine the quality of geographic information?
- How can ground truth of geographic information be collected empirically and how can this data serve as a reference dataset?
- What is the quality of *Openstreetmap*, *Locatus* and *Google Maps* datasets in the study-areas?
- To what extent can differences in quality between the test datasets be noted?

1.4 Study-area

The study-area of this research is the city of Utrecht in The Netherlands. Within this city a comparison between the city centre (study area 1) and suburban shopping area *Nova* (study area 2) in the district *Kanaleneiland* is drawn. This comparison is made, because according to Zielstra & Zipf (2010) there data quality is significantly higher in inner cities indicating that city centres receive more attention by mappers than suburban areas. The focus on retail POI has resulted in the choice of a shopping area in the suburbs, because that is where retail POI are present. Figure 2 and figure 3 present both study areas and their locations within the city. Note that the city centre is much larger than the suburban shopping area. Therefore not the complete centre, but a sample is researched. The details on the sample selection can be found in chapter 4: ground truth.

Figure 2: Study area 1: City centre of Utrecht (divided in neighbourhoods)

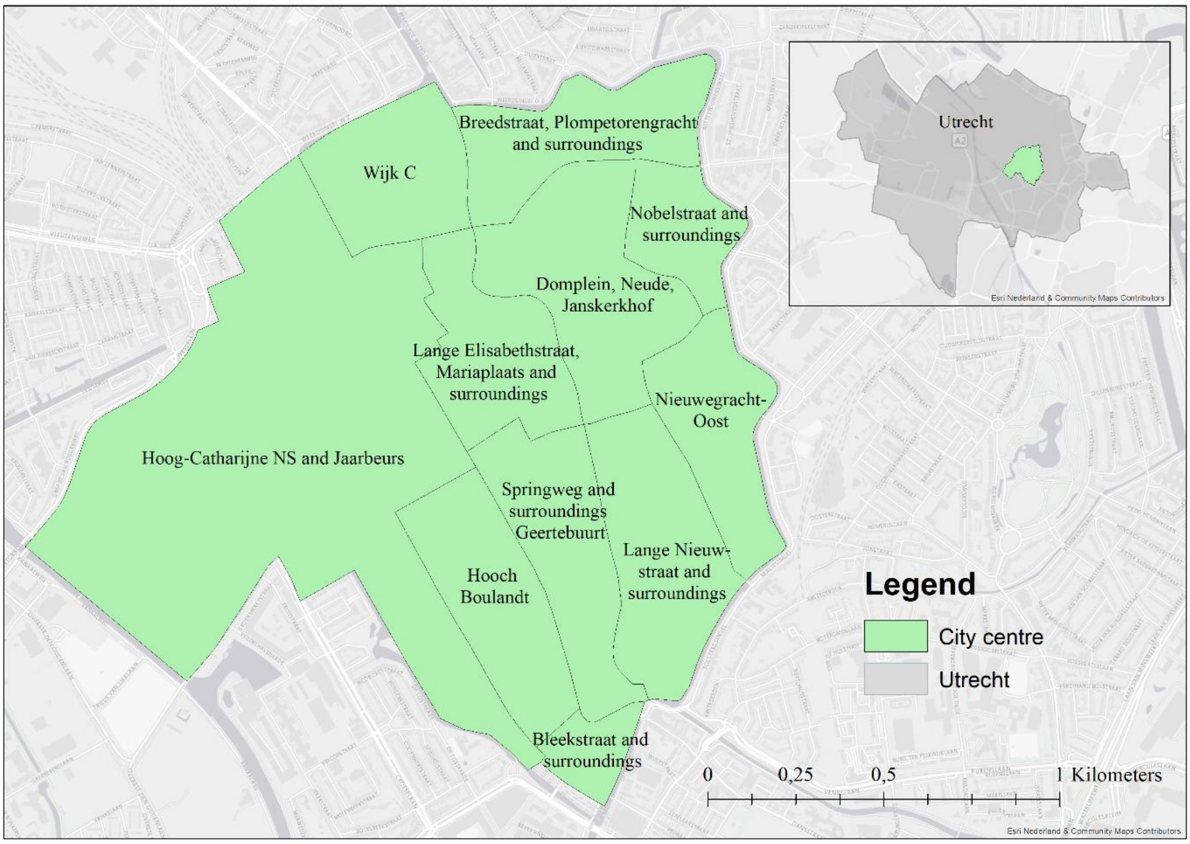


Figure 3: Study area 2: Nova shopping area

1.5 Report structure

This thesis is structured in nine chapters with this introduction chapter as the first. Chapter two briefly outlines the most critical similar researches that have been conducted in the field of VGI quality assessment. The third chapter also discusses scientific literature, but fixates more on the background theories instead of considerably similar researches. This chapter starts with historical context on how VGI has emerged, then discusses quality assessment of geographic information in general and in particular VGI and ends with a conceptual model. In the fourth chapter the collection methodology of the ground truth dataset is discussed after which the fifth chapter, methodology, discusses the research procedure and operationalisation. Here is disclosed in what methods the research objectives are reached. The sixth chapter discusses the background and collection procedure of the test datasets that are central in this research. The results are then presented in chapter seven, followed by the discussion of the results (chapter eight) and conclusion of this thesis (chapter nine).

2. Related work

Researching the quality of VGI is a hot topic in scientific literature. Especially the quality of the *Openstreetmap* platform is often attempted to assess. In their paper on existing literature regarding the quality of VGI, Senaratne et al. (2017) have reviewed various quality measures and indicators for VGI and existing quality assessment methods based on a selection of 56 scientific papers. They quantify the papers based on the quality metrics that are discussed and the methodology that is applied. Girres & Touya (2010) have attempted to create the most extended assessment of VGI (based on 8 geo-data quality metrics) for OSM in France. In most other literature the focus is on one or a few quality metrics and most researchers do not distinguish between VGI and professional spatial data metrics.

The most researched quality metrics in scientific literature are completeness and positional accuracy. As Haklay (2010, p. 687) states: "*positional accuracy is an ancient issue in mapping science and therefore must be tested and completeness is an outstanding issue in VGI, because there is no top-down coordination to ensure systematic coverage*". Zielstra & Zipf (2010) focussed solely on the completeness of OSM in and around 5 German cities where the *TeleAtlas MultiNet* dataset served as reference. Their conclusion was that there is a very strong heterogeneity in the OSM data and that completeness is significantly higher in inner cities compared to rural areas. They link this result to the presence of more active contributors in larger cities. Helbich et al. (2012) on the other hand focussed in their research solely on positional accuracy in an unnamed German city where the accuracy of OSM is concluded high and suitable for small and medium scale applications. Haklay (2010), Girres & Touya (2010) and Cipeluch et al. (2010) researched both completeness and positional accuracy in respectively London, France and Ireland. Haklay found a positional accuracy of 5.8 meters for 100 sample points with small differences within the city and Girres and Touya resulted with a positional accuracy of 6.7 meters for the same sample size in France. In their work the conclusion is also that the quality of the OSM dataset is strongly connected to the number of volunteers in a certain area. Hochmair et al. (2018) researched the positional accuracy of Google (N=51) and OSM (N=55) sample POIs in downtown Salzburg where they did not find any positional errors. This is due to their assumption that a POI located in front of the correct building is correct and is assigned an offset distance of 0.

The research of geo-data quality metrics requires a reference dataset which can serve as 'ground truth' in the assessment of several metrics (completeness, positional accuracy, attribute accuracy). This reference dataset is a professional dataset of which the quality is assumed to be of a high level. Examples of researches using a reference dataset are Girres & Touya (2010) who compared the quality of OSM in France with BD TOPO data; Kounadi (2009) and Haklay (2010) who did the same for respectively Athens and London with HMGS and Ordnance Survey datasets as reference; Helbich et al. (2012) who compared Tele Atlas and OSM in a German city and used Ordnance Survey datasets as reference. However, "*In many cases there is no access to the correct data. Therefore, we would be forced to look for an alternative method*". Esmaili et al. (2013, p. 13) suggest an alternate method to determine positional accuracy based on comparing the existing data of the same place with each other according to the metadata that their creators have given. Another alternative is the collection of ground-truth data in the field. Cipeluch et al. (2010) compare Google Maps, Bing Maps and OSM data in Ireland with ground-truth data and local knowledge. The paper unfortunately lacks to explain the methodology of the ground-truth data collection.

The overall conclusion of most papers is that there is a lot potential in VGI, especially because it is easily and freely accessible, but the heterogeneity of the datasets result in a lot of uncertainty regarding the data quality. This quality can be quite high as Girres & Touya (2010) and Haklay (2009) state about the positional accuracy of OSM. However the quality seems to be strongly connected to the number volunteers in a certain area and the mapping locations, indicating that the VGI quality is higher in densely populated areas (Cipeluch et al. 2010; Zielstra & Zipf, 2010).

3. Theoretic background

3.1 The development of Web 2.0 and the corresponding revolution in Geography

Since the beginning of the Internet, about 25 years ago, it has increasingly been used to provide geographic information to the public. The first online geographic application was the *Xerox PARC Map Viewer* which was introduced in 1993 and allowed requests for Global maps or maps of the USA which returned a HTML-file including an image of the requested map (Haklay et al., 2008). This first attempt of making maps accessible on the Internet has evolved over the years with a lot of projects like *Multimap* (1995), *GeoInfoMapper* (1997) and *ArcIMS* 3.0 (2000). The fraction of the Internet where geographic information is provided to the public is termed the ‘Geographic World Wide Web’ or the ‘Geo-Web’. The (Geo-)Web has over time developed from a place where professionals provided the public with information into a place where the public could create, develop, share and use information through many different applications (Haklay et al., 2007). The result of this development is called *Web 2.0* and was first described by O’Reilly (2007, p.16) as ‘*the network as platform, spanning all connected devices*’ where it is possible for users to ‘*provide their own data and services in a form that allows remixing by others*’. In relation to the Geo-Web the result of the web development is termed *Web-Mapping 2.0*.

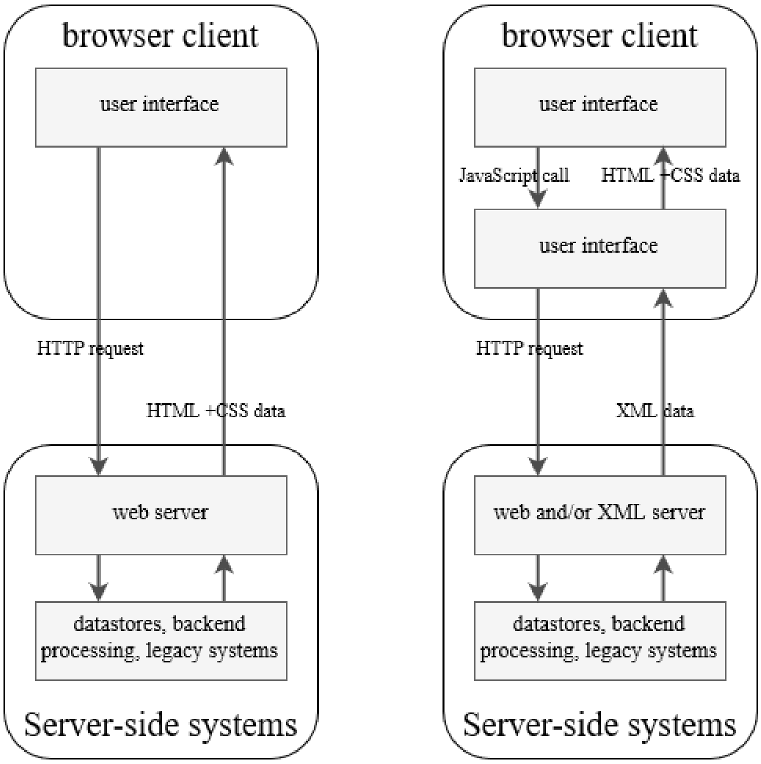
The evolution of the Geo-Web caused a revolution in the scientific Geography field where the distinction between ‘expert’ and ‘amateur’ already started to blur before Web-Mapping 2.0 due to globalisation and the familiarisation of people with more than just their local environment. The geographic knowledge of the public increased and could be expressed on the Geo-Web with the capabilities of the Web 2.0. This allowed anyone, expert or not, to create geographic information on the Internet. Non-experts are still framed as being ‘amateur geographers’, claiming that their contributions are inferior to that of professionals, where in fact being a professional does not necessarily guarantee high quality and vice versa (Goodchild, 2009, p. 91):

> “*If one takes professional qualifications as the distinguishing characteristic of the professional, Darwin was by modern standards an amateur ornithologist, Banks was an amateur taxonomist and Galileo an amateur astronomer*.”

The outcome of the revolution is named Neogeography, which is in short the ability of people to use and create their own maps from their own perspective and to mutually share location related information (Turner, 2006). For geographic features on the internet, three groups of technologies were of significant importance in the development into Web-Mapping 2.0: the Global Positioning System (GPS), Web 2.0 technologies that could be used in combination with geographic application and the development of Application Programming Interface (API). On May 1st of the year 2000, US president *Bill Clinton* announced the removal of the then existing selective availability of the GPS signal. According to Hayley et al. (2008) this is could be seen as the official birthday of Neogeography. This removal allowed the public to make use of a low-cost GPS-receiver with an accuracy of 6 to 10 meters (in normal conditions) in contrast to approximately 100 meters before. The Web 2.0 technologies are first of all the decrease in costs of big data transfer capacity, second of all the creation of standards for data transformation (by among others the OGC) based on XML and third of all the development of the AJAX approach (Haklay et al., 2008). This approach has ended the ‘start-stop-start-stop’ nature of the classic web application model by adding an AJAX engine which is responsible for both rendering the interface and communication with the server (see figure 4). This allows the user to communicate with an application independent of the server which steeply increases the processing time (Garrett, 2005).

The development of APIs allowed users with limited programming and server management knowledge to access more web-applications which increased the number of people that could create edit and delete geographic information on the web. Furthermore several APIs made it possible for this group to easily access a lot of databases which contain geographic information (Haklay et al., 2008).

Figure 4: The traditional (left) and the AJAX (right) web application model



Source: Garrett (2005)

3.2 Volunteered Geographic Information

The revolution in Geography as a result of the Web 2.0 created the opportunity for the public to collect its own geographic information. Data collected by the public is termed by Goodchild (2007) as 'Volunteered Geographical Information' (VGI). The Neogeographers that collect this data are largely untrained, their actions are mostly voluntary and the accuracy level of the process is uncertain. The biggest VGI projects nowadays are *Wikimapia*, *Openstreetmap* and *Flickr*. *Wikimapia* encourages users to 'describe the whole world'; *OSM* is developing a free digital map of the world; and *Flickr* is compiling a vast resource of georeferenced photographs all around the world (Goodchild, 2009). VGI datasets are complex to implement, because they are created by a heterogenous group of people who all come from different backgrounds. That is why this data is sometimes termed Asserted Geographic Information. The content of the data is asserted by the creator without citation, reference, or other authority (Goodchild, 2007; Mocnik et al., 2017; Mummidi & Krumm, 2008). All the different contributors to a VGI project have different perceptions of their real-world environment. As a result, the created data can be interpreted differently by others who have a deviant perception of the same environment. This difference in mental representation can lead to confusion and means that this type of data is only useful when the users understand the producer's interpretations of the real world. Mocnik et al. (2017, p. 2) describe the difference between VGI and non-VGI in the following way:

> *"As VGI is, in contrast to non-VGI, created by many people, many contexts and groundings exist simultaneously. This is in contrast to non-VGI data, which is created by a small group of*

people or even only one person, often using only one perception method and having a fixed taxonomy or ontology in mind."

The heterogeneity of VGI does not necessarily have to be a disadvantage for the data quality. *Linus' Law* states that "*given enough eyeballs, all bugs are shallow"* (Raymond, 2001, p. 19), implying that the more people get involved in a project, the more individual mistakes can be eliminated. Haklay et al. (2010) researched if *Linus' Law* is valid for OSM road data in England by comparing the average positional accuracy with the number of contributors. Even though this relation is not linear, there seems to be a connection between the two variables which legitimises *Linus' Law*. Other advantages of VGI data is that it is mostly free of charge, up-to-date and sometimes is the only existing source of geo-information (Goodchild, 2007; Goodchild & Li, 2012).

3.3 Quality assessment of geographic information

It is important to be aware of the quality of geographic information, because it is impossible to make a perfect representation of the earth. This means uncertainty is inevitable and quality control is necessary (Longley et al., 2011). In order to define the quality of any type of information source, it is important to start with a relevant definition of the term 'quality'. The International Standardisation Organisation (ISO) created guidelines which describe general procedures for evaluating the quality of geographic information. This standard, geographic information – data quality (ISO 19113-15), indicates two main categories of geo-data quality. The first, internal quality, includes six quality metrics of intrinsic characteristics which influence the quality of geographic data: completeness, thematic accuracy, logical consistency, temporal quality, positional accuracy, and usability (Dorn et al., 2015; Jonietz & Zipf, 2016). The second, external quality includes the fitness for use of the data. This quality is determined by the usability in applications and depends very much on metadata and documentation (Bucher et al., 2016).

Where the standard of the ISO divides geographic data quality in internal and external quality, Criscuolo et al. (2016) propose an additional category especially for crowdsourced resources: the trustworthiness of geographic data. The focus of this category is on the data producer and applies specifically on VGI data sources, because these are heterogeneous of nature and therefore the data producer is less trustworthy. The inclusion of one VGI data source in this research makes this category very relevant and therefore it will be taken into account. In this classification the internal quality is *termed intrinsic quality*, the external quality is termed *pragmatic quality* and the new trustworthiness group is termed *extrinsic quality*. This classification of Criscuolo et al. (2016) is a guideline for this research.

3.3.1 Intrinsic quality metrics

"A *specific intermediate quality concept is needed to document inherent characteristics of geographical data that will be useful for every user to evaluate their ability to fulfil their application requirements*" This is how Bucher et al. (2016, p. 133) charachterise internal data quality. Van Oort (2006) created a extensive list of intrinsic quality metrics in his PhD research on geographic data quality at *Wageningen University* (table 1). His list was acknowledged by many researchers in this field (Haklay, 2010; Haklay et al., 2010; Helbich et al., 2012; Van Exel et al., 2010) and was based on Aronoff (1989) who interpreted the draft of USA-SDTS from a management perspective, USA-SDTS (1992) which is The United States of America spatial data transfer standard, ICA (1995) which is the *International Cartographic Association*, CEN/TC287 (1998) which is the Technical committee 287 of the *Comité Européen de Normalisation* (CEN) and ISO/TC211 (2002) which is the Technical committee 211 of ISO. Van Oort (2006) classified all elements existing in the sources in three categories: explicitly recognised in the spatial data quality section of the metadata (S), explicitly recognised as an element (M), in another section of the metadata and implicitly recognised as an element (I). In total 11 elements are recognised at least once. Five of these metrics (lineage, semantic accuracy, variation in quality, meta-quality and resolution) are not included in the ISO 19157 standard. Semantic accuracy is only explicitly recognised as a metric by ICA and CEN where van Oort (2006, p. 15) states: "*now that the development of the ISO standards is completed, CEN will adopt the ISO standard so that the element will most likely disappear altogether*". Because of this, semantic accuracy is furthermore disregarded. The last three missing metrics are often encountered not as individual elements but as sub-elements of other metrics. This leaves lineage as the only missing metrics in the ISO standard. Because the focus of this research is on assessing the quality of point data the metric logical consistency is also not taken into account. This metric determines the quality of the topology of the nodes and polygons in the dataset and not the points. According to Goodchild & Li (2012) and Bucher et al. (2016) positional accuracy, attribute accuracy, logical consistency and completeness are the four fundamental dimensions of intrinsic data quality. Therefore these dimensions (apart from logical consistency which is irrelevant) are the focus of this research.

Table 1: Intrinsic quality metrics of geographic data

	Aronoff (1989)	**USA-SDTS (1992)**	**ICA (1995)**	**CEN TC287 (1998)**	**ISO TC211 (2002)**
Lineage	S	S	S	S	S
Positional accuracy	S	S	S	S	S
Attribute accuracy	S	S	S	I	S
Logical consistency	S	S	S	S	S
Completeness	S	S	S	S	S
Semantic accuracy	-	-	*S*	*S*	-
Usage, purpose, constraints	S	M	-	S	M
Temporal quality	S	M	S	S	S
Variation in quality	-	I	I	S	I
Meta-quality	-	I	I	S	I
Resolution	S	I	I	I	M
S =	explicitly recognised in the spatial data quality section of the metadata				
M =	explicitly recognised as an element				
I =	in another section of the metadata and implicitly recognised as an element				
- =	not recognised				

Source: van Oort (2006)

The positional accuracy of a dataset refers to the accuracy of the included spatial components. Researching accuracy is in fact in most cases researching error with the proposition: a lower error correlates with a higher accuracy. For point data, error is usually defined as the discrepancy between a test location and a reference location. Error can be measured horizontal (X and Y) and/or vertical (Z). The most common ways to determine the spatial error of a set of points is by the calculation of the Mean Error (ME) or the Root Mean Squared Error (RMSE). The ME is logically the mean distance between a test points and a reference point. This method can however be sensitive to outliers. The RMSE measures the error between two datasets by quantifying the differences of the values. The smaller the error is, the more accurate the data (Chrisman, 1991; Veregin, 1999).

The most common method of assessing attribute accuracy is creating a 'classification error matrix' which allows for the calculation of this accuracy. This methodology has its origins in the field of remote sensing where it is used to determine the accuracy of image classifications (Veregin, 1999). This matrix tests the categories of a classification in a structured manner. Reference data is compared to the classified data and expressed in an error matrix where the classification would be perfect if apart from the major diagonal (correct classifications) all records are 0. The overall accuracy of the attributes can be calculated by dividing the sum of the major diagonal by the total number of records.

Figure 5: An example of the classification error matrix

		Reference Data			
		F	W	U	Row Total
Classified Data	F	28	14	15	57
	W	1	15	5	21
	U	1	1	20	22
Column Total		30	30	40	100

Sum of the major diagonal = 63

Overall Accuracy = 63/100 = 63%

Source: Story & Congalton (1986)

In addition the accuracy of the classes separately can be determined in order to note where the largest error has occurred. The producer's accuracy is calculated by dividing the correctly classified samples by the total number of reference samples in that category. On the other hand, the user's accuracy is calculated by dividing the correctly classified samples by the total number of classified samples in that category. It is important to understand the difference, because a high producer's accuracy like class 'F' in figure 5 (28/30 = 93%) does not necessarily indicate a high user's accuracy (28/57 = 49%). This means that 17 records are falsely classified as 'F' which are in fact 'W' or 'U' (Story & Congalton, 1986).

The next metric, completeness, assesses the extent of the errors of omission and the errors of commission in a database. When in a certain database features are excluded that should have been included, this indicates an error of omission. Likewise finding a feature in the dataset that should not have been included indicates an error of commission (van Oort, 2006; Veregin, 1999). An example:

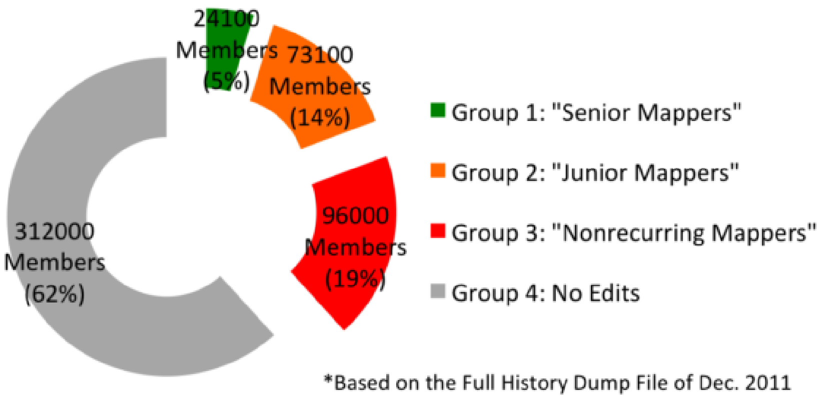
> "*An abstract universe states that a dataset contains hospitals and defines a hospital as a building in which one or more doctors are employed and where the average annual number of patients is at least 100. A building with an annual average of 90 patients does not belong to this abstract universe. If this building is present in the dataset then that is an error of commission (resulting in overcompleteness). A building with average 110 patients and with doctors employed does belong to the abstract universe. If this building is not present in the data set then the data set is incomplete due to an error of omission*" (van Oort, 2006, p. 16).

3.3.2 Extrinsic quality metrics

The extrinsic quality of geographic data relates to the credibility of the author and the reliability of the information (Criscuolo et al., 2016). Where many scientific papers agree that VGI can be assessed with the same metrics as professional geographic datasets (Senaratne et al., 2017), some argue that VGI specific quality metrics are required on top of the standard metrics, because of the heterogenic nature of VGI (Mooney et al., 2010).

VGI specific indicators are in general more abstract and mostly relate to the data creator. The most mentioned metrics are: trustworthiness, credibility, vagueness, local knowledge, experience and recognition (Senaratne et al., 2017). Trustworthiness of data is based primarily on subjective factors (e.g. reputation, reliability, trust) and determines whether the data receiver can trust the quality of the data. As such, the outcomes of this indicator for the same dataset may be differently, because different people can have different attitudes towards the data. Together with expertise the trustworthiness determines the credibility of VGI data. In order to assess this, it is necessary to know who the data-creator is which is not in all cases possible (Flanagin & Metzger, 2008; Senaratne et al, 2017). Vagueness determines to what extent the perception of a data-collector is geographically uncertain (e.g. what are the boundaries of a certain area?). This indicator focusses on perception rather than measurements (De Longueville et al., 2010). The last three VGI quality indicators (local knowledge, experience, recognition) are created by Van Exel et al. (2010) who introduced the concept of Crowd Quality (CQ). This is based on a two-dimensional approach where existing metrics are complemented by user-related metrics. These metrics are local knowledge, experience and recognition. The first one, local knowledge, influences user-related quality, because awareness of the real-world environment allows a contributor to correct data or add missing data more easily. The second indicator, experience as a contributor, determines the level a data producer is at and therefore the quality of this person's contributions. This can be quantified by looking at register time or number of added or adjusted features. Goodchild & Li (2012) suggest volunteers with a lot of experience can act as gatekeepers who maintain and control the quality for a certain area. The third indicator, recognition, is a quality indicator which is determined by the ranking a contributor receives from the online community. Examples are the rating of a member on *Ebay* or the reputation level on *Stack Overflow* (Maué, 2007; Van Exel et al., 2010).

Figure 6: Classification of the experience of OSM contributors



Source: Neis & Zipf (2012)

Even though there are six VGI-specific metrics according to the literature, only one of them seems to be quantifiable for this research: 'experience'. Trustworthiness, credibility and vagueness are subjective concepts and difficult to research. Furthermore, local knowledge requires a lot of information of the data suppliers which is out of scope for this research and recognition is not relevant, because OSM does not have some sort of ranking among its contributors. The experience metric however can be researched by quantifying the number of contributions. Neis & Zipf (2012) quantified the contributions in four groups: senior mappers (>1.000 created nodes), junior mappers (10-1.000 created nodes), nonrecurring mappers (<10 created nodes) and no edits (0 created nodes). This last group is by far the largest with 62 percent.

3.3.3 Pragmatic quality metrics

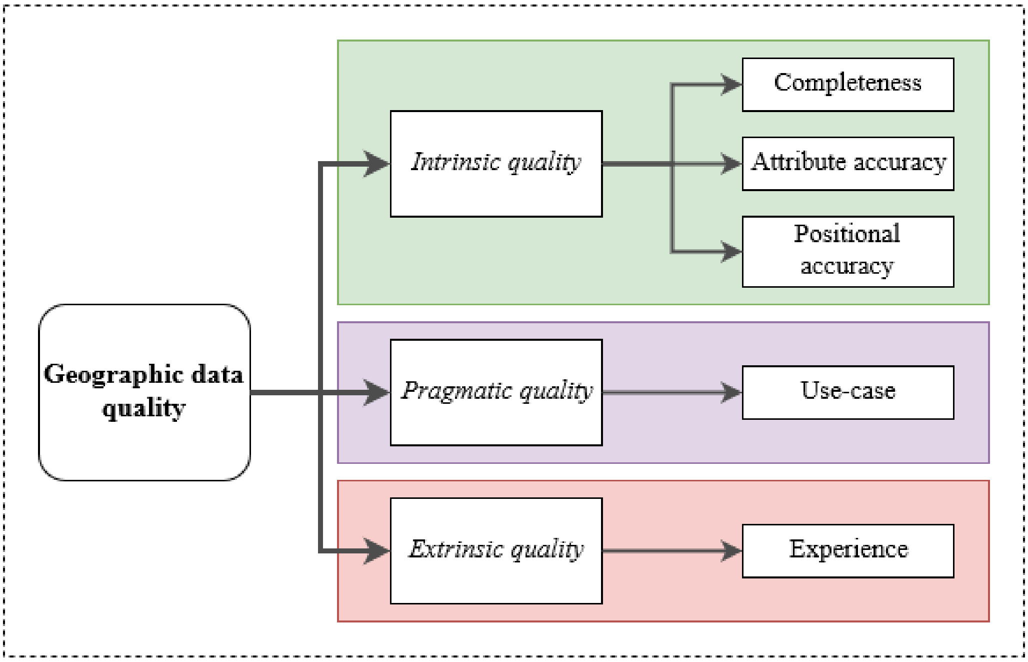
The pragmatic quality of a dataset is very dependent of the intended application it will be used in. A research of the shortest path from point A to point B has very different data requirements in comparison to a research of finding sufficient primary school locations. Where the data focus in the first research is on roads and connectivity, the focus of the second research is on factors influencing primary school locations. This example points out that the pragmatic quality of a dataset is related to the intended use-case (Bucher et al., 2016).

3.4 Summary

With the development of the Web into a place where the public is enabled to create, develop, share and use information (‘O Reilly, 2007), an opportunity emerged for geographic information on the Internet. Now, not only professionals, but everyone was able to provide this type of information (Haklay et al., 2007). Geographic data created by the public is termed by Goodchild (2007) as Volunteered Geographic Information (VGI) and is difficult to use, because the data source is a big heterogeneous group of volunteers from different backgrounds. Professional data on the other hand is created by a small group with mostly a singular perception methodology (Mocnik et al., 2017). However this does not necessarily mean that professional data is of higher quality than VGI data as the heterogeneity can also be seen as an advantage, because a lot of people can reach a high quality control (Haklay et al., 2010).

The uncertainty of the quality of VGI in relation to professional geographic information is central in this thesis. In order to determine quality, a clear definition for professional and VGI datasets is required. The conceptual model of figure 7 is an overview of this definition for this thesis. It is determined by three main categories of geographical data quality (Criscuolo et al., 2016). This is first of all the intrinsic quality, determined by quantitative metrics completeness, attribute accuracy and positional accuracy. Second of all, pragmatic quality is determined by the usability of a geographic dataset within a use-case. And third of all, extrinsic quality applies only on VGI data sources and is determined by the experience of the data producer. Mostly this is not necessary for professional data, because this is created by a homogeneous group of data collectors (van Exel et al., 2010).

Figure 7: Conceptual model



4. Ground truth

This chapter extensively explains the collection methodology of the ground truth dataset in the study areas which serves as a reference for the determination of the intrinsic and pragmatic data quality of the test datasets. The choice for empirical collection of ground truth instead of using a professional data source as reference is made, because the assumption that professional geographic information is of higher quality than VGI is problematic. Nowadays the quality of professional data is not necessarily higher than the quality of VGI which indicates that a professional dataset should not be used a reference (Jonietz & Zipf, 2016). Goodchild (2009, p. 88) criticizes the culture of professional products being of high quality and the framing of non-professional as incompetent:

> *"In the case of surveying and the creation of cadastral information, there is a longstanding tradition that the products of professional surveyors are of high quality. Terms such as professional convey an immediate sense of care, attention to detail and adherence to rigorously applied standards, whereas the very term amateur suggests poor quality and is even used pejoratively."*

In addition to the critic on the quality of professional data, the empirical collection of reference data ensures the data is focussed on this particular research and collected with a structured methodology. The ground truth dataset will be collected within a fraction of the city centre of Utrecht and within shopping area *Nova* in the suburbs. It is out of scope to collect a complete dataset of all existing retail POI in the city centre, because this is overly time-consuming. Therefore a random sample of the streets of the centre will be constructed where all existing retail POI will be collected.

For the sample to be created, first a database of all existing streets in the city centre has to be constructed. The number of addresses in a street determines the probability the street will be selected in the sample. The basic registration of addresses and buildings (BAG) which contains all addresses in The Netherlands and is maintained by the municipalities is adopted as data source. From the BAG all addresses within the city centre are selected and stored in a local database where the number of addresses for each street is counted and ordered descending (see appendix III for the result table). In total there are 301 streets where *Oudegracht* is absolutely the largest with 957 addresses. This means that this street has the biggest chance of being selected for the sample with a probability (P) of 7.19 percent. A random sample generator in *Python* (see appendix IV for the script) is used to select 10 streets based on their probability (number of addresses). In the streets selected for the sample all retail points are collected empirically and used as reference data. Table 2: shows the selected streets.

Table 2: The selected streets in the city centre of Utrecht

ID	Name
110	Choorstraat
11	Hartingstraat
35	Willemstraat
56	Minrebroederstraat
47	Oudkerkhof
49	Oranjestraat
1	Oudegracht
16	Vredenburg
93	Wolvenstraat
82	Lange Jufferstraat

The buildings in the streets selected in the sample are extracted from the BAG and are used as a basemap in the collection process. Figure 8 and figure 9 show the selected buildings in the Wilemstraat and Oudegracht. Retail POI are only collected for the selected buildings, all other buildings are not included in the sample and therefore not relevant.

Figure 8: BAG buildings Oudegracht

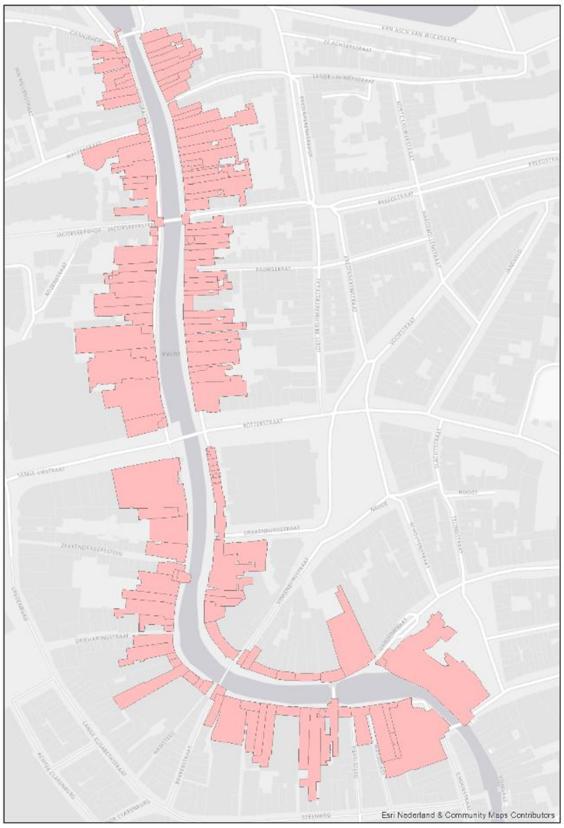


Figure 9: BAG buildings Willemstraat

When the sample is created, the next step is to establish a classification for the data that will be collected in order to assess the attribute accuracy of the test data. As the focus of this research is on retail data, the Standard Industrial Classifications (SBI) created by the Dutch Central Bureau for Statistics (CBS) is used as a guideline in the data collection process. This classification is based on the activity classification of the European Union (NACE) and on the classification of the United Nations (ISIC). Only the first three digits of relevant classes are used for this research, because zooming in too much can result in reclassification errors (see 4.2.1). The focus is on the 3-digit classes presented in table 3. Appendix V provides an extended description of the smaller digits for all used classes which makes the reclassification of the test dataset more straightforward.

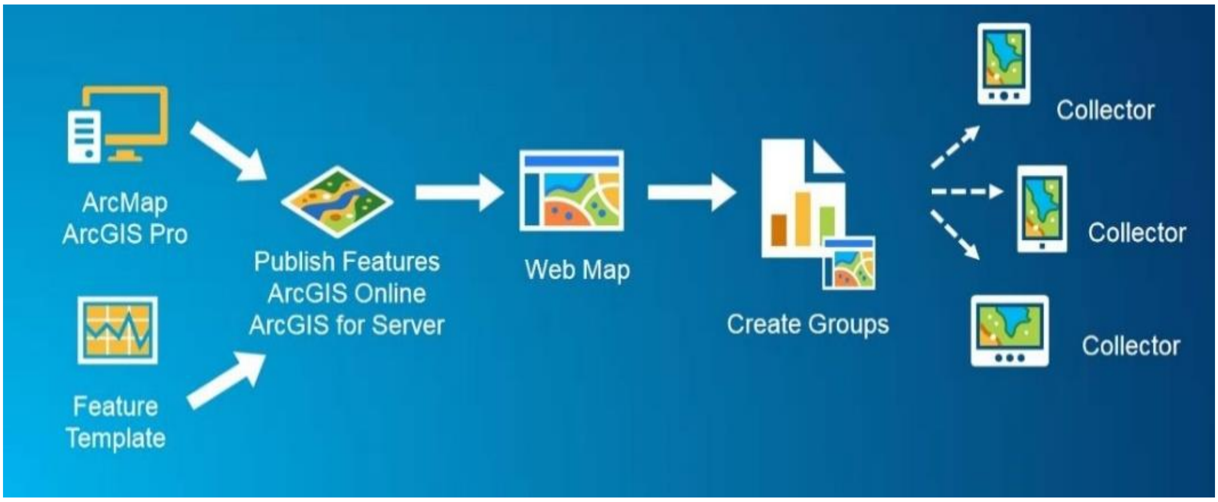
Table 3: Relevant 3-digit SBI classes

SBI class	Name
471	Retail sale in non-specialised stores
472	Specialised shops selling food and beverages
474	Shops selling consumer electronics
475	Shops selling other household equipment
476	Shops selling reading, sports, camping and recreation goods
477	Shops selling other goods
551	Hotels and similar accommodation

552	Renting of holiday bungalows and apartments; youth hostels and tourist camps
559	Other provision of lodgings n.e.c
561	Restaurants
563	Bars
910	Lending of cultural goods, public archives, museums, botanical and zoological gardens and nature reserves activities
931	Sports activities
932	Other recreation

The software that allows for easy collection of data in the field is the *Collector for ArcGIS* application, created by *Esri*. This application makes it possible to collect and update data in the field with a smartphone or a tablet by using an existing map or GPS. In this case the BAG buildings will be used as base layer to prevent measurement errors which can occur with GPS. Before collecting data in the field a geodatabase needs to be created where feature classes with the correct types are added in order to be able to store the data in a correct manner. Furthermore the information model has to be configured to meet the requirements of the data collection form. Once configured, a map should be built in *ArcMap* and published as a feature service on the *ArcGIS Online* platform where it can be shared with the collectors. In this case there is only one collector. When opening the application on a mobile device, the created map pops up and data (locations and names) can be collected within the indicated categories.

Figure 10: Workflow of Esri's Collector for ArcGIS application



Source: Esri (2017)

In the collection process all selected buildings are analysed in detail where after it is determined whether or not retail POI is present and in what category this point belongs. In some cases there is no clear distinction between a 'restaurant' (SBI 561) and a 'bar' (SBI 563), because a POI can be somewhere in the middle of both. Therefore in the collection of the ground truth, all POI where it is possible to have dinner are classified as 'restaurant'. Another important applied method is that all the points are placed near the entrance of the POI. When the entrance is not located in the selected street, the POI is not included in the ground truth dataset.

5. Methodology

5.1 Intrinsic quality metrics

This first paragraph of the methodology explains how the relevant intrinsic geographic data quality metrics described in the theoretic framework are operationalised. The paragraph ends with a flowchart summarising all steps that are conducted.

5.2.1 Completeness

From the four mentioned types of completeness (data, model, attribute and value) in the theoretic background, the focus is merely on data completeness since the focus of this research is on identifying to what extent the points of the reference dataset are included in the test datasets. The first step in finding the data completeness is a manual comparison between the reference and test datasets. The reason this step carried out manually is because this avoids errors related to an automatic process (Girres & Touya, 2010; Haklay, 2009). An example of an error that would occur with an automated process is a different spelling of the correct POI in the datasets. A fast food restaurant in the Nova shopping area is named in the reference dataset as the '*HFC café*'. This same restaurant is however named '*HFC Utrecht Kanaleneiland'* in the *Google Maps* dataset and '*Halal Fried Chicken*' in the *Locatus* dataset. These are three totally different spellings for the same POI. An automated process would detect an error of omission, because the spelling is different. However this is not an error, but just a deviating spelling of the same POI. To avoid these errors, the completeness is determined manually.

Determining the data completeness of the test datasets starts with the reclassification of the datasets into the SBI classes and filtering out the points that are not classified as one of the classes (meaning they are not relevant for this research). Thereafter, the manual comparison of the names in the databases takes place. To all correct POI a new ID corresponding to the same point in the reference dataset is assigned. The POI in the reference dataset which are not included in the test dataset indicate an error of omission.

$$\boldsymbol{error\ of\ ommision} = \frac{\text{Number of incorrect points}}{\text{Total number of points}} \times 100\%$$

Because the focus is on completeness, this error will be subtracted from 100 percent to result with a completeness percentage.

$$\boldsymbol{Completeness} = 100\% - (\text{error of commission})$$

5.2.2 Attribute accuracy

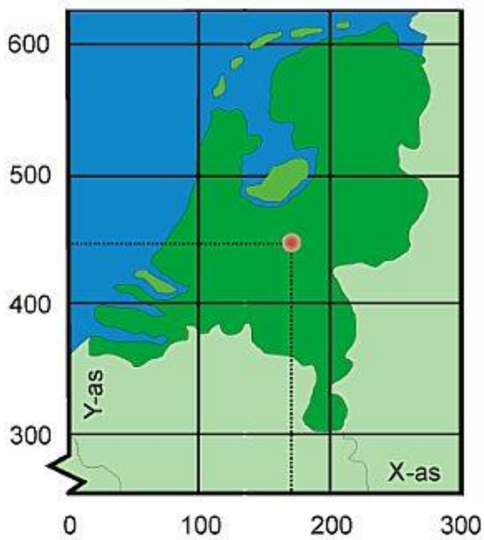
The reclassified correct points which resulted from the assessment of the completeness are used to create a classification error matrix (Story & Congalton, 1986). This matrix provides an insight in the attribute accuracy of the points that are included in the test and the reference dataset. The correct points are linked to the same points in the reference database with a simple SQL query (see appendix I). This results in two databases with the same points, classified as the SBI codes. From this data the matrix is created with the crosstab function in *SPSS Statistics*. From this error matrix the overall attribute accuracy can be calculated by dividing the sum of the diagonal by the total number of records.

5.2.3 Positional accuracy

The same 'correct points' used to determine attribute accuracy are used to determine positional accuracy. It is first of all necessary the change the coordinate system from a geographic to a projected system as this allows you to calculated distances in meters instead of in degrees. The *RD New projection* is applied which is focussed solely on The Netherlands. After the transformation the geometry (X and Y coordinates) of the reference (X_1 and Y_1) and test (X_2 and Y_2) points are calculated. These coordinates are subsequently used to calculate the Euclidean distance between each pair of points with the following mathematical function in *Python*:

```
math.hypot(!X1!-!X2!, !Y1!-!Y2!)
```

Figure 9: The RD New projection



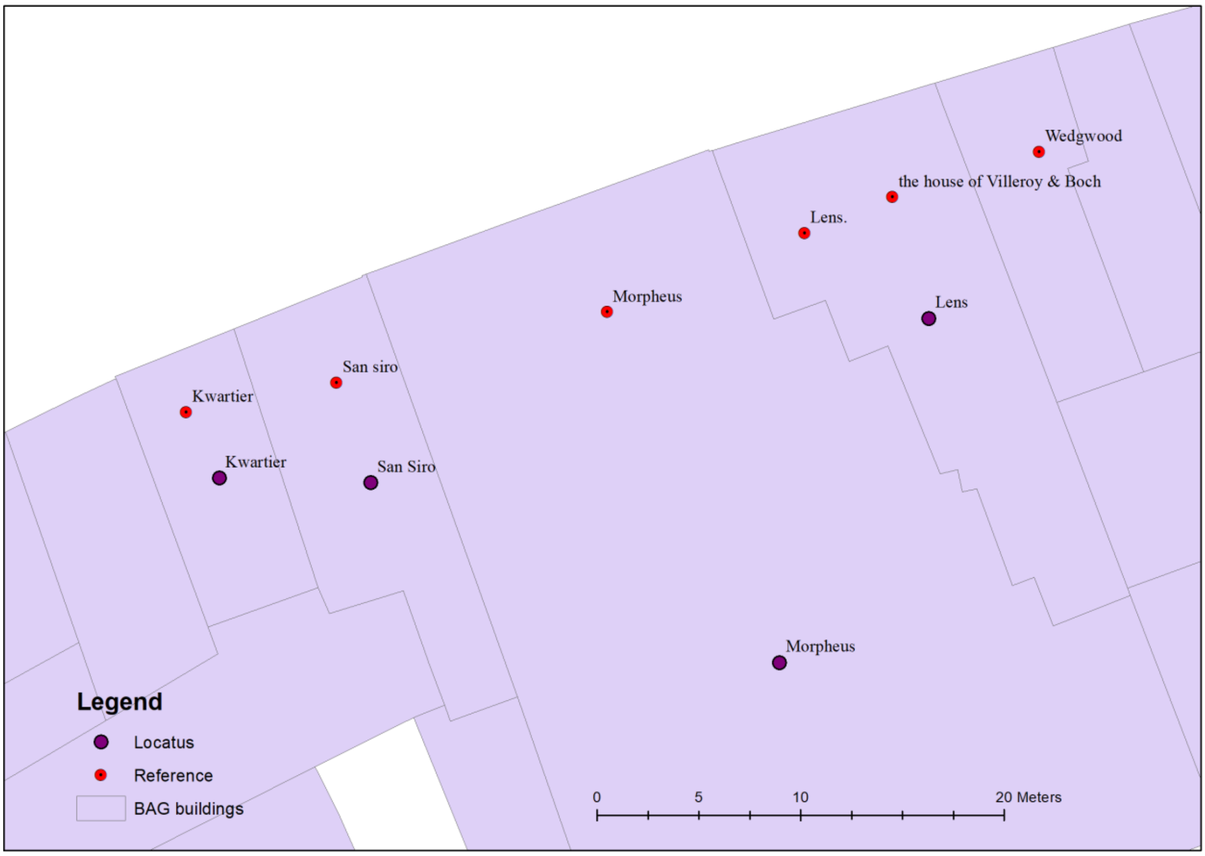
Source: Allesovergps (2017)

As stated in chapter 3, the most common way to determine the spatial error of a set of points is by calculation the Mean Error (ME) or the Root Mean Squared Error (RMSE). Because of the sensitivity of the ME to outliers both the ME and the RMSE of the test datasets are calculated. What applies for the positional accuracy of the test datasets is that the lower the ME and the RMSE, the higher the accuracy. This is a good indicator for the comparison between different datasets, because the numeric outcomes directly show which dataset is most accurate. However when dealing with POI it is important to keep in mind that a deviation in distance between a test point and a reference point does not directly indicate an error. This can be a consequence of a different placement method of the point within a building as can be seen in figure 11. The point 'Morpheus' is located in the correct building by Locatus, but is still 20 meters away from the reference point.

In order to determine the RMSE of the test datasets the coordinates of the collected points in the reference dataset are compared to the coordinates of the same points in the test datasets. Obviously, a difference in coordinates indicates a positional error. The actual RMSE can be calculated by means of the following formula where Σ = summation; P = predicted (expected values or unknown results); O = observed values (known results); and n = sample size:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P - O)^2}{n}}$$

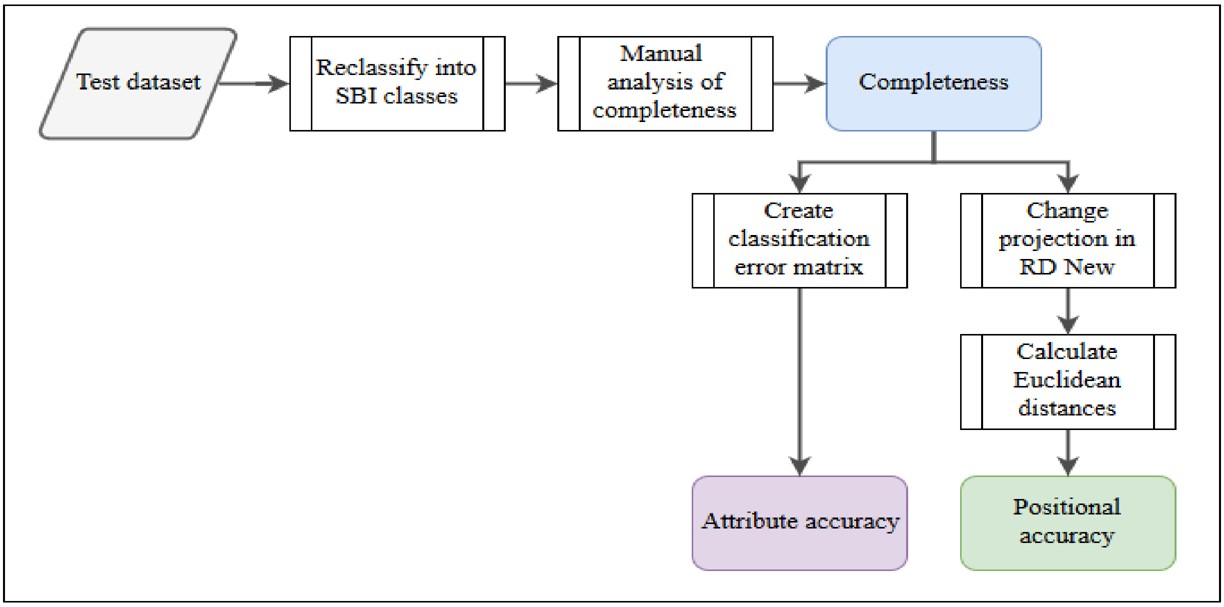
Figure 11: Deviating placement methodology in a part of the city centre of Utrecht



5.2.4 Flowchart intrinsic quality

Figure 12 is flowchart presenting a simplification of the methodology of the intrinsic quality analysis of the test datasets. It is striking that the determination of the completeness is a critical step in the analysis of the attribute and positional accuracy. In this first step irrelevant points are filtered out of the datasets and the correct points receive a new ID corresponding to the same point in the reference dataset. This allows for the linkage of the points which is necessary for the classification error matrix and the calculation of the Euclidean distances.

Figure 12: Intrinsic quality methodology



5.3 Pragmatic quality metrics

The pragmatic quality of geographic data can be determined by testing its usability in a practical environment. In this case, this environment is the *Healthy Urban Living Project* discussed in the introduction. One of the researches within this project requires data of fast food outlets in order to investigate the relations between the residential fast-food environment and the individual risk of cardiovascular diseases in The Netherlands. Fast food outlets generally sell food that consists of processed meat, refined carbohydrates and is high in salt, saturated fat and calories (Jaworowska et al., 2013). The rapid growth in the number of fast food outlets in the recent years can be a factor in the risk of cardiovascular diseases. Here the hypothesis is that proximity to fast food outlets influences the individual risk of cardiovascular diseases (Poelman et al., 2018). This study applies Tobler's first law of geography stating that "*everything is related to everything else, but near things are more related than distant things*".

For this particular research it is very important that the used dataset has accurately distinguished fast food outlets from other types of (food) outlets. In addition, it is important that the positional accuracy of the fast food outlets is high, because the location is used as a variable in the determination of the relation with cardiovascular diseases. Therefore the methodology of the pragmatic quality of the test dataset is somewhat similar to the methodology of the intrinsic quality. In this case the metrics only apply to fast food outlets (see table 4). The ground truth dataset will again serve as reference in order to determine the usability of the test datasets for this research.

Table 4: Pragmatic quality metrics

Completeness	Are the fast food outlets included?
Attribute accuracy	Correctness of the classification as fast food
Positional accuracy	Correctness of fast food locations

5.4 Extrinsic quality

The extrinsic quality is only determined of the VGI test dataset (*Openstreetmap*), because this is the only dataset where reliability of the data producer is relevant, because of the heterogenous group of data collectors. The only metric determining the extrinsic quality is identifying the experience of the data producers. Due to the difficulty of limiting the experience research to the identified study areas within the city of Utrecht, for this metric the entire city is used as a study area.

The first step in identifying the experience of the VGI contributors in the study area is classifying the number of contributions. The classification of mappers by Neis & Zipf (2012) in 'senior mappers', 'junior mappers', 'non-recurring mappers' and 'no-edits' will be the guideline for this metric. However, the users without edits will be disregarded since they do not influence the quality of the data and only contributions of the last five years are taken into account. The assumption here is that the area in which a user creates nodes is the area of residence of this user.

Table 5: Percentage of mapping categories worldwide

Category	Number of edits	Percentage
Senior mappers	> 1000	12
Junior mappers	10 – 1000	38
Non-recurring mappers	< 10	50

Source: Neis & Zipf (2012)

Table 5 shows the division of OSM contributors over the three categories worldwide in 2012. According to Neis & Zipf (2012) only the senior mappers, the mappers with more than 1,000 edits, contribute to the OSM project in a productive way. However, this disregards that a large group of contributors between the 500 and 1,000 edits can also make a difference in an area. That is way a new group is added to the table: 'medior mappers'. This group has between 500 and 1,000 edits which lowers the maximum number of edits for the junior mappers to 499 edits.

To determine the experience of the OSM users in Utrecht, data of edit history of all users of last five years is combined. This data can be collected in an online application called 'user editing summaries' (Anderson, 2016). This application divides the world in small tiles and allows to view the number of edits per user in a tile for each year. So in this case the number of edits per user of the four tiles dividing Utrecht is taken into account for the last five years. This data is stored in a local database where first the edits of users that have contributed in more than one year are combined. And second the edits of users for each square are combined (see appendix IX).

6. Data

This chapter discusses the test datasets in detail and describes how they are collected and if necessary pre-processed for analysis. Three different datasets are researched in this thesis where *locatus* and *Google Maps* are professional data sources and *Openstreetmap* is a VGI source. Other differences between the datasets is that *Locatus* is the only dataset that is costly to collect (the other two can be collected free of charge) and that *Google Maps* data cannot be collected in a *Esri* shapefile format.

Table 6: data sources overview

Organisation	Type	Costs	Format
Locatus	Professional	Yes	Shapefile
Google Maps	Professional	No	XML/JSON
Openstreetmap	VGI	No	XML/Shapefile

6.1 Data description

6.1.1 Locatus

Locatus is an organisation that provides professional datasets concerning retail information. Their database is updated weekly and they create reports for customers concerning vacancy and passer-byes. Most importantly, *Locatus* collects POI data in shopping areas focussed on retail. According to their website the *Locatus* database contains every sales point in the Benelux (city or village) with metadata (name, address, formula and sector). This data is collected by their own professionals in the field who visit another part of the Benelux every week. This data normally is costly to collect, but a fraction of it can for this research be used free of charge.

6.1.2 Google Maps

The reason *Google Maps* data is included in this research is because it is in contrast to OSM professionally collected data and in contrast to *Locatus* (to a certain extent) freely available. The difficulty however with collecting *Google Maps* data is that it is only accessible through an Application Programming Interface (API). Most of today's software projects heavily depend on the use of API
libraries and they improve code reuse, reduce development cost and promote programmers' productivity (Qiu et al., 2016). Third-party applications can use APIs to take advantage of or extend the functionality of existing services. An example is an embedded *Google map* on a website, which can be achieved by using the *Static maps API*, *Places API* or *Google Earth API*. In this research the *Google Places API Webservice* and the *Google Maps Geolocation API* will be used. These APIs are able to find detailed information about places across a wide range of categories, including retail POI and detect the coordinates of these places. The data will be accessed with the programming language *Python* in a *JSON* or a *XML* format and the standard usage limit is 2,500 API requests per day which is enough for this research, but could be a constraint for a more demanding project (Google Developers, 2017).

For the use of *Google's* APIs an identification key is required. In this case for the *Google Places API Webservice* and the *Google Maps Geocoding API*. *Google* allows its users to receive this key for free on their personal API console page. The API key will be used in *Python* by importing the *Google Places* library and connecting with *Google* by inserting the API key into a statement (see appendix II for the *Python* script). Several requests can be made with the API:

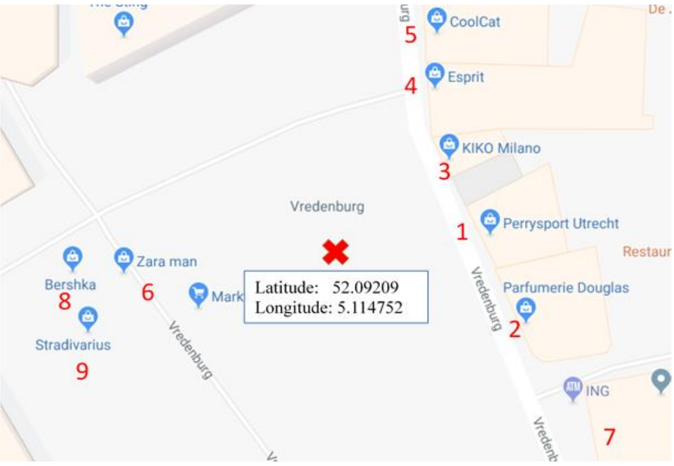
- Place searches
- Place details

- Place add
- Place photos
- Place autocomplete
- Query autocomplete

The data required for assessing the geo-data quality is the location and metadata of POI within the study areas. This can be accessed with *place searches* and *place details*. The API enables users to access data with a nearby search request or a text search request. The first lets the user search for places within a specified area. The second is a web service that returns information about a set of places based on a string (for example "museum in Amsterdam"). Due to the clear demarcation of the study-areas, the nearby search request is most useful. Centre coordinates are used in combination with a 'rank-by distance' function which returns POI increasingly further away from a single location. This allows you to systematically pick several locations and filter irrelevant results in order to create a complete database.
Furthermore the details of the requested places can be accessed which extends the information of the places with for instance the type, ratings or opening hours. In this research only the name, coordinates (in a Google Web Mercator projection) and types of the requested places are useful.

Figure 13 and corresponding table 6 are an example of a places search for 'stores' ranked by distance starting at the red cross in the middle. The table shows that the stores surrounding the coordinates, are the first ones in the results. This proves that by selecting points in the middle of selected streets in the study areas (for large streets multiple points are selected) all relevant POI are selected.

Figure 13 & Table 6: Example of a Google Places rank-by-distance 'store' search

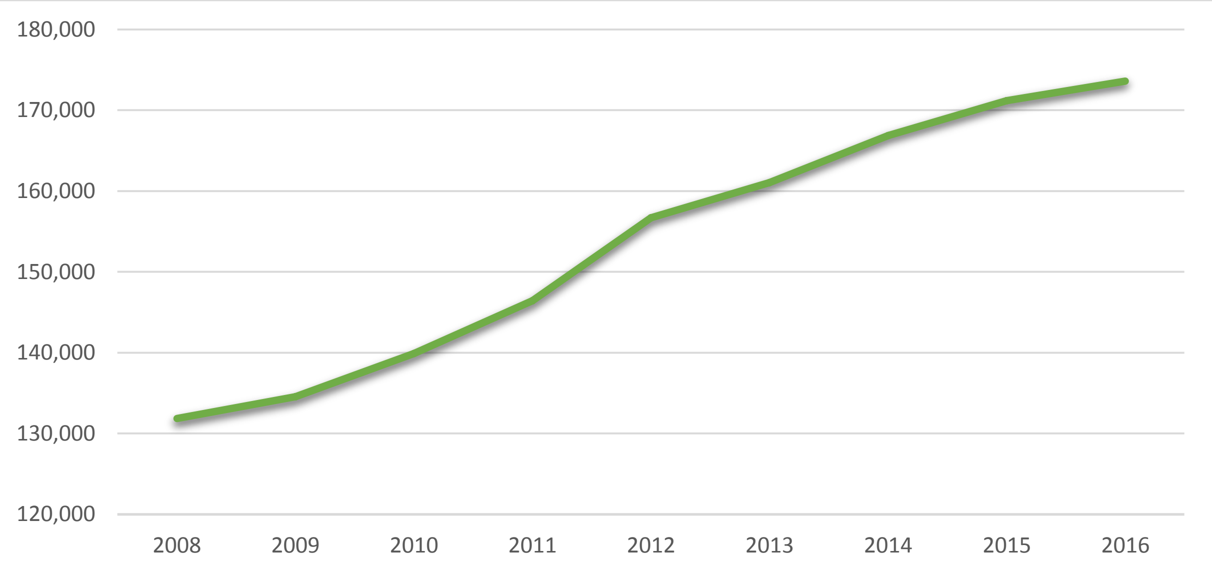


ID	Name	First type	Latitude	Longitude
1	Perry Sport	Shoe store	52.0922	5.1151
2	Douglas	Clothing store	52.0921	5.1152
3	Kiko Milano	Clothing store	52.0923	5.1150
4	Esprit	Clothing store	52.0925	5.1149
5	CoolCat	Clothing store	52.0925	5.1149
6	Zara Men	Clothing store	52.0922	5.1141
7	ICI PARIS XL	Clothing store	52.0918	5.1154
8	Bershka	Clothing store	52.0922	5.1140
9	ZARA	Clothing store	52.0926	5.1145
10	Stradivarius	Clothing store	52.0921	5.1140

6.1.3 Openstreetmap

In 2004 Steve Coast started the *Openstreetmap* project in the United Kingdom and since that over a million volunteers have contributed to the goal of making global geographic data freely available. Especially in Europe where accurate digital geo-data is considered to be expensive this goal was very important. Since 2006 the initiative has become a foundation and received support of among others *Automotive Navigation Data* and *Microsoft*. A team of approximately 40 of the volunteers are the core of OSM and maintain the server and the infrastructure. Furthermore this team works on improvements of the project (Haklay & Weber, 2008; OpenStreetMap Wiki, 2017). The OSM data is stored in the database as three types of objects: nodes, ways and relations. In this research the focus is on the nodes which contain a latitude and a longitude. All object-types are furthermore divided into fixed OSM categories. In The Netherlands the number of OSM users has steeply increased in recent years from just above 130,000 in 2008 to more than 170,000 in 2016.

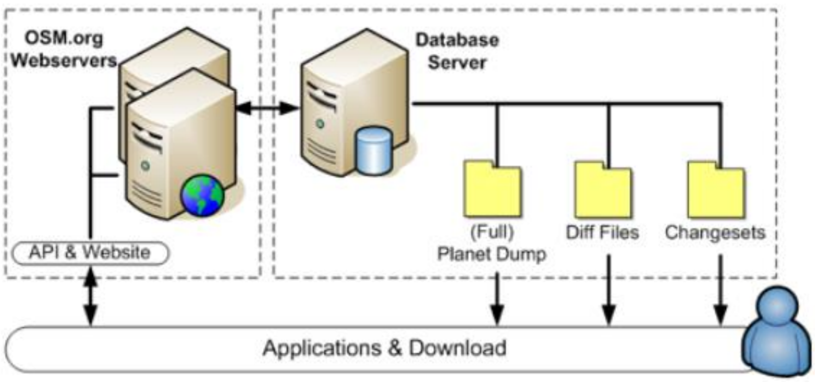
Figure 14: Development of OSM users in The Netherlands (2008 – 2016)



Source: Openstreetmap (2017)

OSM data can be collected with different methodologies. First of all a so-called 'dump file', which is updated every week, can be downloaded from the OSM database server. Second of all, the changes in the database can be collected in minutes, hours and days which are called 'diff files'. Third of all the changes made by OSM contributors can be collected by downloading the 'changeset' file from the server. These files are all enormous, because they cover the entire world (Neis & Zipf, 2012). In addition to the database server downloads, data can also be collected from third parties which offer data in a shapefile format and allow to select a country before downloading which greatly reduces the size of the download. Data for this research is collected at third party *Geofabrik* which is a German organisation that extracts selects and processes open geo-data for free.

Figure 15: Retrieving OSM data



Source: Neis & Zipf (2012)

6.2 Data inspection

When inspecting the three test datasets it stands out that the categories over which the data is distributed differ a lot. *Locatus* and OSM have the most extensive classification with both three levels of categories. At *Locatus* they are: sector, category and activity and at OSM there are not defined. The focus of *Locatus* is purely on retail which explains this high degree of detail. OSM and *Google Maps* provide a broader spectrum of data, but especially *Google Maps* is less specific in its classification. The nature of their classification is more general and there is no clear division of the categories in detail level. Most POI in the *Google Maps* data have more than one data type where the first type seems to be the most specific. However it is possible that a type that is first for one point is second or third for another. This is in contrast with other two datasets where this is not possible. As table 7 points out, Locatus has more than twice as many retail classes in the study area in comparison to Google Maps and OSM.

Table 7: Number of retail classes per dataset in the study areas

Dataset	**Number of retail classes**
Locatus	70
Google Maps	26
Openstreetmap	31

6.3 Data processing

After collecting and inspecting the test datasets, the data must be prepared for analysis. The first step is reclassifying the categories into the 3 digit SBI codes (table 3) where the more detailed SBI categories of appendix V are a guideline when there is uncertainty. The categories that cannot be classified as one of the SBI codes are classified as 'irrelevant' and are filtered out of the dataset. The next processing step is selecting the data within a distance of 50 meters from the selected BAG buildings in the study areas which filters out points which are irrelevant because they are not located within the study area.

7. Results

This chapter discusses the results of the data quality assessment of the three test datasets in the two study areas. First of all the intrinsic quality is discussed where the completeness, attribute accuracy and positional accuracy are assessed. Second of all the extrinsic quality is discussed which is only relevant for VGI data (Openstreetmap). Third of all the pragmatic quality is discussed where the fitness for use of the datasets in the context of the 'healthy urban living project'.

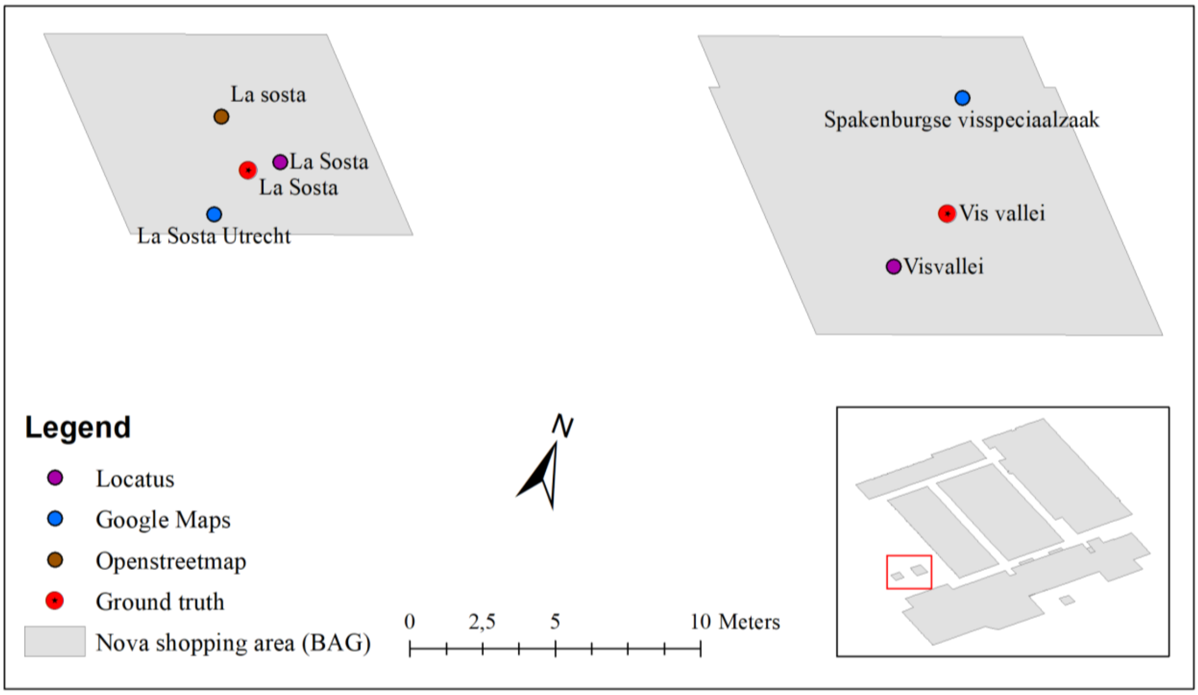
7.1 Intrinsic quality

For the determination of the intrinsic quality, a reference dataset is collected in both study areas. In the Nova shopping area, 43 retail POI are collected and in the selected sample of the city centre of Utrecht, 194 retail POI are collected. This paragraph describes the intrinsic quality of the test dataset based on this ground truth sample as reference.

7.1.1 Completeness

The decision to assess completeness manually is undoubles the most accurate methodology. Most POI have a different spelling in all or some of the test datasets which would have indicated errors in an automated process. Figure 16 is an example where a manual analysis results in only one missing POI in the OSM dataset (Visvallei) and an automatic process would also see 'La Sosta' and 'Spakenburgse visspeciaalzaak' in Google Maps as missing, because of a different spelling.

Figure 16: A fraction of the results in the Nova Shopping area



The results of the manual analysis of the completeness show that in both of the study areas the completeness of the Locatus dataset is most accurate. With 93.0 percent completeness in the Nova shopping area (40 of the 43 points in the reference dataset) and 84.0 percent in the city centre of Utrecht (163 of the 194 points in the reference dataset) this dataset is by far the most complete. Locatus performs approximately 30 percent better in Nova and more than 10 percent in the city centre in comparison with Google Maps and Openstreetmap. Google Maps performs slightly better with respectively 62.8 and 73.1 percent in Nova and the city centre against 60.5 and 67.4 percent for Openstreetmap.

Table 8: Completeness of the test datasets in the study areas

	Nova		City centre	
Dataset	**N**	**Percentage**	**N**	**Percentage**
Locatus	40	93.0	163	84.0
Google Maps	27	62.8	141	73.1
Openstreetmap	26	60.5	130	67.4

Where Google Maps and Openstreetmap score better in the city centre, this is not the case for the Locatus dataset which scores 9.0 percent worse. However it can still be concluded that Locatus is by far most complete in both the Nova shopping area and the city centre of Utrecht.

7.1.2 Attribute accuracy

Before discussing the results of the attribute accuracy it is meaningful to examine the classification of Google Maps. Like Locatus and Openstreetmap, a POI in Google Maps can consists of several types. However in the classification of Google Maps all types are on the same level of detail which results in confusion. For the determination of the attribute accuracy only the first type is taken into account, because this is in most cases the best fit. Nonetheless it should be noted that the different types allocated over the POI are sometimes not very accurate. The 'Esprit Store' for instance is classified as 'cloting store' as first type (which is correct) and as 'physiotherapist' as second type. This makes absolutely no sense. Furthermore there are 19 POI with 'clothing store' or 'shoe store' as first type and 'food' as second type. The performance of Google Maps is fairly good based on the first types, but the errors in the other types have to be kept in mind.

The overall attribute accuracy, based on the reclassification of the test datasets into the SBI classification of the Dutch Central Bureau for Statistics is reduced from the error matrices in appendix VI. It is striking that the attribute accuracy of the Openstreetmap dataset performs best in both study area with 96.3 and 90.8 percent respectively in Nova and the city centre. In the Nova area this is quite a difference in comparison to the Locatus dataset which performs 13.8 percent worse (an accuracy of 82.5 percent). This difference is significantly smaller in the city centre where the attribute accuracy of Locatus is 87.8 percent, indicating a difference of only 3 percent. The accuracy of Google Maps is most stable with a score of 92.3 percent in the Nova shopping area and 87.4 percent in the city centre. However it has to be noted that Google Maps consists of a less detailed classification which reduces the likelihood of reclassification errors.

Table 8: Attribute accuracy of the test datasets

	Nova		City centre	
Dataset	**N**	**Percentage**	**N**	**Percentage**
Locatus	40	82.5	163	87.8
Google Maps	27	92.3	141	87.4
Openstreetmap	26	96.3	130	90.8

Zooming in on the error matrices gives insight in the type of existing errors in the datasets. For the Nova shopping area, most errors of Locatus are due to incorrectly classifying 'shops selling other goods' as 'shops selling other household equipment' (4 times). This error in the Locatus dataset is noted in the centre as well (4 times). In this study area the most striking errors are 2 ‘restaurants’ classified in the Google Maps dataset as ‘shops selling other goods’ where there is no uncertainty regarding the type of these POI (‘The Pancake Bakery Muntkelder’ and ‘STACH food’). Furthermore a lot of errors are the result of confusion between a ‘bar’ and a ‘restaurant’. As stated in chapter 4, all POI with the possibility to consume food inside (with take-out restaurants and food outlets as exceptions) are considered a restaurant (SBI code 561). A clear example is ‘Quignon’ which is classified as a bar in OSM and as a restaurant in the ground truth. The front of this POI says: ‘kitchen & bar’, indicating a place where you can consume food inside. In total this type of error returns 6 times in OSM, 7 times in Google Maps and 8 times in Locatus.

In general, all test dataset have high attribute accuracies with Openstreetmap as most accurate. Even though Google Maps performs very well, in this dataset the most striking errors occurred. Furthermore there is a minor difference in accuracy between the study areas which cannot be explained.

7.1.3 Positional accuracy

As stated in chapter 4.2.2, the positional accuracy of POI is somewhat difficult to determine, because this depends on the placement of a point within the building. In the reference dataset all points are placed where the entrance of the building is, but the test datasets could have used a deviation placement method. For this reason it is acknowledged that a Euclidean distance between 0 and 20 meters between a test point and a corresponding reference point does not necessarily indicate an error. Nonetheless this is the only method to determine positional accuracy of points and high differences between test and reference points clearly indicate errors and influence the results.

It stands out that in both study areas the mean distances of all test datasets are beneath 20 meters indicating a very good positional accuracy (see figure 17 and figure 18). There are however in all dataset and in both study areas points with a distance larger than 20 meters included. In the Nova shopping area the positional accuracy is higher than in the city centre. The point of attention here is in the Google Maps dataset where some outliers exist. Of these outliers ‘Bristol’ is the maximum with a distance of 84.0 meters from the reference point. The high standard deviation of this dataset indicates that this is not the only outlier. In total 7 points are further than 20 meters away from the matching reference point. In the city centre Locatus is the dataset with most outliers. ‘Humphrey’s’ is the maximum (67.4 meters) of in total 15 points with a distance higher than 20 meters.

Figure 17: Positional accuracy Nova shopping area

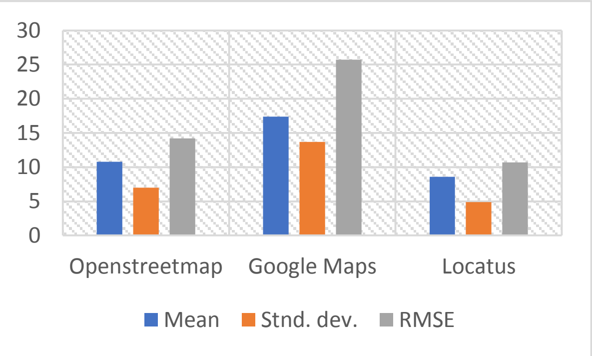
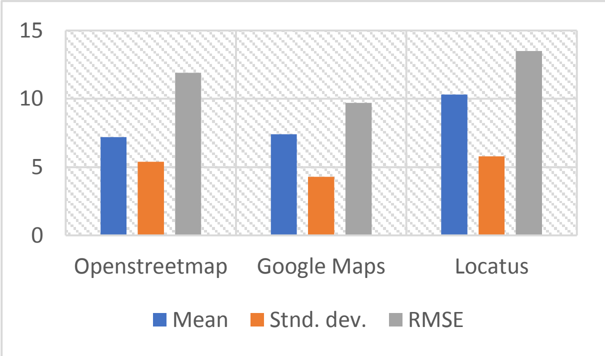


Figure 18: Positional accuracy Utrecht city centre



7.2 Pragmatic quality

The first step in assessing the pragmatic quality of the test datasets is to identify the usability of the classification for the case-study: the relation between the residential fast-food environment and the individual risk of cardiovascular diseases in The Netherlands. So there has to be a clear distinction of fast food outlets in the test datasets. Locatus the following classes related to fast food outlets: fastfood; delivery/take away; and grillroom/shoarma. The only difficulty here is that a grillroom or a shoarma outlet can also be classified as fastfood and delivery and takeaway are not necessarily fast food. However since the study only requires data on fast food outlets, the three classes can be reclassified into one 'fast food outlet' class. In the Openstreetmap dataset all fast food outlets are accommodated in the class 'Fast food'. This is very easy to use and does not require any data processing. Google Maps is the only dataset without a fast food class. Only 'meal delivery' and 'meal takeaway' are included, meaning that fast food outlets are accommodated in either one of these classes or in 'restaurant'. This makes Google Maps data unsuitable to use for this study. Nonetheless this dataset is still taken into account when checking the fast food completeness and positional accuracy.

In total 12 POI in the study areas are identified as fastfood outlets (6 in the Nova shopping area and 6 in the city centre). Even though the number of outlets is low, this still approves for a detailed analysis (see appendix XIII). In terms of completeness the three dataset perform very well. Locatus and Google Maps only miss 1 outlet and OSM misses 2. Both Locatus and Openstreetmap have not included 'Dunkin' Donuts' which opened in October 2017 (4 months ago). This indicates that the temporal quality of both datasets is not very accurate. The other missing outlet in the OSM dataset is in the Nova shopping area, just like the one missing in the Google Maps dataset.

Table 9: Pragmatic quality results

	Locatus	**Openstreetmap**	**Google Maps**
Completeness (%)	92.3	84.6	92.3
Attribute accuracy (%)	100.0	81.8	-
Positional accuracy (m)	6.4	6.5	7.0

The attribute accuracy of Locatus is free of errors. All fastfood outlets are identified as one of the classes related to fastfood. Openstreetmap on the other hand identified two fastfood outlets as 'restaurant' which would not be taken into account in the study. As stated before, the Google Maps dataset does not have a specific fastfood category and therefore it is irrelevant to determine their pragmatic attribute quality. 3 of the fastfood outlets were identified by Google Maps as 'meal takeaway', 1 as 'café' and the rest as 'restaurant'. In terms of positional accuracy of the fast food outlets, all datasets are very accurate with Euclidean distances of 6.4, 6.5 and 7.0 meters. While keeping in mind that the placement method of the points within a building can result in approved distances these results are excellent.

In all elements the dataset of Locatus performs best. This dataset is therefore the best fit for the research of the relation between the residential fast-food environment and the individual risk of cardiovascular diseases. The performance of Openstreetmap is acceptable, but would not result in the most accurate outcomes of the study. Lastly, the pragmatic quality of the Google Maps dataset for this study is weak. There is no classification of fast food in general which makes it impossible to perform this study and on top of that their positional accuracy is more than 20 meters worse than Locatus and Openstreetmap.

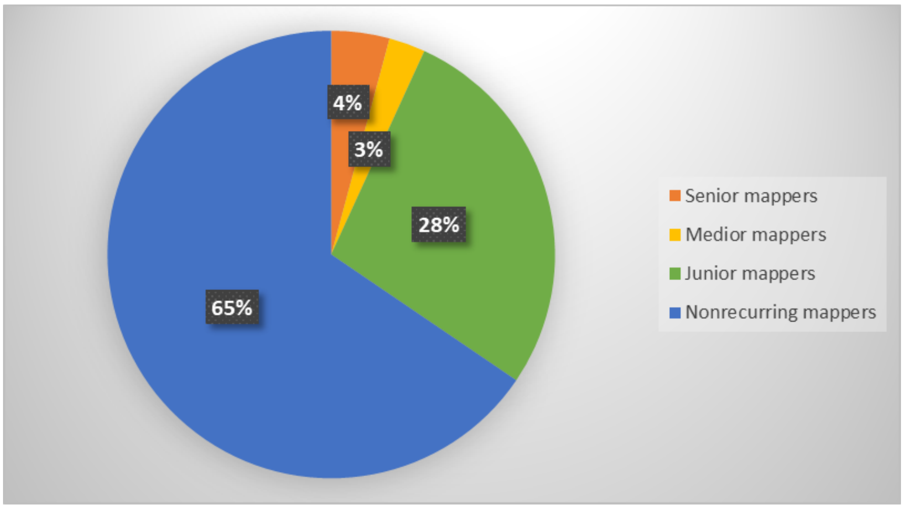
7.3 Extrinsic quality

The metric 'experience' is selected to determine the extrinsic quality of the VGI dataset: Openstreetmap. For this analysis all changesets from 2013 to 2017 in Utrecht are taken into account.

In total there have been approximately 450,000 edits in this area by 571 volunteers. If all volunteers contributed equally this indicates 795 edits per user.

The distribution of the edits over the users is however not equally. Classifying the contributors in senior, medior, junior and nonrecurring mappers result in distribution of figure 19. The nonrecurring mappers are by far the largest group of volunteers with 65 percent. The junior mappers are second with 28 percent and the medior and senior mapper categories consist of only 3 and 4 percent of the total number of volunteers.

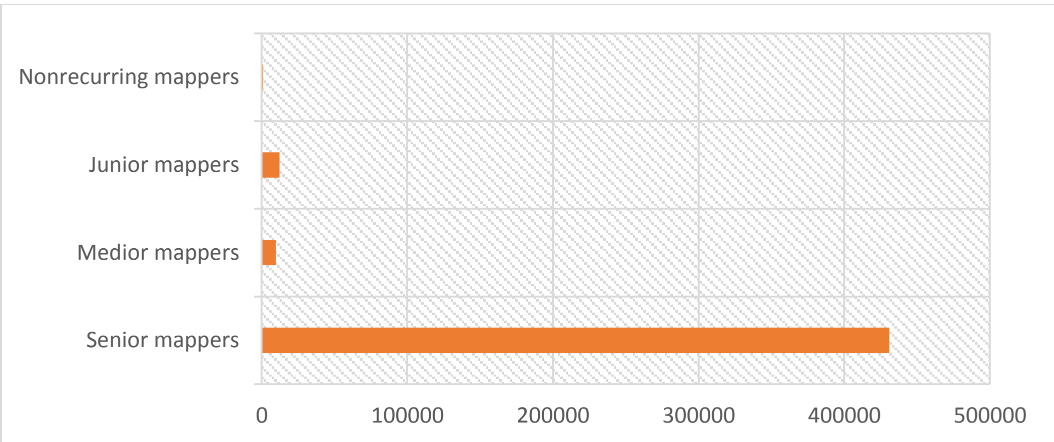
Figure 19: Experience OSM users in Utrecht (2013-2017)



Even though more than half of the contributors in Utrecht can be classified as a nonrecurring mapper this group is responsible for not even 1 percent of all contributions (figure 20). The 4 percent of volunteers classified as senior mappers on the other hand are responsible for 95 percent of all edits. Within this group there are 5 users with more than 20,000 edits and the biggest contributor is solely responsible for more than 40 percent of all contributions with 184,355 edits.

So the fact that the group of medior and senior mappers is small does not influence the experience of the OSM data, because 95 percent of all edits are still done by the most experience senior mappers. Even if the 93 percent nonrecurring and junior mappers with little experience make mistakes in the dataset, the experienced and very active senior mappers can resolve these errors.

Figure 20: Number of edits per experience class of OSM users in Utrecht (2013-2017)



8. Discussion

In this chapter the results are discussed by comparing them to other scientific studies and substantiating debatable choices that have been made. In the scientific literature the research focus is mostly on VGI quality and in particular the road quality of OSM. Therefore not all results can be compared to other researches. Some studies also included Google Maps as a test dataset and focussed more on POI quality. For these studies it is possible to draw a comparison and place the results of this thesis in scientific perspective. The results of the Locatus dataset can unfortunately not be compared with other studies, because this is the first research of the quality of this dataset.

The assumption when selecting the study areas of this research: (the city centre of Utrecht and a suburban shopping area) was that there is a difference in geographic data quality between the centre and the suburbs, because data of the centre generally is more important. Nonetheless there have been no results to support this statement. Perhaps a second study area in a less densely populated area of The Netherlands (outside of the Randstad) would have resulted in more interesting outcomes. Furthermore the fact that this research is only conducted in Utrecht result in that no general statements about the data can be made for other places in The Netherlands and definitely not outside of the country. However this methodology includes a manual inspection which is very time-consuming when scaling up the research area. Automating the process and using a professional source as Locatus as a reference could be a solution, but as the results of the completeness analysis shows an automated process results in a lot of errors and furthermore the entire results chapter proves that a professional source as Locatus is not free of errors.

8.1 Intrinsic data quality

In this research three metrics have been used to determine the intrinsic data quality of the test datasets where in the literature more metrics are mentioned. However the selected metrics are characterised as the fundamental dimensions of intrinsic data quality (Bucher et al, 2016; Goodchild & Li, 2012) and have a quantitative nature, meaning the results are easy to compare.

The completeness of the test datasets is determined by manually linking points to the reference dataset and calculating the percentage of missing points indicating an error of omission. The error of commission is disregarded, because it is difficult to determine whether a point is in the right location, but not present anymore or in the wrong location, but still present (which indicates a positional error). For this reason the focus has been solely on the error of omission. In the literature the metric completeness mostly focusses on road completeness (Girres & Touya, 2010; Haylay, 2010) instead of the completeness of POI. The reclassification of the attributes of the test datasets into the 3-digit SBI class is a difficult, but the only way to determine attribute accuracy. It is possible that the existing classification matches poorly with the SBI classification unfairly resulting in lower attribute accuracy. Yet this methodology is the singular way of assessing the attribute accuracy in a quantitative manner which was the goal of this research. The reclassification process has been very prudent in order to end up with the most reliable results. The applied methodology is mentioned in the literature (Dorn et al., 2015; Veregin, 1999), but is only sparsely applied in practical research.

The results of the positional accuracy analysis are very good. In the scientific literature of POI quality research is mostly focussed on positional accuracy of OSM, so the results of this metric can be compared to those researches. The results of OSM in this research (ME's of 10.8 in Nova and 7.2 in the centre) correspond to the results of Haklay (2009) who noticed a positional accuracy of 5.8 meters in London and Girres & Touya (2010) who found an accuracy of 6.7 meters in France. Cipeluch et al.

(2010) researched both OSM and Google Maps in five case study locations and found no no consistent accuracy for both datasets. A very recent study of Hochmair et al. (2018) in the centre of Salzburg, however, found no positional errors in a POI sample of both datasets. This result corresponds with the results of this research in the centre of Utrecht where Google Maps and OSM have similar high accuracies (ME's around 7 meters). The applied methodology for determining positional accuracy of the POI datasets is very influential on the results, because as pointed out in chapter 5.2.3, deviating placements within a building result in errors with this methodology where they are in fact correct. Still these errors usually are not higher than 20 meters, but for example working with polygons instead of points would solve this problem.

8.2 Extrinsic data quality

The trustworthiness of the data producer, classified here as extrinsic data quality, is in this research determined by the experience of the OSM contributors with the number of edits a user have made as indicator. However it is questionable whether or not this is an indicator of experience. A user with lots of edits could still have a lot of difficulties contributing and users with only a few edits on the other hand could be a quick learner. Still this is the only method of analysis concerning the extrinsic quality of Openstreetmap. As stated in the theoretic background, other platforms where volunteers interact like *Ebay* and *Stack Overflow* keep track of the reputation users within the community. This is a more straightforward metric of extrinsic quality, but unfortunately not applicable on Openstreetmap. The contributors of OSM are now classified in four categories with senior mappers as most experienced. As Goodchild & Li (2012) argue, experienced mappers can be the gatekeepers in a certain area where they can control the data quality. As the senior mappers in Utrecht are responsible for 95 percent of all edits this indicates a high quality control and thus a high trustworthiness.

8.3 Pragmatic data quality

As Bucher et al. (2016) point out; the pragmatic quality of a dataset is dependent on the intended application it will be used in. In this research the application to test the pragmatic quality has been the relations between the residential fast-food environment and the individual risk of cardiovascular diseases. So the focus has been on the fast food outlets in the datasets. This resulted in a very high quality for Locatus, a moderate quality for OSM and a poor quality for Google Maps. However it is possible that another use-case would have resulted in a different division.

8.4 Research value

In the motivation paragraph of the first chapter was stated there is no real quality framework to determine which dataset to use within for example the Healthy Urban Living project. In this research a framework is created (figure 7) and the quality of three prominent datasets in The Netherlands are assessed in a structured manner. In contrast to most scientific studies the focus has not been solely on a VGI dataset (OSM), but also on two professional datasets (Google Maps and Locatus) to see what the quality of VGI is in comparison to the professional datasets.

Where the results of this study show that the Locatus dataset is most accurate, this does not mean that this dataset should at any time be used in practice instead of OSM or Google Maps. In reality the factor quality is not the only determinant of data choice. Locatus is for example in comparison to the other datasets very costly to use. Furthermore Locatus is not useful in a use-case without a retail theme. In that environment Google Maps or Openstreetmap could be more valuable. So the practical quality of the datasets is very dependent on the data purpose.

9. Conclusion

In this thesis the retail POI data quality of Locatus, Google Maps and Openstreetmap is researched on the basis of the following research question:

How can the quality of professional geographic information and Volunteered Geographic Information of Points-Of-Interest be assessed and to what extent does the quality of this information differ within the city of Utrecht?

Three main categories of geographic data quality are identified (Criscuolo et al., 2016): intrinsic quality, extrinsic quality and pragmatic quality; and two study areas within the city of Utrecht are selected: the city centre and shopping area Nova in the suburbs.

The assessment of the intrinsic quality has been the largest part of this research. Three most impacting quantitative metrics are selected as indicator of this quality: completeness, attribute accuracy and positional accuracy. The importance to research two of these metrics is addressed by Haklay (2010, p. 687) who state that "*positional accuracy is an ancient issue in mapping science and therefore must be tested and completeness is an outstanding issue in VGI, because there is no top-down coordination to ensure systematic coverage*". In terms of completeness Locatus is in both study areas significantly better than the other datasets with a completeness of 93.0 percent in Nova and 84.0 percent in the city centre. Google Maps and Openstreetmap both miss almost 40 percent in the Nova shopping area and around 30 percent in the city centre which is quite a lot. The attribute and positional accuracy can only be determined for the points assessed as 'complete'. The attribute accuracy of these complete points is assessed based on the Standard Industrial Classification of the Dutch CBS (three digit detail). On this metric the results are very good with accuracies varying from 82.5 to 96.3 percent in the Nova shopping area and from 87.4 to 90.8 in the city centre. In both study areas Openstreetmap performs best. Overall, the intrinsic quality of Locatus is highest due to good results on all three metrics. Especially in terms of completeness the intrinsic quality of Google Maps and Openstreetmap leave a lot to be desired. Furthermore there seems to be no significant difference in intrinsic quality between the centre and suburban shopping area Nova.

Table 10: Results of the intrinsic quality assessment

	Nova			City centre		
Dataset	**Completeness**	**Attribute accuracy**	**Positional accuracy**	**Completeness**	**Attribute accuracy**	**Positional accuracy**
Locatus	93.0%	82.5%	8.6 m	84.0%	87.8%	10.3 m
Google Maps	62.8%	92.3%	17.4 m	73.1%	87.4%	7.2 m
Openstreetmap	60.5%	96.3%	10.8 m	67.4%	90.8%	7.2 m

The pragmatic quality of the three datasets is determined by testing their usability in a study of the Healthy Urban Living project at the University of Utrecht. This study researches the relation between the residential fast-food environment and the individual risk of cardiovascular diseases in The Netherlands and requires data of fast food outlets. The completeness, attribute accuracy and positional accuracy of fast food outlets is analysed and in all elements Locatus performs best. The performance of Openstreetmap is acceptable, but would not result in the most accurate outcomes of the study. The pragmatic quality of the Google Maps dataset is weak. There is no classification of fast food in general which makes it impossible to perform this study and on top of that the positional accuracy is more than 20 meters worse than Locatus and Openstreetmap due to a large outlier.

Only the extrinsic quality of Openstreetmap is analysed, because this is a VGI dataset created by a heterogeneous group of contributors. The metric determining this quality is the experience of the volunteers in the city of Utrecht. A high degree of experienced contributors is a positive influence on the trustworthiness of the data, because these users are expected to be very accurate and correct mistakes of less experienced contributors. In the city of Utrecht 4 percent of all users can be classified as senior mapper (> 1,000 edits) and 65 percent as nonrecurring mapper (0-10 edits). The fact that the group of senior mappers is small does not influence the experience of the OSM data, because 95 percent of all edits are still done by the most experience senior mappers. Therefore Openstreetmap has a high level of experienced contributions in Utrecht, meaning a high extrinsic quality.

Overall, the Locatus dataset is relatively of the highest quality in the study areas. The intrinsic as well as the pragmatic quality of this dataset is better than that of Openstreetmap and Google Maps. This last dataset turned out to have a poor pragmatic quality. Furthermore there is no significant difference noted between the two study areas in the city of Utrecht. However as the last chapter of the discussion points out, the recommendation of this research is not to always use Locatus instead of OSM of Google Maps. This is very dependent on the environment the data is used. This thesis can be a decision support when considering one the tested datasets and provides a framework to test other datasets in other study areas.

10. Bibliography

Bucher, B, Falquet, G., Metral, C. and Lemmens, R. (2016). Enhancing the management of quality of VGI: contributions from context and task modelling. In: *European Handbook of Crowdsourced Geographic Information*, pp. 131–142.

Cambridge Dictionary. (2017). Lineage Meaning in the Cambridge English Dictionary. Retrieved November 21, 2017, from https://dictionary.cambridge.org/dictionary/english/lineage

Cipeluch, B., Jacob, R., Winstanley, A., & Mooney, P. (2010). Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resuorces and Enviromental Sciences* (pp. 1-4). United Kingdom: Leicester

Chrisman, N. (1991). The error component in spatial data background: error and spatial data definitions. *Geographical Information Systems*, *1*, pp. 165-174

Criscuolo, L, Carrara, P, Bordogna, G, Pepe, M, Zucca, F, Seppi, R, Oggioni, A and Rampini, A. (2016). Handling quality in crowdsourced geographic information. In: *European Handbook of Crowdsourced Geographic Information*, pp. 57–74.

De Longueville, B., Ostländer, N., & Keskitalo, C. (2010). Addressing vagueness in Volunteered Geographic Information (VGI) – A case study. *International Journal of Spatial Data Infrastructures Research*, *5*, pp. 463-1725.

de Tré, G., van Britsom, D., Matthé, T., & Bronselaer, A. (2013). Automated cleansing of POI databases. *Intelligent Systems Reference Library*, *50*, pp. 55-91

Dorn, H., Törnros, T., & Zipf, A. (2015). Quality Evaluation of VGI Using Authoritative Data—A Comparison with Land Use Data in Southern Germany. *ISPRS International Journal of Geo-Information*, *4*(3), pp. 1657-1671

Esmaili, R., Naseri, F., & Esmaili, A. (2013). Quality Assessment of Volunteered Geographic Information. *American Journal of Geographic Information System*, *2*(2), pp. 19-26

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal*, *72*(3–4), pp. 137-148

Garrett, J. J. (2005). Ajax: A New Approach to Web Applications. *Adaptive Path*, pp. 1-5

Girres, J.-F., & Touya, G. (2010). Elements of quality assessment of French OpenStreetMap data. *Transactions in GIS*, *14*(4), pp. 435-459

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), pp. 211-221

Goodchild, M. F., & Hunter, G. J. (1997). A simple positional accuracy measure for linear features. *International Journal Geographical Information Science*, *11*(3), pp. 299-306

Goodchild, M. F., & Li, L. (2012). Assuring the quality of Volunteered Geographic Information. *Spatial Statistics*, *1*, pp. 110-120

Google Developers (2017). Google Places API Web Service. Retrieved October 12, 2017, from https://developers.google.com/places/web-service/

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*, *37*(4), pp. 682-703

Haklay, M., Singleton, A., & Parker, C. (2008). Web Mapping 2.0: The Neogeography of the GeoWeb. *Geography Compass*, *2*(6), pp. 2011-2039

Haklay, M., & Weber, P. (2008). OpenStreetMap: user-generated street maps. *Pervasive Computing*, *7*(4), pp. 12-18

Helbich, M., Amelunxen, C., & Giscience, P. N. (2012). Comparative Spatial Analysis of Positional Accuracy of OpenStreetMap and Proprietary Geodata. In *Proceedings of GI_Forum* (pp. 24-33)

Hessler, J. W. (2006). Warping Waldseemüller: A phenomenological and computational study of the 1507 World Map. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *41*(2), pp. 101-114.

Hochmair, H. H., Juhász, L., & Cvetojevic, S. (2018). Data Quality of Points of Interest in Selected Mapping and Social Media Platforms. In *LBS 2018: 14th International Conference on Location Based Services* (pp. 293-313)

Hui, M. (2016). Volunteers work together to map neighborhoods and tell their stories. *The Washington Post*. Retrieved from https://www.washingtonpost.com/local/trafficandcommuting/mapping-a-city-and-telling-its-story/2016/07/24/905dd6d0-4c40-11e6-aa14e0c1087f7583_story.html?utm_term=.3ff4f9e56dd0

International Organisation Standardization (2013). ISO 19157:2013 - Geographic information - Data Quality. Retrieved October 4, 2017, from https://www.iso.org/standard/32575.html

Jaworowska, A., Blackham, T., Davies, I. G., & Stevenson, L. (2013). Nutritional challenges and health implications of takeaway and fast food. *Nutrition reviews*, 71(5), pp. 310-318

Jonietz, D., & Zipf, A. (2016). Defining Fitness-for-Use for Crowdsourced Points of Interest (POI). *ISPRS International Journal of Geo-Information*, *5*(9), pp. 1-12

Kounadi, O. (2009). Assessing the quality of OpenStreetMap data. Msc Geographical Information Science: University College of London Department of Civil, Environmental and Geomatic Engineering

Maué, P. (2007). Reputation as tool to ensure validity of VGI. In *VGI specialist meeting* (pp. 13-14), California: Santa Barbara.

Mocnik, F., Zipf, A., & Fan. H. (2017). The Inevitability of Calibration in VGI Quality Assessment. In *Proceedings of the 4th Workshop on Volunteered Geographic Information: Integration, Analysis, and Applications (VGI-Analytics)*, The Netherlands: Wageningen

Mooney, P., Kildare Ireland, C., Corcoran, P., & Winstanley, A. C. (2010). Towards Quality Metrics for OpenStreetMap. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems* (pp. 514-517). New York: New York

Mummidi, L. N., & Krumm, J. (2008). Discovering points of interest from users' map annotations. *GeoJournal*, *72*(3–4), pp. 215-227

Neis, P., & Zipf, A. (2012). Analyzing the Contributor Activity of a Volunteered Geographic Information Project - The Case of OpenStreetMap. *ISPRS International Journal of Geo-Information*, *1*(2), pp. 146-165

O' reilly, T. (2007). What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *Communications and Strategies*, *65*(1), pp. 17-31

OpenStreetMap Wiki. (2017). Retrieved October 5, 2017, from http://wiki.openstreetmap.org/wiki/File:Osmdbstats1_log.png

Poelman, M., M. Strak, O. Schmitz, G. Hoek, D. Karssenberg, M. Helbich, A. Ntarladima, M. Bots, B. Brunekreef, R. Grobbee, M. Dijst & I. Vaartjes (2018). Relations between the residential fast-food environment and the individual risk of cardiovascular diseases in the Netherlands: a nationwide follow-up study. *Manuscript in preparation*.

Qiu, D., Li, B., & Leung, H. (2016). Understanding the API usage in Java. *Information and Software Technology, 73*, pp. 81-100

Raymond, E. (2001). *The Cathedral and the Bazaar*. California: Sebastopol: O'Reilly Media.

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C., & Haklay, M. (Muki). (2017). A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, *31*(1), pp. 139-167

Story, M., & Congalton, R. G. (1986). Remote Sensing Brief Accuracy Assessment: A User's Perspective. *Photogrammetric engineering and remote sensing*, *52*(3), pp. 397–399

Van Exel, M., Dias, E., & Fruijtier, S. (2010). The impact of crowdsourcing on spatial data quality indicators. In *Proceedings of the 6th GIScience international conference on geographic information science* (pp. 213-217), Switzerland: Zurich van Oort, P. (2006). Spatial data quality: from description to application. PhD thesis: Wageningen University

Veregin, H. (1999). Defining data quality. *Geographical Information Systems, 1,* pp. 177-189

Wroclawski, S. (2014, January 14). Why the world needs OpenStreetMap. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2014/jan/14/why-the-world-needs-openstreetmap

Zielstra, D., & Zipf, A. (2010). A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany. In *13th AGILE International Conference on Geographic Information Science* (pp. 1-15). Portugal: Guimarães

Appendices

Appendix I: MySQL database creation

This appendix includes the *MySQL* commands that used in order to store, sort and query the required data for the 'experience' metric.

Database creation

```
CREATE DATABASE experience
USE experience;
```

Table creation

```
CREATE TABLE one (
      ID int(10),
      User VarChar(30),
      Edits int(10)
      );
```

Inserting the data

```
INSERT INTO one (ID, User, Edits) VALUES
      (1, 'Zugführer', 612),
      (2, 'Gertjan Idema', 346),
      (3, 'Imergis', 109),
      (4, 'Maarten Deen', 65),
      (5, 'sander79', 65)
      …
      (296, ‘Bert Koning’, 1)
      ;
```

This is repeated for the other three squares

Query

The following SQL-query results in combining the number of edits from a user in multiple years and orders it descending:

```
CREATE TABLE linksonder_final
      SELECT user, SUM(edits) FROM linksonder Group By user Order By
      SUM(edits) DESC;
```

Appendix II: Google Place API Python script

An example of query in *Python* which makes use of the *Google Maps* API library. This query selects restaurants within the city of Utrecht and prints their name and their coordinates.

```
from googleplaces import GooglePlaces, types, lang

YOUR_API_KEY = 'YOUR_API_KEY'

google_places = GooglePlaces(YOUR_API_KEY)

query_result = google_places.nearby_search(
        location='Utrecht',
        radius=10000,
        types=[types.TYPE_RESTAURANT])

for place in query_result.places:
     print place.name

if query_result.has_next_page_token:
    query_result_next_page =
google_places.nearby_search(pagetoken=query_result.next_page_token)
```

Appendix III: Number of addresses per street in the city centre of Utrecht

Name	Count	Probability
Oudegracht	957	7,19
Arthur van Schendelstraat	413	3,10
Lange Nieuwstraat	308	2,32
Van Sijpesteijnkade	292	2,20
Springweg	254	1,91
Nieuwegracht	246	1,85
Voorstraat	245	1,84
Jan van Foreeststraat	234	1,76
Catharijnesingel	224	1,68
Moreelsehoek	209	1,57
Hartingstraat	198	1,49
Lange Lauwerstraat	152	1,14
Keizerstraat	149	1,12
Twijnstraat	142	1,07
Mariaplaats	133	1,00
Vredenburg	130	0,98
Vinkenburgstraat	126	0,95
Breedstraat	123	0,92
Kromme Nieuwegracht	121	0,91
Herenstraat	117	0,88
Jansveld	109	0,82
Groenestraat	107	0,80
St.-Jacobsstraat	107	0,80
Croeselaan	101	0,76
Nobelstraat	100	0,75
Godebaldkwartier	96	0,72
Nieuwekade	96	0,72
Wittevrouwenstraat	95	0,71
Korte Nieuwstraat	95	0,71
Lange Koestraat	94	0,71
Nicolaas Beetsstraat	94	0,71
Henriëtte Roland Holststraat	87	0,65
Haverstraat	87	0,65
Wijde Begijnestraat	87	0,65
Willemstraat	86	0,65
Plompetorengracht	86	0,65
Justus van Effenstraat	85	0,64
Waterstraat	84	0,63
Stationsstraat	83	0,62
Lijnmarkt	83	0,62
Steenweg	81	0,61
Graadt van Roggenweg	80	0,60
Van Asch van Wijckskade	78	0,59
Lange Smeestraat	78	0,59
Oudegracht aan de Werf	76	0,57
Hollandse Toren	76	0,57
Oudkerkhof	75	0,56
Achter St.-Pieter	73	0,55
Oranjestraat	70	0,53
Zuilenstraat	69	0,52
Stationshal	69	0,52
Visschersplein	69	0,52
Hamburgerstraat	68	0,51
Westerkade	68	0,51
Loeff Berchmakerstraat	65	0,49
Minrebroederstraat	64	0,48
Vaartsestraat	64	0,48
Radboudkwartier	63	0,47
Donkerstraat	62	0,47
Buurkerkhof	25	0,19
Pelmolenweg	25	0,19
Begijnehof	24	0,18
Stationsdwarsstraat	24	0,18
Jan Meijenstraat	24	0,18
Brouwerstraat	23	0,17
Ridderhofstad	23	0,17
Agnietenstraat	23	0,17
Vrouwe Justitiaplein	22	0,17
Catharijnekade	22	0,17
Telingstraat	22	0,17
Catharijne Esplanade	21	0,16
Geertekerkhof	21	0,16
Drift	21	0,16
Boogstraat	21	0,16
Lepelenburg	21	0,16
Kroonstraat	20	0,15
Nicolaasstraat	20	0,15
Fentener van Vlissingenkade	20	0,15
Carry van Bruggenstraat	20	0,15
Korte Rozendaal	20	0,15
Molenstraat	20	0,15
Mineurslaan	19	0,14
Wijde Doelen	19	0,14
Alendorpstraat	19	0,14
Fockstraat	18	0,14
Plompetorenbrug	18	0,14
Wijde Begijnehof	17	0,13
Schutterstraat	17	0,13
Bruntenhof	17	0,13
Kockstraat	16	0,12
Hieronymusplantsoen	16	0,12
Lichtegaard	16	0,12
Radboudtraverse	16	0,12
Willemsplantsoen	16	0,12
Nathanaëlspoort	15	0,11
Lijnpadstraat	15	0,11
Sterrenhof	15	0,11
Muntstraat	15	0,11
Donkeregaard	15	0,11
Andreashof	15	0,11
Clarenburgplein	15	0,11
Hoog Catharijnepassage	15	0,11
Dirck van Zuylenstraat	15	0,11
Visscherssteeg	15	0,11
Catharijneplateau	14	0,11
Pasteurstraat	14	0,11
Teugelhof	14	0,11
Korte Minrebroederstraat	14	0,11
Oranjehof	14	0,11
Jaarbeursplein	14	0,11
Kleine Geertekerkhof	14	0,11
Vredenburgpassage	14	0,11
Lauwersteeg	13	0,10
Catharijnesteeg	13	0,10
Pastoor van Nuenenhof	13	0,10
Rijnkade	13	0,10
Veemarktplein	13	0,10
Servetstraat	13	0,10

Ridderschapstraat	61	0,46		Hoogt	13	0,10
Korte Jufferstraat	60	0,45		Stationspassage	12	0,09
Lange Jansstraat	59	0,44		Achterom	12	0,09
Bleekstraat	59	0,44		Westerstraat	12	0,09
Janskerkhof	59	0,44		Voor Clarenburg	11	0,08
Predikherenstraat	58	0,44		Jacobskerkhof	11	0,08
Strosteeg	57	0,43		Stationstraverse	11	0,08
Domstraat	57	0,43		Catharijnepoort	11	0,08
Wolvenplein	55			Albert Verweystraat	11	0,08
Zadelstraat	54	0,41		Jacobsgasthuissteeg	11	0,08
Stationsplein	54	0,41		Wed	11	0,08
Vrouwjuttenstraat	52	0,39		Abraham Dolesteeg	11	0,08
Brigittenstraat	52	0,39		Van Zijstweg	10	0,08
Korte Lauwerstraat	51	0,38		Korte Elisabethstraat	10	0,08
Vrouwjuttenhof	51	0,38		Hemdsmouwsteeg	10	0,08
Mariastraat	50	0,38		Bijlhouwerstraat	10	0,08
Boterstraat	49	0,37		Singelsteeg	9	0,07
Vismarkt	49	0,37		Walsteeg	9	0,07
Rozenstraat	49	0,37		Voetiusstraat	9	0,07
Jeruzalemstraat	48	0,36		Nicolaasdwarsstraat	9	0,07
Kloksteeg	48	0,36		Keistraat	9	0,07
Keukenstraat	48	0,36		Lauwerhof	9	0,07
Lange Jufferstraat	48	0,36		Kintgenshaven	9	0,07
Brandstraat	47	0,35		Kleine Slachtstraat	9	0,07
Boven Clarenburg	47	0,35		Karmelietenhof	9	0,07
Nobeldwarsstraat	47	0,35		Hamsteeg	9	0,07
Predikherenkerkhof	46	0,35		Zoutmarkt	8	0,06
Domplein	46	0,35		Hofpoort	8	0,06
Lange Rozendaal	46	0,35		Hekelsteeg	8	0,06
Annastraat	46	0,35		Jacobskerksteeg	8	0,06
Achter Clarenburg	45	0,34		Pelmolenplantsoen	8	0,06
Geertestraat	44	0,33		Wijde Watersteeg	7	0,05
Vredenburgplein	44	0,33		Bruntensteeg	7	0,05
Wolvenstraat	44	0,33		Slachtstraat	7	0,05
Lange Elisabethstraat	44	0,33		Spoorstraat	7	0,05
Nieuwekamp	44	0,33		Nieuwegracht aan de Werf	7	0,05
Drieharingstraat	43	0,32		Wittevrouwenkade	7	0,05
Schoutenstraat	43	0,32		Catharijnepassage	7	0,05
Boothstraat	43	0,32		Eligenhof	7	0,05
Jeremiestraat	42	0,32		Dichtersbaan	7	0,05
Pieterskerkhof	42	0,32		Vaartsehof	6	0,05
Bergstraat	42	0,32		Boven Catharijnepoort	6	0,05
Pastoor van Nuenenstraat	42	0,32		Korte Koestraat	6	0,05
Jacobijnenstraat	42	0,32		3e Buurkerksteeg	6	0,05
Korte Smeestraat	42	0,32		Leidseveer	5	0,04
Smakkelaarsveld	41	0,31		Nicolaaskerkhof	5	0,04
Pauwstraat	41	0,31		Veemarktstraat	5	0,04
Zakkendragerssteeg	40	0,30		Waterpoort	5	0,04
Zwaansteeg	40	0,30		Stadsplateau	5	0,04
Potterstraat	40	0,30		Moutstraat	4	0,03
Choorstraat	39	0,29		Pausdam	4	0,03
Jansdam	38	0,29		Massegast	4	0,03
Pieterstraat	38	0,29		Jan Meijenhofje	4	0,03
Magdalenastraat	37	0,28		1e Buurkerksteeg	4	0,03
Sterrenbos	37	0,28		Herman Gorterstraat	4	0,03
Boven Vredenburgpassage	37	0,28		Clarenburg	4	0,03
Lange Viestraat	37	0,28		Laan van Puntenburg	3	0,02
Trans	36	0,27		3e Achterstraat	3	0,02
Begijnekade	36	0,27		Tolsteegbrug	3	0,02
Noorderstraat	36	0,27		Vredenburgkade	3	0,02
Doelenstraat	36	0,27		Stadhuisbrug	3	0,02
Mariahoek	36	0,27		Korte Hamstraat	3	0,02
Neude	36	0,27		2e Buurkerksteeg	3	0,02

Reguliersteeg	35	0,26		Zonnenburg	3	0,02
Ambachtstraat	35	0,26		Peterseliesteeg	3	0,02
A.B.C.-straat	34	0,26		Laddersteeg	3	0,02
Eligenstraat	34	0,26		2e Achterstraat	3	0,02
Korte Jansstraat	34	0,26		Tolsteegbarrière	3	0,02
Moreelsepark	34	0,26		Vlaamse Toren	3	0,02
Gildenkwartier	34	0,26		Smalle Begijnestraat	3	0,02
Varkenmarkt	33	0,25		Drakenburgstraat	2	0,02
Schalkwijkstraat	33	0,25		Westplein	2	0,02
Servaasbolwerk	33	0,25		Pauwhof	2	0,02
Tuinstraat	33	0,25		Kroonsteeg	2	0,02
Twijnstraat aan de Werf	31	0,23		Diaconessenstraat	2	0,02
Dorstige Hartsteeg	30	0,23		Manenburg	2	0,02
Hardebollenstraat	30	0,23		Knooppassage	2	0,02
Dorstige Harthof	29	0,22		Daalseplein	2	0,02
Daalsesingel	29	0,22		Daalsetunnel	1	0,01
Schroeder van der Kolkstraat	28	0,21		Nauwe Blindesteeg	1	0,01
Geertebolwerk	28	0,21		Reviuskade	1	0,01
Kalverstraat	28	0,21		Sterrenburg	1	0,01
Oudekamp	28	0,21		Nauwe Watersteeg	1	0,01
Zilverstraat	28	0,21		Begijnesteeg	1	0,01
Achter de Dom	28	0,21		Suikerstraat	1	0,01
Zakkendragershof	27	0,20		Zilverberghof	1	0,01
Bakkerstraat	27	0,20		Nieuwe Daalstraat	1	0,01
1e Achterstraat	27	0,20		Abraham Dolehof	1	0,01
Andreasstraat	27	0,20		Van Asch van Wijckstraat	1	0,01
Lucasbolwerk	26	0,20		Hanengeschrei	1	0,01
Ganzenmarkt	26	0,20		Rondpoort	1	0,01
Hof van St.-Jan	26	0,20				

Appendix IV: Random number generator

This module selects a random sample (size=5) from a list of input data (vec) based on their probability (P) to be selected.

```
import numpy as np

vec =[1, 2, 3, n]
P = [0.071944, 0.031048, 0.023154, n]

print np.random.choice(vec,size=5,replace=False, p=P)
```

Appendix V: extended description of CBS classification

47 Retail trade (not in motor vehicles)

- **471 Retail sale in non-specialised stores**
    - 4711 Supermarkets, department stores and similar non-specialised stores
    - 4719 Department stores and similar non-specialised stores with non-food
        - 47191 Department stores
        - 47192 Non-specialised stores with non-food (no department stores)

- **472 Specialised shops selling food and beverages**
    - 4721 Shops selling potatoes, fruit and vegetables
    - 4722 Shops selling meat and meat products, game and poultry
        - 47221 Shops selling meat and meat products
        - 47222 Shops selling game and poultry
    - 4723 Shops selling fish
    - 4724 Shops selling bread, pastry, chocolate and sugar confectionery
        - 47241 Shops selling bread and pastry
        - 47242 Shops selling chocolate and sugar confectionary
    - 4725 Shops selling beverages
    - 4726 Shops selling tobacco products
    - 4729 Other specialised shops selling food
        - 47291 Shops selling cheese
        - 47292 Shops selling natural and health food
        - 47293 Shops selling foreign food
            - 47299 Specialised shops selling other food n.e.c.*

- **474 Shops selling consumer electronics**
    - 4741 Shops selling computers, peripheral equipment and software
    - 4742 Shops selling telecommunication equipment
    - 4743 Shops selling audio and video equipment, eventually combined with white goods
        - 47431 Shops selling audio and video equipment
    - 47432 Shops selling a combination of brown and white goods

- **475 Shops selling other household equipment**
    - 4751 Shops selling clothing fabrics, household textiles and haberdashery
        - 47511 Shops selling clothing fabrics
        - 47512 Shops selling household textiles
        - 47513 Shops selling knitting wool, fancywork and haberdashery
    - 4752 Shops selling do-it-yourself articles
        - 47521 Shops selling hardware
        - 47522 Shops selling paints and wallpaper
        - 47523 Shops selling building and garden materials of wood
        - 47524 Shops selling tiles
        - 47525 Shops selling kitchens
        - 47526 Shops selling parquet, laminate and cork floors
        - 47527 Specialised shops selling other do-it-yourself articles
        - 47528 Builder's merchants and other shops selling various building materials
    - 4753 Shops selling floor coverings and curtains
    - 4754 Shops selling electrical household appliances and parts of it
        - 47541 Shops selling white goods
        - 47542 Shops selling sewing and knitting machines
        - 47543 Shops selling other electrical household appliances
        - 47544 Shops selling parts of electrical household appliances
    - 4759 Shops selling furniture, articles for lighting and other household articles n.e.c.*
        - 47591 Shops selling furniture
        - 47592 Shops selling articles for lighting

        - 47593 Shops selling various home furnishings
        - 47594 Shops selling musical instruments
        - 47595 Shops selling glassware, china and pottery
        - 47596 Specialised shops selling other household goods n.e.c.*
        - 47597 Non specialised shops selling household articles

  - **476 Shops selling reading, sports, camping and recreation goods**
    - 4761 Shops selling books
    - 4762 Shops selling newspapers, magazines and stationery
    - 4763 Shops selling music and video recordings
    - 4764 Shops selling bicycles and mopeds, sports and camping goods and boats
      - 47641 Shops selling bicycles and mopeds
      - 47642 Shops selling water sports goods
      - 47643 Shops selling sports goods (not for water sports)
      - 47644 Shops selling camping goods (no caravans)
      - 4765 Shops selling toys

  - **477 Shops selling other goods**
    - 4771 Shops selling clothes and clothing accessories: textile supermarkets
      - 47711 Shops selling menswear
      - 47712 Shops selling ladies' wear
      - 47713 Shops selling outerwear and clothing accessories (non-specialised)
      - 47714 Shops selling baby- and children's clothes
      - 47715 Shops selling various baby articles
      - 47716 Shops selling underwear, foundations etc.
      - 47717 Shops selling clothing accessories
      - 47718 Textile supermarkets
    - 4772 Shops selling footwear and leather goods
      - 47721 Shops selling footwear
      - 47722 Shops selling leather goods
      - 4773 Dispensing chemists
      - 4774 Drugstores and shops selling medical and orthopaedic goods
        - 47741 Drugstores
        - 47742 Shops selling medical and orthopaedic goods
      - 4775 Shops selling perfumery and cosmetic articles
      - 4776 Shops selling flowers, plants, seeds, garden material, pets and articles for pets
        - 47761 Shops selling flowers, plants, seeds and garden material
        - 47762 Garden centres
        - 47763 Shops selling pets and articles for pets and angling
      - 4777 Shops selling jewellery and watches
      - 4778 Shops selling other goods n.e.c.*
        - 47781 Shops selling photographic equipment
        - 47782 Shops selling optical articles
        - 47783 Shops selling paintings, frames, pictures, art, religious articles
        - 47789 Specialised shops selling other goods n.e.c.*
        - 4779 Shops selling antiques and second-hand goods
        - 47791 Shops selling antiques
        - 47792 Shops selling second-hand clothing
        - 47793 Shops selling second-hand goods (no clothing)

55 Accommodation

- **551 Hotels and similar accommodation**
  - 5510 Hotels and similar accommodation
  - 55101 Hotels with restaurants
  - 55102 Hotels without restaurants
- **552 Renting of holiday bungalows and apartments; youth hostels and tourist camps**

    - 5520 Renting of holiday bungalows and apartments; youth hostels and tourist camps
        - 55201 Renting of holiday bungalows and apartments
        - 55202 Youth hostels, tourist camps, group accommodation
- **559 Other provision of lodgings n.e.c.***
    - 5590 Other provision of lodgings n.e.c.*

56 Food and beverage service activities

- **561 Restaurants**
    - 5610 Restaurants
        - 56101 Restaurants
        - 56102 Fast-food restaurants, cafeterias, ice cream parlours, take-out eating places etc.
- **563 Bars**
    - 5630 Bars

91 Lending of cultural goods, public archives, museums, botanical and zoological gardens and nature reserves activities

- 910 Lending of cultural goods, public archives, museums, botanical and zoological gardens and nature reserves activities
    - 9101 Lending of cultural goods and public archives
        - 91011 Public libraries
        - 91012 Lending of works of art
        - 91019 Lending of other cultural goods and public archives
    - 9102 Museums and art galleries
        - 91021 Museums
        - 91022 Art galleries
    - 9103 Preservation of historical buildings
    - 9104 Botanical and zoological gardens and nature reserves activities
        - 91041 Botanical and zoological gardens, children's zoos
        - 91042 nature reserves activities

93 Sports and recreation

- 931 Sports activities
    - 9311 Sports facilities
        - 93111 Swimming pools
        - 93112 Sports halls and gymnasiums
        - 93113 Playing fields
        - 93119 Other sports facilities
    - 9312 Outdoor sports 93121 Football
        - 93122 Outdoor team sports (no football)
        - 93123 Athletics 93124 Tennis
        - 93125 Horse riding and manèges
        - 93126 Cycling
        - 93127 Motor sports
        - 93128 Winter sports
        - 93129 Other outdoor sports
    - 9313 Fitness facilities
    - 9314 Indoor sports
        - 93141 Individual indoor sports
        - 93142 Indoor team sports
        - 93143 Power lifting and combat sports
        - 93144 Bowling, skittles, billiards etc.
        - 93145 Playing chess, draughts, bridge etc
        - 93146 Sports schools (combat sports)
        - 93149 Other indoor sports, omnisports
    - 9315 Water sports
        - 93151 Swimming and diving
        - 93152 Rowing, sailing and surfing
    - 9319 Other sports activities

        - 93191 Own account sportsmen
        - 93192 Sport fishing
        - 93193 Organising boat trips for fishing
        - 93194 Supporters clubs (sports)
        - 93195 Organisation of sports events
        - 93196 Umbrella organizations, cooperative and advisory bodies in the field of sports
        - 93199 Other sports activities n.e.c.*
- **932 Other recreation**
    - 9321 Amusement parks and theme parks
        - 93211 Amusement and theme parks
        - 93212 Carnival attractions
    - 9329 Other recreation n.e.c.*
        - 93291 Marinas
        - 93299 Other recreation n.e.c.* (no marinas)

Appendix VI: Classification error matrices

Nova shopping area

Openstreetmap

Reference

-	471	472	474	475	476	477	561	562	563	910	931	932	Total
471	3	0	0	0	0	0	0	0	0	0	0	0	**3**
472	0	3	0	0	0	0	0	0	0	0	0	0	**3**
475	0	0	0	1	0	0	0	0	0	0	0	0	**1**
476	0	0	0	0	2	0	0	0	0	0	0	0	**2**
477	0	0	0	0	0	12	0	0	0	0	0	0	**12**
561	0	0	0	0	0	0	5	0	1	0	0	0	**6**
562	0	0	0	0	0	0	0	0	0	0	0	0	**0**
563	0	0	0	0	0	0	0	0	0	0	0	0	**0**
910	0	0	0	0	0	0	0	0	0	0	0	0	**0**
932	0	0	0	0	0	0	0	0	0	0	0	0	**0**
Total	**3**	**3**	**0**	**1**	**2**	**12**	**5**	**0**	**1**	**0**	**0**	**0**	**27**

Google Maps

Reference

-	471	472	474	475	476	477	561	562	563	910	931	932	Total
471	1	0	0	0	0	0	0	0	0	0	0	0	**1**
472	0	2	0	0	0	0	0	0	0	0	0	0	**2**
475	0	0	0	1	0	0	0	0	0	0	0	0	**1**
476	0	0	0	0	0	0	0	0	0	0	0	0	**0**
477	2	0	0	1	0	11	0	0	0	0	0	0	**14**
561	0	0	0	0	0	0	7	0	0	0	0	0	**7**
562	0	0	0	0	0	0	1	0	0	0	0	0	**1**
563	0	0	0	0	0	0	0	0	0	0	0	0	**0**
910	0	0	0	0	0	0	0	0	0	0	0	0	**0**
932	0	0	0	0	0	0	0	0	0	0	0	0	**0**
Total	**3**	**2**	**0**	**2**	**0**	**11**	**8**	**0**	**0**	**0**	**0**	**0**	**26**

Locatus

Reference

-	471	472	474	475	476	477	561	562	563	910	931	932	Total
471	3	0	0	0	0	0	0	0	0	0	0	0	**3**
472	0	4	0	0	0	0	0	0	0	0	0	0	**4**
475	0	0	0	1	0	0	0	0	0	0	0	0	**1**
476	0	1	0	0	1	0	0	0	0	0	0	0	**2**
477	0	1	0	4	0	16	0	0	0	0	0	0	**21**
561	0	1	0	0	0	0	8	0	0	0	0	0	**9**
562	0	0	0	0	0	0	0	0	0	0	0	0	**0**
563	0	0	0	0	0	0	0	0	0	0	0	0	**0**
910	0	0	0	0	0	0	0	0	0	0	0	0	**0**
932	0	0	0	0	0	0	0	0	0	0	0	0	**0**
Total	**3**	**7**	**0**	**5**	**1**	**16**	**8**	**0**	**0**	**0**	**0**	**0**	**40**

City centre

Openstreetmap

-	471	472	474	475	476	477	551	561	563	910	931	932	Total
471	1	0	0	0	0	0	0	0	0	0	0	0	**1**
472	0	0	0	0	0	0	0	0	0	0	0	0	**0**
474	0	0	3	0	0	0	0	0	0	0	0	0	**3**
475	0	0	0	3	0	1	0	0	0	0	0	0	**4**
476	0	0	0	0	1	1	0	0	0	0	0	0	**2**
477	1	0	0	1	0	65	0	0	0	0	0	0	**67**
551	0	0	0	0	0	0	1	0	0	0	0	0	**1**
561	0	1	0	0	0	0	0	32	3	0	0	0	**36**
563	0	0	0	0	0	0	0	3	8	0	0	0	**11**
910	0	0	0	0	0	0	0	0	0	1	0	0	**1**
932	0	0	0	0	0	0	0	0	1	0	0	4	**5**
Total	**2**	**1**	**3**	**4**	**1**	**67**	**1**	**35**	**12**	**1**	**0**	**4**	**131**

Reference

Google Maps

-	471	472	474	475	476	477	551	561	563	910	931	932	Total
471	0	0	0	0	0	0	0	0	0	0	0	0	**0**
472	0	0	0	0	0	0	0	0	0	0	0	0	**0**
474	0	0	1	0	0	0	0	0	0	0	0	0	**1**
475	0	0	0	5	0	1	0	0	0	0	0	0	**6**
476	0	0	0	0	0	1	0	0	0	0	0	0	**1**
477	0	0	2	1	0	75	0	0	0	0	0	0	**78**
551	0	0	0	0	0	0	1	0	0	0	0	0	**1**
561	0	1	0	0	0	3	0	31	5	0	0	0	**40**
563	0	1	0	0	0	0	0	2	11	0	0	0	**14**
910	0	0	0	0	0	0	0	0	0	0	0	0	**0**
932	0	0	0	0	0	0	0	0	1	0	0	1	**2**
Total	**0**	**2**	**3**	**6**	**0**	**80**	**1**	**33**	**17**	**0**	**0**	**1**	**143**

Reference

Locatus

-	471	472	474	475	476	477	551	561	563	910	931	932	Total
471	1	0	0	0	0	0	0	0	0	0	0	0	**1**
472	0	0	0	0	0	0	0	0	0	0	0	0	**0**
474	0	0	2	0	0	1	0	0	0	0	0	0	**3**
475	0	0	0	7	0	0	0	0	0	0	0	0	**7**
476	0	0	0	0	3	0	0	0	0	0	0	0	**3**
477	0	0	0	4	1	79	0	0	0	0	0	0	**84**
551	0	0	0	0	0	0	1	0	0	0	0	0	**1**
561	0	3	0	0	0	0	0	40	0	0	0	0	**43**
563	0	0	0	0	0	0	0	8	6	0	0	0	**14**
910	0	0	0	0	0	0	0	0	0	1	0	0	**1**
931	0	0	0	0	0	0	0	0	0	0	0	1	**1**
932	0	0	0	0	0	0	0	0	2	0	0	4	**6**
Total	**1**	**3**	**2**	**11**	**4**	**80**	**1**	**48**	**8**	**1**	**0**	**5**	**164**

Reference

Appendix VII: Positional accuracy tables

Nova shopping area

Openstreetmap

Name	Distance in meters
La Dima	37.0
La Chica	32.3
Gall & Gall	30.5
KIK	20.7
Bristol	16.6
0	15.1
Vomar	14.3
Brasserie Vrienden	13.6
Brood by Alex	10.6
HEMA	10.0
Albert Heijn	9.2
Wibra	8.8
Etos	7.8
Blendz	7.7
Hans Anders	7.1
Zeeman	6.9
Eethuis Ensar	6.9
Intertoys	6.1
Handyman	6.0
Bloemenshop Rozeneiland	5.1
New York Pizza	4.7
Pearl Opticiens	3.9
Slagerij Coskun	3.0
Primera	2.9
La sosta	1.8
Bram Ladage	1.6
Bursa Kebab Restaurant	0.8

Google Maps

Name	Distance in meters
Bristol Utrecht Kanaleneiland	84.0
Holland & Barrett	54.8
vanHaren	41.6
AH Hammarskjoldhof	39.1
Action	36.6
Op=Op Voordeelshop	25.3
Handyman - Onderdelenhuis	18.7
Big Bazaar	15.8
Kruidvat	12.8
Gall & Gall Utrecht	12.7
Simitci Dunyasi	11.7
Bram Ladage	10.8
Okay, fashion & jeans	10.5
Hema	10.3
Blendz Shoes	10.0
Zeeman Utrecht Kanaleneiland	9.1

HFC Utrecht Kanaleneiland	8.3
Brasserie Vrienden	7.2
New York Pizza	6.5
Etos	6.2
Bread by Alex	5.2
Kik Textile Utrecht	4.1
Spakenburgse visspeciaalzaak	3.8
Bursa Kebap Restaurant	3.3
Hans Anders Opticien Utrecht	2.1
La Sosta Utrecht	1.7

Locatus

Name	Distance in meters
Albert Heijn	28.6
Big bazaar	24.1
Action	21.1
Kruidvat	18.6
Hema	17.7
Gall & Gall	13.5
Op = op voordeelshop	13.4
HFC café	13.2
Okay	12.2
Bazar oriental	11.9
Bristol	11.2
van Haren	10.5
Kik	10.3
Wibra	9.7
Eethuis Ensar	9.4
Vomar	9.3
Blendz	9.2
Zeeman	9.1
Brood by Alex	9.0
Bursa	8.2
Brasserie Vrienden	7.8
Holland & Barret	5.6
Pearle opticiens	5.5
Intertoys	4.9
La Chica	4.9
Hans Anders	4.3
Vivánt	4.2
Primera	4.1
Etos	3.8
Stam	3.7
Coskun	3.7
Bram Ladage	3.5
Rozeneiland	3.4
New York pizza	3.1
La Dima	2.8
Restore	2.0
Vis vallei	2.0
Handyman	1.6
Simitci Dunyasi	1.5
La Sosta	1.1

City centre

Openstreetmap

Name	Distance (m)
Aphrodite	84.2
Il Pozzo	44.9
La Fontana	29.5
De Oude Muntkelder	29.2
Zara	26.2
Denham	24.5
Humphrey's	20.2
Broese	19.2
image	18.5
Tilt !	17.3
Apollo Hotel Utrecht City Centre	16.3
Café De Stad	15.6
McDonald's	14.5
Café Ome Willem	14.2
Esprit	14.1
Nuestro Secreto	13.9
Perry	13.7
Scapino	13.4
Zara Home	12.6
KFC	12.3
Den Draeck	12.2
Quignon	11.6
KPN	11.2
The Sting	11.2
pipoos	10.8
Urban Outfitters	10.5
Broodje Mario	10.1
CoolCat	9.8
Didi	9.2
Timberland	9.0
Anna van Toor	8.9
Punte	8.3
Nespresso	8.2
Esprit	8.0
Manfield	7.5
The Society Shop	7.3
Covers	7.3
Morpheus	7.1
Mahanakorn	6.9
Van Dalen	6.6
Douglas	6.6
Expresso	6.5
WE	6.4
Lola&Liza	6.4
Mej. Janssen	6.2
K-Sjot	6.1
Bibliotheek Utrecht	5.9
La Cubanita	5.7
Hunkemöller	5.6
Meneer Smakers	5.5
Guts & Gusto	5.3
Basis	5.2

Name	Distance (m)
Belsimpel.nl	4.9
H&M	4.8
Floris van Bommel	4.7
Gandhi	4.6
Sacha	4.6
HEMA	4.5
Lens	4.4
Kuijper’s Hobbyhuis	4.4
L'Occitane	4.4
Rembrandt	4.4
Scotch & Soda	4.3
Kasteel Oudaen	4.3
Cocon	4.2
Tafel aan de Gracht	4.1
Nieuwe Dikke Dries	4.1
Winkel van Sinkel	4.1
Club Maggy	4.0
Mango	3.9
Claudia Sträter	3.8
Pearle Opticiens	3.7
Mutsaers	3.7
Pull & Bear	3.7
Vans	3.6
Buck's BBQ House	3.6
Gauchos	3.5
Kruidvat	3.4
Sacha	3.3
Los Argentinos	3.3
PK Bar & Kitchen	3.2
Manfield	3.2
Saffraan	3.2
Stöpler optiek	3.1
Chasin'	3.1
Just Brands	3.0
& Other Stories	2.9
El Borne	2.9
Broadway	2.8
Suitsupply	2.8
Vanilla	2.4
Manneken Pis	2.4
Hollister	2.3
Fusta d'Oro	2.2
Margaret Wines	2.1
DENHAM	2.0
Beers & Barrels	1.9
Arthur & Willemijn	1.9
San Siro	1.9
Bodytalk	1.9
Blue Phone	1.8
Image	1.8
La Cantina Di David	1.8
Pracht.nl	1.7

Flying Tiger	5.2
Veltman Liesting	5.2
De Potdeksel	5.1
Nelson	5.1
Zeeman	5.1
Caroline Biss	5.0
Superdry	5.0
G-Star Raw	4.9
La Grotta	1.1
Café Flater	1.0
Schiller Theater	1.0
Umami	0.9
Café Kalff	0.9

Supertrash	1.5
Pauw	1.5
New Tailor	1.5
Costes	1.3
Strand West	1.3
De Zwarte Vosch	1.3
W. Pijper	1.1
Balkan Grill Boro	1.1
Toque Toque	0.8
Broodnodig	0.6
Only	0.6
Erich De Gilde	0.5
Stach	0.3

Google Maps

Name	Distance (m)
il pozzo	39.6
Zara	36.4
La fontana	33.2
Kartoffel	21.7
Den Draek	21.1
Nuestro secreto	19.5
Tilt	18.9
Schrandt Koffers	18.8
Cafe de stad	17.0
Esprit	16.5
Elizabeth Wakefield styling	15.3
De Vossenpoort	15.0
KFC	14.9
Didi	14.6
Timberland	14.0
La senorita	13.6
Cantina di David	13.5
Kimmade	13.2
Cafe ome Willem	12.8
Coolcat	12.8
Veltman Liesting	12.7
't oude Pierement	12.7
Perry	12.6
Nespresso	12.6
the society shop	12.1
Oudaen	11.6
Caroline Biss	11.2
Lens.	10.9
McDonald's	10.5
Manfield	10.5
broodje Mario	9.9
Esprit	9.8
van Dalen	9.7
The Sting	9.6
Hollister	9.6
Pipoos	9.6
Hunkemoller	9.4
Bolia.com	9.4
Hästens	9.2

Name	Distance (m)
Be one	7.0
Flying tiger	6.8
& other stories	6.8
La Cubanita	6.7
pk	6.5
Quignon	6.5
San siro	6.5
't koffieboontje	6.4
pull & bear	6.2
Only	6.1
Aphrodite	6.1
Kwartier	6.0
Kruidvat	6.0
Colori	6.0
Mahanakorn	5.9
Gaucho's	5.9
Loetje	5.9
Paperbird	5.8
Chasin'	5.8
Meneer smakers	5.5
Sissy-boy	5.4
Lola liza	5.4
H&M	5.4
Anna van Toor	5.4
Mango	5.4
de Muntkelder	5.0
El borne	4.9
urban outfitters	4.8
Balkan grill Boro	4.8
Cocon	4.7
The Bluzone	4.7
Douglas	4.5
Strandwest	4.4
Essentiel Antwerp	4.4
Stach	4.2
Image	4.2
Jack's casino	4.1
Body talk	4.1
Vanilla	4.1

Ici Paris XL	8.8
WE	8.8
Covers couture	8.3
Scapino	8.3
Winkel van Sinkel	8.2
La grotta	7.8
Just brands	7.7
Vans	7.7
G-star raw	7.7
Pentik	7.7
Guts & gusto	7.6
Ghandi	7.5
SuperdryStore	7.3
Hermans	7.3
Riviera Maison	7.1
Claudia sträter	7.1
Casanova	3.3
pracht.nl	3.1
Only	3.0
India port	2.9
Maggy	2.9
Brasserie 't zusje	2.8
Oebens	2.7
Dagelijks lekker	2.7
Humphrey's	2.6
Denham	2.5
Nelson	2.5
Repeat	2.4
Pauw	2.3
Aspect	2.2
Cafe Flater	2.1
Saffraan	2.0

Mannekenpis	3.9
Manfield	3.9
Kookai	3.8
L'occitane	3.7
Kuijper's hobbyhuis	3.7
Sacha	3.7
Broodnodig	3.6
Broekmans & van Poppel	3.6
Buck's BBQ house	3.6
Cafe Kalff	3.6
Floris van Bommel	3.5
New Tailor	3.5
Suitsupply	3.4
Margaret wines	3.3
Nieuwe dikke Dries	3.3
Arthur & Willemijn	3.3
Peter Kaiser	2.0
Studio	1.9
beers and barrels	1.4
Umam	1.4
Zwaluwbar	1.4
Dunkin' Donuts	1.4
Los Argentinos	1.3
Suits 99	1.2
De zwarte vosch	1.2
Toque Toque	1.1
Mutsaers	1.0
De grammophoon winkel	1.0
Costes	0.9
Pemba	0.7
Erich de gilde	0.5

Locatus

Name	Distance (m)
Humphrey's	67.4
Pipoos	50.6
Hema	37.4
Apollo hotel	32.5
La fontana	30.1
Esprit	27.9
Image	27.0
& other stories	26.8
Broerse	24.6
il pozzo	22.8
MAC	22.8
SuperdryStore	22.1
Nespresso	21.7
de bibliotheek	21.0
urban outfitters	20.4
Perry	20.0
Morpheus	19.0
Didi	18.6

Name	Distance (m)
KPN	13.2
Mej Janssen	12.9
beers and barrels	12.7
Loetje	12.7
Winkel van Sinkel	12.6
La senorita	12.4
Mahanakorn	12.4
de Workout	12.1
Werftheater	12.1
K-sjot	12.0
Riviera Maison	11.9
Broadway	11.9
KFC	11.8
de werfkring	11.4
India port	11.3
van Dalen	11.3
Gaucho's	11.2
El borne	11.2

Esprit	18.3		Nelson	11.0
The Sting	18.1		Le Connaisseur	11.0
H&M	18.0		Buck's BBQ house	11.0
Manfield	17.8		Timberland	11.0
Tilt	17.3		Douglas	10.3
Anna van Toor	17.0		La Cubanita	10.3
pull & bear	16.8		Swordfish & friends	10.2
Cafe de stad	16.0		Maggy	10.2
Zara home	15.0		Sacha	10.2
Kiko	14.9		the north face	10.0
Rembrandt	14.7		Saffraan	10.0
't oude Pierement	14.6		Hunkemoller	9.9
Den Draek	14.6		Los Argentinos	9.8
Coolcat	14.5		Balkan grill Boro	9.8
Nuestro secreto	14.5		Kruidvat	9.7
Hollister	14.5		Taverna	9.5
Costes	14.1		Scotch & Soda	9.5
Cafe ome Willem	13.9		Basis	9.4
Ghandi	13.9		Paperbird	9.4
Vanilla	13.8		W. Pijper	9.3
tafel aan de gracht	13.7		Lola liza	9.3
Zara	13.7		Suitsupply	9.2
Mango	13.5		Scapino	9.2
Expresso	9.1		Meneer smakers	4.8
Cantina di David	9.1		de Muntkelder	4.4
Pearle opticiens	9.0		Blue phone	4.3
Casanova	9.0		Pauw	4.3
Kookai	8.9		Stöpler optiek	4.1
Repeat	8.7		Kwartier	3.8
Pentik	8.4		Broekmans & van Poppel	3.6
Be one	8.4		Sacha	3.6
Bolia.com	8.0		Caroline Biss	3.4
G-star raw	8.0		Aspect	3.3
Chasin'	7.8		WE	2.9
Suits 99	7.7		Broodnodig	2.9
broodje Mario	7.6		De zwarte vosch	2.8
Fusto d'oro	7.6		Claudia sträter	2.8
McDonald's	7.6		Guts & gusto	2.8
Moscow	7.5		The Bluzone	2.8
Aphrodite	7.5		Only	2.6
Quignon	7.3		Peter Kaiser	2.5
Belsimpel.nl	7.1		New Tailor	2.5
Strandwest	7.0		Denham	2.3
Arthur & Willemijn	6.9		de potdeksel	2.3
Lens.	6.9		Body talk	2.1
Flying tiger	6.5		La grotta	1.9
Schiller theater	6.5		pracht.nl	1.6
't koffieboontje	6.4		Mannekenpis	1.6
Oebens	6.4		Covers couture	1.5
Elizabeth Wakefield styling	6.3		No vintage phobia	1.5
Sissy-boy	6.2		Mutsaers	1.4
Nieuwe dikke Dries	6.1		Schrandt Koffers	1.4
the society shop	6.0		Erich de gilde	1.4
Just brands	5.9		Margaret wines	1.3
pk	5.7		Veltman Liesting	1.2

Colori	5.7		Umam	1.2
Manfield	5.6		Cocon	1.0
Hermans	5.5		Stach	1.0
Only	5.4		Cafe Flater	0.9
Toque Toque	5.4		L'occitane	0.9
Oudaen	5.3		Cafe Kalff	0.9
Floris van Bommel	5.0		Swirl's ice cream	0.8
San siro	4.9		Vans	0.8
Kuijper's hobbyhuis	4.9		Ici Paris XL	0.7

Appendix VIII: Pragmatic quality table

ID	Name	Study area	Locatus_type	Locatus_dist	OSM_type	OSM_dist	Google_type	Google_dist
1	Dunkin' Donuts	Centre	-		-	-	Cafe	1.4
2	Meneer Smakers	Centre	Fastfood	5.0	Fastfood	5.5	Restaurant	5.5
3	McDonald's	Centre	Fastfood	8.0	Fastfood	14.5	Meal takeaway	10.5
4	KFC	Centre	Fastfood	11.5	Fastfood	12.3	Restaurant	14.9
5	Broodje Mario	Centre	Fastfood	8.5	Fastfood	10.1	Restaurant	9.9
6	Manneke Pis	Centre	Fastfood	0.9	Fastfood	2.4	Restaurant	3.9
7	La Sosta	Nova	Grillroom/shoarma	0.9	Fastfood	1.7	Meal delivery	1.9
8	Bursa	Nova	Fastfood	7.6	Restaurant	0.6	Restaurant	3.5
9	New York Pizza	Nova	Delivery/Takeaway	3.0	Fastfood	4.9	Meal delivery	6.6
10	HFC café	Nova	Fastfood	13.1	-	-	Restaurant	8.6
11	Bram Ladage	Nova	Fastfood	3.4	Fastfood	1.6	Restaurant	10.3
12	Eethuis Ensar	Nova	Grillroom/shoarma	9.0	Restaurant	6.7	-	-

Appendix IX: OSM edits per user in Utrecht (2013 – 2017)

user	edits
Gertjan Idema_BAG	184,355
rivw_BAG	91,388
Zugführer_BAG	38,484
Gertjan Idema	25,469
Sander H_BAG	24,701
sander79	8,620
Hendrikklaas	7,348
Zugführer	5,741
ArjanO	5,343
Martin Borsje_BAG	5,245
It's so funny_mechanical	5,118
AnkEric	4,202
Its so funny_mechanical	3,716
de vries	3,002
AEelderink	2,768
Sander H	2,474
The Maarssen Mapper	2,420
Christoph Lotz	2,311
rivw	1,749
Imergis	1,469
Maarten Deen	1,462
fsteggink	1,406
cartinus	1,182
brbbl	1,046
Jasmon	875
HanW	838
Martien Sch	767
jeij	695
It's so funny	694
gvb	685
JJJWegdam	671
padvinder	669
Davio	591
PeeWee32	568
Commodoortje	562
AlbertP	533
CJTmmr	528
cartinus_BAG	520
ligfietser	511
VRU1	485
WJtW	466
KartoGrapHiti	457
mtrosm	439
Pander	432
Skywave	360
AlbertP_BAG	352
Andre Engels	329
Jozzy	321
nilodo	295
Martin Borsje	272
martijnschmap	272
BAGgeraar	231
OliverH	222

user	edits
Jvanbiezen	220
pjdebruin	206
VictorVan	193
sebastic	185
jaimemd	183
Sven Witte	178
Karlsmam	177
wimvantklooster	171
HYS	168
marczoutendijk	165
Kippii	164
Dutch Mapper Mechanical	157
groothuisss	155
Math1985	140
Herman56	130
eggie	127
GercoKees	121
HourOfTheWulf	114
dvdhoven	108
Amaroussi	107
kannix	106
fx99	101
gjp_osm	99
HarryS	97
Wisse Jelgersma	90
wvdp	88
drMerry	84
dionysus1975	76
tizzossos	75
Walter Schlögl	74
Virtugon	73
FvGordon	72
Kingigi	67
maggot27	65
Jan Westerhof	60
raldee	60
mboeringa	59
the Sandinator	57
Whimself	56
ijsb	56
glenn1236	55
elmarburke	55
milovanderlinden	55
srebbe	54
albiobola	54
nimapper	52
David_456	52
Vincent Vandalon	51
IIVQ	50
paulbe	50
peterthorn	49
Dirk V	48
Voidmapper	45

Graeme Herbert	43
Jerrol	42
hendrix123	41
mprins	40
edhaas1	39
emvee	38
brandmeester	38
slhh	36
micjon	35
xnyhps	34
HenkL	34
wheelmap_visitor	33
DRW	32
Vincy	32
lucb1e	32
woodpeck_repair	31
bu11frog	31
Heliosmaster	30
nicorikken	30
UserUtrecht	29
SomeoneElse_Revert	29
Vanmeule	29
Cavit	28
Cees Mechielsen	28
malcolmh	28
Maarten van der Veen	28
Lucacker	27
b3r3nd	27
bdiscoe	26
Mcke	26
CreaProdu	26
byckel	25
Frank314	23
Danny Oscar Kip (The Doctor)	23
apeb12	23
JaVaWa	21
arvdk	21
SPTH	21
Boeicop	19
Pim Immink	19
robbert-t	19
Dutch Mapper	19
DTeelde	19
Freek Bes	19
Polyglot	19
Polarbear	18
borishag	18
Martjan	18
Math1985_mechanical	18
jaap de vries	18
Allroads	17
hankwang	17
MadDogSailor	17
Manu1400	17
Zandlopertje	17
hatzfeld	16

MacTarmac	16
oini	16
bklooster	15
Classical geographer	15
Hjart	15
bas(sie)	14
pipodeclown12	14
groen_duiven	14
flierfy	14
hark	14
captain007	14
KawaiiFubuki	14
JWegdam	13
tovvie	13
David Vrolijk	13
Léon	12
edhelas	12
Kruising	12
TomMoekotte	11
Marcel030	11
Glipglob	11
The Beum	11
WesleyE	11
Nearo	11
R0bst3r	11
Tubeliar	11
Henry Tian	10
raymondnijssen	10
Maarten Bosmans	10
georgfuchs	10
Wout van Dijk	10
Remy Blom	10
Menno Tammens	10
edugof	10
Real Shadowman	10
mueschel	9
PatrickBos	9
Pieter Spiekman	9
jack1968	9
unknown365	9
Roger de Graaf	9
Med	9
Ratverlegh	9
Martijn Gastkemper	9
GeniusDex	8
irateh	8
Rudolf Mayer	8
andrewpmk	8
ChrissW-R1	8
JosvR	8
tookregor	7
MHCvT	7
tekenis	7
dok33	7
tijmen w	7
EstherH	7

tmatz81	7
jeroen020	7
YogiRambo	7
absmn	7
mrvanes	7
BikeAway	7
NDW Servicedesk	7
Sint E7	7
peSHIr	7
gblab	6
stroet43	6
CeesW	6
Magnatic	6
BCNorwich	6
JorisBerg	6
Snorski	6
WKOSM	6
ElGrandeBanana	6
piderman31415	6
Blauwdruk	6
wheelmap_android	6
John-map53	6
aytfadc	6
topoSheep	6
element100	6
Koen Veenenbos	6
ratrun	6
ajkeizer	6
uboot	6
Spinpoot	6
PiHe	6
Tom Freijsen	6
itserik	5
Koen Rijnsent	5
Batareikin	5
LivingWithDragons	5
zacmccormick	5
Bart Schulte	5
g j roelvink	5
ErikG1974	5
dylan_mulder	5
rstok	5
Krille von Stralau	5
Jeroen Muris	5
PimVerver	5
Jedrzej Pelka	5
Its so funny	5
sauls	5
tomasvandemeent	5
user_5359	5
ulmtuelp	5
highflyer74	5
H4rdcore!	5
StevAlba	4
highbuilder	4
Abius	4

Mats Hofman	4
tbicr	4
Michaelvo	4
WOOO	4
Johnb53	4
Klaas Kruidhof	4
marczoutendijk_mechanical	4
Mentam	4
Webmapper	4
NiekvdMaas	4
seagrave	4
Marc Weustink	4
Joost Claessen	4
Lekiamnavkniredlaa	4
RJL1987	4
Jur_	4
Harmen1993	4
Martin2009	4
RedFox	4
Kratakooi	4
Floris Jan	4
Kosturnix	4
HugoI	4
escada	4
West Erop	4
Rien_23	4
Sanshinkai Aikido Utrecht	4
jutezak	4
tosher	4
user_568	4
Kerstkonijn	4
ijsbaan	3
Mart81	3
lesko987	3
wouterdriezen	3
gpsfietser	3
Eupeodes	3
choess	3
bibi6	3
eribog	3
koen88	3
ligtvoet	3
Sybren	3
sara shafiee	3
Roel Stemerdink	3
Rienk Mebius	3
patrickvhoeijen	3
Rens Wientjes	3
mzve	3
utack	3
pyth	3
Matthijs	3
ErikO	3
A67-A67	3
Bullroarer	3
WouterW	3

whateverm8	3
rainygarden	3
patzi	3
whturner	3
Thibaud	3
RoyBraam	3
Mike Bosschaert	3
venerdi	3
Wim van Dorst	3
jurvre	3
Sannetjaah	3
PinkfloydHL	3
mapmapdotnl	3
Jan Maat	3
Martin_GPS	3
skunk16	3
Wesley In Der Maur	3
Darkklaw	3
Tvw	3
hvdwolf	3
poppei82	2
Siebrand	2
MacFreek	2
jawaagmeester	2
JaapK	2
robbiex	2
micheldejongh	2
Myquandro	2
GrooveBooster	2
spesnova	2
Robin_p	2
maria_jcb	2
cbrace	2
gnijholt	2
Essex_Boy	2
cdavila	2
gormur	2
Rag	2
CitymapperHQ	2
Jacco Lammers	2
thetornado76	2
Reintje	2
mapper999	2
rayquaza	2
JeanFred	2
ricodejong01	2
bvbever	2
Crimon8ten	2
andreklaver	2
manyac	2
Test360	2
paulanca	2
JanFi	2
Steven Vance	2
rolphvkuijk	2
look4book	2

nammala	2
Hexaedair	2
Area-controler	2
Paulext	2
Rien van der Laan	2
MapEdG	2
ToffeHoff	2
cantece	2
Macumba Macaca	2
habhan	2
kasteel	2
Reinout	2
VKR DeBilt	2
Creek	2
eyebidem	2
DoubleA	2
mennoowh	2
ff5722	2
Aleks-Berlin	2
GarethD	2
Luuk B	2
ediyes	2
kisaa	2
cbdavis	2
Zeehond	2
mr-g	2
Thijs Brentjens	2
RicoZ	2
David Crochet	2
JelleZijlstra	1
Inca0	1
Rudus	1
ter-burg	1
cruiserpaule	1
Cearon	1
BuganiniQ	1
Berf	1
Gerard61	1
Myckel	1
MappingDog	1
thorum	1
yasio90	1
you-zs	1
cengelen	1
Jano John Akim Franke	1
linuzer	1
Alecs01	1
ziTneY	1
BergingM	1
Victorvk	1
vademecum	1
Parie	1
Markus59	1


Points of interest of three different data sources in the Nova shopping area, Utrecht Kanaleneiland (2018)
N
GIMA
Legend
Reference
471
472
475
476
477
561
Openstreetmap
471
472
475
476
477
561
563
Locatus
471
472
475
476
477
561
Google Maps
471
472
475
477
561
BERNADOTTELAAN
BENELUXLAAN
VAN VOLLENHOVENLAAN
DE LOUTERLAAN
5 MEIPLEIN
KONINGIN WILHELMINALAAN
CHURCHILLLAAN
TRUMANLAAN
0
12.5
25
50 Meters
Description:
This map shows the result of a geographic data quality research of Locatus, Google Maps and Openstreetmap in the city of Utrecht. The map shows the locations of retail points of interest of the test datasets and of a reference datasets. The points are classified in the Standard Industrial Classification of the Dutch Central Bureau for Statistics. Corresponding points are presented with the same number.
Reference points that are absent in the test datasets indicate a completeness error, points with a deviating classification in comparison with a corresponding reference point indicate an attribute error and points with a different location in comparison with a corresponding reference point indicate a positional error.
Sources:
Locatus: Sales points Utrecht
Check date: April, 2017
Google Maps API requests
Accessed: January, 2018
Geofabrik: Openstreetmap data of The Netherlands
Accessed: January, 2018
Kadaster: Basisregistraties Adressen en Gebouwen
Version: December, 2017
Stefan van den Berg
Printed: Utrecht University Feb 28rd, 2018
Coordinate System: RD New
Projection: Double Stereographic
Datum: Amersfoort
False Easting: 155,000.0000
False Northing: 463,000.0000
Central Meridian: 5.3876
Scale Factor: 0.9999
Latitude Of Origin: 52.1562
Units: Meter
Esri Nederland & Community Maps Contributors
134800
134900
135000
135100
135200


Points of interest og three different data sources in a selection of the city centre of Utrecht (2018)
GIMA
Geographical Information Management and Applications
N
Legend
Reference
471
474
475
476
477
551
561
563
910
931
932
Openstreetmap
471
474
475
476
477
551
561
563
910
932
Locatus
471
474
475
476
477
551
561
563
910
932
Google Maps
474
475
477
551
561
563
932
Description:
This map shows the result of a geographic data quality research of Locatus, Google Maps and Openstreetmap in the city of Utrecht. The map shows the locations of retail points of interest of the test datasets and of a reference datasets. The points are classified in the Standard Industrial Classification of the Dutch Central Bureau for Statistics. Corresponding points are presented with the same number.
Reference points that are absent in the test datasets indicate a completeness error, points with a deviating classification in comparison with a corresponding reference point indicate an attribute error and points with a different location in comparison with a corresponding reference point indicate a positional error.
Sources:
Locatus: Sales points Utrecht
Check date: June, 2017
Google Maps API requests
Accessed: January, 2018
Geofabrik: Openstreetmap data of The Netherlands
Accessed: January, 2018
Kadaster: Basisregistraties Adressen en Gebouwen
Version: December, 2017
Stefan van den Berg
Printed: Utrecht University Feb 28rd, 2018
0
37.5
75
150 Meters
Coordinate System: RD New
Projection: Double Stereographic
Datum: Amersfoort
False Easting: 155,000.0000
False Northing: 463,000.0000
Central Meridian: 5.3876
Scale Factor: 0.9999
Latitude Of Origin: 52.1562
Units: Meter
Esri Nederland & Community Maps Contributors
136100
136200
136300
136400
136500
136600
136700
136800
136900
137000
137100