MASTERARBEIT / MASTER’S THESIS

Titel der Masterarbeit / Title of the Master‘s Thesis

„Methods of Quality Assurance for Crowdsourced Geodata“

verfasst von / submitted by

Magdalena Koubek, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 856
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Kartographie und Geoinformation
Betreut von / Supervisor:	Ass.-Prof. Mag. Dr. Andreas Riedl

Table of Content

List of Tables

List of Figures

List of Abbreviation

BEV	Bundesamt für Eich- und Vermessungswesen
CSGI	Crowd Sourced Geographic Information
CRS	Coordinate Reference System
FRP	Fire Rescue Path
GI	Geographic Information
GIS	Geoinformation System
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
INSPIRE	Infrastructure for Spatial Information in the European Community
ISO	International Organization for Standardization
OGD	Open Government Data
OSM	OpenStreetMap
POI	Point of Interest
QA	Quality Assurance
UGC	User Generated Content
VGI	Volunteered Geographic Information
WWW	World Wide Web

Abstract

Crowdsourcing has become a major part of the world of geographic information. Several aspects of spatial data can be covered with crowd users, volunteers or non-experts. In this certain project I work with a crowd which is not voluntarily but works for a financial remuneration. I want to find out which aspects of quality are important to this project and which methods there are for the process of quality assurance. The goal of this thesis is to determine the need and the benefit of each method. I will show how much time each method needs to be executed and which result comes along with this method. The mentioned project was an attempt to map fire rescue paths in a certain area in the city of Munich. This project was initialized by the company Parkbob GmbH. Parkbob offers digital information about parking restrictions in certain cities all over the world. As fire rescue paths are widely spread over Munich and as parking and standing is absolutely forbidden on these areas, it was essential to include those into the parking datasets. Unfortunately, there was no dataset about the fire rescue paths, therefore Parkbob decided to hire a crowd which should collect all fire rescue paths of the research area. In this thesis I want to explain how the project was built up and what aspects of the framework need quality assurance. Then I will shortly have a look at crowdsourcing in general, I will give examples of other projects and extract the methods which they were using. Also, I want to explain the parameter of quality of spatial data based on literature research. After that I will take the methods I can use for this project and test them all on the same dataset of the collected fire rescue paths. The results show what approaches did not bring the desired results and which methods would have reduced the invested time. Finally, I will discuss the advantages and disadvantages of these methods and explain critically what can be done differently in similar future projects to increase the quality of the data.

Kurzfassung

Crowdsourcing ist bereits seit Jahren fester Bestandteil der GIS-Welt. Es gibt zahlreiche Aspekte die mittels Freiwilliger, Laien oder einer Crowd abgedeckt werden können. In diesem speziellen Projekt arbeite ich mit einer Crowd die nicht freiwillig partizipiert, sondern für eine monetäre Bezahlung. Ich möchte mit dieser Arbeit herausfinden, welche Aspekte der Qualität für dieses Projekt von Bedeutung sind und welche Methoden existieren, um eine Qualitätssicherung durchzuführen. Das Ziel dieser Arbeit ist den Nutzen und die Kosten für jene Methoden zu bestimmen. Daher möchte ich zeigen, welche Aufwendungen benötigt werden, um jede einzelne dieser Methoden durchzuführen und welches Ergebnis diese zeigen.

Das erwähnte Projekt war ein Versuch sämtliche Feuerwehrzufahrten in einem bestimmten Gebiet in München zu digitalisieren. Erstellt und durchgeführt wurde dieses Projekt von der Firma Parkbob GmbH. Parkbob bietet digitale Informationen über den Parkraum verschiedener Städte auf der ganzen Welt. Da Feuerwehrzufahrten sehr häufig in München vorkommen und Parken und Halten hier absolut verboten ist, war es für Parkbob besonders wichtig diese Informationen in das Dataset zu inkludieren. Leider existierte bisher noch keine Informationen über die Standorte der Feuerwehrzufahrten, daher hat sich Parkbob dazu entschlossen selbst eine Crowd zu engagieren die diese Aufgabe übernimmt.

Mit dieser Arbeit möchte ich die Rahmenbedingungen dieses Projekts beschreiben und erklären welche Aspekte Bedarf an eine Qualtitätssicherung haben. Ich möchte auch auf Crowdsourcing allgemein eingehen, verschiedene Projekte, die damit arbeiten beleuchten und deren Methoden extrahieren. Außerdem werde ich Qualität in Bezug auf Geodaten anhand von Literarturrecherche genauer beschreiben. Anschließend werde ich jene Methoden, die für das München-Projekt verwendbar sind, praktisch anwenden und vergleichen. Das Ziel ist es zu sehen, welche Methoden die gewünschten Ergebnisse bringen und bei welchen Methoden der zeitliche Aufwand nicht in Relation zu dem Ergebnis steht. Zum Schluss möchte ich die Vor- und Nachteile der einzelnen Herangehensweisen analysieren und das Projekt kritisch betrachten. Somit kann anhand dieser Arbeit Konsequenzen für zukünftige Projekte gezogen werden, um die Aufwendungen zu verringern und die Qualität zu erhöhen.

1. Introduction

Currently there is a high impact on GIScience from the increasing number of VGI (Volunteered Geographic Information) contributions from individuals (Jiang, 2012). Many companies see a chance of using user generated content (UGC) to improve their services faster and cheaper than with conventional sources. This is not limited to geodata, companies from every kind of industry source out part of their work to profit from the great mass of short-term tele-workforce. This is also true for spatial data and shows a shift from the conventional top-down approach for data acquisition, where few experts are gathering the data, to a bottom-up approach, where a higher number of non-experts are collecting data. (Howe, 2009)

In the last 20 years the way geodata was created changed a lot due to the phenomenon of user generated content. Innovation in technologies when it comes to handheld devices or geographic information technologies and the popularization of the internet lead to this appearance. After the emerging power of VGI platforms there is also a commercial interest in crowdsourcing data. Companies may need crowdsourced spatial data for their daily work. It can happen that companies need specific data. They are also willing to pay for high quality datasets. However, sometimes the need of specific data cannot be satisfied as there is still data which has not been mapped so far. During my work at Parkbob, we often came across the problem, that there are objects which have no corresponding spatial dataset. There are already companies (Coord, 2020, Waze, 2020) which are using crowdsourcing for gathering specific spatial data, and Parkbob also wanted to become one of them.

As it is then up to non-experts to collect this data, the quality can vary a lot. Companies which use this data commercially, have the liability to build their service up on reliable data. Here, the question about the quality of the data created by non-experts, comes up. There are papers handling different kind of methods for assessing the quality of VGI like Senaratne et al (2017) show in their paper. Other research like Exel, Dias, & Fruijtier (2014) are analysing the indicators of quality of crowdsourced data by comparing it with already existing commercial data or institutional data.

In most cases the data is compared with commercial or administrative datasets which, however, are not always accessible due to the lack of availability, contradictory licensing restrictions or high procurement costs. (Barron, Neis, & ZIpf, 2014)

Barron (2014) presents some ways to evaluate data without comparing it to already existing datasets. He calls those intrinsic approaches, which means that no reference dataset is needed to control the quality. This is also what I need for the FRPs project. As there is no reference dataset, how can the quality of the data be validated?

Some of the mentioned papers are high in value when it comes to the theoretical background of spatial data and more specific in spatial crowdsourced data and its quality assurance. But when it comes to a certain use case, the validation proposals need to be adapted. Therefore, I wanted to check if these quality assurance (QA) methods are also usable in a practical environment.

During my work at Parkbob I came along missing datasets for specific objects quite often. During the data collection of the City of Munich the struggle about the missing information of Fire Rescue Paths (FRPs) was big, so a crowdsourcing project to collect this information was established. However, the QA of the data was very time intensive, and I saw that we would not be able to scale up this approach to other cities. As the need of letting crowd users submit spatial data will also occur in other cities, I had to find solutions how the QA of this kind of data can happen faster.

1.1. Objectives and Research Questions

The aim of this thesis is to compare different approaches for quality assurance of spatial data and especially of crowdsourced spatial data. In the first step those approaches will be collected and extracted from papers and from companies and projects directly, as they are doing it in practice. There are companies which are already gathering spatial data via crowdsourcing, which I want to list within this thesis and show their approaches and methods of QA. I will check if they can be adopted for the use case in this thesis as well.

Within the thesis there should be a clear overview of the parameters of quality of spatial data and of the indicators of quality for crowdsourced data. Additionally, I want to give different methods how the measurements and indicators can be determined and focus here on intrinsic methods, where no existing reference dataset is needed. These methods will be tested within the framework explained in chapter 2, So, this thesis shows a practical example of QA of crowdsourced spatial data. A company should be able to decide how they want to verify the spatial data based on the results of this thesis. The main results should show time expense and quality measurements of the methods for this certain dataset. Based on this aim, I generated following research questions:

- Theoretical research
    - What are measurements of the quality of crowdsourced spatial data?
    - What methods of QA of spatial crowdsourced data exist?
    - Which QA methods are scalable?
- Project based conclusion
    - Which is the cheapest method of QA for crowdsourced data?
    - Which method of QA gives the most trustful result?

My personal goal is, that after this thesis is written, it is obvious for me and for Parkbob how future crowdsourced projects can be established and to have a customized framework for the desired level of quality.

1.2. Outline of the Thesis

In order to answer the previously defined research questions, I will first give an overview of the crowdsourcing project which is the framework of this thesis. I will explain the research area, the crowd users and how the dataset was established. In the third chapter I will focus on the background of crowdsourcing and spatial data. The characteristics of crowdsourced geodata and its problems will be topic here. I will also list some examples of projects and companies who are already dealing with the problem of the quality assurance of crowdsourced spatial data in this chapter. I want to give an overview of the methods they are already using. In the following chapter theoretical knowledge about the quality of spatial data and about the indicators which come along crowdsourcing spatial data are discussed. After that, I will write about the categories of methods and the actual methods of QA.

In chapter 6 I want to present the results of the methods for the use case. Finally, in the last chapter I will discuss these result and write about the findings and conclusions and moreover, answer the research questions. Least, I will give suggestions for future work and what I expect has the potential to become a topic of further research.

2. Fire Rescue Paths Project

This master thesis is created due to a project at the company Parkbob. Here in this chapter I want to explain the framework of this thesis and explain why there was the need for further research into the topic of quality assurance of crowdsourced spatial data. Moreover, I will explain the study area and give insight in the crowd users' characteristics. Then I give an overview of the data which is used for the methods application in the chapter 5. I want to give details about how the data was collected, stored and validated in first place.

During my work at Parkbob GmbH, I was part of a team which was responsible for the collection and transformation of data for the parking app Parkbob offers. Parkbob offers services concerning parking data in several cities worldwide. The main workflow includes the collection of data, the transformation into a heterogenous dataset and the publication through an app for Android and IOS smartphones. In most cases the company use public data sources like OGD (Open Government Data) sources, however in some cases it is not possible to take existing datasets due to usage restriction or simply because they do not exist. I was focusing on data collection of the city of Munich, when the project was started.

Fig. 1: Parkbob Logo. Parkbob (2019)

While collecting parking information for Munich, Parkbob realized that there was a lack of information of fire rescue paths (FRP). On FRPs parking, standing and stopping is absolutely forbidden and as they are widely spread all over the city of Munich, it was very important to Parkbob to include those parts of the street in a parking dataset. While researching for a dataset which includes exactly these information, Parkbob realised that the data does not only not exist, however there are many institutions, like the fire brigade of Munich itself, who would be highly interested to have this kind of information for themselves. Because of the need of this data, the Munich Crowd Project was established. The idea was that the information was collected by crowd users who are living on the spot. These crowd users submitted FRPs via a third-party app named *GISCloud*. The submitted points where collected and manually validated by Parkbob. Then the data was included into the overall dataset of Parkbob, to publish it via the Parkbob app.

Fig. 2: FRP in Munich. Captured 2018

2.1. Area

Fig. 3: Area of collected FRP in Munich. Map: OSM

The area where the FRPs where collected does not include the whole city of Munich, but an area which is inside of the so-called 'Mittlerer Ring' which is the state route B2R. Within this area the streets are part of the parking management system of the city of Munich. (Stadtverwaltung München, 2019)

As Parkbob covers not the whole city but focuses on the areas where parking management is active, the study area is limited to the space within the pink line shown in Fig. 3. The study area has a size of 44.58 km^2. The structure of the city is very heterogenous. It varies from a high density of inhabitants with many buildings and little space to parks and green banks besides the river Isar. Also, more rural parts with single family homes are within the study area. This diversity has impact on the QA of the project as you can see in chapter 5.5, as the FRP are not equally distributed over the city but are dependent from buildings and streets.

2.2. User

Those people who were contributing to this project are called crowd users or contributors. The project team only has little information about the contributor. Besides name, birthdate and email address only the payment information was collected. To contribute to the project, crowd users did not need any special preconditions or qualifications. Only the possession of a smart phone with Android or IOS software was needed. To find people who want to contribute to this project, Parkbob announced a job offer via job platforms and mini job groups on Facebook. Within one month, there were 58 requests to the Job offer, but only 14 people took the next step and signed the working contract, which is about ¼ (24%). From 14 people who signed the contract only 7 people finally contributed to the project.

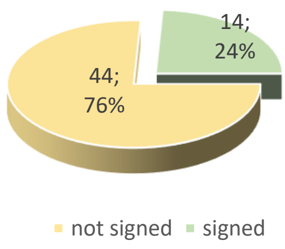


Fig. 4: Contract status.

Username	Contributions numeary	% from all contributions
Emr.	78	1.42%
Lei.	87	2.00%
Bah.	185	5.06%
Sam.	362	9.91%
Ser.	485	12.37%
Fru.	452	13.28%
Azu.	2044	55.95%
Total	**3653**	**100.00%**

Table 1: Contributions in percent and absolute numbers.

Also, among those who contributed to the project is a lot of variety between the users when it comes to the number of contributions. The participation inequality which is appearing within this case is not a new problem. It is *"the phenomenon that a very small percentage of participants contribute a very significant proportion of information to the total output",* as Haklay (2016) wrote. It is a very common problem among crowdsourced projects and within online communities in general and has been observed over decades on different platforms (Haklay M. , 2016). Not only VGI projects are concerned, as an example Nielsen (2006) shows that on the crowdsourced platform Wikipedia, 0.003% of all users contribute two-thirds of the content. In general, the participation inequality follows a 90-9-1-Rule:

- 90 % of all users do not contribute, but only read or observe
- 9 % of all users contribute not regularly and not a lot
- 1 % of the users are responsible for the main content

When comparing the numbers with our use case, 7 users from 58 contributed to the project which is 12.07%. 100% of all contributions were submitted by 12.07% of all users and 55.95% was submitted by the most active user who makes only 1.72% of all users. To give an overview over the numbers, I summed the rounded numbers up in Table 2. Seeing these numbers makes it clear that this project is not an exception from the participation inequality and follows almost perfectly the 90-9-1 rule. Barron (2014) explained the biggest problem with this phenomenon: *"The less contributors that are responsible for the major proportion of the data the higher the dependence on those few."*

Username	Contribution in %	Percentage from all users
other users (51)	0%	88%
emr	1%	
Lei	2%	
Bah	5%	
Sam	10%	10%
Ser	12%	
Fru	13%	
azu	56%	2%

Table 2: Demonstration of the 90-9-1 Rule.

The dependency on only one user "azu" within this project is alarmingly. Without that one user, 56% of the contributions would not exist. This experience could help the project team for the next project to implement tools which help to get a bigger range of well contributing crowd users.

2.3. Data

The data which was collected during the project was submitted by crowd users and includes information about the FRP, the position and information about the smartphone itself. I want to explain three phases of the establishment of the database: First step was the collection phase, when a user submits the FRP to our server. Second, the data cleaning, as I did not want to include some of the points in the data analysis the dataset has to be cleaned. Least, the validation phase, which I will explain in detail in chapter 5, however here I will give an overview.

The following figure shows the dataflow of the project. After the data was created, the data was stored within the GISCloud environment at first place. Then, an automated file transfer from the GISCloud server had to be created to download the information on a Parkbob internal system. There the validation process happened. After the submissions had been checked by Parkbob, all positively validated points had been uploaded to the GISCloud server, so the crowd users can see which FRP have already been submitted. The same points also got processed to lines and included in the overall dataset of Munich in the Parkbob environment.

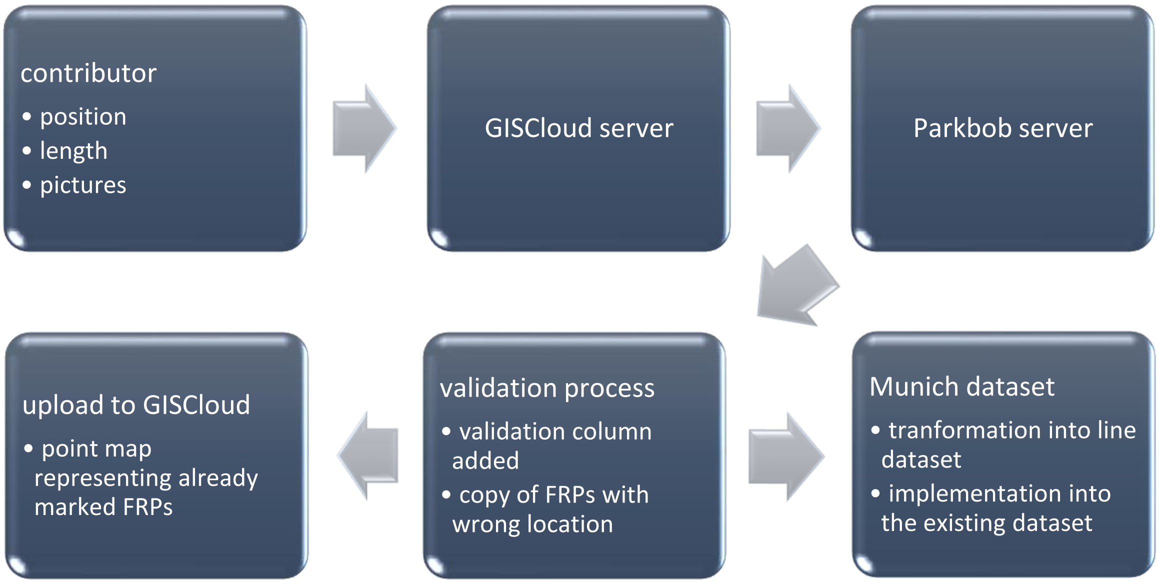


Fig. 5: Workflow of the Munich FRPs project.

2.3.1. Collection

The data itself was collected by people who are living around the research area and use their smartphone to submit FRP while they get along in their daily routine. The preconditions to become a crowd user should be the least limiting as possible, so many people could join in. In this case the user needs to download an App GISCloud (GISCloud, 2019), which is available for Android and IOS. The idea was to keep the collection process as simple as possible and allow the users to take part in the project while they continue living their normal lifestyle, with only stopping by shortly when a FRP is spotted.

The submission includes two types of information, the automatic tracked information which includes phone type, camera type, GPS location, position accuracy and date and the information which the user must submit manually like a picture of the FRP, the length of the FRP and the manual pinned location of the FRP as you can see in Fig. 6.

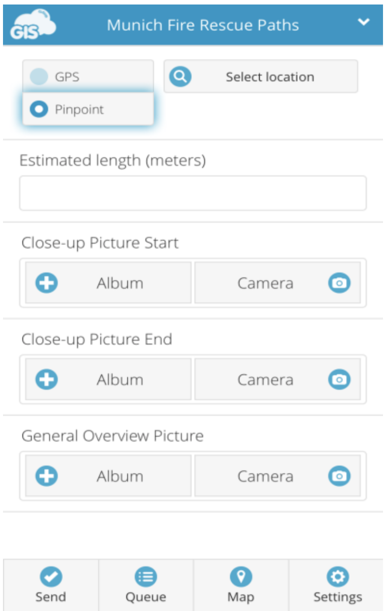


Fig. 6: Screenshot from the GISCloud app.

This manual information is mandatory to include in the submission, as it would be not possible to validate process the submission without. The information about the location is submitted twice, on the one side automatically via the GPS signal on the phone and on the other side the user drops a pin on a map the position where the FRP is located. This helps to identify low quality of spatial information about the FRP. In Table 3 I want to show two example records of the data derived from the app when someone submits a FRP. In Table 3 all fields of a record are shown. Most of the fields get the information during the submission process. *Latitude*, *longitude*, *altitude*, *accuracy*, *altitude_Accuracy* and *time* are parameters which are filled up by the information from the GPS signal of the mobile phone. The *compass* information is created from sensors from the smartphone. *Model*, *version*, *platform* and *app_version* present information about the smartphone itself and the app. The *username* is based on the log-in information within the GISCloud app. The *estimated_length_meters* is an empty field which the user has to fill up during the submission process (see Fig. 6). This number should represent the length of a FRP in meters. *Close_up_start*, *close_up_end* and *general_overview_picture* are the paths of the pictures which the user had to take during the submissions and show the path of the storage.

The fields *validation* and *notes* are created during the postprocessing. These are filled up by the project team. If a submission was accepted the *validation* field is filled with passed, if the submission can be used but further processing is needed with passedw (passed but wrong), if the submission fails the quality assurance, this field get filled with failed. In that case also the field *notes* will be filled up with the reason why the point failed. After the project ended 3692 submissions were recorded with a total estimated length of 20 725 m.

estimated_length_meters	**7**	**3**
_created	43087,4075	43088,61213
close_up_start	P:\GIS\CrowdSourcing\998…	P:\GIS\CrowdSourcing\998…
close_up_end	P:\GIS\CrowdSourcing\998…	P:\GIS\CrowdSourcing\998…
general_overview_picture	P:\GIS\CrowdSourcing\998…	P:\GIS\CrowdSourcing\998...
validation	failed	passed
notes	outside of service area	
latitude	4,82121E+13	4,81463E+13

longitude	1,15544E+13	1,15843E+13
altitude	483	506
compass	1,09317E+14	4,34655E+13
accuracy	3,43447E+14	10
altitude_accuracy	10	3
time	1516410813	1516722044
username	ser	azu
model	iPhone7,2	iPhone8,4
version	1021	1122
platform	iOS	iOS
app_version	18400	18400

Table 3: Use case data table example.

2.3.2. Cleaning

Before using the data for any analysis, I cleaned the data. Initially there were 10 active users. However, two of them were co-workers who tested the app under real conditions. So, they created their own account and started submitting points via the app. As they knew exactly what was expected the submissions were highly accurate. This would have influenced the statistics towards a more positive result; therefore, their submissions were not included in the analysis. Additionally, one user signed up twice with a different username. So, I had to sum up all submissions from both usernames under one user. After the data cleaning there were 3653 records from 7 users left.

2.3.3. Validation

GISCloud offers not only an app, where the users can create submissions, but also a webservice, where all the submissions can be visualized and reviewed. The submissions are geolocated on a map, and with clicking the points you will get all the related information. In the next step the data was compared with satellite images or street view images. There are some issues when using this method for QA: first, sometimes there is no Street View of the specific area. Second, even if there is a Street View, it might not be useful because it could be too old, or the desired spot is covered e.g. by a construction site or trees. Alternatively, the submissions where crosschecked on Mapillary (www.mapillary.com/app) or Bing Maps (https://www.bing.com/maps/). This method is very time consuming and not very satisfying as it is dependent on the quality and the currentness of Google Street View, Bing Maps and Mapillary. The total time expanse of only crosschecking the submitted data adds up to approximately 50 working hours. This is also the reason why I decided to write this thesis about QA of crowdsourced data. This approach was very time consuming and was not scalable as it was very cost intensive. Together with the communication and pay out with the users, it was meant to be a full-time job for at least one person.

Fig. 7: Example of a FRP in Munich, comparing a) Google Street View with b) Mapillary imaginary. Captured: 2020-07-01

3. Quality of Spatial Data

The quality of spatial data can differ in many aspects. It is easy to create spatial data, but it might be on interest, when spatial data is meant to have good quality.

"All geospatial data are then, at different levels, imprecise, inaccurate, out of date, incomplete...". (Devillers & Jeansoulin, 2006)

Based on this citation, geospatial data can have a certain quality, but if this quality is rated as good or not sufficient, needs to be determined by the usage of the data. But why is the quality of spatial data so important?

"Users must always conduct a systematic investigation of whether a given data set is sufficiently accurate for a given use." (Elwood, Goodchild, & Sui, 2012)

This means that data is always subjectively high or low in quality, depending what further purpose the data has. Data might be useful for a project but for another project it could be worthless as more precision or more details would be necessary.
To assure the quality of spatial data is an important thing as the trust in the accuracy of the data increases also its use for specific purposes. And vice versa would the unknown quality status of the data make it useless. Here I want to have a closer look on how the quality of spatial data can be measured and in the next step, what additional elements need to be concerned when talking about QA for crowdsourced geodata.

3.1. General Quality Concepts of Spatial Data

When it comes to working with geodata, quality plays a major role. This is true for the production, assessment or exchange of the data. The *International Organization for Standardization* (ISO) defines several guidelines and principles concerning geodata. The ISO 19157 contains quality measurements and possible procedures for quality evaluation of digital spatial data and is the revised version of the ISO 19113:2002 which is also cited several times in this thesis and replaces the ISO 19138:2006 and ISO 19114:2003 as well. (ISO 19157:2013, 2020)
The ISO 19157 standard defines as the quality criteria for spatial data (ISO 19157:2013, 2020):

- Completeness
- logical consistency
- positional accuracy
- temporal quality
- thematic accuracy
- usability

The presence and absence of data or features and attributes and relationships is described within the **completeness** criteria. By keeping the consistency of the data structure, the **logical consistency** is in most cases easily obtained by an appropriate acquisition environment. The **positional accuracy** is influenced by two main parameters. The first thing is how the position was collected. For example, if the location was tracked with a GPS device, if the location was manually pinned on a map or picked in an aerial image. Second, it is also depending on the wanted type of location information. The possibility of making positional mistakes varies depending on the vector type. The question is, if it is asked for a point, GPS coordinate or an object within a specific area.

When it comes to **temporal quality**, the accuracy can vary a lot depending on the usage of the objects. Sometimes it is necessary to record the exact timestamp or even real-time integration (e.g. traffic applications) is necessary, however in some cases it is enough to know within which year the data was collected. The **thematic accuracy** refers to the accurate classification to the attributes of objects which don't have a quantitative character. Here, individual perception can lead to diversity in the classification. For example, can a person rate a street as a main street, while the same person rates another street with the same characteristics as a minor street. This can happen without intention; however, the person thinks that these streets belong to different categories due to lack in experience or subjective interpretation.

Last of all, I want to describe the **usability** which includes all the before mentioned dimensions, but in concern of how they fit the users need. It is depending on the specific needs of the consumer and can therefore vary from use case to use case.

Similar to the ISO 19157 standard, already in the 1980s, the US Government defined five fundamental dimensions of geospatial data (Goodchild & Li, 2012):

- Position Accuracy
- Attribute Accuracy
- Logical consistency
- Completeness
- Lineage

The **position accuracy** shows how precisely the coordinates measurement was done. **Attribute accuracy** measures the information of the data besides the positional information. The internal relations of the database are tested by the **logical consistency**. **Completeness** shows how complete the collection of objects is. **Lineage** includes the metadata about where the data comes from and how it was derived. (Li, Zhang, & Wu, 2012).

Besides those parameters, other dimensions are meanwhile discussed and added to this list. This includes temporal and semantic accuracy. (Goodchild & Li, 2012)

Not only the US government created a recommendation for spatial data QA, also the **European Union (EU)** created an initiative for a common spatial information infrastructure among the European community – INSPIRE (Infrastructure for Spatial Information in the European Community). The aim of INSPIRE is to establish a network of spatial applications from and for all EU-countries. The interoperability of the network should be finalized in October 2020. (INSPIRE Austria, 2020)

Within the INSPIRE project also quality standards were defined which are mostly based on the ISO 19157. The data quality elements are defined in the technical guidelines of the corresponding themes. Spatial data elements are categorized into 34 groups, each with its own technical guideline and data quality elements. I decided to have a closer look in the data specification on *Transport Networks*, as "*all topographic features related to transport by road, rail, water and air"* belong to this group, and in my opinion, this addresses also FRPs. (INSPIRE - Transport Networks, 2020)

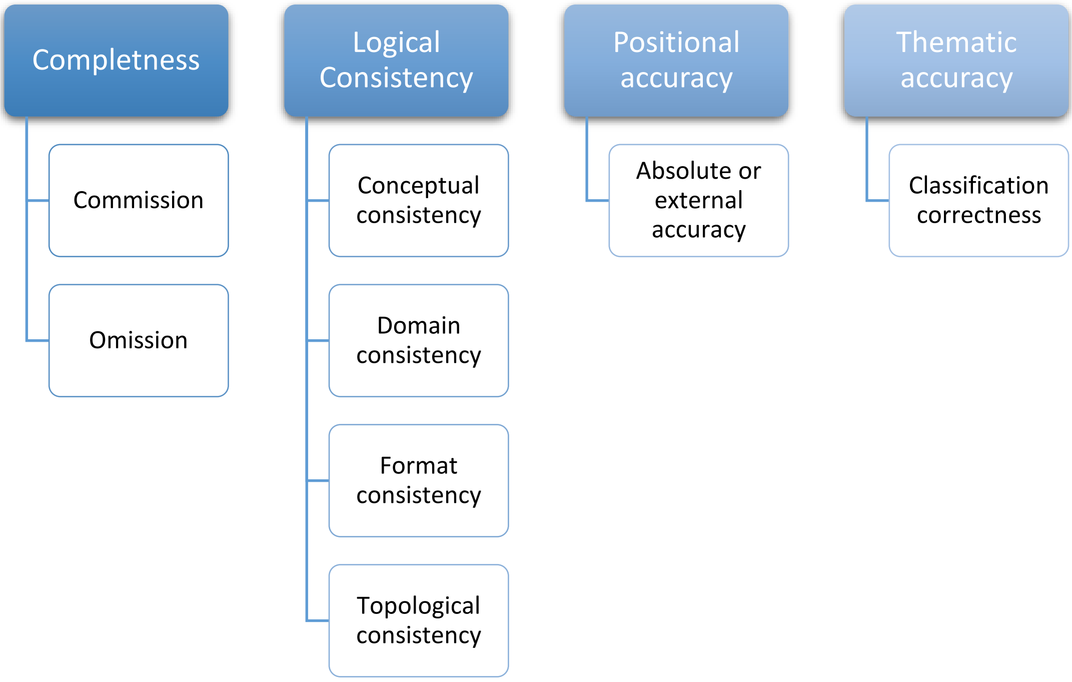


Fig. 8: Data quality items of Transport Networks in INSPIRE. Authors' Illustration after: (INSPIRE - Transport Networks, 2020)

As shown in Fig. 8, INSPIRE splits up the data quality into four parts: completeness, logical consistency, positional accuracy and thematic accuracy. When it comes to completeness, INSPIRE differentiates between commission (the excess of items or duplicated items) and the omissions (missing items). The logical consistency consists of the conceptual, the domain, the format and the topological consistency. The conceptual schema should be predefined in the planning phase. Objects which are not compliant to this scheme are then counted as invalid according to the conceptual inconsistency. The consistency of the domain represents the conformity of the value of an object, for example the absence of a value. When objects are in conflict with the physical structure of a dataset, they are not consent to the format consistency. The topological consistency represents invalid overlapping surfaces, missing connections, invalid silver areas, invalid self-intersections and invalid self-overlapping errors within the dataset.

The positional accuracy shows the mean value of positional uncertainties of objects. Positional uncertainties are described as the difference between the measured position and the corresponding true position. Finally, the classification correctness is expressed with the thematic accuracy.

Devillers & Jeansoulin (2006) summed up several of the concepts of quality of spatial data and established a figure which represent the basic idea of data quality. It does not show the dimensions of the quality of data, however it expresses that quality cannot be measured absolutely, it changes by the users' needs.

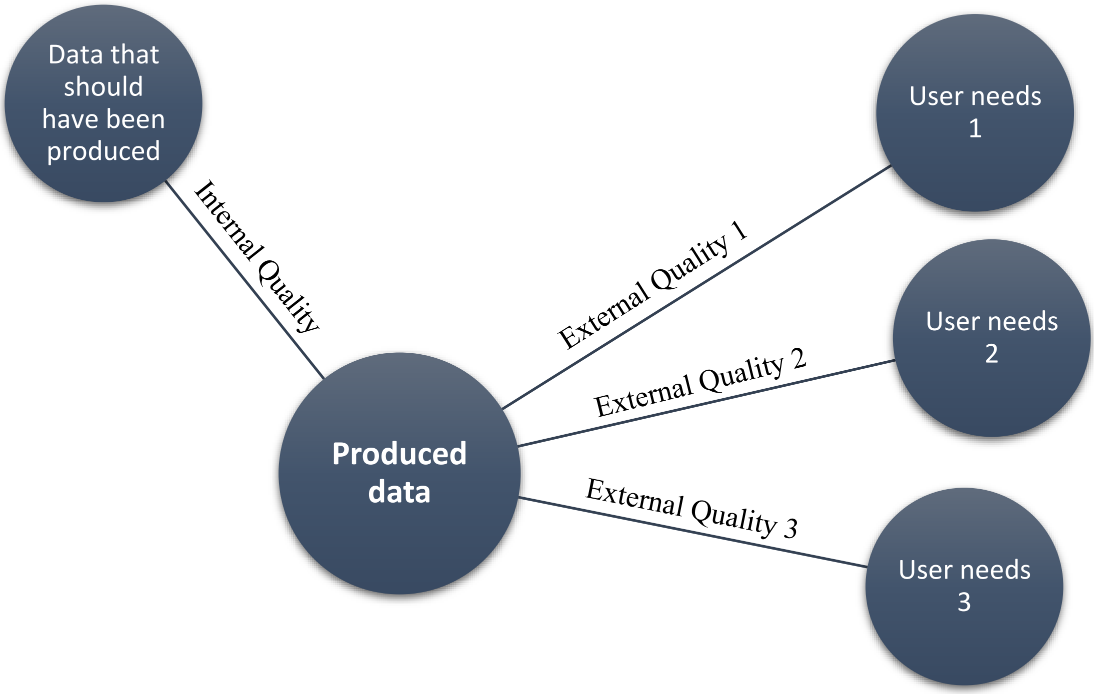


Fig. 9: Concepts of internal and external data quality. Author's illustration, based on Devillers & Jeansoulin, (2006)

It presents especially the concept of internal and external quality. The internal quality refers to how much the created data is similar to the data which should have been produced. The external quality refers to how much the created data can satisfy the corresponding user needs. This is called the fitness-for-use. As there might be more than one user, there are also different user needs and so, there might be also different kinds of external quality. The question is, if the data is good enough for the desired purpose.

"*Therefore, the concept of external quality implies that quality is not absolute, and the same product can be of different quality to different users.*" (Devillers & Jeansoulin, 2006)

After explaining different approaches of quality concepts, there are a lot of intersections between them. All the mentioned quality concepts will be faced on page 19 where all concepts of traditional geodata and quality indicators of crowdsourced geodata are summed up compared.

3.2. Quality Indicators of Crowdsourced Spatial Data

The ISO 19157 represents a reference standard for GI; however, the quality of crowdsourced GI presents some additional aspects which require new indicators to be standardized and evaluated. (Exel, Dias, & Fruijtier, 2010)

The quality of crowdsourced data has been widely discussed in the last years. To evaluate crowdsourced GI, data has often been compared to already existing datasets like Haklay (2010) did with OSM data in Great Britain or Girres & Touya (2010) did with OSM data and french datasets. Goodchild & Li (2012) mention that such studies give "*useful insight into the accuracy of VGI, but only indirectly help to identify mechanisms for assuring the and improving the quality"*. This is also true for areas, where crowdsourced GI happens already for a longer period and the crowdsourced GI dataset is more complete, accurate and dense than the corresponding authoritive dataset. (Vandecasteele & Devillers, 2015)
In this chapter I want to emphasize on additional aspects of the quality of crowdsourced spatial data compared to conventional spatial data. Of course, the data itself has the same quality measurements as mentioned in the previous chapter. Here, it should be clarified what additional indicators there are to measure the quality of crowdsourced spatial data.
Exel et al (2010) defined the quality indicators for spatial data as crowd quality, in my opinion this term expresses the reality very precisely, as the quality of crowdsourced geodata is not only determined by the data itself, but also by its contributor.
According to Devillers & Jeansoulin (2006) there are two main categories of quality – internal quality and external quality - as discussed in the previous chapter. However, there is one important part missing when it comes to crowdsourced geodata: the trustworthiness of information (Criscuolo et al 2016). This means when working with crowdsourced spatial data, not only the quality of the data itself must be analysed but also the production of the data as well as for which purpose it was created. As those factors cannot lead to an accurate measurement, it presents an approximation of the quality which is also referred to as proxy measure for data quality (Dai, 2008). One of these proxy measures, or so-called indicators, is the reliability of a user. This trustworthiness is a powerful and often used indicator within crowdsourcing systems (Fogliaroni, D'Antonio, & Clementini, 2018).

Moreover, Antoniou & Skopeliti (2015) define not only measures for crowdsourced spatial data which can be compared to authoritive data, but also indicators which help to evaluate crowdsourced data when the access to authoritive data is not possible. Therefore, not only the data itself, also the way how and by whom the data was collected was analysed. This means besides the quality parameter discussed in Chapter 3.1, these indicators show how likely a submission is to be correct. Antoniou & Skopeliti (2015) mention four indicators:

- Data Indicators
- Demographic Indicators
- Socio-economic Indicators
- Contributors' Indicators

Data Indicators show the quality by analysing the data only without concerning any additional parameter. This can happen based on the density, attributes or history of the data. Additionally, I would add the way the data is collected to this group as the positional accuracy is highly depending on how the data was gathered. This includes the GPS signal preciseness or if the data was manually placed, the personal preciseness of the contributor.

The second group of indicators are the **demographic indicators** which are generated by relations between the data and the demographic environment of an area. Haklay (2010) found out that there is a correlation between the population density and the number of contributions and consequently with the data completeness.

Socio-economic indicators include factors as population age and income within the region.

The last group are **contributors' indicators**, which include the experience and local knowledge of the contributor itself. Neis & Zipf (2012) defined by four different contributors' groups based on the number of created nodes:

- "Senior Mappers" (contributors with 1 000 and more created Nodes),
- "Junior Mappers" (contributors with at least 10 and less than 1 000 created Nodes),
- "Nonrecurring Mappers" (contributors with less than 10 created Nodes)
- contributors with no edits.

The question is, if a high number of created nodes comes along with high quality. Besides the number of nodes created, the contributors' quality can also be determined by other activities of the users. The amount of responses in community forums or the amount of time the user is already registered can also be possible measurements. (Exel, Dias, & Fruijtier, 2010)

There had been several studies among the trustworthiness of the contributors. It seems like age and education (Delaney, 2007), background and experience (Galloway, 2006), how well the user is trained and if there is a supervisor locally (Fitzpatrick, 2009) have impact on the trustworthiness of the user.

Comparing these indicators with the categories of spatial data by Criscuolo (2016) in Fig. 10, it seems like there is of course a lot of overlapping items between the quality of spatial data and the quality of crowdsourced spatial data. It is not surprising that the data indicators and intrinsic indicators are describing the same elements. Additionally, also the content of extrinsic quality has a lot of intersections with other mentioned factors.

Based on the ISO 19113 (note: meanwhile revised as ISO 19157) Criscuolo et al (2016) tried to sum up the measurements into the following three main categories:

- intrinsic quality
- extrinsic quality
- pragmatic quality

The intrinsic quality describes the quality of the data itself and its informative content. The extrinsic quality is depending on external parameter which are independent from the data itself. The pragmatic quality describes how useful the data is for specific needs. This means, the quality of a dataset can only be defined when concerning the purpose of the data. In some cases, less detailed metadata is sufficient for a specific use, therefore the quality is relatively rated higher. The same dataset might be rated lower when a higher accuracy is needed. For example: the dataset which is described in Chapter 2 has a positional accuracy within several meters which is good enough for its purpose to show on a map where fire rescue paths are. If the task would be only to count all the fire rescue paths within a specific area, for example the outer circle of Munich, the positional accuracy would be highly exaggerated and less positional accuracy would be enough. Absolute statements about the quality can therefore only be done solely for the intrinsic and the extrinsic quality without concerning the pragmatic quality.

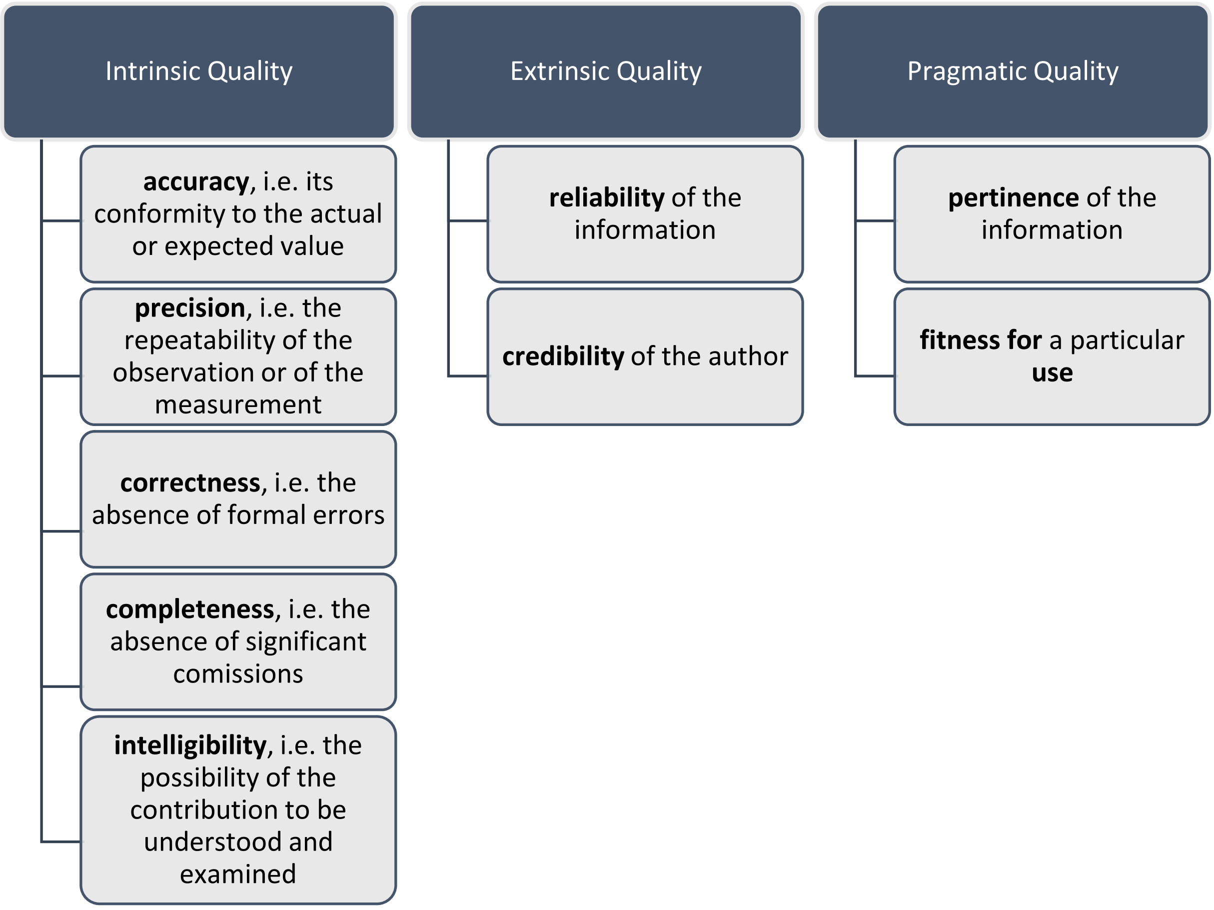


Fig. 10: Quality Categories of crowdsourced spatial data. Author's illustration, based on Criscuolo et al (2016)

In Fig. 10 the three categories are described by several examples and properties which fit in the corresponding category. Criscuolo et al (2016) sketched this figure to give the reader an idea which elements can influence the quality of data. To evaluate the quality of a dataset, these elements need to be defined first. However, they also mention, that this collection of factors always needs to be adapted for specific applications and needs. This is also what I have done in the following chapter.

3.3. Proposal of a Quality Concept

There have been a lot of discussion about the quality of spatial data and the quality of crowdsourced spatial data in the previous chapters. Here, I want to give my own proposal of quality measurements and indicators for crowdsourced spatial data. By analysing the different approaches of categorisations of quality measurements, there are several items overlapping in most of the literature.

In the following table I tried to sum up all the concepts of quality indicators described in Chapter 3.1 and chapter 3.2 While there are all concepts side by side it shows up that there are a lot of overlapping criteria in the different theories. The internal, intrinsic quality or data indicators have all in common that they are focusing on the data or the dataset itself without concerning anything about how, by whom or for which purpose the data was collected. For the further use in this thesis I will call these criteria **intrinsic quality** criteria, as it fits its meaning most accurate in my opinion. According to the theories, this category includes attribute accuracy, position accuracy, completeness, logical consistency, lineage and temporal accuracy.

The second group shows how the dataset meets specific needs of the usage. Although there are many different names used, all of them want to answer the question, if the dataset can fulfil the wanted criteria of the use case. For this thesis I want to sum them up with the **usability** of the data.

The third part of the quality concepts refers to everything which concerns the user who created the content. This question was not important for conventional data sources as the trust in authoritive geodata companies was always at a very high level and the quality of the data rarely questioned. Thus, it is not surprising that these categories did not show up in quality concepts concerning conventional sourced spatial data. Here, I want to sum those elements up into **contributor quality** – as they are all referring to the contributor's characteristics and using 'extrinsic' or 'external' as a term could be confusing for the reader. The credibility of the user was already discussed in Criscuolo et al (2016)

Overview of (crowdsourced) Spatial Quality Concepts											
Traditional data quality measurements											
Devillers & Jeansoulin 2006	Internal Quality						External Quality				
Goodchild& Li 2012	Position Accuracy	Attribute Accuracy	Logical Consistency	Completeness	Lineage						
Barron 2014 / ISO 19156	Position Accuracy	Thematic Accuracy	Logical Consistency	Completeness		Temporal Quality	Usability				
INSPIRE (Transport Networks)	Position Accuracy	Thematic Accuracy	Logical Consistency	Completeness							
Crowd quality measurements and indicators											
Antoniou & Skopeliti 2015	Data Indicators								Demo-graphic Indicators	Socio-economic Indicators	Contributors' Indicators
ISO 19113	Intrinsic Quality						Pragmatic Quality		Extrinsic Quality		
Crisculo 2016	Accuracy	Precision	Correctness	Completeness	Intelligibility		Pertinence	Fitness for use	Reliability		Credibility
Koubek 2020	Intrinsic Quality						Usability		Contributors Indicator		
	Position Accuracy	Attribute Accuracy	Logical Consistency	Completeness	Lineage	Temporal Accuracy			Motivation		Credibility

Table 4: (Crowdsourced) spatial data quality concepts.

In Fig. 11 I established my own indicator tree based on the various concepts mentioned in the previous pages. This tree includes all elements that are part of the quality of the data of the FRPs use case.

There is one aspect which was not discussed so far in this paper, the **motivation** of the crowd user.

For myself a very important question is the motivation behind the data. As it seems that it is not covered by most of the analysed sources which I mentioned before in this chapter.

Antoniou & Skopeliti (2015) already mentioned, an analysis of the motivation of the contributor could give more insight into the data and its quality as well. This is also why I differentiate between VGI and commercial crowdsourcing as I explain in chapter 4.1. I think the question about why someone is contributing to crowdsourced GI plays a very important role.

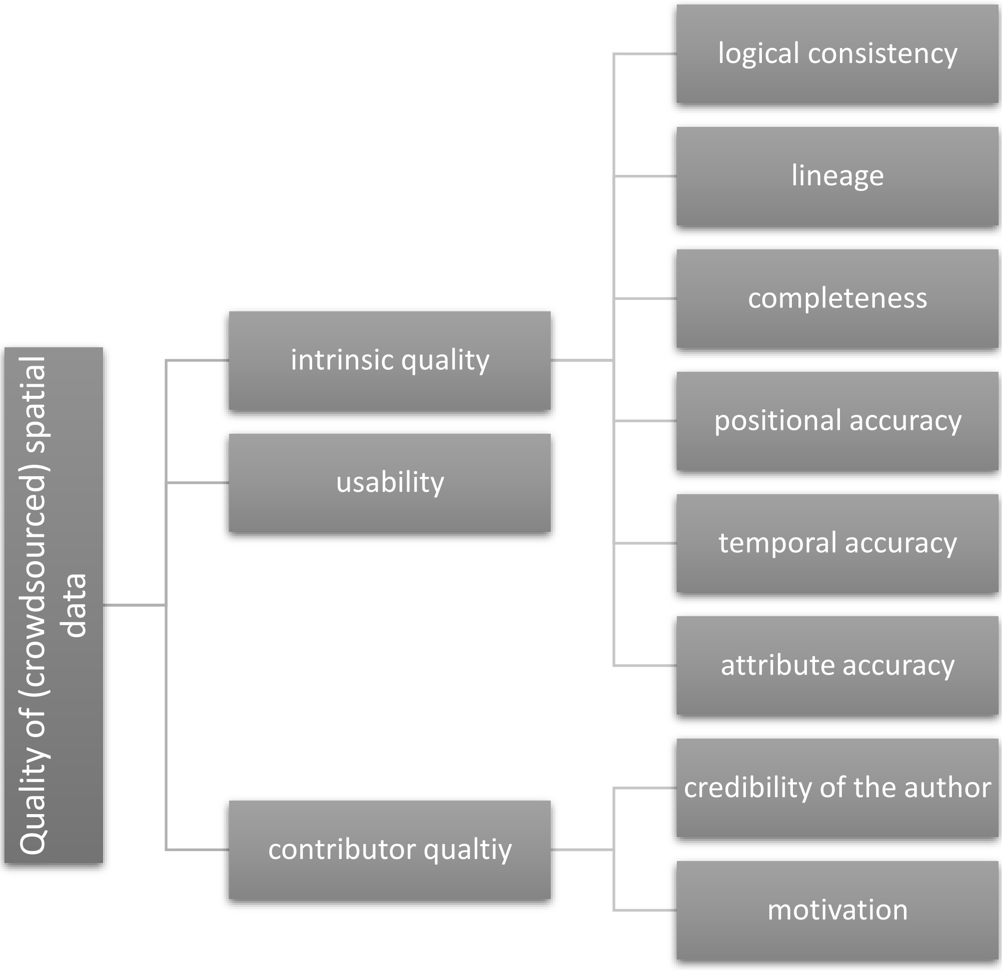


Fig. 11: Quality concept adapted for the FRPs use case.

4. Crowdsourcing and Spatial Data

Spatial data has been gathered traditionally by professionals who were working for national mapping agencies or other institutions which focused on spatial data. Quality standards where defined and is was expected that all cartographers and mapping agencies are fulfilling these standards and obtain the minimum quality. Non-authoritive data collection processes like crowdsourcing were most of the time not taken seriously as the scientific legitimacy was missing. (Sieber & Haklay, 2015)
Due to the evolving phenomena of Web 2.0 in the 2000s, Volunteered Geographic Information (VGI) could be formed out of the development of different web services which are based on the concept of user created content. (Goodchild M. , 2007)

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. [..] I term this volunteered geographic information (VGI), a special case of the more general Web phenomenon of user-generated content. (Goodchild M. , 2007)

In scientific context, the questions of social changes between traditional user and producer arose. Or, as Budhathoki et al (2008) expressed it as the **Produser.** People are no longer only the consumers of static information on the world wide web (WWW), nowadays they are very likely to share their own knowledge and contribute actively on the content of the internet. (Barron, Neis, & ZIpf, 2014)
Crowdsourced geodata is also part of the user generated content (UCG) generation and has become a remarkable new source of geographic information in the last years. Wikimapia, Flickr, Geonames and the most well-known OpenStreetMap are just some well-established examples which built their services around crowdsourced (geo)data.

Citizens are using handheld devices to collect geographic information and contribute it to crowd-sourced data sets, using Web-based mapping interfaces to mark and annotate geographic features, or adding geographic location to photographs, text, and other media shared online. These phenomena, which generate what we refer to collectively as volunteered geographic information (VGI), represent a paradigmatic shift in how geographic information is created and shared and by whom, as well as its content and characteristics. (Elwood, Goodchild, & Sui, 2012)

An enormous amount of data got available via this approach which could be used for all kind of purposes – and I am confident that a lot more will be available. It is not only of interest for GIScientists but also for the field of human or physical geography. Additionally, advanced technology, when it comes to web services and mobile devices, are helping people to contribute without considering their education or knowledge.

VGI, also referred to as crowd-sourced geodata, is defined as the collaborative acquisition of geographical information and local knowledge by volunteers, amateurs or professionals. (Goodchild M. , 2007)

Budhathoki (2010) mentioned in his dissertation several aspects why someone would participate in a VGI project. The motivation varies from self-representation in the WWW, fun to technology interest, or – as in the FRPs use case – financial interest. The difference between traditional GI and VGI has several aspects. Elwood et al (2012) named the content of the information, the technologies for the acquisition, quality issues and the social process around VGI as variations concerning VGI and traditional spatial data. On the following pages I want to focus on the aspects of the collection of geographic information and how CSGI differs from traditional GI. In the next chapter I want to show that it also bears some difficulties which should be considered when working with crowdsourced geodata. In the last part of chapter 3 I want to give examples of crowdsourcing and how QA can be dealt with.

4.1. Aspects of Geographic Information Collection

Within this thesis I work with crowdsourced GI data, which was not collected voluntarily, but a monetary compensation was given for the collected data. Most of the research which covers crowdsourcing is limited to volunteered geographic information. As the contributors of the FRPs project got paid, the term 'voluntarily' does not fit here. Therefore, it is important for me to point out, that the terms 'crowdsourced spatial data' and 'volunteered geographic information' cannot be used interchangeably. Goodchild & Li (2012) write in their paper that VGI is

"a version of crowdsourcing in which members of the general public create and contribute georeferenced facts about the Earth's surface and near-surface to websites where the facts are synthesized into databases."

Thus, VGI is meant to be a part of crowdsourced geoinformation. Criscuolo et al (2016) mention that there are two kind of contributors of CSGI: the contributors to scientific initiatives and also VGI projects and unaware contributors on social networks when GI is retrieved from texts, geo-tagged pictures and points of interest (POI). Obviously the FRPs project does not fit in one of those two groups. VGI contains the term "volunteered" that refers to the consciousness of the publication of the information without a renumeration in a financial sense. It is often not clear if contributors got rewarded for their contributions, because the benefits are not always directly connected to the project itself. Also, contributors offering data are not always aware of the geographic context the information can be used in, thus they are not conscious about providing VGI. Criscuolo et al (2013) explained the more relevant term "incidental data". They defined three types of data resources based on the parameter expertise, intention and reward:

- Official data: Experts gather intentional data and get rewarded
- Volunteered data: Data is gathered intentional without rewards.
- Incidental data:
    - Non-experts gather intentional data with rewards.
    - Unintentional.

Based on this theory the data of the use case in this thesis refers clearly to the third group and is defined as incidental data where non-experts gather intentional data with rewards. Also, Heipke (2010) defined contributors that work for a monetary payment as *Mechanical turks* and refers with that term to the Amazon crowdsourcing marketplace, the Amazon Mechanical Turk (see chapter 4.3). There people can contribute to certain tasks and get a financial reward. (Mechanical Turk, 2020)

Regarding this, I will not use the term VGI in connection to this use case, instead I will stick to the term crowdsourced geodata, crowdsourced spatial data or crowdsourced GI.

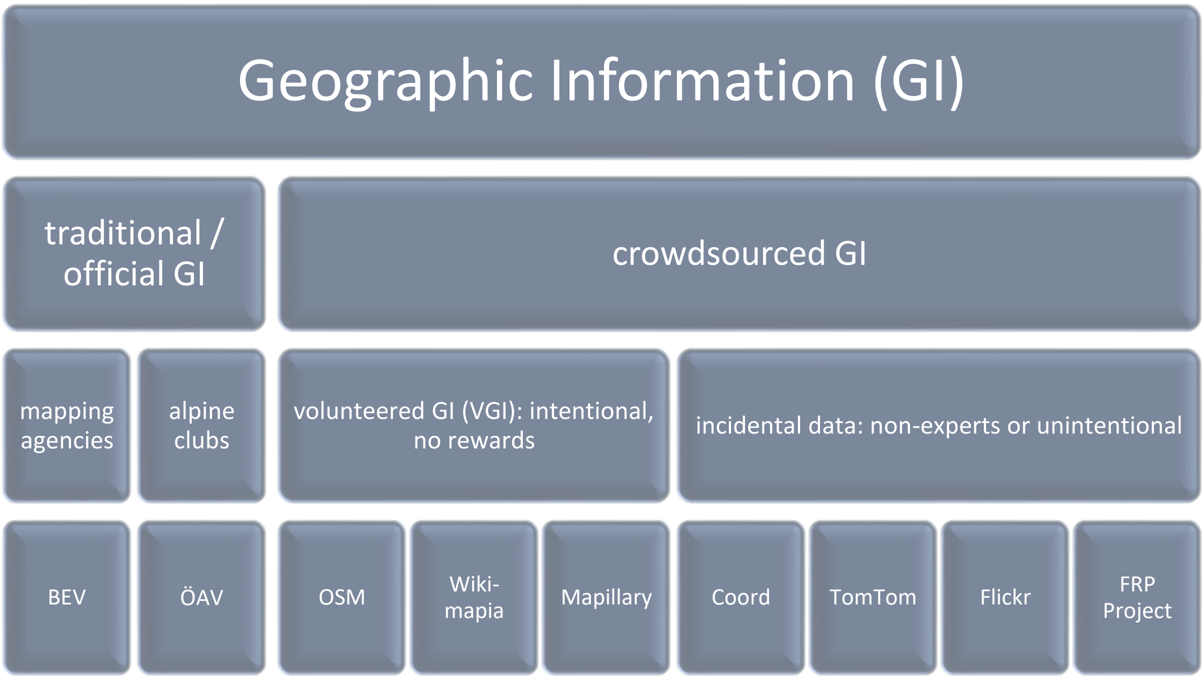


Fig. 12: Different aspects of collecting spatial data.

In Fig. 12 I summed up the different aspects of GI when it comes to the collection of the data. Traditional GI is collected by experts who get paid for the collection. GI experts are well educated and have a common sense of the quality standards of GI. Therefore, quality assurance takes only place intrinsic. The trustworthiness of the contributor is not doubted. Institutions belonging to this group are often national mapping agencies like the Bundesamt für Eich- und Vermessungswesen in Austria or the mapping agencies with a long tradition in map making like alpine clubs as the Alpenverein (ÖAV). CSGI can be split up into VGI and incidental data. VGI platforms are for example OpenStreetMap, Wikmapia or Mapillary. Incidental data can be intentional be non-experts, like it happens in the FRPs project, but also at TomTom or Coord (see chapter 4.3). Incidental data can also be unintentional, so the contributor did not submit data with the consciousness that the geoinformation can be used in another context. This can happen when geotagged information on social media is analysed, as it happened with Flickr data, tweets on Twitter or with YouTube videos. (Andrienko & Andrienko, 2016), (Borge et al 2018)

4.2. Problems connected to Crowdsourcing and Spatial Data

Crowdsourcing in general was recognized by several companies as a mass collaboration system which allows them to quickly and inexpensively accomplish tasks. However, crowdsourcing has some aspects which should be concerned before companies decide to outsource their work.

Criscuolo et al (2016) evaluated three main problems when it comes to crowdsourcing and spatial data:

- Metadata
- Processing data
- Trustworthiness

Metadata can be a big problem when using crowdsourced geodata which was not established for a specific use case, but data was taken that is already existing within the WWW. For example, 15% of all social media activities are georeferenced and can therefore be attractive be GIS experts. However, only having a spatial context, does not qualify data to be reusable in a geographic context. Traditional data offers information about the domain in which the data should be used (spatial resolution, temporal aspect, etc…). As crowdsourced data can lack in those information, parts or whole procedures of quality assurance cannot take place.

Second, processing crowdsourced data can be a very difficult task. The reason is not only the missing metadata, as mentioned above. Merging heterogenous data can lead to a challenging processing tasks due to different datatypes, reference systems, accuracy etc…

The biggest problem that crowdsourced geodata faces could also be the reason why it is not overtaking the traditional way of spatial data collection - the quality evaluation. Traditionally spatial data was gathered by national mapping agencies or commercial companies. To maintain their reputation of good quality data and to keep up business, it was essential that the provided data had continuously high quality. Meanwhile the procedure of the collection of crowdsourcing data is not that much divergent from a professional approach. Mapping agencies also often use handheld devices, with similar accuracy as a modern smartphone in positioning. It is the knowledge and experience of a professional cartographer which also influences the quality of the data and the trustworthiness from potential customers.

However, other parameters got high in interest. Antoniou & Skopeliti (2015) mention local knowledge and currentness as some of the reasons why crowdsourced geodata got attention besides conventional sources. Already in 2007 Goodchild pointed out the difference between conventional geographic information and crowdsourced geodata. He insists on emphasizing on the role of crowdsourced geodata as augmenting our knowledge of the geographic world on several levels.

The main character of crowdsourcing is that many people are contributing to a dataset. This leads to a variety in quality. Senaratne et al (2017) mention as reasons for this

- Heterogeneous contributors
- Different technologies
- Different tools
- Level of details and precision
- Heterogeneous purposes
- Lack of gatekeeper

Geographic Information which is created from non-experts or is coming from non-traditional sources comes along with problems which should be concerned when working with such data.

As a consequence, the contributor itself arose to another quality parameter additionally to the other aspects of quality of traditional collected spatial data.

These problems that are described by Criscuolo and Senaratne are especially true when it comes to unintentional data. (Criscuolo, et al., 2016) As the dataset of the FRPs was not created unintentional but intentional by non-experts, the question is how those problem occur in the project. Having a predefined technical environment, the metadata problem and the processing problem can be eliminated. What is left, is the question about the contributors' quality indicators.

Having these aspects in mind, crowdsourcing can be very powerful, when well established methods help to overcome the mentioned struggles.

4.3. Examples

Crowdsourcing and spatial data are well related in GI scientists' minds, since at least the rise of Open Street Map. However, there are also other examples which use the power of crowdsourcing in a different manner. Here I want to give some examples of how crowdsourcing can be combined with spatial data and what methods are used to verify the data collected by individuals.

There are numerous examples which I cannot all cover here. Therefore, I decided to take examples which are different in the way they are using crowdsourcing so a big part of the spectrum can be covered. Also, I want to have examples for financial and for non-financial remuneration. As it is interesting to see how to deal with real-time information, I wanted to have an example with that setup as well. For the FRPs project the analysis of the crowd itself was very useful, therefore I wanted to have an example which has an established crowd, and an example which still needs to establish a crowd and can therefore not use QA tools of the crowd itself. I wanted to have examples which use intentional data only as I only want to have examples where the contributors can be concerned as well in the validation process. To have a relation to my project, I decided to take examples which are GIS-related only, with one exception which is Mturk. I want to mention here, that the list is of course not complete, however it covers some of the most well-known applications.

	Financial remuneration	*Real-time integration*	*Established crowd*	*Intentional contribution*	*GIS related*
Coord *www.coord.com*	X			X	X
Mapillary *www.mapillary.com*		X	X	X	X
Mturk *www.mturk.com*	X		X	X	
OSM *www.openstreetmap.org*			X	X	X
Streetspotr *www.streetspotr.com*	X		X	X	
TomTom *www.tomtom.com*		X	X		X
Waze *www.waze.com*		X	X	X	X
Wikimapia *www.wikimapia.org*		X	X	X	X
Wikipedia *www.wikipedia.org*		X	X	X	

Table 5: Examples of crowdsourcing applications.

Coord is a company focusing on curb information, where the collection process has similarities to the collection of the framework of this thesis. I decided to have a closer look at Coord as it has the most in common with Parkbob when it comes to the final product of the data. Also, it does not have a pool of crowd users they can approach, as they use always small areas and the data collection is depending on their clients, therefore, crowd must be newly established for every project.

Mapillary is an open source platform for raster data in sense of georeferenced street pictures. Mapillary fulfils the same criteria as Waze, therefore I only wanted to keep one of those two in the list and due to the more diverse functionalities I decided to go for Waze.

Mturk is not only geo-related but claims to be a platform for crowdsourcing in general. Also, **streetspotr** is similar to Mturk, however it is more focused on mini jobs on the go, so the users do small jobs while they are in their normal environment and can decide on their own via an app, when a job pops up nearby, if they go for it or not. This approach is very similar to the FRPs approach. Both Mturk and streetspotr are not applications to collect data, but to find workers for their clients offering mini jobs. This can be every kind of job – also data collection - however streetspotr is offering almost only jobs depending on the location of the user, were Mturk also offers jobs that can be fulfilled from everywhere. As Mturk has a broader audience and a wide scope of use cases I decided to include Mturk as well, as it is a platform offering mini jobs, not compulsory with spatial background but it provides an uncomplicated framework to find crowd users for any kind of project.

The VGI platform **OSM** also finds a place in this list as it is mentioned several times in this thesis and plays a very important role in many scientific publications.

TomTom and **Waze** are both specialized in car navigation, the difference is that waze actively includes their users in the process of the creation of the real-time integration of data. TomTom extrudes the information passively from GPS signal and accelerometer in mobile devices and cars. As Waze is free for the consumer and can be installed as an app on every smartphone it is closer to the FRPs project as the more traditional approach from TomTom as software producer for navigation systems.

Wikipedia and its GIS-counterpart **Wikimapia** are both based on the contribution from volunteers. Both are based on the many-eyes principle, so many users see and correct possible mistakes. As Wikimapia fulfils the same criteria as Waze, I decided to prefer Waze due to real time integration and a more modern approach via an app.

4.3.1. Open Street Map

Open Street Map (OSM) is one of the biggest projects when it comes to VGI. The goal is to build a digital world map with VGI from many different contributors, which is free and open for anybody. Contributors collect spatial information using GPS devices or aerial imagery. All information from all users are than merged together into one database. The current number of the nodes, users and and other key elements can be retrieved any time from https://www.openstreetmap.org/stats/data_stats.html as I did in Table 6.

Fig. 13: OpenStreetMap Logo. (OSM Stats, 2020)

OSM Stats, 2020-05-04	
Number of users	6 411 872
Number of uploaded GPS points	781 6197 703
Number of nodes	5 982 817 627
Number of ways	660 880 532
Number of relations	7 792 887

Table 6: OSM Stats. (OSM Stats, 2020)

With more than 6 million users and almost 6 billion nodes this database became a serious source for spatial data worldwide. All the data produced is released with open-content license, so everybody can use it for their own purpose.

As Fig. 14 shows, the number of new data uploads is stable over the last ten years. However, the number of edits which were made, was growing up to 50 000 edits per month in 2017 but then the count continues to be steady on a lower level between 35 000 and 45 000 edits per month. The data quality that is

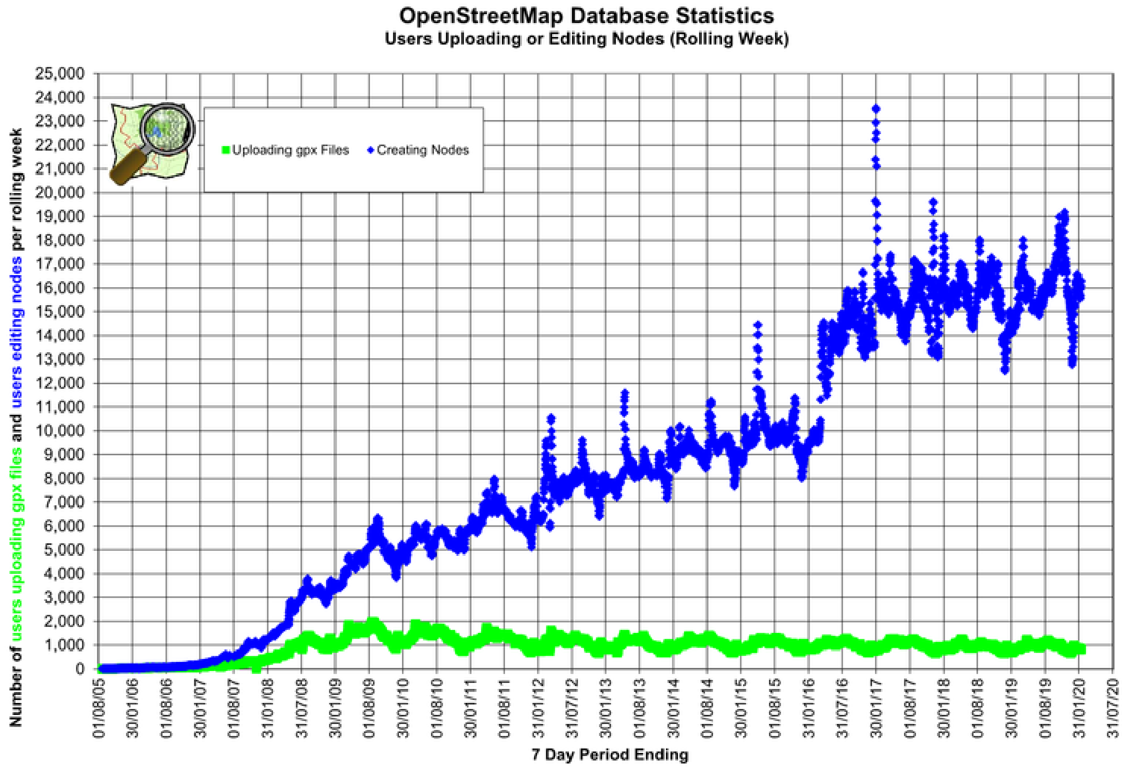


Fig. 14: User upload and editing statistic. (OSM Stats, 2020)

expected is depending on the density of the data. In areas with low coverage the accuracy does not need to be as high as in areas where data has already been refined a lot. In general, OSM says that every edit should make the map overall more accurate than before. The motivation of the contributors has a broad range from self-expression, interest in new technologies or the conviction that geodata should be freely available for everyone. (Neis & Zipf, 2012)

The increasing numbers of edits is also part of the answer of how the QA happens at OSM. OSM has manual tools like “Notes” and “MapDust” which allows users to repost bugs manually and highlight them in the map. On the other side, there are detection tools, which find errors like inaccuracy or sparsely mapped places automatically. These findings can then be checked by a user and if needed fixed. OSM lists 34 tools on their wikipage (OpenStreetMap Wiki, 2018) for automatic error detection, *JOSM Validator*, *Osmose* and *Relation Analyzer* – to name some of them. All these tools should help other users to find mistakes and to fix them. To avoid wrong submissions, OSM established overtime a very detailed beginners guide which helps crowd users to contribute wrong or inaccurate data. However, the whole system is based on the approach of many eyes and many submissions.

Apart from that, there are some automatic approaches that have tag recommendations for the contributors or take over the whole tagging process. As most mistakes in the attributes come from misspelling or wrong classification due to the lack of regulation and standardized classifications. (Senaratne, et al, 2017)

Besides the 4-eyes principle which is useful for attribute and positional accuracy, OSM recommends different approaches of how the completeness of a specific area can be measured. It makes a difference if the method is intrinsic, so it does not need a reference dataset, or extrinsic which compares the OSM data with data from a different source. I am going to explain these methods in detail in chapter 5.5 but want to give here an overview:

- Extrinsic Methods
    - One-to-one comparison of undifferentiated features.
    - One-to-one comparison of matched features
    - Population-based estimation
- Intrinsic Methods
    - Cross-location comparison
    - Asymptote representation
    - Number of features per area

4.3.2. Coord

Fig. 15: Coord Logo. (Coord, 2020)

Coord is a US company which focused on collecting data of curb spaces. They established a software which helps interested people to capture and update several aspects of curb spaces such as parking information, ramps, signs etc. Coord itself manages the data and offers it via API for free for non-commercial use. Moreover, they offer their surveyor app to urban planners, consultants and public agencies. Those entities can employ people to submit the curb information with this app or ask for volunteered submissions. Coord published a use case from Philadelphia where 101 curb miles where collected within 240 working hours. In this use case, they collected every information twice to determine possible inconsistencies in the dataset and systematically errors. This means also here, the many-eyes approach was applied.

Fig. 16: Coord Surveyor App. (Coord, 2020)

4.3.3. Waze

Fig. 17: Waze logo. (Waze, 2020)

Waze is an application for smartphones which promises to make driving and routing more convenient. It is maybe the most successful application which uses the crowd sourcing approach for traffic. (Lendák, 2016)

The app uses GPS information and the accelerometer to determine the speed of the user and uses this information to predict the driving time for other users. However, beside the passive use of the spatial information, each user is also able to submit active information about the traffic. It is possible to report accidents, hazards on the street, police traps and map issues. This can be very helpful as the app promises to have real-time information to avoid unnecessary delays during driving.

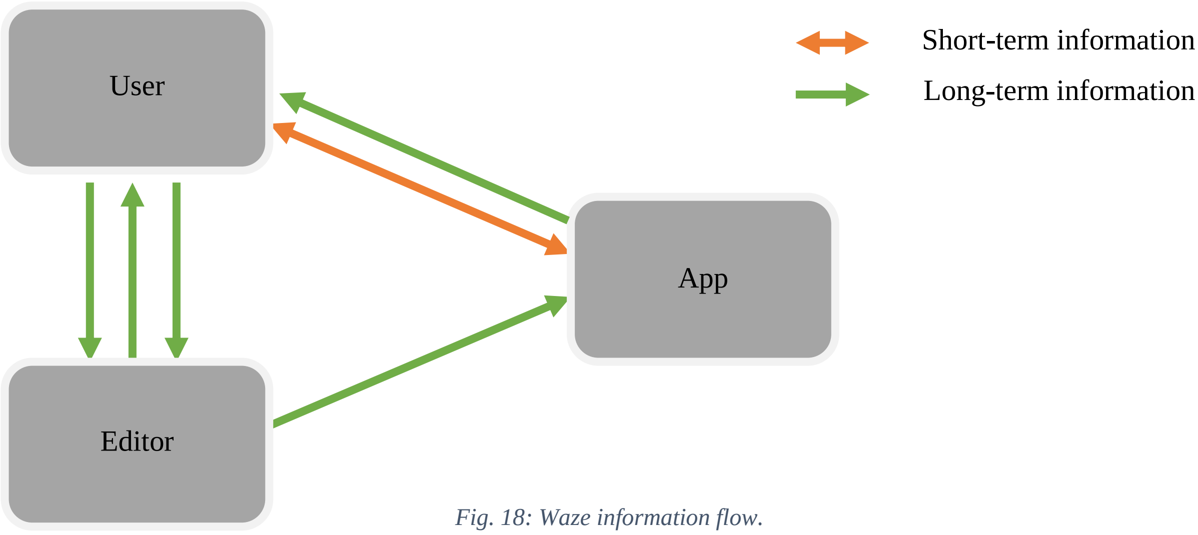


Fig. 18: Waze information flow.

In general, there are two kind of crowdsourced information the app uses: the continuous information and the current information. The first group is the long-term information and includes especially street map information like street location, type of street, number of lanes etc. As it is long lasting information, reports from users will not affect the map immediately. As you can see in Fig. 18 reports are edited within the map editor where users voluntarily review and authorize the submissions. After the new data was reviewed and the data is included in the existing dataset, the map inside the app is updated.

Editor rank table

Rank	Required edits
1	None
2	3 000
3	25 000
4	100 000
5	250 000
6	500 000

Table 7: Waze editor rank table. (Wazeopedia, 2020)

Sometimes the new information is compared to the existing data automatically and based on several parameter the dataset gets updated without been reviewed by a humankind. This can happen when for example a street was rebuild and shifted, the algorithm, based on the GPS signal of several users, changes the position to the new location. When editing reports manually in the map editor, the volunteers get rewarded by editor points.

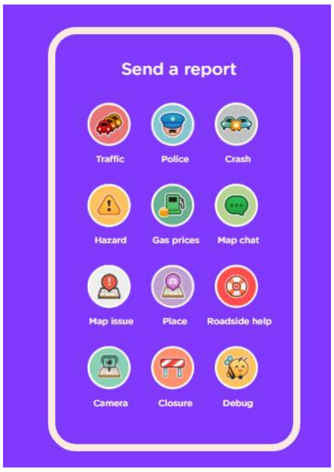


Fig. 19: Screenshot from the Waze App. (Waze, 2020)

Depending on their edits, they obtain a specific rank which comes along with a bigger radius to edit and with superior rights in the editing process. The number of edits that must be taken are shown in Table 7.The other kind of information, the current information can be reported directly via the app by the users. These submissions will be included in the map immediately. Besides the active reporting, also the speed information is collected by the app to predict the drive time for other users. As every user can submit information, fraud is not possible to avoid. However, as the information is updated immediately, the developer claim that users will report if wrong information was included in the system, so the information will be updated with the correct information promptly as a user recognises the fraud. The preconditions to make this system work are of course a certain number of users and that they use the app actively. It is also possible, that Waze could have a powerful solution for parking problems in urban areas where there is no possibility to implement an advanced infrastructure of parking sensors in the streets, as the app can detect by different sensors in the smartphone if a car parks in or out. If the count of users reaches a critical number, there can be a significant economic impact the app obtains. (Lendák, 2016)

4.3.4. MTurk

Fig. 20: Mturk Logo. (Mechanical Turk, 2020)

Amazon Mechanical Turk is a job marketplace that helps companies and individuals to outsource their work. It calls itself a *"crowdsourcing marketplace"* and aims businesses to profit from a global workforce. MTurk is a general job platform which is not limited to spatial information. It advertises crowdsourcing as less time consuming, expensive and difficult to scale, which is exactly the purpose of the project explained in chapter 2. Also, this platform works in first place with a financial remuneration. MTurk is also well used for scientific research. Chandler & Kapelner (2013) created a framework, where they wanted to find out about how they can increase the quality of crowdsourced data via MTurk. So, they wanted to implement additional motivational aspects besides the financial remuneration. They found out, that bringing meaningfulness to a task increases the quality of the results clearly. Also, they mention, that with decreasing the meaningfulness, the quantity remained the same but the quality decreased.

The use case in this paper is more likely to be part of the gig economy-environment – as MTurk - than crowdsourced spatial data-environment. Platforms which are built for gig economy are defined as "*digital, on-demand platforms that enable a flexible work arrangement, influence local entrepreneurial activity."* (Burtch, Carnahan, & Greenwood, 2018)

It gives the employees a very high degree of flexibility and the possibility to decide on their own the working hours and earnings. This also comes along with a big disadvantage: when it comes to crowdsourcing, there is no real employment between the contributors and the employer. In this case, the contributor is an independent contractor who gets paid by submission. Consequently, it is also the contributors' decision how many contributions will be made. This means, the employer cannot influence the users' activity directly.

MTurk is only a platform where employers can find potential crowd users, this means the QA is always on the side of the employer. Still I wanted to include this web service in this list, as it works similar as the FRP project.

5. Methods of Quality Assuring Crowdsourced Spatial Data

In the previous chapters I give an insight in the quality of spatial data and crowdsourced spatial data. The next step is now the quality assurance (QA), which means the practical way the quality of the data is analysed.

"Data quality assurance aims at identifying, correcting and eliminating errors." (Jacobs, 2016)

This means QA is not only the process to find out the potential sources of errors but also to change and update the data to achieve a flawless dataset. QA helps to find errors and depending on what element is concerned, correct or eliminate them. There are already several projects which face the manual validation procedures, also the FRPs project used the manual approach in first place. The problem with it is, that it is very time consuming and it is not possible to scale it up. To be able to prove large amounts of new spatial information through crowdsourcing (semi-) automatic processes have to be developed.

Additionally, a closer look at the dataset's characteristics give an insight in its suitability for the purpose of use. Dickinson (2010) described that not all aspects of the purpose of use can be determined beforehand. Several elements appear during the data collection or even after.

In the following pages I will first explain to approaches of QA categorisation, one after Goodchild & Li (2012) and one after Criscuolo (2016).

After that I will explain existing methods of QA and how they are applied to the FRPs project (chapter 5.4 - 5.8). In the final part of chapter 5 I will sum up the practical usage of the methods. As not all methods fit for the purpose of this project, an overview of all methods which I came across during my research will be listen in Fig. 59 on page 76.

5.1. Approaches of QA for Geodata after Goodchild & Li (2012)

While I did research for this thesis, I came across this paper quite often when it comes to intrinsic methods of QA for crowdsourced geodata. Based on these approaches many works have been created (Senaratne, et al, 2017). Therefore, I want to include it here as well, although I was critical about how it would fit to the project. Goodchild & Li (2012) identified three basic approaches how methods of intrinsic QA can be categorized which are shown in Fig. 21. It seems like this list fits only for VGI platforms as OSM but is not suitable for the use case in this thesis. However, I will describe the categories and actual use them as methods themselves in the further analysis.

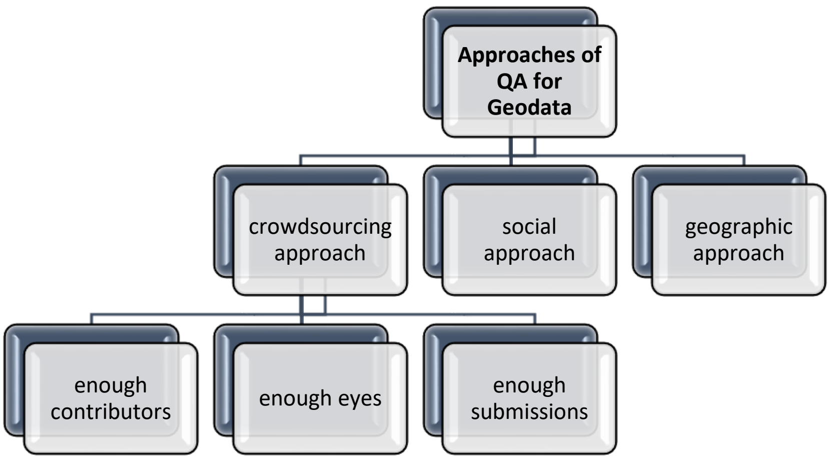


Fig. 21: Approaches of QA for spatial data. Author's illustration based on Goodchild & Li (2012).

5.1.1. Crowdsourcing Approach

Goodchild and Li defined this approach by the number of people contributing to a submission. They assume that many people might be able to solve a problem which one person, even if highly qualified, could not solve. Besides the initial creating of data, the second point is the correction of mistakes made by others. The higher the number of eyes which have the potential to see an error, the likely it is to be found. In the context of software engineering, Raymond (1991) defined the Linus' Law with *"given enough eyes, all bugs are shallow".* This is also the approach of *Waze* (see chapter 4.3.3) to maintain the short-term information in their app. Also, Neis & Zipf (2012) point out the importance of further maintenance of data when analysing the OSM database. The currentness of data should ideally be "*continuously, homogeneously, throughout and is not limited to specific features".* (Barron, Neis, & ZIpf, 2014) Nevertheless, when more users are correcting data, so-called tag wars can be initiated by insisting on different opinions – as happened several times on OSM. (Mooney, 2011)
The third way weights the reliability based on the number of similar contributions. However, there is a maximum of 13 contributions when highest quality is reached and would not raise anymore by increasing the number of contributors. (Haklay, et al, 2010)
In the analysis of the FRPs project only the *enough eyes* approach was used in terms of the 4-eye principle.

5.1.2. Social Approach

This approach is based on the fact that the number of contributions is not equally distributed along the contributors. This is a common phenomenon among social media and in online communities which is known as the "Participation Inequality". (Nielsen, 2006)
The Participation Inequality is also known as the 90-9-1 Rule:

- 90 % are consumers without any contribution
- 9 % give contribution a lower priority but still submit occasionally
- 1 % are the heavy contributors who are responsible for the major part of the content.

In chapter 2.2 the contributors of the use case are analysed based on this theory. However, due to this phenomenon, Goodchild & Li (2012) propose to give rewards to individuals who are making prolific contributions with a higher rank. So, they are allowed to act as moderator or gatekeeper for further submissions from other users. This kind of system is already established among several collaborative projects. In OSM there are for example two kind of users: ordinary users and the Data Working Group who takes care of copyright, disputes and so on. The ordinary user is only allowed to add and edit features. In fact, this approach simulates the structure of a traditional mapping authority as your rank in hierarchy comes along with more responsibility and is honoured with higher reliability and/ or salary. For this project the users were rated due to the quality of their submissions, but no further step – like increase user rights - was taken. The project team thought about the possibility to let superior users control submissions of other users. As the technical setup would have been to extensive, the team decided to stick to the manual reviews by the project team and postponed the idea to a potential future project.

5.1.3. Geographic Approach

The geographic approach is based on the geographic context of the data. Just like Tobler (1970) defined the *First Law of Geography*: "*All things are related, but nearby things are more related than distant things*". Thus, the data is compared with existing datasets to check if the collected data is likely to be true. For this approach, existence of datasets is a precondition, which is unfortunately not always the case. However, it is not meant, that you compare the collected data with data which represents the same information. With this method you see the crowdsourced data in its geographic context. To give an example for the dataset which is described in chapter 2, it wouldn't be logical if a fire rescue path is positioned within a forest or far away from the street. So, the submitted points are compared with an existing dataset – in this case a street dataset or a surface dataset – to check certain logical components of the data.

5.2. Categorization of QA Methods after Criscuolo et al (2016)

Already in the previous chapters it is obvious that the topic of validation of crowdsourced spatial data was already discussed widely in the past 20 years. In this chapter I want to analyse an attempt to put the methods into categories to get a clear overview. I have chosen the most general approaches of defining the categories – from my perspective. At the end of the chapter I want to define my own categories as an aggregation of the different approaches described.

To determine the approaches of quality control, I want to specify the categories in which a method can be put after Criscuolo et al (2016). He describes that the types of quality control can be divided into three groups:

- by time: when does the quality control take place?
- by whom: who does the quality control?
- what happens with the controlled features?

In the following figure I summed up the different approaches:

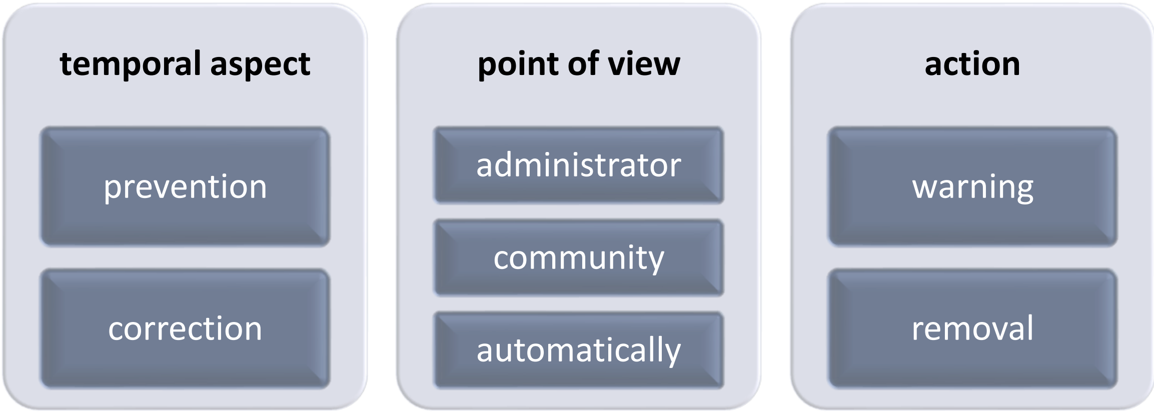


Fig. 22: Aspects of quality control. Author's illustration, based on Criscuolo et al (2016)

5.2.1. Temporal aspect

The temporal aspect of QA can be split up into two categories: the prevention and the correction phase. The prevention approaches' main character is to avoid incorrect data by intrinsic and usability means (compare page 17) by explaining and leading the producer of the data in the right direction before the data is submitted. One way to do so, is by handbooks or manuals where the submission procedure is explained (Crall, 2011). Also, technical support belongs to this category when you give the users no freedom to fill out information by themselves but to provide dropdown lists or autofill options. (Popescu, 2009). Additionally, the pre-election of the users also belongs to this group.

Contrary to the prevention approach, the correction takes place after the data was submitted. This means this category contains all approaches concerning reviews, filters and algorithms. Every approach that improves the submissions after it was created or removing unnecessary data belongs this group.

5.2.2. Point of View

The review of the data by an expert or the administration team gives in most cases more trustfulness in the data, of course, only when assuming that people are acting in conformity to the standards and results which are wished to achieve with the dataset. The expertise in GIS and the knowledge of the needed pragmatic quality is an advantage of this group. Criscuolo et al (2016) mention that a combination of review by administration and by community can provide significant improvements. This is a consequence of the advantage of local knowledge that citizens might have. Also, the objectivity for the overall project as they are not involved in the further processing of the data is reasonable.

The third point of view is the view from the technical automation. This way large amount of data can be verified with low time expanse. However, it is questionable if the result reaches up to the degree of human validation.

5.2.3. Action

The next step is the question about the action which should be taken with the data. Removing the data might let you keep a clean and consistent dataset. However, it means that you lose a part of the information you might want to keep, if not for the primary purpose for the dataset but maybe for further use. Keeping the doubtful information results in a corrupted dataset. Criscuolo et al (2016) argues also with the needed storage for non-useful information. In my opinion this is a void argument due to the endless storage possibilities which have emerged over the last years.

5.3. Categorization for the QA of FRPs

On the pages above I explained an approach of how quality measurements can be categorized. For this thesis I want to use parts of both approaches (Criscuolo 2016, Goodchild & Li 2012) to obtain the best results fitting to my use case.

Crisculo et al (2016) asks the questions when? who? and what? to every method. In my opinion this idea makes sense, as every method can be exactly determined by this proposal. The first approach is not universal usable as there are methods that can’t be situated in one of the categories. The mentioned categories are actually methods focusing on crowdsourced data. The crowdsourcing and the social approach are usable for crowdsourced data only, as it would be too expensive for traditional data collection to hire several contributors who create a lot of redundant data. The social approach would not be a new phenomenon for traditional mapping agencies, as it would represent a common hierarchy within a company. The geographic approach as an intrinsic method can be used anyway not only by a crowdsourcing project.

As described before, this approach is not usable for categorizing the methods, they are methods which I will apply to the use case in the upcoming chapters.

Here, I want to try to put all methods which I use within the project in one of the categories by Criscuolo et al (2016). I used the list I created in chapter 3.3 and added the methods to each element. This resulted in Table 8 where I want to show to which category they belong to after Crisculo et al (2016). The logical consistency, the lineage, the temporal accuracy and the motivation where all already determined in the planning phase of the project. For the logical consistency and the lineage, the technical environment was created which does not allow the crowd users to have impact on this, so these are preventative methods and executed by the project team / administrator. The question about the action is not that easy to answer. If taking the removal option, I do not have the possibility to review the data and maybe miss information which can be reused anyway. Therefore, I decided to categorize them to the warning group, because not able to submit a point can also be understood as a warning, without notifying someone. Completeness was checked during the project, actually it is hard to decide which temporal aspect this element has. As we decided that some parts of the study area are not as well mapped as others, we wrote the crowd users specific emails where we explained them, where there is potential for more FRP. Therefore, the completeness approach implemented a correction process during the ongoing project. The count of submissions per certain area was an automatic process, however, as the city structure can’t be labelled to be homogenous, additional review by the administrator was needed. So, it can be categorized as a semi-automated approach.

Positional and attribute accuracy had two kind of QA methods, for both there was detailed information in the handbook of the project which explains the crowd users how mistakes can be avoided. Anyway, both aspects where controlled after the submissions where taken over by the project team to make sure all mistakes where found. Instead of deleting the wrong data, the information was updated manually. The history and the experience approach are also both categorized as semi-automated methods, like completeness. Last but not least, analysing the motivation of the crowd users happens already during the planning phase, as the project team decided to use financial remuneration to motivate people to contribute to the project.

		Temporal Aspect		Point of View			Action	
		prevention	correction	administrator	community	automatically	warning	removal
Logical consistency	Technical environment	X		X			X	
Lineage	Technical environment	X		X			X	
Completeness			X	X		X	X	
Positional accuracy	Handbook	X			X		X	
	Manual verification		X	X				
Temporal accuracy	Technical environment	X		X		X	X	
Attribute accuracy	Handbook	X			X		X	
	Manual verification		X	X				
Credibility	History		X	X		X	X	
	Experience		X	X		X	X	
Motivation		X		X			X	

Table 8: Methods of QA categorized.

5.4. Prevention

During the planning of a crowdsourcing project, the project team might think about elements that can potentially bear errors and where mistakes can happen. As described in chapter 5.2.1, there are two temporal aspects when it comes to QA of spatial data: to correct the data after it was created or to avoid potential mistakes by introducing tools which help to prevent the dataset from errors. Here in this chapter I want to explain the approaches which follow the second temporal aspect – the prevention. For the FRPs project following elements have a preventative approach, although some of them are also subject to the correction approach:

- Logical consistency
- Lineage
- Positional accuracy
- Temporal accuracy
- Attribute accuracy
- Usability
- Motivation

To achieve the desired quality, a technical environment was created, which did not allow anyone to make errors within the mentioned elements. The whole project was built up to fit a specific purpose, therefore the idea of the data was already very well established. This means, we could exactly determine how the final data should look like and therefore the usability was not a question anymore. We knew that we finally want to have lines on the curb spaces of the street which has the attribute FRP. As the app we used for the contribution did not allow to record lines but only points, a code was created which snapped the submitted point during the postprocessing to the next street section, cut this section by the attribute length of the FRP and changed the attribute to FRP. Therefore, it was obvious, that we need the position and the length of the FRP. To be able to verify the information we decided to add imaginary to the record which should show an overview and a close-up of the FRP. Without these information, the users were not allowed to submit any point. This resulted in a very consistent dataset which was perfect suitable for the project. Lineage is also not an important topic in this project, as there are no other sources than the created submissions from the crowd users. Also, the temporal aspect of the data is nonrelevant as the project had a specific time frame of four months.

Positional and attribute accuracy do also partly fit in the prevention phase. We created a detailed handbook which should answer all the question which might come up during the contribution phase. It includes guidelines to pin the location point correctly (see Fig. 23), how pictures should be taken (see Fig. 25) and how a FRP length should be measured (see Fig. 24). To create the handbook an estimated time expense of 600 minutes was needed, including the creation of the first scratch, testing by co-workers and the finalisation. In the end, the handbook consisted of 22 pages. As the data evaluation showed, the handbook itself was maybe a good start, but as the mistakes that where taken are quite numerous, it is obvious that this alone won't lead to a high-quality dataset.

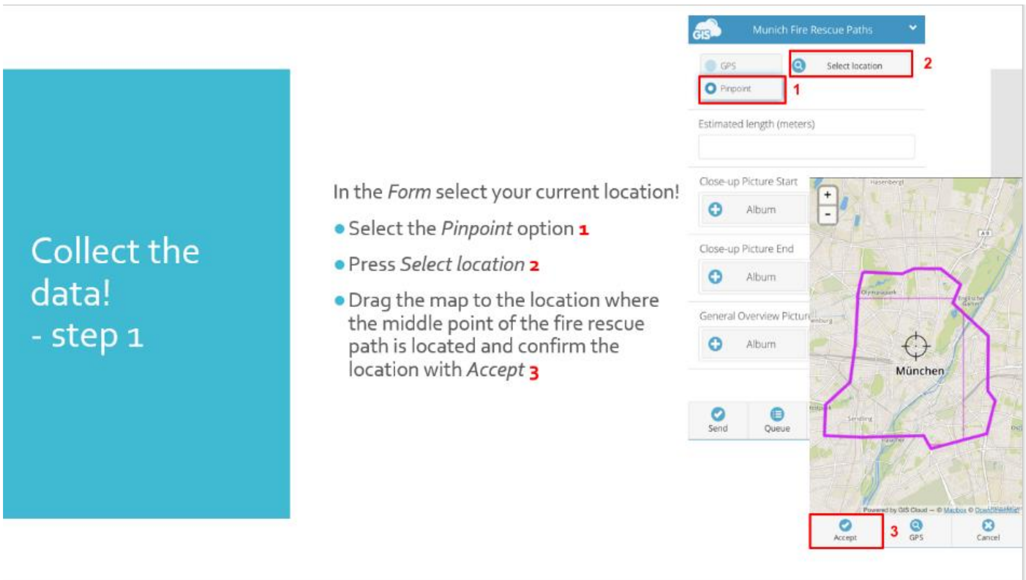


Fig. 23: Screenshot from the FRP handbook, pinpoint details.

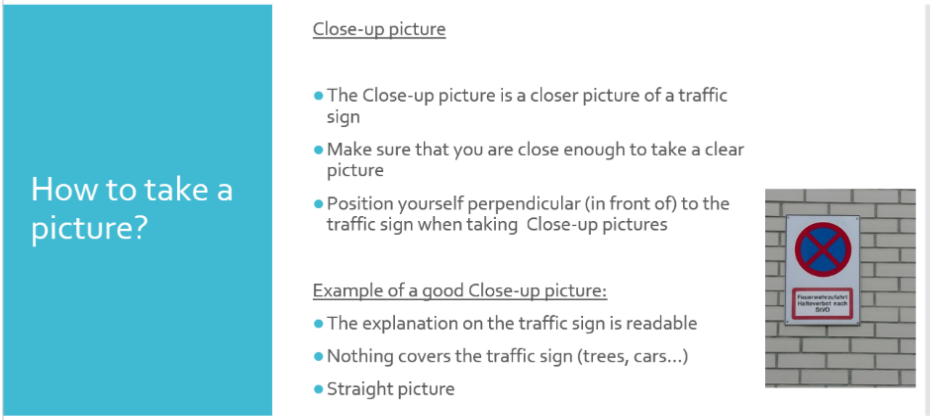


Fig. 25: Screenshot from the FRP handbook, close-up details.

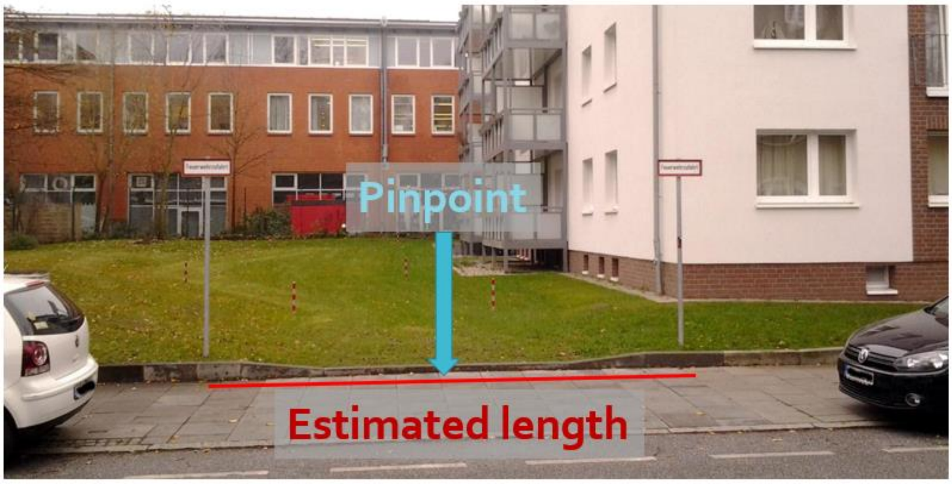


Fig. 24: Screenshot from the FRP handbook, FRP length measurement.

The technical environment was more complicated to build up. Here I do not only include the setup of the GISCloud app itself, also the brainstorming process before, when the project team thought about how the data can then be included in the overall dataset and the coding to transform the submitted points into lines which cut the curb lines of the street dataset had to be developed.

The GISCloud app can be configurated in several ways to fit a specific purposes. Following mandatory fields were created in the app:

- Pinpoint location
- Estimated length in meter
- Close-up picture start
- Close-up picture end
- General overview picture

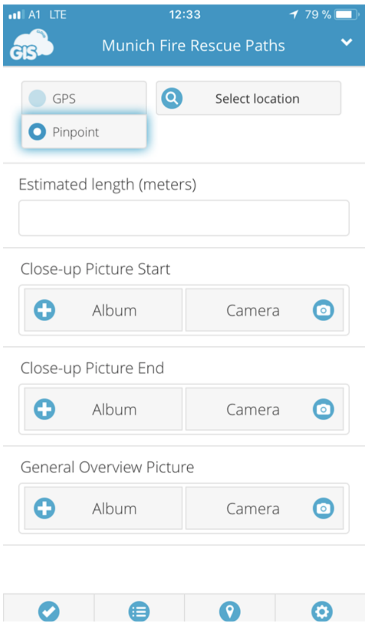


Fig. 26: Screenshot from the GISCloud app with all mandatory fields.

This means that the prevention of mistakes in the FRPs project consists of two parts: the handbook and the technical environment. This led to a dataset with an attribute and positional accuracy of 85% (see chapter 6.2).

5.5. Completeness

The measurement of the completeness of a dataset is an ongoing challenge especially in the crowdsourced world of spatial data. There are several techniques how completeness is estimated. In general, I want to separate the techniques into two groups: one group uses reference datasets to estimate the completeness, the other group does not use additional data for the evaluation.
Examples for the first group are (OSM, 2019):

- The simplest way is the comparison of the data with an external source of the same data. For this approach it is simply needed to compare the counts within a certain area.
- Another way is the one-to-one comparison of undifferentiated features. Here is not only the count important. A buffer around the new features is created to find the corresponding feature in the reference dataset. Thus, not only the count of the dataset but also the location of the features is checked.
- One-to-one comparison of matched features is a similar way as the previous approach. However, here it is necessary that the features of the new dataset have a corresponding feature within the reference dataset. The next step is then to compare the individual records to find inconsistencies.
- The last method within this group is the population-based estimation. If population data is available, an approximate number of data per certain number of population can be estimated or derived from existing mapped units.

In the second group there are all approaches which avoid using external data. The capability to define the quality of a dataset by intrinsic methods is a very important tool. Especially in areas where current spatial information is lacking. Moreover, features which had no importance so far in official data – as the data in the use case within this thesis can potentially be evaluated. Following methods could be collected during my research:

- If an area is already well-mapped it might be feasible to use this area as a reference for a **cross-location comparison**. Counts for features can be estimated by comparing areas with similar economic status. *"for instance, European cities with populations between 250 & 500k are likely to be largely comparable; US & Canadian cities of similar size may or may not & so on"* (OSM, 2019)

- New features in already well-mapped areas are following certain trends when reaching an **asymptote representation**, the so-called Feature accumulation. The closer the number of features comes to 100% the less submissions are expected.
- Another technique is to check **the number of features per area**. This helps to compare certain areas and to show where there might be a lack of data.

As for this use case there is no comparable dataset, only the three methods from the second group will be used and evaluated on the following pages.

5.5.1. Location Comparison

Unfortunately, there is no dataset available which can be used for this method. During my research I realised, that most cities are not responsible for the labelling of fire rescue paths. It is possible to buy a standardized sign online and bringing it up without an administrative review. This means, there is no chance of finding representative numbers without counting them one by one locally.

5.5.2. Asymptote Representation

With this approach, I already expect in advance that a total representation of the reality, so 100% of all FRP in the study area, is very unlikely to be reached. The number will increase and the total count of FRP will get closer to the absolute maximum, but never reach it. This phenomenon is also named **Feature Accumulation** (Mapillary, 2019). In mathematical words, *'the point moves along the branch of the curve to infinity'.* (Kuptsov, 2020). With that knowledge we can say, the closer we get to completion, the lower the number of new submissions will be. Therefore, I took a closer look at the number of submissions over time.

The use case started on the 18. December 2017 and ended on the 9. April 2018, so it lasts for 112 days. Fig. 27 shows the number of submissions that where taken day by day. The counts started low in the beginning of the project and there was even a gap where no submissions at all where done. The reason for this could be the Christmas holiday shortly after the start of the project. In January the submissions where increasing until they reached a peak on the 29th January with 335 submission on a single day. After that the counts kept being low with several exceptions. Due to the asymptote representation we can assume that the percentage of the completeness is already on a high level.

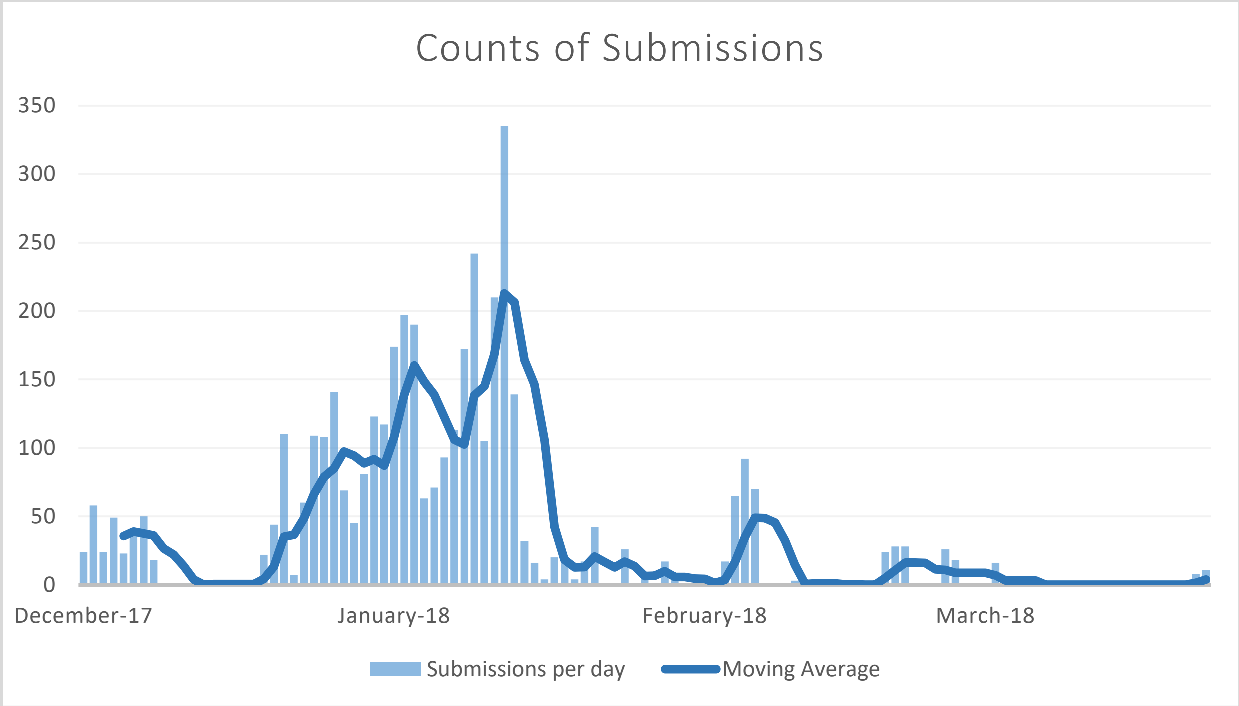


Fig. 27: Counts of submissions per day.

In the following figure I wanted to compare the submissions made with the total amount of submissions. The blue line is a smooth representation of the number of submissions made per month. The orange line represents the total number of submissions. The x-axis shows the months, where I choose to exchange the months names to numbers to increase the readability of the function. As the project started in December, the number 1 represents November 2017, number 2 represents December 2017 and so on. The figure shows, that when the total number of submissions reaches around 4000, the amount of monthly submissions drops radically. Based on the theory of the feature accumulation, it is possible to say, that the number of submissions is already very close to 100%. The question is, when it makes sense to stop the project. Only based on the theory of the feature accumulation, the project team still has to decide, when no big growth in submissions is estimated. This can be very hard when there is no reference data. As Barron (2014) said while finding a quality analysis for a road network: "*Absolute statements on the completeness of the road network are only possible with the help of a ground truth reference dataset*".

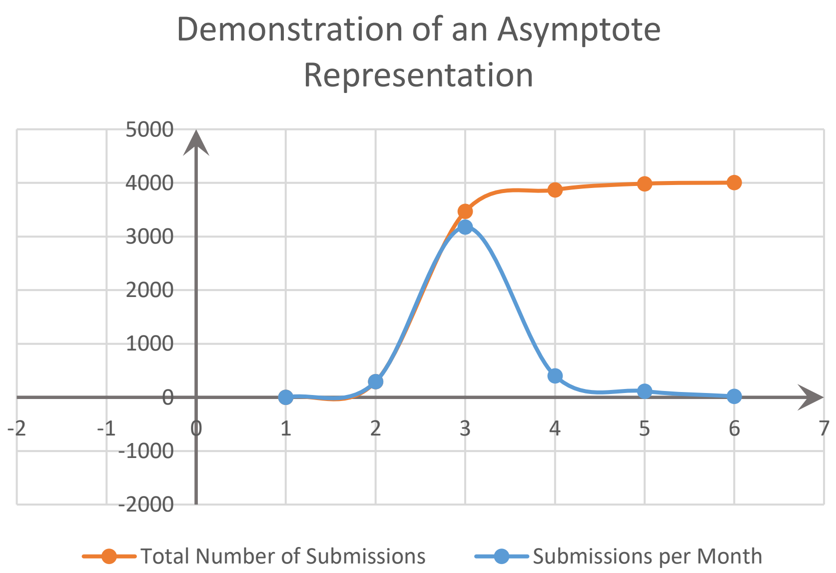


Fig. 28: Demonstration of an asymptote with the project numbers.

5.5.3. Number of Features per Area

This method helps to identify inequalities of the quantities within the dataset. The study area is split up into smaller regular parts, thus it is possible to compare certain locations and find out if there are differences in the number of features per area. For this analysis QGis 3.4 was used. The first step was to create a grid to split up the study area into smaller squares with consistent size of 1x1 km, or 1 km². For this step the tool 'Create Grid' (see Fig. 29) was used. The grid extend is based on the border of the study area which was created along the 'outer ring', as explained in chapter 2.1. For the sizing I decided to take 1km as it is a number which is easy to calculate with, and most people are able to imagine the size of one square kilometer. The importance while using this tool is, that a metric CRS is used as otherwise the size will be in degrees and the area cannot be divided into cells with specific number of meters. Therefore, the UTM 32 N (EPSG: 4647) is used. The result of the tool is shown in in Fig. 30. The problem with this result is that we have now more squares than needed, as the tool always creates the same number of squares horizontally as vertically. So, I deleted manually all squares which are not within the study area (see Fig. 30 / b). The next step was to cut the polygons along the border line (see Fig. 30 / c), so no space outside of the study area will be included in the future analysis.

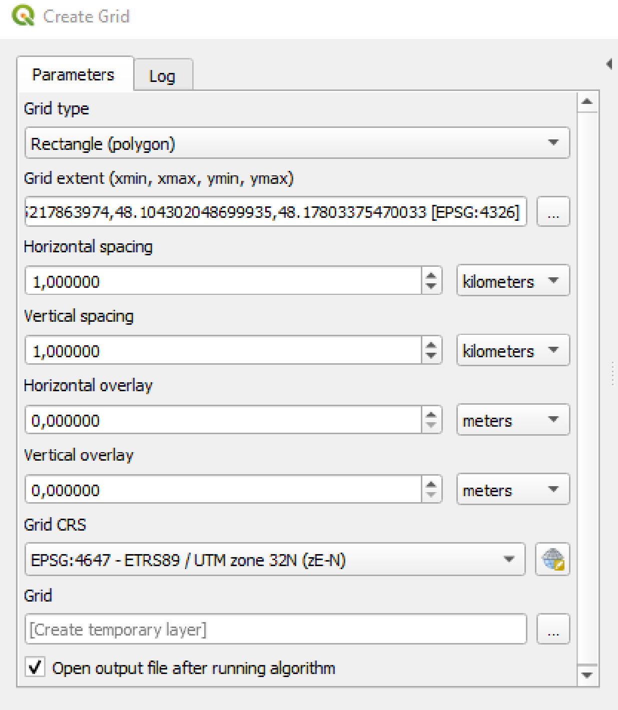
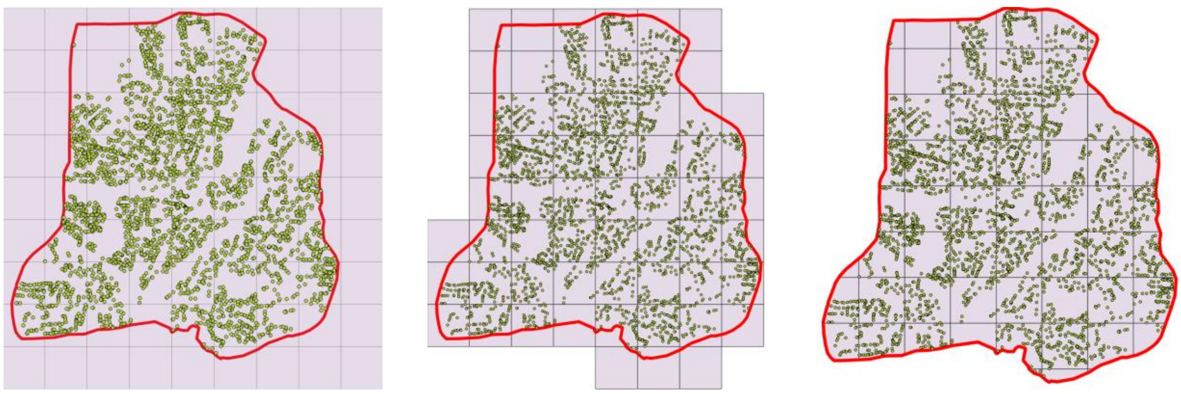


Fig. 29: Screenshot of the Parameters of the Tool 'Create Grid'.



a b c

Fig. 30: Comparison of the squares. a: after 'Create Grid'. b: after deleting the additional squares. c: after cutting the squares along the border.

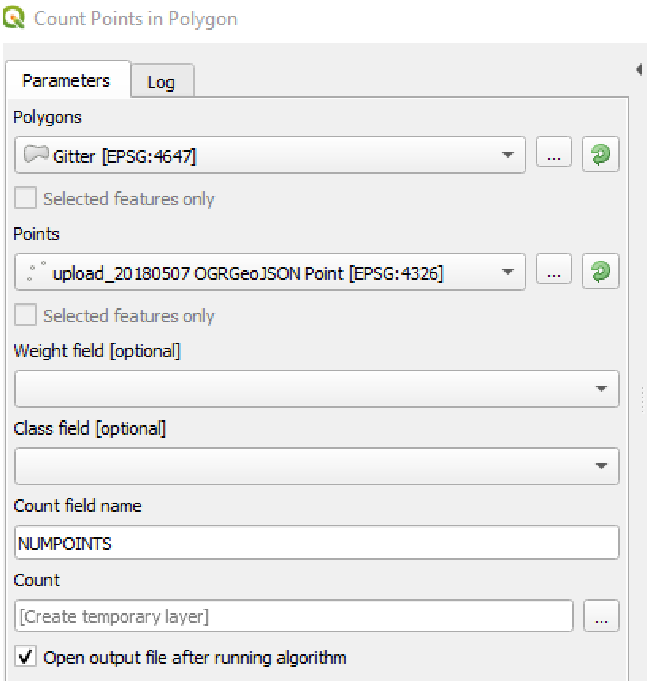


Fig. 31: Screenshot of the 'Count Points in Polygons' tool.

After that, I calculated the number of points which are within the individual squares. Therefore, I used the tool ‘Count Points in Polygon’ as Fig. 31 shows. The result is saved as a new column “NUMPOINTS” in the attribute table as you can see in Fig. 32.

As not all squares are totally fitting into the study area, not all squares have the same size. I wanted to overcome this problem as I calculated the number of points relative to the size of the square. Therefore, I calculated *NUMPOINTS / km2*. This way I can compare all squares, even if they do not have the same size.

The result was saved in a new column named “p pro km^2”. In Fig. 32 a part of the attribute table is shown. The first column represents the id of each square, the second column shows the number of points which have been calculated with the ‘Count Points in Polygon’ tool. The third column shows how many points there are per square kilometer. The last column is the size in square kilometer.

id	NUMPOINTS	p pro km2	km2
10	2	10.417	0.192
11	0	0	0.395
12	40	98.039	0.408
13	67	149.554	0.448
14	66	119.134	0.554
15	69	81.176	0.85
16	77	77	1
17	75	101.215	0.741
19	0	0	0.619
20	12	12	1
21	60	60	1

Fig. 32: Example of the attribute table of the squares

The importance of the relative comparison of the points becomes obvious when comparing the absolute and the relative counts side by side as shown in Fig. 33. Several squares belong to another category when changing absolute to relative numbers and the whole map looks more complete when only the relative numbers are considered. This is important if the project team decides to lead the crowd to specific areas to increase the level of completeness.

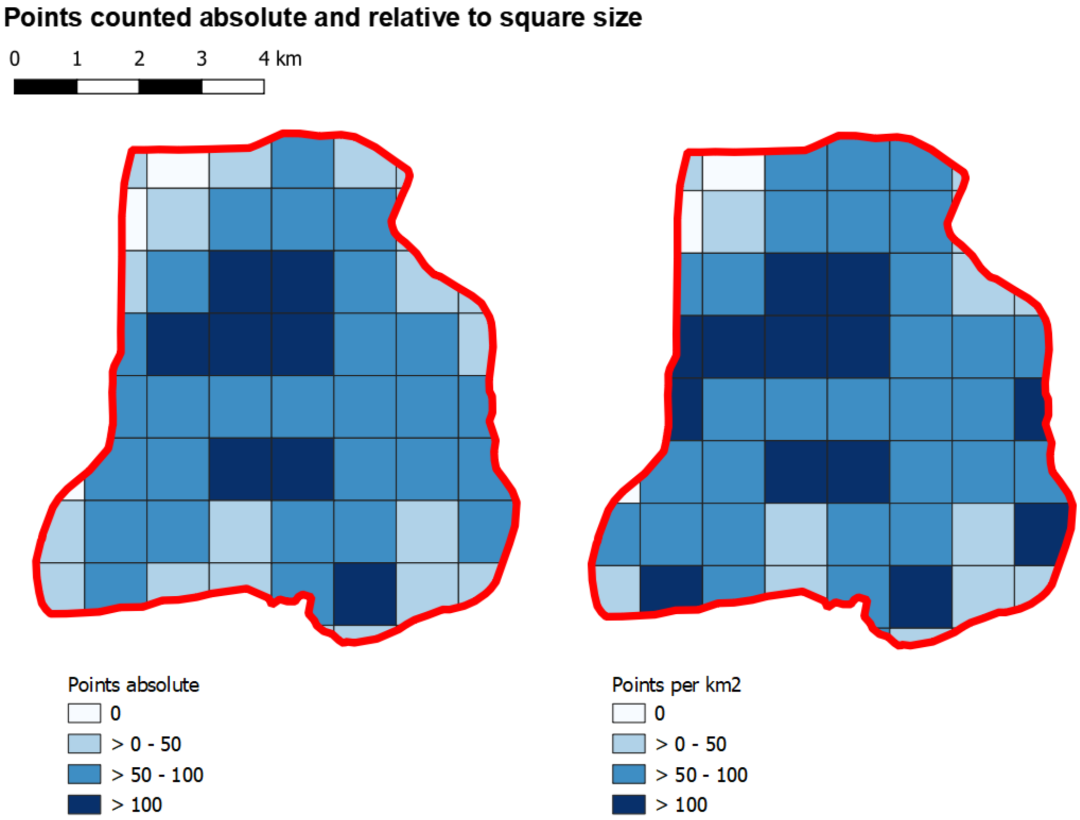


Fig. 33: Point density, absolute counts compared to points per km².

For the next step I used an OpenStreetMap as background to see, if I can find out why there are such inequalities within the research area. For the easier understanding, I labelled every square with an UID, a unique identification, as you can see in Fig. 34, so the reader can follow my analysis.

When I look closer to the environment of the different locations, several divergences of the counts of the points became obvious. There are no points at all in square 6, 19 and 56 which might be a consequence of almost the whole area is covered by a park and almost no buildings are placed here. However, the missing points in square 11, 57 and 66 need a closer look, as there is potential of missing fire rescue paths, due to the existence of buildings and streets. In the very small square 55 it is truly possible that there are no FRPs, but also here we cannot be totally sure. In squares 20, 34, 35 and 71 there is a lot of green area which can result in the lower number of FRPs. My first proposal would be, to tell the crowd users to go to areas 8, 54, 61, 62 and 66 as the potential of finding new FRPs might be the highest due to the number of buildings and roads. In my opinion this method is lacking in the consistency of the city structure. As there is not an equal number of buildings and streets in every part of the city, not every area is comparable. My proposal would be to recreate this comparison and count the points relative to the meter of streets in every section. Still this might not the ideal solution, as there are also streets without FRPs like highways, but the accuracy would be higher than only comparing equally sized areas. Especially when it is possible to allocate the submissions to a street section, the parts of a street which are not covered so far can be defined on a bigger scale.

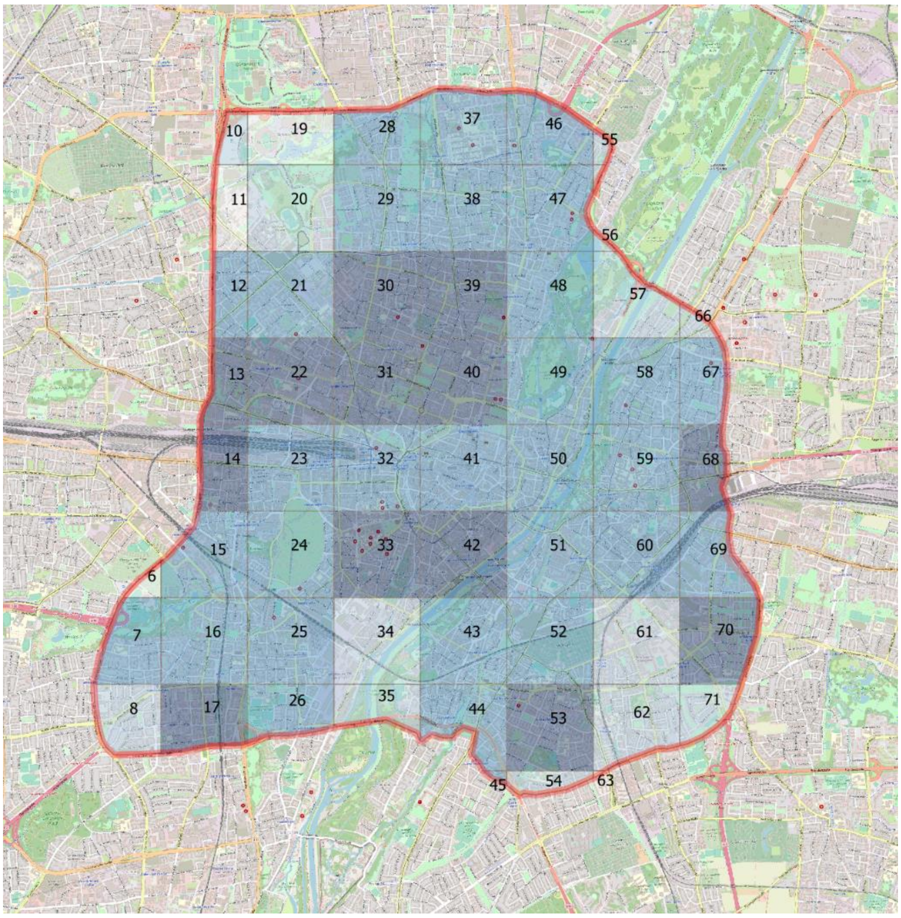


Fig. 34: Overview with OSM in background and UID for each square.

5.5.4. Number of Features per Street Kilometer

As a consequence, I took the next step and created the number of features per street kilometer analysis. Therefore, I downloaded the street dataset from OSM via Geofarbik (2020). To reduce the amount of data I immediately cut out the street dataset by the border of the research area. There was a total length of all road of 234 928 000 000 km left. Of course, I needed to clean out the dataset first.

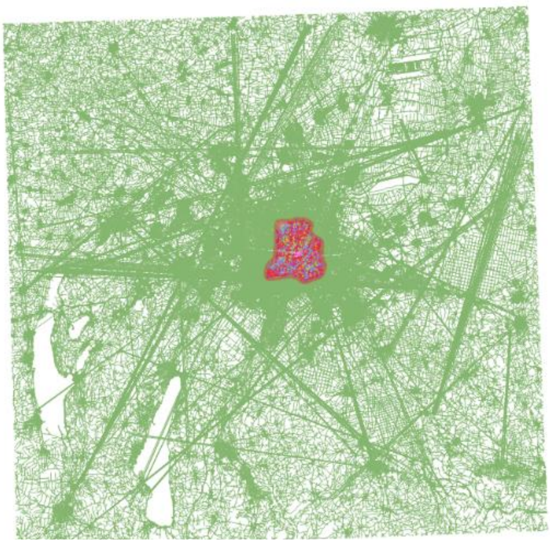


Fig. 36: Street dataset compared to the research area.

Fig. 36: Streets of Munich with wrong tagged features as highlighted in red.

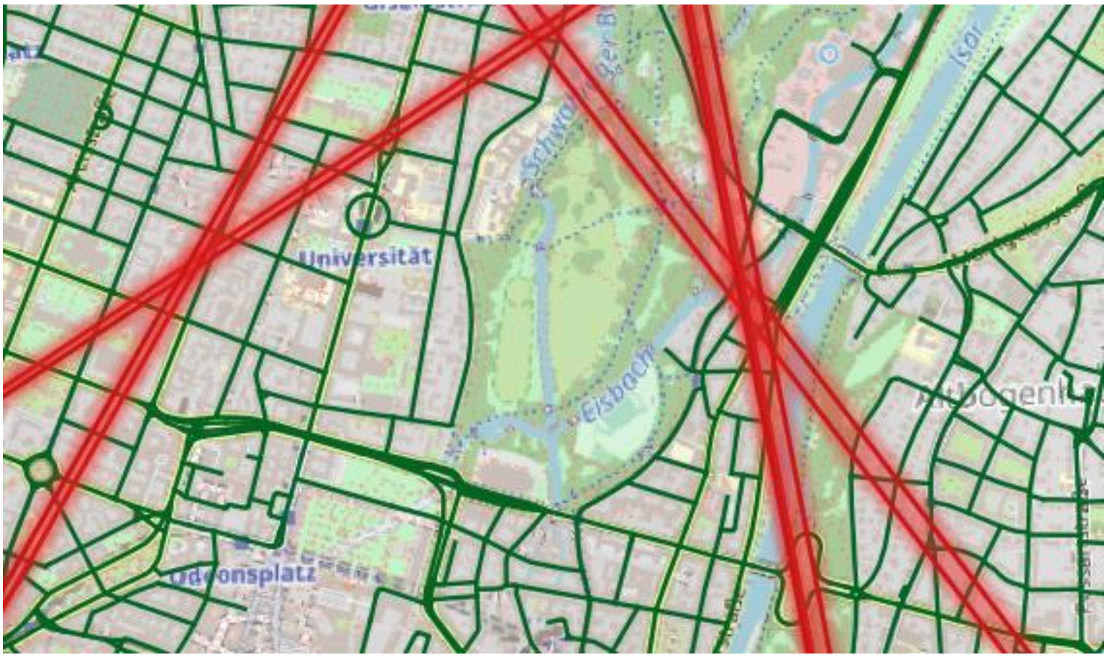


Fig. 37: Close-up of the street selection. Green streets are kept, red lines are deleted. Map: OSM

I decided to keep only streets that are relevant for the target group of Parkbob, which are car users. Therefore, I deleted all features that were tagged as bridleway, corridor, cycleway, elevator, footway, proposed, steps, service or path. All those tags describe roads for horses, pedestrians, bicycles or for roads that are off street like inside buildings or garages. After that, I realized that the dataset includes some lines which are not actual roads as highlighted in red in Fig. 36 and Fig. 37. As it seems that those lines do not follow a certain scheme, I decided to delete them manually. They all had different tags and I was also not able to identify which pattern they follow. As there are several of those lines, I wanted to find an approach where I do not have to select each line manually. As the lines are positioned all over the dataset, I thought that they have to be way longer than any actual street. Therefore, I sorted the street dataset by length and selected all lines which are longer than two kilometers. After controlling the selection, I could delete all selected features as I only those wrong lines were detected. Finally, there are 592 636 meter of street sections left.

In the next step I calculated the length of the lines which are inside a grid polygon. Therefore, I used the tool ‘Sum Line Lengths’. To put the counts of the points in relation to the length of the street sections in every square, I calculated *NUMPOINTS / km*. This way I have the number points per one kilometer of street section in one square. As we see in Fig. 39 there are no points in square 6, 11 and 19. For 19 and 6 I would agree with the not existence of points, as there

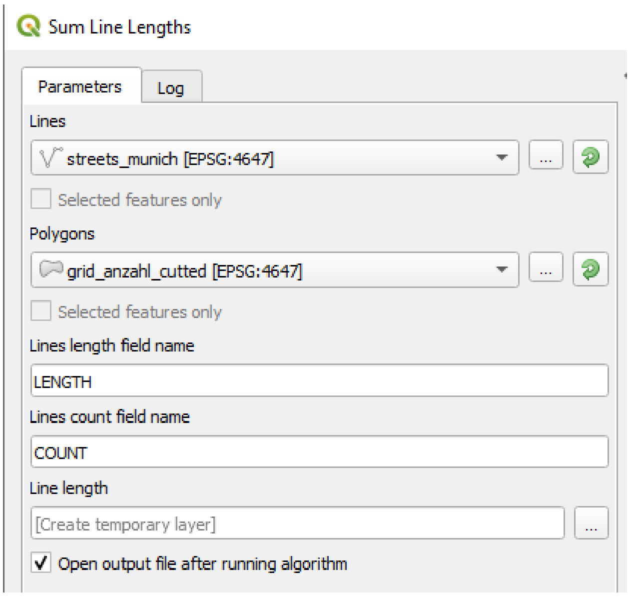


Fig. 38: Screenshot from the tool 'Sum Line Lengths'.

are also almost no street sections. However, it is possible that there are still unregistered FRPs in 11. When only concerning this map, the low number of points is also reasonable in 10, 20, 35, 56 and 57 due to the less dense street network.

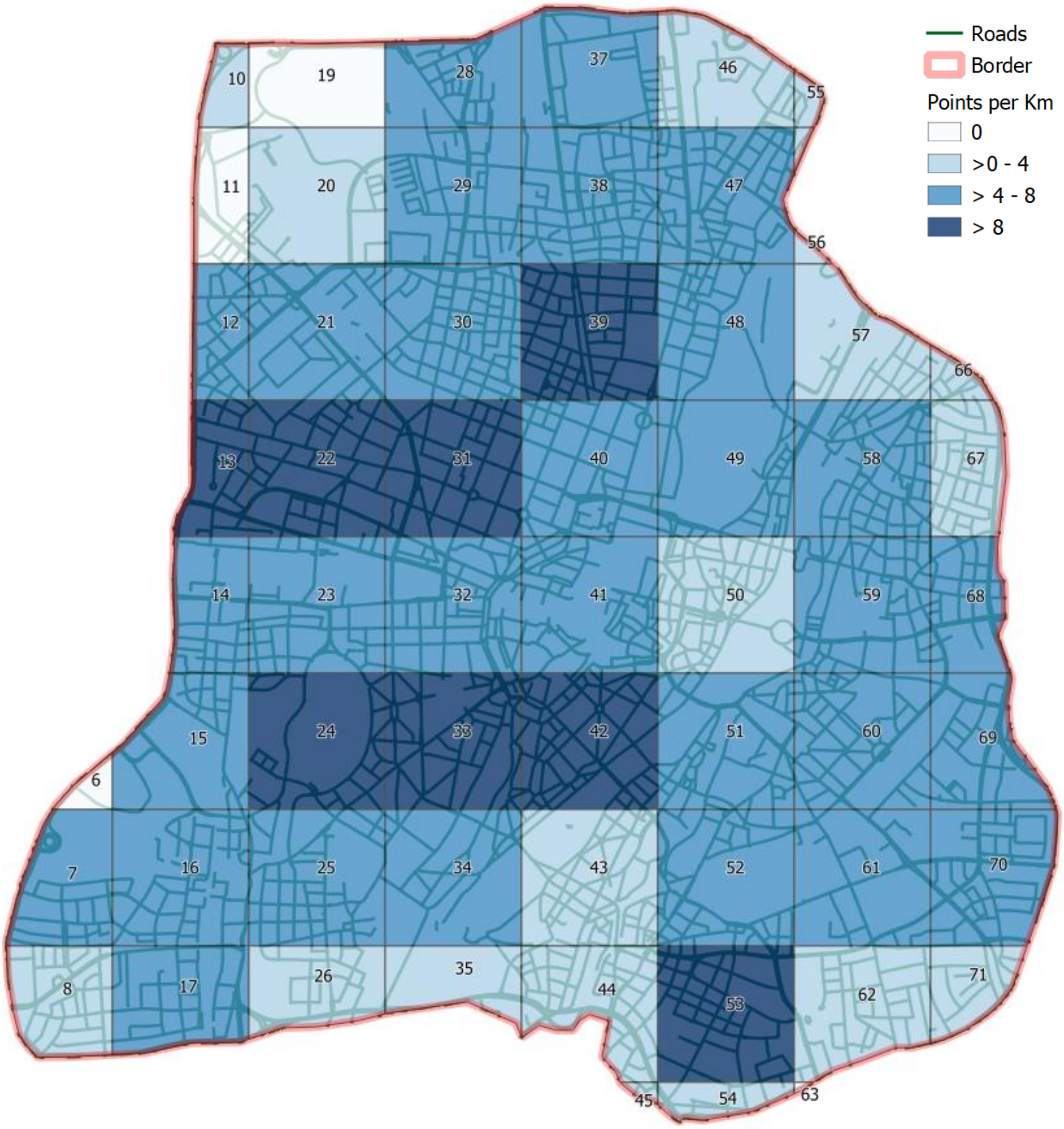


Fig. 39: Points per km overview.

To be able to compare both approaches. I coloured the squares based on the standard deviation.

	per area	*per km*
standard deviation	38.98	3.02
variation	1519.59	9.11
average	68.86	5.09

Table 9: Variation measures.

$$standard\ deviation = \sqrt{\frac{\sum_{n=1}^{n}(x_{i-\bar{x}})^2}{n-1}}$$

The standard deviation shows the square root of the variation of each FRP dataset. In this case we only need the lower standard deviation, as outliers higher the standard deviation are actually good for us as it shows squares with a higher density. The negative outliers give us a clear signal, that there are less FRP than in most of the other squares.

	STAD	1ST BREAK (AVERAGE – STAD)	2ND AVERAGE	3RD BREAK (AVERAGE + STAD)
P / KM²	38.98	29.88	68.86	107.84
P / KM	3.02	2.06	5.08	8.10

Table 10: Statistical calculation of the breaks.

When comparing both approaches, we see that most of the results are similar. In the previous analysis I suggested to send the users to square 8, 54, 61, 62 and 66. When I take closer look in Fig. 40 that the results are very close. However, the result is here more explicitly as all white squares are below the standard deviation and belong therefore to the outlier group.

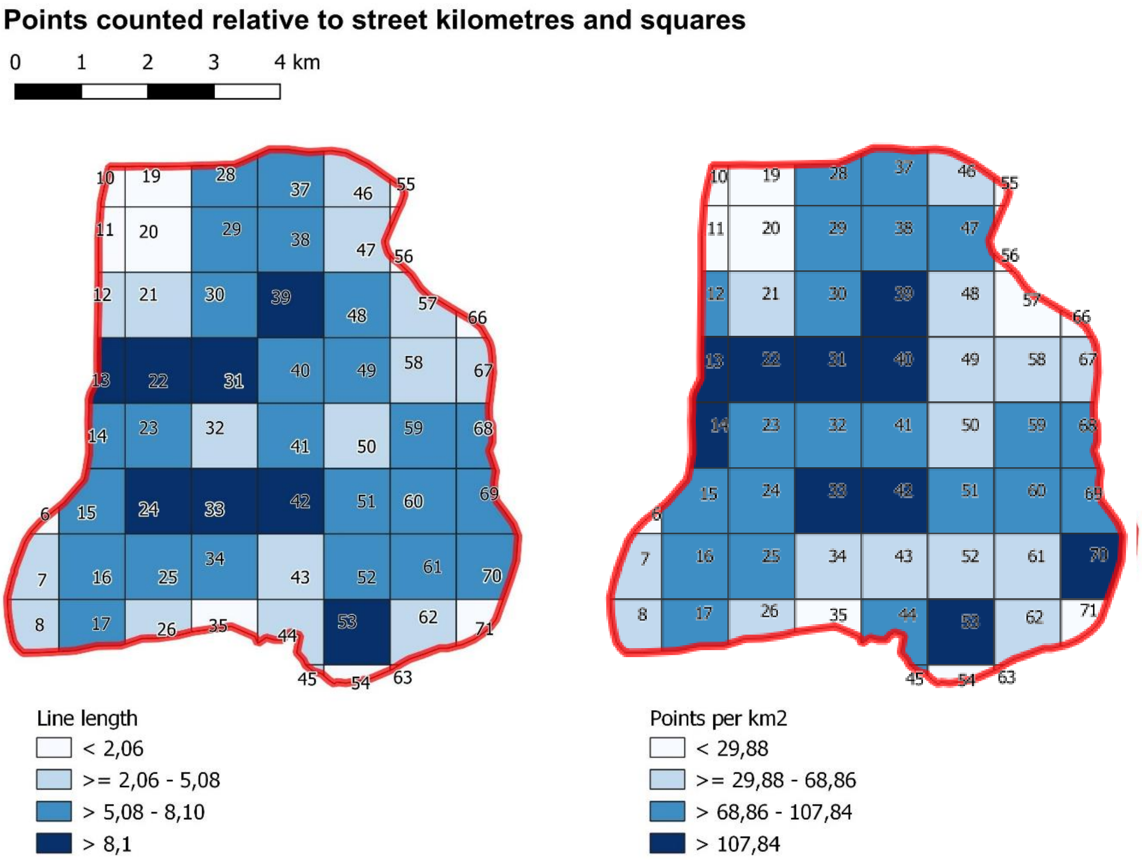


Fig. 40: Comparison of points per street km and points per km².

To see if my visual impression is close to the truth and to be able to make quicker recommendations to the crowd, a numeric approach might be preferred. The above analysis is based on personal interpretation, which can be without doubt a powerful tool, however, to keep this approach scalable, I also want to propose a way to calculate the presumptive neglected areas.

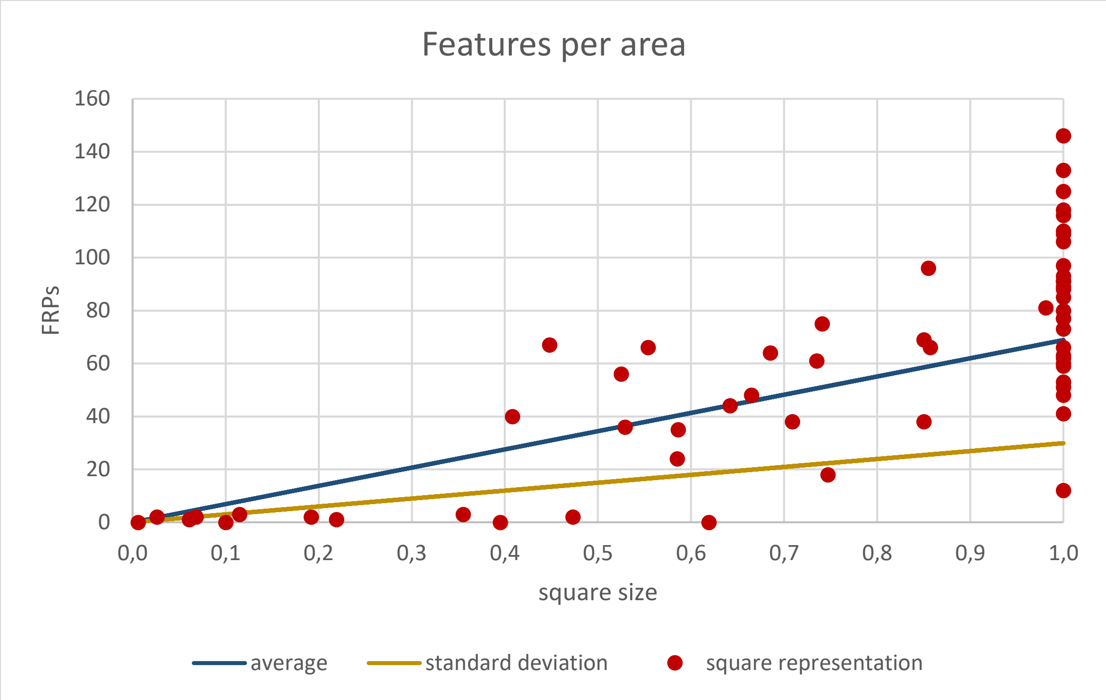


Fig. 41: Overview of the FRPs count relative to the square size.

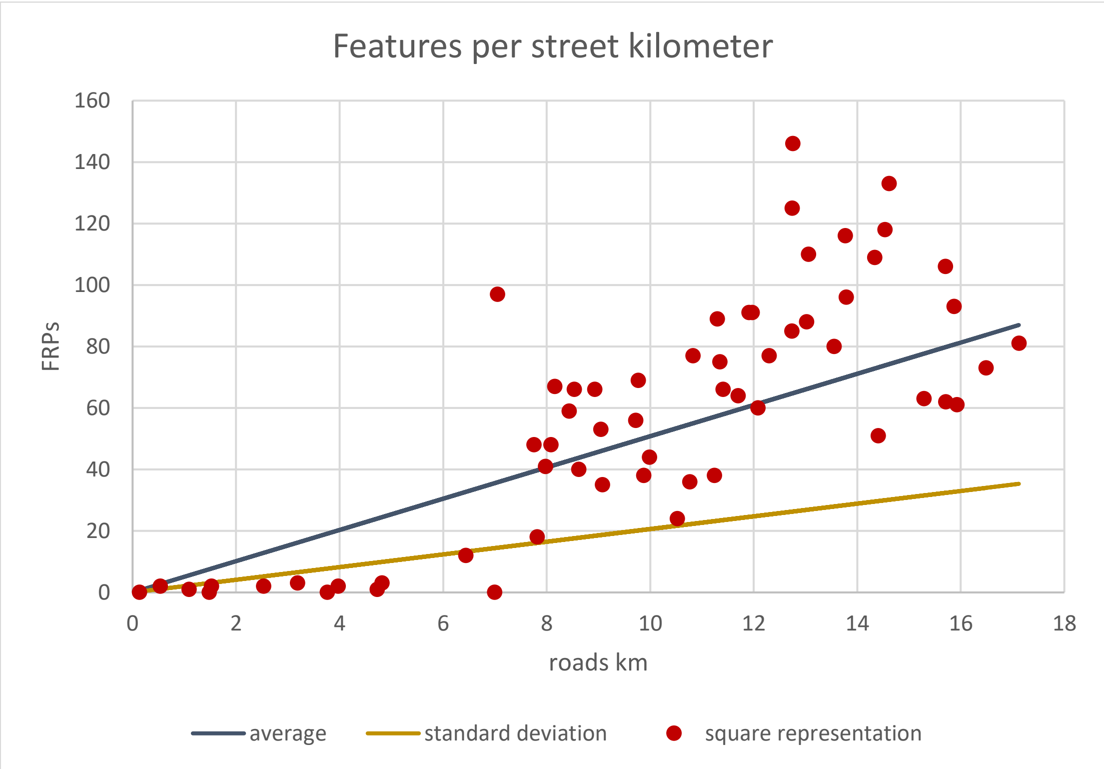


Fig. 42: Overview of the FRPs relative to the street kilometers.

Each red dot in Fig. 41 and in Fig. 42 represents a square with its FRPs count per the square size or per kilometers of roads. To find out, which squares a lacking in submissions I decided to take the standard deviation in account. The variation measures in Table 9 are calculated based on the FRPs per km^2 and the FRPs per road kilometer. As I want to have the standard deviation and the average relative to the cell size, I multiplied the numbers by the cell size. I did the same calculations for the road kilometers analysis respectively with the street kilometers instead the cell size. This way I get the measurements I need to compare each square. Fig. 41 and Fig. 42 show the average number with the blue line and standard deviation with the yellow line. Based on that, I would conclude that all squares below the yellow line, have not been as good mapped out as other parts of the city. To use a calculated method, I subtract the standard deviation from the number of points in every square. Every object that has a solution smaller than zero, belongs to the outlier group. Table 11 shows some example records from the calculation to give a better overview.

id	km²	FRPs	FRPs / km²	average / km²	staD / km²	FRPs – StaD/km²
6	0.10	0	0.00	6.89	2.99	-2.99
7	0.64	44	68.54	44.21	19.18	24.82
8	0.59	24	41.03	40.28	17.48	6.52
10	0.19	2	10.42	13.22	5.74	-3.74
11	0.40	0	0.00	27.20	11.80	-11.80
12	0.41	40	98.04	28.09	12.19	27.81
13	0.45	67	149.55	30.85	13.38	53.62
14	0.55	66	119.13	38.15	16.55	49.45

Table 11: Example of the variation calculation. FRPs per km²

id	km	FRPs	FRPs / km	average / km	staD / km	FRPs - StaD/km
6	1.48	0	0.00	7.51	3.05	-3.05
7	9.98	44	4.41	50.72	20.60	23.40
8	10.52	24	2.28	53.47	21.71	2.29
10	2.53	2	0.79	12.85	5.22	-3.22
11	6.99	0	0.00	35.52	14.42	-14.42
12	8.62	40	4.64	43.79	17.78	22.22
13	8.15	67	8.22	41.43	16.82	50.18
14	8.53	66	7.74	43.34	17.60	48.40

Table 12: Example of the variation calculation. FRPs per street km.

Based on this calculation, in both approaches 12 squares were identified to be beneath the standard deviation: 6, 10, 11,19, 20, 35, 54, 55, 56, 63, 66, 71 and additionally a 13th square, number 57 when it comes to the street kilometer analysis. Eight of these squares were already identified by the visual analysis. For most of them, the explanation for the few numbers of submissions was the lack of buildings and streets. This means the calculated approach can be executed very fast and both, the features per area approach and the features per kilometer approach lead to satisfying results in a decent time.

To sum it up, it is an easy and very fast tool which gives an overview of the completeness of the dataset. This can be a reliable information, when the city structure is concerned as well. For the features-per-kilometer approach the data cleaning can be a tricky step in the setup of the tool. To achieve accurate results a careful pre-processing of the data must be done.

5.6. Positional Accuracy

The positional accuracy is defined '*as the degree to which the digital representation of a real-world entity agrees with its true position on the earth's surface.'* (Harding, 2005)
How accurate the data should be is of course always depending on the specific use. Apart from that, there are some approaches that can be concerned when identifying the quality of the position.

- **Many eyes**: This means that besides the contributor another person checks the information that was submitted. This can happen by the administrator team – as it happened during this project – or it can happen by other contributors. To allow other users to change submitted data, the technical environment needs to be well established which is not always the case, and it was not the case in the FRP project.
- **Enough submissions:** To rely on enough submissions would mean to allow the contributors to submit data more often. After a certain amount the same record was collected, the homogenous information of all submission will be extracted. If more submissions contain the same information about the position, it is more likely to have the correct location.
- **Geographic approach:** Here the data is viewed in its geographic context. This approach was also used during the FRPS project, as the logic of the data was aligned with streets and buildings. This can happen automatically or manually. In this case it was a manual approach, as it was always checked if the FRP was close enough to desired street section.
- **Social approach:** The social approach separates users into groups with different rights. Advanced users can be gate keeper or can obtain an administrative role. Also, higher rewards are possible to increase the motivation. This approach is explained in more detail in chapter 5.8.

The accuracy is always depending on the needs of the actual project or the users' needs (see Fig. 9 and Fig. 10). The question is how accurate do you need the data? In our case we explained the user in advance some aspects they should consider before submitting a point. The most important thing is that a FRP is always located on the side of a street. In this phase, it was already planned to snap the points in the postprocessing process to the closest street section. This means it was important that the point lies unambiguous next to one street section. If the point was submitted somewhere in the middle of the street where it would not have been clear to which side of the street the point would be snapped to, this would have caused a positional accuracy which would not meet the needs of the project. Therefore, we moved it manually closer to the desired street section as it happened in the processing phase of the project.
The accuracy of the location is very much depending on how the submissions were created. This can happen in two ways, first, the submission is created automatically by saving the actual position by a Global Navigation Satellite System (GNSS) or manually, by dropping a pin on the desired location. The study area is located in the centre of the city of Munich where there are a lot of high buildings. This might influence the accuracy of the smartphones' GNSS signal and can lead to divergences of the tracked location and the real location of 100m and even more. To avoid this problem, the location had

to be picked by the contributor manually and not automatically by the app. Anyway, we tracked the GPS location as well to be able to analyse the difference of the GPS location and the manual entered location. In most cases there was not a lot of variation between the measured location and the pinned location as the close-up screenshot in Fig. 44 shows.

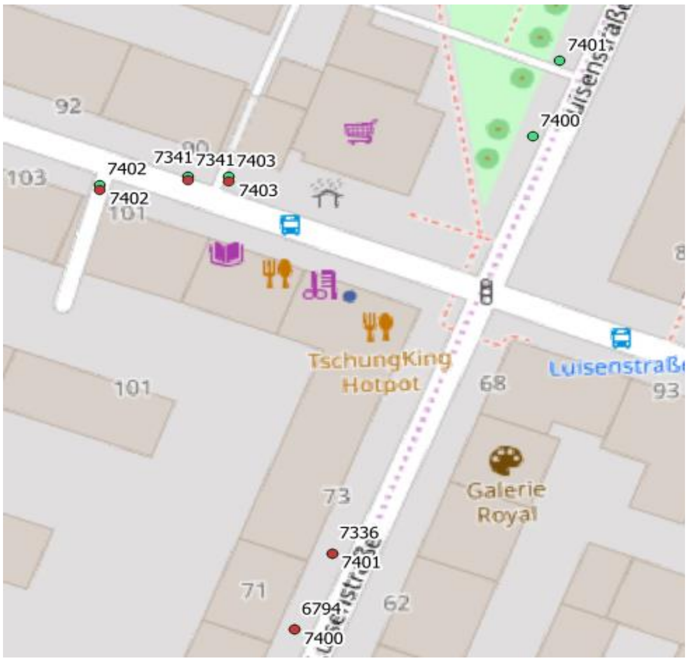


Fig. 43: Comparison of the GPS (red) and the manual entered location (green). Map: OSM 2019.

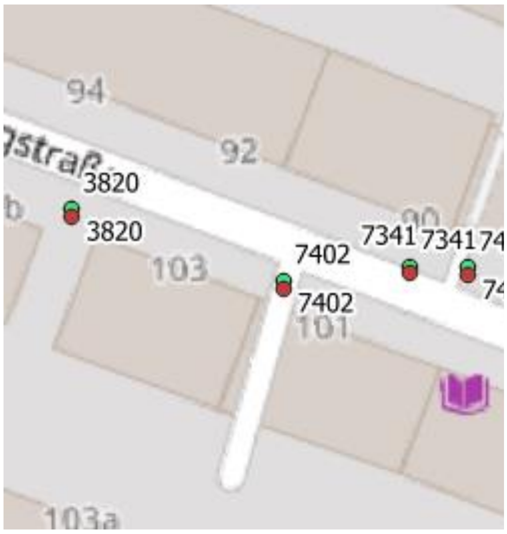


Fig. 44: Comparison of the GPS (red) and the manual entered location (green) close-up. Map: OSM 2019

Nevertheless, it can happen, that the difference between the two locations is more than just some meters as the screenshot in Fig. 43shows. The red dots represent the GPS location and the green dots represent the location pinned by the users. Most of the submissions have differences of just some meters as you can see in Fig. 44. However, the inequality of the location of point 7400 and 7401 in Fig. 43 to its corresponding real location is approximately 100 meters.

So, the first step of the quality assurance of the positional accuracy was not to accept automatic GPS location, this method is a prevention method and is named here as **handbook method**, as the problem is explained to the users in advance through the handbook. The users needed to submit a manual location.

Fig. 45: Location mistakes red – pinned location, green – real location.

Anyway, we did not expect that the contributor gives perfect locations either. This means also those locations had to be checked after the submission was made. In Fig. 45 there is the case where the manual pinned location was not done carefully. The two red points are the pinned location and the green points represent the real locations. The differences are approximately 20 and 30 meters which would make an important difference when these points are added to the dataset. Therefore, a manual check needs to be done and all locations need to be crosschecked. Here the points where moved manually by the project team.

The **manual verification** is a very time-consuming method. It means that every single contribution must be checked individually. As there is no reference dataset which can be used for crosschecking, the way how to verify the data is personally and one by one. This means a person has to visit personally the exact spot where the submission was taken and check if it has the correct position and semantic information. As this is not very practical and would make the crowdsourcing approach redundant, alternatively online services were used. There are free services where you can get insight in streets of cities without being physical present. Examples for those services are:

- Google Street View
- Mapillary
- Google Maps, Bing Maps Aerial Imaginary

Crosschecking with those services is only possible if the desired location has data. No-one of the above-mentioned services does cover 100 % of the research area. The next thing to consider is, that the imaginary available through those services is not always up to date. Within city areas, there is a lot of construction going on: buildings are teared down, new buildings are built, passenger areas are created and so on. Also, the time of the year the imaginary was captured influences the quality. If the street has avenue characteristics with a lot of trees on the sidewalks, it can happen that due to the fully leafed trees, the needed information is covered. During the validation process, the Google Street View imaginary was on average 9 to 10 years old

This means, the verification is depending on two major parameters: the completeness and the currentness of the online service. In this use case, most parts of the city are covered with the mentioned services above. In Fig. 46 you can see a screenshot from Google Street View in the Map View from the Schyrenviertel in Munich. Every street which has a blue line has captures from the street. Some streets however are not covered, like Freibadstraße or Humboldtstraße.

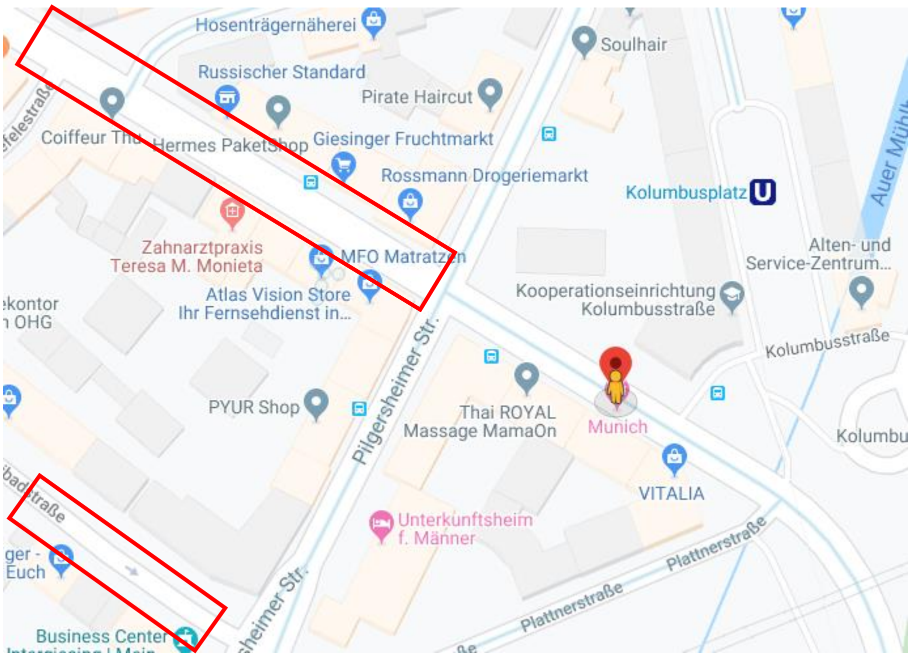


Fig. 46: Screenshot from Google Street View Overview. Captured 2020-01-28

Fig. 47: Screenshot from QGis with OSM in background. Captured: 2020-01-28

To get fast to the exact location, where the submission was taken, I used the plugin ‘Street View’ for QGis which opens a new window in a web browser and pins immediately the location you clicked in QGis. Once the window is open, you only need to check if the FRP is where you expect it. This way the crosschecking with the imaginary gets really fast and submissions can be approved within less than 30 seconds. In the example of Fig. 47 the point with the id 4874 is checked and in Fig. 48 you can see the imaginary from Google Street View for the same location.

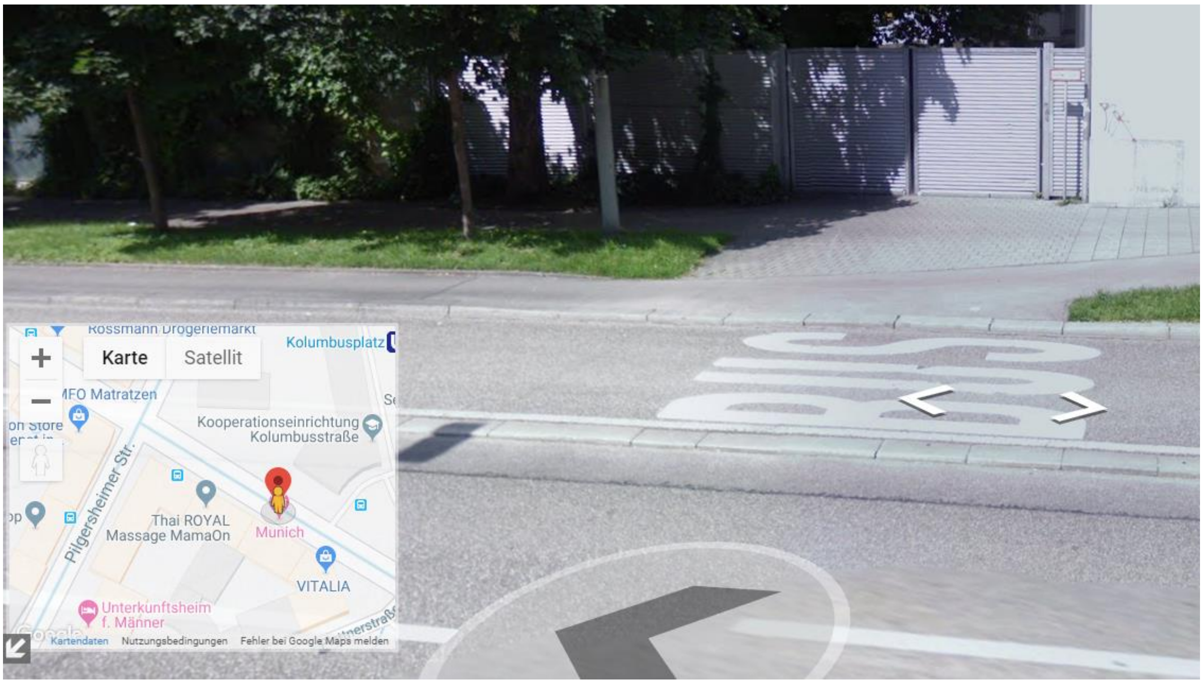


Fig. 48: Screenshot from Google Street View. Captured 2020-01-28

Google Street View is in most cases the first choice as the quality of the imaginary is most of the time the highest. However, it can happen that the desired location is not covered by Google Street View. In this case my second choice was Mapillary. The plugin ‘go2mapillary’ for QGis allows you to see all imaginary of Mapillary inside your QGis project as you can see in Fig. 49. This way you get a quick overview whether you can verify the FRP with Mapillary or if there is no coverage within the desired area.

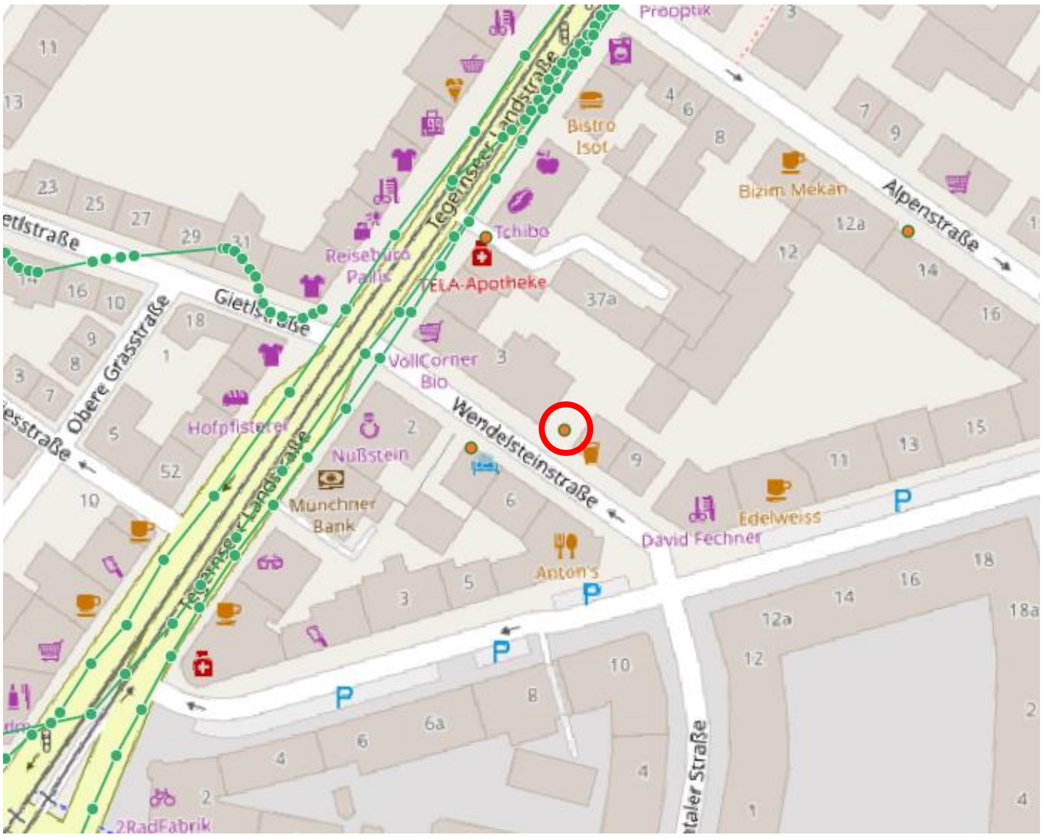


Fig. 49: Screenshot from QGis with the go2mapillary Plugin activated. Captured 2020-02-17

This is the most time-consuming part in the evaluation. It becomes very challenging when there is no Google Street View and no Mapillary imaginary on certain location like for the red circled submission in Fig. 49. In that case I used the Google Maps 3D View which exists worldwide. It is not very clear and if there is a FRP can only be assumed. With comparing the picture of Google Maps 3D View (Fig. 50) with the picture of the submission of the user, in most cases it is clear if the submission is on the correct location.

Fig. 50: Google Maps 3D View. Captured 2020-02-18

The final problem is then, if the imaginary data is outdated. Most of Munich imaginary by Google Street View was created in 2008 and 2009. If there were constructions and sections were rebuilt the submissions cannot be compared. If there is also no possibility to crosscheck with Mapillary, I decided to approve these points in favour of the crowd user, as I am not able to prove them wrong. This case, that there is no way of crosschecking the points, happened about 10 times during the whole project. By the total amount of 3653 submissions this small amount is negligibly.

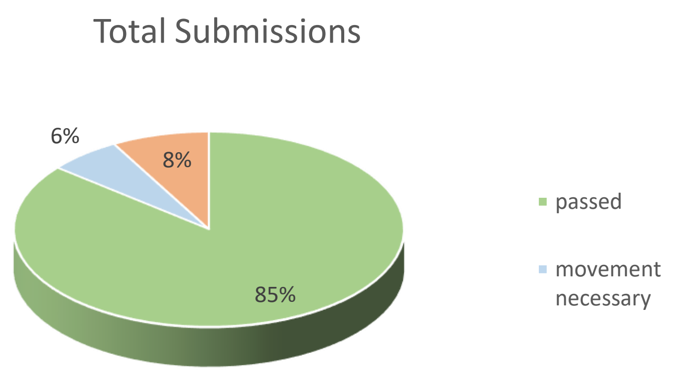


Fig. 51: Share of submissions with wrong location.

During the manual QA of the data, 313 FRP where identified on the wrong location. Sometimes it happened that the points were not positioned totally wrong, but not as accurate that they would be relocated to the correct street section. All points which the project team could identify as located wrongly, and the actual location could be found, were moved. If the location could not be detected, the point became invalid. This however was not the usual case. 6 % of all submitted points where shifted to the correct location as you can see in Fig. 51.

5.7. Attribute Accuracy

The methods of QA of the attribute accuracy are similar to the ones described in the previous chapter.

- **Many eyes:** This approach doesn't change a lot when it comes to attributes rather than positioning. Only that here, the attributes are checked. For the FRP project this approach is used.
- **Geographic approach**: Attributes of geographic data can also have a geographic context. For example, it is possible that cities within a dataset have the attribute of the corresponding country. Or the data has attributes that can be approved based on the geographic context they are situated in, like a railway station which can only be located along a railway.
- **Enough submissions:** Also, this approach is similar to the one with positional accuracy. If an attribute is repeatedly the same in several submissions, it is more likely to be correct.
- **Social approach:** The social approach can allow certain users to get superior rights or even other tasks, based on their experience or history.

Within this project the attribute information is limited to the length information. The way the data is collected, no other attribute information can vary in its quality. Every submission must be a fire rescue path and all other information about the users and the used technology is gathered automatically. The only exception is the length of the fire rescue path, which each user must estimate by itself. Thus, in this chapter I will propose methods to determine the attribute quality of the length of the geometry.

During the first step, the prevention phase, I tried to limit the mistakes the users can make to a minimum. Therefore, I want to keep it as simple as possible. In the manual, I explained that the minimum length of a FRP is under any circumstances at least three meters (Stadtverwaltung München, 2020). The accuracy of the length should be inside the variation of two meters. The reason for this is, that the future dataset will not have more accurate information and however, during the usage of the app for finding a parking space, the information does not need to be more accurate than two meters.

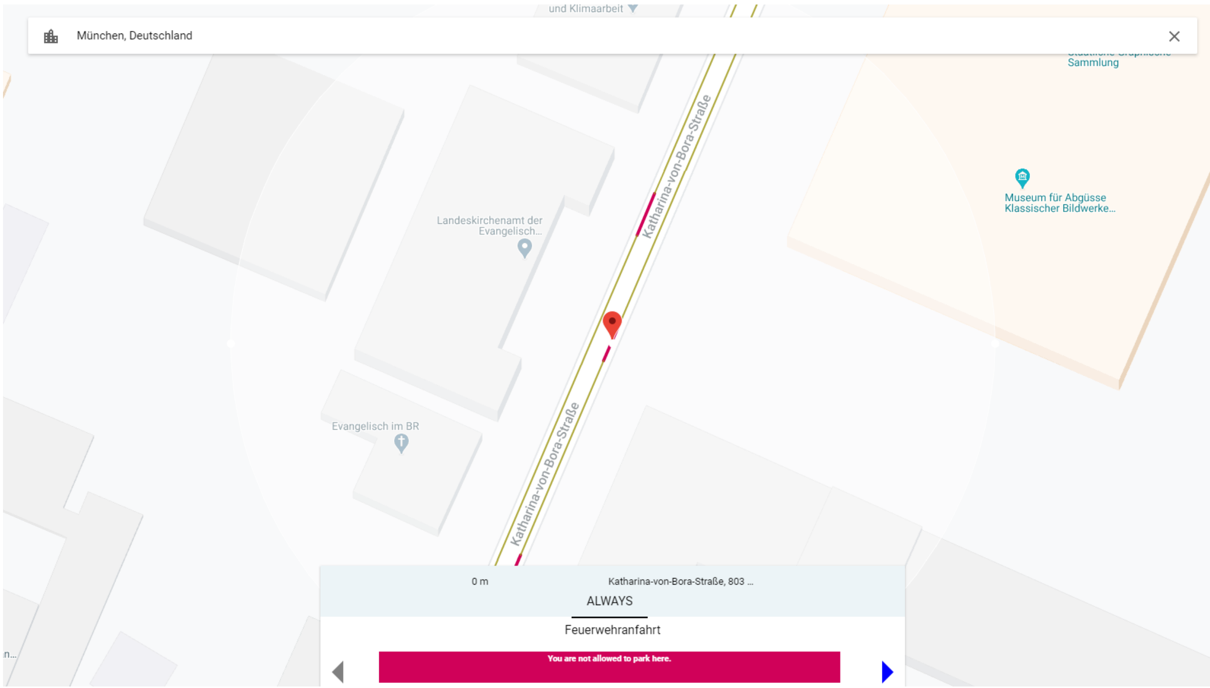


Fig. 52: Screenshot of the Parkbob application in Munich. Captured 2020-02-17

The reason for using the Parkbob app is to know what parking rules apply on the street sections. The consumers of the app will not be able to define the exact length of the sections, therefore it is enough to show that there are restrictions within a certain area, to give the user the possibility to compare the information with the reality. In Fig. 52 a street section in Munich, including two FRPs is shown. The selected FRP has an estimated length of five meters. However, imagine being a user of the app, this information is not very important, as I have to check the reality anyway. It would not make a difference if the FRP is 5 meters or 7 meters long.

On the other side, it makes a big difference for the users if the FRP is way longer as shown in Fig. 53. The selected FRP is estimated to be 50 meters long, so the whole street section is no parking area. However, also in this case it is not important if the FRP is exactly 50 meters long, but the user gets the idea if only parts of the street sections are no parking area, or if the whole street is a no parking area.

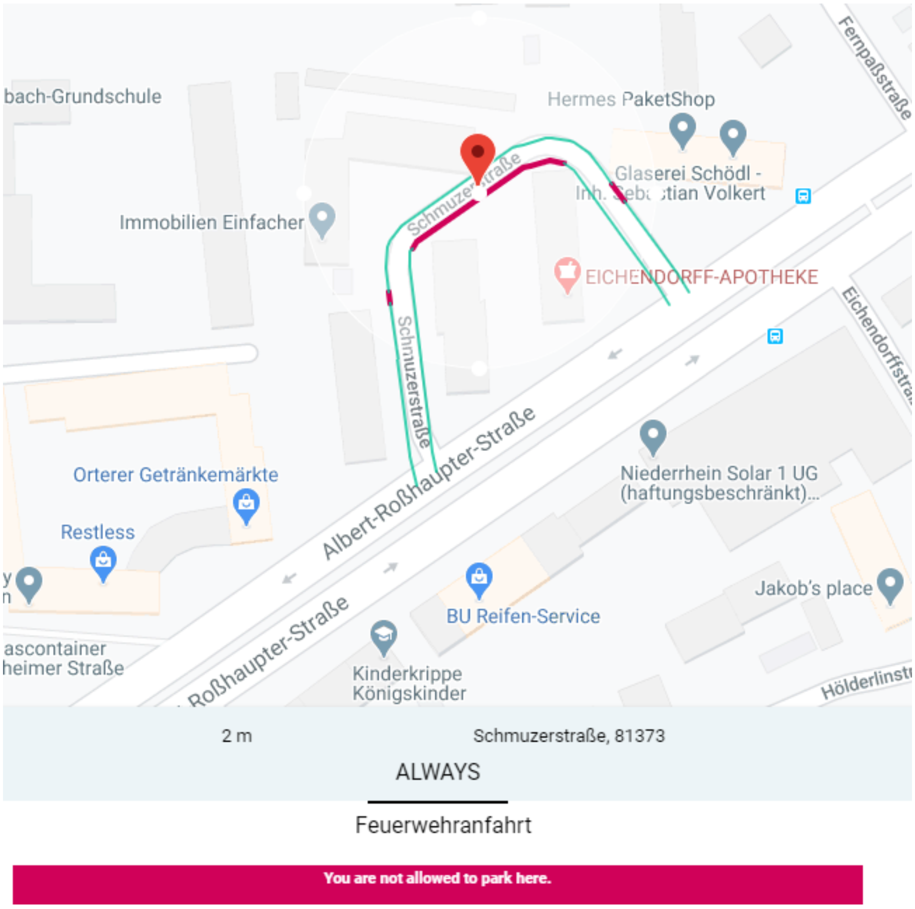


Fig. 53: Screentshot of a FRP in Munich in the Parkbob app. Captured 2020-02-17

I expect the users to be able to estimate the length within a tolerance of two meters. However, I was not able to test the contributors in advance if they can assume the right length while standing in front of a FRP. The attribute check of the submissions was made manually and went along with the manual positional check. This means the estimated length was compared to the submitted picture and to the imaginary from an online service. With those information I was able to evaluate the attribute accuracy within the mentioned tolerance.

5.8. Contributor Quality

One of the quality indicators is the contributor quality. In can be separated into the users' credibility and the users' motivation. The first group includes methods which are based on the categorization of the contributors. In chapter 3.2 I already explained the way Neis & Zipf (2012) defined user groups based on their experience (Senior Mappers, Junior Mappers, Nonrecurring Mappers and contributors with no edits). However, I ask myself, if the number of created nodes comes along with high quality. Besides the experience also the history influences the credibility of the user. History represents contributions of a single user that was already crosschecked. As a separate indicator, the motivation of the contributor should be concerned as well. These three approaches can be mixed to create an overall rating for the users: experience, history and motivation.

5.8.1. Experience

The experience of a user is measured on the amount of work that happened through that user so far. It is not necessarily the quality of the work, but the knowledge and experience that the user gained from a certain number of submissions. It is expected, that with the increasing number of submissions, the quality rises, so there is a correlation between the contributors' quality and his or her experience in contributing to the project. (Exel, Dias, & Fruijtier, 2010)

When using this method for the use case, the recommended categories of users from Neis & Zipf (2012) would result in only one *Senior Mapper*, six *Junior Mappers* and 51 *contributors with no edits*. Therefore, I decided to adapt the numbers for this project to get a better insight in the consequences of the experience rating. I changed the limit for the *Senior Mappers* to 400, so the users of the project split up more equally.

	Nodes according to Neis & Zipf (2012)	***Number of users***		***Nodes needed for this project***	***Number of users***
"Senior Mappers"	>1000	1		>400	3
"Junior Mappers"	10 - 1000	6		10 – 400	4
"Nonrecurring Mappers"	< 10	0		< 10	0
contributors with no edits.	0	51		0	51

Table 13: Experience Level of the Users

Based on this evaluation, the project team can decide how the Senior Mappers get rewarded. Senior Mappers get basically higher reputation which can result in great trust and automatic approval of further submissions.

5.8.2. History

The history is defined by the value of the contribution of a user so far. Based on the manual evaluation, the user gets a certain status which allows the project manager to distribute the resources of crosschecking more effective. The performance of a user can be considered when estimating the future quality of the submissions of the individual user. This means, if a user's contributions had a certain quality, I expect the quality to stay the same level for future contributions as well.

To use the history of the users I will take the first 20 submissions of each user to determine his or her credibility. I calculated the percentage of the correctness of these 20 submissions (see Table 14) and then I can compare the numbers to the percentages from all submissions of the corresponding user (see Table 15).

Username	Contributions	Approved	Dismissed	Approved in %
Azu	20	20		100%
Bah	20	15	5	75%
Emr	20	13	7	65%
Fru	20	19	1	95%
Lei	20	19	1	95%
Sam	20	13	7	65%
Ser	20	13	7	65%
Total	**140**	**112**	**28**	**80%**

Table 14: User rating based on the first 20 submissions.

Username	***Contributions***	***Approved***	***Dismissed***	***Approved in %***
Azu.	2044	1847	197	90%
Bah.	185	139	46	75%
Emr.	52	21	31	40%
Fru.	452	366	86	81%
Lei.	73	56	17	77%
Sam.	362	339	23	94%
Ser.	485	338	147	70%
Total	**3653**	**3106**	**547**	**85%**

Table 15: Total user rating

First step is to define a limit which a user needs to reach to be rated as trustworthy. In this case I established two categories: the basic and the superior mappers. All contributors start in the first group. Every submission from users in this group needs to be crosschecked. When the level of correctness reaches 90%, the individual user gets promoted to the superior group. I define the minimum number of submissions to be rated at 20. This number is chosen randomly, however the number needs to be low enough, so resources of the project can be saved, and high enough to get an impression of the contributor. The ideal case is, that users in the superior group do not need any further crosschecking and all their submission count automatically as valid. The limit of 90% can vary from project to project and is not necessarily fixed during one project. The project manager can define the number at 80% first, and then decide that the number can even be higher and change the expectations of the quality.

Username	**% first 20**	**% total**
Azu	100%	90%
Bah	75%	75%
Emr	65%	40%
Fru	95%	81%
Lei	95%	77%
Sam	65%	94%
Ser	65%	70%
Total	**80%**	**85%**

Table 16: Comparison of the user evaluation.

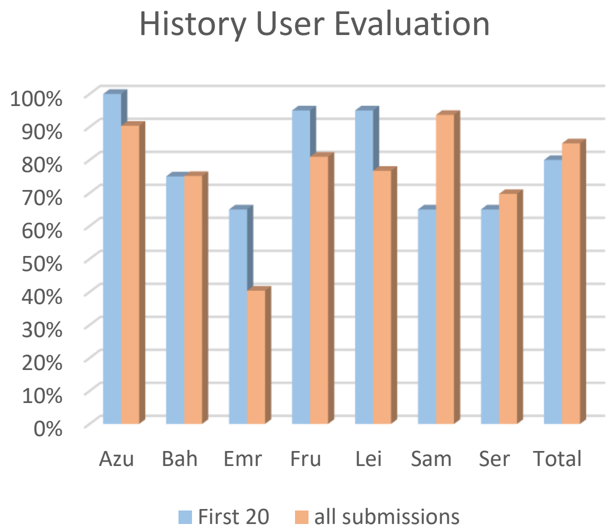


Fig. 54: History comparison.

In Table 16 all contributors who reached the 90% limit are coloured in green. After 20 submissions three out of seven people got promoted to the superior group. One further user reached the goal of 90% when all submissions are concerned. However, two users could not keep the quality of the all over the project. When only accepting a correctness of 90% after 20 submission and trusting these three users, the resources for crosschecking would have been reduced by 70% (see 6.3). As you see, the problem is, that the quality of the contribution did not stay continuously high. Two out of three users did not keep their level of quality during the project. Consequently, more reviews during the project would be necessary to track of the quality and to avoid ending up with a very low quality.

5.8.3. Motivation

Lendák (2016) identified several motivation mechanisms when it comes to crowdsourcing applications:

- appealing to our built-in altruistic nature (i.e. it is natural to people to try to help others)
- gamification
- in-app games
- social networking
- financial motivation

Altruism as the main factor of crowdsourcing is especially true for crowdsourcing apps which specialized on natural hazard (e.g. Ushahidi) but also routing apps like Waze, where users can help other to avoid traffic jams or police fines. Chandler & Kapelner (2013) found out that the workers effort could be increased significantly by increasing the meaningfulness of its work.

Gamification is a way of letting the user feel like the app and submitting information is a game, as you have to fulfil tasks and get rewarded. **In-app games** motivate the user to stay longer within an app and the possibility to interact with their peers is the **social networking** method.

The **financial aspect** is also the one used in this use case. It assumes that people see crowdsourcing as a kind of work. As the collaboration of employee and employer will only continue if the employee submits work in an expected quality, the employer might have little trust issues when it comes to quality. Many platforms use more than just one motivation to increase the work of their contributors like gamification and social networking.

The motivation in the use case is from financial nature. None of the other mentioned motivation indicators mentioned before were applied.

5.8.4. User overall rating

The overall rating of the users is based on two parameter: the experience and the history or the quantity and the quality of the contributions. When it comes to the quantity of a users' activity, 400 is the amount of submissions that is needed to reach the break point, to become a senior mapper. For this project, this means that submissions from senior mappers do not longer need to be crosschecked as we expect the quality to be on a continues high level within this group.

The quality of the submissions is calculated two times: after the first 20 submissions of each user and after the project has finished and all submissions could be concerned. The "trust index" of the user has a range of 0 to 1, where 1 express the highest rank a user can achieve and 0 means the user is registered but did not contribute in any way.

I first thought about how to calculate a rating based on these to parameters. I wanted the result to help to determine when crosschecking is necessary and when the trust in a user is high enough to abstain from it. I will call this group in the following the "trust group". To be rated as trustful, the user must reach at least 0.90. If you are a "Senior Mapper", the quality of your submissions needs to be under 0.9 to let you fall out of the "trust group". The rating for the quality is based on the percentages itself expressed in an index that has a range from 0 to 1.

To express the experience of the user, I also used an index from 0- 1. I decided to give "Senior Mappers" the highest possible value which is 1. For "Junior Mappers" I decided to go with 0.85. The reason is, that if the user submits more than 10 contributions, the user is already very close to the desired threshold of 0.9, however to belong to the "trust group" the quality rating needs to be at least 0.95, so slightly higher than if you only concern the quality of the user.

To sum it up, the overall rating is based on following ideas:

- If you are a "Junior Mapper" you need to show higher quality than if you are a "Senior Mapper" to belong to the "trust group".
- If you are a "Senior Mapper" your level of quality can be lower as you compensate it with quantity.
- Even if you submit a lot of points and become a "Senior Mapper", if you don't take care of the quality and your quality rating is below 0,78, you will not belong to the "trust group".

When having a look at the situation after the first 20 submissions of each user shown in Table 17, three people reached the threshold of 0.9.

Username	% first 20		Experience		Rating
Azu	100%	1.00	Jun. Mapper	0.85	**0.93**
Bah	75%	0.75	Jun. Mapper	0.85	0.80
Emr	65%	0.65	Jun. Mapper	0.85	0.75
Fru	95%	0.95	Jun. Mapper	0.85	**0.90**
Lei	95%	0.95	Jun. Mapper	0.85	**0.90**
Sam	65%	0.65	Jun. Mapper	0.85	0.75
Ser	65%	0.65	Jun. Mapper	0.85	0.75

Table 17: User rating after 20 submissions.

As we only go for the first 20 submissions all users get the experience rank "Junior Mappers". Here the combination of the two parameter shows clearly how the ideas that I have written above work out: User Azu reached the threshold easily as the quality was on a very high level. Also, two other users reached the "trust group". However, the main part of the users had lower quality rating and thus did not reach the desired 0.9 due to the lack in experience. Of course, this approach seems to be irrelevant as all users belong to the same experience group. However, this should demonstrate how a user rating can work and has to be adapted individually for every other project.

Username	% total		Experience		Rating
Azu	90%	0.90	Sen. Mapper	1,00	**0.95**
Bah	75%	0.75	Jun. Mapper	0.85	0.80
Emr	40%	0.40	Jun. Mapper	0.85	0.63
Fru	81%	0.81	Sen. Mapper	1.00	**0.90**
Lei	77%	0.77	Jun. Mapper	0.85	0.81
Sam	94%	0.94	Jun. Mapper	0.85	0.89
Ser	70%	0.70	Sen. Mapper	1.00	0.85

Table 18: User rating total.

In Table 18 I want to show how the overall rating was at the end of the project. This way the reliability of the user rating after the first 20 submissions can be confirmed. We see that two out of three users could keep their rank in the "trust group". Only one person – Lei - could not obtain its quality. In fact, the quality rating dropped from 0.95 to 0.77, so decreased by 17,1%. This is very unusual, as it is expected, that the submissions of a user obtain higher quality with the experience a user gets. We also see that Ser did reach the "Senior Mapper" group, however he was not able to increase the quality of his submissions, so he did not level up to the "trust group".

5.9. Methods Summary

For this use case, not all the described quality indicators are important. In chapter 3.3 I summarized those quality indicators which are relevant for crowdsourced spatial data in general. However, it is always necessary to see the context of the data. Here, for this use case the logical consistency and lineage are already predefined due to the software environment, therefore the QA happens already in the prevention phase. The user does not have the possibility to change the data structure as the user does not have access to the data once it is submitted and a submission is only valid under defined circumstances. Completeness and positional accuracy are very important indicators which were checked on the previous pages with more than just one approach. As the project was only open for a specific time range and is already closed, the temporal accuracy QA also happened already during the prevention phase. The attribute accuracy needs to be checked as well, as the user is asked to define the length of the FRP. This is the only semantic information where the diligence of the user is important. The usability is already clear in advance, as it is mandatory that the data submitted has the necessary information. The motivation of the author can be neglected, as all the users were found via job platforms, so the intention was in every case to earn money. Last but not least, the credibility of the users is measured by a combination of experience and history to include not only quantity but also quality into the QA.

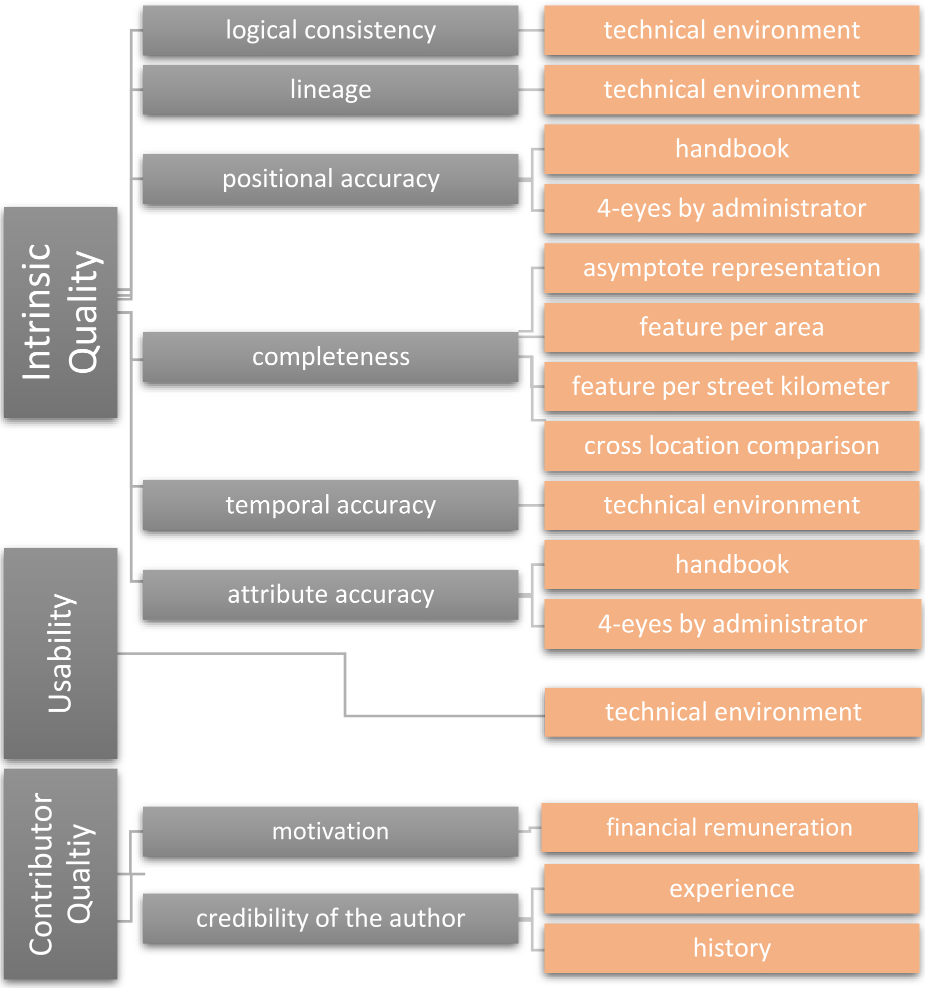


Fig. 55: Methods of crowdsourced spatial data QA.

6. Results

After using several of the methods in the previous chapter, I want to show here how the different methods result in terms of time and accuracy. Some of the aspects of quality which I talked about in chapter 3 are not relevant for this use case. As already mentioned on the previous pages, the database and the number of contributors is relatively small. Still this thesis gives an important insight how QA can be adapted and how it can be done with a dataset that has individual characteristics of crowdsourced geodata.

6.1. Completeness

When it comes to completeness, the intrinsic approaches were confirming the intentional thought about the dataset which the project team used in first place. The result of the *asymptote representation* goes along the idea of how the contributions are distributed over time in a crowdsourcing project. The decision about stopping the project was taken due to the same issues which I then mentioned in chapter 5.5, but only by reviewing the data situation without strengthening the idea with numbers. This means the analysis supports the intention of take down the project at the right time. Nevertheless, the result is not absolute and without any reference dataset this would also not be possible to determine absolute numbers about the completeness of the dataset

The additional comparison with another city would have been interesting, but as it seems, Munich is not the only german city without a dataset about the FRPs.

The methods of *the number of features per area* approach gives a clear first insight in the spatial distribution of the submissions. This approach can be a good automated process to identify regions where the density is not as high as in other locations. However, due to the diversity in the city structure it is not a satisfying result without a manual review. The adopted version *number of features per street kilometers* overcomes the problem with the heterogenous city structure. However, it comes to a similar result as the initial approach. It is possible that this approach becomes a fully automated process in future projects, as the setup is established in a decent time range and the results allows the project team to have a structured insight in the completeness of the dataset.

6.2. Positional and attribute accuracy

The attribute and the positional accuracy were both analysed the same way. By using the many-eyes principle almost all mistakes could be identified, recorded and then – if possible – eliminated. In the following figure is an overview of all errors that could be revealed during the QA. The most common mistake was that the FRP was positioned on a wrong location. In most cases, it was possible to correct this mistake. The second biggest error was that submissions were made on locations where we already had FRPs. At this point of the project, we didn’t want to deal with duplicate submissions as it would have caused to much effort to merge those contributions into one single datapoint. Other mistakes which happened rarely were submissions that were actually no FRP (happened 4 times), submissions that were taken outside of the study area (5 times) and submitted pictures that were not useful (3 times). In Fig. 56 the percentage of each mistake is shown. On the left diagram you see the total amount of submissions, split up into correct records and those records where errors could be identified. Approximately 15% of all FRPs were uploaded without sufficient quality in positioning and attributes. Almost 60% of those errors were location mistakes where nearly all of them could be updated by the project team.

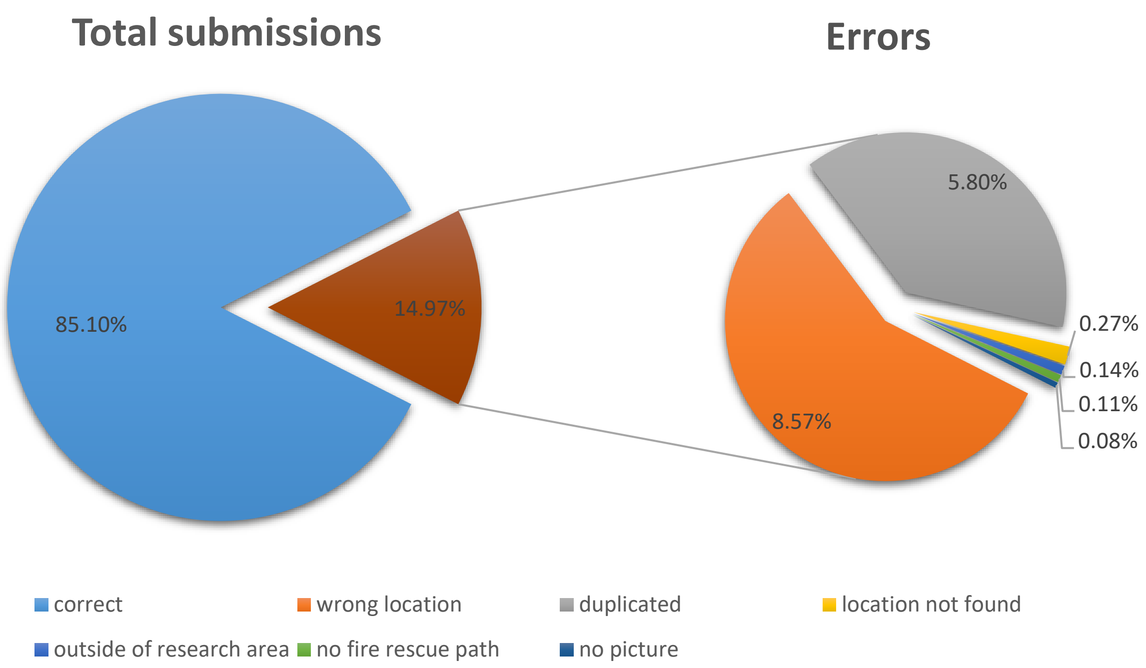


Fig. 56: Errors overview.

MISTAKES	NO	%
wrong location	313	57.22%
duplicated submissions	212	38.76%
location not found	10	1.83%
outside of research area	5	0.91%
no picture	3	0.55%
no fire rescue path	4	0.73%
total	547	100.00%

Table 19: Mistakes statistics.

The main mistake was that the location was not pinned correctly which makes 59.05% of all errors, including the wrong locations that could be corrected (57.22% of all mistakes) and the locations which could not be found (1.83% of all mistakes, 0.14% of all submissions). For the project team this is a very important information, as they can focus on this problem when it comes to improvement of the project environment. As duplicated submissions made 38.76% of all mistakes, it is very interesting, why crowd users did not realize the existence of those points while they were submitting. All other mistakes happened only occasionally and have therefore less priority when it comes to improvement of the project. To the group of wrong attributes, I count those records without a useful picture and those points which did actually not represent a FRP. It never happened that the project team realised a mistake when it comes to the length of the FRP. This means the attribute accuracy is at 99.19% as only 7 submissions in total have been mistakes concerning their attributes. This makes 1.28% of all mistakes and 0.19% of all submissions.

6.3. Contributor Quality

Based on the user rating, the project team can decide, if the positional accuracy and the attribute accuracy can be trusted and no further manual comparison needs to be done. I want to demonstrate which level of quality can be achieved by using the user rating. The key element here is the contributor quality, because it allows to decide if certain submissions can be trusted or if they need revision.

In the user overall rating in chapter 5.8 I determined three crowd user who belong to the "trust group". The meaning of this group is, that their submissions are very likely to have a high positional and attribute accuracy and can thus be accepted without crosschecking. To demonstrate how the trust group would influence the QA in terms of time and the quality level if the total result, I compare the dataset in the following table.

Username	***Contributions***	***Approved***	***Dismissed***	***Approved in %***	***Time expense Seconds***	***Time expense Minutes***
Azu.	2044	1847	197	90%	100717	1678.62
Bah.	185	139	46	75%	9116	151.93
Emr.	52	21	31	40%	2562	42.70
Fru.	452	366	86	81%	22272	371.20
Lei.	73	56	17	77%	3597	59.95
Sam.	362	339	23	94%	17837	297.29
Ser.	485	338	147	70%	23898	398.30
Total	**3653**	**3106**	**547**	**85%**	**180000**	**3000.00**

Username	Contributions	Approved	Dismissed	Approved in %	Time expense Seconds	Time expense Minutes
Azu.	2044	2044	0	100%	0	0.00
Bah.	185	139	46	75%	9116	151.93
Emr.	52	21	31	40%	2562	42.70
Fru.	452	452	0	100%	0	0.00
Lei.	73	73	0	100%	0	0.00
Sam.	362	339	23	94%	17837	297.29
Ser.	485	338	147	70%	23898	398.30
Total	**3653**	**3406**	**247**	**93%**	**53414**	**890.23**

Table 20: Submission overview with concern of the "trust group" with time expense. Rounded numbers.

Table 20 shows the difference between using a "trust group" and without using it. The upper part shows the needed time to make the QA for each user. The lower part of the table shows how the time spend would change, if some user would be in a superior group and all corresponding submissions are accepted without crosschecking. As we would accept all submissions from the *trust group* without review, 100% of the FRPs would be rated as approved. The total amount of approved submissions would increase by approximately 8% up to 93.26%. The number of hidden fails rises to 300, which is 8.21% of all submissions and even 54.84% of all failed submissions. These numbers look on the first sight very frightening, however it is necessary to put them in relation to their use. Around 80% of the undetected wrong points belong to one user, Azu. As this user submits the main part of the contributions, it is obvious that with the increasing amount of submissions also the total amount of failed points increases. I already explained in chapter 2.3, that the total time expense of crosschecking the submissions sums up to approximately 50 hours or 3000 minutes. With a total amount of 3653 submissions this would mean the time expense for each submission was 49 seconds. When we compare the time expense, we see that it drops down to 890.23 minutes This is a decrease of 70.64% or approximately 35 hours. This means by decreasing the time expanse by 70.64% we will not be able to determine 300 wrong points. This would lead to a dataset with level of quality of 91.79%. Consequently, the expected quality threshold of 90% could be reached with far less effort. Table 21 gives an overview of how the trust group would change the data quality and how much time would be saved.

	without trust group manual crosschecking	**with trust group**
submitted data	3692	3692
detected errors	548	248
time expense	50h	15h
dataset quality level	99%	90.79%

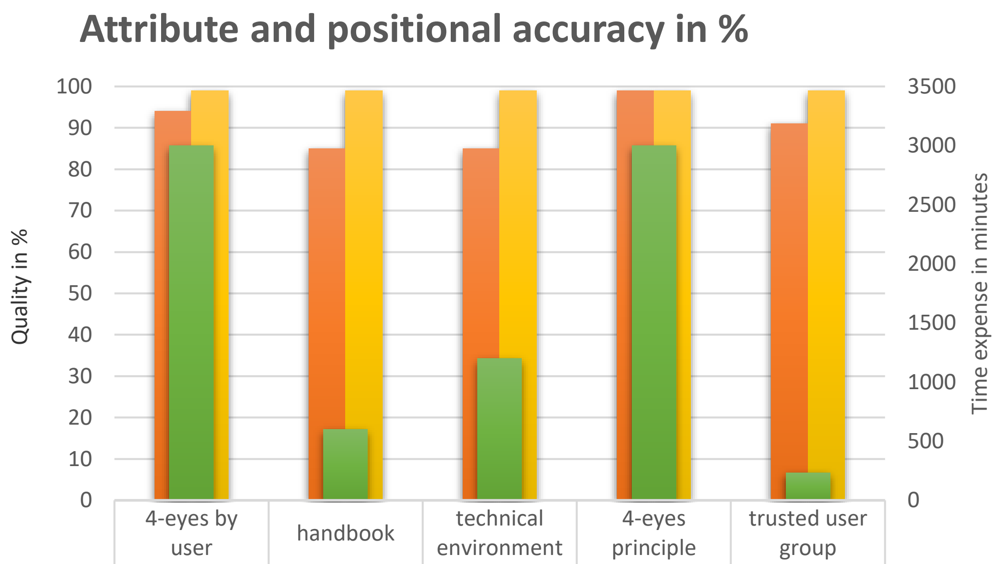
Table 21: Comparison of the dataset with and without trust group.

The reason why the dataset quality reaches 99 % when all submissions are checked is, that then the detected mistakes can be corrected or – if correction is not possible – deleted from the dataset. The last 1% include those mistakes which have been overseen. This number can only be estimated and of course we can also expect a dataset with 100% correct records, however my experience leads to the conclusions that expecting some mistakes of human failure is closer to the reality than not. To achieve 90.79% with the trust group, I calculated the 99% - 300 undetected mistakes (8.21%).

The last part of the user contribution is the influence of the motivation into the quality of the dataset. Unfortunately, we do not have the chance to take a closer look in different approaches of motivation, as the project is already closed and therefore a use case specific research is not possible.

6.4. Comparison of the Methods

As one goal of this thesis is to help other crowdsourcing projects to decide on possible methods for the QA, I use this chapter to compare the results in terms of reached quality and time expense. However, it is important to mention, that in case of prevention methods, they always have an impact on the post-collection methods like the 4-eyes method. There are of course cases when a technical environment or a handbook were not created for the users, as already existing crowdsourced data is used, for example when data is taken from the OSM project. As I want to focus on the FRP project, this is not the case here. I included the methods of analysing the quality of the attribute and positional accuracy, prevention and of the contributor. In Fig. 57 I want to show how much time was needed to implement each of the methods and effecting the mentioned elements.



	4-eyes by user	handbook	technical environment	4-eyes principle	trusted user group
positional accuracy in %	94	85	85	99	91
attribute accuracy in %	99	99	99	99	99
time expense (min)	3000	600	1200	3000	234

Fig. 57: Dataset quality level of after methods with time expense.

The prevention methods of creating a detailed handbook and creating a technical environment leads to positional accuracy of 85% and an attribute accuracy of 99%. We know these numbers as all records had been reviewed manually, and every mistake has been noted. This means with the 4-eyes method the positional accuracy increased up to 99% as all mistakes that could be repaired were done so. As already mentioned before, the last 1% represents mistakes that could not been identified, these can also be named as human failure which can be expected when working with a manual approach. Only 6% of all submissions were not repairable in terms of positioning and therefore are invalid submissions. As it is possible to delete this points from the dataset, the result is an almost error free collection. When using the trusted user group, the quality of the dataset reaches 92%, with a far less effort when it comes to time issues.

The method of the trusted user needed 234 minutes (see Table 22), including the time for validating the first 20 submissions of each user and the estimated time I needed to setup the user rating.

users	7
x Submissions checked per user	x 20
Submissions total	**140**
x time needed to check one submission	x 49
Time needed to check 140 submissions	**6860s (114.33 minutes)**
+ Time needed to set up user rating	+ 120 minutes
Time needed for the trusted user method	**234.33 minutes**

Table 22: Calculation of the time expense for the trusted user method.

On the first glance, all methods lead to a similarly high level of quality. The differences in the prevention approaches concerning attribute and positional accuracy are in my opinion subject to the simplicity of determine the length of something contrary finding the right spot on a map. The project team underestimated how hard it is for non-GIS-experts to read a map properly in times of navigation systems. It has been analysed that people with geographic background like Geography students have a much better performance when reading maps than others. (Ooms, et al., 2016)

As the whole project team consists of geography experts, a view from a different perspective on the setup could have prevented the dataset from positional mistakes. The main statement that can be deducted from Fig. 57 is that the 4-eyes principle gives the most trustful results, but also with less time intense methods a relatively high level of quality can be reached.

7. Discussion

After the theoretical research and the application of different methods of QA on the FRPs project, I want to sum up and discuss the results and the findings. In the first part I will talk about the conclusions I draw from the results of this thesis. Second, I will answer the research questions I asked in the beginning of the thesis. Finally, I will think about future work and which topics have the potential to become an object of deeper research.

7.1. Conclusion

The question about the quality of spatial data has already been discussed a lot within the literature. Actually, I realized just during writing this thesis, that this is a complex topic which has probably as many variations as use cases exist. With the additional aspect of the creator of the geodata, it became even more extensive and reaches out over QA of conventional data by far. There is not one quality. The quality of spatial data has several aspects which all result in a degree of quality which meets the specific need or not. This can be very confusing in first place as different aspects of quality need different kinds of methods for QA.

Based on the results on the previous chapter, there is one thing which really stands out to me: the contributor quality. Intrinsic QA is depending a lot on the credibility of the contributor in this case. It seems like to do QA with intrinsic methods only comes to its limits, when the number of contributors is low. Especially when there is no history of the data itself which can be concerned. However, the time saved when increasing the level of trust in specific users is tremendous. Moreover, when they are able to monitor submissions from other users, QA could be easily scaled up to other cities as the time expense for the project team is limited to the setup of the project and not to its validation.
Crowdsourcing projects are known for the big number of contributors. In this case, there were seven different people who submitted data. Unfortunately, this amount seems to be too low, however even here the methods worked out well and as you can see in chapter 2.2, also a small group shows characteristics of a crowdsourcing group. Admittedly, a higher number of crowd users would possibly lead in a more precisely result. However, I think I, Parkbob and other readers can still learn how the mechanics of the QA work and that also a small group of contributors can offer useful data.

Also, I must admit, that a financial remuneration looks nice in first place, but is not a guarantee for high quality. I would recommend thinking about additional motivational systems like increasing the meaningfulness. In this use case, for example, I expect that with telling the contributors that the fire brigade of Munich would profit from this dataset and therefor all citizens of Munich as well, contributors would have been more cautious when working for the project. This would lead to a higher quality of the submitted data without crosschecking and could possibly reach the threshold of 90% without additional editing by the project team. Therefore, this approach should be concerned by future projects.

By analysing the mistakes that were taken by the crowd during the FRP project, I found out that the main error was located in positioning accuracy. For me this means that in future project this has to be in focus of the team, to establish an environment which leads to a minor number of mistakes in this aspect. I already mentioned that the reason for the high numbers of wrong positioning could be, that non-GIS-experts are often unfamiliar to use maps in this kind of context. Modern technology allows us to use maps passively, getting directed by navigation systems without being aware of the environment we are surrounded by. To use the map actively in a city area happens very rarely nowadays. Therefore, this could lead to an overload of information which the users cannot handle. A more detailed handbook, learning lessons and screenshots are possible prevention methods to make the users more familiar with the map. Another possibility would be to search for a crowd which is already comfortable with spatial data. This can happen in certain forums, for example on OSM related websites or with an explicit remark in the job description.

The second biggest mistake was that a submission was made from a FRP which has already been mapped before. It was clearly explained to the users, that if a FRP has already been submitted, they won't get paid for an additional submission of the same FRP. Anyway 212 times which make 40% of all mistakes have been in that kind of way. When a point is submitted, it becomes immediately public for all users. Therefore, it is not so clear why duplicated submissions happened so often. An explanation could be, that the users did not recognize that there was already a point on the same location inside the map. Another possibility is a combination with a wrong positioning, so the users thought that the existing point is on a different location. To reduce the amount of duplicated submissions, the theoretical training can be extended with a more intense focus on the existing points. As Parkbob clearly explained in the handbook that duplicated submissions will not be financially compensated, the consequence of working without remuneration was not very effective.

7.2. Answering Research Questions

The overall goal of this thesis is to answer the research questions defined in chapter 1.1. To do so, literature research was needed to collect different methods which have already been discussed theoretically as well as in practical examples. After analysing the use case, I could filter the methods and determine which I could use for the FRP project. The practical application of the methods to the FRPs project gave the final results, so I could then answer all the research questions.

7.2.1. What are measurements of the quality of crowdsourced spatial data?

The first question I wanted to answer with this thesis was the one about the quality measurements for crowdsourced geodata. I analysed the literature and defined in chapter 3 the aspects of geodata which need to be considered. I realized, that crowdsourced geodata has the same issues when it comes to quality measurements as traditional geodata, but there are some additional factors which cannot rate the quality of the data directly: the proxy measures or the indicators. So, besides the intrinsic quality and the usability, the contributor of the data influences the quality as well. All three categories include several measurements, I extracted them from different readings and displayed them all in the following figure.

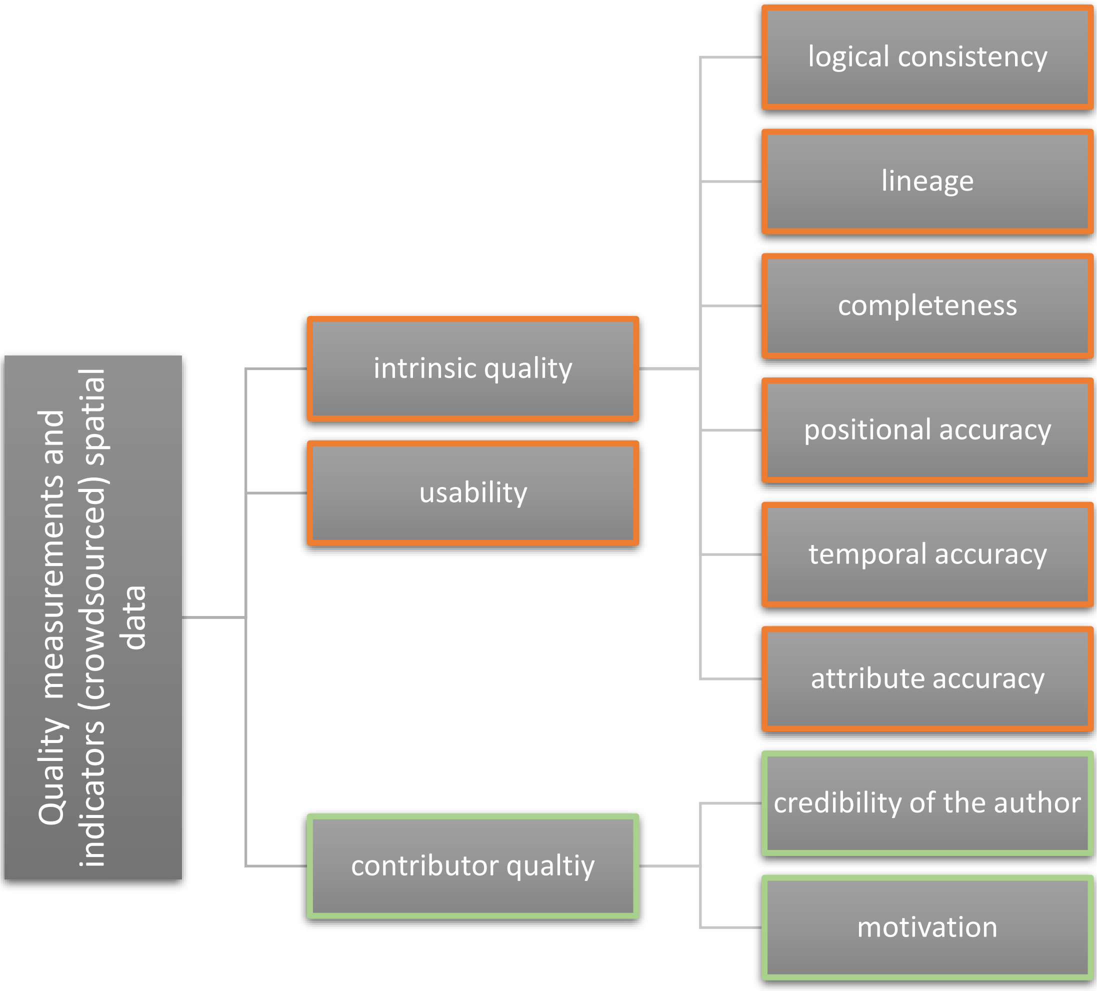


Fig. 58: Quality measurements (orange) and indicators (green) of (crowdsourced) geodata.

7.2.2. What methods of QA of spatial crowdsourced data exist?

There are numerous different methods to do the QA for crowdsourced geodata. The categorization of the collection of methods can done in several ways. The obvious one is, when the methods are allocated to each quality element as I did in chapter 5.5 to 5.8. Some methods will be named more often, as they can be used for more than just one element. The second possibility is to categorize the methods like Criscuolo et al (2016) based on the questions when? who? and what? Both options might result in redundant information as methods do not always fit into only one of these categories. All methods that were used for this project were displayed in Fig. 55 on page 66. However, I also discussed several other methods which were not suitable for the FRP project, still they are methods which might be useful in a different environment. Therefore, I decided to make an overview (see Fig. 59) of all methods which I came across during my research to this thesis. I explained all of them in chapter 4 and chapter 5. Here, also those methods are included which use a reference dataset for the QA. The methods in the first column are suitable for most elements of crowdsourced spatial data. The second column includes all methods concerning the completeness of a dataset only and the final column contains all methods to analyse the credibility of the users.

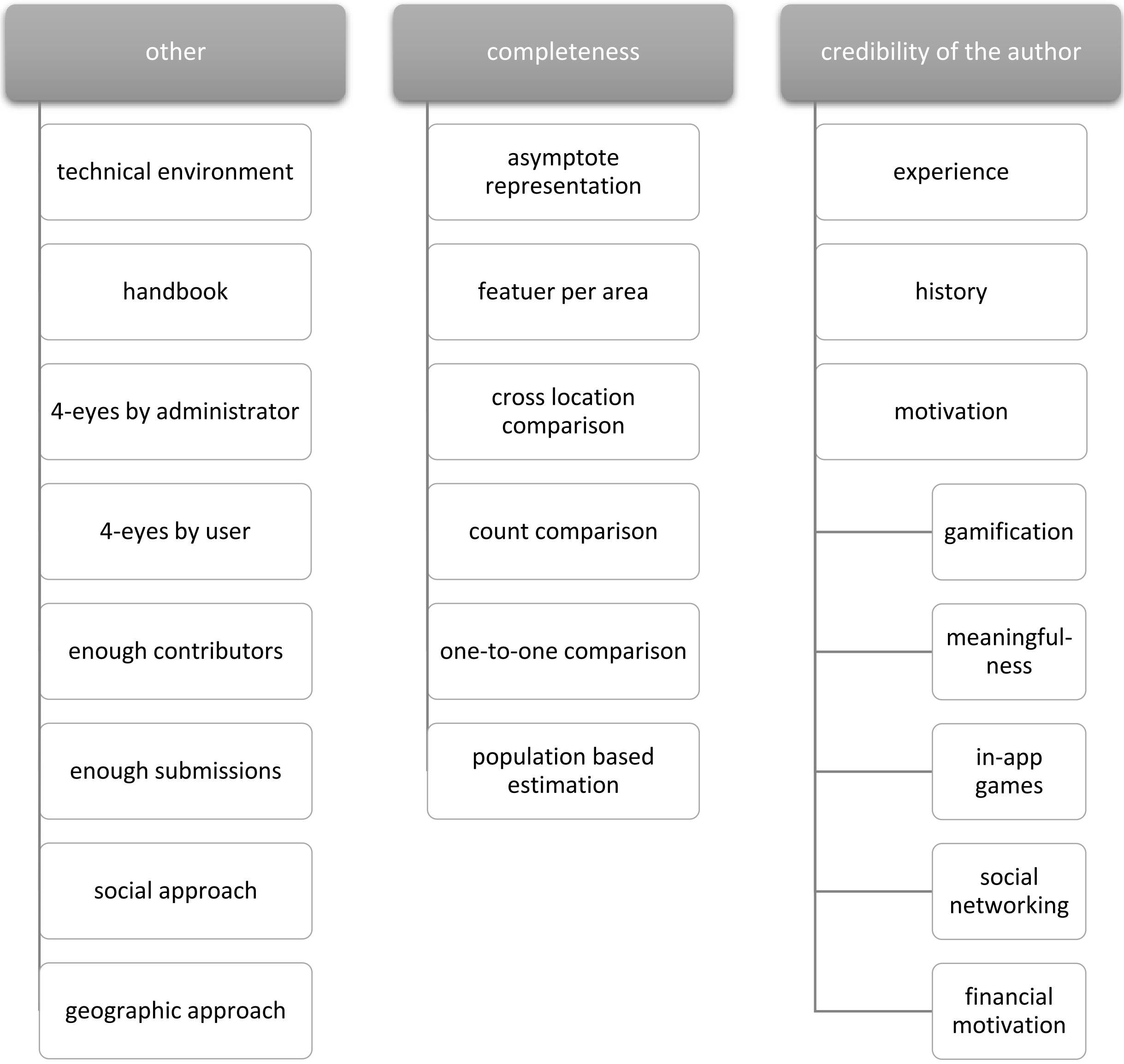


Fig. 59: Methods overview.

7.2.3. Which QA methods are scalable?

If a method is scalable is very much depending on the project and on the goal of the QA. For the FRP project, I want to have a dataset which reaches a level of quality of 90%. This thesis aims to find ways, to make sure similar projects in other cities can be implemented and the quality of the data reaching at least the threshold of the desired 90%. How can Parkbob make QA of several datasets parallel? The manual method of crosschecking each record is very time intense and does not allow to do several projects at the same time. By analysing the users' quality, some of the contributors would not need any further revisions when a small percentage of mistakes is accepted. This approach can help to scale up

crowdsourcing and maintain a certain level of quality. Still this approach cannot happen solely automatically. Random samples help to make sure the quality is kept up and (semi-) automatic approaches for checking the QA of completeness can help to redirect the contributors to specific areas. Analysing the contributor quality is not only helpful by reducing the manual crosschecking by the administrator. Knowing which crowd user submits higher quality can help to decide on other methods of QA by the contributors. For example, can the administrator decide if those user with higher ranks get superior tasks, higher remuneration or – as explained in chapter 5.8 – they become trusted users.
Based on the methods used in this project, using an accurate technical environment and a detailed handbook does the first step to an accurate dataset. The user rating will help to increase the quality of the dataset without a high investment of time. Therefore, those three approaches have the ability to be scaled up to several projects by keeping the needed time at a low level.

7.2.4. Which is the cheapest method of QA for crowdsourced data?

To answer this question, I want to clarify that 'cheap' is not always an expression of a monetary issue. As people get paid different in different projects and not every kind of work has the same financial remuneration, comparisons in monetary terms can be very hard. As payment and time is often related, I prefer to change the question to 'Which methods is the least time intense?'. Also, here, we have to make a difference what time means. Is the solely workforce with the invested hours meant or the time span of the whole project? The amount of working time will however stay the same, but the overall time span would decrease be enlarging the workforce. This means, the more contributors I have, the faster the project is finished. Here, it is always important, that the aim is, that crowdsourcing can be expanded without increasing the traditional workforce too much. Therefore, it is obvious that every automated process will decrease the time spend on QA when scaling up crowdsourcing. Consequently, I can build up a relation between scalable processes and cheap methods. To answer the question: all prevention methods and the definition of the contributor quality have a high impact in the time of the project team spend on the QA and therefore do those methods decrease the monetary investments. Moreover, automated process can reduce the overall time span and have impact on future projects as well, as once the setup for an automatization is done, it will be reusable perpetual.

7.2.5. Which method of QA gives the most trustful result?

The method of the *many-eyes* principle is the one method which could provide the most trustful results in this case. As every single contribution is checked, almost all mistakes can be detected. Only those errors are left, that were undetected due to human failure. In my opinion it is not important by whom this method is made. Administrator as well as other users can make mistakes and oversee errors. It could be argued that crosschecking by the administrator team has the advantage of high expertise of the project and therefore it would lead to a better result. When it comes to scaling up and implement other projects in other regions, it will be necessary to use the crowd for the crosschecking, if the 4-eyes method is desired. This method can also be improved by not only crosschecking once but increasing the number of eyes on data would increasing also the detection of mistakes. However, I want to mention that also the prevention methods lead to a dataset with 85% correctness in position and attributes. Together with the user rating, this approach can be a far cheaper and less time intense alternative with acceptable results.

7.3. Outlook

After all the research and the finding for this thesis, a closer look into the applied methods on other projects would be on high interest. Also, a broader use of the methods, in different a environment and other studies would increase the general usability of crowdsourced spatial data.

Personally, I would expect a detailed research about the different approaches of motivation for crowd users. As my guess is, that the growing gig economy will also influence the world of geodata, this can be a very important topic for the future. As I found some minor approaches to determine the relation of motivation and quality, a detailed statistical approach to evaluate the influence of the motivation is missing.
Moreover, an improved process of the crowd user rating can be helpful for several future projects, not necessarily in a geographic context only. Not necessarily including history and experience only, but also motivation, geographic background or other parameters.

Crowdsourcing projects can be very different in terms of aim, setup and technical environment. I guess there is potential in setting up a modern classification of crowdsourced geodata, to remodel Criscuolos' et. al. (2013) approach, or even establish a new categorisation of types of crowdsourced geodata.
Based on the divergence in crowdsourcing projects, also the way the QA is done can vary in different ways. Especially when it comes to intrinsic methods, the way the data was collected, the data structure and the crowd itself have high impact in the QA. Summarizing, there is a lot of potential in other or new methods of QA of spatial data and in the creative use of crowdsourced data.

Finally, I want to mention that almost all of the methods can also be applied to traditional collected geodata. Especially the contributors' indicators have not been concerned in conventional geodata institutions so far. Including a profile of the responsible cartographer in the metadata, can maybe lead to the possibility to find connections between creator and dataset. In my opinion it would be interesting to use the mentioned techniques not only for crowd users but also for professionals.

8. Sources

Andrienko, G., & Andrienko, N. (2016). Gaining Knowledge from Georeferenced Social Media Data with Visual Analytics. In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, & R. Purves, *European Handbook of Crowdsourced Geographic Information,* pp. 157-167.

Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: an overview. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 345-351.

Barron, C., Neis, P., & ZIpf, A. (2014). A Comprehensive Framework for Intrinsic OpenStreetMap Quality Analysis. *Transactions in GIS, 18*, pp. 877-895.

Borge, J. L., Moura, A. C., de Paula, P. L., & Casagrande, P. B. (2018). Big-Data to support the study of the environmental disaster in Mariana, MG: From VGI to social media gegraphic information. *Revista brasileira de cartografia,, 69*(8), pp. 1586-1597.

Boutsis, I., & Kalogeraki, V. (2014). On Task Assignment for Real-Time Reliable Crowdsourcing. In *International Conference on Distributed Computing Systems*, pp. 1-10.

Budhathoki, N. (2010). *Participants' motivations to contribute geographic information in an online community.*

Budhathoki, N., Bruce, B., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the user of spatial datainfrastructure. *GeoJournal*(72), 149-160.

Burtch, G., Carnahan, S., & Greenwood, B. (2018). Can You Gig It? An Empirical Examination of the Gig Economy. *Management Science, 64*(12), pp. 5497-5520.

Chandler, D., & Kapelner, A. (2013). Breaking monotony with meaning: Motivation in crowdsourcing markets. *Journal of Economic Behavior & Organization, 90*, pp. 123-133.

Coleman, D., Georgiadou, Y., & Labonte, J. (2009). Volunteered Geographic Information: The Nature and. *International Journal of Spatial Data Infrastructures Research, 4*, pp. 332-358.

Coord. (2020). Retrieved from https://coord.co/

Crall, A. W. (2011). Assessing citizen science data quality: an invasive species case study. *Conservation Letters*, pp. 433-442.

Criscuolo, L., Carrara, P., Bordogna, G., Pepe, M., Zucca, F., Seppi, R., Rampini, A., Oggioni, A. (2016). Handling quality in crowdsourced geographic information. In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, & R. Purves, *European Handbook of Crowdsourced Geographic Information*, pp. 57-74.

Criscuolo, L., Pepe, M., Seppi, R., Bordogna, G., Carrara, P., & Zucca, F. (2013). Alpine Glaciology: An Historical Collaboration between Volunteers and Scientists and the Challenge Presented by an Integrated Approach. *International Journal of Geo-Information, 2*(3), pp. 680-703.

Dai, C. D. (2008). An Approach to Evaluate Data Trustworthiness Based on Data Provenance. *Secure Data Management: 5th VLDB Workshop*, pp. 82-98.

Delaney, D. G. (2007). Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions, 1*, pp. 117-128.

Devillers, R., & Jeansoulin, R. (2006). Spatial Data Quality: Concepts. In R. Devillers, & R. Jeansoulin, *Spatial Data Quality*.

Dickinson, J. L. (2010). Citizen Science as an Ecological Research Tool: Challenges & Benefits. *The Annual Review of Ecology, Evolution, & Systematics*, pp. 149-172.

Elwood, S., Goodchild, M., & Sui, D. (2012). Researching Volunteered Geographic Information. *Annals of the Association of American Geographers*, *102*(3), pp. 571-590.

Exel, M., Dias, E., & Fruijtier, S. (2010). *The impact of crowdsourcing on spatial data quality indicators.*

Fitzpatrick, M. C. (2009). Observer bias and the detection of low-density populations. *Ecological Applications,* pp. 1673-1679.

Fogliaroni, P., D'Antonio, F., & Clementini, E. (2018). Data trustworthiness and user reputation as indicators of VGI quality.

Galloway, A. W. (2006). The Reliability of Citizen Science: A Case Study of Oregon White Oak Stand Surveys. *Wildlife Society Bulletin*, pp. 1425-1429.

Geofarbik. (2020). Retrieved from http://download.geofabrik.de/europe/germany/bayern/oberbayern.html

Girres, J.-F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. In *Transactions in GIS,* pp. 435-459.

GISCloud. (2019). Retrieved from https://www.giscloud.com/

Goodchild, M. (2006). In R. Devillers, & R. Jeansoulin, *Fundamentals of Spatial Data Qualtiy,* pp. 13-17.

Goodchild, M. (2007). Citizens as Sensors: The World of Volunteered Geography. *Geojournal*, *69*(4), pp. 211-221.

Goodchild, M., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial Statistics,* pp. 110-120.

Google Street View. (2019). Retrieved from https://mapstreetview.com/

Haklay, M. (2010). How Good is volunteered geographical information? a comparative study of OpenStreetMap and ordnance survey datasets. In *Environment and Planning B: Planning and Design,* pp. 682-703.

Haklay, M. (2016). Why is participation inequality. In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, & R. Purves, *European Handbook of Crowdsourced Geographic Information,* pp. 35-44.

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? the validity of Linus's Law to volunteered geographic information. *Cartographic Journal, 47*(4), 315-322.

Harding, J. (2005). Vector Data Quality: A Data Provider's Perspective. In R. Devillers, & R. Jeansoulin, *Fundamentals of Spatial Data Quality*, pp. 141-161.

Heipke, C. (2010). Crowdsourcing geospatial data. *ISPRS Journal of Photogrammetry and Remote Sensing, 65*, pp. 550-557.

Howe, J. (2009). Crowdsourcing: why the power of the crowd is driving the future of business. *Three River Press*.

INSPIRE - Transport Networks. (2020). Retrieved from https://inspire.ec.europa.eu/theme/tn

INSPIRE Austria. (2020). Retrieved from https://www.inspire.gv.at/Allgemein/Allgemeines-zu-INSPIRE.html

ISO 19113:2002. (2020). Retrieved from https://www.iso.org/standard/26018.html

ISO 19157:2013. (2020). Retrieved from https://www.iso.org/standard/32575.html

Jacobs, C. (2016). Data quality in crowdsourcing for. In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, & R. Purves, *European Handbook of Crowdsourced Geographic Information,* pp. 75-86.

Jiang, B. (2012). Volunteered Geographic Information and Computational Geography: New. In D. Sui, M. Goodchild, & S. Elwood (Eds.), *Crowdsourcing Geographic Knowledge. Volunteered Geographic Information in Theory and Practice,* pp. 125-138.

Kuptsov, L. (2020). *Encyclopediaofmath.org*. Retrieved from https://www.encyclopediaofmath.org/index.php/Asymptote

Lendák, I. (2016). Mobile crowd-sensing in the Smart City. In C. Capineri, M. Haklay, H. Huang, V. Antoniou, J. Kettunen, F. Ostermann, & R. Purves, *European Handbook of Crowdsourced Geographic Information.*

Li, D., Zhang, J., & Wu, H. (2012). Spatial Data Quality and Beyond. *International Journal of Geographical Information Science, 26*(12), pp. 2277-2290.

Mapillary. (2018). Retrieved from https://www.mapillary.com/

Mapillary. (2019). Retrieved from https://www.mapillary.com/

Mechanical Turk. (2020). Retrieved from https://www.mturk.com/

Mooney, P. (2011). OpenStreetMap. *Hengstberger Symposium Towards Digital Earth: 3D Spatial Data Infrastructures,* pp. 7-8. Heidelberg.

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

Neuhold, G., Ollmann, T., Buló, S., & Kontschieder, P. (2018). *Mapillary*. Retrieved from https://research.mapillary.com/img/publications/ICCV17a.pdf

Nielsen, J. (2006). *Nielson Norman Group.*

Ooms, K., De Maeyer, P., Dupont, L., Van Der Veken, N., Van de Weghe, N., & Verplaetse, S. (2016). Education in cartography: what is the status of young people's map-reading skills? *Cartography and Geographic Information Science*, pp. 134-153.

OpenStreetMap. (2018). Retrieved from https://www.openstreetmap.org

OpenStreetMap Wiki. (2018). Retrieved from https://wiki.openstreetmap.org/wiki/OSM_Tasking_Manager/Validating_data

OSM. (2019). *OSM Completeness*. Retrieved from OSM: https://wiki.openstreetmap.org/wiki/Completeness

OSM Stats. (2020). Retrieved from https://wiki.openstreetmap.org/wiki/Stats

OSM Stats. (2020). Retrieved from OSM: https://www.openstreetmap.org/stats/data_stats.html

Parkbob. (2018). Retrieved from https://www.parkbob.com/

Popescu, A. G. (2009). Mining a Multilingual Geographical Gazetteer from the Web. *International Joint Conference on Web Intelligence and Intelligent Agent Technology*, pp. 58-62.

Raymond, E. (2000). The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. *Information Technologies and Liberies, 2*, pp. 23-49.

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

Sieber, R., & Haklay, M. (2015). The epistemology(s) of volunteered geographic. *Geo: Geography and Environment, 2*, pp. 122-136.

Stadtverwaltung München. (2019). Retrieved from https://www.muenchen.de/int/en/traffic/cars-and-traffic/parking-zones.html

Stadtverwaltung München. (2020). Retrieved from https://www.muenchen.de/rathaus/Stadtverwaltung/Kreisverwaltungsreferat/Branddirektion-Muenchen/Einsatzvorbeugung/Feuerwehrzufahrt.html

Tobler, W. (1970). A computer movie simulating growth in the Detroit region. *Economic Geography, 2*, pp. 234-240.

TomTom. (2020). Retrieved from TomTom: www.tomtom.com

Vandecasteele, A., & Devillers, R. (2015). Improving volunteered geographic information quality using a tag recommender system: The case of OpenStreetMap. In J. Jokar Arsanjani, A. Zipf, P. Mooney, & M. Helbich, *GIScience: experiences, research, applications*, pp. 59-80.

Waze. (2020). Retrieved from Waze: https://www.waze.com/

Wazeopedia. (2020). Retrieved from Wazeopedia: https://wazeopedia.waze.com/wiki

Statutory Declaration

Herewith I affirm,

- That this master thesis is entirely my own work and that I have not used any other sources than those indicated.
- That this master thesis was not submitted as an examination paper before.
- That this master thesis is identical with the version evaluated by the supervisor

---

Eidesstattliche Erklärung

Hiermit versichere ich,

- dass ich die vorliegende Masterarbeit selbstständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubter Hilfe bedient habe,
- dass ich dieses Masterarbeitsthema bisher weder im In- noch im Ausland in irgendeiner Form als Prüfungsarbeit vorgelegt habe
- und dass diese Arbeit mit der vom Begutachter beurteilten Arbeit vollständig übereinstimmt.

Wien, am……..