TECHNISCHE UNIVERSITÄT WIEN
Vienna University of Technology

MASTERARBEIT

MOTORWAY NETWORK DEPICTION FROM OPENSTREETMAP DATA IN ACCORDANCE WITH CARTOGRAPHIC DEMANDS

Ausgeführt am Institut für
Geoinformation und Kartographie
der Technischen Universität Wien

unter der Anleitung von
Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Wien
Univ.Ass. Dipl.-Ing. Florian Ledermann, TU Wien

durch
Sigita Grīnfelde
Erzherzog-Johann-Platz 1, 1040 Wien

Wien, 1 September 2016

____
Unterschrift (Student)

MASTER'S THESIS

MOTORWAY NETWORK DEPICTION FROM OPENSTREETMAP DATA IN ACCORDANCE WITH CARTOGRAPHIC DEMANDS

Conducted at the Institute for
Geoinformation and Cartography
Vienna University of Technology

under the supervision of
Univ.Prof. Mag.rer.nat. Dr.rer.nat. Georg Gartner, TU Wien
Univ.Ass. Dipl.-Ing. Florian Ledermann, TU Wien

by
Sigita Grīnfelde
Erzherzog-Johann-Platz 1, 1040 Wien

Wien, 1 September 2016

____
Signature (Student)

Declaration of Authorship

I, Sigita GRĪNFELDE, declare that this thesis titled, “Motorway network depiction from OpenStreetMap data in accordance with cartographic demands” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ________________________________

Date: ________________________________

VIENNA UNIVERSITY OF TECHNOLOGY

Abstract

Department of Geodesy and Geoinformation

Master of Science

Motorway network depiction from OpenStreetMap data in accordance with cartographic demands

by Sigita GRĪNFELDE

Roads account for a large and by far one of the most important parts of a map. Main roads, such as motorways, are of a special significance as they represent city and country arteries, as well as connect major locations across countries. Final map products go above exploitation of static base maps and limited map extents. As data amounts and number of representations within a single application grow, so does the difficulty to maintain a neat visualization at all times.

Crowd-sourced projects such as OpenStreetMap (OSM) often offer an up-to-date, easy to access base map in areas that are not covered by authoritative data at the level of detail necessary. The world-wide collaboration of OSM users is powerful, yet OSM faces issues regarding data quality and consistency due to the large amount of contributors and vague guidelines of data annotation.

In this thesis efforts to answer the question, how to improve depiction of motorways when using OSM map, are made. The result is achieved through coping with challenges and empowering the benefits of OSM data characteristics.

This research, firstly, defines cartographic demands regarding motorway depiction, and examines how well OSM major roads comply with the demands. Finally, it offers a few methods for improvements of motorway identification attribute accuracy and extraction of motorway topological relation knowledge. Further these techniques are adapted to motorway network in Austria, tested and the results discussed.

Keywords: Road depiction, Motorway network, Cartographic demands, OpenStreetMap highways, Attribute accuracy, Attribute completeness, Topological integrity

Acknowledgements

I am truly thankful to Dr. Georg Gartner and Dipl.-Ing. Florain Ledermann, under which supervision and guidance this work was conducted. Their insights and valuable feedback are an integral part of the present work. Moreover, I would like to express a spatial thanks for their great encouragement that helped me to keep the motivation whenever I was facing difficulties.

Also, I owe a profound gratitude to my partner, that not only untiringly expressed his support but also shared ideas and gave technical advice and help.

Finally, I must thank my family for the faith in me, as well as for boosting my confidence when it was needed during the time of my writing and throughout all the years of my education.

Contents

List of Figures

List of Tables

List of Abbreviations

AND	**A**utomative **N**avigation **D**ata
ASFiNAG	**A**utobahnen- (und) **S**chnellstraßen- **F**inanzierungs- **A**ktiengesellschaft
bmvit	**B**undesministerium (für) **V**erkehr, **I**nnovation (und) **T**echnologie
OSM	**O**pen**S**treet**M**ap
POI	**P**oint**O**f**I**nterest
SQL	**S**tructured**Q**uery**L**anguage
VGI	**V**olunteered **G**eographic **I**nformation
ÖAMTC	(Der) **Ö**sterreichische **A**utomobil-, **M**otorrad- (und) **T**ouring **C**lub

Chapter 1

Introduction

1.1 Introduction and Motivation

Road networks have a significant role in maps and their applications, such as, motorway maps, spatial analysis and routing. Before creating a high quality map or using it to perform tasks, the uttermost important thoughts should be focused on the source of the data. Questions, such as, is the required data accessible, what are the expenses to obtain the data, what is its quality, is the level of detail sufficient for the purpose, what preprocessing and cleaning tasks will it require, must be considered. This list can continue much further, since the underlying data is the crucial foundation to determine the development and usability of any application that is build on top of it.

A decade ago, a volunteered geographic information community started its work by creating a free editable map of the world, OpenStreetMap (OSM). By now over 2 million of registered users can enrich the geographic data in OSM with their local knowledge (OpenStreetMap Wiki, 2015e). OSM is widely used and applicable as a base map for variety of purposes as a result of its high level of detail, free accessibility and easy integrity with processing tools.

On the downside, OSM can at times expose to issues, such as incorrect topology, inconsistent data description, varying coverage and level of detail. Consequently, OSM is not in a ready-to-use state and may require substantial amount of data cleaning and generalization. To this point there are several efforts made to asses the data quality and automatically detect errors. However, there is yet not much done in order to tackle the issues the nature of OSM road data introduce.

This work is rooted in the idea of creating maps, that depict visually appealing, high quality motorway networks. Based on this vision, the necessary data qualities to achieve it are examined. More specifically, the main focus is concentrated on determining, whether OSM motorway data has a potential to produce maps, that are in line with a set of required cartographic demands. Additionally, it is directed to addressing techniques to improve the output where issues arise. Methods developed within this work are first fine-tuned using motorway data in Vienna and later tested on the entire Austrian motorway network.

1.2 Research question and objectives

Primary, this thesis will answer the question - what is the relation between cartographic demands and OpenStreetMap motorway data? In order to outline an answer, following partial questions need to be considered - what are the cartographic demands of road depiction, what is the potential of OSM motorway data deriving maps, that are in conformity with these demands. Finally, what methods could improve the result. As a result the identified objectives are:

- To define a list of cartographic demands regarding road depiction,
- To examine if OpenStreetMap motorway data have a potential to derive maps that are in line with cartographic demands,
- To explore methods for improving OpenStreetMap motorway data compliance with cartographic demands;

1.3 Structure of the thesis

The rest of this thesis is structured as following: In Chapter 2 the necessary background on important aspects regarding road visualization are explained. A special focus is pointed towards road depiction and use at different scale. As a result cartographic demands for road depiction at large and medium/small scales are defined. In chapter 3 the OpenStreetMap data model is explained with the focus on motorway relevant elements and attributes. In Chapter 4 OpenStreetMap conformity to the cartographic demands, using specific examples, is discussed. In Chapter 5 efforts to tackle issues complicating OpenStreetMap motorway depiction are made. Several methods to improve the attribute accuracy and topological descriptiveness are developed, tested and results discussed. In Chapter 6 conclusions and suggestions for further improvements are given.

Chapter 2

Cartographic demands

The end goal of a map from users' perspective is to provide the necessary information. Therefore, a map should transfer a trustworthy, correct knowledge that can be read in an effortless manner. Cartographers, on the other hand, are concerned with the best methods to formulate and present this information. There are many questions to be considered when creating a map: What is the topic and purpose of the map? What is the best way to present this information? The physical world is very complex and detailed, however, drawing the reality on a map is possible only in an abstract manner due to the constraints of map media. Examples of such constraints are the limitation of space, the limitation of the chosen scale and the human perception abilities of abstracted objects.

To maintain a good readability at different scales, a map cannot be scaled down or up by simply reducing or increasing the zoom level, without changing the level of detail (LoD) (Cecconi and Galanda, 2002). As a result it would be either too detailed and unreadable or too abstract and missing details. Therefore, it is very important to apply generalization on a map. This means that at different scale levels, the map will have different cartographic demands to reach a good representation of reality.

First, this chapter attempts to classify generalization constraints based on previous literature. Next, map use at large and medium/small scales have been collected and road symbolization at various scales analyzed using gathered examples. Finally, based on all the previously discussed topics, cartographic demands for motorway topic have been grounded.

2.1 Classification based on previous literature

Wolf, 2009 defined constraints in two groups. Constraints in one group are limitations of the map display device (paper, digital) characteristics, while constraints in the other group come from human perception abilities of abstracted data. A similar classification from a different perspective is based on the idea of geometry and semantics, two fundamental aspects of cartographic generalization (Pearson and Janodet, 2009). Geometric generalization draws parallels with limitations of the map display device. The decisions to modify map, for example perform elimination, are based solely on feature size or proximity to others features. Semantic generalization, on the other hand, is based on human perception. Using the example of road elimination, the process looks at the importance of features. Thus, at a certain scale only major road classes may be preserved.

Using generalization constraints as a reference, it is possible to define the respective cartographic demands. In general words, cartographic demands are map requirements directed towards successful representation that are scale and topic dependent. However, we can go deeper in these concepts and give a more specific categorization. Classification of the constraints extracted from map generalization literature, allows defining cartographic demands on top of this formal knowledge and further list their properties.

Two essential categories where cartographic generalization splits are geometric and semantic constraints(Wolf, 2009; Zhang, 2012). Furthermore, several authors have named readability (legibility) and preservation as examples of the major generalization constraints' types (Zhang, 2012; Stoter et al., 2009, Stoter et al., 2009). Apart from readability and preservation, another category of frequently explained constraints is representation constraints, or in other terms, constraints of different representation levels (Zhang, 2012; Harrie and Weibel, 2007; Duchêne, Barrault, and Haire, 2001). Further each of them is explained in more detail.

Semantic constraints

Cartographic generalization can be divided into two fundamental parts, that are modification of objects based on geometry and modification of objects based on semantics (Wolf, 2009). Constraints of the transformation do not solely depend on object's size or shape, but mainly on its importance in the context of the map and surrounding features. For example, aggregation operator can be executed on two adjacent objects if the attributes of both are same or similar (Liu et al., 2003).

Geometric characteristics of objects are usually explicitly known. Higher order phenomena such as semantics and patterns, that are traditionally interpreted by cartographers, need to be obtained when generalization is handled by computers. If this information is not know, it can be acquired using external data sources or by data enrichment techniques (Zhang, 2012).

Additionally, importance of semantics can be seen in multi-scale databases. Maintaining relationships between representations of the same real word object at different scales is crucial to keep consistency, propagate updates, perform analysis of data and other tasks. (Devogele, Trevisan, and Raynal, 1996). An illustrative example of influence of semantics at scale transition is building amalgamation resulting in a settlement area.

Readability and preservation constraints

Readability constraints, as the name suggests, are there to improve the readability of the map and determine topology, position, orientation, shape, and distribution (Stoter et al., 2009). On the other hand, the preservation constraint takes care that important objects, relationships and patterns are kept in the map. Preservation constraints are satisfied before generalization but are violated when readability constraints are being solved and vice versa. This means, both constraints will never be completely satisfied at the same time and an optimal solution should be always found (Zhang, 2012), (Stoter et al., 2009).

Within the topic of road maps, common readability constraints are minimum distance between segments of a single road or different roads, width of a road symbol, topology of a road network. Preservation constraints depending on the scale could determine that, for instance, all streets with traffic intensity over a certain threshold should be preserved or all highways starting with secondary class should be preserved.

By the few examples it can be argued that aforementioned constraints are dependent on, firstly, the map topic and, secondly, the scale. Consequently, the next paragraphs describe the representation constraints.

Representation constraints

In previous literature representation constraints have been assigned also with other names, such as, different levels of analysis (Duchêne, Barrault, and Haire, 2001) or different granualities (Zhang, 2012). No matter the name chosen, a common terminology is used to describe the constraints: the concept of -micro, -meso and -macro level. The different levels within the representation constraint mean the different levels at which for instance a road topic is analyzed. Consequently, at micro level cartographic conflicts within an object are analyzed, at meso and macro level groups and feature classes are considered (Zhang, 2012, Duchêne, Barrault, and Haire, 2001). Respectively we define demands at these levels - demands within individual road, demands between 2 roads or even the whole network.

Individual road cartographic constraints

According to (Duchêne, Barrault, and Haire, 2001) the two main symbolization conflicts regarding scale, that occur within one road are coalescence and granularity conflict. Duchêne, Barrault, and Haire, 2001 was concerned with the fact that even within a single road there are heterogeneous parts that would need a different treatment. A single road can have sections that carry no conflict or have either of the major conflicts. The first conflict, coalescence conflict occurs during scale transition, when the road symbol has become too thick for the new scale and overlaps itself. This is a very common situation within mountains where roads contain very sharp curves (Figure 2.1). The second conflict, granularity conflict, happens when the road is too detailed for the scale.

Additionally to the aforementioned micro level conflicts, Duchêne, Barrault, and Haire, 2001 name issues regarding positional accuracy, internal topology and "holes" in the road symbol.

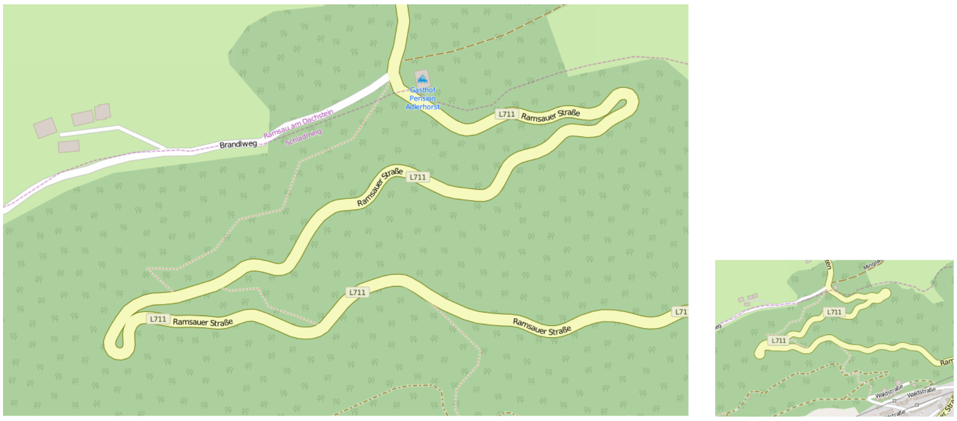

FIGURE 2.1: Self overlapping secondary highway in mountainous area. With the decrease of scale self overlapping increases.

Road network constraints

The main concerns within meso level, are overlaps between close by roads that visually result in misleading topological connections not only seemingly connecting roads that are not connected in reality (Figure 2.2), but also reducing clarity at junctions (Duchêne, Barrault, and Haire, 2001) (Figure 2.3). Similarly like with individual roads, the overlap issue occurs when road symbols are increased at scale transitions, with absence of minimal distance requirement.

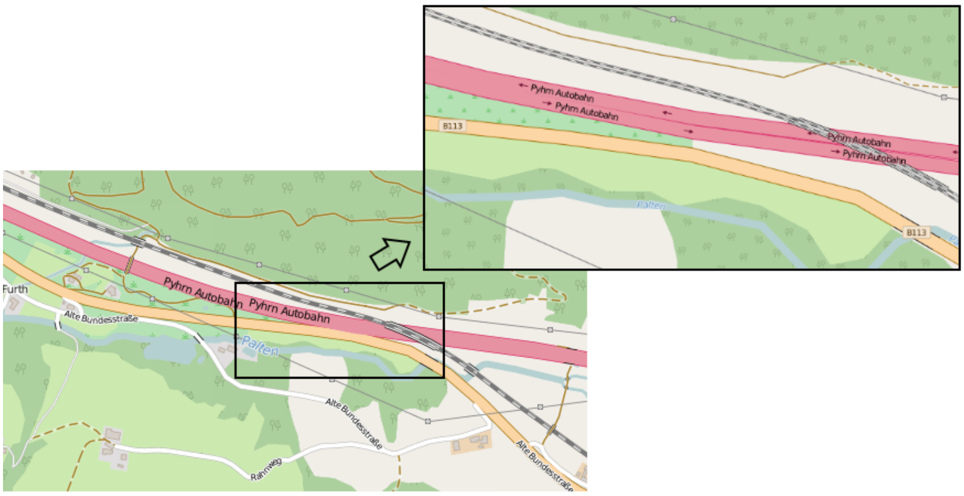

FIGURE 2.2: False connection when the scale is reduced - road symbols overlap

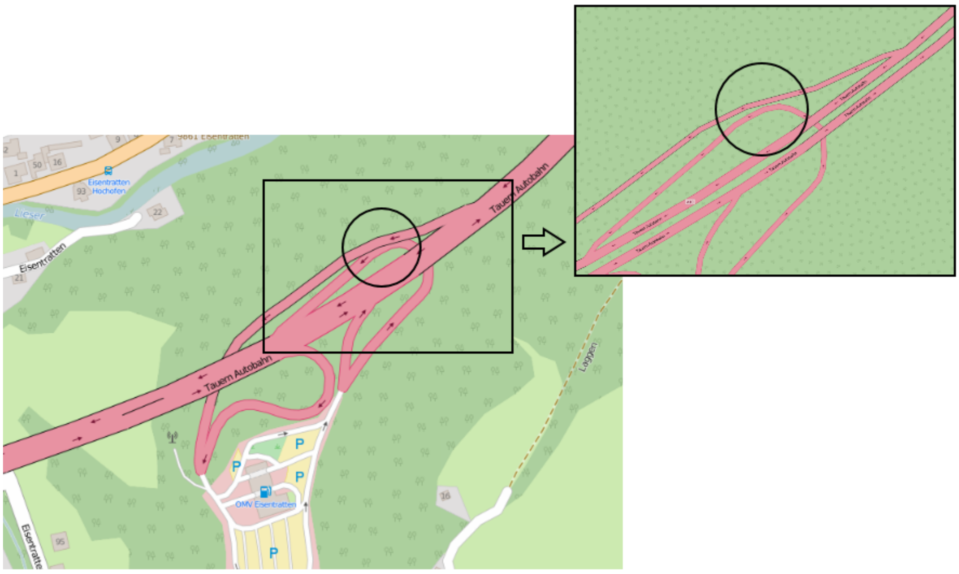

FIGURE 2.3: Reduced clarity at junction when scale is reduced - a segment of a link overlaps another segment while in a larger scale it is visible that these segments have a distance in between.

2.2 Scale dependent road depiction

The need for scale comes due to the limited size of map display. Wherefore, cartographers have to spend time to understand the human perception abilities and, based on that, define appearance of the desired map. Scale transition also exposes to possible conflicts between objects and thus guides development of the cartographic demands.

Figure 2.4 indicates how important it is to adopt a map to its target scale in order to reach a good understanding of the scene depicted. Simple scaling of the map, without adaption of symbolization, result in a weak representation. Wolf, 2009 writes about division of map generalization in structural or quantitative and conceptual or qualitative generalization. Quantitative generalization here is context unaware reduction of map content with scale change, while qualitative is transforming map from elementary to more abstract elements. The bigger the scale the closer to reality symbols can look, the more separate objects per area can be shown. The smaller the scale, the more abstract becomes symbolization and the more limited the number of objects. It is necessary to identify in the jungle of real life objects those elements, that should be eliminated from the map after scale transition, and how to symbolize the rest.

When we talk about depiction of objects based on the map scale, it is important to specify the feature class (points, lines, areas) and the features in question (for example, line features - motorway or river), as the treatment of those objects will depend on their semantics. For example, the width of

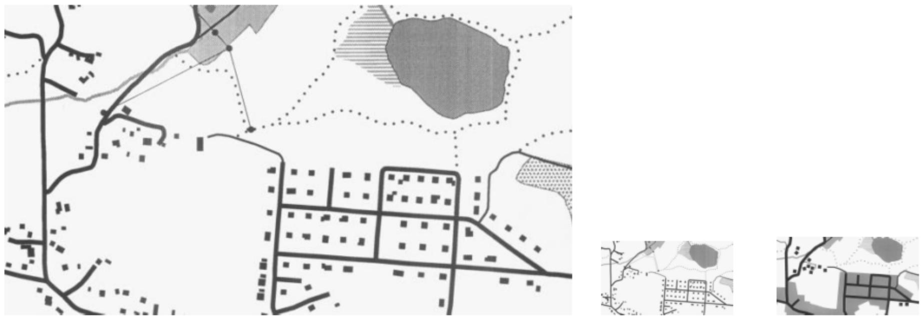

FIGURE 2.4: Three maps of the same area; (a) at a scale of 1:10 000, (b) same map scaled down to 1:50 000, and (c) generalised from (a), but for presentation at a scale of 1:50 000. © Lantmäteriverket, Sweden (2005/4413). Source: Harrie and Weibel, 2007

a river will be reduced with reduction of the scale, until it will not be possible to depict its width on a map proportionally to its width in real life. At that moment it will be symbolized by a single line and eventually on an even smaller scale map, possibly deleted. A fjord, on the other hand, will be reduced in width until it is not sufficient to represent it as two lines anymore. Then it will get deleted as it is not depicted by a single line based on cartographic convention (Buttenfield, 1991). Road symbols are treated even differently than rivers and fjords. When the scale becomes small enough, road symbol will become proportionally larger than it is in reality. Some displacements of nearby roads will be performed to avoid overlap or roads with lower importance will get deleted.

2.2.1 Map use

There is an enormous number of different use cases for maps that are somewhat related to road topic. All of these use cases have the optimal scale at which they can be represented the best. Efforts to classify rather common use cases for various scales have been made.

Scale	**Use case**
< 1: 1000	Planning and maintenance of utility lines in the surrounding of streets and other roads
1: 2000	Topographic maps for cities and other dense settlements for real estate management needs, area planning, detail planning, utility line maintenance
1: 5000	Topographic maps for planning purposes, detail planning, utility maps
1: 10 000	Topographic maps including roads, orienteering maps, access maps

1: 12 000	Military city maps
1: 25 000	Mountain navigation, walking maps, access maps
1: 50 000	Mountain navigation, cycling and mountain-biking maps, military land navigation maps, city maps, road maps, utility maps (high-voltage power lines, oil, gas pipelines), national parks, reservoirs
1: 100 000	Topographic maps of a city and its surroundings, cycling and mountain-biking maps
1: 250 000	Road atlases, general purpose maps, rural areas, administrative borders, tourist road maps, motorcycle maps
1 : 1 000 000	Maps depicting motorway and primary route network, cities and major towns, administrative borders
> 1 : 15 000 000	Main road arteries connecting countries

TABLE 2.1: Example map use cases based on scale

The example use cases mentioned in the table 2.1 are dependent on the source data. For example, utility maps can be created only if the data is at the necessary detail and contain the expected object information. In this thesis focus is directed to VGI, more specifically OSM map, which is a general use map. Based on the list of OSM based services[1] top 3 service categories are 67 general purpose maps (such us, showing points of interest and routs), 35 sports related maps (such as, hiking and biking maps) and 29 maps offering services (examples include very specific information search, such as, vegan restaurants or playground spots). Apart from the previously mentioned services built on top of OSM, the OSM base map is often used in order to navigate in the city and areas with active rural tourism using the plain OSM map or such applications as MAPS.ME[2]. One of the reasons is that OSM offers, for example, very detailed information on points of interest in cities and paths in popular urban areas within Europe (Neis, Zielstra, and Zipf, 2013). It can be said that the overall application of OSM map is different types of navigation. For this reason, areas at different scales from OSM map are presented and potential use cases discussed.

[1] http://wiki.openstreetmap.org/wiki/List_of_OSM-based_services
[2] http://maps.me

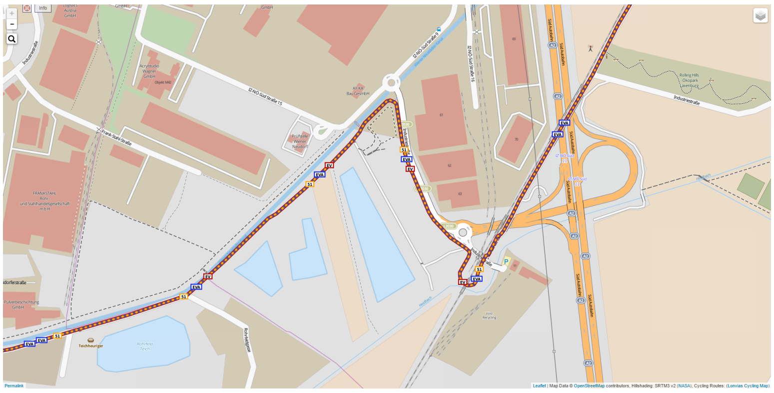

FIGURE 2.5: Cycling track in Vienna at scale 1:4100. Source: http://hikebikemap.org/

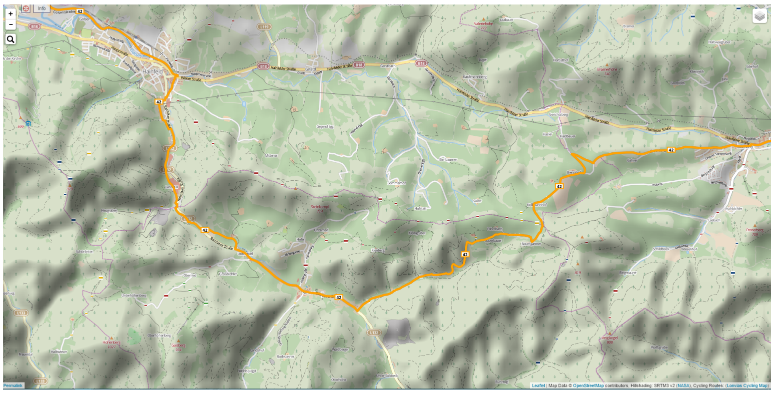

FIGURE 2.6: Mountainous cycling track in a rural area in Austria at scale 1:41 000. Source: http://hikebikemap.org/

Figure 2.5 and figure 2.6, both depict cycling tracks, one in urban but the other figure in rural area. The optimal scale at which to display navigation map depends on factors such as the complexity of the scene, meaning density of a road and path network, if there are any landmarks and the speed at which the map reader is moving. Thus, cycling through a complex junction, might require detailed large scale view on the scene, while moving along a mostly straight path along valley, might only need the general direction and major turns. A clear example can be observed between the automatically set scale in Google Maps app preview for navigation going with different means of transportation and thus speed - on foot and by car (figure 2.7). In the first case the scale is set to a scale around 1:2000, while for cars the scale is decreased to 1:7000.

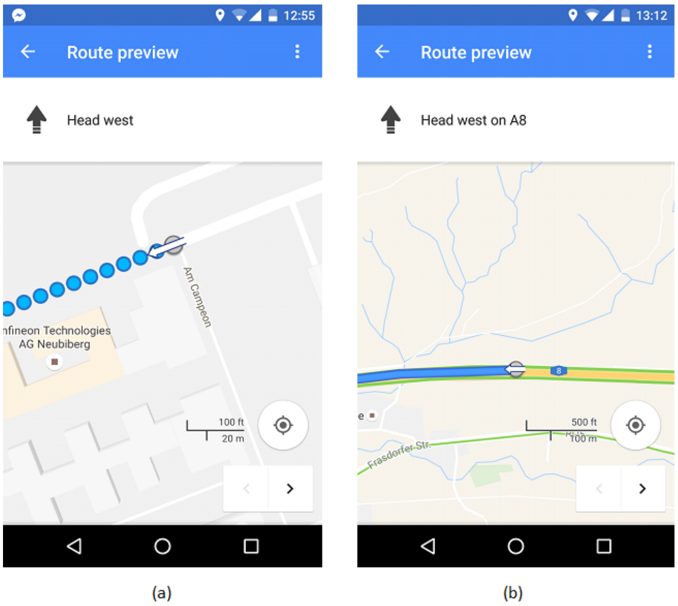

FIGURE 2.7: Image *(a)* - preview of navigation on foot in scale around 1:2000, image *(b)* - preview of navigation by car in scale around 1:7000. Source: Google Maps app.

For maps based on another line feature type that falls under transportation, national railway maps, an obvious indicator for scale change is the transition in the level of rail displacement and the number of displayed stations. For example, the large scale National Rail network map[3] of UK could be used to plan a trip to one of the minor importance stations, while the small scale map[4] gives a better overview when planning a route between major cities (figure 2.8 and figure 2.9). In the small scale map most of the in-between stations are omitted and the rail representative line geometry simplified. Additionally, city names in the small scale map are depicted more clearly by omitting the different stations within one city and replacing them with a single city name.

OpenStreetMap Wiki page lists 20 OSM-based routing services. Even tough not all of them are intended for car navigation, it is undoubtedly one of the most common use of maps. Based on the scale in which the map reader asks the question, depends the scale of the map in which the answer can be found. Thus different scale is needed to answer a question such as how to reach the tanking station from this motorway, than the scale needed to present a major connection between two cities. This statement relates to previously in this chapter discussed scale dependency on the scene complexity. This is because the scene complexity grows, on one hand, with aspects, such as the density of the road network and landmark presence, on the other hand, with the scale of the asked question. Table 2.2 lists

[3] http://www.nationalrail.co.uk/static/images/structure/css/OfficialNationalRailmaplarge.pdf

[4] http://www.nationalrail.co.uk/static/images/structure/css/OfficialNationalRailmapsmall.pdf

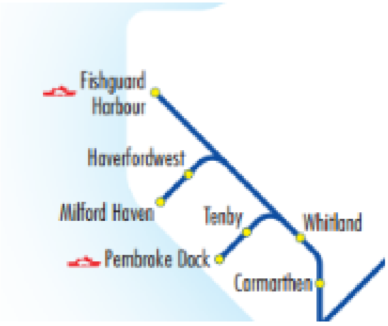

FIGURE 2.8: Cutout of UK National Rail network at large scale. Source: National Rail

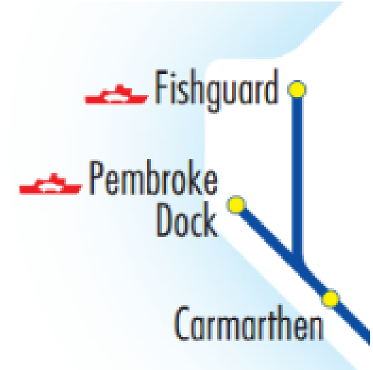

FIGURE 2.9: Cutout of UK National Rail network at small scale. Source: National Rail

some possible map tasks for large scale and medium/small scale motorway maps.

Large Scale	**Medium/Small Scale**
Navigation through complex junctions	Navigation along a long straight motorway
Navigation in a city	Navigation in rural areas
Answering question:"Where are the slip roads on the route connecting to towns?"	Answering question:"Which motorways connect these *n* cities?"
Answering question:"Where on my route there are rest stops?"	Answering question:"What is the traffic situation on the alternative routes?"
...	...

TABLE 2.2: Required map scale for motorway map use-case examples

2.2.2 Symbolization

Symbolization for the map objects assigns visual variables, such as the color, size, shape, according to the map scale and the object semantics (Cecconi and Galanda, 2002). Appropriate symbolization at each scale is part of ensuring readability of the map. The color choice and size of the symbols should unambiguously imply the hierarchy of the road classes. A couple of examples of main road symbols, such as motorway, primary and secondary road symbols, from different map legends at several scales have been collected. Changes in symbols over scale transition in the given examples have been analyzed.

Example legends for scale 1 : 10 000

FIGURE 2.10: Legend cutout of topographic map. Source: Landesvermessung und Geobasisinformation Brandenburg[5]

FIGURE 2.11: Legend cutout of digital topographic map. Source: Landesamt für Vermessung und Geoinformation Sachsen-Anhalt[6]

An example legend for scale 1 : 25 000

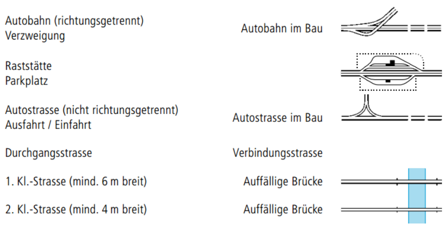

FIGURE 2.12: Legend cutout of topographic map. Source: Bundesamt für Landestopografie swisstopo[7]

FIGURE 2.13: Legend cutout of digital topographic map. Source: Landesamt für Vermessung und Geoinformation Sachsen-Anhalt[6]

[5] http://www.geobasis-bb.de/pdf-Dateien/TK10-V.pdf
[6] http://www.geoportal-th.de/Portals/0/Downloads/Geoproxy/Legende_GeoClient_DTKcol.pdf

Example legends for scale 1 : 50 000

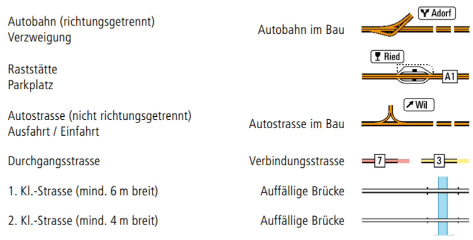

FIGURE 2.14: Legend cutout of country map Switzerland. Source: Bundesamt für Landestopografie swisstopo[7]

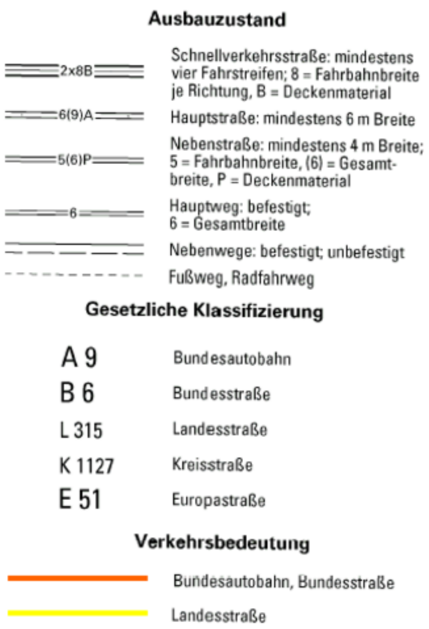

FIGURE 2.15: Map. Source: Landesamt für Vermessung und Geoinformation Sachsen-Anhalt[6]

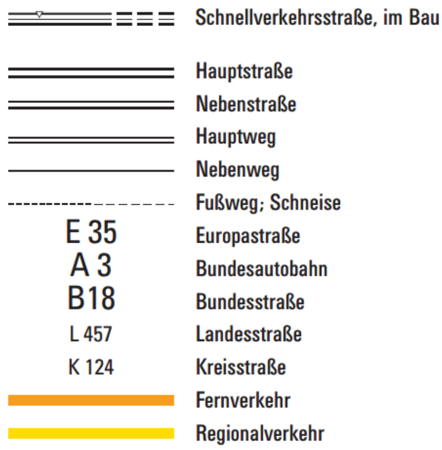

FIGURE 2.16: Legend cutout of a topographic map. Source: Landesamt für Geoinformation und Landentwicklung[8]

[7] http://www.swisstopo.admin.ch/

[8] https://www.lgl-bw.de/lgl-internet/web/sites/default/de/07_Produkte_und_Dienstleistungen/Galerien/Dokumente/Legende_TK50.pdf

Example legends for scale 1 : 100 000

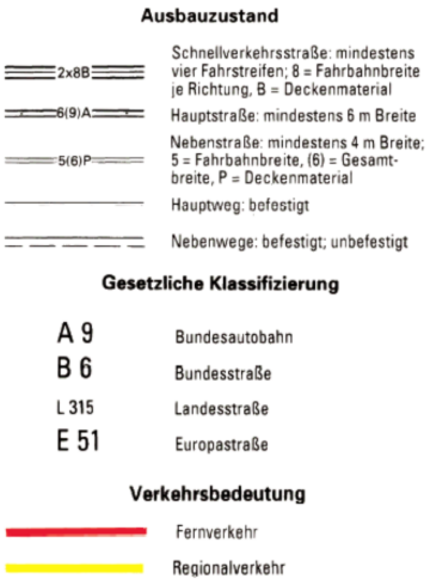

FIGURE 2.17: Map. Source: Landesamt für Vermessung und Geoinformation Sachsen-Anhalt[6]

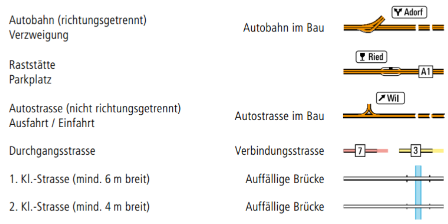

FIGURE 2.18: Map. Source: Bundesamt für Landestopografie swisstopo[7]

An example legend for scale 1: 250 000

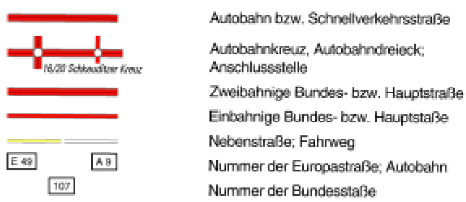

FIGURE 2.19: Topographic overview map. Source: Landesamt für Vermessung und Geoinformation Sachsen-Anhalt[6]

An example legend for scale 1: 600 000

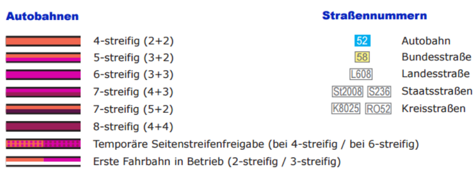

FIGURE 2.20: Motorway map of Germany. Source: Patrick Scholl[9](need to check if I can use this image)

[9] http://autobahnatlas-online.de/

Road class decrease in all examples above is indicated by reduction of the symbol width, line thickness, color change and line style change. Color choice within the collected examples can be observed to be fairly similar. Motorway color is either orange, red, or another reddish dark tone. Subsequent road classes obtain an increasingly lighter color, such as yellow and at last black and white. Commonly interstate and regional highways are colored, while other roads may have no color assigned. Also geometry pattern of each road class is repeating. Motorway symbol consists of two outer lines and a thinner center line. Subsequent major roads that are at least 4m wide, are represented with two outer lines that enclose an increasingly thinner area as the importance of a road decreases. In some cases an indication of number of lanes is added to the major road symbols (figure 2.15 and figure 2.17).

The most distinctive legend is in figure 2.20. This legend is part of a German motorway map and has a large color variation added to symbols in order to describe in details the number of lanes for each motorway. Swisstopo map legends additionally demonstrate the symbolization of motorway junctions, exits and entrances (figure 2.12, 2.14 and 2.18). Over scale change a rather little transition in these symbol details can be observed. The most noticeable symbol transition over scale within maps from the same source, can be seen between legend cutout for a Saxony-Anhalt Office for Surveying and Geographic Information map of scale 1 : 100 000 (figure 2.17) and 1 : 250 000 (figure 2.19). In the scale 1 : 250 000 only the color red is assigned to all important roads, the additional information, such as the number of lanes and road width indications, have been omitted. Nevertheless, it is important to note that here presented collection of legends is sourced from maps of European areas. The patterns of road symbol appearance may be different in maps made in other parts of the world.

Overall, only small variations in symbology over scale can be observed - such as, eliminating the number display of county roads (*Kreisstraße*) at the scale 1 : 100 000 (figure 2.17) that is visible in scale 1: 50 000 (figure 2.15), or the previously discussed color scheme change between maps in scale 1 : 100 000 (figure 2.17) and 1 : 250 000 (figure 2.19). When depicting a major road on various scales, the main transitions come down to, firstly, the change in proportion between width of the symbol and width of the road in reality, secondly, elimination of less important roads.

2.2.3 Demand definition

Based on the use-cases listed in the table 2.2 and review of related literature, certain demands that can be used to guide evaluation of motorway depiction suitability to the scale, have been extracted in the table 2.3. The selected demands have been chosen as the most significant aspects in motorway depiction.

Large Scale Demands	**Medium/Small Scale Demands**
1. Sufficient level of detail At large scale maps are commonly used for navigation. It is important to depict all parts of motorways (road, tunnels, bridges), all lanes and turns, etc.	**1.** Generalization level suiting the scale At medium and small scale not all map objects can be represented proportionally to their real life size, therefore, symbolization or elimination is necessary.
2. Detailed and correct labeling Labels indicate the current location and way to the destination. Accurate motorway names and numbers, as well as motorway link direction information can be crucial for map use.	**2.** Appropriate symbolization At these scales roads are not anymore represented proportionally to their real life width. This brings a need for a good symbolization, that in a well readable manner depicts the road class (such as, motorway or secondary highway) and importance within the network.
3. Topologically descriptive In order to satisfy many tasks performed on a map, junctions, connections to different POIs and populated places should be accurate and clearly legible.	**3.** Solved generalization conflicts As described in the section 2.1 the main concerns regarding generalized map are readability and preservation aspects. Thus demands, such as minimum distance between road segments, geometry (size, orientation, etc.), junction clarity should be respected. Also topology should be preserved after generalization tasks are performed.
4. Geometrically close to reality Map may be frequently compared to the real life scene in front of the map reader. Thus depicting similar shaped, sized and oriented road parts, such as curves, is of a high importance.	

TABLE 2.3: Motorway map demands based on scale

Table 2.3 consists of demands separated by two categories - large scale maps and medium/small scale maps. All the listed demands to some level have to be respected in map off any scale. However, here demands have been grouped by the importance for the scale. For instance, labeling of roads is necessary for any scale, however, it is very important when navigating using a large scale map, to have a detailed labeling, to take the correct turn and to be surrounded by enough references. On the other hand, also appropriate symbolization is required in any case. Using medium and small scale maps the need for a good symbolization only grows as the number of elementary objects decreases and objects on the map become more and more abstracted.

Table 2.3 is made with an assumption that the necessary data for all scale maps is available and complete. The demands regarding data amount put forward are either sufficient level of detail of the data at large scales, or suitable generalization level at medium and small scales.

Chapter 3

OpenStreetMap data model and quality aspects

In 2006 a volunteered geographic information community started its work on creating a free, editable map of the world, OpenStreetMap (OSM). The ever growing number of active users, that last year alone reached the 160 000 mark, ensures a wide community behind enriching the geographic data in OSM with their local knowledge (OpenStreetMap Wiki, 2015e). OSM is broadly used and applicable as a base map for variety of purposes as a result of its high level of detail, free accessibility and easy integration with processing tools.

Apart from separate user contributions, OSM have experienced a number of valuable imports from other sources, that have helped greatly to improve the data amount and quality. Examples of such imports are: aerial photography provided by Bing, MapBox, entire Automotive Navigation Data (AND) street data of the Netherlands, as well as AND road networks of China and India and more (OpenStreetMap Wiki, 2016j, OpenStreetMap Wiki, 2014a).

Data, edited by separate users or imported from external sources, have to follow a certain data model. Understanding its structure is a base knowledge for working with the data. More about the data model, challenges introduced by characteristics of data collection and structure, as well as data quality aspects can be found in the next sections.

3.1 OSM data model

3.1.1 General Concepts

The elementary object of OpenStreetMap conceptual data model is called element. This abstract concept is further divided in more specific ones - nodes, ways and relations. All of the elements can have assigned tags that describe some significant characteristic of the element (OpenStreetMap Wiki, 2015a).

Nodes

Nodes are points in space that are defined by coordinates and an id value. However, they can contain additional information and tags. Oftentimes nodes will build up ways, for instance, by being part of a road or building. However, they can exist on their own, representing a point feature, such as a bench.

Ways

Ways are always constructed by at minimum two nodes. They can be either open, closed or called areas. Open ways are simple polylines, often observed as roads and railways, and their starting point and end point do not match. Closed ways, on the other hand, always have a matching starting and end point. A well representative example of a closed way is a roundabout. Area is a filled section defined by a closed way. Typically, those are parking areas, parks, buildings and similar objects. In most cases, specifying the kind of an object, implies that it is an area. In some cases, a dedicated attribute *area* is necessary to be marked with a *yes* value (OpenStreetMap Wiki, 2015c).

Relations

Relations can contain an ordered list of both, nodes and ways, as well as other relations. Its primary goal is to describe either logical or geographical relationship between objects. Any of these objects can also have an assigned role that determines what meaning this element has in the relation. For example, relation *route* can be assigned to several values, such as *hiking*, *bus* and *bicycle*. In case of the value *bus* and *bicycle*, a way that is a member of this relation can have a role *forward* or *backward*. A point being part of the *bus* relation can have a role *stop*. To this moment according to the OSM Wiki page, there are fifteen established relation types (OpenStreetMap Wiki, 2015f).

Tags

A tag describes an important semantic characteristic of an element by using a key-value pair. Usually elements have more than one tag. For example, building of the Vienna University of Technology has tags, such as *building = yes*, *amenity = university*, *name = TU Wien, Hauptgebäude*, and a few more.

By default some tags may be implied using others. For instance, to define a bicycle path, one can specify tags:

highway=path;

bicycle=designated;

foot=no;

surface=paved;

However, it is possible to use *cycleway* value that already implies bicycle path and a paved surface:

highway=cycleway;

foot=no;

The general term for such a social tagging system is *folksonomy*[1]. A peculiar feature of folksonomy and tags is their free format definition, meaning that users can create new tags themselves. Nevertheless, the OSM community has approved a set of commonly used tags, ensuring a certain level of consistency. New tags are proposed, discussed and may result in being widely used or not. However, this process does not stop anyone from using their own tags (Zhang and Tinghua, 2015).

Free tagging causes positive consequences as well as negative. On one hand, better, more descriptive tags can be found over time, on the other hand, not all users use the tags in a consistent manner. Free tagging allows to harmonize differences across countries - OSM is used world wide and, while certain attributes are valid in Europe, they might not be fitting to describe the same features in Asia. However, a number of confusing tags and tag combinations with similar meaning may be introduced. This is disadvantageous when one is querying data. In the best case scenario, the tags, involved in the query, are common and well established tags within the OSM community. If it is not the case, all the possible tags should be found and accounted for. To draw an example of the variety in tagging, the previously mentioned university building has six different name keys, each with its own purpose - *name*, *alt_name*, *int_name*, *loc_name*, *name:de*, *name:en* (OpenStreetMap Wiki, 2016g).

When creating maps based on OSM data, tags are used for many tasks starting from extracting the information you need from a map, such as motorway network within a certain region, to applying algorithms to the data, such as generalization algorithms. Thus tags can be used to build queries, extract the necessary or remove the unnecessary data, classify, symbolize, be used in algorithms, etc.

3.1.2 Motorways

Motorway in the context of OSM data model is a tag value of a key *highway* *(Tag: highway = motorway)*. *Motorway* is the single chosen term by OSM representing, in the context of physical world, a controlled-access highway, in other English speaking parts of the world also called expressway and freeway (OpenStreetMap Wiki, 2016d).

Nodes construct ways that serve as motorway segments. A way can contain a maximum of 2000 nodes but normally in case of motorways this

[1] Vander Wal, 2007

number is much smaller (OpenStreetMap Wiki, 2016k). Motorways are conventionally represented by two parallel carriageways, each pointing in the direction of traffic (OpenStreetMap Wiki, 2016i). The separate motorway segments are part of a motorway relation. For example, all motorway *West Autobahn* segments should belong to a relation that identify them with the particular motorway, such as relation *West Autobahn* or a possible relation reflecting the national motorway code *A1*. However, the same segments can be part of several relationships classifying them in certain groups. For example, it is common in Germany to create a motorway relation based on its national highway code and European road code, and a relation as roads between certain cities by each direction (eg. relation *A 9 Halle/Leipzig - Berlin [positive]*).

Motorways belong to the most significant highway class. Other high importance highways are trunks (most important roads that are not motorways), primary (commonly connecting larger towns) and secondary (commonly connecting towns) roads. While many basic attributes of these highway classes are the same, there are differences in some of the attributes and implications that come along with the road class. For example, *highway=motorway* and the later explained tag *highway=motorway_link* are the only ways that imply the road quality. Ways that are not paved should never use these tags (OpenStreetMap Wiki, 2015d).

For the case of motorways, there are two tags that contribute to the identification of a specific motorway. They are *name* for the motorway name and *ref* for the motorway national code. In some countries the *name* value is documented identically to the *ref* value in case motorways do not have specific names. Other tags in the tagging scheme give information about the number of lanes, maximum and minimum speed, destination, carriageway reference number for motorways in United Kingdom, etc. (OpenStreetMap Wiki, 2016i).

Motorway class implies few tag key and value pairs. One of the implied tags is *access=no*, that denies access for general public. This means that additional tags to indicate allowed vehicle types, are necessary. In case of motorways the next implied tag explains exactly that. *motor_vehicle=yes* specifies that the allowed transportation mode is restricted to all motor vehicles. Next key value pair that should always be present for motorways is *oneway=yes*, that denotes - each way can be used only in one direction (OpenStreetMap Wiki, 2016e).

3.1.2.1 Special sections and objects of motorways

highway=motorway is not the only tag that is crucial for motorways. Link roads, that are used to access or exit a motorway, are tagged with *highway=motorway_link*. *motorway_junction* is a point feature designated to be placed at motorway exits (between motorway carriageway and motorway link) to hold the exit number or name depending on each country specifics. Furthermore, the first *motorway_link* following the *motorway_junction* node should hold a tag *destination:ref* to indicate the reference number of a motorway this exit leads to.

There are two more essential and common parts of a motorway that have to be taken into account - bridges and tunnels. In the simplest cases, their segments are tagged with *bridge=yes* and *tunnel=yes*, respectively. To identify to which motorway a bridge or a tunnel belongs, is not always a trivial task. Despite the fact that tunnels and bridges have a dedicated name key *tunnel:name=* and *bridge:name=*, in countries such as Austria and Germany, the motorway name attribute is commonly occupied by the name of the tunnel or bridge.

In more complex scenarios when several roads represent a bridge, the bridge carries additional features or it has multiple levels there are alternative tagging methods. One way is to create a separate bridge outline defined as *man_made=bridge*. The other option and the only in case of bridges with several levels, is creation of a bridge relation (OpenStreetMap Wiki, 2016f).

Following table 3.1 summarizes the most important objects and their attributes regarding motorway networks.

Object	Attribute	Use	Example
motorway	name	name of a motorway, motorway bridge, tunnel or another object	Ost Autobahn
	ref	reference number	A4
	bridge	indicates a bridge on a motorway	yes
	tunnel	indicates a tunnel on a motorway	yes
motorway_link	ref	reference number of motorway the link is exiting	A23
	destination:ref	reference number of destination motorway	A23
motorway_junction	name	junction name specified on the sign	Knoten Schwechat
	ref	exit number specified on the sign	16

TABLE 3.1: Summary of the most important objects and attributes regarding motorway networks

Other objects that are often found along motorways are roads and areas for rest and service stations. *highway=service* specifies a service to access a fuel station or a service way on a parking lot. A very similar tag *highway=services* is used for areas along a road with a fuel station and commonly also other services such as food. *rest_area*, on the hand, is dedicated for rest areas along a road without any services.

A couple of tags are defined to indicate security objects. *escape* is used to specify an emergency lane for trucks in case of a braking failure. Emergency phones that are often located on motorways should be tagged with *emergency=phone* specifying also its number to mark the location of the caller.

3.1.2.2 Road restrictions

There is a high usage of navigation applications nowadays, and many routing services choose OSM to build their applications. With this use case in mind, OSM is adapting the data, especially attributes concerning roads.

The simplest specified features are maximum and minimum speed, number of lanes, surface and obstacles, such as, crossings. Few more crucial attributes regarding navigation are turn restriction tags, indications of one way streets, documented forward and backward lanes. Restrictions can be indicated by turn restrictions using *Relation:restriction* or @ sign to add a certain condition. Restriction relation can reveal which turns are allowed and which not, for what vehicle type the restriction is applied, exceptions and time when it applies. The latter restriction attribute can be mapped also by using the conditional restrictions with @ sign. This sign allows to indicate the vehicle type and time of the restriction as well as road condition dependent restrictions, allowed vehicle attributes and other rules (OpenStreetMap Wiki, 2016c).

As a conclusion, it is evident that motorways as well as other highway types can potentially contain a large amount of additional information. This information, only if filled out consistently and is present for all highways, can greatly contribute to road related data analysis and navigation services.

3.2 OSM data challenges

OpenStreetMap is a Volunteered Geographic Information System, that exposes to a number of unique challenges compared to commercial map data. Starting from the way the data is collected to how it is structured and maintained, needs new methods and creativity in solving issues.

Unlike any commercial data provider, behind OSM stands an enormous number of mapmakers. To this moment there are over 160 000 active contributors (OpenStreetMap Wiki, 2015e). This is seemingly positive fact, as all of them coming from different areas of the world can share their local knowledge by using the map. However, this fact also brings several issues regarding consistency, completeness, level of detail and accuracy. The high number of contributors is not the problem itself, it rather comes in combination with the difficult management of such a high number of people from different backgrounds.

The OSM Wiki is offering a broad documentation of how to map and attribute objects. Moreover, Wikipedia contains a number of WikiProjects - collaborative projects carried out by groups of people interested in a specific topic, location or task. Some countries have a motorway related WikiProject that can give suggestions on what is the correct way to map and tag motorway related objects, for example, motorways and motorway bridges. Examples of WikiProjects are projects for motorways in Germany[2] and Austria[3], projects for Austrian *Schnellstraße*[4] and motorway bridges[5], as well as European E-road network[6]. The listed motorway WikiProjects contain a complete list of the elements and the current mapping state (completeness, errors) for each of them. Following these country specific guidelines can notably improve the consistency and quality of the motorways. On one hand, users can see which motorways need some improvements, on the other hand, they serve as guidelines, for example, WikiProject *Austria/Schnellstraße* gives directions on when to tag the so called *Schnellstraße* (road of a similar category as motorway, translated "motorway") as motorway and when as trunk.

Apart from the documentation, also many tutorials and answers to questions within the wide OSM community can be found. However, to reach a high quality mapping style and routine, time and effort to study the documentation and gain experience are needed. The diversity in amounts of skills and energy contributors are willing to input, as well as the varying number of active users in different areas, result in issues, such as inconsistency and various levels of accuracy.

As already described in the section 3.1.1, OSM maintains a free tagging system that only suggests the use of certain tags and allows creating new ones. While this freedom can bring benefits as the model is adaptable to user needs over time, it can be problematic for many applications - for navigation purposes or querying certain kind of data using tag information (Zhang and Tinghua, 2015).

Senaratne et al., 2016 discusses additional reasons of the different data quality. Some examples of the causes are usage of various technologies and tools, map serving heterogeneous purposes, heterogeneity in coverage due to vague concepts, lack of standardization that results in quality variations across different data sources, such as text, image and map. One of the consequences Senaratne et al., 2016 notes is different level of detail in different

[2] http://wiki.openstreetmap.org/wiki/WikiProject_Germany/Autobahn
[3] http://wiki.openstreetmap.org/wiki/WikiProject_Austria/Autobahnen
[4] http://wiki.openstreetmap.org/wiki/WikiProject_Austria/Schnellstra%C3%9Fen
[5] http://wiki.openstreetmap.org/wiki/WikiProject_Austria/Autobahnbr%C3%BCcken
[6] http://wiki.openstreetmap.org/wiki/WikiProject_Europe/E-road_network

areas. Some of the most detailed areas, have been mapped to distinction of separate trees and benches. As the map can be used for various purposes, this level of detail can be valuable for some applications, while unnecessary for others. This means that most applications will require querying the data to extract what is really needed. That process, though, asks for attribute consistently and completeness.

Finally, Senaratne et al., 2016 mentions lack of gatekeepers. This suggests that there should be more experienced OSM contributors that perform quality checks on existing data. There is a hope for it to improve in the future, since in many areas the tendency of OSM contributions changes from creating more new data towards improving quality of the existing data (Hashemi and Ali Abbaspour, 2015, Sehra, Singh, and Rai, 2014).

Quality assessment is for the largest part dependent on contributor visual inspection. This method has many downsides, for instance, it is ineffective and in many cases cannot be carried out by remote contributors, as the local knowledge is necessary to spot many of the errors. Moreover, issues with wrong or missing data that is needed for specific applications, such as navigation purposes, can hardly be found with visual map examination. These problems can be spotted during the actual development or usage of an application (Zhang and Tinghua, 2015). With this said, it is clear that after defining and documenting the application specific needs, automated checks could significantly reduce the efforts and give more reliable results.

There are many existing error detection tools. These tools support automated or manual error recognition and relay on other contributor manual fixes. Automated edits should be done only by experienced users through consultation with the rest of the community and following *Automated Edits code of conduct* in order to protect the OSM data base (OpenStreetMap Wiki, 2015b, OpenStreetMap Wiki, 2016b).

3.3 Aspects of data quality

In the chapter 2 focus was on answering questions - how to depict roads on a map, what should be considered when depicting roads. Based on these questions it was possible to extract and summarize cartographic demands for road depiction at different scale levels. In this chapter, on the other hand, specifically road depiction in OSM is discussed. Furthermore, in this sub-chapter the attention is shifted towards another view on road depiction - from how to depict roads to is the depiction successful.

3.3.1 Classification

Many studies concerning VGI data quality assessment point to five main data quality aspects. Completeness, positional accuracy, temporal accuracy, logical accuracy and attribute/semantic accuracy (Haklay, 2010; Barron, Neis, and Zipf, 2013; Sehra, 2014; Sehra, Singh, and Rai, 2014; Arsanjani et al., 2013; Hashemi and Ali Abbaspour, 2015).

- **Completeness** - The measure that indicates the sufficiency of data objects and attributes available as well as the level of excess data (Sehra, Singh, and Rai, 2014). Incomplete map data leads to a wrong perception of the scene represented via map. At smaller scale it would be imprecise understanding of object distribution patterns. At larger scale incomplete map data can complicate the orientation using the map. If landmarks, streets, street names and other objects or object attributes are missing, the scene in the map differs from the actual scene, giving a misleading information.

- **Positional accuracy** - A degree to which coordinates of an object in map differs from the object in reality. (Sehra, Singh, and Rai, 2014; Haklay, 2010).Violations of positional accuracy may lead to relatively wrongly positioned objects or object groups, such as overlaps and false connections.

- **Temporal accuracy** - The level to which map data is up-to-date. Additionally, it can be defined as a ratio to which changes in data correspond to the changes in real life (Haklay, 2010). Similarly to completeness aspect, unsatisfactory temporal accuracy, results in a wrong scene representation. Objects and attributes that are missing, changed or do not exist anymore violate map temporal accuracy.

- **Logical accuracy** - Accuracy of internal data logical relationships. The component of logical accuracy, that is of special importance in VGI, is topological consistency. Unconnected roads, undershoots, overshoots, duplicated lines are few of very common issues (Hashemi and Ali Abbaspour, 2015). While there are tools to detect possible topology issues, before fixing an issue, the editor has to be sure that, for instance, certain roads meet in reality. In some cases one of the roads can be with its end point very closely located to another road, but in map as well as in reality not connected. Also a road can seem to have a mistakenly dangled endpoint, that in fact is a road with a dead end.

- **Semantic accuracy** - The measure of how well data attributes represent the meaning of the data in the real world. In case of OSM, to evaluate semantic accuracy, it is necessary to closely look at the used tagging system and how well objects are tagged. The two important parts of semantic accuracy regarding roads are road type and road attributes (Arsanjani et al., 2013). From a visual perspective, violations of semantic accuracy lead to inappropriately symbolized or wrongly labeled objects.

3.3.2 Quality assessment methods

Quality assessment can be carried out using two main approaches, external or extrinsic and internal or intrinsic. The first one uses reference data, that is considered as the "correct" data to compare the map in question with. The latter method does not use any third party data but either specifications or internal data (Zhang and Tinghua, 2015).

The downside of external evaluation is the need for external data source. Example of external method is buffer growing, where the distance between a line feature in one data set and the same line in another is compared. Completeness can be evaluated using grid based method, detecting the discrepancy between data sets within a certain sized grid cell (Sehra, Singh, and Rai, 2014). First, it cannot be guaranteed that the authoritative reference data is correct itself or even of a higher quality and, second, official data in many cases associates with high expenses. Another argument is that maps may be created for various purposes, thus, not adequate to be compared (Loidl and Keller, 2015). Additionally, if there is too high disagreement between the two data sets, this method is not sufficient and other techniques have to be applied (Zhang and Tinghua, 2015).

Internal evaluation, on the other hand, is a data-centered approach. Loidl and Keller, 2015 explains that, for example, in context of attribute consistency, internal approach is based on the redundancy and inherit logic.

3.3.3 Reference data

Extrinsic techniques using reference data are rather common in research quality assessment and data enrichment tasks. When a high quality external data is available, the process can be made fairly easier. Typically data sources are commercial or governmental, that can be closed, requiring access in return for money, or open. The reference data, depending on the task, can be either textual or non-textual data, such as maps.

In case of this thesis textual data is used as the reference for some methods in order to improve the given OSM motorway data of Austria. The considered sources are following. *ASFiNAG* is the Austrian state-owned company that plans, finances, builds, maintains, operates and collects tolls on the primary road network. They offer various information on motorway planning, construction[7] and operating, as well as reports[8] that can possibly be useful for this work. The other very important source is *bmvit* - the Austrian Ministry for Transport, Innovation and Technology. In their website under transportation section they offer statistics and other information on all kinds of transportation, including motorways [9], tunnels[10]and planned projects. The third potential data source is *ÖAMTC*, the Austrian Automobile, Motorcycle and Touring Club. This association covers topics regarding traffic, mobility and safety on Austrian roads.

[7] https://www.asfinag.at/unterwegs/bauen

[8] https://www.asfinag.at/documents/10180/13369/de_Buch+30+Jahre+ASFINAG.pdf/8af1a7eb-9bda-4f3c-86a9-81181d4bacb9

[9] https://www.bmvit.gv.at/verkehr/strasse/autostrasse/index.html

[10] https://www.bmvit.gv.at/verkehr/strasse/tunnel/index.html

3.3.4 Existing quality assurance tools

There is a long list of error detection and bug reporting tools, also called quality assurance tools, for OSM data error reporting, checking and editing. These tools vary in their focus and, therefore, identify a different range of errors.

Examining the cartographic demands derived in the chapter 2, demands that to some extent could potentially be checked for their validity using an error detection tool, are the ones directed towards sufficient level of map detail, topology and detailed labeling. Further in the text three applications that include information on highway issues, mainly tag based problems, are shortly discussed.

The first tool is *Qa.poole.ch*[11]. Its *NoName map layer*, as the title suggests, shows location on road segments with a missing name or reference value. Also it highlights segments with the tag *access=yes*, that is considered harmful due to undefined semantics in access tagging. The last layer this tool provides shows buildings that have no house number, name or other alternative identification.

layers.openstreetmap.fr[12] is a quality assurance tool that provides a number of different layers each marking several errors. The most remarkable regarding motorways in the *structure* layer are detection of sudden highway type change, broken highway level continuity, big relation, duplicate geometry and unconnected way. There are also several checks for name tag - missing values, highway reference value written in the name tag, name value containing two names and, finally, given language specific name while no default name is filled out.

OSM_Inspector[13] offers a number of different layers structured in views. The geometry view highlights long ways that are constructed from more than 19000 nodes, a number that comes close to the maximum, 2000 nodes. Also ways consisting of only one node or a duplicate node are marked. Further self-intersecting roads and the nodes at which they self-intersect are detected. Highway view additionally to missing name or reference value, looks for unknown highway types, ways tagged with *oneway*, *lanes*, *maxspeed* and *maxheight* speed but missing values. Routing view displays issues crucial for routing purposes, for example, unconnected, duplicate roads and road islands (Figure 3.1).

[11] http://qa.poole.ch/
[12] http://osmose.openstreetmap.fr/en/map/
[13] http://tools.geofabrik.de/osmi/

FIGURE 3.1: Error detection with OSM Inspector. In purple - road island (road that is not reachable), in red - unconnected highways with distance between them up to 1m

3.3.5 Conclusion

As a crowd-sourced project with a large community of contributors, OpenStreetMap exposes to a different set of data challenges compared to commercial data. Therefore, also another set of quality aspects deserve the highest attention. Completeness, semantic accuracy, logical accuracy are few priority points in quality assessment. While a number of quality assurance tools for OSM are developed, they are detecting mainly data errors and missing attribute values, where only manual fixes can be applied.

Chapter 4

OpenStreetMap data compliance with cartographic demands

In chapter 2 road depiction has been discussed and, based on the conclusions, in sub-chapter 2.2.3 scale dependent cartographic demands for motorway topic have been extracted. The aim of this section is to conclude if OSM data have a potential to derive maps, that are in line with the defined cartographic demands for highway depiction. With the focus on motorways and major highways, various aspects of highway data characteristics, illegible visual examples and motorway related objects are extracted and studied in order to test OSM data compliance with previously proposed demands.

4.1 Completeness

The table 2.3 summarizing cartographic demands, displayed in the previous chapter, assumes that the available data is complete. This is a crucial precondition for most map use cases. However, OSM data, as discussed earlier in 3.2, is a product of volunteer collaboration all over the world. This results in a map that has various data completeness between different regions, countries, cities, rural and urban areas, etc. For this reason, additionally to the defined demands of road depiction, completeness is discussed.

Generally, Europe is one of the worlds regions where contributor activity and number of contributors is comparatively high both per area and per number of population (OSMstats, 2016, Pascal Neis, 2012). Numbers regarding official road length data in Austria in comparison with the length of mapped roads by OSM contributors are used to derive a conclusion of road data completeness in Austria. The source of the official road length is based on combination of *bmvit* (Austrian Ministry for Transport, Innovation and Technology) data for years 2011[1] and 2012[2].

Motorway length (m)

	bmvit	**Doubled**	**OSM**	**+ Links**
All motorways with reference A#	1719000	3438000	3408794	3512574
All motorways (reference A# and S#)	2182000	4364000	4073918	5005633

TABLE 4.1: Motorway length comparison between official data and motorways mapped in OSM (Austria)

All road length (m)

	bmvit	**Doubled**	**OSM**
All roads + links	124510000	126692000	210465432
All roads - group 1*	124510000	126692000	27830879
All roads - group 2*	124510000	126692000	27754273
All roads - group 3*	124510000	126692000	29053692

TABLE 4.2: All road length comparison between official data and roads mapped in OSM (Austria)

group 1* = ways tagged as highway = *motorway_link*, *trunk_link*, *primary_link*, *secondary_link*, *tertiary_link*, *raceway*, *steps*, *path*

group 2* = group 1 - *cycleway*

group 3* = group 2 + all links (*motorway_link*, *trunk_link*, *primary_link*, *secondary_link*, *tertiary_link*)

Both, table 4.1 and table 4.2, include the road length in meters given by *bmvit* data and roads mapped by OSM users. The length calculation in OSM data was done by exploiting PostGIS *ST_Length* function, that returns length of *LineString*[3] geometry. The queries using *ST_Length* were applied on the road segments with tags and tag values of the particular interest.

As the highest road classes with reference values *A#* and *S#* (# stands for any number, such as, *A21*) in OSM are depicted using 2 lines (for each direction), also in the table columns called *Doubled* calculation of highway

[1] https://www.bmvit.gv.at/verkehr/gesamtverkehr/statistik/downloads/viz_2011_gesamtbericht_270613.pdf
[2] https://www.bmvit.gv.at/verkehr/gesamtverkehr/gvp/faktenblaetter/umwelt/fb_strasse_schiene_netz.pdf
[3] *LineString* - an object built from sequence of points and segments connecting them.

official length multiplied by two is shown. Table 4.2, compared to table 4.1, additionally includes lower class roads that are drawn in OSM using only one line and links that are mostly single direction. For this reason column *Doubled* in this particular table is a sum of the doubled motorway length from table 4.1 and all other segment length, that are not roads marked with either reference *A#* or *S#*, counted only one time. This is necessary so that the official road length would be comparable to the mapped one. There is still a risk of mistakenly duplicated ways in OSM, that would be included in this comparison.

Based on the data in table 4.1, it is visible that the official motorway *A#* length is very close to the mapped - difference is only around 700 meters, that could be considered negligible on the background of total length. However, when roads with reference *A#* are added to motorways with reference *S#*, the difference between mapped road length and the given one, significantly grows to around 29 kilometers. That could indicate that, since motorway *A#* class has a higher priority than *S#*, also in mapping it has received more attention. Also missing reference values could have caused this discrepancy. Another possibility is that some segments at the beginning and end of links in OSM are considered in *bmvit* data as parts of motorways.

Since it is not known how the official data of motorway length is measured, it is not clear if motorway links are included in the length calculation or not. For this reason, link length is added to calculation of the OSM mapped motorways for comparison with the ground truth. In the result it can be observed that in both cases mapped roads are significantly longer than the number in comparison, therefore, most likely, links are not counted as parts of motorways in the official data. Nevertheless, it should be noted that the reference data available is from year 2013, while length of the roads mapped in OSM is from the year 2016. Furthermore, OSM users are known for being very quick in mapping changes, hence, OSM data is more up to date (Arsanjani et al., 2013).

In the conclusion, it can be said, that OSM is very close to the official data in completeness of motorways marked with *A#*. Motorway counterpart, roads marked with *S#* are not as close to the ground truth data. However, as discussed in previous paragraphs, it is difficult to definitely state that it is incompleteness in sense of unmapped ways because the calculation of the mapped road length is not clear. This calculation solely relays on the road reference value *A#* or *S#*. The possible issue may be rooted in the way elements are tagged. It can be the fault of missing reference values or segments tagged as links instead of roads. Also other structures like tunnels and bridges can be labeled using their name without adding the reference of motorway they are a part of, thus, being left out of the calculation of mapped roads. This dependency on the correctness and style of element tagging, clearly indicates a correlation between the completeness measure and semantic accuracy. The major topics in semantic accuracy regarding OSM data are the road type and attributes (Arsanjani et al., 2013). Any imperfections in both of the two can lead to gaps in completeness.

4.2 Sufficient level of detail

Data sets with high level of road network detail can be beneficially used for traffic planning, geographical analysis, LBS, emergency evacuation as well as simply navigating through a city (Li et al., 2014).

Looking at the world map, it can be said with confidence that the mapped detail level is not sufficient. It varies across areas with different population density - urban and rural areas, as well as across countries (comparison United Kingdom and Iraq) (Senaratne et al., 2016, Al-Bakri, 2015).

The idea of sufficient level of detail collides with the completeness, analyzed in this chapter earlier, and one of the following points discussed later in the chapter - detailed and correct labeling demand. Apart from attributes analyzed under completeness and detailed labeling factors, there is a number of other attributes, such as maximum speed, one way indication and motorway bridges and tunnels, that are important details of motorways. Further these are analyzed.

Attributes regarding motorway maximum speed and one way indication are *maxspeed* and *oneway*. Both of them show good results for Austrian motorways - out of 11824 motorway segments, 10389 segments have information on the maximum driving speed and only one segment is missing the *oneway* tag.

Important parts of motorways that are not the roads themselves, are bridges and tunnels. Unfortunately, bridge names from the *ASFiNAG* (corporation of Austrian motorway maintenance) list of bridges (OpenStreetMap Wiki, 2014b), after a close inspection, appear to be unclear. Many of those names include some description of their location and, therefore, the information is insufficient to be thoroughly checked against OSM data. However, Austrian motorway tunnel names from *bmvit* tunnel list[4] have been extracted and compared with OSM data. *bmvit* provides a record of 161 motorway tunnels in utilization while, after querying OSM motorway data, 148 distinct tunnel names in the area of Austria were found.

Analyzing OSM data, it was found that despite the dedicated tunnel name key *tunnel:name*, only a small part of the motorway segments tagged with *tunnel=yes* actually holds the tunnel name in *tunnel:name*, majority use the motorway *name* key instead. 140 distinct tunnel names are found in *name* tag versus 28 are found in *tunnel:name*. However, 20 of all tunnel segments have their names indicated in both tags. Independently from the key used for tunnel name value, all motorway segments with *tunnel=yes* and *bridge=yes* are further referred to as tunnels and bridges.

114 of the tunnels mapped in OSM match approximately to some tunnel name in the *bmvit* tunnel list. 27 out of the 55 tunnels in OSM, that did not find any matching name, are holding in the name field either solely motorway name or motorway name in combination with a tunnel name that

[4] http://www.bmvit.gv.at/verkehr/strasse/tunnel/downloads/tunnelliste2016.pdf

could not be found in the official data. Comparison of the two lists was not trivial due to the different way of writing the same name. While *bmvit* data is using shortened names, OSM mostly contain full names and sometimes redundant information, commonly a motorway name. Examples of matching entries from both lists are shown in table 4.3.

bmvit **name**	**OSM name**
Farchern West UT	Süd Autobahn Unterflurtrasse Farchern-West
Ofenauer	Ofenauer Tunnel
Gschwendnerberg	Pyhrn Autobahn Gschwendnerbergtunnel
Bruck	Tunnel Bruck

TABLE 4.3: Few examples of matched tunnel name entries between *bmvit* and OSM data. Difference in writing the sames in different data sets is noticeable.

4.3 Detailed and correct labeling

Detailed and correct labeling demand in its root requires attribute accuracy and completeness. On the surface it is important, when talking about road depiction, to display correct and consistent labels on roads in a map. On a deeper level, attribute accuracy and completeness ensures that each road can be recognized as a whole - it can be identified. This aspect is necessary for tasks, such as visualization (for example, the previously mentioned labeling), descriptive statistics (for example, classification) and spatial analysis (for example, routing) (Loidl and Keller, 2015). Also some generalization algorithms, such as selection, may need to exploit data attributes.

Figure 4.1 shows an example of a secondary road segment in Austria missing its name. This way it becomes unidentifiable without further inspection. Next figure 4.2 is a cut out of a layer generated on top of OSM by a quality assurance site *Qa.poole.ch*[5], where red segments indicate an area of road segments missing names. This particular scale was chosen to better illustrate the differences between rural and urban areas. The pieces of map with almost completely named roads are surroundings of two cities Linz and Wels, while the more red sections are rural areas.

The following tables and figures are unfolding the results of identification key, name and reference, value completeness assessment within two areas - Vienna and Austria. First, the number of segments with and without reference value is compared for two groups of roads - only motorways and group of roads that consist of motorways, trunk roads and primary roads (table 4.4 and table 4.7). In Austria reference value is relevant only for these types of roads. This is because lower class roads may have an official reference value, however, they are expected to be commonly used with their

[5] http://qa.poole.ch/

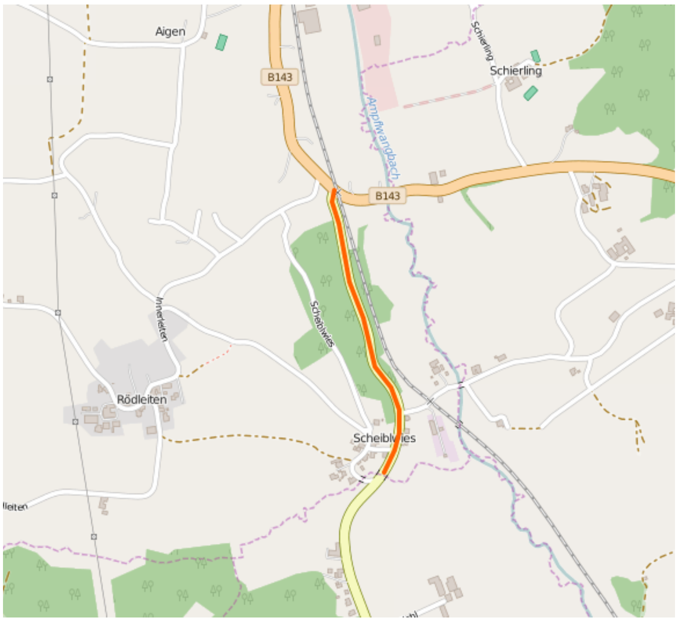

FIGURE 4.1: Secondary road segment missing any identification attribute

name, while reference value serves administrative purposes. Secondary roads in Austria are equivalent to road class called *Landesstraße*. Second, the number of segments with and without name is compared for motorways and a wider group that consists of motorways, trunk roads, primary and also secondary and tertiary roads as all of them should be named (table 4.5 and table 4.8). The same two road groups are used for the check of combined name and reference value existence in table 4.6 and table 4.9. The purpose of the latter figures is to answer the question - how many important roads are missing their identification attributes at all.

Test area: Vienna

Tag:ref=*

Road class	Value exists	Missing value
Motorways	1743	8
Major roads (from primary)	5672	134

TABLE 4.4: Number of segments of different types of roads, that are missing reference values within Vienna

FIGURE 4.2: Overview of roads with no name. Source: http://qa.poole.ch/ (an experimental quality assurance tool)

Tag:name=*

Road class	**Value exists**	**Missing value**
Motorways	1673	78
Important roads (from tertiary)	4405	1478

TABLE 4.5: Number of segments of different types of roads, that are missing name values within Vienna

Tag:ref=* and Tag:name=*

Road class	**Values exist**	**Missing values**
Motorways	1665	0
Important roads (from tertiary)	4257	54

TABLE 4.6: Number of segments of different types of roads, that are simultaneously missing reference and name values within Vienna

The map in figure 4.3 give an overview of missing name and reference values of motorway segments in Vienna.

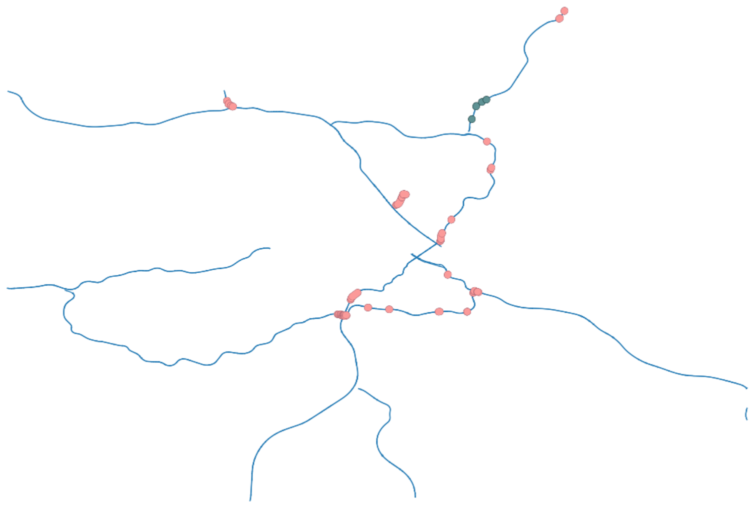

FIGURE 4.3: Map of Vienna motorway network representing segments with missing attribute values. In purple - segments missing name value, in dark blue - segments missing reference value. Other segments have both attribute values.

Test area: Austria

Tag:ref=*

Road class	**Value exists**	**Missing value**
Motorways	11808	16
Major roads (from primary)	46634	1438

TABLE 4.7: Number of segments of different types of roads, that are missing reference values within Austria

Tag:name=*

Road class	**Value exists**	**Missing value**
Motorways	11519	305
Important roads (from tertiary)	58014	22842

TABLE 4.8: Number of segments of different types of roads, that are missing name values within Austria

Tag:ref=* and Tag:name=*

Road class	Value exists	Missing value
Motorways	11514	11
Important roads (from tertiary)	51757	1223

TABLE 4.9: Number of segments of different types of roads, that are simultaneously missing reference and name values within Austria

The numbers were retrieved using simple SQL queries, that checked if segments of specific highway types hold values of reference or name keys. The number of rows, where values are not equal to *null* and where values are equal to *null*, gave the answers. In case of table 4.6 and table 4.9, it was checked if values of both keys are not *null*, as well as if both are *null*, to detect entirely unidentifiable segments.

Regarding the results, in table 4.4 it is visible that motorways in Vienna are well referenced, only a small portion constituting 0,5% are missing. Between major roads the completeness decreases to 2,3% of missing values. With names, as the table 4.5 shows, the numbers are worse - 4,5% and 25,1% of motorway segments and important road segments respectively are missing name. It can be speculated that few of those roads may, in fact, be unnamed roads, however, it is rather unlikely in these road classes. Looking at the next table, it can be seen that there is no unidentifiable motorway - there exists either a name or reference value for all motorway segments. Also in case of 98,7% important roads, there exist at least one identification attribute.

Looking at data of the entire Austria results stay very close to the extracted results of Vienna. A fairly similar portion, 0,1%, of motorways and 3% of major roads have a missing reference value. A slightly smaller percentage of motorways have a missing reference number in whole Austria compared with Vienna - 2,6%. On the other hand, missing names appear in moderately bigger percentage of cases, 28,3%, between important road segments. While the area of entire Austria, unlike Vienna does not have 100% of marked motorways, just eleven segments with no name and reference value leave the number as high as 99,9%. 4,6% of important roads have no identification attribute.

To visually compare the amount of missing identification tags per each road type in Austria as a whole, figure 4.4 was drawn. Purple bars represent the length of segments without name and reference value as percentage of the road type in question. The red bars show the portion each road type fill out of the total length of all motorways, trunk roads, primary, secondary and tertiary roads.

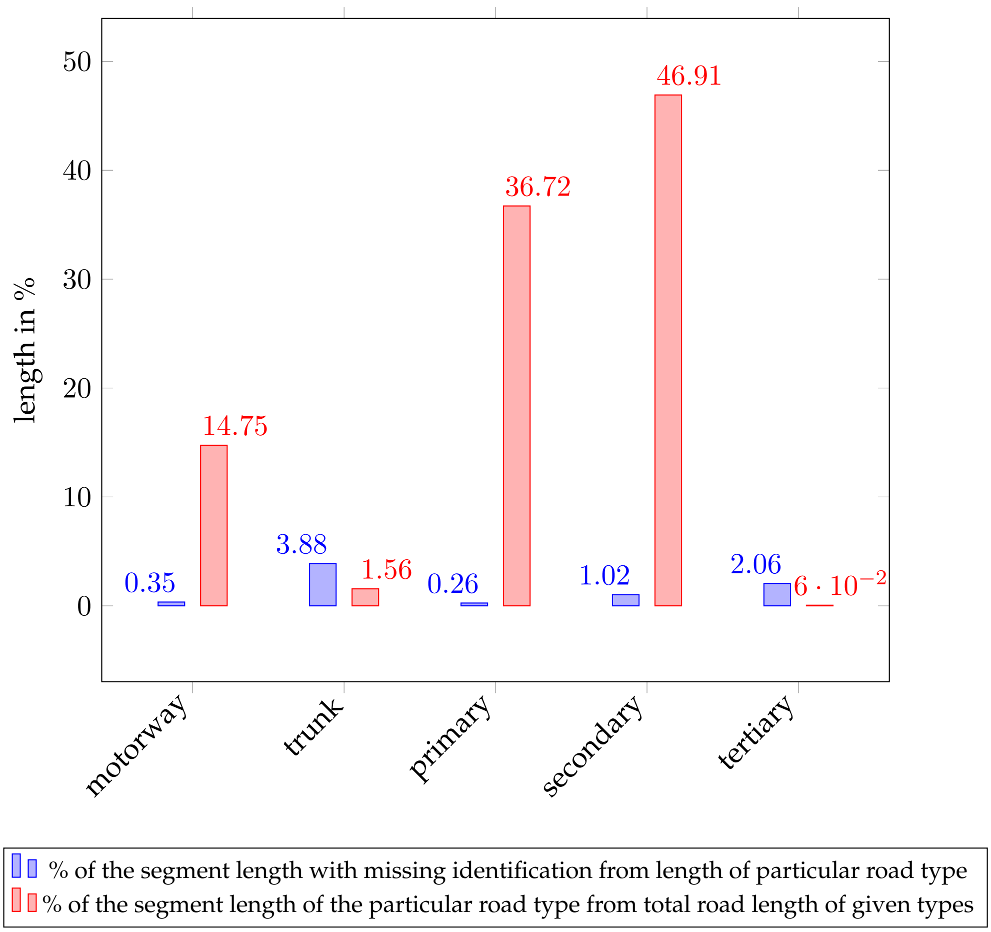

FIGURE 4.4: Bar plot using purple color represents percentage of road length that is missing identification attributes - name and reference value - from the total length of that particular road type. Bars in red display the road length per road type in percentage from total length of all here mentioned road types. Extent: Austria.

Whilst analyzing the results, it should be kept in mind that these numbers do not prove that, for example, in Vienna 98.7% of important road segments are named correctly or 100% of motorway segments are identifiable. Instead, it just means that some name or some reference is assigned and 1,3% of important road segments in Vienna are certainly unnamed. The numbers of problematic segments are expected to be higher if typos and wrong or inconsistent names and reference values would be considered. A necessity for correct road name and reference value detection can be clearly observed, specially within the important road classes. Chapter 5.1 further assesses attribute accuracy of motorways - correctness of names and consistency between name and reference value combinations.

4.4 Topologically descriptive

Motorway connections to other motorways or lower class roads are described by map objects - motorway links and motorway junctions. Using just the available attribute information, topological descriptiveness can be analyzed based on presence of certain key and value pairs. Results are shown in the following tables.

Test area: Vienna

Motorway link

Sum	**Tag:ref=***	**Tag:destination:ref=***	**Both tags**
1399	112	234	15

TABLE 4.10: Number of *motorway_link* segments with reference value or/and destination reference value in area of Vienna

Junctions

Sum	**Tag:name=***	**Tag:ref=***	**Both tags**
239	238	110	110

TABLE 4.11: Number of *motorway_junction* nodes with name and reference values in area of Vienna

Test area: Austria

Motorway link

Sum	**Tag:ref=***	**Tag:destination:ref=***	**Both tags**
6056	792	1031	94

TABLE 4.12: Number of *motorway_link* segments with reference value or/and destination reference value in area of Austria

Junctions

Sum	**Tag:name=***	**Tag:ref=***	**Both tags**
1091	1075	618	618

TABLE 4.13: Number of *motorway_junction* nodes with name and reference values in area of Austria

In order to retrieve the numbers regarding *motorway_link* tag, SQL queries checking nodes, where reference value is not *null*, where destination reference value is not *null* and where values of both keys are simultaneously not *null*, were applied. Only those *ref* values are considered as valid, that follow the motorway reference number convention *A#* or *S#*. Similarly to *motorway_link*, nodes with the tag *motorway_junction* were checked by querying values of each key, *name* and *ref*, individually and, finally, selecting rows, where values of both keys are not *null*.

Reference value on a link segment should serve as an indicator of the motorway identity that the link road exits. Table 4.10 shows that only 9% of all motorway links in Vienna have a reference value specified. The same phenomenon can be observed for area of entire Austria - just 13% of referenced motorway links (table 4.12). The tag *destination:ref* is dedicated to specify the reference value of motorway the link leads to. Number of segments with the tag *destination:ref* increased compared to segments with the tag *ref* in both areas. If values *ref* and *destination:ref* are present then the linked motorways are known. Indications of both, source and target motorway identities are negligibly low - under a 2% mark in both areas. This means that the rest of motorway link segments cannot be associated with any particular motorway or indicate particular motorway connections. Also these separate link segments cannot be recognized as parts of any united link based on the attribute information and without making use of OSM defined relations.

Motorway junction is a point feature that can be well exploited to recognize motorway topology. These points should be placed at motorway link road start point, where motorway exits to another motorway or lower class road. Motorway junctions give valuable information especially to routing maps. In tables 4.11 and 4.13 it can be clearly seen that in Austria names of these junctions are more important than the reference value. Out of 239 junctions in Vienna and 1091 in Austria, 99,6% and 98,5% junctions respectively have an existing name. Junction reference values are junction numbers and do not imply any identification of a motorway. However, motorways in Austria can be recognized through junction names using third party data, such as *ASFiNAG*[6].

Additionally, there are known topology issues that are not related to the attributes but geometry itself (figure 3.1). Problems, such as duplicate features, undershoots and overshoots occur in the OSM data (Senaratne et al., 2016, Hashemi and Ali Abbaspour, 2015).

The analysis shows that most of the motorway link segments cannot indicate particular motorway connections. In order to extract connected motorway information it is necessary to, firstly, recognize the separate link segments in groups. Then it is potentially possible to make use of the existing attribute information each link group holds, the information on motorways these link groups intersect or examine OSM defined relations.

To make the extraction of connection information easier, two options of OSM model improvement can be proposed: (a) introducing relation that

[6]https://www.asfinag.at/documents/10180/13369/de_Buch+30+Jahre+ASFINAG.pdf/8af1a7eb-9bda-4f3c-86a9-81181d4bacb9

groups link segments between two motorways in each direction, (b) encouraging in a more obvious manner *ref* and *destination:ref* key usage on each link segment. The issue with (a) is, that relation creation is an advanced editing and requires a certain level of expertise. The suggestion (b) is more pragmatic. At the moment using OSM in-browser editors, *ref* is one of the default keys that shows up in user interface during the editing process of major road related elements. *destination:ref*, on the other hand, belongs to the tags that should be specifically added by the user. Adding keys, that enhance information on topological relations, to the default tag list during the editing process, would encourage users to fill in values of those keys. This would allow grouping of links that use the same *ref* and *destination:ref* values, as well as easier extraction of connected motorways and other major highways.

4.5 Geometrical closeness to reality

OSM map consists of two types of inputs - imports of external data from open-licensed sources and contributions mainly done on the basis of areal imagery and by collecting GPS tracks (OpenSteetMap Wiki, 2016). Areal imagery with global coverage, that can be used for OSM tracing purposes, is provided by such services as *Bing* and *MapBox* (OpenStreetMap Wiki, 2016a). Apart from these sources, there is a number of country wide areal imagery sources, not limited to only satellite images but also orthophotos (OpenStreetMap Wiki, 2016a). In the area of Austria such a source is provided by *GeoImage.at*. These images are gathered by professionals of the federal governments, therefore, are assumed to be georeferenced at precise locations (OpenStreetMap Wiki, 2015g). Since big part of data edits is done on basis of areal imagery, it can be also assumed that, although errors in digitization exist, the result is geometrically close to the scene in life and the provided imagery. Detailed inspection of this aspect is beyond the scope of this work.

4.6 Generalization conflicts

Even though, two separate demands were proposed, generalization level suiting scale and a demand regarding solved generalization conflicts, here they both are combined examined within the same sub-chapter. The first demand is concerned with objects being appropriately symbolized or eliminated when representation proportional to their real size is not possible. The other demand looks at cases exposing generalization conflicts, such as violation of minimum distance between two roads and self-intersecting roads. In the end, most of the generalization problems are caused by violating one of these demands.

Examples illustrating few problems regarding OSM major road network generalization have been presented already in figure 2.2 and figure 2.3. In these figures roads overlap either giving impression of connection between them or reducing clarity at a junction. Both of these problems could be solved by generalizing the scene or introducing a constraint of minimum distance between road segments.

To illustrate also other type of problems that occur within road network, figure 4.5 and figure 4.6 are presented.

FIGURE 4.5: Various generalization issues - overlapping ways and visually inconsistent road levels over scale transition

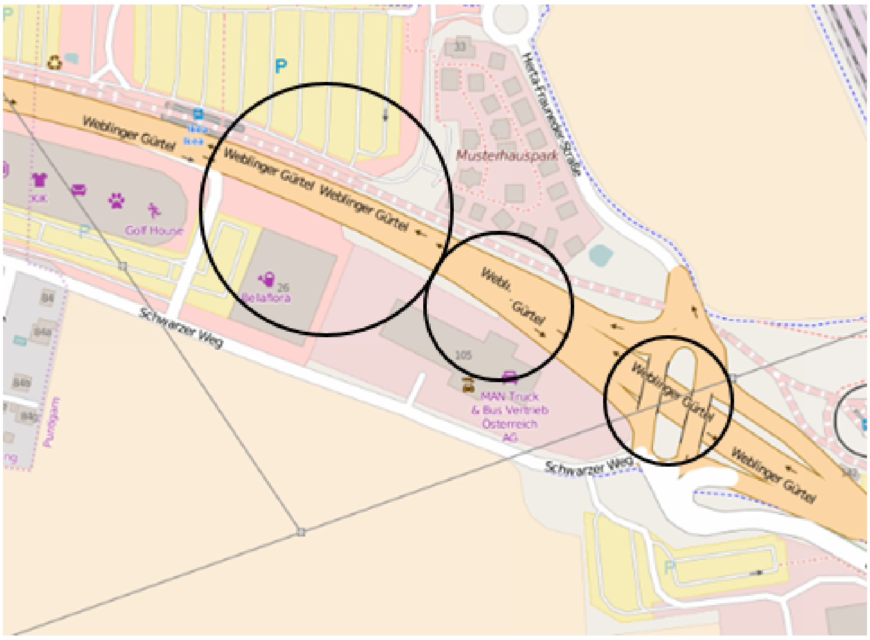

FIGURE 4.6: Reduced clarity at junction when scale is reduced - a segment of a link overlaps another segment while in a larger scale it is visible that these segments have a distance in between.

Examining figure 4.5 at least two issues can be observed over scale transition. First, on the top part of the map cut-out are several parallel roads within a city. These roads during a transition to smaller scale form a structure, that resemble an area instead of lines. On the lower part of the image a runabout (in yellow color) and a motorway (in red color) are located at the same two dimensional coordinates but on different levels. However, changing the map zoom, it becomes unclear which of the two roads passes through the bottom level and which through the top. On the most left map cut-out it seems that the motorway is higher, possibly due to the label that covers the roundabout, as well it looks to be the same case in the most right image. However, analyzing the image in the middle, it is very clear that roundabout is located on a higher level. Such a situation causes unclear junctions and deceptive conclusions. Figure 4.6 presents a consequence of collision of many parallel ways. As a result these parallel segments display highway name labels that also collide with each other, thus making this part of map illegible.

Nevertheless, the classical OpenStreetMap layer style considers different representations of road classes at major scale changes. Figure 4.7 shows secondary highway symbol change and elimination over the transition of zoom level.

FIGURE 4.7: Secondary highway symbol change over three zoom levels - zoom 8 (not visible), zoom 11 (gray, thin line) and zoom 12 (yellow fill with gray case).

Even though some level of control using symbol elimination and style change is present, violations of constraints, such as maintaining minimum distance, can be found in OSM rendering. More efforts should be focused in the direction of generalization to attain well readable map representations at all scales.

4.7 Appropriate symbolization

From the data point of view, in order to satisfy the demand of appropriate symbolization, requirements go hand in hand with detailed and correct labeling demand. The uttermost important preconditions to symbolize roads according to their class and to adjust them to the scale are completeness and attribute accuracy, discussed in the previous points of this sub-chapter.

From the perspective of visual map rendering, the options are limitless. OSM community offers a number of rendering software with the ability to add custom style-sheets and control the symbol visual characteristics at any scale (OpenStreetMap Wiki, 2016h).

4.8 Conclusion

According to the methods used in this analysis, OSM data compliance with cartographic demands vary from demand to demand. Good results were found in motorway completeness based on the length of mapped roads and occurrence of motorway identification attributes - name and reference number value. However, the correctness of these identification attributes has not been assessed and needs to be examined further.

The biggest issues were found in topological descriptiveness. No sufficient method to extract knowledge of motorway connections solely based on attribute information was found. To test the given topological information two objects were chosen - motorway links and motorway junctions. The result showed that motorway links are poorly attributed and cannot be used to describe motorway interconnections. Most of the link segments hold no reference value of the motorway they exit, nor the destination motorway. Motorway junctions, on the other hand, are well attributed with their name. This information can further indicate motorway mutual connections, using external data on specific motorway junction locations.

During evaluation, assessment of several demands was found to be interdependent. For instance, imperfections in semantic accuracy, such as wrongly assigned road type, lead to gaps in completeness. Assessment of sufficient level of detail demand found certain objects, bridges and tunnels, to be important parts of motorways. Missing the attribute indications of these objects, impacts the level of detail.

To analyze OSM motorway network conformity to cartographic demands, each of the demands was assessed individually using OSM element and tag information for quantitative measures, where possible. In general, results showed high completeness of mapped motorways and data model, that encourages rich data description. However, inconsistency of tag usage, missing information on name accuracy and poor topological descriptiveness motivate further inspection and problem assessment.

Chapter 5

Improvements in attribute accuracy and topology extraction

The cartographic demands for depicting motorways determine the usability of data for different applications. Analysis of OSM data compliance with these demands indicated the strengths and weaknesses within the data. In this chapter a further look into two issues is carried out - correct and detailed labeling and topological descriptiveness.

Roads in maps are expected to be easily recognized by their name label or reference number. To be able to show these features on a map correctly and consistently, to be able to define a certain road as a whole, high attribute accuracy is necessary. Tasks, such as visualization, descriptive statistics and spatial analysis, benefit from consistent and accurate attributes (Loidl and Keller, 2015). Even though during assessment of OSM data in chapter 4.3 completeness of identification attributes in case of motorways was found to be good, the result did not imply any information about the correctness of those attributes. In order to group highway segments by the road identity, an ideal attribute accuracy is necessary.

Another expected road characteristic in a map is to represent clear junctions and connections between locations. Also some algorithms such as generalization using graphs, require clear knowledge of linked roads. These goals become more complicated when major roads that do not intersect directly but interconnect using link roads, are involved. In these cases, methods to extract the topology information should be found. In this regard, OSM data assessment in chapter 4.4 was done. As a result it was concluded, that there are great difficulties to detect which motorways are interconnected. No obvious way to obtain this information was found, therefore, topological descriptiveness was also selected as a subject of further inspection.

Since OSM as a crowd-sourced data source is very different from conventionally created maps, the issues regarding its incompleteness have to be approached using methods specifically fitting the nature of OSM data. Solutions for addressing the lack of fulfillment of the two demands are proposed and tested.

The basic setup for tests is following:

Test area: Method development and parameter values are set based on tests for area of Vienna; Further tests of the developed methods carried out for area of entire Austria.

Data source and tools:

- OSM data of Vienna and area of Austria were extracted using Geofabrik[1] download server.
- To query and change data attributes PostgreSQL was used.
- To visualize and edit spatial data mainly QGIS in combination with Python was used.
- Reference data of motorway names and reference numbers taken from bmvit, Austrian Ministry for Transport, Innovation and Technology.

Tasks:

- Improvement of motorway identification attributes - name and reference number - completeness and correctness using intrinsic and extrinsic methods;
- Exploitation of OSM data characteristics in order to retrieve knowledge of motorway topology;

5.1 Attribute accuracy

Quality of applications, such as navigation and spatial analysis, depend on geometric and semantic modeling of the underlying road network data (Zhang and Tinghua, 2015). As it was examined in the previous chapter, a high heterogeneity in the data creation, require an automated procedure not only for detecting issues but also helping to solve them.

In the chapter 3 motorway name and reference number existence was examined. However, the measurement still does not say anything about the correctness of the existent names. In this sub-chapter important analysis points of OSM motorway data are discussed and an automated procedure for detecting and fixing inconsistent and missing motorway names and reference numbers is proposed.

Few of the main name tagging issues also recognized in Zhang and Tinghua, 2015 are:

[1] http://download.geofabrik.de/

- Missing name - as already analyzed in chapter 3, there are cases of missing motorway segment names. The proposed solution here involves finding names based on the given motorway reference number or neighboring segment attribute information.
- Incorrect name - some contributors have used motorway reference number or direction of the highway, or even all of this information combined in the motorway name field. Additionally, typos can be found.
- Inconsistent name - there are many language specifics that have to be accounted for, when analyzing name errors. For example, common inconsistency in German language is writing names as compound words or separate words. This fact brings in additional issues with name consistency that would not occur otherwise.

5.1.1 Methodology

Firstly, rules that should be followed by the used methods, based on the expected outcome, are defined. Then the procedure of the method implementation is established. Sub-chapter 3.3.2 discussed the two categories of quality assessment - extrinsic and intrinsic. In order to improve the attribute accuracy, both of these methods are used. Further, consequential results, advantages and disadvantages of each method are discussed. Within a country there is a high number of motorway segments, therefore, in order to avoid necessity for manual checks, the procedures are automated. The steps are worked out by analyzing the nature of data and then adapting the tasks for area of Vienna. Next, the resulting procedure is tested on entire Austria and outcome discussed.

Rules

Following two rules present the reference number and name interdependence on which decisions to rename segments or add missing values are based.

1.) Every distinct motorway name, that already includes the word *Autobahn* or *Schnellstraße*, should have only one corresponding reference value. Example – segments of *Nord Autobahn* either have reference value *A5* or it is missing for some segments. Assumption is made that *A5* is the correct reference value and is added, where reference value for *Nord Autobahn* is missing.

2.) Each distinct motorway reference value should have a single corresponding motorway name. Example – all motorway segments that hold the reference value *A5* should also be named as *Nord Autobahn*.

Characteristics of Austrian motorway data

When analyzing names and especially, when trying to match similar names, the language specific system of how they are built within a particular context is important.

In German motorway is called *Autobahn* and this string is expected to be part of motorway names. In Austria there are highways that in most aspects are equivalent to motorways, as well as they are tagged *highway=motorway* within OSM data. These highways are called *Schnellstraße* and they are considered within this thesis on the motorway level. The language and country specific knowledge influences the choice of a later discussed technique, fuzzy string matching, parameters. If two names, for instance, *Nord Autobahn* and *Ost Autobahn* are compared, depending on the parameters, the result can give a high matching score, since the sub-string *Autobahn* is already a perfect match. Language specifics are also to be considered when building look-up lists of motorway names contained in OSM data.

Compound words are a common practice in German language. For this reason, many inconsistencies in OSM motorway names can be found - part of segments are named as, for example, *Ost Autobahn* and another part *Ostautobahn*. In order to bring consistency within the data, as a reference an official list of motorway names should be used when possible. For this project information from two sources was gathered and compared. One of them is bmvit[2], the Austrian Ministry for Transport, Innovation and Technology and the other is ASFiNAG[3], the Austrian state-owned company that plans, finances, builds, maintains, operates and collects tolls on the primary road network. However, even between these two sources inconsistencies in names were found. Differences were observed in names, such as *Pyhrn Autobahn* and *Süd Autobahn*. It was discovered that the same names can be inconsistent even within a single ASFiNAG document - at times written as compound words, at times separately. This means, that inconsistencies in writing names is not only crowd-sourced information issue but can be even caused by the varying information from official sources. Due to the contradictions found in ASFiNAG and the fact that motorway names generally are more often found to be written as separate words, bmvit document is used as a reference for the correct motorway names and their reference numbers.

It is worth noting that in not all countries motorways hold names, instead only reference numbers are used. The presence of motorway names in Austrian network, can be seen as an advantage, since each of the two can be corrected or filled using the knowledge of the other attribute.

[2] http://www.bmvit.gv.at/verkehr/strasse/autostrasse/planung/downloads/kategorisierung.pdf

[3] https://www.asfinag.at/documents/10180/13369/de_Buch+30+Jahre+ASFINAG.pdf/8af1a7eb-9bda-4f3c-86a9-81181d4bacb9

5.1.2 Methods and results

5.1.2.1 Intrinsic methods

Intrinsic methods allow to perform tasks independently from other data sources. This is crucial when no reliable third party data is available or easy to access. Therefore, in order to improve attribute accuracy, first, two intrinsic techniques, using only the available information within OSM data, are carried out. Motorway names or reference values are corrected based on other segments that have either one matching attribute (name or reference value) or, when matching is not possible, based on the neighboring segment attributes. Each of the methods have their own strength as well as imperfections, therefore, it is suggested to combine them.

a) Attribute based method

Description and initial test on area of Vienna

Due to the use of solely inherent data, a reliable look-up list of correct name and reference number pairs from given data is extracted. The primary idea of this method is, using the entries of the look-up list, to replace wrong or missing names and reference numbers. Further the process sequence is shortly described.

1.) **Creating a look-up list.** The first step of the look-up list creation is based on an SQL query stating a couple of conditions. Using regular expressions, all motorway names and reference values are selected, where motorway name consisted of a case insensitive sub-string *Autobahn* or *Schnellstraße*. Additionally, in this selection the reference value is not allowed to be *null* since such a case would not contribute to the look-up list. Distinct pairs of name and reference number are filtered and entries, where a single reference number within the list appears more then once, analyzed. These entries cannot be used in a look-up list without inspection, since they contain an error or a special case. It was found that selecting the shortest motorway name from all variations corresponding to each reference number, gives the best result. This is because OSM contributors in many of the wrong entries have added additional information to the actual name, such as direction, reference number etc., that results in a wrong name value.

For the look-up list strictly motorways with strings *Autobahn* or *Schnellstraße* in their name are used. This is done in order to filter and exclude different motorway objects, for instance, bridges and tunnels.

2.) **Assigning reference value when missing or different than in look-up table.** Finding all the motorway segments, whose names match the names from the look-up table. When a match has been found, assigning reference value from the look-up table. This adds reference value where it was missing and corrects it where it was wrong.

3.) **Assigning name when missing or different than in look-up table.** Finding all the motorway segments, whose reference value match the reference value from the look-up table. When a match has been found, assigning name from the look-up table. This adds name where it was missing and corrects it where it was wrong.

This step also includes exchanging bridge and tunnel names for motorway names. 80% of all tunnels in Austria exploit the motorway *name* key for the name value, instead of dedicated tag *tunnel:name* (chapter 4.2). Similar tendency can be observed with bridges. However, these names should be replaced with the motorway names on which they lay. Depending on the application, it might be necessary to keep information on bridge and tunnel names. In these cases they should be moved to their specific tag.

These steps were first applied on motorway network of Vienna. As a result all segments received their missing names and reference values, except for 8 segments that required location based fixing (figure 5.4). Further, the method is tested on motorway network of entire Austria.

Results after application on Austria

The intrinsic technique used in this thesis can find and solve some of the issues without the use of ground truth data. It can detect errors, such as missing data, point out to complex cases, when name or reference value cannot be guessed and give suggestions based on the information accessible in OSM data. Results of method application are further described step-by-step.

1.) **Creating a look-up list.** The extracted look-up list of Austrian motorways contains 29 rows.

2.) **Assigning reference value when missing or different than in look-up table.**

This step exposed the incompleteness of the look-up table. There are two cases when a motorway was not added to the look-up list. Both of those had no segment with a reference value, nor were they tagged as a motorway.

Completeness of reference values in Austria is substantial (sub-chapter 4.1) - only 16 segments were missing their reference value initially. After application of the function two segments found their reference value.

3.) **Assigning name when missing or different than in look-up table.**

After this function was applied, 1328 segments were renamed constituting 163 distinct road name entries from the wrong data set, 117 segments or 12 distinct roads obtained a missing name. Yet 7 motorways are still missing a name. It is necessary to note, that in this test also motorway objects, such as tunnels and bridges, got renamed. Excluding changes regarding these objects, 1050 actual motorway segments were renamed, constituting 44 distinct incorrect road name entries.

For the largest part of the roads that stayed unnamed, reference value is not consisting of the pattern *A#* or *S#*. This means, these segments have been mistakenly tagged as motorways but belong to a lower class roads.

Figure 5.1 presents a result of step-by-step name finding. A cutout of the initial motorway name list can bee seen in the image *(a)*. Next image shows existing name correction - for the most part compound names were split in two parts. In the final image the result, where empty rows are filled based on given reference numbers, can be observed. One can notice, that because *Südosttangente* in OSM data was consistently called with this exact name, and no external data was used, it did not get renamed like it is found in official Austrian motorway list (*Autobahn Südosttangente Wien*). Further improvements using third party data and fuzzy string matching can be done.

name text	name text	name text
Westautobahn	West Autobahn	West Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Wiener Außenring Autobahn	Wiener Außenring Autobahn	Wiener Außenring Autobahn
Südosttangente	Südosttangente	Südosttangente
Ostautobahn	Ost Autobahn	Ost Autobahn
Donauuferautobahn	Donauufer Autobahn	Donauufer Autobahn
Wiener Außenring Autobahn	Wiener Außenring Autobahn	Wiener Außenring Autobahn
Donauuferautobahn	Donauufer Autobahn	Donauufer Autobahn
Südostautobahn	Südostautobahn	Südostautobahn
Ostautobahn	Ost Autobahn	Ost Autobahn
Ostautobahn	Ost Autobahn	Ost Autobahn
Donauufer Autobahn	Donauufer Autobahn	Donauufer Autobahn
		Donauufer Autobahn
		Donauufer Autobahn
Westautobahn	West Autobahn	West Autobahn
Talübergang Wolfsgraben	Talübergang Wolfsgraben	West Autobahn
Westautobahn	West Autobahn	West Autobahn
Talübergang Wolfsgraben	Talübergang Wolfsgraben	West Autobahn
Westautobahn	West Autobahn	West Autobahn
Westautobahn	West Autobahn	West Autobahn
		Donauufer Autobahn
		Donauufer Autobahn
Südosttangente	Südosttangente	Südosttangente
Südosttangente	Südosttangente	Südosttangente
Donauuferautobahn	Donauufer Autobahn	Donauufer Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Süd Autobahn	Süd Autobahn	Süd Autobahn
Südosttangente	Südosttangente	Südosttangente
Laaerbergtunnel	Laaerbergtunnel	Laaerbergtunnel
Südosttangente	Südosttangente	Südosttangente
Laaerbergtunnel	Laaerbergtunnel	Laaerbergtunnel

(a) (b) (c)

FIGURE 5.1: Step-by-step motorway name finding. *(a)* - a cutout of the initial motorway list, *(b)* - existing name correction, *(c)* - empty cell filling.

In conclusion, the attribute based intrinsic method can find missing and correct wrong segment name and reference values solely based on existing attribute information. The necessary extraction of look-up list allows all the renamed segments of each highway hold consistent new attribute values.

As a disadvantage can be named the required analysis of the context and name characteristics in order to extract the look-up list. Names can be inconsistent and hold additional or wrong information, that can complicate the creation of the look-up list. The major disadvantage is that each segment needs some identifying attribute. Either name or reference value can be missing but not both. If this requirement is not met, the attribute values cannot be assigned. It should be also mentioned, that in case a segment is by mistake not tagged as motorway, it will not be checked and stay in its current state.

b) Location based method

There can be cases, when segments have neither name, nor reference number, or a single motorway name correspond to more than one reference value in the OSM data. Additionally, when tunnel and bridge names are wrongly added in motorway name tag, it would still be possible to retrieve the reference value of those motorway segments. It can happen, that different bridges or tunnels have the same name, even though they are located on different motorways (figure 5.2). Thus, if there are segments of those bridges without a given reference value, it cannot be fixed solely based on attribute information. A possible solution for these problem cases would be a location based renaming, where the neighbors of the segment in question can be checked. If both neighbors have the same name and reference value, their attribute information can be used to rename or assign a new reference value to the problem segment. A function to check for names with more than one possible reference value in the data, as well as location based renaming process was created an applied.

	ref character varying	old_name character varying	new_name character varying
1	A5	Absbergtunnel	Nord Autobahn
2	A2	Absbergtunnel	Süd Autobahn
3	A22	Absbergtunnel	Donauufer Autobahn
4	A21	Absbergtunnel	Wiener Außenring Autobahn
5	A4	Absbergtunnel	Ost Autobahn
6	A1	Absbergtunnel	West Autobahn
7	A4	Erdberger Brücke	Ost Autobahn
8	A21	Erdberger Brücke	Wiener Außenring Autobahn
9	A5	Erdberger Brücke	Nord Autobahn
10	A22	Erdberger Brücke	Donauufer Autobahn
11	A2	Erdberger Brücke	Süd Autobahn
12	A1	Erdberger Brücke	West Autobahn
13	A21	Grünbrücke	Wiener Außenring Autobahn
14	A1	Grünbrücke	West Autobahn
15	A4	Grünbrücke	Ost Autobahn
16	A5	Grünbrücke	Nord Autobahn
17	A2	Grünbrücke	Süd Autobahn
18	A22	Grünbrücke	Donauufer Autobahn

FIGURE 5.2: A piece of a table showing one tunnel or bridge name corresponding to several reference values. This particular table is a result of renaming motorway objects to motorway names. Some *Grünbrücke* segments were missing reference value - their reference number could not be filled without inspecting segment location.

Using the data of Vienna, there was a case found of two motorway bridges having the same name *Grünbrücke* and some segments of one or the other bridge having no reference value detected (figure 5.3). An assumption is made, that a segment between other segments with consistent name or reference value belongs to the same road. Therefore, a missing name or reference value can be obtained based on the neighboring segments. Such a method was found to be used also by Funke, Schirrmeister, and Storandt, 2015 in solving cases when street segments are missing name in OSM data. Here author goes one step further and considers more aspects, such as, whether unnamed edge is on the shortest path between two disconnected name components with the same name, the number of close-by named components, shortest path to the closest name component and other factors that are more relevant for street networks where data nature differentiates from motorways.

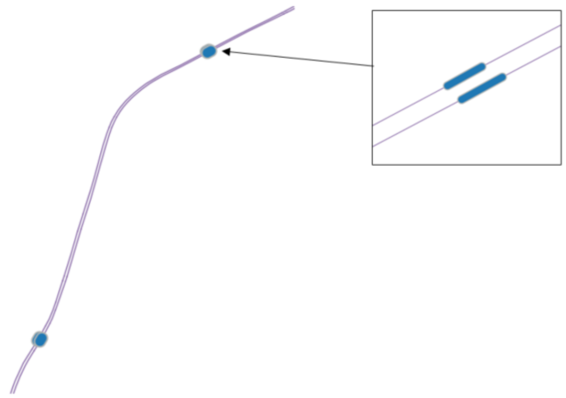

FIGURE 5.3: Location based renaming - case of *Grünbrücke* on motorway *A5*. In blue are marked segments with no reference value and name *Grünbrücke*.

The test area Vienna exposed 8 segments that did not have assigned reference value yet. Figure 5.4 shows the result before (a) and after (b) applying the script for neighbor check and updating the values. The rule is that only in case both neighboring segments have the same reference value, it would be assigned to the segment with missing value. This can be adopted to be more flexible in case there are too many segments missing reference number. For example, instead of checking only the direct neighbors, the new reference value can be retrieved from the two closest segments with an existing reference value. Vienna data is complete enough and all segments were updated with their new reference value based on direct neighbor attribute information.

	ref text	name text
1		Grünbrücke
2		Grünbrücke
3		Nord Autobahn
4		Nord Autobahn
5		Grünbrücke
6		Grünbrücke
7		Nord Autobahn
8		Nord Autobahn

(a)

	ref text	name text
1	A5	Grünbrücke
2	A5	Grünbrücke
3	A5	Nord Autobahn
4	A5	Nord Autobahn
5	A5	Grünbrücke
6	A5	Grünbrücke
7	A5	Nord Autobahn
8	A5	Nord Autobahn

(b)

FIGURE 5.4: Location based renaming - all missing reference numbers in Vienna are filled

Results after application on Austria

In figure 5.5 image *(a)* the motorway rows missing reference value are shown. Image *(b)* presents the output after applying the script for neighbor dependent value filling. As it can be seen, the result is not as promising as it was for Vienna data tested previously - only four segments found their reference value. This indicates that several segments in a row have been missing reference number and it is necessary to adjust the script to look for the closest neighbors with an existing value, instead of solely the direct neighbors.

	ref text	name text
1		Rheintal/Walgau Autobahn
2		
3		
4		
5		
6		
7		A7
8		A7
9		A7
10		
11		
12		
13		
14		
15		
16		Semmering-Schnellstraße

(a)

	ref text	name text
1		
2		
3		
4		
5		
6		A7
7		A7
8		
9		
10		
11		
12		
13	A14	Rheintal/Walgau Autobahn
14	A1	A7
15	A22	
16	S6	Semmering-Schnellstraße

(b)

FIGURE 5.5: The result of location based renaming applied on entire Austria. Image *a* - before, image *b* - after application of the script.

This same script was applied for segments with missing name value. The result reduced the number from 305 unnamed motorway segments to 276 unnamed segments, thus, consequently 24 segment names were found using their direct neighbor information.

The advantage of location based method compared to the attribute based intrinsic technique is that segments without both, name and reference value, can be identified. Furthermore, the procedure can be applied also for other highway classes. However, potentially more successful results can be achieved on long distance highways where the chance of finding a number of well attributed segments is higher. The method can search also for other attributes than name and reference value. Examples are pavement type and maximum speed tags. Nevertheless, the procedure should be applied with caution in order to limit the amount of assumptions, when applying new values to the attributes.

On the downside, this method compared to the attribute based intrinsic method cannot ensure attribute consistency within one road. Here it is not necessary to extract a look-up list, that, on one hand, speeds up the procedure, on the other had, looks at the attribute issue locally, ignoring the rest of the highway segments.

Judging from the results of the investigation, use of direct neighbor attributes is sufficient only when the data is close to fully complete. Besides, it is not the only precondition - result is better also if faulty segments are located disperse. On the contrary, if there are many segments in a row without attribute information, they cannot be fixed. The direct neighbor rule was found too strict for an optimal solution. This can be well observed when comparing results for reference value search and name search in Austria. Completeness of reference values is higher, yet the result of fixed attributes is better for names because of the different distribution of faulty segments. An improvement of the algorithm would be necessary, allowing to check not only the direct neighbors of each segment but instead the closest segments with an existing attribute value.

5.1.2.2 Extrinsic method

Intrinsic method is the only possibility when there is no external data available. However, extrinsic method can significantly improve the results and also increase the reliability of the outcome. For the extrinsic method additionally a list of official Austrian motorways and their reference numbers, provided by *bmvit* (Austrian Ministry for Transport, Innovation and Technology), is used. Attribute accuracy enhancement is based on fuzzy string matching, where similar word patterns are searched for. If the difference between strings is small enough according to a defined rule-set, and it is the best match found in the reference list of names, it is assumed to be the correct name.

Fuzzy string matching

Description and Vienna test area

Fuzzy string matching is a procedure of comparing two strings and searching for an approximate pattern match. The closeness of a match is called edit distance or *Levenshtein* distance. It is the minimum number of edit operations, such as insertion, deletion and substitution of a character, that would be needed to convert one string to be equal to the other (Onifade and Osofisan, 2011; Rodriguez, Kandel, and Bunke, 1997).

PostgreSQL offers its own fuzzy string matching module - *fuzzystrmatch*[4]. It contains several functions to measure distance between strings, and one of them also Levenshtein distance function. It is possible to define a specific cost parameter for every type of difference between strings (character insertion, deletion, substitution) in order to control the algorithm and fit it to the characteristics of the specific data. On the other hand, a Python package called *FuzzyWuzzy*[5] offers an easy interface with four possible Levenshtein based functions, where parameters are preset. This is the approach, that was chosen for matching similar OSM motorway name strings within this thesis. Next, each of the *FuzzyWuzzy* functions are shortly described.

1.) String Similarity (Ratio) - calculates the edit distance.

2.) Partial String Similarity (Partial Ratio) - ignores if the two strings in comparison have different length. If the shorter string is a and the longer string is b, this algorithm searches for the best matching sub-string of length a within string b. For example, two motorway names *Südosttangente* and *Autobahn Südosttangente Wien* would be a good match.

3.) Token Sort (Token Sort Ratio) - order of words in the two strings in comparison does not matter for the result (example, *West Autobahn* and *Autobahn West*).

4.) Token Set (Token Set Ratio) - first, words that are common in both strings are detected and then the reminder, in case one of the strings was longer than other, evaluated.

In order to find the best fitting function to match motorway names, experiments on data of Vienna were made (figure 5.6, 5.7, 5.8 and 5.9). Based on the result, the optimal method was chosen and applied on the data of entire Austria.

First the four functions described above were tested using given motorway names. Various cases were given to all of the functions - comparison of similar but different names (for example, *Ost Autobahn* and *Nordost Autobahn*), comparison of different length strings where first string is a substring of the other (*Südosttangente*, *Autobahn Südosttangente Wien*), comparison of similar strings when one of them has a typo, comparison of compound strings and single word strings, etc. A piece of experiment output is shown in figure 5.6 and figure 5.7.

[4]https://www.postgresql.org/docs/9.1/static/fuzzystrmatch.html
[5]http://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-in-python/

```
>>> fuzz.ratio('Ost Autobahn', 'Ostautobahn')
87
>>> fuzz.partial_ratio('Ost Autobahn', 'Ostautobahn')
82
>>> fuzz.token_sort_ratio('Ost Autobahn', 'Ostautobahn')
70
>>> fuzz.token_set_ratio('Ost Autobahn', 'Ostautobahn')
70
```

FIGURE 5.6: Resulting score of comparison between compound and separately written motorway name.

```
>>> fuzz.token_sort_ratio('Nordost Autobahn', 'Ost Autobahn')
86
>>> fuzz.token_set_ratio('Nordost Autobahn', 'Ost Autobahn')
86
>>> fuzz.partial_ratio('Nordost Autobahn', 'Ost Autobahn')
92
>>> fuzz.ratio('Nordost Autobahn', 'Ost Autobahn')
79
```

FIGURE 5.7: Resulting score of comparison between similar motorway names, where one of them is a sub-string of the other.

After the experiments, certain rules were defined to determine, whether a motorway name match to any name from a look-up list.

for name in OSM motorways **do**
  for lookupName in lookup list **do**
    if $fuzz.ratio(name, lookupName) = 100$ and $fuzz.token\_set\_ratio(name, lookupName) = 100$ **then**
      This is an absolute match
    else if $fuzz.ratio(name, lookupName) \geq 90$ and $fuzz.token\_set\_ratio(name, lookupName) \geq 50$ **then**
      This is 90 50 match
    else if $fuzz.ratio(name, lookupName) \geq 80$ and $fuzz.token\_set\_ratio(name, lookupName) \geq 60$ **then**
      This is 80 60 match
    else if $fuzz.ratio(name, lookupName) \geq 70$ and $fuzz.token\_set\_ratio(name, lookupName) \geq 90$ **then**
      This is 70 90 match
    else if $fuzz.ratio(name, lookupName) \geq 60$ and $fuzz.token\_set\_ratio(name, lookupName) = 100$ **then**
      This is 60 100 match
    else
      This is not a match

As it can be seen in pseudo-code, a combination of the two functions Ratio and Token Set Ratio were chosen to decide whether a match has been found. In figure 5.8 a part of the initial test results can be seen. The colored rows are matched names based on the defined rules. After a visual inspection the results were marked: in green – a valid match, in yellow – an invalid match where names are fairly similar, in red – a clear false positive. This mismatch can be observed with rather short names where the edit distance is small. For this reason, the search was adjusted to always give only the best match found.

As no official list of tunnel and bridge names is used, segments, that hold names of these objects in motorway name tag instead of dedicated name tag, are excluded from the string matching. Therefore, the sum of motorway segments here is different from the sum indicated in previous methods.

Figure 5.9 shows results of the best match search. Within Vienna 5 distinct motorway names have found absolute matches (when the name was already initially correctly written, such as *Wiener Nordrand Schnellstraße*) and 7 wrongly written motorway names have found their best match from the look-up list. Altogether those are 697 perfectly matched and 880 approximately matched segments respectively. Thus, out of 1577 motorway segments, that have an assigned name in the name field, all 1577 have found their best match from the look-up list.

Compared name	Reference name	Ratio	Token Set Ratio (TSR)
Wiener Außnring Autobahn	Wiener Außenring Autobahn	98	98
Weinviertel Schnellstraße	Weinviertler Schnellstraße	94	94
Donauuferautobahn	Donauufer Autobahn	91	51
Südostautobahn	Südost Autobahn	90	59
Süd Autobahn	Südost Autobahn	89	88
Ostautobahn	Ost Autobahn	87	70
Süd Autobahn	Nord Autobahn	80	84
Südosttangente	Autobahn Südosttangente Wien	67	100
Absbergtunnel	Südost Autobahn	29	30
Absbergtunnel	Ost Autobahn	32	32
Absbergtunnel	Nord Autobahn	23	31
Absbergtunnel	Nordost Autobahn	28	28
Absbergtunnel	Mühlkreis Autobahn	19	27
Absbergtunnel	Innkreis Autobahn	20	33
Absbergtunnel	Pyhrn Autobahn	22	30
Absbergtunnel	Tauern Autobahn	29	36
Absbergtunnel	Karawanken Autobahn	19	31
Absbergtunnel	Inntal Autobahn	21	36

FIGURE 5.8: Fuzzy string matching between given OSM motorway names in Vienna and reference data of correct names. Rows colored in green show a valid match, other colors - an invalid match.

Final pseudo-code with the search for the best match:

```
for name in OSM motorways do
    best_ratio = 0
    for lookupName in lookup list do
        if fuzz.ratio(name, lookupName) = 100 and
        fuzz.token_set_ratio(name, lookupName) = 100 then
            This is an absolute match
        else if fuzz.ratio(name, lookupName) ≥ 90 and
        fuzz.token_set_ratio(name, lookupName) ≥ 50 then
            This is 90 50 match
        else if fuzz.ratio(name, lookupName) ≥ 80 and
        fuzz.token_set_ratio(name, lookupName) ≥ 60 then
            This is 80 60 match
        else if fuzz.ratio(name, lookupName) ≥ 70 and
        fuzz.token_set_ratio(name, lookupName) ≥ 90 then
            This is 70 90 match
        else if fuzz.ratio(name, lookupName) ≥ 60 and
        fuzz.token_set_ratio(name, lookupName) = 100 then
            This is 60 100 match
        else
            This is not a match
        ratio_sum = best_ratio + best_token_set_ratio
        if ratio_sum > best_ratio then
            best_ratio = ratio_sum
```

Wrong Name	Matched Refrence Name	Ratio	Token Set Ratio
Südosttangente	Autobahn Südosttangente Wien	67	100
Donauuferautobahn	Donauufer Autobahn	91	51
Absbergtunnel	No match	-1	-1
Erdberger Brücke	No match	-1	-1
Grünbrücke	No match	-1	-1
Grünbrücke	No match	-1	-1
Grünbrücke	No match	-1	-1
Hochstraße Inzersdorf	No match	-1	-1
Hochstraße Kaiserebersdorf	No match	-1	-1
Laaerbergtunnel	No match	-1	-1
Praterbrücke	No match	-1	-1
Prater Hochstraße	No match	-1	-1
Schrägseilbrücke	No match	-1	-1
Talübergang Brentenmais	No match	-1	-1
Talübergang Wolfsgraben	No match	-1	-1
Tunnel Eibesbrunn	No match	-1	-1
Tunnel Kreuzenstein	No match	-1	-1
Tunnel Rannersdorf	No match	-1	-1
Tunnel Rustenfeld	No match	-1	-1
Tunnel Schwechat	No match	-1	-1
Tunnel Stetten	No match	-1	-1
Tunnel Tradenberg	No match	-1	-1
Tunnel Vösendorf	No match	-1	-1
Ostautobahn	Ost Autobahn	87	70
Südostautobahn	Südost Autobahn	90	59
Weinviertel Schnellstraße	Weinviertler Schnellstraße	94	94
Westautobahn	West Autobahn	88	64
Wiener Außnring Autobahn	Wiener Außenring Autobahn	98	98

FIGURE 5.9: Fuzzy string matching between given OSM motorway names in Vienna and reference data of correct names in order to find the best match. Rows colored in green show a valid match.

Results after application on Austria

After fuzzy string matching parameter adaption to the characteristics of motorway names in Austria and performed tests within Vienna motorway network, the same functions are applied to motorways in entire Austria. Absolute match was observed for 21 and approximate match for 10 distinct motorway names. That accounts for 7907 perfectly matched motorway segments and 2956 approximately matched segments. This makes it 10863 matched segments out of 11208 with existent name value. The result is shown in figure 5.10. The best match was determined by the sum of resulting points in functions *Ratio* and *Token Set Ratio*.

The existence of name in data does not indicate the correctness of the name value. Following table 5.1 presents comparison of the name correctness based on the official motorway names in Austria and the name value presence analyzed in sub-chapter 4. Tunnels and bridges in this analysis are excluded, however, it is not known if the segments missing names belong to any of these objects.

Compared name	Best match	Ratio Sum	Compared name	Perfect match	Ratio Sum
Donauuferautobahn	Donauufer Autobahn	142	Brenner Autobahn	Brenner Autobahn	200
Karawankenautobahn	Karawanken Autobahn	146	Brucker Schnellstraße	Brucker Schnellstraße	200
Ostautobahn	Ost Autobahn	157	Burgenland Schnellstraße	Burgenland Schnellstraße	200
Semmering-Schnellstraße	Semmering Schnellstraße	196	Donauufer Autobahn	Donauufer Autobahn	200
Südostautobahn	Südost Autobahn	149	Innkreis Autobahn	Innkreis Autobahn	200
Südosttangente	Autobahn Südosttangente Wien	167	Inntal Autobahn	Inntal Autobahn	200
Tauern Autobahn - A10	Tauern Autobahn	183	Kremser Schnellstraße	Kremser Schnellstraße	200
Weinviertel Schnellstraße	Weinviertler Schnellstraße	188	Mattersburger Schnellstraße	Mattersburger Schnellstraße	200
Westautobahn	West Autobahn	152	Mühlviertler Schnellstraße	Mühlviertler Schnellstraße	200
Wiener Außenring Autobahn	Wiener Außnring Autobahn	196	Murtal Schnellstraße	Murtal Schnellstraße	200
			Nordost Autobahn	Nordost Autobahn	200
			Pyhrn Autobahn	Pyhrn Autobahn	200
			Rheintal/Walgau Autobahn	Rheintal/Walgau Autobahn	200
			Semmering Schnellstraße	Semmering Schnellstraße	200
			Stockerauer Schnellstraße	Stockerauer Schnellstraße	200
			Süd Autobahn	Süd Autobahn	200
			Tauern Autobahn	Tauern Autobahn	200
			Welser Autobahn	Welser Autobahn	200
			West Autobahn	West Autobahn	200
			Wiener Außenring Schnellstraße	Wiener Außenring Schnellstraße	200
			Wiener Nordrand Schnellstraße	Wiener Nordrand Schnellstraße	200

FIGURE 5.10: Fuzzy string matching results for Austrian motorway network. On the left side are shown distinct motorway names that have been corrected as they found their best match from an official motorway name table. On the right side are motorway names that found a perfect match, thus, were initially correctly written.

Name value

Area	All segments	Have a name	Originally Correct	Match found
Vienna	1655	1577	697	1577
Austria	11513	11208	7907	10863

TABLE 5.1: Comparison of number of segments with present and originally correct names, as well as names that have found either perfect or approximate match.

Many of the segments for which a match could not be found, are either containing their reference value in the cell which is supposed to hold the name, or the direction of the road (figure 5.11).

The main strength of the fuzzy string matching is its ability to find typos and correct wrongly written names. To do so, the method does not require additional knowledge of segment identity, such as reference number, compared to attribute based intrinsic method discussed earlier. Fuzzy string matching can be applied also on other highway classes, however, difficulties can arise with street names where many closely similar names are possible.

The disadvantage of the method are parameter settings. Parameters need to be tested and fine-tuned to fit the characteristics if the data. Furthermore, fuzzy string matching does not deal with missing names, therefore, additional methods are necessary.

A7 Linz Prag - Mühlkreisautobahn	A7
A7 Prag Freistadt Wiener Straße	A7
A7 Prag Freistadt Wiener Straße	A7
A7 Prag Leonfelden voestalpine	A7
A7 Wien Salzburg Linz SW	A7
A7 Wien Salzburg Linz-Zentrum	A7
A7 Wien Salzburg Linz-Zentrum	A7
Achselgraben-Galerie	A10
Achselgraben-Galerie	A10
Autostrada Alpe-Adria	A23
Autostrada Alpe-Adria	A23
Autostrada Alpe-Adria	A23
Autostrada Alpe-Adria	A23
Autostrada del Brennero - Brenner Autobahn	A13
Autostrada del Brennero - Brenner Autobahn	A13
Autostrada del Brennero - Brenner Autobahn	A13

FIGURE 5.11: Segments that failed to be matched to any name from the look-up list

5.1.3 Conclusion

Undoubtedly, the way towards semantic homogeneity in OSM data is laborious. Motorways, as the most important roads within any country, represent only 1,8% [6] of the whole Austrian road network data. Characteristics of each highway class have to be analyzed separately in order to create appropriate methods for automatic error detection and attribute information improvements for the entire road network.

Similarly like Zhang and Tinghua, 2015 have acknowledged, also this thesis concludes that the collaborative nature of OSM data collection inevitably leads to tag inconsistencies. Compared to the *bmvit* reference data, around 33% of all motorway segments were either incorrectly written or missing. Automated error detection tools are a huge help in the improvement of data. However, there are tasks when clean data is necessary all at once and needs to be manipulated before any further task can be performed. In these cases, automated methods, such as the ones described in this chapter, can be applied. Advantages and disadvantages of each of the proposed methods in their respective sections are discussed. The intrinsic methods are able to assign a reference value or name to a segment missing those attributes, without the need for ground truth data. Specially outshines the location based method that can assign missing values to segments with no attributes but they highway class. On the other hand, the extrinsic method, fuzzy string matching, can correct wrongly written name attributes without the knowledge of their reference number and information on neighboring segments. Potentially, one of the optimal solutions would be based on the combination of all here described techniques.

[6] https://www.bmvit.gv.at/verkehr/gesamtverkehr/gvp/faktenblaetter/umwelt/fb_strasse_schiene_netz.pdf

5.2 Topological data structure

In chapter 2 one of the cartographic demands stated that, in order to satisfy many tasks performed on a map, junctions, connections to different points of interest and populated places should be accurate and clearly legible. Another demand expressed that generalization of the map should be suiting the scale and generalization conflicts should be solved. To achieve all these goals in an automated way, it is necessary to know the underlying topology.

Theobald, 2001 refers to topological data structure as data structure in which the inherent spatial connectivity and adjacency relationships of features are explicitly stored. Maintaining a topological structure offers a number of benefits. Few of the advantages, that relate to road network, are explicit neighborhood relations and a possibility to apply automated methods dealing with self intersections, overshoots, undershoots and gaps (Theobald, 2001).

Topology describes how the features can be spatially related. Therefore, it can be used to analyze spatial relationships to, for example, dissolve road segments based on a common attribute value (ESRI, 2016). It also can be used to enforce integrity rules (no disconnected graph parts, gaps between two highways with the same name, etc.), support topological relationship queries and edits based on these relationships (for example, select all link roads between motorway *A23* and *A21* and dissolve) (ESRI, 2016).

In the past three decades extensive efforts towards development of various automatic generalization methods have been made. Selection operator by various attributes, such as road length, importance, connectivity, density, is widely used within road networks. In medium to small scale motorway maps selection operator becomes the handiest in harmonizing the motorway density between rural and urban areas in country wide maps (Chaudhry and Mackaness, 2005). A wide number of studies concerning this topic exist. Commonly, algorithms are based on graph theory forming edges and vertices to assure a connected graph – topologically connected road network - after use of selection operator (Mackaness and Mackechnie, 1999). These are just a few use cases that stress out the importance of topological descriptiveness of road networks within the underlying data to support visualization tasks.

OSM offers a data model element - relations. As described in chapter 2, relations model logical or geographic relationships between objects. Use of relations can have a significant role in describing topology. However, it can be often observed that gaps in relations, overlapping relations, relations with questionable importance and incomplete relations exist in the data (figure 5.12). Therefore, alternative solutions to extract topology information should be examined. In this thesis topological descriptiveness of the OSM motorway network is viewed from a perspective of two questions - if it is possible to detect which motorways connect to each other and how to do it. This information can be crucial for tasks like use of generalization selection operator or spatial analysis on motorways.

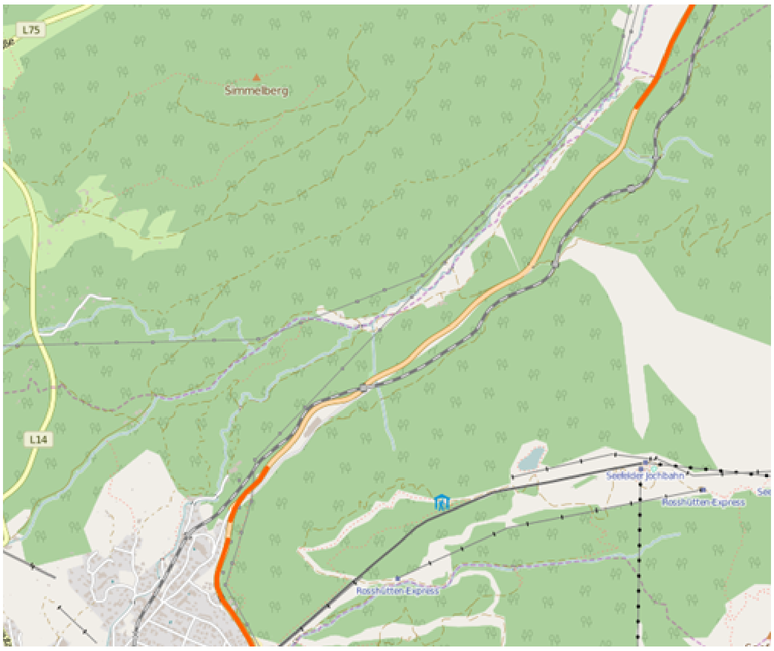

FIGURE 5.12: A hole in a relation in Austria, Tirol. Dark orange - the relation, orange - the primary highway

5.2.1 Methodology

Detection of "linking areas" of motorways

As discussed in chapter 3, link roads tagged with *highway=motorway_link* indicate an exit to surface streets (motorway to non-motorway) or an exit to an interchange (motorway to other motorway). For most tasks on motorway data, such as motorway generalization, it is necessary to identify the latter ones. Based on the analysis done in chapter 3, it is known that only 13% of all motorway links have a reference value that indicates which motorway the link is exiting. The dedicated tag *destination:ref* that specifies the reference number of motorway the link leads to, is used only in few cases.

Depending on the goal of the map, these areas can serve various purposes. For instance, when creating car navigation maps, it is possible to extract information on which motorways are supposed to be linked in particular linking areas and thus either broken motorway topology can be recognized and fixed, or complex junctions simplified. Another use case is representing road network connections for network generalization. This is especially useful in cases when one motorway merges into another. Here one motorway segments do not geometrically intersect the other motorway segments as they are separated by motorway links (figure 5.13). Therefore, detecting linking areas using the motorway links is crucial, in order to obtain the information on topology between the motorways.

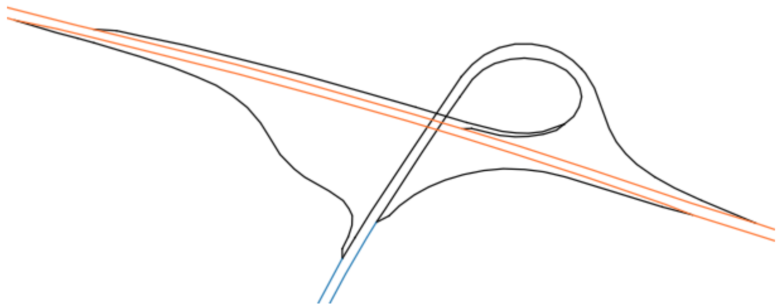

FIGURE 5.13: Motorways are linked with each other using segments tagged as *highway = motorway_link*. These segments almost never hold information on which highways they exit and enter. In blue and orange - two different motorways, in black - motorway link segments.

Finding linking areas consists of several steps. First issue is that motorway exits in OSM data consist of many separate segments that do not have any common attribute to identify them in groups. To detect which highways should be connected, highway links should be grouped based on their segment intersection with each other.

Step 1 – Collecting all link segments that intersect/ touch one another in sets. If a set of such link segments intersects at least two different motorways, it is considered as a significant set (figure 5.14). Significant sets serve as proofs for certain highway connectivity in certain area. The differentiation between significant and insignificant link segment sets is due to complex junctions, where many segment sets either do not touch a motorway or touch only one of them. In these cases they do not carry the important information of connectivity, or they connect a motorway with a lower class road. Apart from that, also links that connect a motorway to a lower class road, are tagged *highway=motorway_link*. These links are not added to the significant link sets.

1.) Collect all the links that intersect some motorway in a list.

for link i in motorway links **do**
  for motorway m in motorways **do**
    if i intersects m **then**
      Add i, m to intersectingLinks list

2.) Collect all the links that intersect one another in pairs.

for link i, j in distinct combinations of i and j in motorway links **do**
  if i intersects j **then**
    Add pair i and j to an array in intersectingLinkPairs list

3.) Group together all link segments that intersect one another.

for pair i in motorway intersectingLinkPairs **do**
  for group in intersectingLinkGroups **do**
    if i intersects group **then**
      update group with i
    else
      append group i to intersectingLinkGroups

4.) For each link that intersects some motorway, count how many times that happens: if more than once, add its whole group to significant link list, if once, add it to potentially significant links.

for link i in motorway intersectingLinks **do**
  count how many times link i intersects some motorway
  if count of motorways i intersects > 1 **then**
    append group in intersectingLinkGroups with i to significantLinks
  else if count of motorways i intersects = 1 **then**
    append group in intersectingLinkGroups with i to potentialSignificantLinks

5.) Check if in a potentially significant link group there are at least two link segments each intersecting some motorway. If such a group is found, add the whole group to significant links.

for link group in potentialSignificantLinks **do**
  count how many links in group intersect motorway and how many different ones
  if count of intersected motorways > 1 **then**
    append group in potentialSignificantLinks to significantLinks

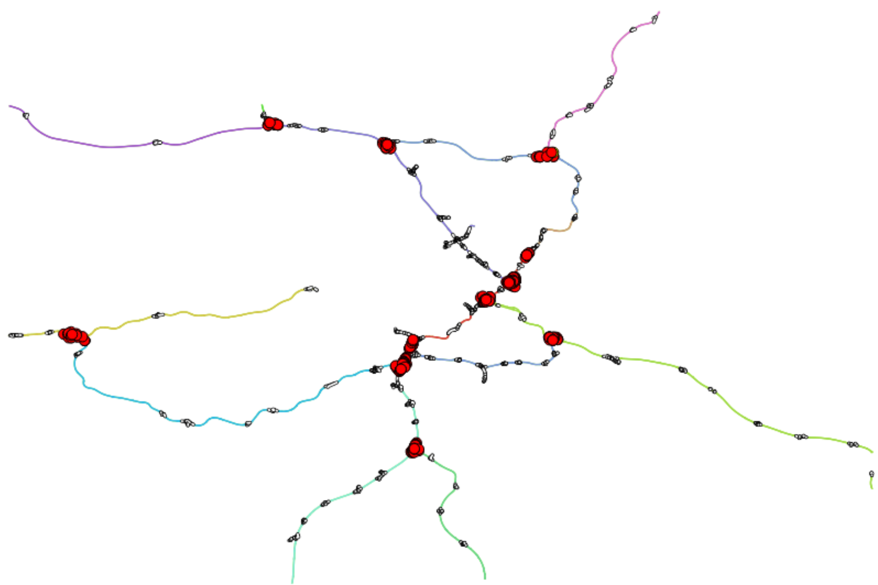

FIGURE 5.14: The result after analysis of significant link detection in Vienna. In red - nodes of significant links, in white with black outline - nodes of insignificant links

Step2 – Defining the "linking area". If a particular task goes beyond the observation of which motorways are connected to each other but also the location where linking happens matters, these areas have to be identified (figure 5.15). This can be done by creating an appropriately sized buffer around grouped links and giving each separate buffer an identification. Identification of separate buffers is necessary since each motorway can have several junctions.

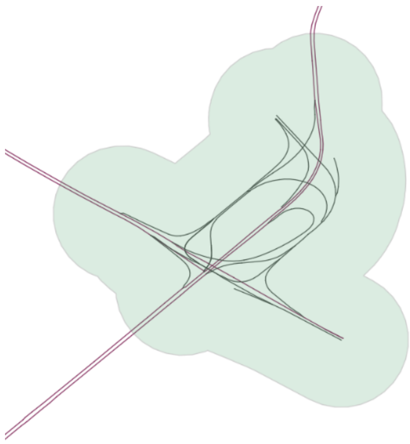

FIGURE 5.15: Motorway linking area

Step3 – Performing further actions with linking areas. Below are examples of detecting road dangles and calculating centroid of linking area.

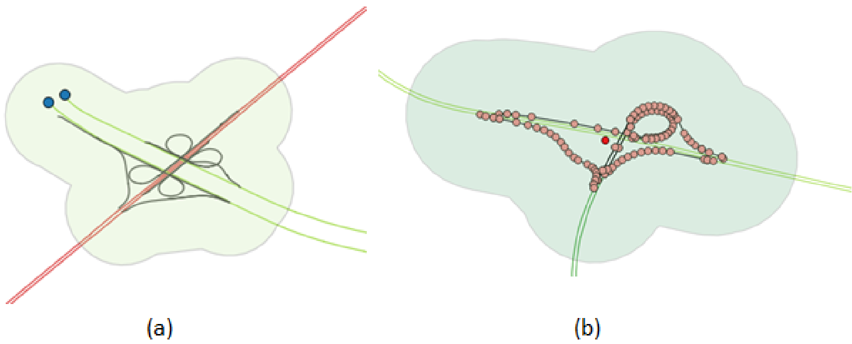

FIGURE 5.16: Further actions on motorway linking areas. Image *(a)* shows detection of motorway dangle within a linking area. Image *(b)* shows detection of centroid of linking area - a point feature representing two motorway intersection. In red - centroid, in light red - link segment vertices.

5.2.2 Method and Results

The outcome of the detection of linking areas can be seen in figure 5.17. Result consisted of 24 correctly identified linking areas, 2 false positives and 2 false negatives. However, it is worth mentioning that one of the false negatives was not supposed to be identified as junction using this method.

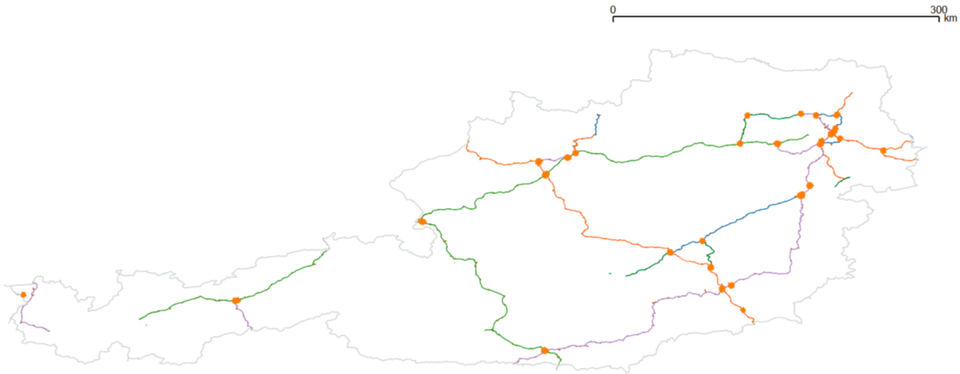

FIGURE 5.17: Overview of detected motorway linking areas in Austria. Areas in orange - linking areas, color-coded lines - motorways.

False positive linking areas (figure 5.18) appear due to various reasons. Image *(a)* contains some motorway segments for which no fitting reference value was found (in lighter green color). From a visual inspection, this row of segments seem to be not belong to motorways and most probably should be removed. Since this group of segments was connected to the motorway links, it was mistakenly identified as motorway junction. In image *(b)*, on the other hand, motorway *(A2)* connects to itself. The orange motorway segments are representing a short motorway *(A2)* extension towards North, that was tagged with a non-existent reference value *(A2z)*.

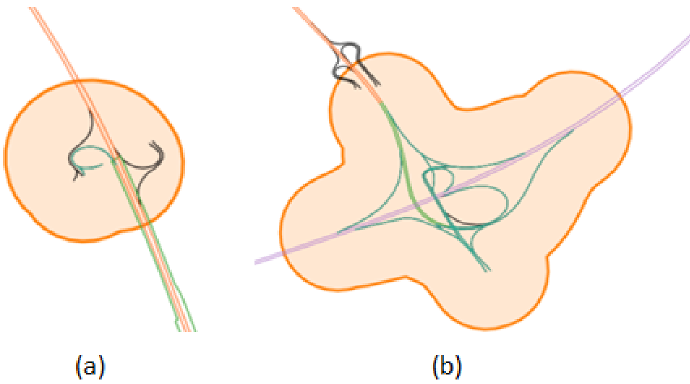

FIGURE 5.18: Detected false positive linking areas. In orange - linking areas, in dark green - significant links, in black - links that were not found significant.

Figure 5.19 image *(a)*, is falsely not detected as motorway junction. Here seemingly the issue is with link segments that do not actually connect to each other, thus they were not added to a group of links that intersect more than one motorway. Image *(b)* represents a case when motorways are connected without use of links. This situation points out that, firstly, all motorway junctions, where motorways intersect directly, should be detected.

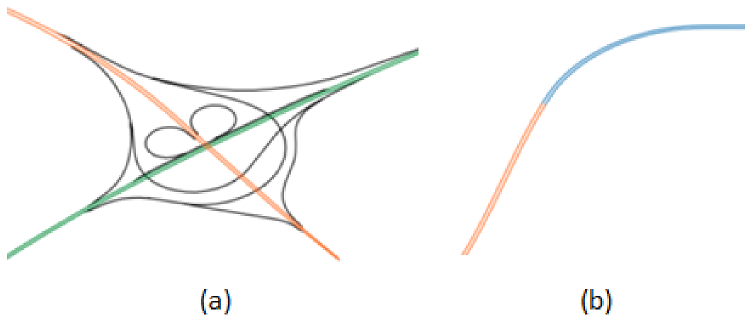

FIGURE 5.19: Detected false negative linking areas. In black - linking areas, that were found insignificant. Image *(a)* - junction not found due to not existent link intersection, image *(b)* - junction not found due to motorway connection without a use of link segments.

5.2.3 Conclusion

Obtaining information on topological relationships between motorways is not a trivial task using OSM data. Junctions in this highway class are commonly complex structures that use link road segments, such as, ramps to connect motorways. Also on the level of data, two motorways are rarely directly intersecting, thus, unless link segments are well attributed, motorway intersection information is not maintained.

The introduced technique, to cope with the issue, was found to work successfully, identifying almost all junctions and connected motorways. The method was tested only on the motorway network. However, this method is useful also for other highway classes that use link roads to connect to other roads. Such highways are trunks, primary, secondary and tertiary roads. Furthermore, as an additional outcome of the procedure all link segments are identified in specific link roads and junctions. Therefore, further tasks such as link simplification can be performed.

On the drawback, the approach heavily relays on OSM contributor precise link segment depiction, that ensures segment connection to each other as well as to the motorways. Also gaps in link roads and cases, where link segment is not defined as *motorway_link*, would break the chain of intersecting link segments. If this kind of an issue would occur at several instances within one junction, a connection might not be recognized.

Chapter 6

Conclusion

OpenStreetMap was a groundbreaking VGI project when it started off a decade ago and still keeps on gaining popularity. It is an impressive geodata collection, visualization and sharing tool, which is used in a variety of business and scientific applications. Yet, it is questionable how straightforward it is to derive cartographically pleasing map visualizations using OSM data. Picking roads as an example of fundamental map elements, there is a set of rules, such as clear junctions, accurate labels and class dependent symbolization, that have to be followed so as to result in a functional and readable depiction. In this regard, the main aim of this thesis was to answer the question: What is the relation between cartographic demands and OpenStreetMap motorway data?

In order to find that out, first, a list of cartographic demands regarding road depiction was defined. The final list was based on researched road generalization constraint classification, common generalization pitfalls and examination of scale dependent map use.

Next, it was studied whether OpenStreetMap motorway data has a potential to produce maps, that are in agreement with the cartographic demands. Examination was based on two areas of different extents - Vienna and Austria. Major results were obtained in points regarding completeness, detailed and correct labeling and topological descriptiveness. Finally, methods for improving OpenStreetMap motorway data were developed, tested and results discussed.

OSM offers its users raw data, that has a considerable amount of adjustments necessary to result in a high quality map. Removing the redundant data to suit the target scale, preserving and emphasizing the important features to fit the intended topic are just few generalization and enhancement tasks that might be necessary. However, to perform those tasks, the imperfections of the underlying data have to be resolved. Thus, for instance, gaps in attribute information influence the result of semantic generalization. Similarly, vague topological information complicate junction simplification.

This work has found that OSM data based motorway map generation encounters difficulties to align with cartographic demands, if the underlying data is not enriched. Two of the defined demands drew a special attention - detailed and correct labeling and topological descriptiveness. Correctness of motorway identification attributes was found to be low. However, it is an important aspect to recognize separate road segments as parts of a certain motorway and perform a good quality road labeling. Regarding the second demand, no obvious way to extract knowledge of interconnected motorways was detected. Yet, to depict clear junctions or perform tasks on a road graph, topological relations should be known.

The described methods to improve OSM data and, thus, depict high quality highway maps, have successfully indicated a good potential. They have shown, that OSM data, in countries with similar major road completeness to Austrian network, is in a sufficient level to use automated data enhancement methods. The results have presented that the procedures described in this work are appropriate for OSM data characteristics and can fix a significant amount of errors

The techniques are applicable also to other elements and tags. Besides using the attribute accuracy improvement methods for motorway name and reference number, they can be applied to other tags, such as maximum speed and road pavement type. Fuzzy string matching can correct any name attributes, if reference data is available. The method for extracting motorway topological information can be additionally used with other major road classes than motorways.

Regarding the research question, what is the relation between cartographic demands and OpenStreetMap motorway data, it has been shown that after just a few surgical changes OSM data has a great potential to be used as a basis for neat motorway network depictions. A number of issues, based on visual observations and underlying data, after the analysis were evident. However, the results of developed methods have demonstrated, that the state of the data is sufficient to cope with the assessed problems.

6.1 Open discussion and future work

Keeping in mind the weaknesses and strengths, derived from OSM data model, the following three aspects may be considered:

- Automated methods, such as those described in this thesis, could be developed further in order to cope with OSM data imperfections.
- Rigorous checks could prevent incomplete or wrong data from being inserted into OSM.
- The OSM data model and the way data is inserted might need improvements.

It is clear that development of perfect, automated methods, that detect and fix data without any manual interference, is unfeasible. Likewise, since the OSM project is highly dependent on contributor motivation, creating a strict data insertion procedure that allows only high quality data imports, might result in a decreased number of volunteers and, thus, OSM popularity.

The future suggestion is to create a system where data insertion is not made more difficult but instead guides the user, that restricts the actions of contributors to a higher level than it is currently but keeps their motivation. Introducing system where users with certain expertise have access to certain actions, such as adding new tags, additionally, cross-referencing existing tags with similar meaning, regulating the creation of new highway relations and similar limitations could possibly improve the quality of data. Better guided and more consistent data import would potentially lessen the necessity for error correction method development and make further work with OpenStreetMap easier.

Moreover, data enhancement methods, such as the ones described in this thesis, cannot make logical assumptions and enrich data without a sufficient amount of already existing data or third party sources. Currently there is a number of quality assurance (QA) tools detecting anomalies and errors in OSM. More active editing of the detected errors, would improve the results of the automated methods. However, users are expected to have their own initiative to use these tools and choose the tasks.

In order to improve the speed of data fixing, the small tasks from the QA tools could be assigned to contributors based on their current location, district or city in a question form. That question could ask to approve or decline correctness of some information in the map and invite to fix it. The more correct data will be contributed to OSM, the easier it will be to tackle the imperfections and, thus, satisfy a number of cartographic demands.

In further research, the rules and parameters of the methods presented in this work can be adapted. Particularly, in case of location based attribute accuracy improvement method, specifying defined rules more thoroughly, would lead to more intelligent decisions. Furthermore, none of the methods individually is able to address all cases of attribute issues. However, a combination of these techniques working together would possibly give impressive results. Enhancing the individual methods and developing a certain work-flow to address the issue as a whole, would be a further challenge to tackle.

Bibliography

Al-Bakri, Maythm (2015). “Ten Years of OpenStreetMap Project : Have We Addressed Data Quality Appropriately ?” In: *Journal of Engineering* 21.4, pp. 158–175.

Arsanjani, Jamal Jokar et al. (2013). “Assessing the Quality of OpenStreetMap Contributors together with their Contributions”. In: *16th AGILE conference*.

Barron, C., P. Neis, and A. Zipf (2013). “Towards Intrinsic Quality Analysis of OpenStreetMap Datasets”. In: *Online proceedings of the International Workshop on Action and Interaction in Volunteered Geographic Information (ACTIVITY) at the 16th AGILE Conference*. Leuven, pp. 1–2. ISBN: 978-3-319-00615-4.

Buttenfield, Barbara P (1991). *Map Generalization: Making rules for knowledge representation*. Ed. by Robert B McMaster. Longman Group United Kingdom, pp. 150–239. ISBN: 0582080622.

Cecconi, Alesandro and Martin Galanda (2002). “Adaptive zooming in web cartography”. In: *Computer Graphics Forum* 21.4, pp. 787–799. ISSN: 01677055. DOI: 10.1111/1467-8659.00636.

Chaudhry, Omair and William Mackaness (2005). “Rural and Urban Road Network Generalisation Deriving 1:250,000 from OS MasterMap”. In: *XXII International Cartographic Conference (ICC2005)*.

Devogele, Thomas, Jenny Trevisan, and Laurent Raynal (1996). “Building a multi-scale database with scale-transition relationships”. In: *Proceedings of 7th International Symposium on Spatial Data Handling*, pp. 337–351.

Duchêne, C, M Barrault, and K Haire (2001). “Road network generalization: A multi agent system approach”. In: *Proceeding of ICC 2001*, pp. 1–9.

ESRI (2016). *Topology basics*. URL: http://desktop.arcgis.com/en/arcmap/10.3/manage-data/topologies/topology-basics.htm.

Funke, Stefan, Robin Schirrmeister, and Sabine Storandt (2015). “Automatic Extrapolation of Missing Road Network Data in OpenStreetMap”. In: *Proceedings of the 2nd InternationalWorkshop on Mining Urban Data*. Lille.

Haklay, Mordechai (2010). “How good is volunteered geographical information? A comparative study of OpenStreetMap and ordnance survey datasets”. In: *Environment and Planning B: Planning and Design* 37.4, pp. 682–703. ISSN: 02658135. DOI: 10.1068/b35097.

Harrie, Lars and Robert Weibel (2007). “Modelling the Overall Process of Generalisation”. In: *Generalisation of Geographic Information*, pp. 67–87. DOI: 10.1016/B978-008045374-3/50006-5.

Hashemi, Payman and Rahim Ali Abbaspour (2015). "Assessment of Logical Consistency in OpenStreetMap Based on the Spatial Similarity Concept Peyman". In: *OpenStreetMap in GIScience*. Ed. by Jamal Jokar Arsanjani et al. Switzerland: Springer International Publishing, pp. 1–20. ISBN: 978-3-319-14279-1. DOI: 10.1007/978-3-319-14280-7.

Li, Qiuping et al. (2014). "Polygon-based approach for extracting multilane roads from OpenStreetMap urban road networks". In: *International Journal of Geographical Information Science* 28.11, pp. 2200–2219. ISSN: 1365-8816. DOI: 10.1080/13658816.2014.915401.

Liu, Y et al. (2003). "Categorical Database Generalization Aided By Data Model". In: August, pp. 10–16.

Loidl, Martin and Stefan Keller (2015). *An intrinsic approach for the detection and correction of attributive onconsistencies and semantic heterogenity in OSM data*. Chicago.

Mackaness, William A and Gordon A Mackechnie (1999). "Automating the Detection and Simplification of Junctions in Road Networks". In: *GeoInformatica*, pp. 185–200.

Neis, Pascal, Dennis Zielstra, and Alexander Zipf (2013). "Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions". In: *Future Internet* 5.2, pp. 282–300. ISSN: 1999-5903. DOI: 10.3390/fi5020282.

Onifade, Olufade F. W. Olufade and Adenike O. Osofisan (2011). "Resolving Ambiguous Queries via Fuzzy String Matching and Dynamic Buffering Techniques". In: *Information Intelligence, Systems, Technology and Management*. Ed. by S. Dua, S. Sahni, and D.P. Goyal. Springer Science, pp. 198–205. DOI: 10.1007/978-3-642-19423-8{_}21.

OpenSteetMap Wiki (2016). *How can a project like this create accurate maps?* URL: http://wiki.openstreetmap.org/wiki/FAQ{\#}How{_}can{_}a{_}project{_}like{_}this{_}create{_}accurate{_}maps.3F.

OpenStreetMap Wiki (2014a). *AND Data*. URL: http://wiki.openstreetmap.org/wiki/AND{_}Data.

— (2014b). *WikiProject Austria/Liste Autobahnbrücken*. URL: http://wiki.openstreetmap.org/wiki/WikiProject{_}Austria/Liste{_}Autobahnbr{\%}C3{\%}BCcken.

— (2015a). *Elements*. URL: http://wiki.openstreetmap.org/wiki/Elements.

— (2015b). *Import/Software*. URL: http://wiki.openstreetmap.org/wiki/Import/Software.

— (2015c). *Key:area*. URL: http://wiki.openstreetmap.org/wiki/Key:area.

— (2015d). *Key:highway*. URL: http://wiki.openstreetmap.org/wiki/Key:highway.

— (2015e). *Stats*. URL: http://wiki.openstreetmap.org/wiki/Stats{\#}Active{_}new{_}and{_}old{_}contributors{_}per{_}year.

— (2015f). *Types of relation*. URL: http://wiki.openstreetmap.org/wiki/Types{_}of{_}relation{\#}Established{_}Relations.

— (2015g). *WikiProject Austria/geoimage.at*. URL: http://wiki.openstreetmap.org/wiki/WikiProject{_}Austria/geoimage.at.

— (2016a). *Aerial imagery*. URL: http://wiki.openstreetmap.org/wiki/Aerial{_}imagery.
— (2016b). *Automated Edits code of conduct*. URL: http://wiki.openstreetmap.org/wiki/Automated{_}Edits{_}code{_}of{_}conduct.
— (2016c). *Conditional restrictions*. URL: http://wiki.openstreetmap.org/wiki/Conditional{_}restrictions.
— (2016d). *Controlled-access highway*. URL: https://en.wikipedia.org/wiki/Controlled-access{_}highway.
— (2016e). *Key:access*. URL: http://wiki.openstreetmap.org/wiki/Key:access.
— (2016f). *Key:bridge*. URL: http://wiki.openstreetmap.org/wiki/Key:bridge.
— (2016g). *Map Features*. URL: http://wiki.openstreetmap.org/wiki/Map{_}Features.
— (2016h). *Rendering*. URL: http://wiki.openstreetmap.org/wiki/Rendering.
— (2016i). *Tag:highway=motorway*. URL: http://wiki.openstreetmap.org/wiki/Tag:highway{\%}3Dmotorway.
— (2016j). *Vertical Aerial Photographs*. URL: http://wiki.openstreetmap.org/wiki/Vertical{_}Aerial{_}Photographs.
— (2016k). *Way*. URL: http://wiki.openstreetmap.org/wiki/Way.
OSMstats (2016). *Edits Per Country*. URL: http://osmstats.neis-one.org/?item=countries{\&}date=6-7-2016.
Pascal Neis (2012). *Distribution of Active Users in OpenStreetMap – Oct-Nov 2012*. URL: http://neis-one.org/2012/11/active-users-osm-nov12/.
Pearson, David W. and Jean Christophe Janodet (2009). "String Distances and Uniformities". In: *Adaptive and Natural Computing Algorithms*, pp. 333–339. ISBN: 3642049206. DOI: 10.1007/978-3-642-04921-7{_}34.
Rodriguez, W., a. Kandel, and H. Bunke (1997). "3D-curve similarity using fuzzy string matching". In: *Proceedings of 6th International Fuzzy Systems Conference* 1, pp. 79–82. DOI: 10.1109/FUZZY.1997.616348.
Sehra, Sukhjit Singh (2014). "Assessing the Topological Consistency of Crowd-sourced OpenStreetMap Data". In: *Human Computation* 1.2, pp. 265–280. ISSN: 2330-8001. DOI: 10.15346/hc.v1i2.13.
Sehra, Sukhjit Singh, Jaiteg Singh, and Hardeep Singh Rai (2014). "A Systematic Study of OpenStreetMap Data Quality Assessment". In: *2014 11th International Conference on Information Technology: New Generations*. Institute of Electrical & Electronics Engineers (IEEE), pp. 377–381. ISBN: 978-1-4799-3188-0. DOI: 10.1109/ITNG.2014.115.
Senaratne, Hansi et al. (2016). "A review of volunteered geographic information quality assessment methods". In: *International Journal of Geographical Information Science*, pp. 1–29. ISSN: 1365-8816. DOI: 10.1080/13658816.2016.1189556.
Stoter, Jantien et al. (2009). "Methodology for evaluating automated map generalization in commercial software". In: *Computers, Environment and Urban Systems* 33.5, pp. 311–324. ISSN: 01989715. DOI: 10.1016/j.compenvurbsys.2009.06.002.
Theobald, David M. (2001). "Topology revisited: representing spatial relations". In: *International Journal of Geographical Information Science* 15.8, pp. 689–705. ISSN: 1365-8816. DOI: 10.1080/13658810110074519.

Vander Wal, Thomas (2007). *Folksonomy Coinage and Definition*. URL: http://www.vanderwal.net/folksonomy.html.

Wolf, Eric B. (2009). "Ontology-driven generalization of cartographic representations by aggregation and dimensional collapse". In: *Lecture Notes in Computer Science*. Springer Science, pp. 990–997. ISBN: 364204929X. DOI: 10.1007/978-3-642-04930-9{_}65.

Zhang, Xiang (2012). "Automated evaluation of generalized topographic maps". PhD thesis. University of Twente. ISBN: 9789061643388.

Zhang, Xiang and Ai Tinghua (2015). *Advances in Spatial Data Handling and Analysis*. Ed. by Francis Harvey and Yee Leung. Switzerland: Springer International Publishing, pp. 143–162. ISBN: 9783319199498. DOI: 10.1007/978-3-319-19950-4.