FAKULTÄT FÜR !NFORMATIK
Faculty of Informatics

Open Government Data in Vienna and London

MAGISTERARBEIT

zur Erlangung des akademischen Grades

Magistra der Sozial- und Wirtschaftswissenschaften

im Rahmen des Studiums

Masterstudium Informatikmanagement

eingereicht von

Dipl.-Ing.(Univ.) Beate Hinterberger, MSc (FH)
Matrikelnummer 9440409

an der
Fakultät für Informatik der Technischen Universität Wien

Betreuung
Betreuer/in: Privatdoz. Dipl.-Ing. Dr.techn. Gerhard Navratil

Wien, 01.12.2014

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

Open Government Data in Vienna and London

MAGISTERARBEIT

Submitted in partial fulfilment of the requirements for the degree of

Magistra der Sozial- und Wirtschaftswissenschaften

in the course of studies

Masterstudium Informatikmanagement

submitted by

Dipl.-Ing.(Univ.) Beate Hinterberger, MSc (FH)
Matrikelnummer 9440409

at the
Faculty of Informatics of the Technical Unitverstity of Wien

Supervision
Supervisor: Privatdoz. Dipl.-Ing. Dr.techn. Gerhard Navratil

Vienna, 01.12.2014 ________________________ ________________________
(Signature Author) (Signature Supervisor)

Technische Universität Wien
A-1040 Wien ▪ Karlsplatz 13 ▪ Tel. +43-1-58801-0 ▪ www.tuwien.ac.at

Acknowledgements

I want to thank my family, parents and friends for their support and for freeing up the time in a working mum's life that was necessary to write this thesis.
And I want to thank my ever-patient supervisor Gerhard Navratil.

I want to dedicate this thesis to my daughter Anna.

Declaration

Dipl.-Ing. (Univ.) Beate Hinterberger, MSc (FH)
1120 Wien

„Hiermit erkläre ich, dass ich diese Arbeit selbständig verfasst habe, dass ich die verwendeten Quellen und Hilfsmittel vollständig angegeben habe und dass ich die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken oder dem Internet im Wortlaut oder dem Sinn nach entnommen sind, auf jeden Fall unter Angabe der Quelle als Entlehnung kenntlich gemacht habe.“

Ort, Datum, Unterschrift

Abstract

The idea of Open Government Data (OGD) is, to make government data available for the public, free of charge and with minimum constraints for reuse. This shall facilitate new business opportunities, enhance government transparency, encourage citizen engagement and distribute the costs of government data processing to communities.

This thesis evaluates, if the Open Government Data offers of the municipalities of Vienna and London can be considered as a replacement for commercially available data. The value of the freely available data will be evaluated by applying two business scenarios of decision support based on spatial data. These are a store location planning and the positioning of mobile communication infrastructure. The thesis investigates, if the OGD portfolios offered by Vienna and London are comparable concerning this application although they follow different approaches for realisation.

The thesis comes to the conclusion that Open Government Data like it is published in the data catalogues of Vienna and London can be used to get an overview on the situation in the two cities. Some of the datasets are suitable for detailed analysis as well, like for example the land use data in Vienna or the data on income in London that can replace the commercial dataset on spending capacity. For a detailed analysis, most of the published datasets are either not detailed enough in terms of spatial resolution, do not contain required attributes or are simply outdated.

Kurzfassung

Die Idee von Open Government Data (OGD) ist, Daten, die von Regierungen erfasst werden, für die Öffentlichkeit zugänglich zu machen. Die Nutzung soll gebührenfrei und mit möglichst wenigen Einschränkungen möglich sein. Das soll neue Geschäftsideen unterstützen, die Transparenz von Regierungen erhöhen und Bewohner dazu ermutigen, sich lokal zu engagieren. Außerdem werden die Kosten, die für die Erstellung der Daten notwendig waren, wieder an die Gemeinschaft verteilt.

Diese Arbeit evaluiert, ob die von den beiden Hauptstädten Wien und London angebotenen offenen Daten als Ersatz für kommerziell angebotene Daten in Betracht gezogen werden können. Der Wert der frei verfügbaren Daten wird dabei evaluiert, indem zwei Geschäftsszenarien herangezogen werden, die eine Entscheidungsfindung auf der Basis von räumlichen Daten erfordern. Diese sind, die Standortplanung für ein Geschäft und die Positionierung von Mobilfunksendeanlagen. Die Arbeit untersucht, ob die OGD-Angebote von Wien und London im Bezug auf diese Anwendung vergleichbar sind, obwohl sie unterschiedliche Strategien für die Realisierung verfolgen.

Diese Arbeit kommt zu dem Schluss, dass offene Daten, so wie sie in den Datenkatalogen von Wien und London publiziert werden, verwendet werden können, um einen Überblick über die Situation in den beiden Städten zu erlangen. Einige der Datensätze sind auch geeignet für detaillierte Analysen. So zum Beispiel die Landnutzungsdaten in Wien oder die Daten über das Einkommen, das in London publiziert und als Ersatz für kommerzielle Datensätze über Kaufkraft verwendet werden kann. Für eine detaillierte Analyse sind allerdings die meisten der publizierten Daten entweder nicht detailliert genug was die räumliche Auflösung angeht, enthalten nicht die erforderlichen Attribute oder sie sind schlichtweg veraltet.

Contents

Contents 7
1 Introduction 10
1.1 Motivation 10
1.2 Aim of the Thesis 10
1.3 Methodology applied 11
2 Background on Information Economy, OGD and Data Quality 13
2.1 Information Economy 13
2.2 History of OGD and LOD 15
2.3 Data Quality 17
3 The portfolios of Vienna and London 25
3.1 Vienna 25
3.2 London 27
3.3 Comparison of the portfolios 28
4 Examples for business scenarios and citizen use cases 29
4.1 Store Location Planning 32
4.1.1 References from Literature 33
4.1.2 Data Requirements for Store Location Planning 38
4.2 Plan a mobile communication network 43
4.2.1 References from Literature 43
4.2.2 Data Requirements for Planning a Mobile Communication Network 51
5 Evaluate differences between open portfolios and commercial offers 53
5.1 Boundary Datasets 55
5.1.1 Boundary Datasets for Vienna 55
5.1.2 Boundary Datasets for London 56
5.1.3 Boundary Data Evaluation 57
5.1.4 Usability of boundary data sets 60

5.2 Population Data 60
5.2.1 Population Data Vienna 61
5.2.2 Population Data London 63
5.2.3 Population Data Evaluation 64
5.2.4 Usability of Population Data 67
5.3 Digital Terrain Model 67
5.3.1 Digital Terrain Model Vienna 68
5.3.2 Digital Terrain Model London 69
5.3.3 Digital Terrain Model Evaluation 70
5.3.4 Usability of Digital Terrain Model 71
5.4 Land Use 72
5.4.1 Land Use Vienna 72
5.4.2 Land Use London 73
5.4.3 Land Use Evaluation 74
5.4.4 Usability of Land Use Data 76
5.5 Buildings 76
5.5.1 Buildings in Vienna 77
5.5.2 Buildings in London 79
5.5.3 Buildings Evaluation 80
5.5.4 Usability of Building Data 83
5.6 Road Network 83
5.6.1 Road Networks of Vienna and London 84
5.6.2 Alternative Road Networks 84
5.7 Employers and Employees 85
5.7.1 Employers and Employees in Vienna 85
5.7.2 Employers and Employees in London 86
5.7.3 Employers and Employees Evaluation 87
5.7.4 Usability of data on Employers and Employees 89
5.8 Commuter Statistics 90
5.9 Hot Spots 90

5.9.1 Hot Spots in Vienna 91
5.9.2 Hot Spots in London 91
5.9.3 Hot Spots Evaluation 92
5.9.4 Usability of Data on Hot Spots 94
5.10 Footfall, Drop In, Shop Mix 94
5.11 Rental rates 95
5.11.1 Rental Rates Vienna 95
5.11.2 Rental Rates London 95
5.11.3 Rental Rates Evaluation 96
5.11.4 Usability of data on Rental Rates 98
5.12 Spending Capacity 98
5.12.1 Spending Capacity Vienna 98
5.12.2 Spending Capacity London 99
5.12.3 Spending Capacity Evaluation 100
5.12.4 Usability of Data on Spending Capacity. 101
5.13 Competitors' locations 102
5.14 Customers 102
6 Summary and Conclusion 104
List of Abbreviations 108
List of Figures 109
Bibliography 111

1 Introduction

Open Government Data (OGD) aims to provide valuable data for unlimited use by citizens as well as companies.[1] In this thesis it shall be discussed, which questions can be answered using the available data that are offered by two major European cities, based on some scenarios that companies currently could face. This thesis will outline the benefits of OGD as opposed to times when no open data was available. Differences will be compared, that can be found in benefits arising from the different approaches for realising OGD by the municipalities of Vienna and London. As a result, this thesis will discuss, what is the actual value of OGD for companies and citizens in terms of data that they do not need to purchase any longer because the catalogue of open data provides a reasonable alternative.

1.1 Motivation

In recent years the data catalogue offered by the city of Vienna under the Creative Commons License has been expanded continuously. Every now and then I tried to find data that would be relevant for answering questions with a geographical context in Vienna, without being thoroughly successful. However, meanwhile the scope of data offered according to the Open Government Strategy is quite big. Therefore a systematic investigation on the potential that lies within the data that is provided seems to be valuable for the consideration of the data within future analysis. In order to get a more representative picture, the offered portfolio will be compared with that of London. London was chosen, because it is another big municipality that is on the same level according to the model of central places by Christaller and it is therefore comparable to Vienna. (Christaller, 1933)London follows a completely different strategy for realising Open Government Data and the United Kingdom (UK) was among the first countries to launch an open government data catalogue.

1.2 Aim of the Thesis

Open Government Data aims to improve open collaboration between governments and citizens as well as economy (Krabina, et al., 2012). Following this idea a whole catalogue

---

[1]compare http://opendefinition.org/od/ last access 2014-10

of data is offered by governments on their web portal. This data can be used under the conditions of a licensing scheme that contains almost no restrictions. The question this paper investigates is, whether the data published as Open Government Data is of actual value for companies. In other words, it will evaluate if the data can be used to replace commercially traded data products for everyday applications in economic reality. This would reduce the costs for a business and therefore be a benefit for the company.

Realising OGD the governments can follow different strategies. Looking at the data catalogues of Vienna and London, one can see that they both claim to apply to the eight principles of OGD as they are published by the OGD group[2], but seem to show differences in the volume presented as well as in how the data is prepared before publication. In order to gain a more representative picture on the value of the freely available data, the data catalogues of these two municipalities will be investigated in comparison.

This thesis will investigate which questions can be answered, using the data that is provided by the two municipalities of Vienna and London. From a company's point of view, this will be investigated based on some business scenarios that companies currently could face. This thesis shall show if the analysis for decision support in the context of store location planning and the positioning of mobile communication infrastructure can be done using open data catalogues. It will investigate, if the portfolios offered by Vienna and London are comparable concerning this application although they follow different approaches for realisation. And it will be evaluated, if their ODG offers can be considered as a replacement for commercially available data.

1.3 Methodology applied

The value of the data will be measured by estimating what kind of questions can be answered using this data in the context of the chosen business scenarios. Therefore the quality and relevance of the data will be crucial. The portfolios of the cities of Vienna and London will be compared regarding the quality and fitness for use of the offered datasets. Parameters that are relevant for the sample business scenarios will be discussed, such as:

- Scope. Which topics or areas of interest are covered by the portfolios?
- Accuracy. How accurate is the data in comparison to data from other sources?

[2] https://public.resource.org/8_principles.html

- Resolution. What is the level of detail that is offered for free?
- Completeness. Does the data comply to the metadata?
- Consistency. Has integrity, consistency completeness etc. been verified?
- Lineage. Are the sources reliable and what is their reputation? How detailed and complete is the metadata (source, creation date, accuracy, checks, updates, responsibilities, etc.)?
- Interfaces. How can the data be integrated into applications (data formats and data structure)? Is it possible to link to a source and get automatic updates? Are they compatible with earlier versions?
- License. Are there any constraints, for example for commercial use?

2 Background on Information Economy, OGD and Data Quality

This chapter will describe the context for the evaluations of Open Government Data in this thesis. The current developments in information economy as well as a short historic overview of OGD will introduce the aims and possible benefits of OGD. The concept of data quality will be used for comparing the offers of Vienna and London. Because there are a number of different approaches and scopes for the definition of data quality this chapter will introduce the aspects that are considered in the analysing chapters later on.

2.1 Information Economy

Some of the rules that Information Economy follows are new to experts on traditional economy. Among these is a certain structure of costs. In the information economy the costs of production are typically sunk costs and in most cases they are high. On the contrary, the costs for reproduction and copying are low or even marginal costs. For example, the production of a software program involves great effort in terms of knowledge and time of software developers that need to be paid for their work. Once the program is finished, it can easily be copied and distributed within seconds over the internet. The same problem arises with the production of data and pieces of information. Collecting data, verifying it, bringing it into a structure and ensuring completeness for a certain topic or area requires time, knowledge and often also special equipment. That makes the production process expensive. Once the dataset is complete, copying and distributing it, is again easy. (Shapiro and Varian 1998)

Kreket al describe in their article a concept for the generation of value with geographic data as an economic good. In addition to the high sunk costs and low marginal costs, the data is a non-rival economic good, which means that it does not wear off, when it is used. Controlling its consumption is very difficult and the fact that it is an experience good requires the consumer to try and test the data in order to assess its value. Krek et al state, that the value of geographic data derives from its use in a decision. Therefore the required data is always highly specialised according to the current decision situation. And usually only a small set of data is required for the individual situation. Producers therefore pack data into products for certain target groups. These products can then be used for certain types of decision situation. The concept of a geo-information value chain therefore begins with the identification of certain user needs or categories of user needs. The possible applications of a data product define its monetary value. In order to create the according data products, a chain of processing steps and transformations usually needs to be performed on the raw material. The paradox of the value chain for geographic data products results from the fact that the highest cost in the chain lie at the beginning, with the production of the raw material where the value of the product is still relatively low. Along the chain, the costs for processing and transforming the data are relatively low, while the values of the data products rise substantially with every step. (Krek, et al., 2000)

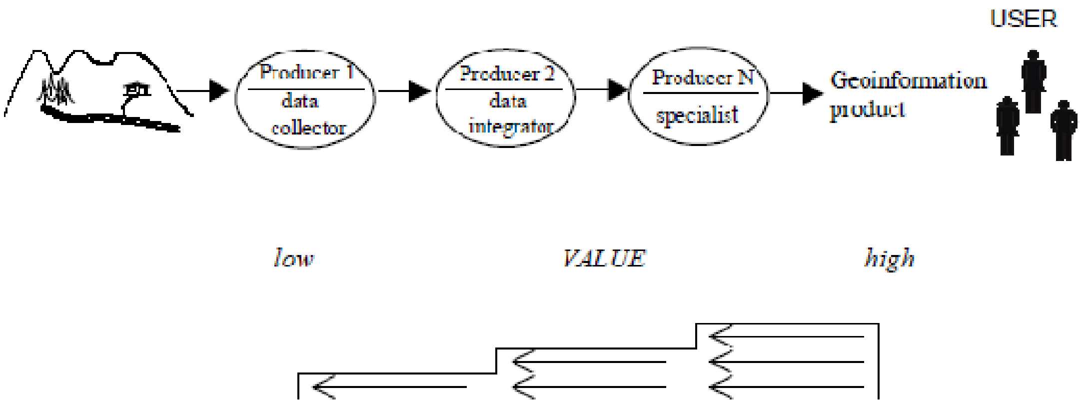


Figure 1: The Value Chain Paradox (Krek, et al., 2000 p. 5)

Shapiro and Varian (1998) outline some principles that can be followed in order to generate additional revenue streams. One of these concepts is what they call a lock-in of customers and suppliers. According to this principle customers as well as suppliers face costs if they want to switch from one information product to another. Therefore they tend to stay with a product, once they are used to it and will likely buy updates and probably also upgrades or extensions. This is valid for software as well as for data products. If employees are used to a certain user interface and a certain bundle of functionality coming with software, they will need training or at least time for adoption to another software product. This is true, even if the functionality is equal and the user interface similar to what they are used to. Also, if switching from one data product to another, one must take into account certain efforts. Most likely interfaces will need to be adopted. And even if the technical implementation can easily be changed in order to integrate an equivalent data product, the content and quality can vary. Even if the extent of this variation is not critical for a business, adoptions will be required for the interpretation of the data in order to generate the same amount of benefit and value from the new data product. These switching costs can be used to bind customers once they started using the products of a certain supplier. As long as the costs for upgrades or maintenance do not exceed those switching costs, customers will likely stay with their supplier. Also monetarily, upgrades and updates are usually much cheaper than a newly purchased license. This gives the customer a good feeling about staying with the supplier he has.

From a company's perspective, there is most likely a need for information, for example to identify and address customers. This thesis will go into further detail examining some business scenarios later in this thesis. If the relevant information cannot be gathered and maintained within the company, one has to participate in a market for data that follows the rules described above. Even if crucial data will become freely available in future, the restrictions coming with lock-in will still be there. The concept of Linked Open Data (LOD) intends to make data coming from different sources comparable, by defining a structure of semantics and linking data sets to certain elements in there. One of the goals of Linked Open Data is it, to enable querying across different sources. This would also reduce the risks of lock-in for customers. This trend will be looked at in more detail in the chapter on OGD and LOD.

One method for increasing possible sales and revenues is versioning a product for different target groups and according to their individual needs. This allows differentiated pricing. An extreme version of versioning is to give away free samples. Also limited versions for very low costs can be distributed. But upgrades of the free samples and limited versions, as well as full versions in combination with a restrictive rights-management can return revenues for the producer. Selling on low price per copy or living on the revenues generated by placing advertisements within free copies can be another strategy to produce revenues. This requires a broad distribution of the product. If one wants to benefit from positive feedback within an information network, a critical mass of customers must be reached quickly. This will require expectation management and sometimes penetration pricing, which means pricing below production costs.

2.2 History of OGD and LOD

The idea of open government data is, to make government data available for the public, free of charge and with minimum constraints for reuse. This shall facilitate new business opportunities, enhance government transparency, encourage citizen engagement and distribute the costs of government data processing to communities.

"*Knowledge is open if anyone is free to access, use, modify, and share it — subject, at most, to measures that preserve provenance and openness*."[3]

The first Obama administration's transparency initiative is said to be the motivation for Open Government Data. But also before that time, the United States of America (USA) have a long tradition on open data. The United States copyright law states the following.

"A "work of the United States Government" is a work prepared by an officer or employee of the United States Government as part of that person's official duties."[4]

And according to section 105 of the Copyright Act, such work is not subject to copyright protection.[5] The idea behind this concept is that the work of the US government employee was already paid from tax money and the product should therefore be available for free.

The data portals of the United States of America (USA) and the United Kingdom of Great Britain (UK) launched their respective data portals in May 2009 and January 2010.[6] (Ding, et al., 2012)The European Union also encourages OGD through a directive on the re-use of Public Sector Information (PSI). The Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information can be found on the homepage of the European Union.[7] The European open data portal was launched in 2012. It offers a data catalogue and links to the national open data portals within the community.[8]More than 30 countries have put online more than 70.000 OGD datasets by 2012. (Ding, et al., 2012)

Erickson and his colleagues report about a project that analysed and documented open data catalogues published by governments around the world. (Erickson, et al., 2013) In 2013 metadata for more than 1.022 million datasets could be found. These were published in over 190 catalogues in 24 languages. More than 450.000 datasets were found in the US catalogue by that time. The UK data catalogue with about twelve thousand datasets ranked on fourth position after France and Canada.

---

[3]http://opendefinition.org/od/ last access 2014-10

[4]http://www.law.cornell.edu/uscode/text/17/101 last access 2014-11

[5]http://www.law.cornell.edu/uscode/text/17/105 last access 2014-11

[6]www.data.gov and data.gov.uk last access 2014-10

[7] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2003.345.01.0090.01.ENG last access 2014-10

[8]http://open-data.europa.eu/en/about last access 2014-10

Another step in the evolution of OGD is to make data from different sources, in different formats and using different vocabulary combinable as Linked Open Data (LOD). (Weiss, 2013) LOD uses Hyper Text Transfer Protocol (HTTP) for data access and Uniform Resource Identifiers (URIs) and the Resource Description Framework (RDF) in order to create a minimum consensus on data representations. Nowadays, the value of LOD deployments is demonstrated by visual mashups on the Web. In theory, in future one can mashup data from different sources, such as crowdsourced information or privately owned data. (Ding, et al., 2012)

2.3 Data Quality

Data is used in decision processes. The quality of the data must be good enough to make a good decision but it should not provide too much detail, in order to save processing time and storage space. There is seldom a definition to be found that describes the measures or attributes that could be used to evaluate the quality of a spatial data set. Some of the articles cited in the chapters on business scenarios refer to high quality spatial data as data where relevant information can be derived from (Byrom, 2005). This seems to be a uniquely applicable requirement for such data sets that shall be used to support decisions. The standard DIN EN ISO 9000 Quality management systems – Fundamentals and vocabulary defines quality as *"degree to which a set of inherent characteristics fulfils requirements"*. (DIN, 2005 S. 18)Data does therefore not need to be perfect in all aspects, in order to be of high quality, but the quality can be quantified in relation to the requirements at hand. This chapter describes some relevant aspects of data quality. Frank and his colleagues described a procedure to select the best dataset for a task. According to their theory the best dataset is the one that suits best for the task at hand. (Frank, et al., 2004)

The value of a dataset can be determined by estimating the improvement of the decision that is made based on the data. In order to find out, which one among a selection of datasets is the best one, different aspects and elements of the products are compared. One main idea of Frank's theory is, that the quality of data dilutes over time. This means, that the accurateness and completeness of a dataset are described in the metadata the way they were when the data was collected, respectively when the metadata was defined. But reality changes over time while the data stays the same. This results in increasing differences between reality and the dataset over time. The presence and quality of certain elements in a data product can influence the value of an analysis result, or they can be crucial to comply to a certain standard, in order to be usable at all. This should also be kept in mind, when evaluating spatial data sets for a certain task. (Frank, et al., 2004)

The quality of data depends on technological possibilities, legal restrictions, and the requirements of the user. The fitness for use of data describes, to what degree the quality of a dataset matches the needs of a task. This is an important factor for the value of a dataset. Technical restrictions usually limit the possible data quality at the time of gathering information. While legal restrictions can also limit the possible data quality by limiting the means for gathering information, laws usually define restrictions for accessing and using data. This also limits the quality of a dataset in the context of a certain task at hand. (Navratil, et al., 2005)

Organisations of cartographers like the International Cartographic Association (ICA) as well as governmental organisations and producers and distributors of spatial data have been trying to define standards, descriptions and measurements for spatial data quality. In the 1990s an ICA Commission of Spatial Data Quality was created that published a recommendation on how to describe a measure for spatial data quality. This description was published in the form of a book containing contributions from experts of different nationality and expertise within the field of spatial information (Morrison, 1995). This book identifies the following Elements of Spatial Data Quality:

- Lineage
- Positional accuracy
- Attribute accuracy
- Completeness
- Logical Consistency
- Semantic accuracy
- Temporal information

Lineage means the history of a dataset. It includes information on the creation of the data as well as every single action that was undertaken to alter or interpret the data.
"Lineage then is a part of the data quality statement that contains information that describes the source observations or materials, data acquisition and compilation methods, conversions, transformations, analysis and derivations that the data has been subjected to, and the assumptions and criteria applied at any stage of its life." (Clarke, et al., 1995 p. 13)

The requirements for positional accuracy for data within a GIS depend on the purpose, the data or the application using the data shall fulfil. Therefore it is important to measure and document the positional accuracy of the entities in a geographic dataset. This ensures that the data can be used in another application or environment than originally defined.
“In the mapping sciences the position of a real world entity is described by values in an appropriate coordinate system. Positional accuracy represents the nearness of those values to the entity’s “true” position in that system.” (Drummond, 1995 p. 32)
Drummond describes in her article the different kinds of errors that occur with spatial information and how they can be measured and documented.

Attributes can take a vast variety of forms. This is due to the fact that an attribute can be defined as a fact about some location, set of locations or feature on the surface of the earth.
“The fact can be the result of measurement with some kind of instrument, such as a thermometer or earth observing satellite; it might be the result of interpretation by a trained observer, such as a land use or soil class; or it might be the outcome of a historical or political consensus such as the names given to lakes or roads.” (Goodchild, 1995 p. 59).

The implementation and use of error handling in GIS tools is not yet spread widely. Many users prefer not to know about the uncertainty in the data they use. (Goodchild, 1995) Potential consequences of ignoring uncertainty are still difficult to estimate. One way is to make small variations in the used data that represent the magnitude of the uncertainty. But to the author’s knowledge, this is not common in everyday business live.

Brassel and his colleagues define completeness of a dataset in relation to the overall information it is supposed to represent. They call this the abstract universe of a dataset:
“Completeness describes whether the entity objects within a data set represent all entity instances of the abstract universe. The degree of completeness describes to what extent the entity objects within a data set represent all entity instances of the abstract universe.” (Brassel, et al., 1995 p. 86).
Depending on the purpose, for what a specific dataset shall be used, the according abstract universe can differ from case to case. Completeness in this context is described as a fitness of use for the special purpose. In contrast, the completeness in the sense of overall data quality refers to an abstract universe that is defined trough the data capturing rules applied when generating the dataset. This kind of completeness shall usually be described in the metadata of a dataset. (Brassel, et al., 1995) For this thesis the general

term of completeness in the sense of data quality is of course relevant and important. However, the actual value of a dataset in the context of decision support is determined by the fitness of use for the individual question that shall be answered. Therefore fitness for use of the dataset found in the different data catalogues will be investigated according to the business cases described.

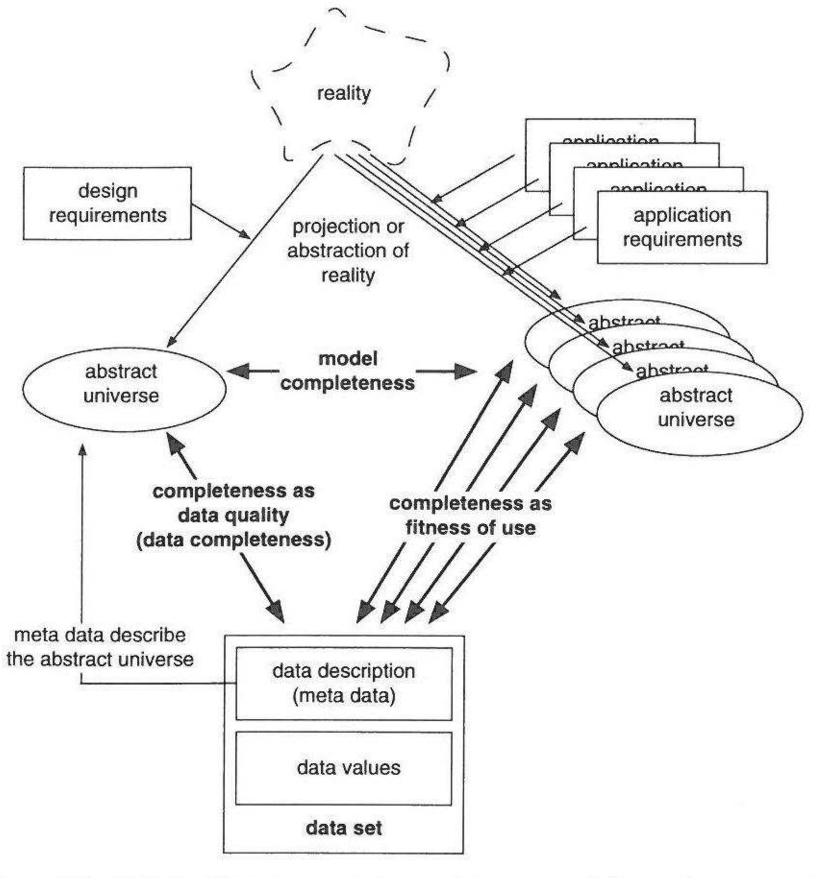


Figure 2: Relationship between data completeness, model completeness and fitness of use (Brassel, et al., 1995 p. 88)

Logical consistency, as described by Kainz (1995) deals with the structure of geographic data and the compliance of attribute information to certain defined rules. It ensures, that data is compatible with another datum in the dataset and that datasets can be combined according to certain rules.

"Logical consistency is the element of spatial data quality that deals with the structural integrity of a given data set. This integrity is based on a formal framework for the modelling of spatial data and the relationships among objects. Mathematical theories of metric, topology, and ordered sets play a major role in the foundation of this framework. From that framework rules and constraints can be derived that are useful for establishing a consistent data set." (Kainz, 1995 p. 134).

The concept of metric is in most spatial datasets inherently present by assuming that the data objects are defined within Euclidean space with Cartesian coordinates. The data that is investigated in this thesis can be assumed to apply to these rules.

Topology investigates characteristics of geometry that do not change under certain transformations. Examples that show these characteristics would be drawings on a balloon. Closed lines remain closed when inflating the balloon, although the exact shape may change. The check for valid topology involves a number of logical and mathematical operations that are described in detail in the article by Kainz. Order is one of the basic structures upon which mathematical concepts are built. A set is said to be ordered when there is an order relation between its elements.

Consistency rues that are well known in the context of relational databases also apply to the attributes of spatial information. Data in a database is required to be accurate, correct, valid and consistent. Consistency is usually ensured by checking the data against a set of consistency rules. Among these can be domains that define a valid range for a datum or referential constraints that ensure that no data is inserted into a table as foreign key without a correspondent entry in the related tables. If all values in a table field must be known, then a check for NULL values will identify invalid entries. Triggers and precondition constraints can ensure certain changes or validations on a dataset that are only performed, when a change is attempted to be made during a transaction.

Semantic accuracy is discussed in an article by Salgé. (1995) This article comes to the following conclusion:
"*Semantic accuracy is composed of several parameters which are not easy to differentiate. An apparent mistake (a house inside a lake) may be due to up-to-dateness (a modification of the shoreline), or to logical consistency (the rule did not take account of possible houses on supporting piles), or to completeness (omission of an island or commission of either the house or the lake). Even more complex, an apparent semantical error may only be the result of geometric mistakes (the house is on the shoreline, not inside the lake).*" (Salgé, 1995 p. 149f)
According to Salgé, semantic accuracy shall describe the semantic distance between the perceived reality and geographical objects. The severity of errors can vary depending on the type of error and the context where the data is used. Salgédescribes the examples of misinterpretations of trees for bushes versus water for buildings. The rates of missing data and over-completeness can be used as one measure for semantic quality. Alternatively the semantic constraints that can be given explicitly or implicitly by the data can be used

for this kind of measurement. An example would be a river that can only flow from its origin to the sea. Attribute accuracy measures the probabilities that attribute values were assigned correctly. Measuring semantic accuracy is only possible by investigating representative samples.

Gutpill describes how the requirements for temporal attribution of spatial information will vary depending on the geographic features and applications (Gutpill, 1995). He comes to the conclusion that temporal attribution will become more important with more sophisticated GIS applications. The article describes three such types for time measurements. First there is the logical time or event time, when a change in reality actually occurred. Second there is the observation time or evidence time when the change has been measured or otherwise observed. And third there is the recording time or database time, when the change has been stored into the database. To ensure high quality of temporal attribution one needs to keep in mind that these different but at the same time valid time attributes do exist for most spatial features and sometimes the individual time measurements cannot be made or are not documented in the data any longer. To make things more complicated, the changes within a dataset should be tracked in order to be able to look at a dataset at a certain point in time. Temporal consistency over a geographic region is often required but almost never perfectly achieved. Surveys for collecting spatial information often are carried out over a long period of time. Usually updates are made more often for small regions where many changes where expected while other regions are not investigated for longer time intervals. The measures for the quality of temporal attribution were yet to be found according to the article by Gutpill the author of this thesis did not find any documented progress in the search for a solution. Temporal attribution is, however, interacting with all the other aspects of data quality discussed in the book on elements of spatial data quality. (Morrison, 1995)

In the final chapter of the book on spatial data quality Veregin and his colleague combine the findings of the earlier articles into an evaluation matrix (Veregin, et al., 1995). The dimensions of this matrix are based on the following two main concepts:

"Geographical observations are defined in terms of space, time and theme. Each of these dimensions can be treated separately (but not always independently) in data quality assessment.

Data quality is an amalgam of overlapping components, including accuracy, resolution, completeness and consistency. The quality of geographical databases cannot be adequately described with a single component." (Veregin, et al., 1995 p. 167)

	Space	Time	Theme
Accuracy			
Resolution			
Completeness			
Consistency			

Figure 3:Evaluation matrix (Veregin, et al., 1995 p. 168)

In the United States a number of standards for data quality is already implemented and in use. The Spatial Data Transfer Standard (SDTS) is based on the idea that users should be able to evaluate the fitness-for-use of a dataset for a certain application from the quality documentation. This requires the metadata to contain detailed information on different aspects of data quality. The standard distinguishes between five components that are similar to the ones discussed in the book:

- Positional accuracy
- Attribute accuracy
- Logical consistency
- Completeness
- Lineage

Accuracy can be tested against a reference source of known higher quality. In the spatial domain the positional component of features is the measure of interest. While for points the location of the x, y and z coordinate can be measured and investigated separately or in combination, this task becomes much more complicated for line or polygon features. For these more complex entities, there exist other models like the epsilon band model which defines error in terms of a zone of uncertainty surrounding a line location. Accuracy in the time domain has not been investigated in very much detail in literature. Measuring a point in time depends on the need to measure time in reference to some agreed upon standard origin. Timestamps in databases and recorded observation time often do not fulfil this requirement. Additionally the distinction between event time, observation time and database time is often difficult or impossible. Temporal accuracy is also affected by temporal resolution. Accuracy in the thematic domain depends on the type of measurement scale of the considered attribute. A classification error matrix can for example be used to evaluate error in a nominal scale. And also for ordinal scale attributes a modified version of this error matrix can be used. Assessing errors in an interval and ratio scales can typically follow the same rules as the assessment of positional accuracy.

Resolution describes the amount of detail that can be observed in space, time or theme. In spatial and temporal dimensions the resolution is limited by the instruments used for observation. But very often data is later generalised to a coarser resolution which needs to be documented in the metadata on the dataset. Resolution in the thematic domain can also be dependent on the capabilities of a measuring instrument. Generally the definition of resolution depends again on the type of measurement scale. Resolution can be defined in terms of fineness and number of categories for an attribute. Soil classes can for example be defined in relatively fine level of taxonomic resolution or in generalized classes of some level.

Completeness describes the relation between the features encoded in the database and the abstract universe of all such features. Assessing for quality of completeness can therefore primarily be done by identifying errors of omission. This requires a precise description of the content of the database. The content can then be tested to what degree it matches the description.

Consistency measures the internal validity of a database. In the spatial domain a violation of consistency refers to a topological error. The same approach can also be used for the temporal domain, by defining rules for temporal topology. In the thematic dimension consistency includes the relations between different attribute values. For example the total population of an area, the number of households and the average population per household of the same area must have a valid relation. There can be numerous relationships of similar kind in a dataset that must be examined. Therefore general tests do not exist.

3 The portfolios of Vienna and London

This chapter will compare the publication strategies of the municipalities of Vienna and London. The respective portfolios are a direct result of these strategies and represent their aims.

3.1 Vienna

The city of Vienna started already many years ago to publish topographic maps of the city. ViennaGIS is a trademark that is used to represent the combined geodata of Vienna. It was introduced in the 1980s when a widespread Geographic Information System (GIS) was introduced. But the foundation for analysing geographic data was already layed in the 1970s, when the need for spatial analysis arose. First, spatial data was collected manually, but soon the Viennese government began to gather and store this information digitally. Based on aerial images, a new city map was created. In order to correct the positions of streets and buildings, all the streets within the city where surveyed anew. The result was the so called “Mehrzweckstadtkarte” (MZK) that covered the whole city and provided high spatial resolution. In 1995 the government of Vienna started publishing data online to be used free of charge. The online city map is still the most visited site on the homepage of the city of Vienna.[9]

Documents on concepts for Open Government Data in Vienna date from 2011 and 2012. The open data catalogue was launched in May 2011 and currently contains data on 246 topics. Currently the City of Vienna offers more than 600 e-Government web-pages and an interactive city map with more than 120 layers.[10]

According to the Open Government Implementation Model the city of Vienna aims to achieve the “ubiquitous engagement of stakeholders” with the help of Open Data. (Krabina, et al., 2012 p. 7) This is achieved through transparency, participation and collaboration. This “Open Government Implementation Model” illustrates the generic procedure for the implementation of Open Government initiatives in Vienna.

[9] http://www.wien.gv.at/viennagis/geschichte.html last access 2014-10

[10] https://open.wien.at/site/open-government-data-in-vienna-2/ last access 2014-10

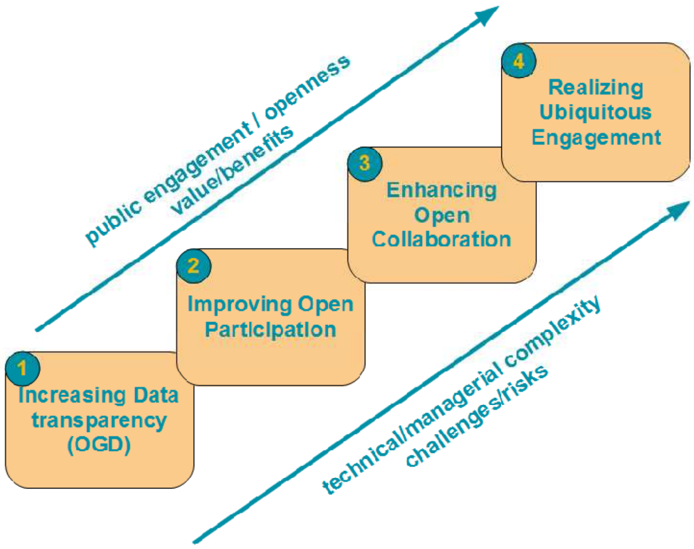


Figure 4: Open Government Implementation Model (Krabina, et al., 2012 p. 8)

The first stage of increasing data transparency focuses on two challenges. One is identifying data with potentially high value or high impact for the addressed stakeholders and their participation in planning processes. And the second is to assure and improve data quality in terms of accuracy, consistency and timeliness. The city government plans to tackle these challenges with the help of internal data monitoring as a central measure. The Open Government Implementation Model describes further on plan how to identify data that shall be published by screening internal catalogues and requirements from users. Then a detailed plan is laid out, how the initial improvement and permanent maintenance of data quality will be guaranteed. Then data will be released an updated in planned releases in coordination with or as a consequence of internal updates. When monitorings of databases show up further potentially valuable datasets for publication this process will be triggered.

The city of Vienna intends to introduce Data Management and Data Governance as new disciplines in public management in order to implement approval cycles with Open Government. The Open Government Implementation Model of the city of Vienna states that *"A focused look at public sector management has been missing so far in Public Management. A control gap has become evident due to the trend toward the release of data in Open Government Data Portals. The Implementation Model is a contribution toward closing this gap."* (Krabina, et al., 2012 p. 6).

3.2 London

The London Data Store (LDS) was launched in 2010. In 2014 a new release called LDS II was launched and the corresponding blog entry reports that at that time it had about 30.000 visitors a month and over 10.000 data sets where downloaded each month.
http://data.london.gov.uk/blog/lds-ii-the-next-step-to-a-city-data-future-2/

The London Data store intends to support collaborations based on a transparency as a governmental principal for the Greater London Authority (GLA).
"The London Datastore is the embodiment of the GLA's ambition to unlock the power of the information held across the GLA family [...] We want communities, public service professionals, anyone, to flex their muscle, to use the data we make available to raise awareness of issues, suggest ways of tackling them, improving lives as they go."[11]

To realise these intentions, the Greater London Authority engages with partners and stakeholders in order to facilitate the use of Open Data. The decision which datasets to publish and how to maintain them lies within the single GLA departments that own the data. The principles and strategy for the London Data Store is published in the Open Data Charter of the GLA.[12]To fill the London Data Store with information, the functional bodies Transport for London (TfL), Met Police, London Development Agency, Olympic Development Authority and London Fire and Emergency Planning Authority (LFEPA) were invited to contribute to the project and they committed to freeing up some of their data.[13]

The blog entry concerning the launch of LDS II states:
"This is why we want LDS II to be a place where ideas for the use of data to solve city issues and capitalise on opportunities to improve services, are born. This is a place in which we want collaborations to start."[14]

---

[11] http://data.london.gov.uk/blog/data-data-everywhere-now-lets-make-difference-it/ last access 2014-10

[12] http://data.london.gov.uk/wp-content/uploads/2014/10/OPEN-DATA-CHARTER.pdf last access 2014-10

[13] http://data.london.gov.uk/blog/hectic-times/ last access 2014-10

[14] http://data.london.gov.uk/blog/lds-ii-the-next-step-to-a-city-data-future-2/ last access 2014-10

3.3 Comparison of the portfolios

The city of Vienna states to carefully select the datasets that shall be published in the Viennese open data catalogue. The data is then updated regularly and the processes for creation and maintenance are defined.

The government of the Greater London Authority invites the different functional bodies to contribute to the London data store. These provide a great number of datasets that are made available online. Consistent processes for publishing and updating the data are not created, but the number of available datasets is comparably high.

Vienna focuses on evaluating and selecting datasets that might have an impact on decision processes or be of value for the addresses stakeholders. The goal of the London government is to improve transparency by publishing as much as possible.

4 Examples for business scenarios and citizen use cases

This chapter will describe some examples for business scenarios and citizen use cases where information in the form of spatially related data is required for decision making. Investigations will be made, which data is suitable to improve the answers to questions that are relevant to businesses and citizens. The consecutive chapters will first outline what data is available besides Open Government Data. This will be based on the author’s professional experience and a search for data that is currently offered for purchase in Austria. Then datasets will be searched in the data catalogues of Vienna and London respectively, that would be usable for fulfilling the investigated tasks. The final chapter will discuss, whether direct or partial replacement of commercial products can be considered. Or if some of the questions under investigation could be addressed in a different way using open data.

Most business models try to optimize the achievable revenue while at the same time investing as little as possible. Businesses’ revenues must cover the costs and produce some profit, in order to allow a company to survive. This requires knowledge about the possible revenue streams, the good that shall be sold, the potential customers, competitors and the preconditions or other requirements that come with the product. To illustrate these dependencies of a business, several models and templates have been created. One of the templates which is widely used for investigation of dependencies before setting up a new business or for improving a business model is the Business Model Canvas by Osterwalder and Pigneur. (Osterwalder, et al., 2010) Filling out this template, one will recognise that quite a lot of information should be gathered. If the uncertainty of the business model estimations shall be reduced, data needs to be purchased or collected upfront an evaluation.

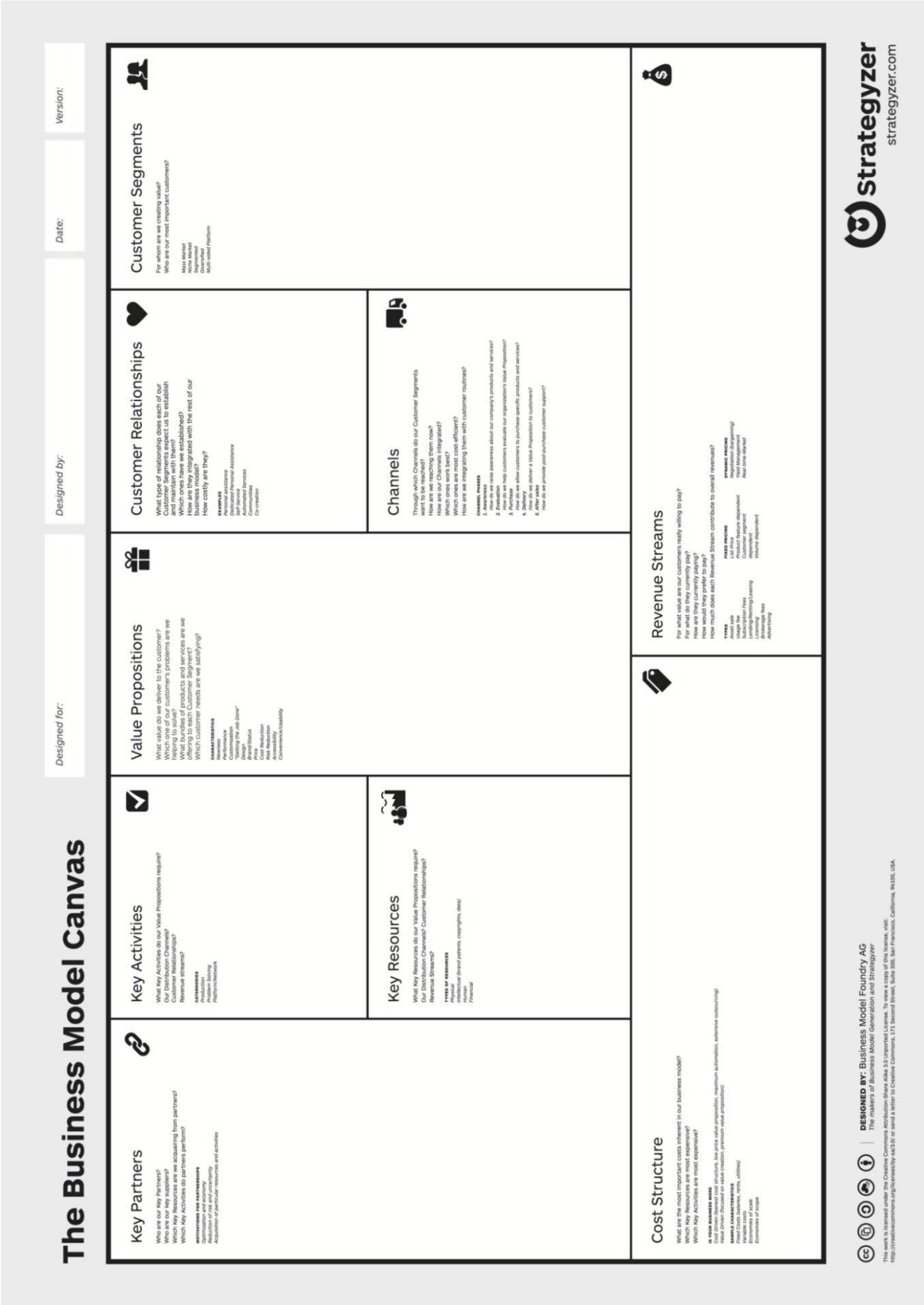


Figure 5: Business Model Canvas ("Business Model Canvas" by Business Model Alchemist[15])

---

[15] http://www.businessmodelalchemist.com/tools last accessed 2014-10

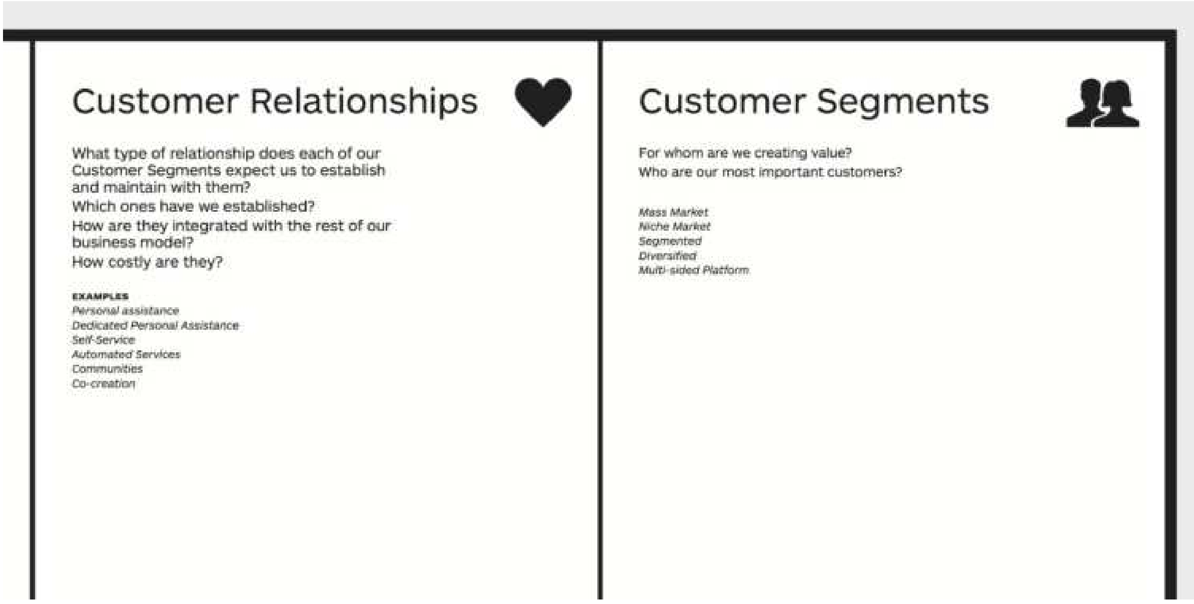


Figure 6: Detail Customer of the Business Model Canvas ("Business Model Canvas" by Business Model Alchemist[16])

Working with the Business Model Canvas, the questions about the customer segments are typically addressed first. And usually this information has a strong relation to space. The template suggests answering the questions “For whom are we creating value?” and “Who are our most important customers?” [17] Customers are typically persons or businesses. Once the target groups for the business are identified, it is crucial to know, how big they are. This will determine the maximum amount of products or services that can be sold. How the customers can be addressed is dealt with in a separate field of the template. So is the question of how to build a customer relationship. One section of the template deals with the channels that shall be used to reach the customers. In most cases the physical location of the customers as well as the point of delivery will be relevant. Online business with no physical product or personal service to be traded may be an exception. This thesis will refer to business scenarios where the customers are persons and companies. Goods or a service shall be delivered to them and their location is therefore relevant.

In this thesis the Business Model Canvas will not be discussed in detail. However, some examples of fictive business will be picked. The data catalogues that will be investigated could be of service for these business models. Consequently this thesis will only

[16] http://www.businessmodelalchemist.com/tools last accessed 2014-10

[17] http://www.businessmodelgeneration.com/downloads/business_model_canvas_poster.pdf last accessed 2014-03-19

investigate those parts of the business models, where data with spatial relevance is involved.

4.1 Store Location Planning

A company that wants to sell to customers directly needs to find suitable locations for stores. To take a closer look at this challenge, this thesis will refer to a fictive company named SampleBakery. This company wants to open up stores in a large city like Vienna or London. According to common business models like the business canvas, the potential customers need to be identified. Typically, different goods sell best within individual target groups. The target groups can be determined based on the value propositions. The Sample Bakery offers the satisfaction of a daily need by selling convenience goods or even staple food like bread. It also offers luxury goods like cakes, cookies and coffee. In a simplified model one can assume that two main target groups can be identified. One will be the people living or working in the close surrounding. They will hopefully come for their daily purchases and sometimes pick up additional luxury goods. The other one will be people who pass by and spontaneously drop in without having a prior intent to come to the shop.

Once the target groups are defined, one needs to find out, how these target groups can be addressed and attracted. In the sample chosen here, this includes the necessity to know where these people live or spend their time respectively. In order to attract the people of the target groups, it would be valuable to know as much as possible about their preferences concerning the goods the shop intends to sell.

Considering the channels the sample bakery wants to use to reach its customer segments, it plans to sell its goods in two different ways. One will be the traditional bakery shop with coffee to go. The other one will be through delivery within a certain distance to the shop. This thesis will not go into detail on the suggestions of the Business Model Canvas on value proposition, key activities, key resources and key partners. One can just assume that there will be a need to have personal for selling and delivery and the goods will need to be produced somewhere and delivered to the shop. The quality of the offered goods responds to the addressed segments. The prices will be on a market level and leave room to do promotions. These factors do not seem to have a strong spatial relevance. However, the competitors on the local market will be important. Also the real estate prices will influence the location decision, since the sample bakery will need to rent a location for its store.

As a result of this small analysis, the need to gain information on consumers, split up into customer segments was identified. To get more information on the residential customers, one needs to know how many people live in the surrounding. Since they could do their daily purchase at the Sample Bakery, it would be valuable to know as much as possible about their preferences. For example what type of bread they prefer would be of interest. Additionally, it would be interesting to know, how many employees work in the close surrounding, assuming that they could buy their lunch at the Sample Bakery or do their shopping on their way home. One also needs to know about local businesses including competitors and rental rates for real estate. In more detail, the data on local businesses should include competitors such as bakeries, stores or super markets selling bakery and bakery deliveries. Furthermore, to estimate the possible drop-in for the Sample Bakery, the characteristics of the shops in the close surrounding and the resulting passenger frequency in front of the shop location, called the footfall, would be of interest. The target groups of the shops close to the investigated location will be attracted to the region and probably drop in when passing by the outlet.

The analysis for a preferable store location therefore requires spatial data on the following topics:

- Residents within a certain distance (within the catchment area)
- Employees within a certain distance (within a catchment area and based on travel to work data)
- Competitors (bakeries, bakery deliveries, super markets)
- Characteristics of shopping streets (shop mix, passenger frequency, respectively footfall data)
- Rental rates for real estate
- Customers preferences (for example from loyalty cards, based on cultural background of similar)

4.1.1 References from Literature

According to available literature similar needs have been identified by several companies. Byrom published the results of an investigation based on a questionnaire done in the United Kingdom (UK) in 2005. (Byrom, 2005) Location planning managers were asked about the internal and external data they used. Focusing on the usage and sharing of data in the planning process, significant differences were identified between various retail and service types. The identified sectors of the retail market were food/beverage, mass merchandise, service, and speciality. The food/beverage sector contained bakeries among fast food outlets, liquor stores and public houses. Companies within the food/beverage sector made less use of internal data derived from the use of loyalty cards than did large mass merchandisers. This result seems not surprising, since mass merchandisers where the first companies introducing loyalty cards. In contrast it is not yet common to use a loyalty card of a local bakery.

Within the mass merchandise and food/beverage sectors data from planning authority applications was frequently used. The authors suspect that this was based on restrictive rules for out of town sites for groceries, which belong to the sector of mass merchandise. The high usage within the food/beverage sector could likely rise from the fact that many outlets where planned according to the major brewer’s public house expansion programs. This will not influence the Example Bakery investigated in this thesis, although it belongs to the same sector.

An article by Wood and Browne also investigates systematically which kind of data is used for retail location planning in the segments of super markets and convenience stores in the UK. (Wood, et al., 2007) It states, that considerable expertise on location decision making is established among the leading supermarket retailers in the UK. One of the reasons identified is a set of restrictive laws these mass merchandise companies need to apply. This finding coincides with the results of Byrom.

In the small store convenience market there is often not so much money available for analysis and legal restrictions are less tight. Additionally, smaller stores tend to have smaller catchment areas. It is more difficult to find or generate data in small granularity for these smaller catchment areas. Wood and Browne explain that modelling of reliable sales forecasts is also more difficult due to the fact that only a small amount of a household’s weekly expenditures are spent in this type of store. Because of these reasons, they suspect that statistical geo-demographic modelling is likely not possible. With our example of bakery stores we will face exactly these problems. While the problem of only partial spending of the households expenditures is still a major drawback, data on local residents became available in more detail since the time when the article by Wood and Browne was published.

According to Byrom again, census data, demographic data and lifestyle data where most often used in the food/beverage and mass merchandise sectors. Especially the service and speciality retailers did significantly less often use detailed information on consumers. The author actually suggests a potential for improvement for the service sector’s retail location planning. For the use of data from managed shopping centres there was no difference between the sectors. Sharing of competitor data was most likely to occur in the service sector. The use of census data was highest in the food/beverage sector. However, this data was often not shared within a company across departments. The data was widely used for location planning but not often handed on to marketing divisions. The finding, that census, lifestyle and demographic data where often used in the food/beverage sector, seems to conflict with the findings of Wood and Browne cited above. The Sample Bakery most likely belongs to the group of smaller stores with smaller catchment areas. Therefore both articles' findings are relevant for this scenario. On the one hand, all three types of data will be valuable for the location decision of a Sample Bakery store. But to be relevant, the data needs to be in small granularity. And the forecasts of spendings in the discussed store will be very difficult to make due to the fact, that only a part of the households' expenditure is going to be spent at the store.

Ten years before the publication of Byrom, Clarke and his colleagues already lined out that external demographic datasets shall be exploited in order to investigate the shopper behaviour in different kinds of site locations. Clarke et al stated in 1995 *"For instance, it will become increasingly necessary for retailers to be aware of shopper behaviour in out-of-town centres, smaller centres and town centre stores on a local level and the way in which these, respectively, affect shopper behaviour in an area. This awareness of external environment and changes requires access to and appropriate exploitation of official and other external demographic datasets."* (Clarke, et al., 1995 p. 5) During the last twenty years the availability of official and other external datasets improved and they can be used for analysis.

Modelling and analysing catchment areas is one of the main tasks within shop location planning. As early as in 1997 Benoit et al (1997) suggest constructing spatial interaction models to analyse the catchment areas of supermarkets. The following parameters where used: expenditure within the customer origin on the goods in question, a measure of the attractiveness of the store and the distance between the customer origin and the store location. The attractiveness of a store was measured by its size. Based on these parameters, the turnover figures of existing stores where predicted and the model calibrated, until the results were close to the actual numbers. Using the so gained model, the turnover of potential new stores could be predicted. The effects that the opening of a new store location would have on the existing stores could be predicted accordingly. This served as valuable decision support information for store location planning. This approach of comparing analogous stores is still valid and applied nowadays. Since the data on road networks and attributes on customers became even more detailed, the definition of catchment areas is also possible with higher accuracy than in was in the 1990s. This approach could therefore also work for the Sample Bakery, if there is already a big network of stores in place. The data collected from these stores could be used for calibrating the model created for describing the customer behaviour and turnover of the new store.

Different techniques for analysing catchment areas and forecasting the expected turnaround for new stores can also be found in a publication of Wood and Browne (Wood, et al., 2007)For example the gravity model for catchment areas is based on the parameters drivetime, attractiveness of the store and competitors nearby. Target groups can be modelled using lifestyle data like the products called MOSAIC, Personcix or ACORN that are available for the UK. By combining these data sources, it is possible to segment the population for example in terms of lifestyle, family structure or income level. The Expenditure and Food Survey (EFS) undertaken by the UK National Statistics helps to estimate the overall turnover of food expenditure originated by the population. The UK National Census provides basic information on the population. And the introduction and analysis of loyalty card data provides the most reliable source of profiling the own customers. Applying a gravity model can construct catchment maps showing drivetime, competitors, demographic as well as psychodemographic information and customer information in combination.

Applying an analogue stores approach includes searching for stores among the own ones that have similar characteristics to the planned one. Looking at their sales history enables to do estimations for the future performance of the planned shop based on own experience. Typical parameters used by Wood and Browne for comparing store locations are penetration or market share in drivetime rings around the store and data from food spending levels within the catchment area or the mentioned drivetime rings. Ideally, estimations of turnover based on analogous stores should be checked by an independent interrogation of census and other spatial datasets. (Wood, et al., 2007)Of course, loyalty cards again provide the most accurate information on the own customers. The strategic use of loyalty card data can therefore always be considered as an alternative. Retailers with few stores cannot use the technique of analogous stores due to the lack of examples.

If geo-demographic analysis is difficult or not reliable, considerations on a micro-scale level become increasingly important. This includes the precise location within a city centre, shopping street or shopping centre. Unquantifiable characteristics of small catchments need to be understood based on site visits rather than by desk-based analysis. Parameters like footfall, car parking lot, visibility, and store design become overwhelmingly important. Gravity models using walk time require a more detailed road network that than for drivetime calculations. For example footpaths need to be included in the data. (Wood, et al., 2007)

As the research on the techniques and data used for store location planning revealed, research in the area of supermarket location planning already progresses towards more complex modelling. Arentze et al (2000) describe an approach to implement a knowledge based system for the development of retail location strategies. They also use spatial models to analyse consumer flow. The parameters they use are accessibility of the shopping location, sales units per floor space and market penetration, among others. Hernandez stated the following already in 2007:

“GIS are used by the vast majority of major retail chains across North America and Europe to provide decision support for a range of location-based decisions.” (Hernandez, 2007 p. 250 referencing Hernandez and Biasotto 2001)

The cited article by Hernandez starts from this premise and focuses on visualisation of spatial information. The representation of changes in market conditions, sales trends, competitors and changed customer profiles so far is usually done by snap shots in time or by comparison of different time periods. The article addresses the lack of interaction with different user groups and suggests dynamic representation of spatio-temporal information. (Hernandez, 2007) In the context of the analysis in this thesis this will remain a forecasting perspective of possible further preparation of information for the purpose of decision support.

The article by Wood and Browne that was mentioned earlier shall be used as a reference for the datasets that would be useful for the SampleBakery as compared to a super market retail location. These are the suggested data sources by Wood and Browne (Wood, et al., 2007):

- Local workforce information
- Footfall in town/city centre
- Population
- Population classifications
- Where people live in relation to where they work
- Food expenditure

- Traffic flow data
- Datasets on the locations of schools, post offices etc. (Landmark information group)

The datasets identified in this article and in the analysis at the beginning of this chapter will be looked at in more detail. The following chapter will evaluate the benefits this data brings for decisions for stores in different locations. These can be a bakery in inner city or one in a village. Super market outlets can be used as examples in comparison where more datasets are already in use for analysis as the examples cited here show.

4.1.2 Data Requirements for Store Location Planning

According to the analysis of business cases, one should be looking for several types of information. This chapter will go into more detail on describing the requirements for the data that shall be used for an analysis. A comparison of data that is offered for purchase and OGD will be the topic of chapters that follow later on. These will include a discussion of the respective usability of the different data products for the described analysis.

About ten years ago the planning and decision making for store locations was widely discussed in literature. One of the articles that can be found on this topic contains the following statement:

> *A key prerequisite for effective locational decision-making is relevant and up-to-date sources of data. If decisions are to be made regarding the location of out lets, then quality data that can be subsequently turned into information (that is data with meaning) are required.* (Byrom, 2005 p. 65)

This is of course still valid and true, although the data available today is likely more accurate than it was ten years ago. The usage of relevant and up to date sources of data is critical for the value that an analysis adds to a decision based on it.

According to the investigation in this thesis on the example store location including the verification using available literature on this topic, requires looking for sources for information on the following topics:

- Population; Residents within a certain distance (within a catchment area)
- Local workforce information; Employees within a certain distance (within a catchment area)
- Competitors (bakeries, bakery deliveries, super markets)

- Footfall in town/city centre / shopping street; characteristics of shopping streets (shop mix, passenger frequency)
- Where people live in relation to where they work (Census travel to work area data)
- Population classifications
- Food expenditure
- Rental rates for real estate
- Street network including footpaths and traffic flow data
- Datasets on the locations of schools, post offices etc. (compare also the landmark information group)

In order to describe the catchment area of the Sample Bakery shop, an estimation shall be done for the travel time at hat is usually accepted by a customer when driving by car or when walking. The typical travel time can only be obtained by experience or through a customer survey. Typically questionnaires could be used to interrogate representative samples from the defined target groups. If a questionnaire is done, some socio-demographic data should be collected in order to have control parameters for the estimations based on purchased or open data. The individual preferences respecting the Sample Bakery`s portfolio would of course also be a part of the questionnaire. The typical estimates for drivetime or walktimethat can be found in literature are around 15 minutes.

- Information needed: estimated acceptable travel time by car or when walking

Once the values for the traveltime of customers by car or when walking are determined, one can use a road network to calculate the extent of the catchment area. The road network needs to be as complete as possible, including paths that can only be used by pedestrians. And for the travel time based calculations in requires to include one-way-streets and information on the speed that vehicles usually drive on a certain road segment. TomTom, that bought the company Tele Atlas and HERE, a brand owned by Nokia who bought the company called Navteq, both offer street networks that are widely used within navigation systems. To the author's experience, these networks are very accurate. They also widely cover paths that cannot be used with a car, but are suitable for use by bikers and pedestrians. As for vehicle traffic, the data usually includes parameters per road segment that indicate the maximum travel speed that is legally allowed. Also, an average speed that vehicles usually show on the respective segment of the street network can be found. The second parameter is very useful for the calculation of travel time distances to investigated shop location. For an analysis based on gravitation models or similar approaches, the distance to the shop will be a measure that is included in the model. It may therefore be reasonable to define zones or characteristic areas that have distinguished distances as well.

- Data needed: road network with navigation details

In parallel to the calculation of a catchment area, target groups need to be defined. Based on the data that is available, one will try to characterise the target groups using parameters that can be found in the data. For the planning of an outlet location this is typically done based on population data. In the case of the Sample Bakery one can expect that the local workforce will also be an important customer segment. Therefore we need to analyse those people as well. Segmentation and definition of target groups involves a lot of experience in order to gain a sophisticated result. Agencies developed their own methods and sell segmentation information. If loyalty cards are being used by the customers, detailed information on the existing customer base can be collected. Using this data, the segmentation of customers can be done by specialised analysts. Regardless of the detail and sophistication of the segmentation, some basic information will be needed. Population data can in Austria be purchased on a great level of detail. The Austrian federal office for statistics, called Statistics Austria publishes and sells statistical data of different level of detail. For this example in a first approach one needs to know the count of inhabitants. For more detailed analysis it could be preferable to know the size of households, the number of children per household and people's age.

- Data needed: Population data, information on age, household size, number and age of children

Besides the local residents one can expect that the local workforce will be an important customer segment. In order to estimate this target group it would be interesting to know how many employees work within the near surrounding and where they live respectively. Based on this information one can conclude, how many people pass by the selected location every day or could drop in during their lunch break because they work within walking distance to the Sample Bakery shop.

- Data needed: Data on employers and the respective number of employees

People who travel over a reasonable distance to work every day will likely spend some money on their way. Especially on their way home, commuters will try to do their shopping for goods of everyday need in an efficient way. This is a potential for the Sample Bakery shop that should be analysed. Besides the number of commuters and their destination and origin, the mode of transport they use would be interesting to know. From this information one could conclude, whether they walk or drive by the shop location that is under investigation.

- Data needed: Commuters, their origin and destination, probably their mode of transport

Another potential group of customers for the Sample Bakery are students that attend nearby schools or universities. They, too, could come to the shop in their lunch break or on their way to or from school.

- Data needed: schools, universities including the number of attendants

The catchment area, is influenced by competitors within or close to the identified area. Nearby competitors will reduce the number of customers that drop into the Sample Bakery shop. In addition, a significant part of the overall household's expenditure on products from the shop's portfolio will likely be spent in other shops than the discussed one. If one uses a gravitation model to evaluate the shop, the locations of competitors will be included into the model. But also, if one only wants to get an informational overview of the environment of the new shop, the spatial relations of the shop location to those of the competitors overlayed with the catchment area and identified customer segments, will provide a first impression of likely losses of the regional potential to the market. The locations of the competitor's stores are likely not provided as a standard data product of a data broker. Big store networks can be part of collections of Points of Interest (POI). They will however require investigation into the data quality and up to dateness. Management consultants may keep lists for their own analysis and might be willing to sell this data. Alternatively the data for an own analysis needs to be collected individually. POI lists can be of help, also the representations of stores in online maps like Google Maps. Usually the companies having chains of stores have a list of locations and contact information online on their webpage. There the data is easily available and most likely up to date.

- Data needed: Locations of competitors (bakeries and supermarkets)

Footfall data about shopping streets, shopping centres, towns and city centres are to the author's knowledge mainly being collected individually when needed. The characteristics of a shopping location tend to change according to changes in the shop mix that is there. But they are also influenced by changes in other shopping destinations. When nearby regions loose or gain attractiveness, people may change their preferred shopping destination. Shopping centres often offer information about footfall and characteristics of customers that visit their centre. Nevertheless it is wise to collect information on a specific location that is in question for a new shop. Even within a shopping centre the individual shop locations may vary strongly in their characteristics. A shop close to an entrance from a public place may have very high frequency in front of the shop and be visible from a great distance also from outside the centre. People will likely enter the building with the intention to spend some time there and the shop will be able to attract them to actually step in and buy something. On the other hand there may be locations in a shopping centre like one next to a passage to a train station, where even more people pass by every day in front of the shop. But most of them will be in a hurry to catch their train or for another destination. They most likely will only step into the shop, if it fulfils an immediate need or at least offers a quick purchase on the way. Such a location may have impressive footfall data in front of the shop, but will only be valuable for certain types of offers. Other shop locations may again be in the basement or upstairs where only a few of the shopping centre's visitors will ever pass by. The frequency of people passing by this location will very much depend on the shops in the close surrounding and is likely to change if the "neighbours" do.

- Data needed: Footfall and estimated drop in, Shop mix

Rental rates of shop locations need to be considered as well. Among the expenses that the expected income must cover, the rental rate can represent a considerable share. The relation of an expensive but good location that generates high income versus the affordability of the rent needs to be analysed carefully before opening a shop.

- Data needed: Rental rates for commercial objects

Food expenditure is related to a statistics on the spending capacity or purchasing power. The spending capacity describes the amount of money that a household has available for consumption purposes. It is calculated by subtracting periodic expenses like rents, insurances and loan payments from the income. This spending capacity can be broken down into the amounts of money that are available for expenses in different sectors of consumption. One of them is of course food. For a documented example of how to use the spending capacity within the location planning process, the master thesis by Richter can be used as a reference. (Richter, 2012) This thesis describes an evaluation of the location of a big shopping centre that was opened in Austria in 2012. A method for calculation of the catchment area is described. Based on the population and the spending capacity within the catchment area, the possible turnaround of the new shopping centre is determined. The influences of rival shopping centres and shopping streets are included in the estimations. Richter also states, that it was necessary to make estimations for some values in his model, since the required data that was not available. Some information has not been collected; other pieces of information are being kept secret. This often makes it difficult to do reproducible and verifiable analysis.

- Data needed: spending capacity and extracted food expenditure (Kaufkraft)

For addressing the potential customers in a way that comforts them, and for building up a relationship it would be useful to gain some knowledge about the preferences of a customer. The best source for this kind of information is of course the customer itself. But there are possibilities to assume some characteristics in advance. Based on census data one can for example determine, if a potential customer lives in an area where a certain mother tongue is spoken by the local population.

- Data needed: Enriched census data

4.2 Plan a mobile communication network

Planning of mobile communication networks is done with the intention to provide the best possible quality of service to as many subscribers as possible at the least necessary costs. Building a mobile communication network is limited by economic and physical preconditions like the features of electromagnetic waves and the legal and physical availability of sites and parcels where base stations can be built. (Benkner, 2007) The quality of service in a mobile communication network depends on three components. First the location where the subscriber wants to use a service must be covered. In Global System for Mobile Communications (GSM) networks, a minimum field strength of -102dBm is required, to be able to use a service outdoors. Second, the interferences with cells using the same or neighbouring channels must be very low. And third, the capacity within the serving cell must be high enough to allow another subscriber to start using a service. Also the neighbouring cells should have free capacity so that the user can move to another location and continue using his service, while he is handed on to neighbouring cells. An occurring lack of capacity will cause blocking or handover blocking leads to a so called dropped call. (Benkner, 2007)

4.2.1 References from Literature

The planning of a network for mobile communication as they are operated by mobile operators requires information of many different sources. The process of planning a site for setting up new transmitters is described in various books and articles. One article that provides a general impression of the process was written by Kriesel (2008) and shall be referenced in this thesis. He worked at the company E-Plus Mobilfunk GmbH & Co. KG, a mobile operator in Germany.

Like others too, Kriesel explains that mobile operators need to make huge investments into licenses for using radio frequencies and the technical infrastructure of their communication networks. If a mobile operator wants to remain competitive, he needs to cautiously plan these investments and try to gain best possible efficiency. Mobile communication networks in Germany contain some 15 to 20 thousand base stations. An ideal network would be homologue concerning distances between base stations, type of antennas and their height, direction and tilt. The resulting cells would be the same size. Since the factors that influence a network in reality are all but homogenous, the network can never be ideal. Usually one observes a roughness of the planet’s surface where terrain height can vary for several hundred meters within a small area. Various buildings and other obstacles block direct lines of sight. Vegetation varies from grass and bushes to trees that can have leaves or not, depending on the season. This influences signal fading and signal runtimes due to reflections. Also the population appears in different densities and likes to move around. Customers even tend to migrate during workdays. For example they move from home to work and back again, creating local and temporal densities.

The planning and design of base stations has to take these influences into account. Kriesel describes that the equipment of a base station for example varies regarding the number of sectors and the height of antenna position depending on the relievo and the height of buildings. Similarly the distance between neighbouring base stations depends on the same parameters, as well as the presence of big obstacles. The building structure in an area is commonly classified into four classes of settled regions. Rural, sub-urban, urban and dense-urban regions are treated differently in the modelling process of network planning as one will see further down. The number of customers in a certain area and the resulting expected volume of calls or data transfer is also an important input parameter for the planning process. In rural areas base stations with a single antenna often are sufficient so serve a large area. Whereas in urban or dense-urban areas small cells need to be built in order to provide the capacity that is required by the number of calls and the volume of data transferred in a small area. The high density of small cells causes the need for cautious planning of frequency channels. Interferences can easily be overlooked. And the presence of walls of the buildings in urban areas also brings the problems of different signal runtimes due to reflections.

Duque-Antón and Benkner published books on communication networks based on electromagnetic waves where they look at the technological preconditions and limitations in more detail. Duque-Antón describes that within mobile phone networks, information is transferred in the form of electromagnetic signals over the air. (Duque-Antón, 2002) The characteristics of the radio channels that are used for the transmission present the preconditions that the technical solutions have to deal with. They therefore need to be taken into account also when planning a radio network for mobile communication. The range of electromagnetic waves depends on the type of wave, its frequency and its power. The higher the frequency, the lower is the range. Under ideal circumstances in open space the electromagnetic waves of a radio signal spread radially symmetrical and the power at the receiver decreases according to the distance from the sender to the power of two. The antennas that are used in mobile communication networks bundle the emitted power into one direction. Additionally to the signal attenuation depending on the distance to the sender, there are other factors that influence the radio wave propagation conditions. Even the atmosphere changes in its characteristics with changing weather conditions. For example higher frequency areas, above 12 GHz, show high signal attenuation when water drops are in the air like it happens when it rains or it is foggy. In this case the electromagnetic waves are heavily dispersed and absorbed by the water drops. Benkner also stresses out that on the way from the sender to the receiver an electromagnetic wave is decayed due to several influences. The distance between sender and receiver has the biggest impact. But also the frequency, respectively the wavelength and the surrounding influence the decay. (Benkner, 2007)

Under ideal circumstances like in outer space a signal das was emitted spherically from a sender would decay to the power of two in relation to the distance from the sender. Most senders are built in such a way that there is a main direction of emission. This directed emission shows higher density of emitted power and allows higher efficiency due to bundling of power. In reality of mobile communication the propagation of radiation always happens near to the earth's surface and can therefore be modelled as propagation along a plane. Additionally this plane is not perfectly flat and reflecting, but there are hills, buildings and vegetation in the way of radiation propagation. According to Benker who refers to models by Okumura and Hata, the power decay can be assumed to be in the dimension of values determined by distance to the power of three to distance to the power of five. The curvature of earth's surface is normally not relevant, since the distances between sender and receiver are small enough. (Benkner, 2007)

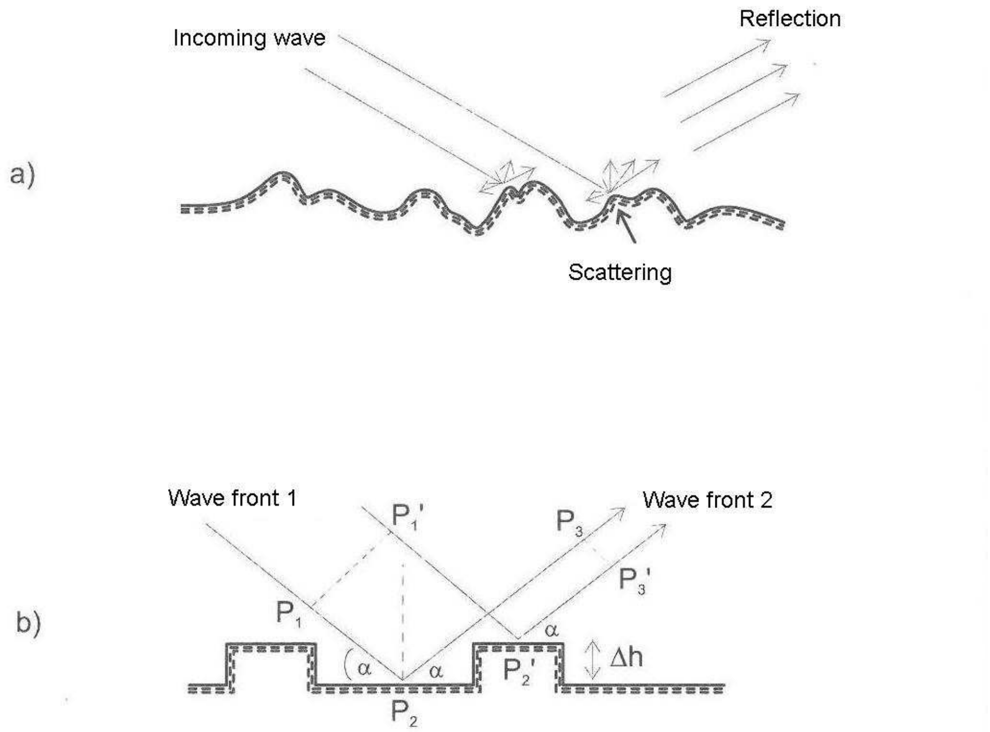


Figure 7:Reflection on rough surface a) real situation, b) ideal model (Benkner, 2007 S. 28, translated)

In reality of mobile communication there are always obstacles and reflecting surfaces in the way of radio propagation or close to it. Therefore, in addition to the signal attenuation depending on the distance there is signal attenuation due to reflection. Diffraction and scattering can be observed at obstacles like hills, vegetation water areas or buildings. Duque-Antón describes that both parts of the signal, the one that takes the direct way and the reflected part overlay when they arrive at the receiver. Due to the fact that the signal spreads over different path (called multipath propagation), the actual decrease of power at the receiver is much higher than it would be under perfect conditions. Depending on the circumstances, the power at the receiver decreases according to the distance from the sender to the power of five in urban built up areas like cities. (Duque-Antón, 2002) Benkner states more generally, that reflection causes higher decay in power. According to him it can be modelled to be about distance to the power of four instead of distance to the power of two like in open space. (Benkner, 2007)

Benkner also describes that along rough surfaces a part of the signal is distracted which corresponds to a diffuse reflection. There is no exact distinction between plain and rough terrain. According to the effect on the signal propagation, an average deviation in height in relation to the wavelength and the angle at which the signal hits the plain can be found. For example, for the frequency of 2 GHz, that is common for UMTS, having a wavelength of 15cm and an incoming angle of two degrees a surface with a standard deviation for heights of more than 3,5m must be considered to be rough.

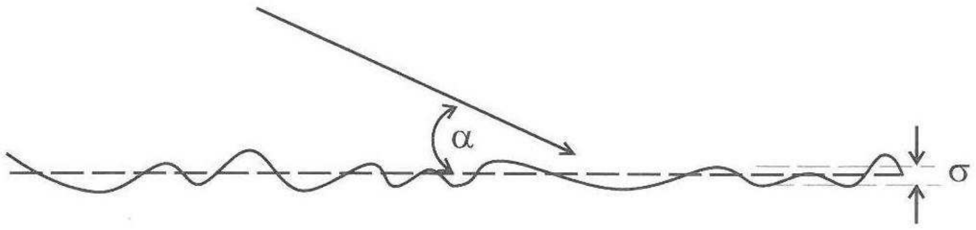


Figure 8: Incoming angle and average deviation in height (Benkner, 2007 S. 29)

Another phenomenon described by Benkner is diffraction. Diffraction of electromagnetic waves is an important phenomenon in the situation where there is an obstacle in the way of propagation that blocks the direct line of sight between sender and receiver. Despite the barrier the receiver can still be reached with a decayed signal. The calculation of diffraction is complex, but important for the design of transmission paths that require over the air connections with best possible quality. In this case the area of the so called first Fresnel Ellipsoid where the signal decay is highest needs to be free of diffracting obstacles. (Benkner, 2007)

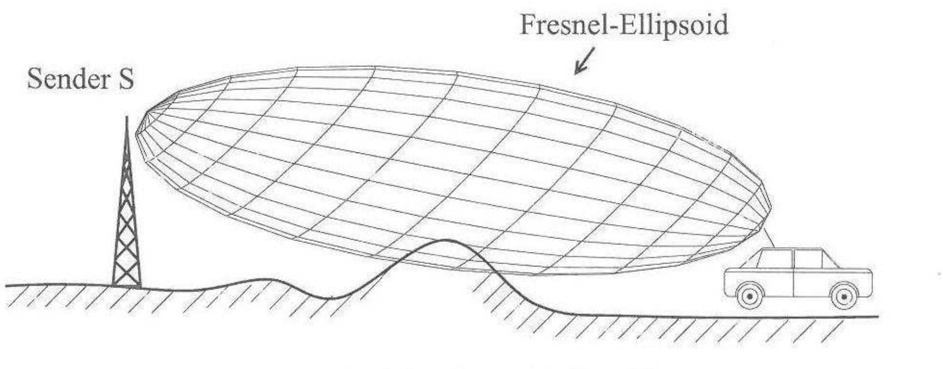


Figure 9: Fresnel Ellipsoid (Benkner, 2007 S. 34)

According to Benkner the signal decay of electromagnetic waves above real terrain with hills, valleys, houses and vegetation cannot be calculated with deterministic models. For the planning of mobile communication networks and the determination of ideal positions for base stations it is important to predict the expected signal power. For this purpose a number of empiric and semi-empiric models for signal propagation have been developed based on series of measurements. Okumura made extensive experiments in the 1960 in the city of Tokyo and its surroundings. He published his results in the form of graphs that built the basis for many models that where developed for predicting signal decay in mobile communication networks. (Benkner, 2007)

Another challenge in the process of planning a mobile communication network is the limited number of frequencies that are available for building mobile communication systems. The World Administrative Radio Conference assigns limited areas of the radio spectrum to single systems. Network operators can then apply for licenses to use certain frequencies. These license for frequencies are very expensive and therefore they need to be used efficiently. For example the GSM-System around 900MHz was assigned a maximum of 125 frequency channels. These can for example be divided by time multiplexing into eight channels each and therefore result in 1000 channels over all. (Duque-Antón, 2002)In order to be able to serve several millions of mobile subscribers, the frequencies need to be used repeatedly with some spatial spreading. This spatial frequency reuse led to the development of a cell based design for mobile communication networks. Applying this technology mobile communication services can be offered with financially worthwhile densities of users while at the same time the probabilities for blockings and the quality of service remain acceptable. (Duque-Antón, 2002)

As Duque-Antón describes, the cell based design of mobile communication networks divides the area that shall be covered into cells. Each of these cells is covered by one base station and a subset of the available frequency channels. No two neighbouring cells are allowed to use the same frequencies. This would lead to interference and result in low quality of the transmitted voice call or other service. Frequencies can only be reused in a cell at a certain distance. Using low emission power at the base station, the resulting cells have a predictable size. Under ideal circumstances, the cells would be circular, which is in model networks approximated by hexagons. In reality the cells have various forms and overlap for about 10 to 15%. The topography (structure of the terrain) and morphology (buildings and built up areas) of the terrain determines the actual size and form of the cells. These parameters are therefore used for creating models of a mobile communication network based on the positions of base stations.

According to practical experience there are different types of models that are typically used to describe different types of cells. Duque-Antón states that macro cells (having a diameter of 3-35km) are typically described by an empiric model like the one described by Omukura in the 1960s. This model distinguishes between the following structures of terrain: Quasi-smooth Terrain and Irregular Terrain. The later one consist of hilly areas without single mountains, single mountains, rising or falling terrain and a mixture of land and water areas. Concerning the morphology the following types are distinguished: Open Area without big obstacles, Suburban Area like villages with houses and trees and Urban with buildings with at least two storeys. Based on this model by Omukura, different variations have been created and integrated into planning tools for mobile communication networks. (Duque-Antón, 2002)

Benkner lists the single steps of planning a mobile communication network as shown in Figure 10. A market analysis should first describe the behaviour and wants of possible customers. It will be the basis for an estimation of numbers of subscribers and the development of a customer base. These estimations highly depend on assumed coverage areas and available capacity. Based on these numbers a first estimation for profitability and feasibility can be made. Once the area that should be covered is defined, regional planners define possible sites for base stations. Next to the physical requirements that need to be fulfilled, the sites must be available and need to be connected to the energy grid and a data network. The analysis of coverage is a complex calculation based on several data sources. The models that are used are usually based on the measurements made by Okumura. But they need to be adapted to the individual needs and surrounding and they are permanently improved and tuned by the mobile operators. Source data that is required for the models includes an extensive database on the topography of the area in question. It needs to provide the altitude of the terrain and information on constructions and vegetation. The data is usually stored and processed in the form of a grid of pixels. For each of these pixels the model will calculate a medium (median) field strength. The resolution of these grids is usually well below 100 meters and can be in the dimension of several ten meters. As a result of the analysis of coverage, every pixel can be assigned the base station that provides the highest signal strength at that location. This base station is also called the best server. The sum of all the pixels that are best served by a cell represent its coverage area. According to the intended quality of service, respectively blocking that will be tolerated in this area, the number of required channels will be determined. Finally the frequencies must be assigned, considering neighbourhoods and overlapping of cells to avoid interferences due to other cells that operate on the same or neighbouring frequencies. (Benkner, 2007)

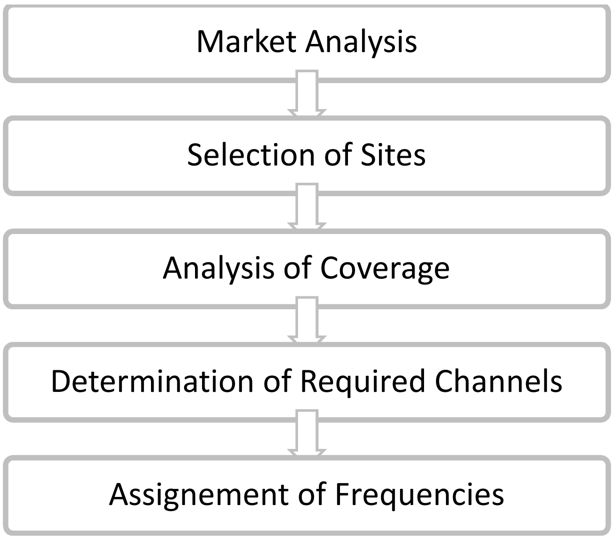


Figure 10: The steps for designing a mobile communication network (Benkner, 2007 S. 149, translated)

In an analogous way, new base stations need to be planned, if dead zones appear in an existing mobile communication network. Kriesel describes his approach if a dead zone that should be closed is identified within the mobile communication network. As a first measure, a search area is defined by the radio network planner. Usually this is done by defining a circle around a centre point and a set of required parameters that a new base station should ideally comply to. Within this area a building or other place where a site can be built should be searched for. The parameters can contain geographic coordinates, height above sea level, building heights, height, azimuth and type of antennas, the technology that should be implemented and other parameters. Based on these parameters predictions can already be calculated in order to estimate the effects that such a new base station would have within the network. The radio planner creates a request to search for a site that fulfils the requirements as good as possible. An acquisition agent then tries to find a spot that fulfils the planners needs and could be rented from the owner in order to build a base station there. Usually a small number of possible sites are identified and investigated in detail. The decision to go for one of the sites found is then a compromise between technical, legal and economic needs. The new base station should erase the dead zone and allow seamless handover to all the neighbouring sites. The later requires an overlap to the coverage area with certain signal strength. Once a site is selected, negotiations with the owner need to result in a contract. Then the construction plans for the new site need to be finalised and a building permit needs to be requested. This can take some time and is a prerequisite for starting the actual construction of the base station. When the structure is built, it needs to be equipped with antennas, feeders and a cabinet containing the IT-components. Only then the site can be switched on and integrated into the network. The process of building and integrating a new base station usually takes months and requires a lot of money. Accurate planning is therefore important in order to establish a profitable business. This planning requires information in the form of data for the analysis of possible sites.

4.2.2 Data Requirements for Planning a Mobile Communication Network

The description above shows, that the data required contains information on the terrain (a terrain model), land use (clutter), population residence, customers and buildings.

The requirements for the terrain model vary in different regions. For unsettled areas a rough estimation of the terrain is generally sufficient, while in densely populated areas the terrain model shall be as accurate as possible. The grid size of the terrain model shall correspond to the prediction modelling for the spread of electromagnetic waves that is done on top of it. Basically the use of building information is nothing else, than the further increasing the terrain information for prediction models in urban areas. Actually using building information in the calculation of prediction models dramatically increases the calculation time. The calculation of signal reflections and resulting signal paths is very consuming in terms of calculation steps and therefore requires huge calculation capacities. This is the reason, why the availability of building data is generally not crucial and can even be treated as an add on information that is not compulsory. Both, the terrain model and building data need to be as up-to-date as possible. While the terrain in rural areas normally does not change much, there can be changes due to construction activity in urban regions. And certainly the buildings that compose a city tend to change regularly.

- Data required: Terrain model (to derive mountains and slopes), optional are details on built up areas

Land use information that can be used in prediction models for mobile communication networks does not require very sophisticated classification. As described in this chapter, a small number of categories of land use need to be distinguished. Among these are open area with no big obstacles, with or without vegetation, suburban areas like villages with houses and trees, urban areas with houses of at least two storeys. This land use information shall ideally be available in a grid dataset having a resolution and accuracy that corresponds to the terrain model and the calculation of the prediction model.

- Data required: Land use information

Population data is available in different forms as it became obvious in the chapter on store location planning. For the planning of a mobile communication network it would be favourable to know how many people reside in each cell of the planning raster and when they are usually there. One source that can be used is the census data. It shows how many persons live at a certain location. Therefore one can assume that these people usually spend a lot of time at that location. However, most people leave their home at some time during the day and go to work or school, do some shopping or meet people at other places. Therefore it is also interesting to know, where the typical locations are, where people reside besides their homes and which roads or means of transport they use to get there. This data is more difficult to find and the actual need must be defined individually. Obviously, information on employers and the numbers of employees would be of interest. Also hot spots like big schools, universities, hospitals, shopping centres and other locations where a great number of people regularly reside might be of interest for the planning of required capacities in a mobile communication network.

- Data required: Population data with high level of detail, number of employees, hot spots

5 Evaluate differences between open portfolios and commercial offers

Based on the examples from the business scenarios in the previous chapter this chapter will suggest an evaluation for the value of the relevant datasets that could be found in the open data catalogues. This will be done by comparing the elements from the respective portfolios regarding a set of parameters that can be used to describe data quality and fitness-to-use for the scenarios described in the earlier chapters. The matrix for evaluating spatial data quality shall be used as a guideline and it will be extended with parameters that seem to be useful in the context of this thesis.

	Space	Time	Theme
Accuracy			
Resolution			
Completeness			
Consistency			

Figure 11: Evaluation matrix (Veregin, et al., 1995 p. 168)

Accuracy in space time and thematic context is documented in literature to be one dimension of spatial data quality. It can only be tested against a reference source of known higher quality. Therefore in the context of this thesis the accuracy of the data found in the data catalogues can only be tested and measured if there is a reference dataset available that is supposed to be of higher accuracy. If such a dataset is available, the locations of coordinates will be compared. As it is described in the chapter on data quality, comparing the measurements of timestamps is difficult and there is no possible application with the datasets investigated in this thesis. Errors in the thematic domain will be addressed according to the description in the chapter on data quality if applicable.

Resolution describing the level of detail for the spatial, time and thematic dimensions will be compared where it is possible to compare the datasets that can be found. Additionally, the consequences for using the data in the scenarios described in this thesis will be elaborated.

Completeness can only be measured by testing the content of a dataset in relation to what degree it matches the description provided by the source, respectively in the metadata. The datasets found in the open data catalogues will be tested for errors of omission and for discrepancies between the content of the dataset and the description in the metadata.

Consistency in the spatial domain refers to topological errors. The data that can be found will be tested for basic topological integrity. Since there are usually no topology rules published with datasets, only some obvious rules can be applied, like polygons need to be closed and have one reference dataset. In theory the same approach can be used for testing temporal topology. Most likely there are no rules to be found that could be applied for this purpose. Therefore test for temporal consistency will not be performed. In the thematic dimension, consistency includes the relationships between different attribute values. Where applicable this type of consistency will be tested.

In addition to these categories defined by the evaluation matrix from Veregin some other parameters seem to be relevant for the usability of the data under investigation. **Lineage**, scope, sources and their reliability can be subsumed under the concept of lineage as it is described by Clarke (Clarke, et al., 1995).

Interfaces including the data formats, data structure as well as accessibility, automatic updates and their compatibility to earlier versions determine how a dataset can be integrated into applications. These parameters also influence the requirements for maintenance of an application that uses this data.

The **licenses** attached to different datasets will limit their usability due to legal restrictions. If several datasets are used in one application the most restrictive licensing model will determine the possible target groups and usecases for the application.

The resulting categories for evaluating the fitness-of-use for the data investigated in this thesis will therefore be:

- Accuracy
- Resolution
- Completeness
- Consistency
- Lineage
- Interfaces
- License

In this chapter the data that is available on a specific topic will be investigated each at a turn. The respective datasets that are offered by the cities of Vienna and London as well as commercially available products will be compared according to the parameters listed above.

5.1 Boundary Datasets

Most of the information discussed in this chapter on data from open data portfolios is provided in the form of tables containing attributes of spatial entities. These entities are mainly the districts or counting districts of Vienna and the boroughs or wards of London. In order to work with this data, the geographic representations of the spatial entities are needed as well. Both data catalogues contain boundary datasets that are provided in a GIS readable format. In this thesis the ESRI shape files where downloaded and used.

5.1.1 Boundary Datasets for Vienna

The catalogue of Vienna offers geographical features in different formats. Among those are for example datasets in ESRI shape file format, as well as offers fin the form of Web Map Services (WMS) and Web Feature Services (WFS). Spatial data layers for the borders of the country, districts and counting districts can be found in the Vienna data catalogue and have been downloaded for this thesis.[18]

The city of Vienna consists of 23 districts that are divided into 250 counting districts.

[18]Country:https://open.wien.at/site/datensatz/?id=86c29af5-c52c-423d-8764-82c1bb037936

Districts:https://open.wien.at/site/datensatz/?id=2ee6b8bf-6292-413c-bb8b-bd22dbb2ad4b

Counting districts: https://open.wien.at/site/datensatz/?id=efbdbab2-a095-422b-acb4-81cc9c380093

Last access 2014-10

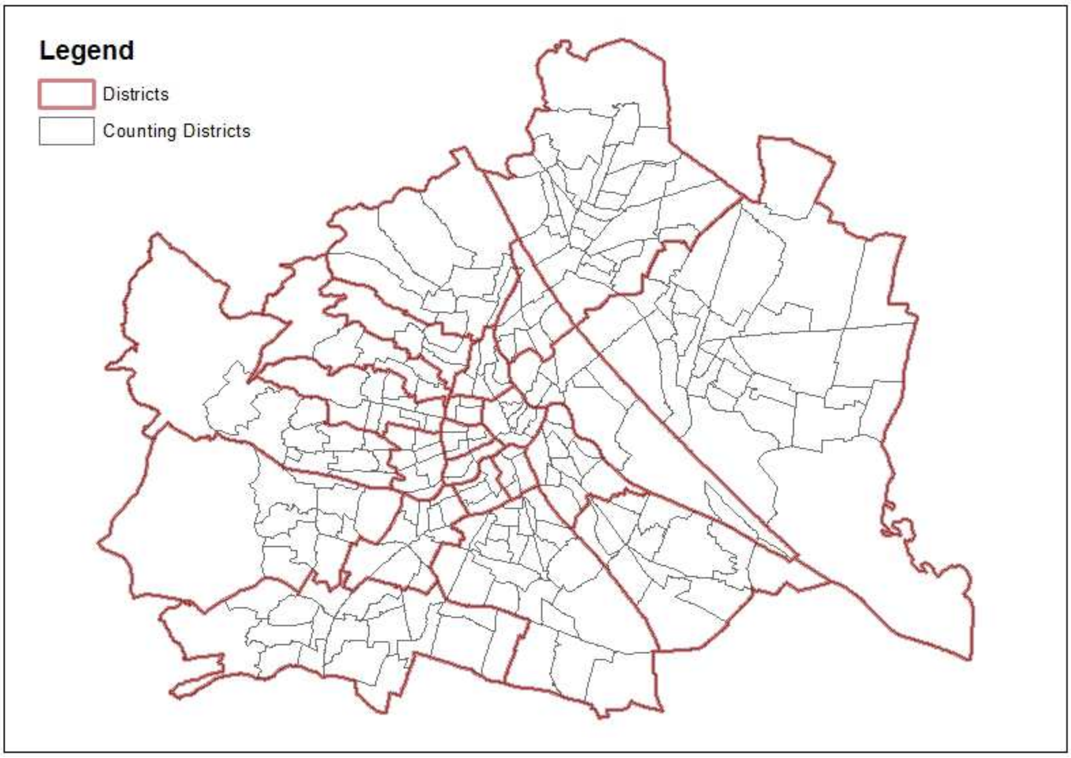


Figure 12: Overview Vienna districts and counting districts

5.1.2 Boundary Datasets for London

The London data store provides two different archives of boundary datasets. One is a compressed archive with a number of different boundary files for the Greater London area.It is called the "Statistical GIS Boundary Files for London", and is dating from October 2013. The source for this information is documented to be the "Opinion Resarch and Statistics" and the data ranges from 2004 to 2011. It contains ESRI shape files and Map Info documents of the London Boroughs, the London Wards with the city of London merged into a single area, Output Area (OA) of 2011, Lower Super Output Area (LSOA) of 2004 and 2011, Middle Super Output Area (MSOA) of 2004 and 2011. The boundaries of the OA and MSOA files have been generalised in order to reduce the file size.[19]This dataset is used and referenced to in this thesis unless stated differently.

Another entry in the London Data Store for boundary files is the also compressed package of "2011 Census Geography Boundary Files", which is dated from November 2012. The Shape Files are grouped into folders for Output Area (OA), Lower Super Output Area (LSOA) and Middle Super Output Area (MSOA). And each category is provided at Extent of the Realm (BFE), Coastline (BFC) and Generalised Coastline (BGC). Each folder

[19]http://data.london.gov.uk/datastore/package/statistical-gis-boundary-files-london last access 2014-10

contains separate files for every borough both in ESRI and MapInfo format. In Total the uncompressed archive uses more than 500MB of disk space although the description on the homepage states that the boundaries of the OA and MSOA files have been generalised in order to reduce the file size. A separate file containing all the London Boroughs could not be found in this archive. Most of the files contain information on the households and number of usual residence as well as the resulting average size of the households and the population density. The source for this information is documented to be the Greater London Authority and the date of the data is of 2011.[20]

The greater London Area consists of 33 boroughs that contain 625 wards.

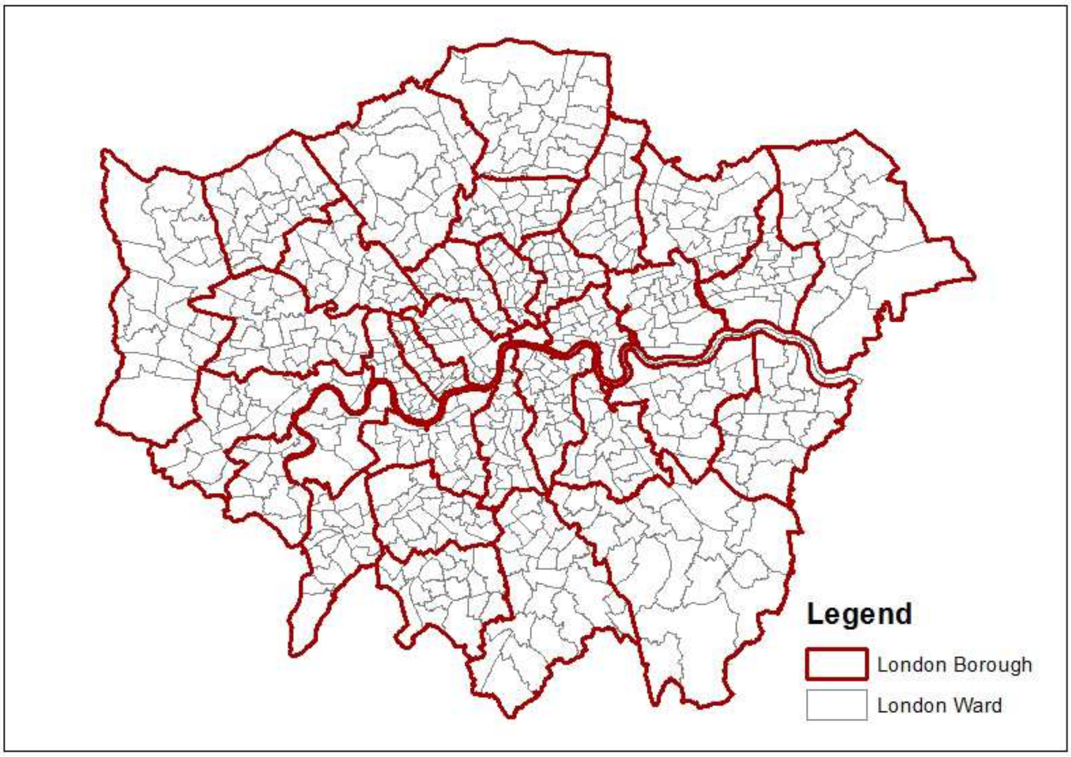


Figure 13:Overview London boroughs and wards

5.1.3 Boundary Data Evaluation

The cities of Vienna and London differ quite a bit in areal size and population. Therefore the boundary data for statistical attributes has a slightly different structure, too.

[20] http://data.london.gov.uk/2011-boundary-files last access 2014-10

	Vienna	London
Population	1.758.159 (sum Zählbez)	8.173.941 (sum UsualRes in OA2011)
Area (square km)	415 square km (sum Bezirke)	1595 square km (sum Wards)

Table 1: Population and aerial size of Vienna and London

Accuracy

Testing the data from the Viennese and London data catalogues is not possible, since there is no dataset of higher accuracy available for comparison. But the metadata states that the boundaries of the OA and MSOA files of the London datasets have been generalised in order to reduce the file size. This should be kept in mind, while working with these datasets.

Resolution

The mean size of the polygons of counting districts in Vienna is 1,66square kilometres (1.659.484 square meters) and 2,56square kilometres (2.551.515square meters)of the London Wards. The London Wards are about 54% larger than the Vienna Counting Districts and the London Boroughs are about 164% larger than the Vienna Districts.

		Mean Aerial Size	Count
Vienna	Districts	18,04	23
	Counting Districts	1,66	250
London	Boroughs	47,68	33
	Wards	2,56	625

Table 2: Count and mean aerial size of boundary polygons in Vienna and London.

Completeness

Testing the boundary data files for completeness does only make sense in the form of checking for omission of identifiers and names for the single polygons that are contained in the datasets. This test has been made for the following datasets and no missing values could be found.

Statistical GIS Boundary Files for London:

- London_Borough_Excluding_MHW
- London_Ward_CityMerged
- LSOA_2004_London_Low_Resolution
- LSOA_2011_London_gen_MHW
- MSOA_2004_London_High_Resolution

- MSOA_2011_London_gen_MHW
- OA_2011_London_gen_MHW

Boundary files for Vienna:

- BEZIRKOGD
- ZAEHLBEZIRKOGD

Consistency

Testing the boundary data for consistency includes tests for a valid spatial topology. Attribute consistency includes the distinct usage of identifiers. Testing for consistency between separate fields in the datasets does not make sense in this context and has therefore been omitted.

Lineage

The boundary files for counting districts of Vienna are created by a department of the city government called Magistratsabteilung 21 - Stadtteilplanung und Flächennutzung. The boundary files for the London Boroughs and Wards are generated by the department of National Statistics. There is no information on planned updates of the data.

Interfaces

Both data catalogues offer the spatial boundary files in common GIS formats like ESRI Shape Files and Map Info Relations. The boundary files for counting districts of Vienna are updated weekly. Interfacing with this dataset is also possible via Web Map Server or Web Feature Server. Besides using the static file format, the data can therefore be integrated into a mesh-up application and will always be up-to-date for the user. The London data is only provided in the form of static files. These can easily be integrated into applications. The London boundary files contain a number of attributes that can on the one hand be useful for different types of applications. On the other hand the great number of attributes containing data from another source, the Ordnance Survey data, increases the probability for the occurrence of troubles when replacing the data with updated versions. On the one hand the format of fields like the data type, header names or the number of attributes may change. And on the other hand, it is more complicated to keep all the attributes up-to-date to the same temporal accuracy, especially if different data sources are involved.

License

The boundary files from the Vienna data catalogue are published under “Creative Commons Namensnennung 3.0 Österreich” This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes.

Both London boundary archives originate from National Statistics data and Ordnance Survey data, both under the Crown copyright and database right. On maps created using the boundaries from the London Data Store entry called “Statistical GIS Boundary Files for London” from October 2013, the copyright must be stated. The wording is provided on the homepage as follows: "Contains National Statistics data © Crown copyright and database right [2012]" and "Contains Ordnance Survey data © Crown copyright and database right [2012]".

The 2011 Census Geography Boundary Files (dated November 2012) where generated from the Office for National Statistics (ONS) 2011Census geography products for England and Wales and need to be licensed accordingly. The Office for National Statistics states on its homepage that the boundary products and reference maps are produced under the Open Government Licence and Ordnance Survey Open Data License. This requires the following copyright statements to be used when reproducing or using the data: “Contains National Statistics data © Crown copyright and database right 2013” and “Contains Ordnance Survey data © Crown copyright and database right 2013”

5.1.4 Usability of boundary data sets

Both cities provide boundary datasets in an easily readable and usable form. As the investigations for different datasets in the London data store unvealed, there was a different coding scheme in use up to 2011, for referencing boroughs and wards of London. Several datasets still use these old codes while the boundary datasets do not contain any clue, on how to combine elder attribute data with the current boundary data.

5.2 Population Data

In Austria, population data can be purchased from a federal institution called Statistics Austria (Homepage: www.statistik.at). There the data is available in various different granularities. The most detailed dataset is based on a 100m raster. In this granularity the following dimensions are available:

- count of inhabitants,
- buildings,
- building with flats and
- count of flats.

This raster is defined according to the directive of the European Union called INSPIRE. Other sizes of raster datasets with more detailed information on the population are 250m, 500m, 1000m, 2000m, 5000m and 10000m. These can be filled with

- people's age,
- nationality,
- country of birth and
- employment status

A set of parameters describing the buildings and flats within are also available. Details on this data product can be found on the product web page at the Statistik Austria homepage.[21]

5.2.1 Population Data Vienna

Within the data catalogue of the city of Vienna a number of datasets containing population data can be found. The information on the current number of citizens is enriched with information on nationality or country of birth. But the greatest level of detail in terms of spatial distribution of citizens is the level of counting districts.

Datasets that can be found contain among others:

- Count of inhabitants
- Nationality[22]
- Country of birth[23]
- Migration background[24]
- Buildings and apartments/flats

On the level of counting districts a classification of the population based on age could not be found. Age is however available on district level. Here, the following classification is used: 0-19, 20-64, 65+ and 85+ [25]

---

[21] http://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/regionalstatistische_rastereinheiten/index.html last access 2014-10

[22] Nationality: https://open.wien.at/site/datensatz/?id=efbdbab2-a095-422b-acb4-81cc9c380093 last Access 2014-10

[23] Country of Birth https://open.wien.at/site/datensatz/?id=5d4f02be-c4bd-4bf8-946d-s715b04a69733 last Access 2014-10

[24] Migration Background https://open.wien.at/site/datensatz/?id=269f3c91-87a9-46ee-8557-d18e216cc4d1 last Access 2014-10

[25] https://open.wien.at/site/datensatz/?id=21783686-d037-4cce-8b2c-86a03df17fbc last access 2014-10

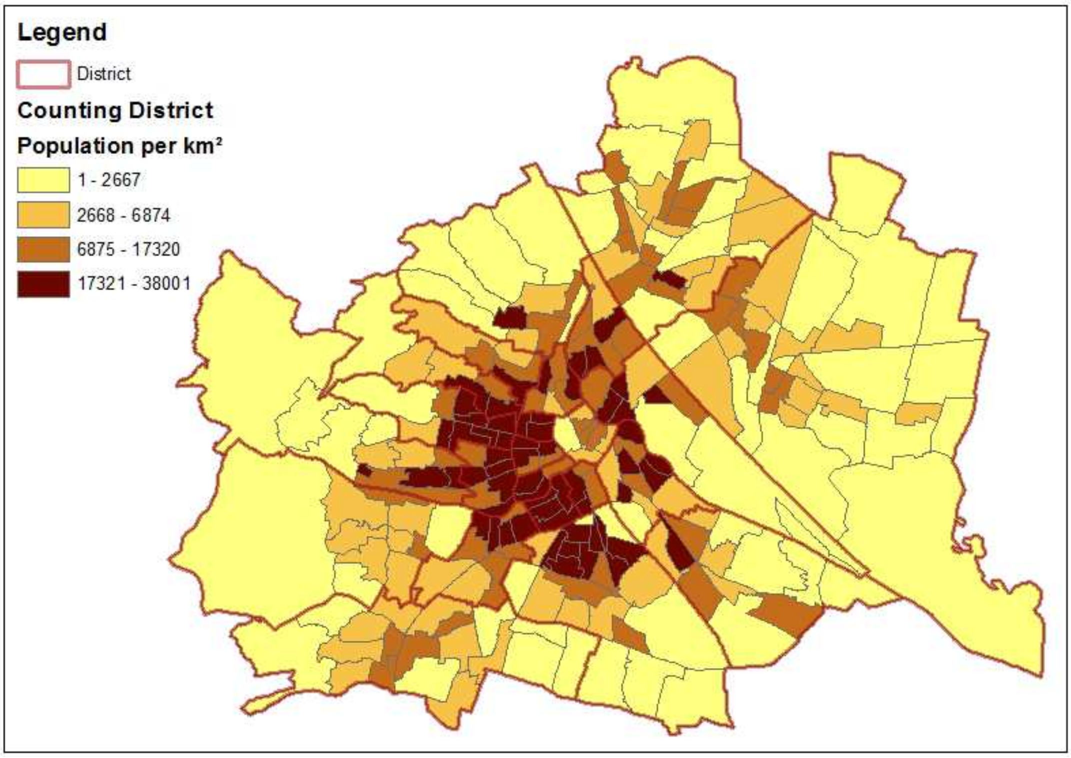


Figure 14: Vienna: population density in counting districts as of January 2013.

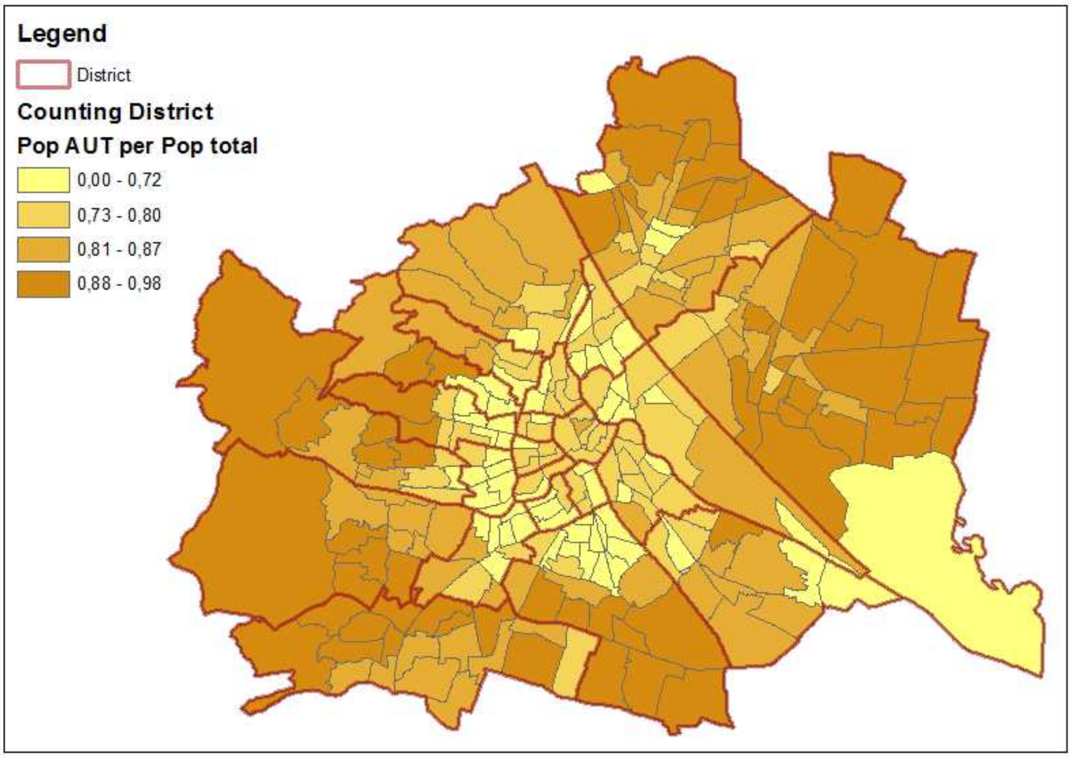


Figure 15:Share of people with Austrian nationality in Viennese counting districts as of January 2013

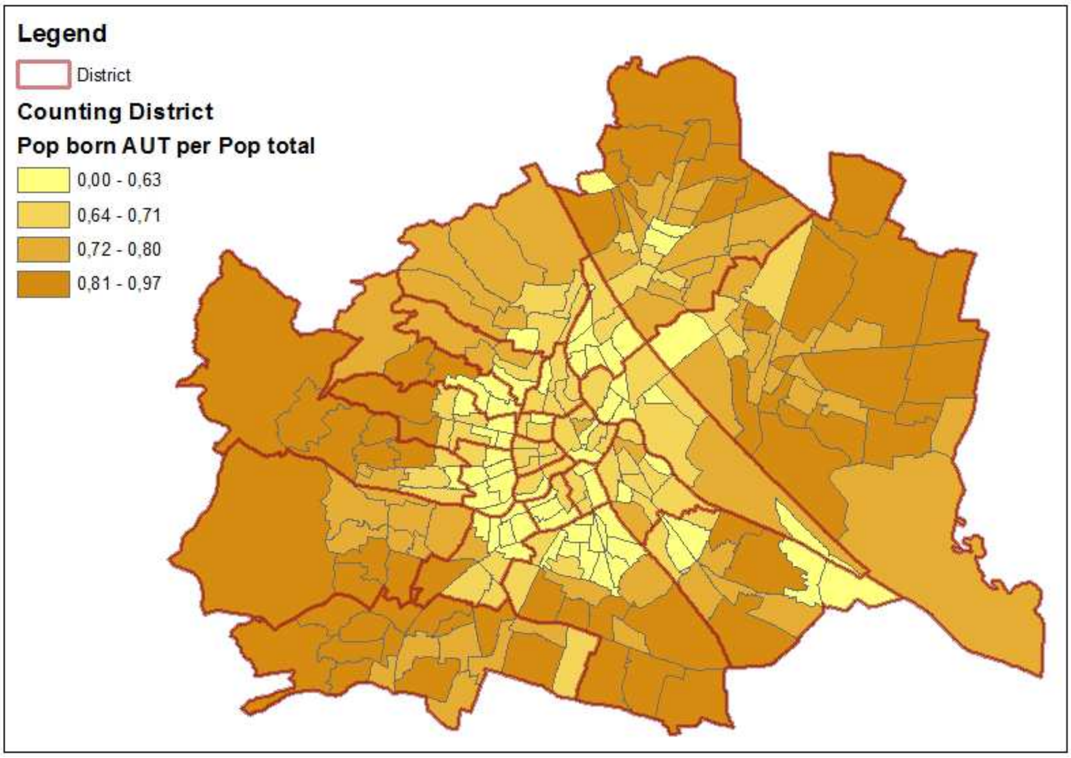


Figure 16:Share of people born in Austria in Viennese counting districts as of January 2013

5.2.2 Population Data London

London provides a set of excel spreadsheets with various information in the granularity of boroughs. This excel sheet contains various information on the population on the London boroughs from different sources and years.[26]

Among others there can be found:

- Count of inhabitants from Greater London Authorities
- People's age divided into the groups of 0-15 Years, working aged (16-64), and 65 years and older as well as the average age
- People's nationality is represented by the first, second and third largest migrant group and their percentage among the population
- Country of birth is not directly included, but the number of people with overseas nationals entering the UK is given as well as migrant rates.
- Employment status is provided in several aspects like male and female unemployment rate, youth unemployment rate and employment rate overall.

Additionally there is information on

[26] http://data.london.gov.uk/datastore/package/london-borough-profiles last access 2014-10

- Education level of the working age population in two categories (persons with no qualification and those with a degree or higher education)
- Annual pay as gross value in general and split up into male and female gross annual pay
- Number of people who volunteered in the past 12 months
- Number of jobs by workplace, percentage of employment that is in public sector, two-year-business survival rate
- Crime rate, fire and ambulance incidents
- Information on house price, ownership or rental status,
- Percentage of greenspace in the borough
- Number of cars, bicycle-usage, public transport accessibility score
- Life expectancy, teenage conception
- Scores for happiness, worthwhileness, anxiety and life satisfaction
- Proportions of seats won in the 2014 election by Conservatives, Labour and Liberals and Democrats

The excel sheet contains links to the various sources of the data presented, but some of the referenced web pages do no longer exist.

Information on the count of apartments is contained in a dataset that is also looked at in the chapter on buildings, the dwelling stock by tenure and condition on borough level.[27]

Information on buildings and flats per building can be found on borough level. The offered data on dwelling stock by tenure and condition is discussed in detail in the chapter on building data.[28]

5.2.3 Population Data Evaluation

Population data can be downloaded from the data catalogues of Vienna and London as well. Additionally, there are commercially offered products and all three datasets differ in various ways.

Accuracy

The datasets that can be found in the catalogues of Vienna and London have not been tested for accuracy since there is no dataset of higher quality available for comparison.

[27] http://data.london.gov.uk/datastore/package/dwelling-stock-tenure-and-condition-borough last access 2014-10

[28] http://data.london.gov.uk/datastore/package/dwelling-stock-tenure-and-condition-borough last access 2014-10

The data that is sold by Statistik Austria could be such a dataset of higher quality that could be compared to the data published by the Viennese government. But this data was not available for this thesis.

Resolution

The datasets that are offered for free show less detailed granularity than the geo-statistical raster datasets offered by Statistik Austria. Both Vienna and London offer datasets that are based on polygons resulting from census areas. The mean size of the polygons of counting districts in Vienna is 1,66square kilometres (1.659.484 square meters) and 2,56 square kilometres (2.551.515square meters)of the London wards. In comparison, the raster data set of Statisik Austria covers squares of 100 meters per 100 meters, which equals 0,01 square kilometers. Additionally to the lower cell size, the raster dataset provides the advantage, that is can easily be used in combination with other raster datasets of similar granularity. This can for example be useful when planning a mobile communication network. The population data of London is only provided on the level of boroughs. But the boundary datasets contain numbers of inhabitants also for wards.

Completeness

The datasets published by the Viennese government where tested for attribute completeness. There are no missing values in the tables. But some of the datasets contain more than the 250 counting districts that where published in the same data catalogue. Most of the districts contain one additional counting district having a number at the end of the districts numbering range. This additional counting district-id refers to a small number of inhabitants. There is no explanation in the metadata or on the homepage of the data source for these additional objects in the tables. One can only suspect that there is some incomplete information in the data source and these persons could therefore not be assigned to one of the counting districts for sure.

Consistency

Attribute consistency was tested for the datasets from Vienna by checking the total values against the different classes and no errors could be found. Consistency of the London datasets was only tested for the corresponding attributes like age and migration background. Among these only for the classes of peoples age totals can be built and checked. No errors could be found.

Lineage

The datasets offered by the city of Vienna are published by a department within the city government that is responsible for statistical data called Magistratsabteilung 23 – Wirtschaft, Arbeit und Statistik. In London the data comes from the federal statistics agency called Office for National Statistics (ONS).The population datasets offered for free by the governments of Vienna and London all are originated from governmental agencies that are responsible for collecting and producing statistical datasets. Therefore in both cases one can assume that the data is as reliable as any comparable dataset that can be purchased from an official authority.

Interfaces

The data from Vienna is published as a comma separated table. But the files contain header rows with explanation that need to be removed, before the data can be imported into a database. Once this is done, the data of different reference points in time can easily be accessed, since all the available data is contained in one table ad simply distinguished by an attribute value. The data from the London data catalogue is provided in the form of a Microsoft Excel file. It contains three different sheets, of which one holds the data that can be used in GIS software. The other two are designed for presenting the data to end users in the form of interactive profiles in the form of tables and overview maps. This makes it easy to get a first impression on the distribution of certain data values. But for importing the data into a database, the table holding the source data needs to be adapted. The changes include the replacement of all the field headers. These are defines as long descriptive texts, containing special characters. At the end of the table there is a number of summary lines that need to be removed as well. Furthermore the table contains data from different years and various data sources. If an update of this table is intended, it can only be done reasonably by adding further rows to the table. This would be very inconvenient, if the data and possible updates shall be used in an application since the underlying database schema would need to be adapted. Using this data source in an application should be avoided if possible. The data sources referenced in the Excel file of the London Borough profile should be used instead, as far as they are still available.

License

Vienna publishes the data under the licence terms called “Creative Commons Namensnennung 3.0 Österreich” This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes. The London Borough profile contains data from different sources and the publication page does not contain explicit information on the overall license agreement.

5.2.4 Usability of Population Data

As we saw in the chapters above, population data is offered on a very detailed level in the spatial dimension by the federal agency of statistics. This kind of information can be very valuable for the planning of base stations for a mobile communication network. Neither of the data catalogues of Vienna or London can fulfil the requirements for this task. In Austria, purchasing data from Statistic Austria will likely be the only reasonable action for planning such expensive infrastructure in the required spatial level of detail.

For the purpose of decision support for the location of a shop, the data can be useful at least for getting a first impression of the situation. Based on the analysis of open data, more detailed information can be purchased, if necessary for the regions that shall be looked at in more detail. The London Data store provides a great number of different attributes for the London Boroughs. These could be used for the identification of certain target groups. Limiting factor remains the spatial resolution that does not allow for analysis in small scale areas. And the fact that there is no licensing statement for the data might be considered as a risk when working with the data.

5.3 Digital Terrain Model

Digital terrain models (DTM) are typically created for governments or governmental projects. If someone wants to buy a DTM for Austria off the shelf, the Federal Office for Surveying and Metrology called Bundesamt für Eich- und Vermessungswesen (BEV) is a place where one would want to look for it. On its homepage this institution offers a DTM in different horizontal solutions that can be purchased for a defined area of interest. The highest resolution that can be purchased is a raster dataset with a cell width of ten meters. The website also provides detailed information about how the data was generated and the resulting errors that can be observed in the positional accuracy. Depending on the type of terrain the accuracy can be considered to be between one and ten meters if the data was generated according to recently developed methods. Data from older processing are stated to have an accuracy of up to 25 meters.[29]

[29] http://www.bev.gv.at/portal/page?_pageid=713,1572954&_dad=portal&_schema=PORTAL last access 2014-10

The city of Vienna also offers a three dimensional city model that is built from the following components: The digital terrain model providing the surface height, the city map holding the actual land use information, the model of building objects consisting of prisms representing the building heights and the model of roofs that contains detailed information on the roof structure in different areas of the city.[30]This 3D model of Vienna, respectively its components, are updated permanently.

5.3.1 Digital Terrain Model Vienna

The digital terrain model (DTM) offered by the city of Vienna is a point raster with a horizontal resolution of ten meters.[31] It was created in October 2012. In the metadata there is no additional information like vertical accuracy or how the dataset was created. A link leads to contact information of the creator who is responsible for the content. The creator is the Viennese department of survey. The dataset lacks information about its coordinate system and spatial extent. Therefore it is somewhat difficult to import the data into a GIS and work with it in combination with other data. If the user follows the link provided in the metadata he finds the homepage of a division of the Viennese government that is mentioned as the author of the dataset, the department of land survey (Stadtvermessung). The websites of this organisation contain descriptions of data products they offer for purchase. Among these is also a digital terrain model that seems to be the source data for the one offered in the open data catalogue. The original product that is offered commercially contains additional information in the form of vectors that describe break lines in the terrain. The documentation of this data product is available online ant it offers detailed information on the methods that are used for creation and the resulting spatial and temporal accuracy. The coordinate system is also provided, including the responding EPSG-Code. 31256.[32]The department of land survey also offers a dataset that contains information on the height of the individual buildings above the terrain. This dataset is suitable for creating a 3D-Model of the city and is presented as part of the governmental online GIS application.

---

[30]https://www.wien.gv.at/stadtentwicklung/stadtvermessung/geodaten/stadtmodell/produkt.html last access 2014-10

[31] https://open.wien.at/site/datensatz/?id=82764cdb-a0e0-4e64-ba8f-31cc9a303c5a last access 2014-10

[32]http://www.wien.gv.at/stadtentwicklung/stadtvermessung/geodaten/dgm/produkt.html last access 2014-10

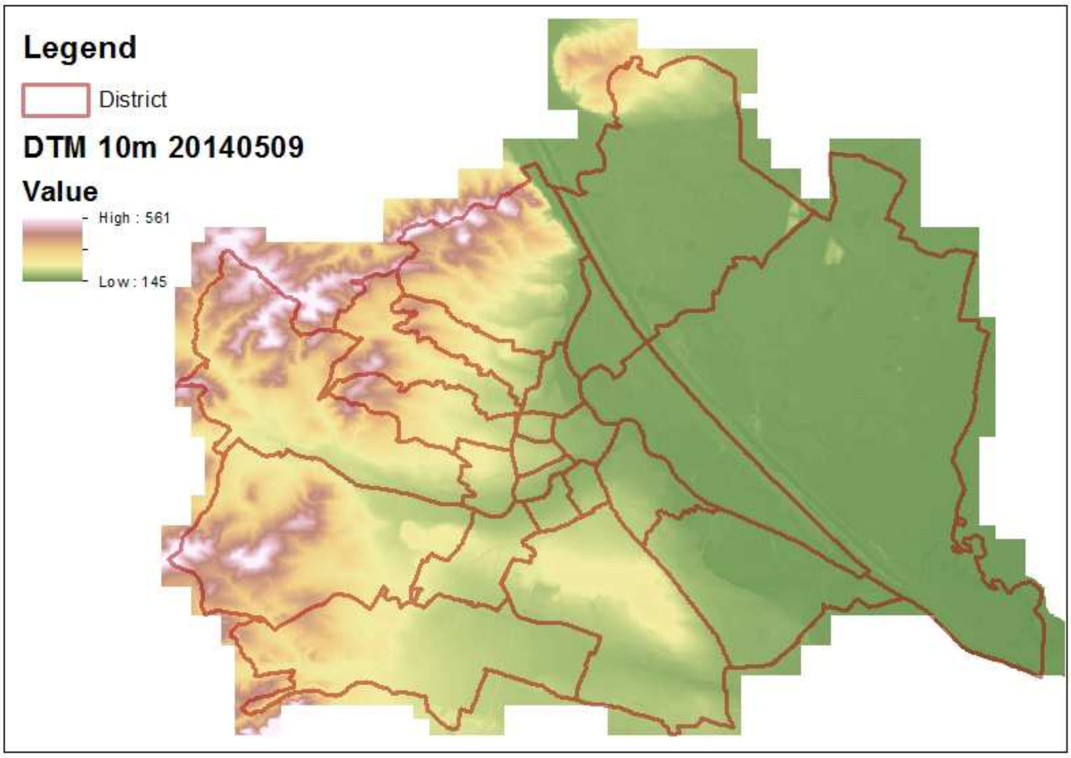


Figure 17: Digital Terrain Model Vienna

5.3.2 Digital Terrain Model London

The open data catalogue of London does not offer a Digital Terrain Model nor does it contain a Digital Surface Model. Searching for terrain models for London in the internet showed that there are commercial offers available for the area of London. The data catalogue of the UK government contains a number of elevation models that are all limited in use by license terms. For example a dataset based on Light Detection and Ranging (LIDAR) measures, can be found there.[33]LIDAR is an Airborne Laser Scanning (ALS) method. This means that a laser scanner is mounted to a plane or helicopter. It then sends laser signals from the air down to the surface. These signals are reflected by the surface or objects on the surface and again detected by the laser scanner. Modern ALS systems detect not only one signal, but a whole footprint that is generated by laser beams that are sent out in defined angles. While the plane moves on, the changes between the registered footprints can be interpreted into a three dimensional model of the earth's surface. A short description of the principle of ALS can be found in the bachelor thesis by Martin Wieser. (2010) The resulting dataset is a Digital Surface Model (DSM) which means, that it does not show the contours of the ground level, but of the surface, including buildings and vegetation.

[33]http://data.gov.uk/dataset/lidar-digital-surface-model last access 2014-10

According to its description the LIDAR data can be used to derive maps of land use and to identify buildings and roads or other infrastructure. The resolution of this data lies between 25 centimetres and 2 meters. Unfortunately the use of the data is limited by licensing and there is no version available that can be used under the conditions of Open Data. The dataset is linked with a detailed set of metadata.[34]

5.3.3 Digital Terrain Model Evaluation

A Digital Terrain Model can only be downloaded from the data catalogue of Vienna, but there are commercially offered products available for both cities. The London Data catalogue does not offer a terrain model. Alternatively the LIDAR homepage could serve as a source for a terrain model but the licensing would need to be clarified.

Accuracy

The datasets that can be found in the catalogue of Vienna has not been tested for accuracy since there is not dataset of higher quality available for comparison. The data that is sold by Bundesamt für Eich- und Vermessungswesen could be such a dataset of higher quality that could be compared to the data published by the Viennese government. But this data was not available for this thesis.

Resolution

The horizontal resolution of the dataset offered by the Viennese government is comparable to the one of the commercially provided dataset from BEV. Both claim to offer datasets that contain raster information with a raster size of 10 meters. The resolution of the terrain model offered by LIDAR is 2 meters and below that. Both the commercial products of the Viennese department for survey and BEV can also be purchased with additional information on break lines in terrain represented as lines. These increase the actual resolution of the data since relevant changes in terrain can be identified with much higher accuracy than it would be possible using the raster dataset alone. Regarding the attributes there could no relevant difference be found in resolution of altitude values.

Completeness

Testing for completeness is not reasonable for this kind of dataset.

---

[34] http://data.gov.uk/harvest/gemini-object/2b1eb4d5-0bb3-4670-b8df-1e90f569afd4 last access 2014-10

Consistency

Attribute consistency is not relevant for Digital Terrain Models and was therefore not tested.

Lineage

The dataset offered by the city of Vienna is published by a department within the city government that is responsible for land survey called Magistratsabteilung 41 - Stadtvermessung. The dataset originates from laser scan information that was enriched with the contours derived from the city map. Additionally, structural lines like stairs and retaining walls have been included. According to the metadata file the LIDAR data originates from surveys carried out by the Environment Agency and represents over 15 years of survey work carried out by a specialist team in remote sensing, based in Bath. The Terrain models offered by the government of Vienna and LIDAR all are originated from governmental agencies that are responsible for collecting and producing land survey data. Therefore in both cases one can assume that the data is reliable.

Interfaces

The data from Vienna is published as an ASCII file. The data can easily be imported into a database. LIDAR data is provided as an ASCII file among other formats as well.

License

Vienna publishes the data under the licence terms called "Creative Commons Namensnennung 3.0 Österreich" This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes. LIDAR provides data under the individual license models. The data is therefore not offered for free out of the box.

5.3.4 Usability of Digital Terrain Model

Only the data catalogue of Vienna contains a Digital Elevation Model for free. The resolution of this dataset is suitable to get a first impression of the terrain of the city. For detailed analysis one will likely consider buying the more detailed Version either from the city government or from the national institution Bundesamt für Eich- und Vermessungswesen. Also, if the building structure of the city or a three dimensional model is of interest, one needs to buy a commercial product. In London there is no alternative to the commercial products of the Environment Agency.

5.4 Land Use

Information of land use is typically derived from satellite or other aerial images. There are companies that offer land use datasets in different horizontal resolution and with different level of detail regarding the categories. This kind of data is often tailored for a special purpose or for a group of potential customers with similar requirements. Therefore this kind of information with high spatial or thematic resolution and reliability is usually rather expensive.

5.4.1 Land Use Vienna

The actual land use in the area of Vienna is available online.[35] According to the metadata, the information offered was generated from the interpretation of aerial images and completed with information based on the experience of employees. 32 categories of land use are hierarchically structured into three levels. The spatial accuracy corresponds to the building blocks of the Viennese spatial reference system, which was used to provide the geometric basic data. The most recent dataset that was published is dated from the year 2012.

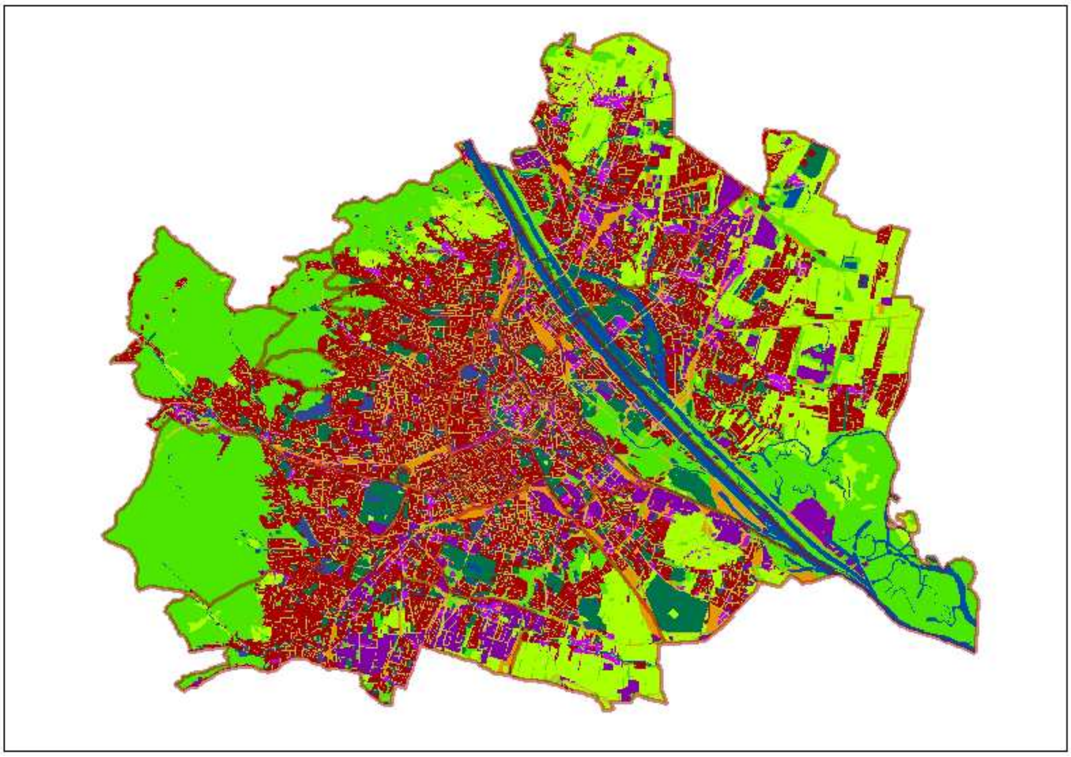


Figure18: Land use in Vienna (Realnutzungskartierung)

[35] https://open.wien.at/site/datensatz/?id=2f5baa1f-208c-42c2-8d04-9ea74aa1b229 last access 2014-10

5.4.2 Land Use London

On ward level, land use information for the area of London is provided in the form of tables that list areas of certain land use per ward.[36]This dataset on ward level was created from data collected in 2005 and there are no updates planned to be made. The statistics is also stated to be in an experimental status and was not yet accredited as a National Statistic. A closer look at the data shows, that the attributal reference system that was used in this dataset is not valid any longer. In 2011 the coding system for administrative units in the United Kingdom was changed by the Office for National Statistics.[37] This dataset can therefore not be linked to the boundary datasets that are available in the London data store any longer.

On the level of boroughs there is a dataset available on land use of previously developed land. This so-called brownfield land may be available for re-development.[38] This dataset contains information from a temporal range of 2004-2009 and is updated annually. But it covers only a very small aspect of land use.

The area of a designated green belt is delivered as a separate data set. The area of designated Greenland per borough is provided in the form of a table and additionally there is a geographic dataset available that can be read with a GIS.[39]

[36]http://data.london.gov.uk/datastore/package/land-use-ward last access 2014-10

[37]http://en.wikipedia.org/wiki/ONS_coding_system last access 2014-10

[38] http://data.london.gov.uk/datastore/package/land-use-previously-developed-land-borough last access 2014-10

[39]http://data.london.gov.uk/datastore/package/area-designated-green-belt-land last access 2014-10

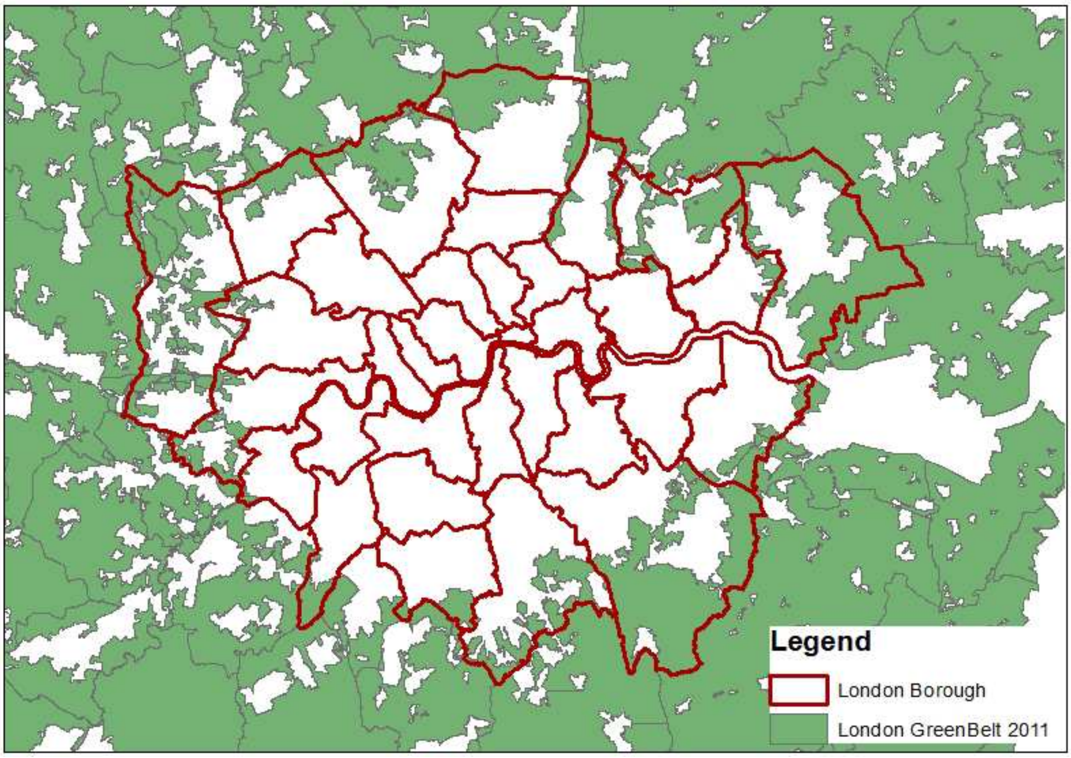


Figure 19: London overview and green belt area

5.4.3 Land Use Evaluation

Some land use information can be downloaded from the data catalogues of both cities, Vienna and London. And there are commercially offered products available, too.

Accuracy

The datasets that can be found in the catalogues of Vienna and London have not been tested for accuracy since there is not dataset of higher quality available for comparison.

Resolution

The datasets that are offered by the city of London for free show less detailed granularity than the data offered by the Viennese government. London offers a dataset that is based on polygons resulting from census areas. The resolution as well as the spatial accuracy of the Viennese dataset corresponds to the building blocks of the Viennese spatial reference system. Only the area of the designated green belt is offered in a higher resolution also for the city of London, but it covers only a small aspect of land use. The attributal resolution offered by Vienna is also higher. There is a list of 32 values that have been interpreted from aerial images. The London dataset lists a number of ten different land use categories that are given as a percentage per ward.

Completeness

The datasets published by the governments of Vienna and London where tested for attribute completeness. There are no missing values in the tables.

Consistency

Tests for attribute consistency are not reasonable for these datasets since the attributes contain only codes and descriptive text.

Lineage

The dataset offered by the city of Vienna is published by a department within the city government that is responsible for city planning called Magistratsabteilung 18– Stadtentwicklung und Stadtplanung. The data was generated trough interpretation of aerial images and completed with additional information from specialists that was used to remove uncertainties in the interpretation. The datasets on previously developed land per borough that is offered by the city of London was generated by the Department for Communities and Local Government (DCLG).This is a ministerial department that is supported by several agencies and public bodies. The data focuses on previously developed parcels, also called brownfield land that could be reused for buildings. Estimates on this potential are made by specialists and published in the form of count of sites and count of dwellings per category in each borough. All the datasets that could be found were published by governmental institutions. They can be trusted to publish information that is reliable. But when using the London data per ward one needs to keep in mind, that it is stated in the metadata to be in an experimental status and not yet accredited as a National Statistic. It will probably never be accredited since the dataset dates from 2005 and has never been changed since then.

Interfaces

The data from Vienna is published in various different file formats. ESRI shape files can be used as well as images. The data is also offered in the form of a Web Feature Service, Web Map Service and other service formats that can be used for a distributed system. The file data can easily be imported into a geo-database. The service formats can be used for integration into an application. The London data store contains the designated green land in the format of ESRI shape files. The data on ward and borough level that is offered in the London data catalogues is formatted in Microsoft Excel files that are designed to be read by end users and not by applications. The data needs to be prepared and formatted so that it can be imported into a database. Since the data of the London Wards is still in an experimental status, one must be prepared, that the format and content of this file could change with the next update. Or that no updates might be available at all. If an update will be made in future, the current coding system for administrative units will be used, as it is nowadays common in the UK. Therefore the versions might not be compatible and will likely need manual adjustment if they shall be used together.

License

Vienna publishes the data under the licence terms called “Creative Commons Namensnennung 3.0 Österreich” This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes. The data offered by the city of London is published under the Crown Copyright license. In the explicit context of Open Government Licensing this requires mentioning the source of the dataset as well as providing a link to Open Government License (OGL). Otherwise the data can be used for non-commercial as well as commercial purposes.

5.4.4 Usability of Land Use Data

Referring to the business scenarios that are used in this thesis in order to evaluate the usability of spatial data, the land use information is required for the planning of a mobile communication network. In this context, the data that is published by the city of Vienna would be very useful. Both, the spatial and the attributal resolution are high enough to derive areas that are relevant in size and that show certain characteristics in this kind o analysis. The data offered by the city of London is not usable for such an application. Both types of resolution in this data are not sufficient for this purpose. The data that is provided on wars level cannot even be linked to the spatial reference system that is provided in the London data store without manual adjustments. And it is likely outdated as well. In reality, however, the planning of a mobile communication network requires consistent information on a greater area than just one big city. The generation of classes and attributes within this data shall be consistent for the whole area that is required for an analysis. Only if these preconditions are provided, the results of a prediction of electromagnetic wave propagation can be optimised to a sufficiently reliable level.

5.5 Buildings

Information on buildings can on the one hand describe information on the physical structures and on the other hand it can inform about the usage and legal situation. Data on the physical structure is strongly related to digital models of the terrain. The availability of information on building height will only be mentioned briefly in this chapter and is discussed in more detail in the chapter on Digital Terrain Models.

The Austrian federal institute for statistics, Statistik Austria, designed a database called Adress-GWR-online that holds a vast list of information related to buildings, parcels and flats. Currently not all the parameters are available yet for every building in Austria, but the database is updated continuously by local authorities.[40] Statistik Austria also offers a raster dataset with a raster size of 100m containing information on buildings and flats within the buildings.[41]

5.5.1 Buildings in Vienna

The Viennese data catalogue contains datasets on buildings and apartments per counting district. The table on buildings provides information on the number of apartments per building, the usage as private home or commercially used objects. The type of ownership is documented as well as the number of storeys and the period when the building was built.[42]The table on apartments provides information about the number of inhabitants per apartment, the number of rooms and the size of the apartment. It also informs whether the apartments are rented or owned by the inhabitants and gives an overview on the sanitary standard of the apartment. The last one includes the type of toilet and water supply in the apartment or on the hallway.[43]

---

[40] Link to the document: http://www.statistik.at/web_de/static/gebaeude-_und_wohnungsregister_ab_29._maerz_2010_041634.pdf Link to detailed homepage at Statistik Austria: http://www.statistik.at/web_de/services/adress_gwr_online/index.html last access 2014-10

[41] http://www.statistik.at/web_de/klassifikationen/regionale_gliederungen/regionalstatistische_rastereinheiten/index.html last access 2014-10

[42] Gebäude in Wien Gebäudemerkmale: https://open.wien.at/site/datensatz/?id=f82153eb-de34-42d8-898c-254d3119e117 last access 2014-10

[43] Gebäude in Wien Wohnungsmerkmale: https://open.wien.at/site/datensatz/?id=54180602-1640-4b26-b9ca-2000c9218307 last access 2014-10

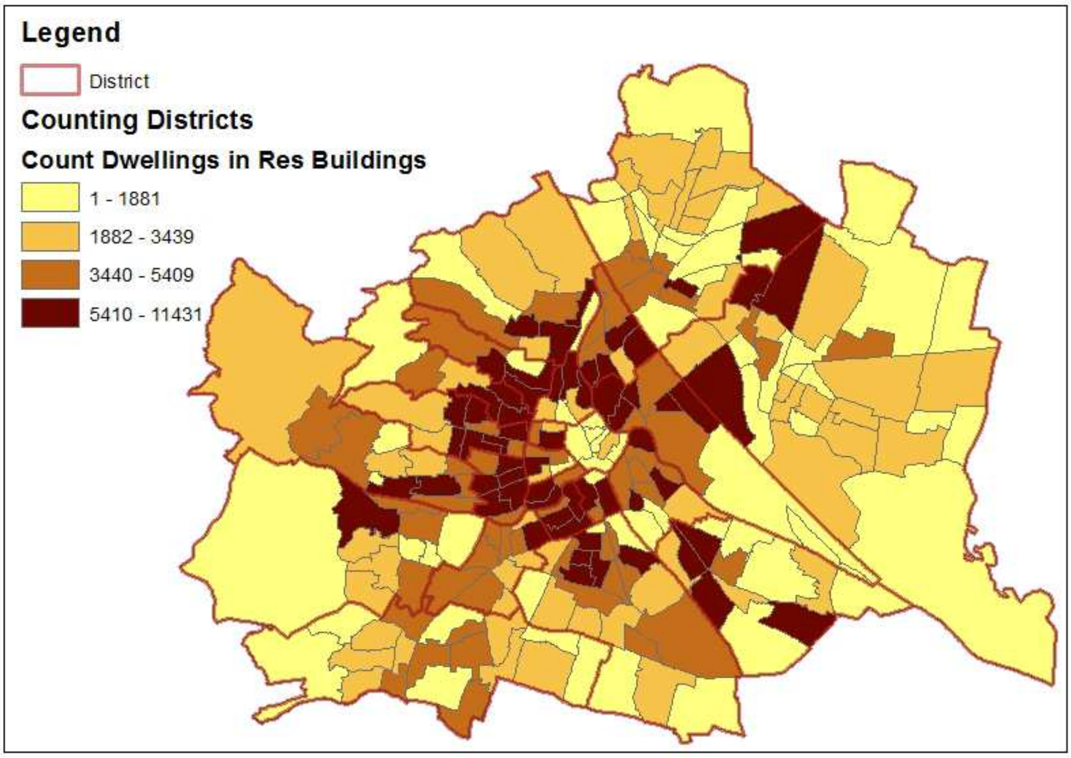


Figure 20: Apartments in Vienna in residential buildings (per counting district)

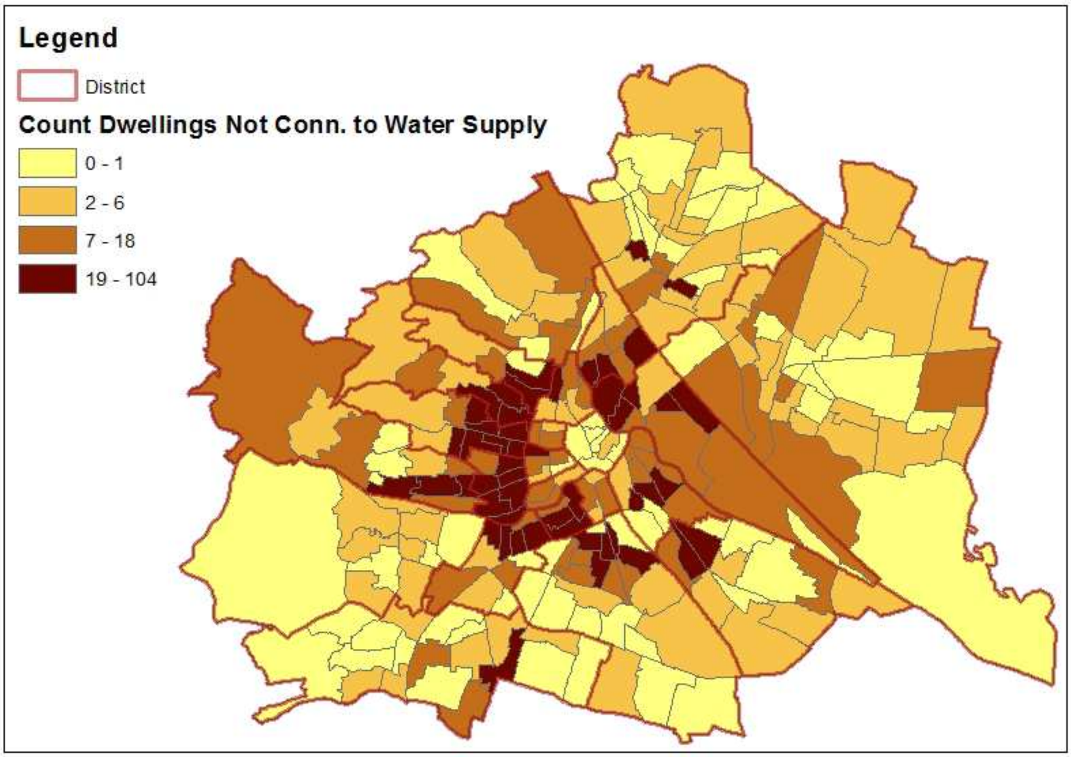


Figure 21:Apartments in Vienna that are not connected to water supply (per counting district)

The city of Vienna also offers a three dimensional city model that is built from the following components: The digital terrain model providing the surface height, the city map holding the actual land use information, the model of building objects consisting of prisms representing the building heights and the model of roofs that contains detailed information on the roof structure in different areas of the city. This 3D model of Vienna, respectively its components, are updated permanently.[44]

5.5.2 Buildings in London

Dwelling stock by tenure and condition is provided per borough.[45]This dataset contains counts and percentages per borough for total dwellings and a number of categories. Among these are public sector dwellings of different kind opposed to owner occupied and private rented dwellings. The condition is represented in three categories: unfit, “Decent Home Standard” and three types of “Capital Type”. The data is available for the period of 2001 to 2011 in the form of one datasheet per year. The data is published by a department of the Office for National Statistics (ONS) called Neighbourhood Statistics. It underlies crown copyright. Unfortunately the codes that are used to identify the single boroughs do not match the current coding scheme that is used with the polygons that are available in the London data store. Therefore the data cannot be linked to the polygons or to other data.

Net additional dwellings per borough are also provided.[46] This dataset contains estimates for count of dwellings per borough, the annual changes and the average persons per dwelling. The data is published by the Department for Communities and Local Government (DCLG). The available data ranges from 2004/05 to 2013 and the next update is announced for 2014.

Information on ward level could not be found in the London data store.

---

[44] https://www.wien.gv.at/stadtentwicklung/stadtvermessung/geodaten/stadtmodell/produkt.htmllast access 2014-10 last access 2014-10

[45] http://data.london.gov.uk/datastore/package/dwelling-stock-tenure-and-condition-borough last access 2014-10

[46] http://data.london.gov.uk/datastore/package/net-additional-dwellings-borough last access 2014-10

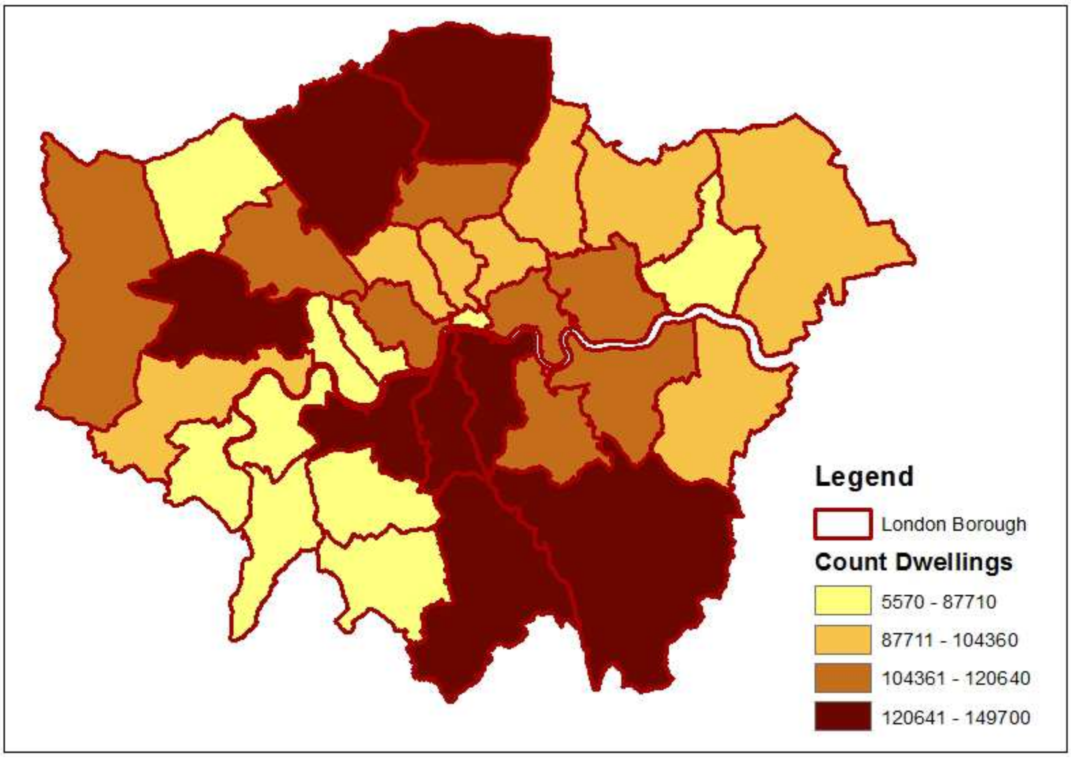


Figure 22: Dwellings per borough in London.

5.5.3 Buildings Evaluation

Information on the physical structure like the building heights are not offered in open data catalogues. But this information is available for purchase for example for the city of Vienna. Information on the usage of buildings, the number of apartments and some other parameters is available in both open data catalogues. As it is mentioned in the chapter on population data, Statistik Austria also offers a raster dataset containing buildings and the flats within the buildings.

Accuracy

The datasets that can be found in the catalogues of Vienna and London have not been tested for accuracy since there is no dataset of higher quality available for comparison.

Resolution

The datasets that are offered for free by the cities of Vienna and London are both based on the polygons that are also used for publication of other statistical information. These are the counting districts in Vienna and the boroughs in London. The attributal resolution offered by Vienna is higher than the one of the London data. In both datasets the attribute information is given as a percentage per borough. But the number of different attributes is higher in Vienna.

Completeness

The datasets published by the governments of Vienna and London where tested for attribute completeness. There are no missing values in the tables of the Viennese datasets and the London dataset on net additional dwellings per borough. The table on dwelling stock by tenure and condition for London borough contains empty fields. There is no information in the metadata or other description why these values are not available. Using this information one must therefore keep in mind, that the data is incomplete. Since fields where changed with the datasets of 2005 and again 2009, a database that holds information from all then years will contain fields with missing values for several years. Furthermore, this excel file only contains data for the period of 2001 to 2011. More recent information is not available.

Consistency

The file on dwelling stock by tenure and condition for London boroughs misses several values. Therefore the attributes in this dataset are not consistent either. Besides that, attribute consistency was tested in the Viennese datasets by checking the totals against the sums of different categories. No errors could be found.

Lineage

The dataset offered by the city of Vienna is published by a department within the city government that is responsible for city planning called Magistratsabteilung 23 –Wirtschaft, Arbeit und Statistik. The data was generated from counts in the governmental registers. A link that is provided for getting more detailed information on the counting methods leads to the homepage of Statistik Austria. One can therefore assume that the data is the same or at least from the same source as the products offered by Statistik Austria on this topic. The data on dwelling stock by tenure and condition in London was created by a department of the Office for National Statistics (ONS) called Neighbourhood Statistics. Detailed information on how the data was created is not available. The dataset on net additional dwellings was created by measuring the absolute increase in stock between one year and the next, including other losses and gains (such as conversions, changes of use and demolitions). The local authorities delivered this information via forms and annual reports. The data for London was delivered by the Greater London Authority. Only the figures from 2001 and 2011 are census figures, the other values are estimates and number have been rounded to the nearest then. The data published in either of the two city data catalogues refers to sources that are governmental institutions and can therefore be trusted to publish information that is reliable.

Interfaces

The data from Vienna is published as comma separated values in text file. The files contain two describing header rows that need to be removed. Then the data can easily be imported into a geo-database.

The data offered in the London data catalogues is formatted in Microsoft Excel files that are designed to be read by end users and not by applications. The data needs to be prepared and formatted so that it can be imported into a database. The file on dwelling stock by tenure and condition is provided as a Microsoft Excel file with separate data sheets for each year that is contained in the data. The field header is descriptive, therefore long and contains special characters. It needs to be replaced before the data can be imported to a database or GIS. Additionally, the number of fields is not the same in every year. In 2005 some of the fields where replaced with other information and from 2009 on these changed fields where omitted. Using this data one needs to keep in mind, that a database schema for the whole time period of available information must be larger than one single data sheet. Furthermore, the codes that are used to identify the single boroughs do not match the current coding scheme that is used with the polygons that are available in the London data store. Therefore the data cannot be linked to the polygons or to other data. The coding scheme was changed in 2011. The data in the table also ranges up to 2011 and due to the fact that the old coding scheme is used, an update cannot be expected.

The dataset on net additional dwellings per London borough is provided as a Microsoft Excel file containing one sheet per topic. These are net additions, all dwellings and persons per dwelling. The sheets have additional header rows that need to be removed, before importing the data to a database or GIS. The header names also need to be adjusted, since they contain special characters. The sheets contain one column per year. Updates will therefore likely be done, by adding a column, which needs to be taken into account, if updates shall be imported into an application. This dataset contains both an old and a new ONS Code. The table could therefore be used to translate data that is coded to the old coding scheme in order to link it to other datasets.

License

Vienna publishes the data under the licence terms called “Creative Commons Namensnennung 3.0 Österreich” This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes. The data offered by the city of London is published under Crown Copyright.

5.5.4 Usability of Building Data

The rather coarse spatial resolution of information on buildings and dwellings in London narrows the usability of this data. Also the fact, that the information is not up to date lowers the value of the data for most types of analysis. If the data shall for example be used to identify target groups for certain products, the calculation of a catchment area will require more detailed spatial granularity for the input data. But the example used in this thesis with the sample bakery will not require information on the detailed information on the standard of dwellings in its catchment area. For planning a communication infrastructure, the information on buildings, their age and estimates for upcoming renovations can be very valuable. If this information shall be used in an analysis, the spatial granularity of boroughs will only be suitable for very rough overviews. One will for sure want to look at this information in more detail. The Viennese counting districts will serve better in this context. But for detailed analysis this information will be needed per building. And the planning will in the end even require to go there in person and to look at certain locations in reality.

5.6 Road Network

For the extent of Austria a road network including travel time information is offered by two commercial data producers. TomTom, the former Tele Atlas as well as NOKIA, who took over the company Navteq offer a database that is permanently updated with changes that occur in the real road network. Among the most important customers for these two companies are the providers of navigation tools like the ones implemented in cars. Therefore the information on restrictions within the road network that are applying to cars and trucks are carefully updated on a regular basis. Nevertheless, the data can also be used for the purpose of routing bicycles and pedestrians by allowing them to pass through pedestrian zones and using pedestrian pathways if applicable.

As the webpage of TomTom explains there is a product available called MultiNet[47]. It provides accurate mapping data in combination with turn by turn navigation information. The data is updated permanently either by employees or by customers who can report changes or errors they observed via a web interface.

[47] www.tomtom.com last access 2014-10

The main competitor in the area of road network information that is commercially generated and distributed is the company Navteq. This company was bought by NOKIA in 2007 and the portfolio was integrated into a product stream called HERE.[48] As it is stated on the webpage, HERE offers data that is relevant for navigation of vehicles and pedestrians. The database is updated permanently by employees and customers can report changes or errors via a web interface. Both HERE and TomTom claim to provide the best map data in the world.

5.6.1 Road Networks of Vienna and London

The Viennese data catalogue does not contain a street network. But it offers information on the network of public transport. Both the lines and the points of stations are provided.[49]The catalogue of London does not offer a road network either. But it also contains information on the Transport for London (TfL) Network including Stations for public transport. The catalogue contains a reference to the TfL Website where life feeds for public transport information are made available after registration.[50]

5.6.2 Alternative Road Networks

For an analysis based on open data, alternative road network data can be considered, like the Open StreetMap data. Open StreetMap Austria is a registered association that intends to support the Open StreetMap concept in Austria and wants to represent the impersonated representation of this global idea in this country (www.openstreetmap.at). It acts as contact organisation for governments, sponsors, companies and individuals. The government of the city of Vienna is listed as one of the contributors of map content on the webpage of Open StreetMap.[51] Regarding the United Kingdom, the same webpage lists Ordnance Survey data to be part of the Open StreetMap database.

---

[48]www.here.com last access 2014-10

[49] https://open.wien.at/site/datensatz/?id=36a8b9e9-909e-4605-a7ba-686ee3e1b8bf and https://open.wien.at/site/datensatz/?id=21fca925-12ac-4215-ba1a-a9c73cb3b082 last access 2014-10

[50] http://data.london.gov.uk/datastore/package/tfl-bus-stop-locations-and-routes last access 2014-10

[51]http://www.openstreetmap.org/copyright/en last access 2014-10

5.7 Employers and Employees

The raster based census data from 2001 that is offered by Statistic Austria contains the parameters number of employees and number of workplaces. Alternatively one can use the data that is published by a producer of phonebooks and yellow pages. The company called Herold offers a dataset called Marketing CD Business that includes companies, their addresses and contact information as well as the coordinates associated with the addresses.[52] A part of the datasets listed there is enriched with detailed data on the company like the number of employees. The data is very up to date and updates can be purchased in intervals of less than a year. Unfortunately only about one third of the companies included in the database contains information on the number of employees. Therefore this information is of limited use for an analysis like the ones described in this thesis.

5.7.1 Employers and Employees in Vienna

The data catalogue of Vienna does not contain information on employers. The number of employees is provided in the form of a commuter statistics per district.[53] This table provides information on the number of working people living in a district, as well as the number of people working in a district. The later one is divided into categories representing the distance that people drive to work. The categories are no movement, moving within the district, moving from another district, moving from another Bundesland or from another State.

---

[52] http://www.ichbinderherold.at/kunden-gewinnen/dialog-marketing/die-adresse/business-to-business/marketing-cd-business/ last access 2014-10

[53] https://open.wien.at/site/datensatz/?id=a9934952-82ff-47b3-a5fc-441d77c420e2 2014-10

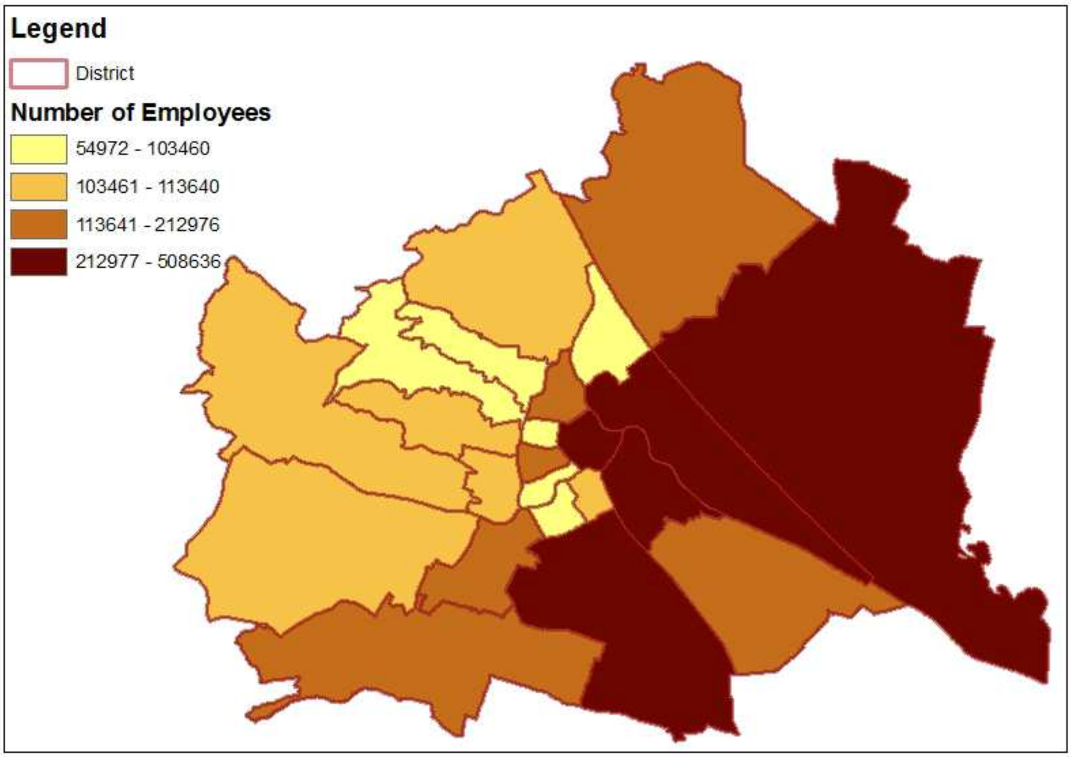


Figure 23:Number of employees per Viennese district

5.7.2 Employers and Employees in London

In the London data store several datasets can be found that contain statistical information on employment and workplaces. All of them are on borough level.[54] The table on Place of Residence by Place of Work contains a cross tabulation showing for the residential population of each London borough, how many of them work in which London borough or more distant local authority. Reading the rows of the cross table, one gets the information, how many workers there are in workplaces within every borough.

[54] http://data.london.gov.uk/datastore/package/place-residence-place-work-local-authority last access 2014-10

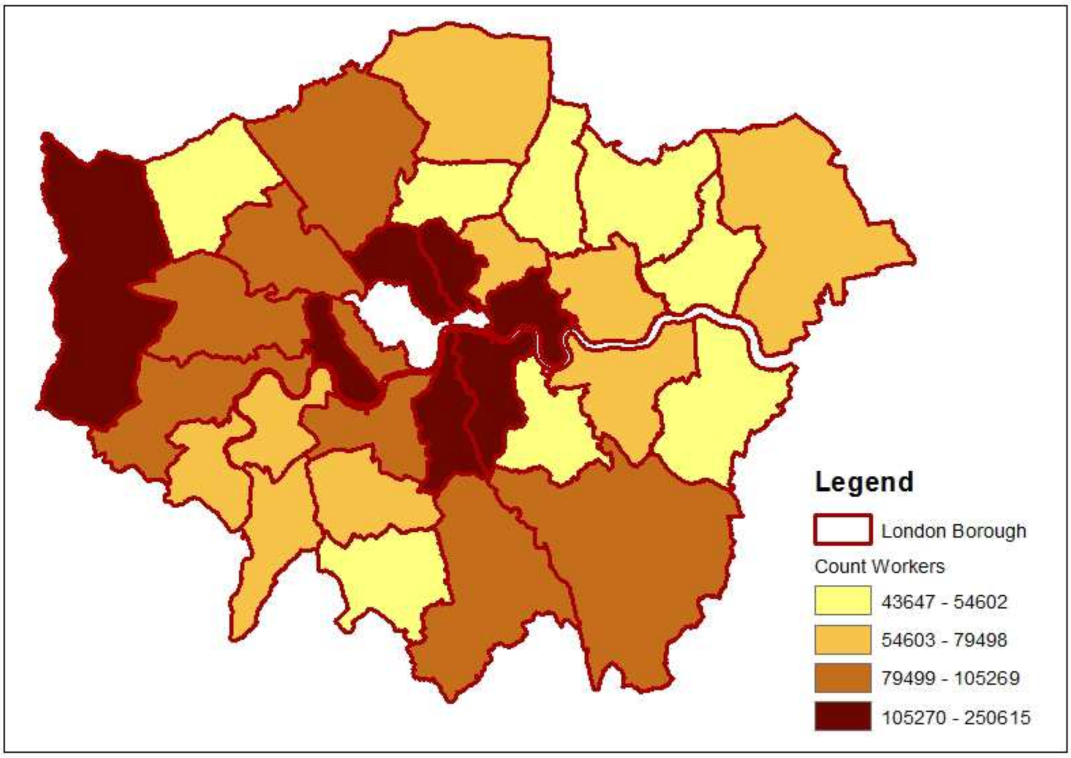


Figure 24:Number of workers per London borough.

Additionally there is a number of sources in the London data store, that provide information on different aspects of employment per borough. For example the daytime population is provided as well as earnings and numbers on employment in public and private sector.[55]

5.7.3 Employers and Employees Evaluation

The data offered by the open data catalogues is derived from statistics on employment and commuters. Detailed information on employers can only be purchased from commercial data providers.

Accuracy

The datasets that can be found in the catalogue of London have not been tested for accuracy since there is not dataset of higher quality available for comparison.

---

[55]http://data.london.gov.uk/datastore/package/workplace-employment-industry-borough
http://data.london.gov.uk/datastore/package/daytime-population-borough
http://data.london.gov.uk/datastore/package/workplace-employment-sex-and-status-borough
http://data.london.gov.uk/datastore/package/earnings-workplace-borough
http://data.london.gov.uk/datastore/package/workplace-employment-publicprivate-sector-borough
all last accesses 2014-10

Resolution

The datasets that are offered for free by the cities of Vienna and London are both based on the polygons that are also used for publication of other statistical information. These are the counting districts in Vienna and the boroughs in London. The attributal resolution offered by the city of London is higher than the information provided by the city of Vienna. London for example provides more detailed information on the source and destination boroughs for commuters. The file on Place of Residence by Place of Work from the London data catalogue contains information from 2011 and will only be updated every ten years. The Viennese dataset contains information from 2009 and 2010. Information on planned updates is not provided.

Completeness

The datasets published by the governments of Vienna and London where tested for attribute completeness. There are no missing values in the tables of the Viennese dataset. The London datasets are missing the values for Westminster and the City of London. There was no hint in the metadata, why the values for these two boroughs are not available.

Consistency

Testing for attribute consistency was not reasonable for these datasets.

Lineage

The dataset offered by the city of Vienna is published by a department within the city government that is responsible for city planning called Magistratsabteilung 23 – Wirtschaft, Arbeit und Statistik. The data was generated from counts in the governmental registers. The Data from London originates from the ONS. Both sources applied a method to protect against disclosure of personal information that includes swapping counts between different geographic areas. Small counts on the lowest geographies might therefore be inconsistent with other sources.

Interfaces

The data from Vienna is published as comma separated values in text file. The files contain two describing header rows that need to be removed. Then the data can easily be imported into a geo-database. The data offered in the London data catalogues is formatted in Microsoft Excel files that are designed to be read by end users and not by applications. The data needs to be prepared and formatted so that it can be imported into a database. The file on Place of Residence by Place of Work contains a cross tabulation including summaries per rows and columns. This file is very rich in content but not only the headers but also the data of this file needs to be edited before it can be imported into a database.

License

Vienna publishes the data under the licence terms called "Creative Commons Namensnennung 3.0 Österreich" This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes. The data offered by the city of London is published under Crown Copyright.

5.7.4 Usability of data on Employers and Employees

The information on companies and the numbers of employees offered by Herold is a valuable dataset. Especially the spatial representation of the data, where the companies are represented by points formed of coordinates is very valuable in the context of planning a mobile communication network. Companies can be located within the serving areas of cells and estimations for the expected traffic and required throughput of data can be made. Although the point usually only represents the entrance to the headquarter building, for small businesses this is usually sufficient for the required estimations. Big companies will likely cover a substantial area with office buildings or production plants. Others will have mobile units or dependencies in other places than the headquarter building. But these will anyway require an individual analysis in terms of possible network coverage. The fact that only a subset of the companies that are contained in the database are attributed with the number of employees is a considerable drawback.

The data offered in the catalogues by Vienna and London is derived from statistics on employment and commuters. The numbers are provided as tables per district or borough. These administrative units are too big to be used for example for a detailed planning of a mobile communication network. But they can give a first impression, in which areas there is potential for high communication traffic during the day time, respectively working hours. Also the potential for attracting business customers could be higher in these areas. For detailed planning of infrastructure and sales forces, more detailed data will be required.

For the purpose of evaluating a shop location, the data available in the open data catalogues can give a first impression on the general areas that would be suitable for a certain type of shop. For the evaluation of catchment areas, more detailed data will likely be required. Unless the goods that shall be sold in this shop will justify driving across some districts or boroughs. For the sample bakery that is cited in this thesis this would not be the case and a detailed analysis of the close surrounding of a potential shop location would require more detailed information like for example the data offered by Herold or like it is part of the raster products of Statistik Austria.

5.8 Commuter Statistics

Information about where people live in relation to where they work is collected by Statistik Austria. A general overview per Bundesland is available online, showing how many citizens do not work in the same community where they live.

Both Vienna and London provide datasets on number of workers per district, respectively borough and categories of the distance to their place of residence.[56] These two files are discussed in the chapter on Employers and Employees above.

5.9 Hot Spots

In the business examples used in this thesis, information on hot spots would be very useful. In the context of this thesis hot spots are places where a big number of people meet regularly. Typically, schools and universities fulfil this criterion and would be of special interest for the sample bakery in this thesis. Concert halls, shopping malls, stadiums or other regularly used event locations can be of interest as well.

In Austria the number of schools per community is available from Statistik Austria on their website. Detailed information about the address and associated location as well as the number of students can be purchased via data brokers. The situation is similar concerning universities. Point of Interest (POI) collections like those of TomTom contain university locations and also some locations of schools. Since the collection and maintenance of POIs are not the core business of a navigation software producer, these lists cannot be trusted to be complete and up to date.

---

[56] Vienna:https://open.wien.at/site/datensatz/?id=a9934952-82ff-47b3-a5fc-441d77c420e2 last access 2014-10
London:http://data.london.gov.uk/datastore/package/place-residence-place-work-local-authority last access 2014-10

5.9.1 Hot Spots in Vienna

The data catalogue of Vienna contains schools as a separate layer or as part of the landuse data (Realnutzungskartierung 2001, 2003, 2005) [57] The universities and universities of applied sciences can be found as a separate dataset as well.[58]

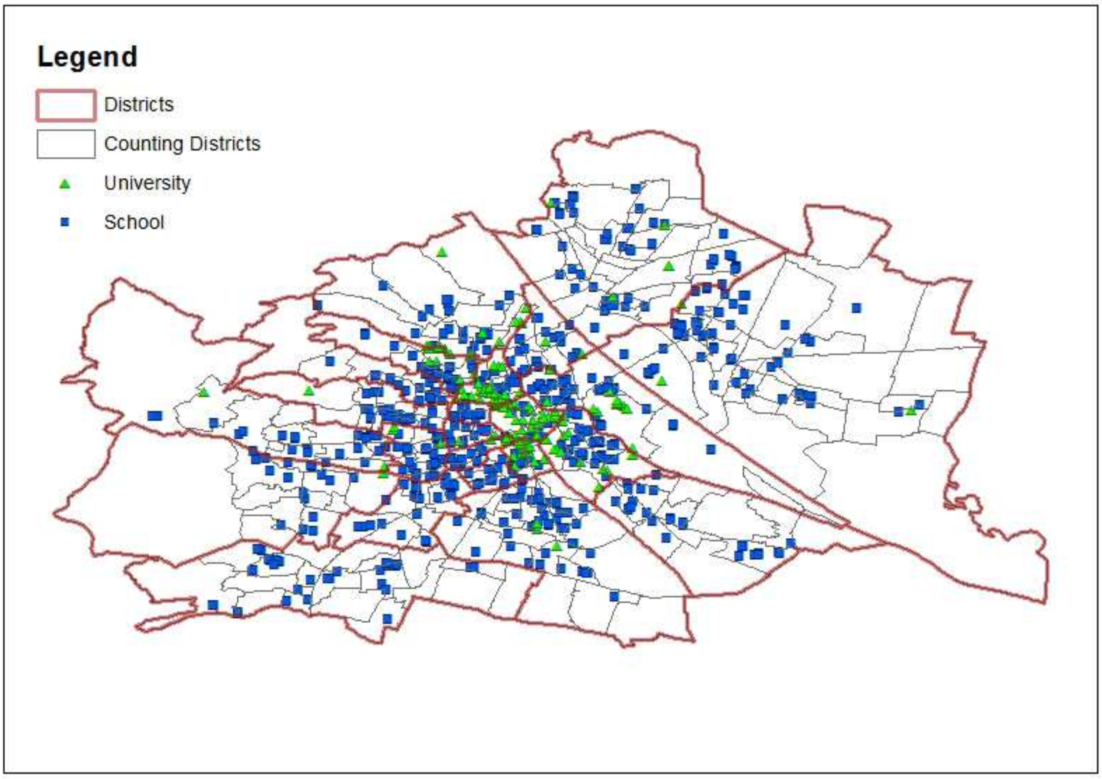


Figure 25: Schools, universities and applied universities in Vienna

Other hot spots like stadiums, concert halls or entertainment centres could not be found in the Viennese data catalogue.

5.9.2 Hot Spots in London

The London data store provides a link to an interactive School Atlas with detailed information on schools and their attendants.[59] The underlying data for this application is also provided in the form of Microsoft Excel workbooks, but the coordinates for the actual locations of the schools could not be found. From tables that are provided in the London

---

[57] https://open.wien.at/site/datensatz/?id=c1ba372b-dba2-4bce-b72e-b5c832eaaf44 last access 2014-10

[58] https://open.wien.at/site/datensatz/?id=3e62e13e-3955-494a-91b3-f00ab9e84b48 last access 2014-10

[59] http://data.london.gov.uk/datastore/package/london-schools-atlas, http://www.london.gov.uk/webmaps/lsa/ last access 2014-10

datastore, it is possible to derive the number of schools and their attendants per borough.[60]

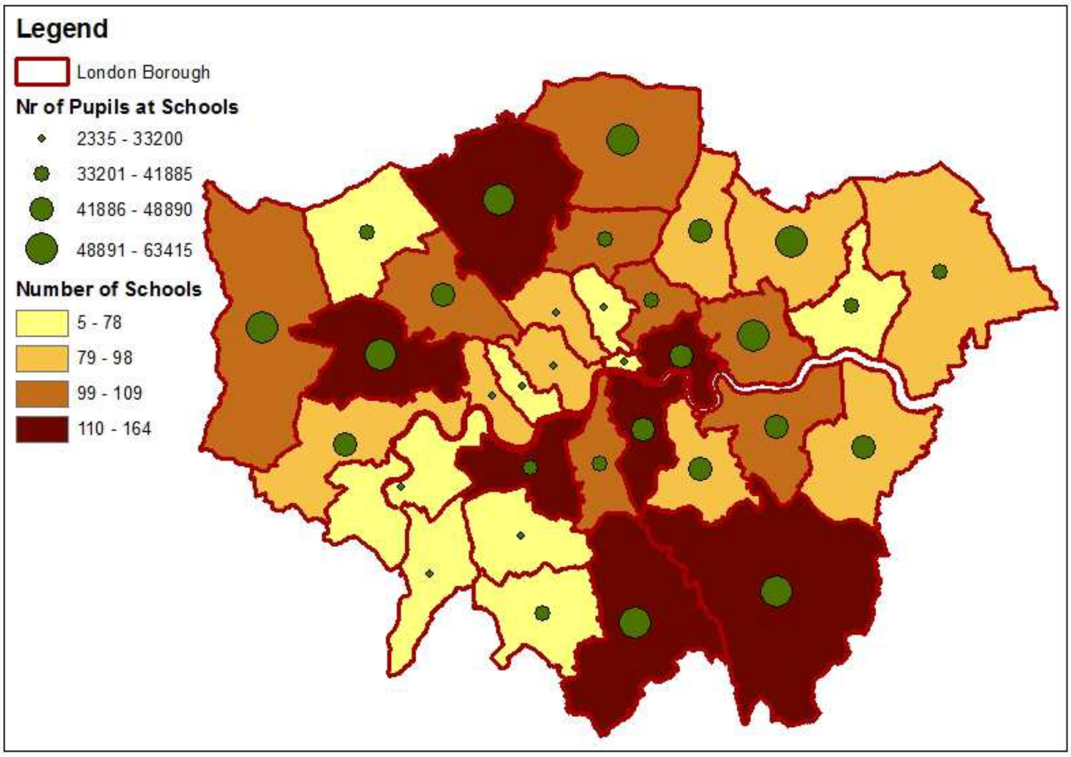


Figure 26: Schools and attendants per London borough.

Other hot spots like stadiums, concert halls or entertainment centres could not be found in the London data catalogue.

5.9.3 Hot Spots Evaluation

The city data catalogues of Vienna and London offer different kind of data on schools and universities. The Viennese catalogue contains them as part of the land use mapping, as well as a separate thematic theme representing the locations of the buildings as points. In this evaluation only the point theme shall be looked at, since the land use layer is discussed in the chapter on land use information. London does not provide the geographic location on the schools as a public dataset. But the data catalogue contains a file with schools and number of pupils per borough.

Accuracy

The datasets that can be found in the catalogues of Vienna and London have not been tested for accuracy since there is not dataset of higher quality available for comparison.

---

[60] http://data.london.gov.uk/datastore/package/schools-and-pupils-type-school-borough last access 2014-10

Resolution

The datasets that are offered for free by the city of Vienna are provided in the form of geometric points with attribute information. There is no information about the spatial accuracy of these points. One can only assume, that dues to the fact the data was published by a department of the Viennese government, it could be based on the building information that is available in the Viennese city model. The resolution of the data responds to the number of buildings identified. The data that is published in the London data catalogue is provided in the form of attribute tables per borough. The spatial resolution is therefore rather coarse. The attributal resolution is in contrast rather high, since the table contains additional information like the number of pupils and different categories for pupils and types of schools. The data is available in separate data sheets for the years 2011 to 1014.

Completeness

The datasets published by the governments of Vienna and London where tested for attribute completeness. There are no missing values in the tables of the dataset.

Consistency

Checking for attribute consistency is not reasonable in these datasets, since the attributes contain descriptive information and single value fields only.

Lineage

The dataset offered by the city of Vienna is published by a department within the city government that is responsible for the schools of Vienna, called Stadtschulrat für Wien, MA 56 – Wiener Schulen, respectively the universities of Vienna themselves. The Data from London originates from the Department for Education and is categorized as an official statistics.

Interfaces

The data from Vienna is published in the form GIS readable datasets like ESRI shape files and in a number of different formats that can also be used for life integration into an application like WMS and WFS. The files can easily be imported into a geo-database. The data offered in the London data catalogues is formatted in Microsoft Excel files that are designed to be read by end users and not by applications. The data needs to be prepared and formatted so that it can be imported into a database.

License

Vienna publishes the data under the licence terms called "Creative Commons Namensnennung 3.0 Österreich" This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes. The data offered by the city of London is published under Crown Copyright.

5.9.4 Usability of Data on Hot Spots

The points for Schools and Universities that are offered in the Viennese data catalogue can easily be integrated into an application. But there is no additional information available like the number of pupils attending the school. The Data published in the London data store on the other hand, contains lots of additional information on pupils and categories of schools, but there is no location information available for the single school buildings. For applications like the analysis described in this thesis, a combination of both types of information would be necessary in order to produce valuable results. Knowing the exact location of a hot spot, but not having any information on its capacity, the potentials arising from this hot spot cannot be evaluated. Knowing the capacity of hot spots in the area of one borough on the other hand does not provide enough spatial detail in order to evaluate, if the capacity is going to be relevant for a certain location. This is valid for both business scenarios that are used as a reference in this thesis. The data on schools and universities can therefore be used to get a first impression, on where the points are, respectively in which general area there are potentials. But for a detailed analysis the missing details in resolution will be necessary. Additionally one would likely be interested in information on other types of hot spots that were not contained in the open data catalogues at all. Stadiums and regularly used concert halls, or highly frequented malls might be relevant for a business as well.

5.10 Footfall, Drop In, Shop Mix

Footfall data describes the frequency of passengers in a certain location. This is usually of interest for the area in front of a shop and describes how many people pass by the store location. The drop in represents the number of persons that actually enter the shop. And the companies and brands that are represented in the shops close to the location in question are described by the shop mix. There are some consulting companies in Austria that offer to do this kind of data collection and will then advise to go for certain locations if they promise to provide the required number of visitors from certain target groups. Footfall data and characteristics of shopping streets, shopping centres, towns and city centres, like for example a shop mix, are likely not available for purchase “off the shelf”.

Neither London nor Vienna does offer any information on shop mix or footfall data of shopping destinations in the city. The only related dataset that could be found is on markets in Vienna. (Märkte in Wien)

5.11 Rental rates

Information on rental rates is collected by various institutions. Real-estate agents can easily collect information from the offers they put together for their clients. Web portals that present real estate offers can easily collect the price information and produce statistics from the data. One example that can be found in Austria is the Immopreisatlas of an Austrian bank.[61] It shows the average rent or price per square meter for purchasing an apartment or house on district level. Various consulting companies are very well informed about the rental rates for shop locations in various regions and individual destinations. They gather this information through their activity as consultants for companies within different branches on the market. This way they collect samples that may be representative for many shop locations in question by new customer companies. A standardised product could not be found for purchase in Austria that would contain reliable information on the rental rates for businesses or for residential use. The governments could analyse the prices from real estate transactions that are documented in the land register. A product that references to that source could not be found in Austria.

5.11.1 Rental Rates Vienna

The city of Vienna does not publish data on rental rates of real estate in the city as part of its data catalogue.

5.11.2 Rental Rates London

For the city of London there is a quarterly updated excel sheet available, containing the average private rental rates observed within the past twelve months.[62] This data is published and updated by the organization called Opinion Research and General Statistics (GLA). It contains the count of rents, average, lower quartile, median and upper quartile for the 12 months preceding every quartile of a year.

[61] http://www.immopreisatlas.at/ last access 2014-11

[62] http://data.london.gov.uk/datastore/package/average-private-rents-borough last access 2014-10

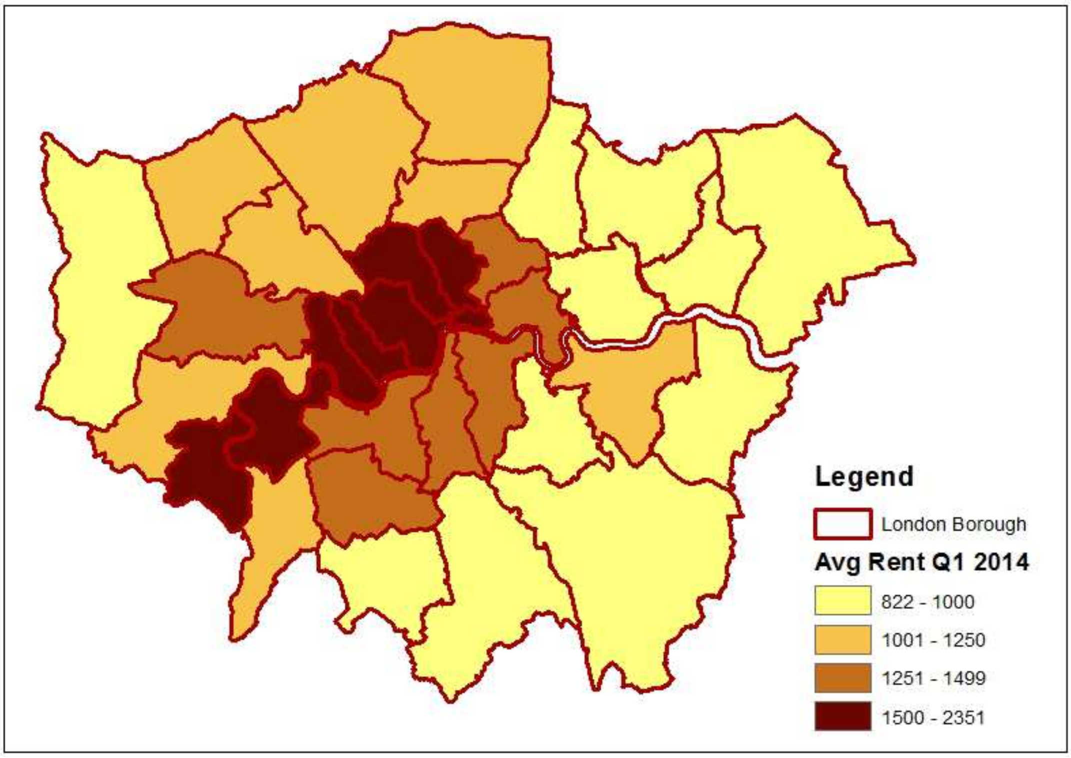


Figure 27: Average rental rates per London Borough in Q1 2014.

Additionally there is a dataset available that contains average house prices gained from sales transactions in the reported time.[63]Information on changes of ownership of dwellings is also provided, but only for the period of 2001 to 2009.[64]London also offers information that is derived from the land registry data.[65] The description of this dataset mentions that the data might not be identical with official statistics on borough level since the data is directly derived from land registry information. And caution shall be used when using the data for small scale analysis.

5.11.3 Rental Rates Evaluation

Vienna does not offer any open data on rental rates or real estate prices, neither for business objects nor for privately used dwellings. The London data store contains several files on rental rates and house prices. Although only one directly contains information on rental rates, the other files on house prices could possibly be used to derive estimations for rental rates that come with certain house prices. But also the London data is either

[63]http://data.london.gov.uk/datastore/package/average-house-prices-borough last access 2014-10

[64] http://data.london.gov.uk/datastore/package/changes-ownership-dwelling-price-borough last access 2014-10

[65] http://data.london.gov.uk/datastore/package/average-house-prices-ward-lsoa-msoa last access 2014-10

limited to privately used objects or the objects are not classified or filtered at all, like the dataset based on Land Registry data.

Accuracy

The datasets that can be found in the catalogue of London have not been tested for accuracy since there is not dataset of higher quality available for comparison.

Resolution

The datasets that are offered for free by the city of London are mainly based on the polygons of London boroughs. So is the most interesting dataset on rental rates.

Completeness

The dataset on rental rates that is published by the government of London was tested for attribute completeness. There are no missing values in the tables.

Consistency

Testing for attribute consistency is not reasonable in this kind of dataset.

Lineage

The dataset offered by the city of London is published by the organization called Opinion Research and General Statistics. The metadata mentions and organization called Valuation Office Agency (VOA) as the creator of the data. Detailed release notes are available online.[66] It states that only cases where there was evidence for a transaction, like the payment of a rent could be found. Cases where there was limited or no evidence where dropped. The release note recommends using the data with caution and as an indication only. And it reminds to be aware that it is not accurately representative for the whole population due to the selection process above.

Interfaces

The data offered in the London data catalogues is formatted in Microsoft Excel files that are designed to be read by end users and not by applications. The data needs to be prepared and formatted so that it can be imported into a database. The file on rental rates per borough contains a sheet that shows the trend of average rental rates by comparing the quarterly numbers. This sheet can quite easily be reformatted and imported to a database. Only the field headers need to be renamed.

---

[66] http://www.voa.gov.uk/corporate/statisticalReleases/120823_PRRM_ReleaseNotes.html#Methodology last access 2014-10

License

The data offered by the city of London is published under Crown Copyright.

5.11.4 Usability of data on Rental Rates

The city of Vienna does not offer any information on rental rates or real estate prices in its open data catalogue. The data offered by the city of London can serve as an overview on the regions where there rents and prices for real estate are high and what are the proportions. The data also offers a trend for the period of some years. In combination with knowledge on the investments or mayor events in different boroughs, this can be used for speculations on the further development of the regions. For the detailed planning of a shop location one will, however, need more detailed information on the prices per square meter that are common in single shopping streets, on certain squares or in shopping malls. Even there, there will be reasonable differences between single locations in close proximity, but typically the characteristics of a location will be reflected in the general price level there. And this kind of information can only be transported in data with higher resolution than borough level, or by an expert, who gathers knowledge and experience by consulting a great number of companies.

5.12 Spending Capacity

Spending capacity is defined as the income of a person or household that is available for the purpose of consuming. Therefore it is calculated by summing up all types of income and reducing regular spending like rents, insurance, loans, etc. Information on spending capacity in Austria is calculated and sold by a company called GfK. Although GfK uses a slightly different definition where they only sum up the income and do not subtract regular spending. The data is usually presented on district level. But it is at least partly also available in more detailed resolution like zip code areas.[67]This is not detailed enough to be used for the evaluation of possible shop locations that lie within close proximity. But the data can be used for the estimation of possible turnover of a new outlet if the catchment area can be compared to similar ones in other parts of the country.

5.12.1 Spending Capacity Vienna

The data catalogue of Vienna does not contain any information on spending capacity.

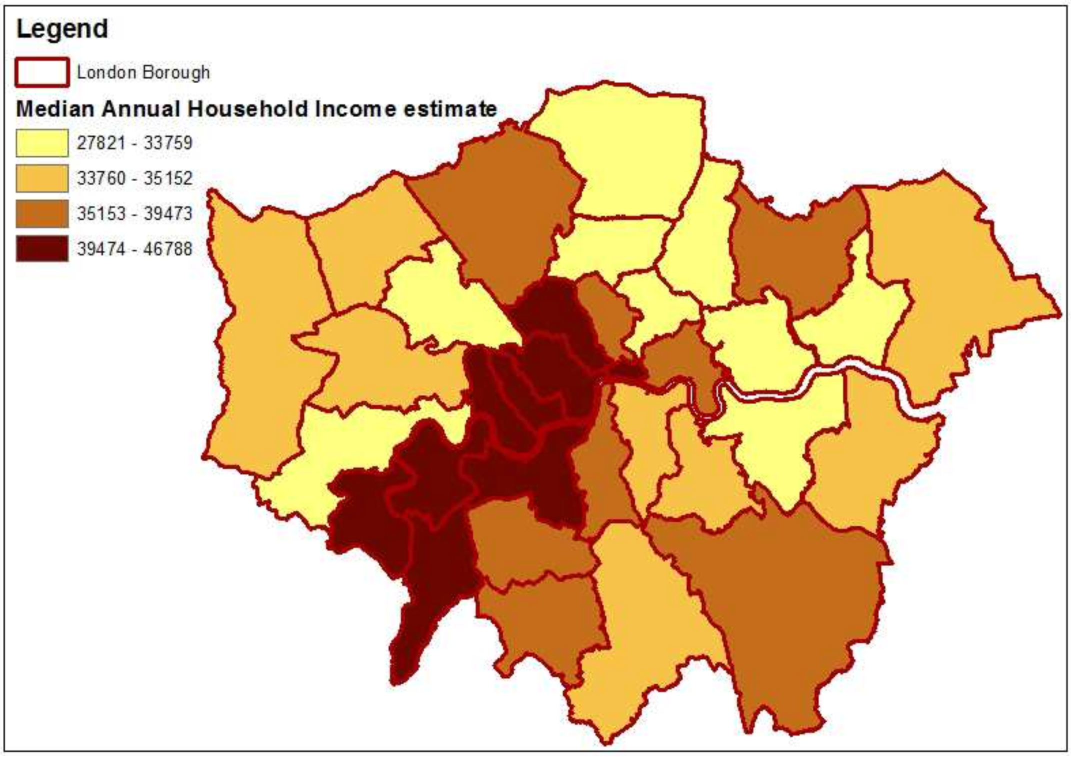
---

[67] http://www.gfk.com/de/documents/pressemitteilungen/2014/20140402_pm_gfk-kaufkraft-dach_fin.pdf last access 2014-10

5.12.2 Spending Capacity London

The London Data Store provides information on people's income per borough.[68] This dataset contains figures for peoples income per hour, week, full time and part time and distinguishes between male and female in some categories. But the referencing codes in this dataset correspond to the old coding scheme of boroughs and can therefore not be linked to the polygons that are provided in the London data store.

Another dataset contains information on household income in small areas.[69] This dataset contains estimates for the total mean and median annual household income. These two values are not only available per borough but also per ward. The data was published by the Greater London Authority and dates from 2011. The metadata describes the methods that were used for the estimation and invites for feedback in order to improve the model. The data is in a testing phase.



[68]http://data.london.gov.uk/datastore/package/earnings-place-residence-borough
last access 2014-10

[69]http://data.london.gov.uk/datastore/package/household-income-estimates-small-areas
last access 2014

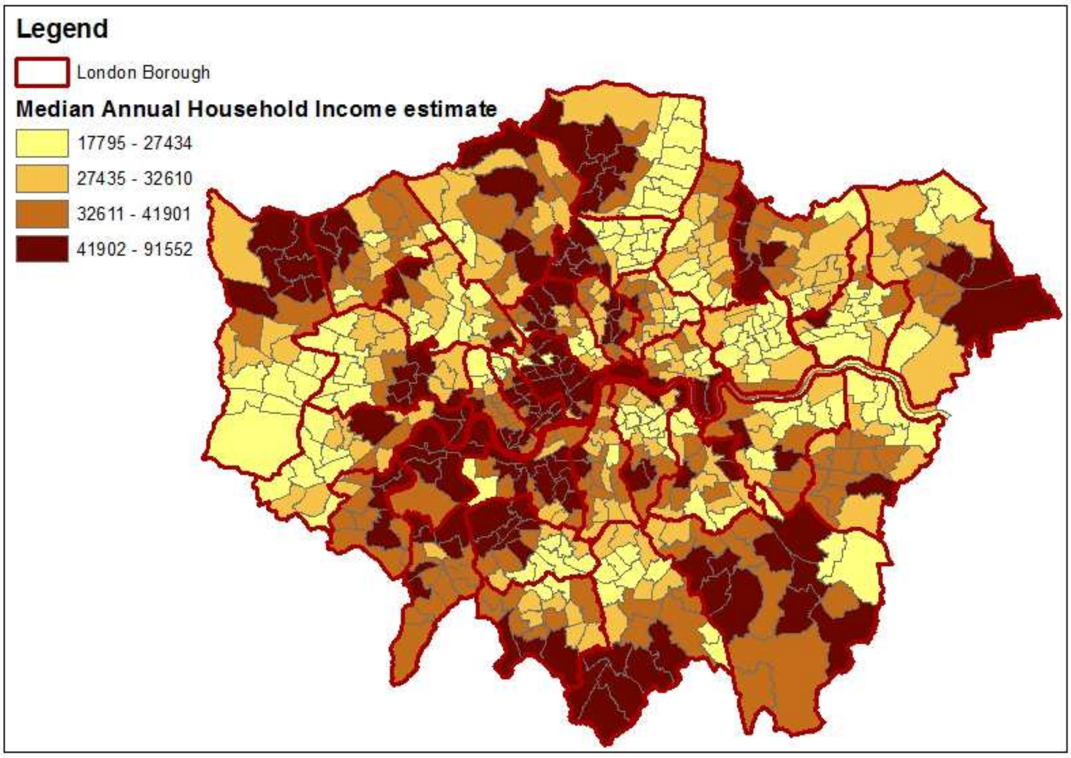


Figure 28: Median annual household income in London on ward level and on borough level.

5.12.3 Spending Capacity Evaluation

Vienna does not offer any open data on spending capacity or income. The London data store contains two files with income data. One of them can actually be linked to the polygons available in the data store. Although the income per household that is available in the London data store is not the same as spending capacity, it represents a major factor. In order to get estimates for the spending capacity one would need to get information on other forms of income like social payments or others. Regular spending are also omitted in the data from GfK. Therefore the absence of this information in the London data is in fact positive for the comparison.

Accuracy

The datasets that can be found in the catalogue of London have not been tested for accuracy since there is not dataset of higher quality available for comparison.

Resolution

The datasets that are offered for free by the city of London are based on the polygons of London wards and boroughs. The resolution of commercial data is about the same.

Completeness

The dataset on income that is published by the government of London was tested for attribute completeness. There are no missing values in the tables.

Consistency

Testing for attribute consistency is not reasonable in this kind on dataset

Lineage

The dataset offered by the city of London is published by the Intelligence Unit of the Greater London Authority. The estimates are based on data that was collected in 2011 and 2012. Different sources where combined to get estimates for all areas and a high resolution. Detailed information on this process in published on the website of the GLA.[70] So far, no update has been announced for the data.

Interfaces

The data offered in the London data catalogues is formatted in Microsoft Excel files that are designed to be read by end users and not by applications. The data needs to be prepared and formatted so that it can be imported into a database. The file contains separate sheets for the data per ward and per borough. These can quite easily be reformatted and imported to a database. Only the field headers need to be renamed.

License

The data offered by the city of London is published under Crown Copyright.

5.12.4 Usability of Data on Spending Capacity.

The data on household income that is offered by the city of London is not exactly the spending capacity according to its definition. It is therefore not possible to use this data to actually calculate the expected spending in a certain segment of consummation. But it could be used as an estimate in the context of store location planning. If the sample bakery has a store network and there are locations with comparable income levels in the catchment area, these can be used as a reference for the new store location. Regarding the resolution, it is of course preferable to get the information as detailed as possible. And compared to commercial products, the dataset offered by the city of London is equally useful. Two drawbacks could be identified with the London data. One is the experimental calculation of the data. In order to have a reliable dataset it would be necessary to verify

[70] http://data.london.gov.uk/documents/update-07-2014-methodology.pdf last access 2014-10

the results of the model. No information on such verifications or any other feedback could be found. And the second drawback is the age of the data (2011/2012) and the fact that there is no update announced.

5.13 Competitors' locations

The location of competitors is a dataset that will vary greatly according to the field of business and question at hand. Depending on the requirements, it is likely that there is no dataset available that can be used off the shelf. But by combining information from standard data products with individual inquiry will likely provide a dataset that can be used for an analysis. For example the competitors for a bakery can be found partially in the company data from Herold, while the mass merchandisers and larger chains of bakeries usually present a list of stores including addresses on their homepage. Google maps can be used for a manual search of potential competitors in the area. And to make sure, every competitor location has been identified and is still a valid feature in the dataset, a physical survey in the area is the only truly reliable source of information.

The data catalogues of Vienna and London do not provide any data on supermarkets or other shops. The only shopping related data set that can be found there locates the 17 markets of Vienna where producers offer their goods. Additionally some information on Christmas markets in provided, which has only seasonal relevance for shop locations.[71]

5.14 Customers

The number and residence of actual customers of a company is valuable information within an individual company. Most likely the owner of this information will not give it away as it represents a commercial advantage to know as much as possible about actual or potential customers. Furthermore, the processing and analysis of personal data is restricted by legal regulations ensuring the protection of the individual's privacy. Therefore customer data is usually only available within a company. Certain habits and characteristics can be derived from data that is traded by address brokers. They collect this information through questionnaires where the questioned person explicitly allows the

[71] Wiener Märkte: https://open.wien.at/site/datensatz/?id=aab0a0be-89e6-442e-8cf9-e762bd18f76b Weihnachtsmärkte: https://open.wien.at/site/datensatz/?id=16c7fa5b-e2ac-4558-b76c-506766d3949d both last accessed 2014-10

commercial use of the given information. The open government catalogues investigated in this thesis do not offer detailed information on individual persons.

6 Summary and Conclusion

As suspected, the data from the open data catalogues can be used for some purpose and mainly as a means to get an overview or first impression. If one wants to do detailed analysis, purchasing datasets will still be necessary.

The data catalogues show a principal difference. In London datasets are online from different sources and of different quality and age. One can find amazing details in the statistics on borough level. But in the mass of data it is sometimes difficult to find the right dataset. Partly there are experimental datasets still online that have not been updated for years. The older datasets use the old coding scheme for administrative units that has been replaced in 2011. If there has not been an update to the data since then, there is a high probability to find this old coding scheme. But the polygons that are online as boundary dataset only contain the new coding scheme. And there was no official translation table to be found. Only by chance, during the investigations for this thesis there was one dataset found, that contained both coding schemes.

The data that is published in the London Data store is formatted in a way that makes the data easy to read for humans. The data is mainly provided in the form of tabular data in Microsoft Excel format with several worksheets. The field headers are always descriptive and cannot be used in a database. The Viennese data catalogue provides information in a form that is more easily readable for applications than for humans. Tables are usually formatted as comma separated files with headers that can be used in a database. Series of data from different categories or years are combined into longer tables where codes distinguish the categories.In the London data very often large, respectively wide tables with many columns are produced, showing all categories in parallel. The Viennese datasets are often provided as Web Map Service or Map Feature Service. These can directly be integrated into an application and will always gather the most up to date dataset from the data catalogue.

To be valuable for the analysis of a shop location for the sample bakery the data needs to be available in small granularity. Information on district or borough level can be used to get an overall impression of the situation in the area. Since the catchment area of a bakery usually does not cover several districts of boroughs, the information on this level cannot be integrated in models based on this technique.

For the success of OGD, the establishment of certain standards seems highly beneficial. One of the standardizing developments is the enrichment of open data catalogues with the necessary metadata information in order to realise the concept of Linked Open Data. This way, the published data does not need to be harmonized between different sources, but the RDF definitions still enable the combination of data from different origin and of different format. Regarding the establishment of standards, Shapiro and Varian (1998) suggest to consider different strategies. Being part of a standardization group requires careful differentiation from the others in order to be recognized by the customers as an individual brand and product. Competition of very similar products will be high in this case. Not participating in a standard can however make it difficult to reach a customer base of sufficient size to return the revenues needed to cover production costs and gain profit. In the case of OGD, being part of a standardization group seems to be beneficial, at least if the effort that comes with the task is ignored. Competition between similar products will be less important than with commercial products, if the data is easy to access and the quality of the data and its description is high. Reaching a big customer base is not crucial for the return of revenues, since the data is given away for free. But it must be in the interest of governments to establish their data as the standard data source if participation and collaboration of stakeholders shall be achieved, as it is stated in the OGD concept for Vienna. (Krabina, et al., 2012)

The city of Vienna seems to consider carefully, which datasets to publish. This is also described in the OGD Implementation Model. (Krabina, et al., 2012) The published datasets are limited in their number, compared to the London data store. But the quality of the data is very high, in terms of up-to-dateness and interfaces for implementation into applications. But in many cases, there was a commercial product available, sometimes from the same department of the city government, which provides higher resolution. One very obvious example is the digital terrain model, where the lineage information can be looked up at the product homepage of the commercial offer. Other datasets, like for example the population data, information on buildings and the workforce, are offered with higher resolution by the national statistics agency, which is called Statistik Austria. In all these cases the required accuracy for the task at hand will determine, if the free data can be used. The London data store contains a vast amount of information. Especially the statistics on borough level seem to cover almost any aspect of city government and planning. But the resolution of these datasets is not high enough for the type of analysis that was investigated in this thesis. There is likely more detailed information available from the originating departments, if one asks there directly. A drawback that comes with the amount of data is the number of old or outdated datasets that can be found in the London catalogue. If a dataset is now ten years old and still in an experimental status, like the data on land use per ward, one should keep this in mind, if using this data for an analysis. The fact that these older datasets still contain a reference system that is no longer published among the boundary datasets, reasonably limits their usability as well.

Open Government Data like it is published in the data catalogues of Vienna and London can be used to get an overview on the situation in the two cities. Some of the datasets are suitable for detailed analysis as well, like for example the land use data in Vienna or the data on income in London that can replace the commercial dataset on spending capacity. For a detailed analysis, most of the published datasets are either not detailed enough in terms of spatial resolution, do not contain required attributes or are simply outdated.

All the data that was found in the Viennese data catalogue was presented under licence terms called “Creative Commons Namensnennung 3.0 Österreich” This requires mentioning the source when presenting the data or products using this data. But there are no limitations for using the data for private or commercial purposes. Most datasets were explicitly produced and owned by the Viennese government. Only the information on dwellings references to the national department of statistics. The data that was found in the London data store originated from different sources. Most of them explicitly comply to the Open Government Data criteria. But there are different requirements of how to refer to the source, when the data is used. The data in the London borough profiles that contains among others information on population and additional characteristics is not attached with an overall licensing statement. The data was collected and combined from different sources. But one can assume that all the data that is published in the London data store actually is open data.

Both strategies for publishing Open Government Data have their positive and negative aspects. In order to do a detailed analysis like the ones described in this thesis, neither of the data catalogues provides sufficient information, to completely replace commercial data products. The two data catalogues do not only show different approaches for selecting and publishing datasets, but they also have different types of datasets available. There is for example detailed information on terrain and land use available in Vienna. Even the individual buildings are shown in the online interactive city map. But there is no information on income levels, rental rates and house prices available. London, on the other hand, offers these statistics with a great number of details in the attributes, at least on borough level. Information on land use or terrain cannot be found in the London data store.

Both data catalogues offer information that would be valuable for identifying the topics, where more detailed data is required.

List of Abbreviations

ALS Airborne Laser Scanning

BEV Bundesamt für Eich- und Vermessungswesen

BFC Boundary - Full resolution, Clipped to the coastline

BFE Boundary - Full resolution, Extent of the realm

BGC Boundary - Generalised Clipped to the coastline

DCLG Department for Communities and Local Government

DSM Digital Surface Model

DTM Digital Terrain Model

EDGE Enhanced Data rates for GSM Evolution

EFS Expenditure and Food Survey

GIS Geographic Information System

GSM Global System for Mobile Communications

HTTP Hypertext Transfer Protocol

ICA International Cartographic Association

INSPIRE Infrastructure for Spatial Information in the European Community

LFEPA London Fire and Emergency Planning Authority

LIDAR Light Detection and Ranging

LOD Linked Open Data

LSOA Lower Super Output Area

MSOA Middle Super Output Area

MZK Mehrzweckstadtkarte

OA Output Area

OGD Open Government Data

ONS Office for National Statistics

POI Point of Interest

PSI Public Sector Information

RDF Resource Description Framework

SDTS Spatial Data Transfer Standard

TfL Transport for London

UK United Kingdom

UMTS Universal Telecommunications System

URI Uniform Resource Identifier

USA United States on America

List of Figures

Figure 1: The Value Chain Paradox (Krek, et al., 2000 p. 5) 14
Figure 2: Relationship between data completeness, model completeness and fitness of use (Brassel, et al., 1995 p. 88) 20
Figure 3: Evaluation matrix (Veregin, et al., 1995 p. 168)....................23
Figure 4: Open Government Implementation Model (Krabina, et al., 2012 p. 8)....................26
Figure 5: Business Model Canvas ("Business Model Canvas" by Business Model Alchemist)....................30
Figure 6: Detail Customer of the Business Model Canvas ("Business Model Canvas" by Business Model Alchemist)....................31
Figure 7: Reflection on rough surface a) real situation, b) ideal model (Benkner, 2007 pp. 28, translated)....................46
Figure 8: Incoming angle and average deviation in height (Benkner, 2007 p. 29) 47
Figure 9: Fresnel Ellipsoid (Benkner, 2007 p. 34) 47
Figure 10: The steps for designing a mobile communication network (Benkner, 2007 pp. 149, translated)....................50
Figure 11: Evaluation matrix (Veregin, et al., 1995 p. 168)....................53
Figure 12: Overview Vienna districts and counting districts....................56
Figure 13: Overview London boroughs and wards 57
Figure 14: Vienna: population density in counting districts as of January 2013. 62
Figure 15: Share of people with Austrian nationality in Viennese counting districts as of January 2013....................62
Figure 16: Share of people born in Austria in Viennese counting districts as of January 2013 63
Figure 17: Digital Terrain Model Vienna....................69
Figure 18: Land use in Vienna (Realnutzungskartierung)....................72
Figure 19: London overview and green belt area 74
Figure 20: Apartments in Vienna in residential buildings (per counting district) 78
Figure 21: Apartments in Vienna that are not connected to water supply (per counting district)....................78
Figure 22: Dwellings per borough in London. 80
Figure 23: Number of employees per Viennese district 86
Figure 24: Number of workers per London borough. 87
Figure 25: Schools, universities and applied universities in Vienna....................91

Figure 26: Schools and attendants per London borough. 92
Figure 27: Average rental rates per London Borough in Q1 2014. 96
Figure 28: Median annual household income in London on ward level and on borough level. 100

Bibliography

Arentze, T, Borgers, A and Timmermans, H. 2000. A knowledge-based system for developing retail location strategies. [ed.] Pergamon. *Computers, Environments and Urban Systems.* 2000, 24, pp. 489-508.

Benkner, Thorsten. 2007. *Grundlagen des Mobilfunks.* s.l. : J.Schlembach Fachverlag, 2007. ISBN 978-3-935340-44-1.

Benoit, D and Clarke, G P. 1997. Assessing GIS for retail location planning. [ed.] Pergamon. *Journal of Retailing and Consumer Services.* 1997, Vol. 4, 4, pp. 239-258.

Brassel, Kurt, et al. 1995. Completeness. *Elements of spatial data quality.* Oxford : Elevier Schience Ltd, 1995.

Byrom, J W. 2005. The use of data in outlet locational planning: a preliminary ex amination across retail and service sectors. *Management Reserarch News.* 2005, Vol. 28, 5, pp. 63-74.

Christaller, Walter. 1933. *Die zentralen Orte in Süddeutschland : eine ökonomisch-geographische Untersuchung über die Gesetzmäßigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen.* Darmstadt ; 2006 Sonderausg. der 2., unverändert. Aufl., Repr. Nachdr. der 1. Aufl., Jena, 1933 : Wiss. Buchges., 1933. ISBN 978-3-534-19736-1.

Clarke, Derek G and Clark, David M. 1995. Lineage. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Clarke, Ian and Rowley, Jennifer. 1995. A case for spatial decision-support systems in retail location planning. *International Journal of Retail & Distribution Management.* 1995, 23, pp. 4-10.

DIN. 2005. *Qualitätsmanagementsysteme - Grundlagen und Begriffe (ISO 9000:2005) Dreisprachige Fassung EN ISO 9000:2005.* Berlin : DIN Deutsches Institut für Normung e.V., 2005. DIN EN ISO 9000.

Ding, Li, Peristeras, Vassilios and Hausenblas, Michael. 2012. Linked Open Government Data. *IEEE Intelligent Systems.* 2012, May/June 2012.

Drummond, Jane. 1995. Positional accuracy. [book auth.] Joel L Morrison. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Duque-Antón, Manuel. 2002. *Mobilfunknetze Grundlagen, Dienste und Protokolle.* Braunschweig/Wiesbaden : Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, 2002. ISBN 3-528-03934-5.

Erickson, John S, et al. 2013. Open Government Data: A Data Analytics Approach. [ed.] IEEE. *IEEE Intelligent Systems.* 2013, September/October 2013, pp. 19-23.

Frank, Andrew U, Grum, Eva and Vasseur, Bérengère. 2004. *Procedure to Select the Best Dataset for a Task.* Maryland : Third International Conference on Geographic Information Science GIScience, October 20-23 2004, 2004.

Goodchild, Michael F. 1995. Attribute accuracy. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Gutpill, Stephen C. 1995. Temporal information. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Hernandez, Tony. 2007. Enhancing retail location decision support: The development and applicationof geovizualisation. *Journal of Retailing and Consumer Services.* 2007, 14, pp. 249-258.

Kainz, Wolfgang. 1995. Logical consistency. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Krabina, Bernhard, Prorok, Thomas and Lutz, Brigitte. 2012. *Open Government Implementation Model v2.0.* Vienna : KDZ Centre for Public Administration Research, 2012.

Krek, Alenka and Frank, Andrew U. 2000. The Production of Geographic Information - The Value Tree. *Geo-Informations-Systeme - Journal for Spatial Information and Decision Making.* 2000, 13, pp. 10-12.

Kriesel, Olaf. 2008. Funknetzplanung, Auswahl und Realisierung von Standorten für GSM-/ UMTS-Basisstationen. [Buchverf.] Jürgen Sieck und Michael A Herzog. *Wireless Communication and Information New Technologies and Applications.* Berlin : Verlag Werner Hülsbusch, Boizenburg, 2008, S. 101-108.

Morrison, Joel L. 1995. Spatial data quality. [book auth.] Stephen C Gutpill and Joel L Morrison. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Navratil, Gerhard and Frank, Andrew U. 2005. *Influences affecting data quality.* Beijing : Proceedings of the 4. Internat. Symposium of Spatial Data Quality (ISSDQ 05) Beijing, August 25-26, 2005, 2005.

Open Government Data Group. Open Government Data Principles. [Online] [Cited: 07 03 2014.] https://public.resource.org/8_principles.html.

Osterwalder, Alexander and Pigneur, Yves. 2010. *Business model generation : a handbook for visionaries, game changers, and challengers.* Hoboken, NJ : Wiley, 2010.

Richter, Florian. 2012. *Markt- und Standortanalyse im Handelsimmobilienbereich anhand des G3 Shopping Resort Gerasdorf.* Wien : Diploma Thesis Technical University Vienna, 2012.

Salgé, Francois. 1995. Semantic accuracy. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Shapiro, Carl and Varian, Hal R. 1998. *Information Rules: a strategic guide to network economy.* Boston, Massachusetts : Harvard Business School Press, 1998.

Veregin, Howard and Hargitai, Péter. 1995. An evaluation matrix for geographical data quality. *Elements of spatial data quality.* Oxford : Elsevier Science Ltd, 1995.

Weiss, Christian. 2013. *Transferring Open Government Data into the global Linked Open Data Cloud.* Vienna : Technical University Vienna Diplomarbeit, 2013.

Wieser, Martin. 2010. *Vergleich von Airborne-Laserdaten aus Bundesforste Bestand mit terrestrisch ermittelten Messdaten in einem repräsentativen Testgebiet im Nationalpark Donauauen.* Wien : Technical University Vienna, 2010.

Wood, Steve and Browne, Sue. 2007. Convenience store location planning and forecasting - a practical research agenda. *International Journal of Retail & Distribution Management.* 2007, Vol. 35, 4, pp. 233-255.