University of Münster

Geoinformatics

Leveraging Georeferenced Open Government Data

Thesis submitted in fulfillment of the requirements for the degree of

Venia Legendi

in the Department of Geosciences
of the Faculty of Mathematics and Natural Sciences
of the Westfälische Wilhelms-Universität Münster

Submitted by
Auriol Degbelo
from Cotonou, Benin

- 2019 -

Dean: Prof. Dr. Harald Strauß
First reviewer: Prof. Dr.-Ing. Christian Kray (supervisor)
Second reviewer: Prof. Dr. Ralf Bill
Third reviewer: Prof. Dr. Liqiu Meng

Abstract

In the recent years, several countries worldwide have been committing to open data principles, and public institutions in these countries have been making their datasets available, free of charge, for re-use. There are currently numbers of issues preventing a full exploitation of open government data (OGD), and this thesis intends to advance OGD research in three areas: user needs (what are needs and wishes of OGD users?), user information (how to effectively inform OGD users?) and user empowerment (how to enable OGD users to effectively re-use OGD?).

User needs have been mentioned in the literature, but not empirically researched. A set of contributions in this work tackles this currently open issue, and offers insights into needs of OGD users in Columbia and Spain. As to user information, there is currently a lack of tools to monitor OGD usage, and a lack of understanding of the effectiveness of different representations regarding information provision. The thesis brings forth two solutions to address these two issues: the Open City Toolkit Transparency Tool, and an empirical comparison of geovisualizations and data tables for information provision in the OGD landscape. Finally, a current issue regarding user empowerment is that of lack of visualization support for OGD visualization. A contribution of this thesis tackling this challenge is an approach (along with its implementation as a tool), which enables people without programming and Cartography expertise to create thematic web maps.

The contributions of this work help advance the state of the art of the (relatively new) research area of OGD. The user needs identified can provide evidence-based benchmarks to researchers, as they are striving to advance OGD scholarship. The contributions related to user information can provide a useful source of information to public institutions currently involved in OGD publishing activities. Finally, the contributions related to user empowerment could be relevant to data consumers (e.g. researchers, data journalists, businesses, employees of city councils, and members of non-governmental organizations) interested in adding value to existing OGD.

Acknowledgements

I thank my Saviour and Lord, Jesus-Christ for concentration, strength and creativity throughout process of putting together this thesis.

I owe much gratitude to Prof. Dr. Christian Kray for giving me the chance to join the Sitcom Lab, his amazing support, and for his idea of an Open City Toolkit which brought me to look into the topic of open government data.

I would like to extend my gratitude to all my co-authors (alphabetical order): Carlos, Chris, Dev, Fernando, Jonas, Joaquin, Nicho, Saad, Sergi, Sven and Tomi. Thank you very much for your inputs and discussions!

Due thanks goes also to Sitcom (Situated Computing and Interaction Lab) and ifgi (Institute for Geoinformatics) members for creating an environment, where work is not a burden, but truly pleasurable. Thanks to Anita, Guiying, Mehrnaz, Samuel and Shivam for preparing this latex template, and kindly sharing it with me!

Finally, I thank my parents (Augustine & Isidore), and my siblings (Anicet, Christ, Marina) for their unconditional support throughout. So nice to have you as a family!

This work was funded by the European Commission within the Marie Skłodowska-Curie Actions (ITN-EJD), Grant Agreement number 642332 - GEO-C - H2020-MSCA-ITN-2014.

Contents

List of Figures

List of Tables

List of Abbreviations

API	Application Programming Interface
GODI	Global Open Data Index
HCI	Human-Computer Interaction
ICT	Information and Communication Technology
NGO	Non-Governmental Organization
OCT	Open City Toolkit
ODB	Open Data Barometer
OGD	Open Government Data
SUS	System Usability Scale
URL	Uniform Resource Locator

1 Introduction

Open Government Data (OGD) has been increasingly available on the Web, thanks to a growing number of countries committing to open data principles. Germany's first Open Data law enabling free access to government data came into effect on July 13, 2017[1]. G8 members signed the Open Data Charter on June 18, 2013[2], G20 members agreed on the G20 Anti-Corruption Open Data Principles in 2014[3], and the International Open Data Charter has been adopted by more than 60 national and local governments at the moment of this writing[4]. At its core, OGD is a set of policies that promotes transparency, accountability and value creation by making government data available to all (OECD, 2018).

Open Data Inception, a platform collecting open data portals around the world, lists at present more than 2600 portals worldwide[5], implying hundreds of millions of open government datasets currently available. The exploitation of these datasets has attracted attention from research, and some recent (systematic) reviews provide insights into the current state of the art of OGD research. Attard, Orlandi, Scerri, et al., 2015 pointed out that OGD initiatives vary in their nature, but most common approaches to their implementation include data portals, data catalogues, and services. Hossain et al., 2016 identified three current drivers of the open data movement, namely political leadership, institutional pressure, and the emergence of digital technologies. Safarov et al., 2017 identified users of OGD to include citizens, businesses, researchers, developers, non-governmental organizations (NGOs) and journalists. Eberhardt and Silveira, 2018 found most commonly used visualization techniques of OGD to be maps and dashboards (made of pie, line and bar charts).

Specific investigations of the state of open government data in Germany were done in (Hunnius et al., 2014; Akyürek et al., 2018; Hinz and Bill, 2018a; Hinz and Bill, 2018b). Hunnius et al., 2014 indicated that administrative departments mainly provide data that is available in a structured format, at a fairly good quality level, not too sensitive, requires little maintenance and thus little effort to be published.

[1] `http://bit.ly/odatagermany` (last accessed: December 18, 2018).
[2] `http://bit.ly/g8opendata` (last accessed: December 18, 2018).
[3] `http://bit.ly/g20opendata` (last accessed: December 18, 2018).
[4] `http://bit.ly/odatacharter` (last accessed: December 18, 2018).
[5] `https://opendatainception.io/` (last accessed: December 18, 2018).

Akyürek et al., 2018 compared the usability of two commonly used platforms to publish open data in the federal state of North-Rhine Westphalia, Germany (i.e. DKAN and CKAN), and reported that CKAN provides a better user experience for new users and experienced maintainers compared to DKAN. Finally, Hinz and Bill, 2018a; Hinz and Bill, 2018b described OpenDataPortal, a one-stop-portal providing an overview about portals for open geodata in Germany, Austria, and Switzerland.

1.1 Promises of OGD

The increasing attention devoted to OGD worldwide, and in Germany can be attributed to the numerous promises of OGD. The total market value of Open Data was estimated between 265-286 billion Euro in 2020 for the EU28+ in (Carrara, Chan, et al., 2015)[6]. Carrara, Chan, et al., 2015 further forecasted about 100,000 jobs directly related to Open Data by 2020, accumulated cost savings of about 1.7 billion Euro for the EU28+ by 2020, and 629 million hours of unnecessary waiting time on the road in the EU saved per year (which is equivalent to 27.9 billion Euro per year). They also pointed out that not all datasets have the same potential for re-use, and that geographic information is the domain with the highest commercial value regarding open data re-use.

Next to these monetary benefits, open data has been conjectured to also yield non-monetary gains. An assumption of existing work is that OGD is a key *enabler* of Open Government (see Ubaldi, 2013)[7]. As Criado et al., 2018 pointed out, there is consensus among authors that there are three key topics of Open Government: transparency, collaboration and participation. In other words, OGD is a key enabler of greater transparency, collaboration and participation in the society. OGD can increase transparency of government institutions (Janssen et al., 2012; Ubaldi, 2013; Hossain et al., 2016; Safarov et al., 2017), increase self-empowerment of citizens (because they can now analyze available data and challenge a government) (Janssen et al., 2012; Ubaldi, 2013; Hossain et al., 2016), promote economic growth through innovation (Janssen et al., 2012; Ubaldi, 2013; Hossain et al., 2016; Safarov et al., 2017), and increase social value (because datasets related to public facilities may be helpful to enhance the quality of social life) (Hossain et al., 2016; Safarov et al., 2017). OGD initiatives were also reported to potentially impact the economy, governance, education, environment, tourism, transport and mobility of cities in (Ojo et al., 2015). A thorough discussion of possible political, economical, and technical benefits of OGD was presented in (Janssen et al., 2012).

[6] EU28+ in (Carrara, Chan, et al., 2015) referred to the 28 European Member States and the European Free Trade Association (EFTA) countries (Norway, Iceland, Liechtenstein, Switzerland).

[7] OGD and Open Government may exist without each other as discussed in (Yu and D. G. Robinson, 2012), but they can also work hand in hand when provision of re-usable data is used intentionally as a means to increase the overall governmental transparency.

Finally, OGD opens the door to multidisciplinary research on societal challenges. Example societal challenges of interest include (see Charalabidis et al., 2016): lack of cross-communities communication, better anticipation of unexpected crises, and improved governance through enhanced collective intelligence. Yan and Weber, 2018 reported that OGD has been used by nearly all research fields, with Medicine, Environmental Science, and Social Sciences being so far the most active fields to have engaged with OGD.

1.2 Problems of OGD

Despite the promises listed above, there are still a number of challenges hindering the exploitation of OGD to its full potential. This section will present key issues currently preventing the *effective use of OGD* (in contrast to those impeding the opening up of OGD discussed for example in (Barry and Bannister, 2014)). The barriers identified in previous work are classified here based on three dimensions - user needs, user information and user empowerment - in line with the user-centered stance adopted in this thesis. Before presenting the issues, the rationale for choosing the three dimensions is introduced:

- User needs: previous work (Vredenburg et al., 2002) reported that user-centered design (UCD) methods are generally considered to improve product usefulness and usability. There is a spectrum of ways in which users are involved in UCD, but the bottom line of all methods is that users are involved in *some* way (see Abras et al., 2004). Thus, identifying user needs and taking them into account in the OGD context could improve existing and future services surrounding OGD;

- User information: as mentioned above, transparency is believed to be one of the benefits of OGD. Michener and Bersch, 2013 identified two key dimensions of transparency - visibility and inferability - both revolving around information. That is, proper information delivery to users is crucial in order to reap the 'transparency benefit' expected from OGD;

- User empowerment: self-empowerment of citizens has been often mentioned in previous work as one of the possible benefits of OGD as said above. Self-empowerment goes beyond the effective informing of citizens, to enable them to build new *products* based on OGD. This self-empowerment is essential to reap the participation and collaboration benefits of OGD, and more generally, enable citizen-powerholder partnerships to move to higher levels of Arnstein, 1969's ladder of citizen participation.

Tab. 1.1: Overview of current issues surrounding OGD use*

Category	Example open issues (in alphabetical order)
User Needs	• **Lack of research on data users' needs** (Zuiderwijk and Janssen, 2012)
User Information	• Data ambiguity (Attard, Orlandi, Scerri, et al., 2015) • **Data fragmentation** (Zuiderwijk, Janssen, and Susha, 2016) • **Data hidden in reports** (Zuiderwijk, Janssen, Choenni, et al., 2012) • Data not findable (Zuiderwijk, Janssen, Choenni, et al., 2012) • Data not understandable to the general public (Zuiderwijk, Janssen, Choenni, et al., 2012) • No information about the update frequency of the data (Beno et al., 2017) • No information about the way the data was produced (Zuiderwijk, Janssen, Choenni, et al., 2012) • No metadata about the quality of the data (Zuiderwijk, Janssen, Choenni, et al., 2012) • Information overload (Zuiderwijk, Janssen, and Susha, 2016) • **Lack of data context** (Zuiderwijk, Janssen, and Susha, 2016) • Lack of search support (Zuiderwijk, Janssen, and Susha, 2016)
User Empowerment	• Data of limited or bad quality (Zuiderwijk, Janssen, Choenni, et al., 2012) • Data not relevant/interesting (Zuiderwijk, Janssen, Choenni, et al., 2012) • **Lack of a good application programming interface** (Zuiderwijk, Janssen, Choenni, et al., 2012) • Lack of support to improve already opened datasets (Zuiderwijk, Janssen, and Susha, 2016) • **Lack of the necessary skills to make use of data** (Janssen et al., 2012) • **Lack of visualization support** (Zuiderwijk, Janssen, and Susha, 2016) • Limited conditions for using data (Janssen et al., 2012) • No license for using data (Janssen et al., 2012) • No statistical knowledge/understanding of the potential/limitations of statistics (Janssen et al., 2012)

*In bold: issues addressed implicitly or explicitly in the thesis

User needs (what are needs and wishes of OGD users?), user information (how to effectively inform OGD users?) and user empowerment (how to enable OGD users to effectively re-use OGD?) are three areas where much work is still needed to advance OGD research. Table 1.1 presents a non-exhaustive list of current open issues pertaining to each of the categories. These issues were identified in previous articles and are summarized here along the three dimensions of interest.

As mentioned in (Zuiderwijk and Janssen, 2012; Zuiderwijk, Janssen, Choenni, et al., 2012), there is a dearth of research on the needs of users of OGD. Data publishers do not do that research (see Zuiderwijk and Janssen, 2012), nor has the scientific community actively endeavored to close that gap. A notable exception is (Beno et al., 2017) who assessed both users' and publishers' views on obstacles regarding open data adoption in Austria. The importance of knowing OGD users - on the road towards truly beneficial OGD delivery for the society - cannot be overstated. For instance, participants in the interviews conducted in (Zuiderwijk, Janssen, Choenni, et al., 2012) reported that "not relevant/interesting" data impedes smooth open data use. This issue is a direct consequence of a mismatch between user demand and providers' offer. Another issue preventing OGD re-use is that the data is not understandable for the general public due, for example, to the use of a jargon (see Zuiderwijk, Janssen, Choenni, et al., 2012). Further issues limiting an effective information of users include data not findable (Zuiderwijk, Janssen, Choenni, et al., 2012), data hidden in reports (Zuiderwijk, Janssen, Choenni, et al., 2012), data ambiguity (Attard, Orlandi, Scerri, et al., 2015), the lack of contextual information about the opened datasets (Zuiderwijk, Janssen, Choenni, et al., 2012; Zuiderwijk, Janssen, and Susha, 2016), data scattered over many portals (a.k.a. data fragmentation) (Zuiderwijk, Janssen, Choenni, et al., 2012; Attard, Orlandi, Scerri, et al., 2015; Zuiderwijk, Janssen, and Susha, 2016), the lack of advanced search support in many portals (Attard, Orlandi, Scerri, et al., 2015), and the impression of information overload (Zuiderwijk, Janssen, and Susha, 2016). The update frequency of existing datasets was also mentioned in (Beno et al., 2017) as a criterion which influences users in their choice of datasets to re-use.

Related to user empowerment, two key issues have been repeatedly mentioned in previous work. The first is that users may not necessary have the required skills (e.g. computing, statistical) to make effective use of open data (see Gurstein, 2011; Janssen et al., 2012; Zuiderwijk, Janssen, Choenni, et al., 2012); the second is that licensing restrictions hamper the usage of exiting OGD (see Janssen et al., 2012; Shadbolt et al., 2012; Hossain et al., 2016). The lack of a good application programming interface (API) and the provision of uninteresting datasets or datasets of bad quality does not favor OGD re-use either (see Zuiderwijk, Janssen, Choenni, et al., 2012). Finally, OGD research and practice would benefit from more data visualization support, as well as more support to improve already opened datasets, so that users are truly enabled to *add value* to existing OGD.

1.3 Scope

From the previous section, there are still many issues - related to user needs, user information and user empowerment - to resolve to take full advantage of OGD. Addressing these issues necessitates different kinds of measures, e.g. technical, cultural, organizational, and legal. The ways to solutions explored in this thesis are primarily technical (i.e. mitigate the issues through technological innovation) and the focus is on four research questions. The motives behind the choice of these questions are now presented in turn.

Clear insights into users' needs precede effective progress on tackling the problems they face, and this has motivated a closer look at the first research question:

- RQ1: What are current challenges preventing users to take full advantage of existing open government data?

In addition, as mentioned in (Zuiderwijk, Helbig, et al., 2014), "Currently, governments find it difficult to monitor use and few tools exist to monitor how and in what ways open data is used". This has led to the formulation of the second research question:

- RQ2: How to monitor applications' use of open government data on the Web?

Besides, there is a general agreement that visualization is an important topic for OGD use (see e.g. Shadbolt et al., 2012; Charalabidis et al., 2016; Eberhardt and Silveira, 2018), but the exact role of geovisualizations in the OGD context has been rarely, if at all, discussed. In particular, there are different media (e.g. data tables, pictures, diagrams, maps) through which OGD can be presented to users, and understanding their respective effectiveness at enabling transparency (i.e. making specific type of information visible to users) could inform data publishers about the strategies to use in a certain context, and ultimately improve the effectiveness of the publisher-consumer communication. This leads to the third research question:

- RQ3: What is the role of geovisualizations in enabling transparency in the open government data landscape?

Last, the importance of visualization for OGD is acknowledged but given that there are millions of OGD available (and more to come), data publishers may not be in the position to generate visualizations for each and every OGD. A more scalable strategy is to empower users to create visualizations of OGD as they like. Since cartographic knowledge is necessary for the creation of meaningful geographic visualizations (and its absence could quickly become a serious impediment to OGD re-use), this work has looked into the fourth research question:

- RQ4: How to enable users with low cartography expertise to create geovisualizations of georeferenced open government data?

Answering these four questions can help make progress in the three key areas introduced above: user needs (RQ1), user information (RQ2 and RQ3) and user empowerment (RQ4). The common denominator of all contributions of the thesis is that they have used or revolved around *georeferenced OGD*. Georeferenced OGD, in line with (Goodchild, 1998), is OGD which has some form of geographical footprint.

1.4 Method

A number of methods were used while finding elements of answer to the four research questions listed above. They are briefly introduced in this section.

Literature review: a literature review provides an examination of recent or current literature and "can cover wide range of subjects at various levels of completeness and comprehensiveness" (Grant and Booth, 2009). A literature review was used while synthesizing challenges preventing users to take full advantage of OGD (RQ1).

Contextual observation: contextual observation is about "observing how people act in the wild" (Tomitsch et al., 2018). It can be used to develop a better understanding of a design problem, or gather feedback about a prototype design (Tomitsch et al., 2018). Contextual observation was used while gathering evidence on current challenges preventing users from taking full advantage of OGD (RQ1).

Social survey: at its core, a survey is a "set of questions administered to a number of respondents" (Secor, 2010). It allows researchers to quickly get a large number of responses from a population of users that is geographically dispersed (Lazar et al., 2010), and can be very useful to get a "snapshot" of a user population (Lazar et al., 2010). Surveys require a sampling strategy of the target population, and two types of sampling strategies were used at various stages of the work: convenience sampling (RQ1) and snowball sampling (RQ3). Convenience sampling draws users from the target population on the basis of their accessibility (Jensen and Shumway, 2010), while snowball sampling begins by finding an entry point and making contact with some members of the target population; these contacts are then asked to provide names of others (Secor, 2010).

Interviews: as Secor, 2010 pointed out, interviews are often used for studies in which participants are viewed as “experts” from which the researcher hopes to learn from. There are three broad categories of interviews (see Tomitsch et al., 2018): structured interviews (i.e. interview script fixed in advance, and followed closely during the interview), unstructured interviews (i.e. the interview is based on open-ended questions and questions which emerge through the conversation), and semi-structured interviews (i.e. a combination of fixed-script questions and open-ended questions). Three to eight participants are recommended for a small interview study, according to Tomitsch et al., 2018. Interviews may use *probes*, i.e. external aids to promote engagement and aimed at eliciting feedback/reactions relevant to a topic at hand (see Lazar et al., 2010). As mentioned in (Lazar et al., 2010), unstructured and semi-structured interviews are most appropriate when the researcher is looking to dig deeper, searching for critical comments or seeking for design requirements. Unstructured and semi-structured interviews give participants the chance to educate the researcher, and the understanding gained from participants’ comments helps better grasp their needs (see Lazar et al., 2010). Thus, semi-structured interviews were used to gather insights from users during this work (RQ2, RQ3 and RQ4).

Experimental design: experiments have been widely used in Human Computer Interaction (HCI) to evaluate different design solutions. True experiments possess the following characteristics (see Lazar et al., 2010): (i) they have at least one testable research hypothesis; (ii) they usually have at least two conditions, with participants randomly assigned to each of the conditions; (iii) the experiment involves independent variables (i.e. causal factors or forces) and dependent variables (i.e. caused responses or effects)[8]; and (iv) dependent variables are usually measured through quantitative measurements and the results are analyzed through various statistical significance tests. Experiments in HCI usually have between 15 and 20 participants (see Lazar et al., 2010). This work has conducted a controlled experiment in the process of answering RQ3.

Usability testing: as Lewis, 2006 commented, usability testing involves representative participants, representative tasks, representative environments, and participants’ activities monitored by one or more observers. Following Lewis, 2006;

[8]The definition of dependent/independent variables is adapted from (Visser and Jones III, 2010).

Lewis, 2014, usability testing may be formative (i.e. have a primary focus on problem discovery) or summative (i.e. focus on task-level measurement). Within HCI, the world of usability testing encompasses (see Lazar et al., 2010): (i) testing prototypes that have only been built on paper (a.k.a. paper prototypes); (ii) testing prototypes that look complete but have a human behind the scenes responding (a.k.a. "Wizard of Oz" technique); (iii) testing working versions of a software before it is officially released; and (iv) testing software that has already been implemented in existing systems. (iii) was primarily used during this work to learn about the usability of the some of the prototypes developed (RQ2 and RQ4). Usability testing in the context of this work has primarily been *summative*.

Linear models and linear mixed models: there are numbers of data analysis procedures and statistical methods, each with their context of use and assumptions (see Lazar et al., 2010, for a review). The key idea behind linear models and linear mixed models is regression analysis, i.e. identifying the relationship between a dependent variable, and one or more independent variables (Pearce, 2009). Known tests such as t-tests or ANOVAs can be viewed as specific cases of regression. Linear models rely on the assumption of independence between data points, while linear mixed models are adequate when this assumption no longer holds (see Winter, 2013). Linear (mixed) models do not need an assumption of a normal distribution, and are flexible dealing with unbalanced designs. These flexibilities have motivated their choice as statistical analysis method during the work.

Triangulation: research strategies have their respective drawbacks and the key purpose of triangulation is "decrease, negate, or counterbalance the deficiency of a single strategy, thereby increasing the ability to interpret the findings" (Thurmond, 2001). Triangulation can be of various types discussed in (Thurmond, 2001): data triangulation (collection of datasets at different times, or different places to see if similar findings occur), investigator triangulation (use of more than one observer, interviewer, coder, or data analyst during the study), methodological triangulation (use of different methods to collect some evidence about the phenomenon of interest), theoretical triangulation (use of multiple theories when examining a phenomenon) and data-analysis triangulation (combination of two or more methods of analyzing data). Triangulation increases confidence in the conclusions reached after a study, and both methodological (RQ3 and RQ4) and investigator triangulation (RQ4) were resorted to during the course of the work. The data collected about user needs while answering RQ1 is relevant to data triangulation (see Section 11.1.1 for the discussion).

1.5 Contributions

Table 1.2 summarizes the main contributions of the thesis. These contributions are discussed in this section along the three dimensions of user needs, user information and user empowerment introduced earlier. The contributions are also annotated using two taxonomies: Wobbrock and Kientz's taxonomy of research contribution types, and Charalabidis et al.'s taxonomy of OGD research topics.

Wobbrock and Kientz, 2016 listed seven types of research contributions in HCI (these are also applicable to GIScience). *Theoretical* contributions consist of new or improved concepts, definitions, models, principles, or frameworks; *empirical* contributions provide new knowledge through findings based on observation and data gathering; *methodological* contributions create new knowledge that informs how scientific work is carried out; *artifacts* contributions are often prototypes (e.g. systems, architectures, tools, toolkits, mockups) which reveal new possibilities or compel us to consider new possible futures; *dataset* contributions provide a new and useful corpus, often accompanied by an analysis of its characteristics, for the benefit of the research community; *survey* contributions synthesize work done on a research topic with the goal of exposing trends and gaps; and *opinion* contributions aim at provoking reflection/discussion/debate on a topic, and changing the minds of readers through persuasion.

Charalabidis et al., 2016 provided a taxonomy listed 35 research topics of OGD. The 35 research topics are not listed here, but pertinent topics to the current work will be mentioned, where appropriate. The reader is referred to (Charalabidis et al., 2016) for an overview of OGD research areas and topics.

1.5.1 User Needs

The first two chapters of this thesis address the current lack of insights into users' needs in the OGD landscape. Chapter 2 puts OGD reuse in the broader context of a smart cities vision. It relies on the assumption that empowering citizens to take full advantage of available open data is a promising way to foster innovation, creativity, and solutions in future cities. The chapter has done a literature review to arrive at six citizen-centric challenges of open data re-use (C1, Table 1.2). C1 is an opinion contribution.

Chapter 2 speaks on behalf of the users, but chapter 3 compiles self-reported issues from users (gathered through an online survey) and problems faced by users while interacting with existing open government data portals (gathered through contextual observation). Four cities within two countries were taken as use case:

Bogotá (Colombia), Medellín (Colombia), Cali (Colombia) and València (Spain). This chapter has resulted in 19 barriers to open data re-use (C2, Table 1.2), a contribution primarily empirical in nature.

C1 and C2 are contributions to the OGD research area of "OGD needs analysis", which was defined in (Charalabidis et al., 2016) as follows: "This research [area] includes studies of OGD users' needs, with respect to both government datasets, and also functionalities of OGD infrastructures, aiming to lead to further developments of OGD strategies of public organizations, and also functionalities of ODG infrastructures/portals". C1 and C2 in combination provide elements of answers to **RQ1**.

1.5.2 User Information

The first set of contributions under the *User Information* theme aims at addressing the lack of tools to monitor the usage of open government data (**RQ2**).

Chapter 4 has surveyed categories of open datasets from 40 European open data catalogs to inform the design of more sophisticated APIs for OGD. The first contribution of this chapter is a dataset (C3) available online at (Degbelo, Trilles, and Bhattacharya, 2016). Since this dataset collects recurrent categories of open government data portals in Europe, it can be used to inform the design of bi-national or transnational APIs for OGD use in Europe. As such, it is a contribution to the OGD research area of "Service interoperability standards" - the investigation of "standards that can be used for seamless interconnection among OGD related services, in order to serve different OGD uses and user scopes" (Charalabidis et al., 2016). Next to this contribution, the chapter also presents the technical components of *semantic APIs* (i.e. APIs which enable OGD retrieval according to their types, increasing thereby transparency). This fourth contribution is an artifact (C4) and its code is openly available on GitHub[9]. It helps advance the OGD area of "Open webservices/APIs", which is concerned with "facilitating and providing well-designed standards for application programming interfaces (APIs) in OGD platforms, in order to ensure the exploitation and re-usability of published data" (Charalabidis et al., 2016).

Chapter 5 presents examples of web maps using the semantic API from chapter 4. The example web maps intend to add value to existing OGD through the conversion of existing OGD into a more descriptive, machine-readable format (i.e. RDF, Resource Description Format), and the use of web mapping libraries to geovisualize the datasets (making thereby hidden insights in datasets more easily 'consumable' to users). The contribution of this chapter is a set of artifacts (C5), which belong

[9] `https://github.com/geo-c/OCT-Core`.

topically to the area of “OGD portals architecture” - research “conducted concerning the development of architectures of ICT Infrastructures that allow for and support application development utilizing OGD” (Charalabidis et al., 2016).

Chapter 6 presents the Open City Toolkit (OCT) transparency tool (C6), a dashboard visualization of OGD dataset usage. The semantic API from chapter 4 and the apps presented in chapter 5 form two main components of the OCT Transparency tool. The chapter presents also two types of evaluation of the tool: usability and utility. Usability is assessed from the developers’ point of view and touches upon tasks such as register an app, register a dataset, and build a first app; utility is assessed through eight semi-structured interviews of employees from the city councils of Lisbon and Münster. The contribution of this chapter is thus primarily empirical and advances the “OGD portals architecture” area. Contributions C3 to C6 provide elements of answers to RQ2.

The second set of contributions under the *User Information* theme addresses the lack of understanding about the effectiveness of different media for transparency enablement in the OGD landscape (**RQ3**). Chapter 7 proposes to use questionnaire-based measurement to assess the effectiveness of visualizations as to transparency enablement. The contribution of this chapter is of type opinion (C7). Chapter 8 performs a comparison of geovisualizations and data tables during a controlled experiment, to extract their respective strengths as to transparency enablement. The contribution of this chapter is thus empirical (C8). C7 and C8 provide elements of answers to RQ3. Topicwise, they advance the “OGD visualization methods and tools” area, which is concerned with “develop[ing] features and tools for facilitating the creation of visualizations by users on OGD” (Charalabidis et al., 2016).

1.5.3 User Empowerment

The lack of visualization support is a pressing issue to advance OGD adoption, and the set of contributions under the *User Empowerment* theme aim at advancing the state-of-the-art in that area.

Chapter 9 proposes to increase the intelligence of existing geovisualizations as a means to make them more flexible and usable by a broader population. The chapter introduces five requirements of intelligent geovisualizations (as an opinion contribution, C9) and these present exciting opportunities for research and practice on “OGD visualization methods and tools”.

Chapter 10 proposes a semi-automatic approach for the creation of web maps, by and for users with no prior training in Cartography. A proof of concept prototype

Tab. 1.2: Key contributions of the thesis

Problem	*Solution*	*C. Type*	*OGD Research Area*
	USER NEEDS		
Lack of insights into challenges of OGD users	C1: six citizen-centric challenges of open data re-use	opinion	OGD needs analysis
Lack of insights into challenges of OGD users	C2: 18 barriers to open data re-use from a user's viewpoint	empirical	OGD needs analysis
	USER INFORMATION		
Lack of tools to monitor OGD use	C3: Categories of datasets of current open data portals	dataset	Service interoperability standards
Lack of tools to monitor OGD use	C4: semantic API to track apps' use of OGD	artifact	Open web services/APIs
Lack of tools to monitor OGD use	C5: example web maps monitoring their dataset use	artifact	OGD portals architecture
Lack of tools to monitor OGD use	C6: Open City Toolkit Transparency Tool	artifact, empirical	OGD portals architecture
Lack of evaluations of geovisualizations' effectiveness regarding transparency enablement	C7: questionnaire-based approach to insight measurement	opinion	OGD visualization methods and tools
Lack of evaluations of geovisualizations' effectiveness regarding transparency enablement	C8: a comparison of geovisualization and data tables for transparency enablement	empirical	OGD visualization methods and tools
	USER EMPOWERMENT		
Lack of flexibility of existing OGD visualizations	C9: five requirements of intelligent geovisualizations for OGD	opinion	OGD visualization methods and tools
Lack of flexibility of existing OGD visualizations	C10: an approach for web map creation by non-experts	artifact, empirical	OGD visualization methods and tools

illustrating the feasibility of the approach is presented (C10, artifact), and that prototype is evaluated through usability testing (C10, empirical). The contribution in this chapter advances also the area of "OGD visualization methods and tools". Both C9 and C10 provide elements of answers to **RQ4**.

Chapter 11 comments on the generalizability of the findings, general limitations of this work, as well as directions for future research. Chapter 12 concludes the thesis.

1.6 Publications

This thesis is cumulative and each chapter is a manuscript either published, or currently under review. Table 1.3 presents the mappings between the manuscripts used as chapters, the contributions and the research questions.

Tab. 1.3: Publications

Chapter	Contribution	Manuscript	Status
Chapter 2	C1 [RQ1]	Degbelo, A., Granell, C., Trilles, S., Bhattacharya, D., Casteleyn, S. and Kray, C. (2016) 'Opening up smart cities: citizen-centric challenges and opportunities from GIScience', ISPRS International Journal of Geo-Information, 5(2), p. 16. doi: 10.3390/ijgi5020016.	Published
Chapter 3	C2 [RQ1]	Benitez-Paez, F., Degbelo, A., Trilles, S. and Huerta, J. (2018) 'Roadblocks hindering the reuse of open geodata in Colombia and Spain: A data user's perspective', ISPRS International Journal of Geo-Information, 7(1), p. 6. doi: 10.3390/ijgi7010006.	Published
Chapter 4	C3, C4 [RQ2]	Degbelo, A., Trilles, S., Kray, C., Bhattacharya, D., Schiestel, N., Wissing, J. and Granell, C. (2016) 'Designing semantic application programming interfaces for open government data', JeDEM - eJournal of eDemocracy and Open Government, 8(2), pp. 21–58.	Published
Chapter 5	C5 [RQ2]	Degbelo, A. and Kauppinen, T. (2018) 'Increasing transparency through web maps', in Champin, P.-A., Gandon, F. L., Lalmas, M., and Ipeirotis, P. G. (eds) Companion of Proceedings of the Web Conference 2018 - WWW '18. Lyon, France: ACM Press, pp. 899–904. doi: 10.1145/3184558.3191515.	Published
Chapter 6	C6 [RQ2]	Degbelo, A., Granell, C., Trilles, S., Bhattacharya, D. and Wissing, J. (2019) 'Tell me how my open data is re-used: increasing transparency through the Open City Toolkit', in Hawken, S., Han, H., and Pettit, C. (eds) Open Data | Open Cities: Collaborative Cities in the Information Era. Palgrave Macmillan.	In Press
Chapter 7	C7 [RQ3]	Degbelo, A. (2017) 'Linked data and visualization: two sides of the transparency coin', in Vo, H. T. and Howe, B. (eds) Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS'17. Los Angeles, California, USA: ACM Press, pp. 1–8. doi: 10.1145/3152178.3152191.	Published
Chapter 8	C8 [RQ3]	Degbelo, A., Wissing, J. and Kauppinen, T. (2019) 'A comparison of geovisualizations and data tables for transparency enablement in the open government data landscape', International Journal of Electronic Government Research.	Under Review
Chapter 9	C9 [RQ4]	Degbelo, A. and Kray, C. (2018) 'Intelligent geovisualizations for open government data (vision paper)', in Banaei-Kashani, F., Hoel, E. G., Güting, R. H., Tamassia, R., and Xiong, L. (eds) 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Seattle, Washington, USA: ACM Press, pp. 77–80. doi: 10.1145/3274895.3274940.	Published
Chapter 10	C10 [RQ4]	Degbelo, A., Sarfraz, S. and Kray, C. (2019) 'A semi-automatic approach for thematic web map creation', Cartography and Geographic Information Science.	Under Review

Part I

User Needs

2 Opening up Smart Cities: Citizen-Centric Challenges and Opportunities from GIScience

This chapter was published in the ISPRS International Journal of Geo-Information as *Degbelo, A., Granell, C., Trilles, S., Bhattacharya, D., Casteleyn, S. and Kray, C. (2016) 'Opening up smart cities: citizen-centric challenges and opportunities from GIScience', ISPRS International Journal of Geo-Information, 5(2), p. 16. doi: 10.3390/ijgi5020016.*

Abstract. *The holy grail of smart cities is an integrated, sustainable approach to improve the efficiency of the city's operations and the quality of life of citizens. At the heart of this vision is the citizen, who is the primary beneficiary of smart city initiatives, either directly or indirectly. Despite the recent surge of research and smart cities initiatives in practice, there are still a number of challenges to overcome in realizing this vision. This position paper points out six citizen-related challenges: the engagement of citizens, the improvement of citizens' data literacy, the pairing of quantitative and qualitative data, the need for open standards, the development of personal services, and the development of persuasive interfaces. The article furthermore advocates the use of methods and techniques from GIScience to tackle these challenges, and presents the concept of an Open City Toolkit as a way of transferring insights and solutions from GIScience to smart cities.*

2.1 Introduction

It is widely recognised that the concept of smart cities is still emerging, and different stakeholders have distinct conceptualizations about what a smart city is or should be. When considering recent smart cities projects, it is obvious that they deal with distinct facets of cities, and that they have disparate objectives and implementation strategies. Some are driven by companies to promote (proprietary) technology-and sensor-intensive cities (e.g. IBM Smarter Cities[1], Microsoft CityNext[2]), while others are run by consortia of universities, companies and city councils and take a collaborative approach to build smart cities (e.g. MK:Smart D'Aquin et al., 2014, CitySDK[3]). In this paper, we adopt the definition by Yin *et al.* Yin et al., 2015: *a smart city is a system integration of technological infrastructure that relies on advanced data processing with the goals of making city governance more efficient, citizens happier, businesses more prosperous and the environment more sustainable*. This definition emphasizes the role of citizens as main beneficiaries (as in e.g. Nam and Pardo, 2011), and places data and advanced data processing (as in e.g. González et al., 2014) at the center. Similar to Ojo et al., 2015; Janssen et al., 2012; Masip-Bruin et al., 2013, this article furthermore builds upon the assumption that open data may yield substantial benefits to (smart) cities. We indeed believe that *opening up cities*, i.e. empowering citizens to take full advantage of available open data, is a promising way to foster innovation, creativity, and citizens-centric solutions for smart cities. In addition, we argue that geographic data and Geographic Information Science (GIScience) may play an important role in shaping smart cities.

With the proliferation of smart city initiatives, the risk of duplicating efforts and re-inventing the wheel increases. To mitigate this risk, there is a need to get a complete picture of what different research fields can offer to tackle smart city challenges. Comprehensively exposing the achievements of different disciplines allows us to identify those areas that can fruitfully collaborate to realize the smart city vision. This article is written with this viewpoint in mind, and articulates what GIScience has achieved and can offer to smart cities. By matching smart city challenges and GIScience achievements we demonstrate that GIScience is essential in addressing citizen-centric challenges in a smart city context. Consequently, our contribution is threefold:

- a synthesis of citizen-centric challenges in the smart city context;
- a collection of relevant key contributions of and opportunities from GIScience to help address the identified challenges;
- a citizen-centric, technology-driven approach to address these challenges (the Open City Toolkit).

In the following sections, we first discuss related work on smart and open cities (Section 2.2). Section 2.3 concisely summarizes key citizen-centric challenges. The

[1] See http://www.ibm.com/smarterplanet/us/en/smarter_cities/overview/ (last accessed: August 6, 2015).
[2] See http://www.microsoft.com/en-us/citynext/ (last accessed: August 6, 2015).
[3] See http://www.citysdk.eu/ (last accessed: August 6, 2015).

opportunities offered by GIScience to address the challenges are introduced in Section 2.4. Section 2.5 presents the Open City Toolkit as an approach to realize these opportunities, and outlines core research directions currently explored within the GEO-C project. Section 2.6 summarizes the main contributions of this article.

2.2 Related work on smart and open cities

Due to the inherent complexity of smart cities, previous work has produced distinct but complementary perspectives on identifying problems, challenges and trends for the effective conceptualization and implementation of smart cities. This section briefly reviews existing literature on smart and open cities, and sets the scene for identifying key challenges and opportunities in the following section(s).

2.2.1 Trends in technologies, architectures, and infrastructures for smart cities

Yin et al., 2015 conducted an exhaustive literature survey of smart cities, which touched upon a number of dimensions or perspectives including application domain, technological infrastructure, system integration and data processing. The authors concluded that some researchers have defined smart cities from multiple perspectives, while others have given a definition covering only one of the four perspectives[4]. In the literature (Yin et al., 2015; Department for Business, Innovation & Skills, 2013a; Boulos Kamel, Tsouros, et al., 2015) many application areas have been discussed such as government (increasing efficiency and transparency through open data, citizen services, smart city platform, heritage monitoring), citizens (increasing happiness, participation and education) and economy (increasing revenues via social wifi, e-commerce, tourism management, mobile marketing, outdoor digital marketing). Further areas include environment (increasing sustainability by providing solutions for energy efficiency), mobility (improving parking, public transit, or traffic management) and public service utilities such as water and waste (increasing efficiency).

A combined physical and digital infrastructure is considered central to shape smarter solutions for application domains such as the ones listed above. For digital infrastructures, Information and Communication Technology (ICT) and Internet technologies can be regarded as a means to integrate and coordinate city subsystems in order to make cities smarter, more livable and more sustainable (Batty, 2013b; Celino and Kotoulas, 2013). Yin et al., 2015 proposed a technological architecture along those lines, which is composed of four layers: a data acquisition layer, a

[4]See also (Nam and Pardo, 2011) for another detailed discussion of the smart city concept.

data vitalization layer, a common data and service layer, and a domain application layer. Data has a prominent role here, which reflects the authors' view that *[f]rom the perspective of computers and information systems, the city is defined by its sensed data*. Nevertheless, the authors also recognize the tension between accuracy and computational costs of models[5]. A second challenging task in smart cities they identified is the re-use of existing datasets for purposes different from the ones they were originally collected for.

An alternative perspective on smart city architectures is proposed by Silva et al., 2013. The authors analyzed 17 technical architectures from literature, and extracted a set of challenges, mostly technical and data-driven, that smart city architectures should address: objects interoperability, sustainability, real-time monitoring, historical data, mobility, availability, privacy, distributed sensing and processing, service composition and integrated urban management, the incorporation of social aspects, and flexibility/extensibility.

In addition to digital infrastructures, physical infrastructures are also vital to realize smart cities. More specifically, the Internet of Things (IoT) can be regarded as a critical enabler of smart cities infrastructures (Boulos Kamel and Al-Shorbaji, 2014). Zanella et al., 2014 discussed implementation strategies for urban systems leveraging the inherent characteristics of IoT to connect and integrate *a large number of different and heterogeneous end systems, while providing open access to selected subsets of data for the development of a plethora of digital services*.

Vlacheas et al., 2013 defines a social layer, which can potentially horizontally connect several application domains, on top of the IoT to simplify the management of huge volumes of objects. Atzori et al., 2012 proposes a similar concept, the so-called "Social Internet of Things" paradigm, in which things borrow the concepts of cooperation and social relationships for the establishment and management of social relationships between smart objects or things. As Chen *et al.* Chen et al., 2014 point out, the widespread deployment of IoT drives the high growth of data both in quantity and diversity, which results in big data. Similarly, the application of big data technology to IoT accelerates research in this area and facilitates the development of new business models for IoT.

2.2.2 Beyond technologies, architectures, and infrastructures

Challenges in smart cities can be explored from other viewpoints than digital and physical infrastructures. For example, Nam and Pardo, 2011 argue that technological innovation is a means to a smart city, not an end. Branchi et al., 2014 also highlight

[5]Very accurate models are desirable, but they are also computationally expensive.

that smart cities are not only about technologies applied to the city and its spaces. They should also take into account the impact technologies have on the inhabitants of cities. For this purpose, Branchi et al., 2014 proposed the Technologies Analysis Matrix (TAM), which can be used to assign scores to technology-related aspects (e.g., usefulness, advantages/disadvantages, risks/benefits), with respect to impact dimensions (e.g., environmental sustainability, economic sustainability and social sustainability). In addition, the authors proposed the Smart City Matrix (SCM) as a tool to assess how a combination of technologies performs on the mobility, energy, and quality of life in a city.

Rather than focussing on technology alone, there is a growing recognition that designing and deploying citizen-centric city services greatly improves the smartness of a city (Department for Business, Innovation & Skills, 2013a). In this sense, cities need to *open up* to their citizens, by offering their public data in an easily accessible and re-usable format. This enables citizens to access exactly the information - and services built upon them - they need, whenever they need. There have been only a few initiatives taking up open, participative and shared development of cities from the perspective of citizens, but the concept is spreading. More often, smart city projects are geared towards corporate solutions and proprietary platforms for smart cities (Kehoe et al., 2011; Menychtas, Kranas, Donovang-Kuhlisch, Heindrichs-Krusch, et al., 2011; Menychtas, Kranas, Donovang-Kuhlisch, U. Schade, et al., 2011; Donovang-Kuhlisch et al., 2011).

Next to the aforementioned citizen-centric viewpoints, various researchers also considered smart cities from a strategical and design point of view. Angelidou, 2014 presents a review of strategies to realize smart cities. The author distinguishes between national and local strategies, hard and soft infrastructure-oriented strategies, new and existing cities, as well as economic sector-based versus geographically-based strategies. She gives examples of cities implementing each of these strategies, and recommends that cities begin the journey towards becoming smarter by selecting a few domains or areas that urgently need to be improved. Batty, 2013b takes an urban modeling approach to synthesize how concepts from complexity science may shape our understanding of today's cities and how cities can be designed in better ways. Zaman and Lehmann, 2011 identified critical factors and challenges for resource efficiency and management, while McGrath and Pickett, 2011 investigated how to properly integrate ecology and urban design in smart cities contexts.

2.2.3 Open city projects and initiatives

One of the projects dedicated to an open and user-driven philosophy was The Open Cities project (Open Cities Consortium, 2011) started in 2011. The Open

Cities project aims to validate how to approach open and user-driven innovation methodologies in the public sector towards Future Internet services for smart cities[6]. The project plan is to leverage existing tools, trials and platforms in crowdsourcing, open data, and open sensor networks in major European cities.

Developed along similar lines of openness, the project CitySDK (Forum Virium Helsinki, 2014) (Smart City Service Development Kit) aims to create a smart city application ecosystem through large-scale demand-driven CityPilots that package and align key smart city application areas to an open source service developer toolkit. Another relevant project is Open311 Open311 Organisation, 2014, which focuses on providing open communication channels for issues that concern public spaces and public services. One key component of Open311 is a standardized protocol for location-based collaborative issue-tracking. By offering free web API access, the Open311 service is an evolution of the phone-based 311 systems that many cities in North America offer.

Another smart city project strongly related to the concepts of openness and smart citizen participation is the Open & Agile Smart Cities (OASC) initiative by The Connected Smart Cities Network organization Connected Smart Cities Network, 2015. This project aims to popularize the use of a shared set of wide-spread, open standards and principles, thereby facilitating interoperability between different systems within a city, and across multiple cities. This in turn should enable the development of smart city applications and solutions to reach many cities at once. OASC conceives smart city platforms as the combination of APIs developed by the FIWARE Platform Usländer et al., 2013 and data models defined in CitySDK, and uses this combination to leverage a driven-by-implementation approach. Cities are meant to use and improve standard data models based on experimentation and actual usage.

2.2.4 Role of GIS and GIScience

Several researchers have pointed out the importance of GIScience in the vision of smart cities. Contrary to Yin et al., 2015, the survey presented by Brauer et al., 2015 has a specific thematic scope: the impact of Green Information Systems on fostering environmental sustainability in smart cities. The authors point out the importance of GIS for collecting and monitoring environment-related data, but also for other

[6] The Open Cities project should not be confused with the partnership of stakeholders having the same name (i.e. open cities). Open Cities (the partnership) is described at (*Open Cities* n.d.) and is interested in the use of open data to produce innovative solutions for urban planning and resilience challenges across South Asia.

smart city dimensions such as transportation, infrastructure, buildings and urban planning[7].

Daniel and Doran, 2013 discuss potential contributions of geomatics to smart cities, with a focus on technologies, and the pervasiveness of geospatial information. They argue that the integration of ICT and geomatics tools is indispensable for the development of a smart city.

Roche, 2014 poses the question *what can GISciences do specifically to make cities smarter?*. He first extracts four dimensions of smart cities: the intelligent city (its social infrastructure), the digital city (its informational infrastructure), the open city (open governance), and the live city (its continuously adaptive urban living fabric). He then argues that: (i) GISciences can support the development of the intelligent city; (ii) GISciences can also support smart cities by dramatically enhancing the digital city dimension, and in particular the urban informational infrastructure; (iii) the governance dimension of smart cities (called 'open city') can benefit from recent advances in GISciences; and (iv) the live city dimension can also greatly benefit from GISciences, and especially from geodesign (Steinitz, 2012).

The work we present in this article is in line with Roche's regarding the importance of GIScience in a smart city context, but there is an important difference in focus. Where Roche's work emphasizes extracting current trends in the smart city context, we focus on tackling citizen-centric challenges using GIScience. We match GIScience contributions to citizen-centric challenges in order to shed some light on possible solutions rather than matching GIScience ideas to the four smart city dimensions, as Roche does. Finally, we point out a subtle but important issue, namely the use of the term GISciences in Roche's work (in plural form; Roche leaves it undefined). In our work, we focus on the commonly accepted field of GIScience[8].

2.3 Challenges

The trends which have been outlined in the previous section suggest that cities are the focus for many disciplines, ranging from social, economic and environmental sciences, architecture, design and urban planning, to social network analyses, sensor networks and human sensors. Regardless of the focus, recent experiences with smart city developments show that an important challenge is to expose, share and use data (Rodger, 2015). Nevertheless, opening up data without compelling incentives for developers, private companies, and citizens, along with a clear strategy and

[7] Proportion (in %) of GIS to all surveyed tools from the literature analysis done in (Brauer et al., 2015): transportation (23%); infrastructure (50%); buildings (67%); urban planning (100%).
[8] For a recent discussion on the scope of GIScience, see (Goodchild, 2010).

committed management by the data providers (e.g., public authorities) is most likely bound to fail (M. Lee et al., 2016).

Masip-Bruin et al., 2013 enumerate three rationales behind the support of open data initiatives: (i) open data makes government more transparent, participative and collaborative, (ii) open data encourages public involvement in data collection, analysis and application, often reducing government spending or improving efficiency accordingly, and (iii) open data creates a new source of economic growth. Janssen et al., 2012 also studied possible benefits of open data initiatives over smart cities. These covered a number of dimensions such as the political and social dimensions (e.g. more transparency, equal access to data), the economic dimension (e.g. simulation of innovation), and the operational and technical dimensions (e.g. external quality checks of data, sustainability of data).

Besides the benefits of open data for governments, citizens and businesses, there are also risks related to its publication that should be managed (Kucera and Chlapek, 2014). Open data is faced with issues in terms of risks, contingency actions, and expected opportunities in terms of governance, economic issues, licenses and legal frameworks, data characteristics, metadata, access, and skills (S. Martin, Foulonneau, Turki, et al., 2013a). Issues such as the unlawful disclosure of data, the infringement of trade secret protection, violations of privacy and breaches of the security of the infrastructure might have a severe negative impact. Therefore the compliance assessment and the quality control of the data being published should be implemented into the open data publication process. Where the primary data contains sensitive data like personal information, anonymization should be applied (Kucera and Chlapek, 2014). One criticism of current open data initiatives is that they are largely supply-driven (when they should be driven by the demand of citizens). Zuiderwijk and Janssen Zuiderwijk and Janssen, 2014 put forward the idea that a context and dataset dependent decision-making model is needed to weigh the benefits of open data (e.g. creating transparency, the possibility to strengthen economic growth), against the risks and disadvantages of open data (e.g. violating privacy, possible misuse, false impressions, mismanagement issues and misinterpretation of data).

S. Martin, Foulonneau, Turki, et al., 2013b state that despite the development of open data platforms, the wider deployment of open data still faces significant barriers. The lack of insight into the user's perspective and the lack of appropriate governance mechanisms can explain the large gap between the promises of open data and what is actually realized (Janssen et al., 2012).

Finally, as T. Roberts, 2012 state, open data may increase the digital divide and social inequality unless approached right. The only sustainable basis for delivering public benefit from public data is therefore to motivate and enable communities themselves to innovate local service provision, social enterprise and job creation.

The challenge of opening up data can be considered at two levels: infrastructure data and citizen data (Rodger, 2015; SmartCitiesWeek, 2015). Unfortunately, most infrastructure data in a city is still locked away, due to a variety of reasons: lack of resources, knowledge, technical skills, vision, etc. The *open data* movement, although gaining traction, has only scratched the surface of *freeing* this type of data Schaffers et al., 2011. There is a need of cheaper, accessible and better solutions to allow cities and infrastructure developers and maintainers to get their data out and expose it (Caragliu et al., 2011). Citizen data is vital for cities – it's a ground truth for citizens' activities and desires – yet people are often unwilling to share data because they are concerned about privacy and trust issues (Hollands, 2008). We need to develop trusted data creators and certifiers, which will allow citizens to feel confident that they have complete control over the data they share (including the ability to revoke data sharing Naphade et al., 2011), and who uses it for what purpose. Caragliu et al., 2011 elaborate on the concept of smart cities as environments of open and user driven innovation for experimenting and validating Future Internet-enabled services. There is a need to clarify the way living lab innovation methods, user communities, Future Internet experimentation approaches (Granell et al., 2016), and test-bed facilities constitute a common set of resources. These common resources can be made accessible and shared in open innovation environments (Batty et al., 2012), to achieve ambitious city development goals. This approach requires sustainable partnerships and cooperation strategies among the main stakeholders (Naphade et al., 2011).

Based on such critical pointers of development gaps in smart cities approaches, research themes and challenges directly tailored to citizens' needs are brought forth in this section. Here, the assumption is that smart cities cannot become a reality unless citizens are central actors in shaping their cities (Van den Bergh and Viaene, 2015). Citizen-focused challenges for smart cities are not entirely new though. A 2015 CJRES's special issue on *Thinking about smart cities* (Glasmeier and Christopherson, 2015), for example, examined current perceptions on the goals, challenges, and limitations of smart cities beyond of infrastructure- and technology-intensive visions, to stress on greater equity, improved quality of life, and citizen empowerment. Also, smart city professionals recently interviewed before a Smart City Event held in Amsterdam (see Hoevenaars, 2015) highlighted similar challenges: collaboration among different stakeholders, adaptation for growth, as well as costs and funding.

The effect of the above citizen-focused vision for smart cities is palpable in our work. Empowering citizens, analytical methods and tools, and citizen-centric

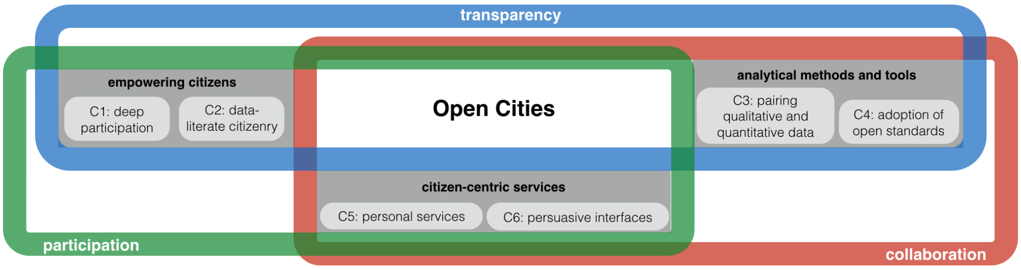


Fig. 2.1: Citizen-centric challenges grouped into three research themes: empowering citizens, analytical methods and tools, and citizen-centric services.

services research themes (Figure 2.1) are useful to improve transparency, facilitate participation, and ease collaboration in a city context. These challenges are not the only ones in smart cities, but they are crucial to better understand the spatiotemporal interactions between cities and citizens. For this reason, we stress in next sections the role of GIScience in the research themes and challenges discussed throughout the paper.

2.3.1 Empowering citizens

Citizen empowerment is a dynamic process, whereby citizens get increasingly engaged with the services a city offers and with other fellow citizens. This process builds upon openness to enable citizens to share data, experiences and skills. It may provide an attractive environment to ultimately fuel transparency and data literate citizenry. Van den Bergh and Viaene Van den Bergh and Viaene, 2015 aptly identify two groups of cities: those that interpret a smart city based on high infrastructure demands, and the ones that opt for a smart citizen focus. The latter vision is consistent with a recent study by Kogan *et al.* Kogan, 2014, which identified citizen empowerment and engagement as the top success factor of a smart city project, thereby pushing ICT into the background. Put simply, without engaged and educated citizens on the access, creation, and interpretation of data and knowledge, a city may only be halfway smart and open.

- Deep participation (**C1**): Recent works (Janssen et al., 2012; Lathrop and Ruma, 2010) have investigated citizen participation in various contexts, including smart cities, where people are often seen as data-collectors improving city services. Yet citizens are more than human sensors collecting data: deep participation is about raising awareness, building capacity, and strengthening communities (Townsend, 2013). There is a need to work *with* the community and not just for, or on, the community (Craglia and Granell, 2014), and this must be reflected in the overall strategy to envision a smart city. Furthermore, city councils must pay special attention to the design and execution of strategies to foster citizen participation at all levels.

- Data literate citizenry (**C2**): Smart cities are not only about ICT and infrastructures; smart cities are about smart citizens, who participate in their city's daily governance, are concerned about enhancing the quality of life, and about protecting their environment. Data literacy should be a skill not just for scientists, but for all citizens. Cities can commit to open data, transparency and ICT as major enablers, but without the appropriate data literacy skills, co-creation and active participation with citizens is unlikely to occur. A key gap relates to how people can gain a sense of control. This necessitates fostering digital inclusion and data literacy skills to interpret and understand the processes and services that drive smart cities.

2.3.2 Analytical methods and tools

Cities need to connect macro (objective, aggregated data) and micro (subjective, citizen-generated data) observations to figure out how global phenomena (transport, mobility, energy, etc.) occurring at city scale relate to multiple citizen observations. Listening to what citizens sense and perceive, and acting consequently is a way of improving quality of life in cities. The analytical methods and tools theme contains the following research challenges:

- Pairing quantitative and qualitative data (**C3**): Analysis methods that are able to integrate quantitative data and qualitative information through citizen science activities, social networks services, and crowdsourcing tools, will have a great impact on the future of our cities as more and more people live in urban areas (Craglia and Granell, 2014). In some cases citizens-generated data takes the form of measurements or quantitative observations (e.g. noise and air pollution measurements). In others such observations are more qualitative or subjective (e.g. opinions, emotions, behaviors) but no less useful. There is a need to move beyond the traditional quantitative analysis of physical phenomena to include also new analytical methods to scrutinize qualitative perceptions of the same phenomena as they are perceived by those who live in and *sense* the city[9]. In addition, the combination of datasets in the big data age needs to cope with a number of challenges listed in (Chen et al., 2014; Fan et al., 2014), for example efficient data representation, redundancy reduction and data compression, spurious correlations, and noise accumulation.

[9]A preliminary look into the rationales as well as challenges involved in the integration of quantitative and qualitative geographic data was provided in (Degbelo, De Felice, et al., 2013).

- Adoption of open standards (**C4**): standards are essential to ensure that underlying technology, systems, data, applications and services are able to interact seamlessly in a coherent manner. Not only does standardization refer here to service interfaces, communication protocols, and architectures but also to data. The adoption of open data standards can dramatically unlock the potential of all citizens to access and use open data. Many cities wrongly assume that making data available, say in pdf format, is enough to be tagged as an open data city. Unless one is a developer that can code a pdf crawler, all this *open* data remains useless to other citizens[10]. Even though many cities are leaders in open data, there still exist barriers (e.g., the lack of open standards) impeding the access and use of such data broadly by people. The point is that small changes towards open standards can eventually lead to big impacts like making city services more transparent, participative and trustable.

2.3.3 Citizen-centric services

The citizen-centric services theme centers on the question of how to redesign existing services and/or provide new services that place citizens at the forefront. Citizen-centric services are emerging as an interaction paradigm linking citizens' needs, skills, interests and their context to data-rich environments like cities.

- Personal services (**C5**): As human beings, we only use a very small part of the retina, called the fovea[11], to see the finer details of objects that we are looking at. The rest of the visual field, which is known as peripheral vision[12], plays a key role even though it does not allow us to distinguish such details. When we detect an object that requires our interest in the sides, we quickly put the fovea on it to identify the object properly. Without the ability to detect the presence of other objects that surround us through our peripheral vision, our vision would be severely limited to a small portion of the visual field.

  Turning back to the smart city context, a research gap is the lack of customized and focused services, i.e. personal services, that are capable of adapting to the peculiarities and needs of each individual citizen, and that help them in performing daily tasks, provide them with up-to-date information, or simply support them in finding their way through the ever-increasing data stream sources available in today's cities by presenting the clearest picture possible of what all this data means. These personal services augment our *peripheral*

[10] See concrete examples at https://www.ted.com/talks/ben_wellington_how_we_found_the_worst_place_to_park_in_new_york_city_using_big_data (last accessed: August 20, 2015).
[11] See https://en.wikipedia.org/wiki/Fovea_centralis (last accessed: October 23, 2015).
[12] See https://en.wikipedia.org/wiki/Peripheral_vision (last accessed: October 23, 2015).

vision, to put in the forefront the pieces of information that could be relevant and might require our immediate attention.

When it comes to personal data and services, privacy is an important issue to tackle. Janssen et al., 2012 mention the unclear trade-off between transparency and privacy values as one of the adoption barriers of open data (and consequently of all benefits associated with the use of open data to make the city smarter). Solove, 2006 discussed the concept of privacy in detail and pointed out that it covers many aspects. Particularly relevant to the current discussion are:

- surveillance: the watching, listening to, or recording of an individual's activities;
- aggregation: the combination of various pieces of data about a person;
- identification: linking information to particular individuals;
- secondary use: the use of information for a purpose other than what it was initially collected for, without the data subject's consent;
- increased accessibility: amplifying the accessibility of information.

Technological progresses, the open data movement, and the trend of big data provide an environment where the risk of privacy harms related to the five aspects above-mentioned is increased. For example, Lyon Lyon, 2014 mentions that *big data practices are increasingly important to surveillance*, and that *in a big data context, the same data are increasingly used for different purposes*. Linked Data, which helps to describe the content and context of resources (see Scheider, Degbelo, et al., 2014), eases aggregation and identification. The open data movement requires increased accessibility. As a result, reducing the risk for privacy violation (e.g. by putting the citizen fully in control of the kind of information s/he would like to disclose) is, in the current context, a major challenge.

Regarding GIScience, the field has focused on location privacy. As Duckham and Kulik, 2006 stress, *[o]ur precise location uniquely identifies us, more so than our names or even our genetic profile*. Challenges mentioned in (Duckham and Kulik, 2006) regarding location privacy include (i) understanding the techniques a hostile person might employ in order to invade another person's privacy, and (ii) the development of truly spatiotemporal models of location privacy.

- Persuasive interfaces (**C6**): City governments pursue novel ways to engage with citizens as to better support their needs and concerns, and to involve them in decisions that affect them. Among the existing methods for getting citizens engaged (e.g. public consultations, local meetings, etc.), the creation of persuasive interfaces is getting importance as user interfaces are seen by citizens as the only *visible interfaces* between city services and themselves. The field of persuasive interfaces is not new; it can be traced back to Tversky and Kahnemann's pioneering work on the prospect theory about framing decisions and psychology of choice (Tversky and Kahneman, 1992). City services need to go beyond traditional interfaces to pay attention to more user-centric interfaces that stimulate and encourage change. From the point of view of GIScience, the challenge lies not so much in design and psychology (which are important aspects), but in creating new types of user experiences that facilitate opportunistic interactions with citizens (Hespanhol et al., 2016), and present information in such a way that citizens are persuaded to change their behavior and take actions accordingly. The stakes here are high, because citizens' behavior plays an inescapable role against today's most pressing environmental issues in cities (Umpfenbach, 2014).

2.4 Opportunities from GIScience

In this section, we look at the contributions from GIScience[13] to address the social and technical challenges and research themes described in Section 2.3.

GIScience has so many influences in multiple aspects of a city that it is a foundational part of smart cities for data acquisition, processing, analysis, representation, and visualization (Vinod Kumar, 2014). This is aptly synthesized by Gruen, 2013 in that *a smart city possesses spatial intelligence*. In the rest of the section, we look at each research theme, and point to existing work (i.e. research contributions, methods and tools) from GIScience that are relevant to address them.

Before going into what the GIScience community is doing, it is worth mentioning that from our perspective, the need to open up the city is a common denominator of many potential solutions to empowering citizens. The open data movement can be regarded as an engine for innovation, economic growth; as a way to create added-

[13] Some of the key contributions of GIScience to date were summarized in (Goodchild, 2010). Those relevant to the current discussion are presented where appropriate, and completed with additional recent contributions from GIScience research.

value services and applications; and to enhance efficiency, effectiveness, and cost savings at city level (Domingo et al., 2013; S. Schade et al., 2015). In this respect, recent case studies (MH Government, Cabinet Office, London, United Kingdom, 2012; Ratti and Townsend, 2011; Lathrop and Ruma, 2010) have demonstrated that concrete actions can help governments to unleash the potential of public data to empower a transparent governance model (e.g. citizens can identify errors, prevent abuses, and inefficiencies), which ultimately builds trust between citizens and their cities (Domingo et al., 2013; Fioretti, 2012). Despite these benefits, open data initiatives are in reality far from operating at their fullest potential. Fortunately, some leading smart cities highlight the fact that citizen engagement and participation are success factors to stimulate the access and reuse of open city data by public and private stakeholders alike (Department for Business, Innovation & Skills, 2013b).

2.4.1 Empowering citizens

Two main research challenges were introduced in Section 2.3 regarding the empowerment of citizens theme: deep participation (**C1**), and data literate citizenry (**C2**). Table 2.1 summarizes key contributions from the GIScience community with respect to empowering citizens.

Tab. 2.1: Matching GIScience contributions to citizen-centric smart city challenges (Theme: empowering citizens). The use of maps is a promising approach to address both the issues of deep participation and data literate citizenry in a smart city context.

Research challenges	*Existing GIScience contributions to tackle the challenges*
Deep participation (**C1**)	• Open Geographic Data & Open GIS • Synchronous distributed online collaborative mapping • Maps as spatial dialogue platforms • Location-based services as means to highlight engagement opportunities both spatially and temporally • Gamification approaches for Volunteered Geographic Information • Insights from GIScience research into contributors' motivation
Data literate citizenry (**C2**)	• Maps as one way of contextualizing and presenting primary data in an understandable way • Insights from spatial thinking research: improving spatial thinking improves STEM achievements

- Deep participation (**C1**): participation at all levels and by all citizens has attracted relatively few attention in the smart cities literature. Public participation GIS (PPGIS) was perhaps one of the first attempts to put geospatial capabilities, tools and applications in the citizens' hands to enhance effective participation and communication among experts and non-experts. Even though PPGIS literature applies to many application domains in cities, decision-making processes in urban planing have quite probably been the domain by excellence for collecting and exploiting local knowledge from citizens through geospatial collaborative tools (Bugs et al., 2010). Geospatial visual methods, in varied forms, have been traditionally used to engage users and enable participation. Fechner and Kray, 2014 proposed an approach which relies on space and time as common integrators, and uses augmented interactive geo-visualizations to facilitate citizen engagement. They introduced three ideas, and exemplary tools, worth exploring in a smart city context: (i) synchronous distributed online collaborative mapping, (ii) the use of maps as spatial dialogue platforms, and (iii) the use of location-based services to highlight engagement opportunities both spatially and temporally.

  Improving deep participation in cities cannot be done without a deep understanding of the motivations of citizens to participate. Coleman et al., 2009 provide a useful summary of contributors' motivations (e.g. altruism, social reward, enhanced personal reputation, or mischief) to willingly produce geographic information. Since the very same contributors of geographic information are also actors (active or passive) in a smart city, deep participation strategies should take into account Coleman et al.'s synthesis about citizens' motivations. Creating this type of win-win situations between city players is a critical success factor for smart cities, whereby city councils and organizations not just collect data and knowledge from citizens, but also give something back that is valued by citizens (Craglia and Granell, 2014).

  Another example of this type of project is the Smart GraphHopper[14], which uses GraphHopper[15] in order to plan routes and subsequently compare them by evaluating different available sensor data, such as noise, air pollution and so on. NoiseTube (Maisonneuve et al., 2009) uses this initiative to gather data from citizens' phones.

  Gamification is a current trend to overcome the limitations of PPGIS tools and applications, and to foster citizen participation and engagement. Martella et al., 2015 have produced a gamification framework for Volunteered Geographic Information (VGI, (Goodchild, 2007)) which has three main parameters: the

[14] See `https://github.com/DIVERSIFY-project/SMART-GH` (last accessed: January 7, 2016).
[15] See `https://github.com/graphhopper/graphhopper/` (last accessed: January 7, 2016).

user, the tasks of the user (data gathering, data validation or data integration), and the types of datasets manipulated by the user. Along the same lines, See et al., 2015 discussed a combination of social gaming, geospatial mobile tools and data collection campaigns to increase the network of volunteers to capture urban morphology for climate modeling purposes.

- Data literate citizenry (**C2**): a data literacy strategy also requires simple and understandable presentations of existing datasets (e.g., in forms of visualizations or geo-visualizations). Fechner and Kray, 2014 argue that maps are one way of contextualizing and presenting primary data in an understandable and engaging way. As such, maps have a key role to play in the improvement of data literate citizenry[16]. Kraak, 2006 points out that maps have the ability to present, synthesize, analyze and explore the real world, and do this well because they visualize it in an abstract way, and only present a selection of the complexity of reality. Wakabayashi and Ishikawa, 2011 present the ability to organize, understand, and communicate with maps as one component of spatial thinking. As a result, insights from spatial thinking research can inform the design of better applications in a smart city context. For instance, the study documented in (Wakabayashi and Ishikawa, 2011) concluded that people associate concrete spatial behavior in their daily lives (such as navigation and wayfinding in space, or sorting of furniture or packaging) with the act of thinking spatially. Uttal et al., 2013 report that improving spatial thinking improves science, technology, engineering, and mathematics (STEM) achievements. This insight implies that part of making citizens smarter is the development of applications which help them improve their spatial thinking abilities.

2.4.2 Analytical methods and tools

Two main challenges were introduced in Section 2.3 regarding the analytical methods and tools theme: pairing quantitative and qualitative data (**C3**), and the adoption of open standards (**C4**). Table 2.2 recaps existing contributions of the GIScience useful to address challenges in this research theme.

- Pairing quantitative and qualitative data (**C3**): relevant to the smart city context is the use of cellular automata to model cities. Cellular automata appear on Goodchild's list of major GIScience achievements. Cellular automata help to model the environment as adjacent cells. Each cell has a state which refers to its attributes, and transitions between cell states are modeled using simple

[16]For measures describing the readability of maps themselves, see (Harrie et al., 2015).

Tab. 2.2: Matching GIScience contributions to citizen-centric smart city challenges (Theme: analytical methods and tools). The suite of OGC open standards is a good starting point for the exchange of (geospatial) information in a smart city context.

Research challenges	*Existing GIScience contributions to tackle the challenges*
Pairing quantitative and qualitative data **(C3)**	• Observation ontologies taking into account both human and technical sensors • Cellular automata as a method for urban growth prediction and simulation • The observation-driven framework for the engineering of geo-ontologies • An algebra for spatiotemporal data • Insights from research in geographic information semantics • Fields as generic data type for big spatial data
Adoption of open standards **(C4)**	• OGC open standards • OGC-based spatial information framework for urban systems and spatial decision-making • OGC SWE & cloud computing • OGC SensorThings API for IoT

rules. They can be interpreted as generators of growth and decline[17]. The wide use of the SLEUTH cellular automata model (for reviews of its applications, see e.g., Clarke, Gazulis, et al., 2007; Chaudhuri and Clarke, 2013) provides evidence that cellular automata is a technique worth considering to predict and simulate urban growth in a smart city context.

So far, GIScience's approach towards the integration of quantitative and qualitative data has been the use of observation ontologies which take into account both (e.g., Kuhn, 2009; Degbelo, 2013; Degbelo, 2015). These works have the concept of observation at the core of their investigations, and are based on the premise that all we know about the world is based on observations (Frank, 2003). To make sense of observation data, GIScience has produced ODOE (Janowicz, 2012), the observation-driven framework for the engineering of geo-ontologies out of observation data. ODOE supports both human and technical sensors, and is therefore useful to consider when pairing quantitative (usually coming from technical sensors) and qualitative data (mostly produced

[17]Put differently, a cellular model assumes only an action space (usually a grid), a set of initial conditions, and a set of behavior rules (Clarke and Gaydos, 1998). For an introduction to cellular automata, see (Batty, 1997).

by humans). Noteworthy also is the algebra for spatiotemporal data (Ferreira et al., 2014) which allows to derive objects and events from the three basic types of observations, namely time series, trajectories, and coverage. Stasch et al., 2014a brought forth a theory which helps to enforce meaningful prediction and aggregation of observations[18]. Kuhn, 2013 proposed eight ideas that many researchers found useful in their work on geographic information semantics:

- experiential realism: people conceptualize reality based on how they experience it through their bodies, sensing and acting in physical environments and in cultures;
- geographic information atoms: the simplest form of a geographic information is a tuple of location and attribute values;
- semantic reference systems: making the semantics of terms explicit and grounding them physically, so that transformations between them can be computed;
- semantic datum: useful to transform between different reference systems;
- similarity measurement: all semantics is context-dependent and can generally not be modeled objectively or even standardized;
- conceptual spaces: provide structures to solve conceptual problems through geometry;
- meaning as process: meaning comes from people using a word, rather than the words having a meaning on their own;
- constraining the process of meaning: tools can only be built to constrain the use and interpretation of terms, not specify *the* meaning[19] of these terms).

Making sense of the wealth of available data in a smart city context can build upon these eight ideas. Finally, GIScience's proposal of field as generic data type for big spatial data Camara et al., 2014 is worth considering when dealing with issues of efficient data representation in a big data context.

[18] Ready-to-use implementations of the theory are still a work in progress.

[19] Specifying THE meaning presupposes a single meaning that one should strive towards defining, but as mentioned above, it is the people who mean something when they use terms in a specific context, not the terms which have a meaning *per se*.

- Adoption of open standards (**C4**): In GIScience, standards of the Open Geospatial Consortium (OGC)[20] are used in a wide variety of domains including environment, defense, health, agriculture, meteorology, sustainable development, and smart cities. Recent works (Percivall, 2015; Li et al., 2013) identify the importance of open location standards to any smart city project and propose a spatial information framework for urban systems and spatial decision-making processes based on the integration of OGC open standards and geospatial technology. The combination of open standards (and APIs) such as OGC CityGML (e.g. 3D spatial city visualization), IndoorGML (e.g. indoor/outdoor navigation/routing to map indoor spaces), Moving Features, and Augmented Reality Markup Language 2.0 (ARML 2.0), would ease the delivery of geospatial features, imagery, sensor observations and geo-referenced social media in a coherent way, and thereby support interoperable and cross-domain city services for urban spatial intelligence, spatial visualizations, and decision making purposes.

  Sensors are crucial for intelligent systems like smart cities (Dameri and Rosenthal-Sabroux, 2014; Hancke et al., 2012) and are well covered by the OGC Sensor Web Enablement (SWE[21]). The OGC SWE standards suite specifies interfaces and metadata encodings to enable real-time integration of heterogeneous sensor networks (Bröring et al., 2011). In this way, most types of sensors can be discovered, accessed and reused for creating web-accessible sensor applications and services (see examples in Devaraju et al., 2015; Tamayo, Huerta, et al., 2009). For example, Mitton et al., 2012 combined cloud-based services to process SWE-encoded sensing data in smart cities.

  When using mobile devices as ubiquitous sensing tools, OGC SWE protocols for data exchange between mobile devices introduce considerable overhead and performance penalties (Tamayo, Granell, et al., 2012). In addition, as SWE standards can be used for creating complex, time-consuming applications, such applications are often limited for resource-constrained devices (Vermesan and Friess, 2014). As a result, and due to the need for compatibility with mainstream technology (e.g. IoT), the OGC has recently delivered the OGC SensorThings API Liang et al., 2015 as a candidate standard. The OGC SensorThings API can be considered as a lightweight OGC SWE profile, that follows a REST-like style, and is particularly well suited for developing IoT-based sensing applications to interconnect resource-limited IoT devices. SEnviro (Trilles et al., 2015), a low-cost, Arduino-based IoT device that monitors atmospheric

[20] See `http://www.opengeospatial.org/` (last accessed: October 19, 2015). The Open Geospatial Consortium is an international not for profit organization which develops open standards for the global geospatial community. See Reed et al., 2015 for further details.

[21] See `http://www.opengeospatial.org/ogc/markets-technologies/swe` (last accessed: October 19, 2015).

variables demonstrated that IoT protocols and the OGC SensorThings API can work together for real-life smart cities applications.

2.4.3 Citizen-centric services

The citizen-centric services research theme comprises two specific challenges: personal services (**C5**) and persuasive interfaces (**C6**). Both model and shape the citizen's personal relationship with a city, its services and places. Table 2.3 summarizes key features from GIScience useful to tackle each challenge.

Tab. 2.3: Matching GIScience contributions to citizen-centric smart city challenges (Theme: citizen-centric services). The seven principles of research into location privacy and the theory of spatialization of user interface can guide research into personal services and persuasive interfaces.

Research challenges	*Existing GIScience contributions to tackle the challenges*
Personal services (**C5**)	• Seven principles of research into location privacy
Persuasive interfaces **(C6)**	• Spatialization of user interfaces • Gestural interaction

- Personal services (**C5**) may be regarded as the new generation of location-based services (LBS). The ability to know the location, both in out- and indoor environments, in real-time paves the way for smart city-specific advances in areas such as location-context systems, real-time tracking and routing, location-based advertising, and so on. Duckham, 2010 identified seven key principles of research into location privacy: (i) geographic space presents constraints to movement, (ii) humans are not random, (iii) large user-contributed datasets are biased, (iv) continuous and snapshot queries are different, (v) location privacy attacks are as important as location privacy protection, (vi) decentralization does not always improve location privacy, and (vii) location accuracy, and location precision are not synonyms (although both can be used to hide information about a person's location). These principles were identified from location privacy research over recent years. Given that location (or the spatial dimension) is a very important component of smart cities (see e.g., Daniel and Doran, 2013; Roche, 2014 for arguments in favor of such a view), privacy research in a smart city context can use these seven principles, as both starting points and guiding insights.

- Regarding user-centric, more persuasive interfaces (**C6**), GIScience offers a theory of spatialization of user interfaces. In pioneering work, Kuhn, 1996 pointed out that *[s]pace is fundamental to perception and cognition because it provides a common ground for our senses as well as for our actions,* and discussed the need for spatial expertise in the field of human-computer interaction. He argued that designers need to be informed about human spatial cognition and properties of spaces in order to design more successful spatialized interfaces. His work introduced two key concepts for the design of intuitive user-interfaces: spatial metaphors and image-schemas. Both concepts are useful to understand how people think about space. A formalization of metaphors and image-schemas in the context of user interfaces was proposed in (Kuhn and Andrew Frank, 1991). Recent work in GIScience (Bartoschek et al., 2014) has looked at gestural interaction with spatiotemporal (linked) open data. In particular, gestures were considered helpful in engaging people with the visualization of complex data (Bartoschek et al., 2014). In summary, incorporating spatial elements and insights may help to provide more effective and intuitive interaction with (personal) smart city services.

2.4.4 Discussion

As the previous sections illustrate, GIScience may help to address citizen-centric challenges in smart cities. Two core pilars of GIScience, namely spatial representation and visualization, and spatial analysis, are particularly relevant for smart cities. GIScience has already developed useful standards, frameworks, formal specifications, techniques, approaches and principles (see Tables 1 to 3) that deal with the representation, understanding, analysing and visualizing spatial aspects of the world. These could be exploited to enforce the spatial component of smart cities. In addition, GIScience may also benefit from smart city initiatives. Indeed, a smart city not only consumes data to produce useful services, but it also generates a broad variety of data[22]. This wealth of data may serve as input for what Miller and Goodchild called data-driven geography (Miller and Goodchild, 2015). Miller and Goodchild commented that with big data, the context of geographic research has shifted from a data-scarce to a data-rich environment. They described data-driven geography as an evolution of geography, and argued that it can provide the paths between idiographic (i.e. description-seeking) and nomothetic (i.e. law-seeking) knowledge.

Tables 1 to 3 also show that maps are a recurrent helpful component to address citizen-centric challenges. The map is explicitly present in approaches which aim at tackling the issues of deep participation (**C1**), and data literate citizenry (**C2**). It is

[22]For example, in a big city like London, about 45 millions journeys a week are generated from the smart card used by rail and buses passengers (see Batty, 2013a).

also implicitly present in approaches for analysis (**C3**), the adoption of standards (**C4**) and the development of persuasive interfaces (**C6**). For instance, maps (and geoanalytics) are often used for visually informing end users about analysis results; OGC standards include the Web Mapping Service and the Web Map Tile Service specifications, both dealing with map rendering (see Percivall, 2015); and maps also play a key role in gestural interaction (Bartoschek et al., 2014). All this indicates that maps are a central component for spatial representation and visualization in smart cities. Other related GIScience work, such as the underlying spatial representation models or alternative visualization techniques, are equally applicable in smart cities.

In addition, spatial analysis is an invaluable part of understanding spatio-temporal data, detecting patterns and making predictions. In today's expanding cities, where an explosive amount of organizational, participatory, demographic, environmental, and spatial data is present, the analysis techniques and solutions developed in GIScience are particularly relevant. Spatial analysis aspects are explicitly present in pairing quantitative and qualitative data (**C3**), adoption of open standards (**C4**), but also relevant for deep participation (**C1**) and personal services (**C5**). Example applications of spatial analysis include crime detection and prediction (F. Wang, 2005), green living and sustainability (Brauer et al., 2015), traffic congestion and control (Barba et al., 2012).

Next to these core GIScience areas, other aspects which are not exclusive for the GIScience field, but have a strong spatio-temporal dimension, offer opportunities to address citizen-centric challenges in smart cities. Tons of geographic data come from citizens through pictures, tweets, geotags, reports, GPS tracks, VGI, (or more generally crowd-sourced data), is increasingly relevant for designing, improving, and assessing city services. Along with VGI, sensor networks and IoT devices are becoming much more pervasive in cities. Such devices are location-based and so location is central to realise context-aware and personal services for a great variety of city services and settings (e.g. outdoor and indoor services). While such IoT devices, sensors, and personal services accelerate the production and consumption of city services, such flows of data also introduce serious privacy and security concerns related to unforeseen uses of citizens' location - an issue already considered in GIScience. Finally, alternative exploration and visualization techniques, such as virtual and augmented reality, provide new ways to present added value information and service, and present a new way to experience smart cities.

In a nutshell, there exists a symbiosis between GIScience and smart cities, and maps are critical in addressing citizen-centric challenges in smart cities. Yet, reaping the benefits of development in GIScience research for smart cities (and the other way round) will not be automatic. It depends on two factors: knowledge transfer, and the availability of open data (open data is the fuel of a data-driven science). The Open City Toolkit - a way of transferring insights and solution of GIScience to smart cities - intents to facilitate this knowledge transfer, and will be introduced in Section 2.5.1.

2.5 Towards the realization of opportunities from GIScience

The previous sections elaborated and discussed opportunities and achievements of GIScience to address smart city challenges, giving supportive evidence that GIScience contributions are key enablers to smart cities. Nevertheless, a full understanding of all the facets, benefits, and possibilities that GIScience can bring to smart cities is still at an early stage. The recently launched EU-funded European Joint Doctorate *Geoinformatics: Enabling Open Cities* (GEO-C[23]) targets a better understanding of this role, from a variety of perspectives. GEO-C's overarching objective is to make substantial scientific progress towards the notion of smart (open) cities. It is worth mentioning at this point that, despite the availability of commercial solutions to tackle smart city issues (e.g., IBM Smarter Planet solutions[24]), there is still a lack of an integrated open source solution to support the move towards smarter cities[25]. Besides the training focus of the GEO-C program, it is also a research project on its own to produce a joint-development of an Open City Toolkit (OCT). Subsequent subsections briefly introduce the vision of the OCT (Section 2.5.1) as well as example research directions at the intersection of GIScience and smart cities (Section 2.5.2).

2.5.1 The Open City Toolkit

In order to realize the opportunities outlined in section 2.4 in a smart (open) city context, different methods can be applied, for example, technology-driven deployments of commercial systems, or citizen-driven participatory design of new urban services. These methods are subject to some limitations. In particular, they usually either favor technology or citizens, but rarely both. In addition, it is often not easy to combine individual solutions, and the transition process from a 'non-smart' to a smart city is neglected. In order to overcome these issues, our research agenda envisions an Open City Toolkit (OCT) at its core, whose working definition is as follows:

[23]See `http://geo-c.eu/` (last accessed: August 20, 2015).

[24]See `http://www.ibm.com/smarterplanet/us/en/smarter_cities/infrastructure/` (last accessed: October 30, 2015).

[25]The Generic Enablers (GEs) built within initiatives such as FIWARE are a good first step, but more is needed, in particular an integrated piece of software which delivers useful services to citizens based on open data (in addition to independent software pieces).

> The Open City Toolkit (OCT) is a collection of tools, processes, specifications and guidelines to empower citizens to participate in and shape the future of their cities, and to deliver services based on open data that are useful for citizens, businesses and governing bodies alike.

An important part of the OCT is an integrated, open source software empowering citizens, providing them with analytical tools and citizen-centric services in the context of a smart city. The OCT is therefore technology-driven and citizen-centric. The usefulness of the OCT is threefold: (i) provide software components addressing the challenges mentioned in Section 2.3; (ii) integrate work done in different facets of smart cities, as detailed further in this section, and (iii) transfer insights from GIScience to smart cities. In essence, five types of components are envisioned for the OCT:

- **A set of tools to improve transparency**: to enable citizens to inspect what data is gathered and how it is used, and to visualize key indicators so that all stakeholders can understand them. This relates to the challenges of deep participation[26] (**C1**) and data literate citizenry (**C2**);
- **A curated set of examples of open source apps, open data and services**: apps and services that are useful to cities/citizens, and relate to the challenges of pairing quantitative and qualitative data (**C3**), as well as the development of personal services (**C5**) and persuasive interfaces (**C6**);
- **An abstract architecture**: describes how apps, processes, services and data can be integrated in order to realize a smart open city. This abstract architecture is to be built upon open standards (**C4**);
- **A "glue" to connect resources, apps and services to realize an open city**: involves a set of APIs and specifications to link components and tap into data. This facilitates further development based on existing resources and artifacts, thereby opening up the smart city's "living" ecosystem.
- **Guidelines on how to realize an open city**: interactive guidelines describing insights about how to facilitate transparency, collaboration, participation using methods from GIScience. The guidelines will also document insights as to how to support the transition to a smart and open city.

[26] Transparency relates to the visibility and inferability of the information (see Michener and Bersch, 2013), while participation relates to the involvement of citizens in city operations. This work assumes that greater transparency will have a positive impact on citizen participation.

By providing such a common, flexible framework/platform, and by fostering transparency, collaboration and participation, we intend to create a bridge between all stakeholders (councils, citizens, companies), between technology and society, and between research and practice. In addition, by incorporating city transformation guidelines and providing set of useful examples for developers and users alike, we aim to facilitate the transition towards smarter cities. Finally, by providing it as open source, any interesting party – be it city authorities, researchers, businesses, practitioners or citizens themselves – can easily obtain, use and/or build on it.

For example, the OCT as a platform will support the integration of existing or novel location-based services such as future transport services or location-based educational apps. When services are realized via the OCT or connected to it, they will benefit from the transparency and participation features built into the framework. These include users being able to identify which data sources are used by which service or being able to configure which services are executed in a smart city and how. Similarly, a broad range of data sources is supported. For example, data produced via a range of sensors using IoT technologies can be easily connected to the OCT. Once this is done, it is accessible for all services and apps running on the OCT, and can also be inspected with the transparency tools built into the OCT.

The OCT is currently being built using web technologies[27]. The primary target users are citizens and city councils, while keeping private companies and governmental institutions as key stakeholders in mind as well. In fact, the GEO-C consortium consists of a mix of city councils and private companies, and foresees links with government institutions and access to citizens via the projects' host cities. All these stakeholders help to define the requirements for the OCT.

The Open City Toolkit will incorporate the results of the various research lines within the GEO-C project. In particular, it will keep all the data, processes, guidelines, standards, ontologies, frameworks and models open, and it will also provide utilities, tools and applications for open smart cities. To facilitate its use, it will incorporate search facilities to retrieve resources according to the specific purpose and needs, as well as browsing and exploration facilities.

2.5.2 Future research directions at the intersection between smart cities and GIScience

In this section, we overview future research directions, summarised in Table 2.4, which are being pursued by the combined team of 30+ doctoral and post-doctoral

[27] The release of the first version of the OCT is due at the end of October 2016.

researchers within the GEO-C project, and provide ample opportunities for other researchers in the field.

One of the research directions worth investigating for fostering citizen participation (**C1**) is the application of the openness principles to ensure that all citizens benefit from and participate in smart cities on all levels. Smart cities need informed and educated citizens who can participate on a deeper level, and can understand how sensed information is being used. Only then, can a win-win situation occur that permits to overcome crucial barriers in accessing, using, and interpreting open data (Janssen et al., 2012).

Promising research directions are the combination of ideas and methods from VGI research, open data and open access, and human-computer interaction to develop hybrid approaches that widely engage diverse groups of people. For example, identifying and understanding the main motivating factors that characterize online citizen participation, and the production and use of VGI by citizens is essential. An interesting case study to explore these issues is the use of public displays as integrators in smart cities. Optimizing two-way information flows between citizens and public displays (i.e. city open data) is central for a timely provision of what they need, with minimal effort. Public displays may facilitate opportunistic and ad-hoc participation in decision-making as well as knowledge creation. Geoinformatics, cartography, maps, visual arts, and design can help citizens to understanding complex interactions by customizing the content that is being displayed. Especially in today's cities, the traditional concept of maps that is strongly coupled to cartography needs to be updated, given that the lines between big data, cartography, and visual arts in mapping are increasingly blurred. Another future line to leverage deep participation is to explore the concept of virtual meeting geo-spaces to bridge the gaps between VGI and PPGIS, i.e., between citizens-driven (bottom-up) and administration-initiated (top-down) approaches. Such virtual meeting geo-spaces would permit a new communication channel to start a dialogue among citizens about a concrete geo-referenced item of interest to all involved participants.

With respect to data literacy (**C2**), the availability of suitable tools to turn citizens (from school children to seniors) into educated and informed citizens of smart open cities is vital to enhance digital literacy. A remarkable example with respect to data literate citizenry (**C2**) is the Open Data Institute (ODI[28]), which carries out mostly training, education, and promotion activities about the consumption and publication of public open data. The ODI's programs are mainly targeted at developers and technically-skilled users who can transfer open data know-how to public and private organizations. This may foster open data literacy as a means to promote economic

[28] See `http://theodi.org/` (last accessed: October 29, 2015)

growth and innovation by facilitating the exploitation of open data capabilities, along the same line of the EC's vision on data-driven economy (Commission of the European Communities the Council, the European Economic and Social Committee, and the Committee of the Regions, 2014).

Future work should complement the ODI's vision by targeting citizens other than skilled developers, ideally in two ways. First, addressing user groups that are typically left out, such as children, disabled or elderly (technologically illiterate) people, is essential. For example, further research in educational tools for children, and accessibility of tools for various target groups, is required to enable all of them to become first-class smart citizens that are aware of their city environment and the city services provided to them, and are able to interact with them. Second, each citizen perceives, interacts with, and senses the city in distinct ways. This suggests that future research could identify and characterize how different groups of citizens perceive and understand cities. Children, elderly, workers, tourists and so on have distinct feelings, needs and perspectives of city services and city open data. The key point here is to identify the main impediments that make current open data, including cartography and geospatial datasets, not understandable and readable by these groups of citizens. This would allow to transform open data into a new type of active, customized open data maps tailored to each group's needs and characteristics to improve user experience and satisfaction.

When it comes to exploring new analytical methods to integrate quantitative and qualitative data (**C3**), one direction of investigation involves the integration of spatiotemporal quantitative measurements and predictions, with qualitative assessments about an individual instantaneous location or usual areas/periods of preferential residence. Expected results included novel analytical methods to compute quality of life indicators based on heterogeneous data sources. Another interesting research avenue is the exploration of new analytical methods to downscale coarse environmental data at city level. This implies novel methods to jointly handle multi-scale, multi-temporal data sources like official climate records with citizen-generated observations.

Predictive modeling is an attractive niche for smart cities. Typical issues in cities such as traffic and pollution can be actively managed by foreseeing possible scenarios and properly reacting to them. In this context, one interesting future research line deals with the modeling of spatiotemporal interactions based on social networks and citizens' digital footprints (e.g. GPS data) to improve the accuracy and timeliness of predictions. Concrete city applications could be predictions about the most likely crime spots and citizens mobility.

There are several opportunities for research on the adoption of open standards (**C4**). For example, there is a clear need for application frameworks for quickly creating and deploying standards-based participatory sensing applications. Such frameworks are crucial to speed up the deployment and delivery of participatory apps to citizens, thereby effectively empowering them in gathering/creating relevant sensory data. This data in turn provides valuable information for governing bodies and other stakeholders to improve city services and operations. In addition, with the increasing rate at which data is generated, the ability to have standards-driven data hubs for accessing and exposing real-time urban data streams coming from multiple sources is an interesting research avenue that may provide added value for a smart city.

The research challenge of personal services (**C5**) covers multiple aspects. We recently observe a growing interest in data privacy, especially related to location-aware applications (Damiani, 2014). In this respect, the identification and analysis of existing and potential scenarios for proximity-based opportunistic information sharing between citizens and/or city services are vital for securing privacy in personal services. Atzori et al., 2012 envision a social layer on top of the IoT paradigm that takes concepts of cooperation and social relationships for the establishment and management of social relationships between smart things. This idea could be extrapolated to determine social roles and relationships that a given device may perform as a function of its actual location (indoor or outdoor) and their relation to other nearby devices or services.

Finally, we envision further developments towards the design and characterization of persuasive interfaces (**C6**). These interfaces can deploy gamification techniques (Deterding et al., 2011) to, for example, stimulate green behavior or green living and to provide gentle but effective incentives to improve performance on a series of health and green indicators. Also, these interfaces can determine the extent to which technologies foster social changes and in behavior, and provoke subsequent action. In the context of green living, for example, it is important to monitor the behavior in space of a citizen, knowing when he/she is walking, riding a bike or driving, and to provide feedback in the form of persuasive messages about the ecological/environmental consequences of his/her actions.

2.6 Conclusion

Smart(er) cities have become a priority topic for academia, industry, government and policy makers alike, and need to be studied from a multi-disciplinary perspective. Given the number of ongoing smart city initiatives and efforts, each with their own focus, there is a risk of duplicating work if these different efforts are not aware of each other, and of the various (other) areas involved in smart cities. This article proposed to expose the outcomes of various relevant research disciplines in a simple but comprehensive manner to alleviate this risk, and used GIScience as exemplary research discipline to scope the discussion.

The paper provided a synthesis of smart city challenges, taking a citizen-centric perspective, and grouped the challenges according to research themes. We considered three research themes (i.e. empowering citizens, analytical methods and tools, citizen-centric services), with two challenges per research theme: empowering citizens necessitates tackling challenges related to deep participation and data literate citizenry; analytical methods and tools involve challenges regarding the pairing of quantitative and qualitative data, as well as the adoption of open standards; and citizen-centric services suggests more work on personal services and persuasive interfaces.

A look into the literature from GIScience has revealed that the field has already provided a number of contributions which are directly relevant to the aforementioned challenges. These include: the use of maps as a both spatial dialogue platforms, and ways of contextualizing and presenting primary data in an understandable way; the use of cellular automata as method for urban growth prediction and simulation; the use of observation ontologies for the integration of quantitative and qualitative (geographic) data; the suite of open standards developed by the Open Geospatial Consortium; the seven principles of research into location privacy; and the spatialization of user interfaces.

The article then proposed a number of future research directions, and introduced the Open City Toolkit as a way of (i) integrating the outcomes of work done along these research directions, and (ii) transferring these research outcomes (and GIScience research outcomes) to smart cities. Several research directions are currently explored within the GEO-C project, undertaken at the authors' universities in collaboration with private companies and city councils. Examples include the use of public displays as integrators of open smart cities, the identification of impediments that make current open data not understandable and readable by certain groups of citizens (e.g. elderly), research in educational tools for children to make them aware of their city environment and the city services, a participatory sensing framework to facilitate citizen participation, explore the concept of virtual meeting geo-spaces to bridge the gaps between VGI and PPGIS, and the formalization of social roles on top of nearby devices and services. We also indicated additional interesting avenues for research.

In summary, GIScience is critical to address citizen-centric challenges in smart cities. Given the breadth of topics covered by both (i.e. GIScience and smart cities), any analysis attempting at clarifying their intersection will ultimately remain limited in scope, and biased towards the research interests of the authors. The article has only scratched the surface of how fruitful the intersection of the two areas could be, and calls for further discussions complementing the views exposed.

Acknowledgments: This work has been funded by the European Commission through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement number 642332, http://www.geo-c.eu/). Carlos Granell and Sven Casteleyn have been partly funded by the Ramón y Cajal Programme (grant numbers RYC-2014-16913 and RYC-2014-16606 respectively). The authors thank Carl Schultz and Malumbo Chipofya who assisted in proofreading the article, and the rest of the GEO-C consortium (PhD students, scholars, and associate partners) for their valuable discussions over the past months on the topic of the paper.

Author Contributions: The idea of the literature review paper was developed jointly by all authors. A.D., C.G. and S.T. contributed to the formulation of research themes, opportunities from GIScience and ongoing scientific work within GEO-C. D.B. contributed to the compilation of smart cities trends, technologies, projects and initiatives, as well as to the description of the Open City Toolkit. S.C. contributed to the related work on smart and open cities, challenges, as well as the description of the Open City Toolkit. C.K. provided the representation of the proposed challenges into research themes, and contributed to formulation of the vision as well as the description of the Open City Toolkit.

Tab. 2.4: Example research directions at the intersection between GIScience and smart cities.

Research challenges	*GEO-C upcoming features beyond the state-of-the-art*
Deep participation (**C1**)	• Identifying and understanding the main motivating factors that characterise online citizen participation • Explore the concept of virtual meeting geo-spaces to bridge the gaps between VGI and PPGIS • Public displays as integrators in open and smart cities; rethink the traditional concept of map as big data analysis, cartography, and visual art
Data literate citizenry (**C2**)	• Educational tools for children to become citizen scientists • Active, customized open data maps that facilitate its full understanding to distinct groups of citizens
Pairing quantitative and qualitative data **(C3)**	• Methods to integrate spatiotemporal quantitative measurements and predictions with qualitative assessments about an individual location • Methods to downscale coarse climatic data at city level • Predictive analytics for improved citizen mobility based on social networks and citizen's digital footprints • Analysis of spatiotemporal interactions of crime data to predict crime hotspots in cities using data provided by the Web 2.0
Adoption of open standards **(C4)**	• Framework for creating and deploying standards-based participatory sensing applications • Standards-driven data hubs for accessing and exposing real-time data streams
Personal services **(C5)**	• Methods for proximity-based opportunistic information sharing and privacy protection • Determining social roles and relationships between nearby devices and/or services
Persuasive interfaces **(C6)**	• Geospatial technology and visual interfaces for green behavior and/or living • Social implications of geospatial technology and location-aware interfaces for behavior changes

3

Roadblocks Hindering the Reuse of Open Geodata in Colombia and Spain: A Data User's Perspective

This chapter was published in the ISPRS International Journal of Geo-Information as *Benitez-Paez, F., Degbelo, A., Trilles, S. and Huerta, J. (2018) 'Roadblocks hindering the reuse of open geodata in Colombia and Spain: A data user's perspective', ISPRS International Journal of Geo-Information, 7(1), p. 6. doi: 10.3390/ijgi7010006.*

Abstract. *Open data initiatives are playing an important role in current city governments. Despite more data being made open, few studies have looked into barriers to open geographic data reuse from a data consumer's perspective. This article suggests a taxonomy of these barriers for Colombia and Spain, based on a literature review, an online questionnaire, and workshops conducted in four cities of these two countries. The taxonomy highlights that issues such as outdated data, low integration of data producers, published data being difficult to access, misinterpretation and misuse of released data and their terms of use are the most relevant from the data consumer's point of view. The article ends with some recommendations to open data providers and research as regards steps to make open geographic data more usable in the countries analyzed.*

3.1 Introduction

Open data holds the promise of "*dramatically reduc[ing] the time and money citizens need to invest to understand what government is doing and to hold it to account*" (The World Wide Web Foundation, 2015). The word "*open*" can be interpreted in many ways (for a recent review, see Pomerantz and Peek, 2016), but throughout this article it is used in line with the Open Definition: "*Open means anyone can freely access, use, modify, and share for any purpose*" (The Open Definition, 2016). Providing datasets freely for access and re-use has received the increasing attention of public bodies and society who see it as a means to improve governance and stimulate knowledge-driven economic growth (Ubaldi, 2013). The concept of open data is now entering the mainstream, with 51 countries (i.e., about 25% of all countries in the world) having an open government data (OGD) initiative according to (The World Wide Web Foundation, 2015). For the purpose of this paper, we use Kučera et al.; Attard, Orlandi, Scerri, et al.'s OGD definition, as a specific subset of data which lies at the intersection of two domains: open data and government data. Empowering citizens to take full advantage of available open data, is a promising way to foster innovation and citizens-centric solutions for cities (see Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016).

The geographic community has carried out considerable effort from international to the local level, developing and implementing an integrated way to promote the sharing process of geographic data. Local Spatial Data Infrastructures (SDI) were since 1992 (Douglas D. Nebert, 2014) a way to tackle issues like standardization, integration, and accessibility shaping a framework that combines institutional arrangements, several technologies, and new policies around geodata. SDIs have been only attractive for limited and specialized geographic communities, and as discussed in (Diaz et al., 2012), could benefit from existing trends (one of these being open, distributed and linked data).

Opening up data is valuable, but using available open data to provide useful services to citizens is equally important. According to Andrus Ansip, Vice-President of the Digital Single Market of the European Commission, "*Data should be able to flow freely between locations, across borders, and within a single data space. In Europe, data flow and data access are often held up by localization rules or other technical and legal barriers. If we want our data economy to produce growth and jobs, data needs to be used. However, to be used, it also needs to be available and analyzed*" [1]. Along the same lines, Janssen et al., 2012 stated: "*Open data on its own has little intrinsic value; the value is created by its use. Supporting use should not be viewed as secondary to publicizing data*". Previous work has investigated various aspects of Open Government data initiatives. These aspects include a business model for Open Government data (OGD) (Ahmadi Zeleti et al., 2016), a measurement framework to quantitatively assess the quality of OGD (Vetrò et al., 2016), an index to measure the maturity of e-government openness Veljković et al., 2014, the use of semantic application programming interfaces as a way of improving access to OGD (Degbelo, Trilles, Kray, et al., 2016), and the motivations of citizens to participate in OGD projects (Wijnhoven et al., 2015), to name but a few. Complementary to these, this work takes a user-centric view, and investigates barriers faced by people when interacting with existing open data portals. The novel contributions of this article are as follows:

[1] http://europa.eu/, accessed July 07, 2017

- Users' reuse barriers in Colombia and Spain. While the literature has listed some challenges with respect to open data and open data portals (e.g., in Janssen et al., 2012; Lourenço, 2015), asking users what actually hinders them has been less often undertaken. Some studies come close to what the current article tries to achieve, but differ substantially either in their method or scope. For example, Lourenço, 2015 did an analysis of open data portals from an 'ordinary citizens' point of view', but his analysis did not rely on inputs from actual citizens. John B. Horrigan and Rainie, 2015 survey Americans' views, Beno et al., 2017 analyzed the Austrian context, and Schmidt, Gemeinholzer, Treloar, et al., 2016 aimed at being global (with participants from over 80 countries), but this article aims at being local and geographic focused - with a focus on Colombia and Spain - and gathers its empirical evidence through both surveys *and* workshops. This work took four cities within two countries as use cases.

- an empirical evidence to validate barriers identified previously in the literature. The empirical evidence is no basis for a validation of barriers from the literature in general. Rather the light shed in the paper contributes to make some conclusions as regards to the four cities examined, namely Bogotá (Colombia), Medellín (Colombia), Cali (Colombia) and València (Spain). In particular, currency, accessibility, terms of use, and data quality were recurrently mentioned by the participants as obstacles. This suggests that these four barriers still deserve close attention from research, and data producers (at least in the cities examined).

The barriers were compiled using three sources: a review of the existing literature on open data, an online survey with 195 participants, and a set of workshops with a total of 155 participants. The participants were developers, analysts, journalists, students, open data experts, professors, politicians, and users with geographic background from cities in Colombia and Spain. The main data users' barriers identified were categorized into six groups: 1) currency, 2) discoverability, 3) accessibility, 4) terms of use, 5) usability, and 6) data quality. The remainder of the paper is organized as follows. Section 3.2 reviews related work on Open Government, barriers from a data producer perspective, as well as open geodata reuse issues. The research method used in this work to compile and validate barriers to open data reuse from a data consumer's perspective is described in Section 3.3. Section 3.4 lists the main findings from literature review, the survey conducted, and a set of participatory workshops. This section also introduces a data user's taxonomy using a fishbone diagram. Section 4.3.1 discusses the results obtained, and the paper ends in Section 3.6 with recommendations for data authorities of the cities surveyed.

3.2 Related Work

This section points to concepts from previous work taken into consideration for this research. It touches upon Open Government data, the value of open data reuse, as well as the relevant role of open geodata reuse and the implications to include data users' perspective at a local level.

3.2.1 Open Government Data

Defined by Kučera et al., 2013 as government-related data that is created and published in the way that meets with the Open Definition (The Open Definition, 2016). OGD is seen as a current trend and a key factor in cities with intersection with Open Data initiatives. Ubaldi, 2013 defined OGD as the combination of government data (as any data and information produced or commissioned by public bodies) and Open Data (data which can be freely used, reused, and distributed by anyone). Meanwhile, in order to understand the different meanings of OGD from bureaucratic, political, technological, and economic perspectives, Gonzalez-Zapata and Heeks, 2015 used the definition of Yu and D. G. Robinson, 2012 through three main foundations—open, government, and data, illustrating three intersection points to determine what OGD means. The result of this combination is government data, open data, and Open Government (actions and government decision-making process should be transparent, collaborative, and participative).

In general, local and national administrations, civil society organizations, the private sector, and overall several stakeholders are taking advantage of the intersection among open and government data. The impact of this combination can be positive in many ways. For instance, the continuous online access to government data is positively associated with knowledge absorption according to J.-N. Lee et al., 2016, who indicated that government data openness could positively affect the formation of knowledge bases in a country; therefore, the level of knowledge base even positively affects the global competitiveness.

However, not only the level of knowledge has been identified. When Open Government initiatives are on the table of public agencies, the expectations to improve the governance processes are certainly high. Moreover, increasing transparency, expanding the public engagement, and improving responsiveness and accountability are the desired goals of most governments. However, determining whether the Open Government initiative is effective or successful could be a challenge for many public agencies. The ambitious aims and expectations of these sorts of initiatives could lead to some failed activities that yield some immediate success but then run the risk of losing steam over time. Identifying the participant, their roles, and including them in the current Open Government initiatives and the way that data is released is illustrated by Williamson and Eisen, 2016 as the key to successful Open Government initiatives. Through a rubric of six questions, Williamson states that even high accessibility levels and well-publicized data are not enough to transform the government processes if people or participants do not have channels to influence it.

Due to the major role of participants, a better understanding of citizens' motivators for engaging in Open Governments actions could guide the current initiatives to get the expected outcomes. Wijnhoven et al., 2015 demonstrated that when citizens feel that their contribution is significant, they are more open for contributing in Open Government projects. Besides, Wijnhoven et al., 2015 found that there is no evidence to suggest that socio-economic factors could affect the participation in those Open Government projects, whereby projects that appear to be well-implemented have a better reaction from citizens than others that only focus their attention on some stakeholders. In the same direction, a well-detailed description and the way in which data is shared could have a positive impact on the participation of citizens. E. Afful-Dadzie and A. Afful-Dadzie, 2017 collected the preferences of media practitioners in five countries in Africa, and observed that online journalists see metadata as the most important factor in a functioning OGD, followed by data format and data quality.

In particular, there are some factors that can influence the level of success of the implementation of Open Government initiatives, especially for authorities that require a solution beyond the current administration. H.-J. Wang and Lo, 2016 looked into some of those factors in Taiwan; their study disclosed that perceived benefits, organizational readiness, and external pressures have a positive effect on OGD adoption. On the contrary, perceived barriers seemed not to have any significant effect on OGD adoption.

Overall, the participation and interaction with the general public, the identification of their needs, and the sustainability of OGD actions are a particular focus of the current studies. Beyond accessibility, the scope is now moving to determine factors and barriers to using or reusing the available public sector data (e.g., crime rates, gas emission, mobility, air quality, or security). Since many available datasets and likewise many cities are in the middle of the implementation stage of open data government initiatives, the aim of authorities is to motivate users to reuse the published data and create a new bunch of services, generating value for data-opening projects. Local experiences of cities, local needs, and kinds of data could bring a differential factor in the way to reuse the available OGD. In the following sections, we will discuss why the reuse of open geographic data is becoming necessary in current open data initiatives and why barriers from a data users' perspective are preventing full advantage being taken of the data release process in which many cities are involved.

3.2.2 Why is the Reuse of OGD Necessary?

Beyond access to OGD, the creation of value is perhaps the most interesting part of open data systems, in which economics, social, and political benefits are being established in local governments. Official entities are trying to increase the transparency of their processes and empower their citizens by publishing a vast list of relevant data. Ubaldi, 2013 provide a work about OGD, in which a list of commonly recognized main beneficiaries of OGD can be found, where the wider economy, the private sector, and the public service marketplace provide the opportunity to increase the innovation expected by official authorities. Access to data by itself does not offer new services or make a difference with other private data provider companies, per se. New value-added services must come in addition to data to bring more opportunities to developer companies to pursue the commercial exploration of OGD. This commercial approach and a new bunch of value-added services are possible when data is reused (Carrara, Vollers, et al., 2017b).

Assuming that greater openness automatically creates value (Ubaldi, 2013), which is a common mistake in many governments. The OGD systems should include the value chain as part of the initiatives, where conditions to develop value-added services and indicators to measure the impact of released data are included as a relevant part of the systems. Local governments and cities overall have an essential role to play in the value chain. Combining published data with data user communities, local authorities are not only playing the role of providers; they also become a partner, facilitator, convener, and enabler of easy reuse. At the same time, empowering the data user communities that have to tackle local issues and deal with reuse barriers on a daily basis, the local level could be the key to transforming the current actions into concrete results. This integrated scenario is only possible when data authorities behind open data initiatives incorporate the reuse as part of their priorities.

Although OGD is a common topic in local governments, most of them have not understood the benefits and value of open data, but mainly the expected benefits for cities' stakeholders. T.-M. Yang et al., 2015 illustrated that data authorities should not only consider data users as the general public, but also their internal departments and other agencies could be beneficial to make the data more reusable and discoverable. OGD actually offers an opportunity for local agencies to carefully survey and identify what datasets they have, which are the most used ones, and what they can share with other departments to improve internal collaboration. Thus, local governments need to educate and empower not only the general public; the first step should be to promote the open data initiatives inside their departments. T.-M. Yang et al., 2015 presented the concept labeled interagency as a foundation of OGD which is the positive impact on cross-boundary information sharing among cities' agencies, where the continuous information sharing is a spiral process to reinforce the communication and at the same time reinforce the OGD initiatives.

3.2.3 Does geographic data has a role to play in open data times in cities?

In a value chain where data user needs are fundamental and cities are a relevant piece of the OGD initiatives, the kind of data also has a role to play. Considering the foundations of OGD developed by Yu and D. G. Robinson, 2012 (data, government, and open), data is also a large concept that could be considered from a specific point of view to identify data users' requirements. The nature of data can influence future barriers, needs, and strategies of OGD initiatives. Based on the concept of Yu and D. G. Robinson, 2012—the intersection of data seen as geographic data (data with a spatial or geographic component), government initiatives, and the definition of open—could be more efficient and interesting in light of city data users' requirements. However, is geographic data a relevant type of data that might bring more effective benefits in local government, and why can the reuse of this kind of data support authorities in their engagement strategy?

According to the *Reusing Open Data* report of the European Data Portal (EDP) (Carrara, Vollers, et al., 2017b), geographic data (25.8%) is the second category only surpassed by the statistical (27%) category that is most reused and also consulted by companies (among 128 domains mentioned) that try to generate revenue from open data reuse in EU member states. This report (Carrara, Vollers, et al., 2017b) also illustrates a strong correlation among open data categories, where "*region & cities*", "*transport*", "*environment*", and "*population & society*" suggest a trend of organizations using those categories together.

Geographic data is considered as one of the most economically relevant data domains for its high demand from re-users across the EU, according to the analytic report of EDP (Carrara, Vollers, et al., 2017a). In (Young and Verhulst, 2016), the impact of current open data was analyzed using 19 use cases around the world. Three involved the geospatial sector, with public authorities in Denmark, Great Britain, United States. Additionally, more use cases geographically related in Singapore, and Uruguay where the impact is assessed in terms of improving services, economic growth, and data-driven engagement. Geographic data has specific characteristics that also demand specific needs from data users; therefore, the identification of those requirements might contribute to improving the OGD initiatives in cities.

In general, there is a great deal of published work regarding the reuse of open data and why it is one of the challenges for current initiatives. Barry and Bannister, 2014 have selected the occurrences of some themes surrounding open data, mentioning that data sharing and reuse are two themes with a high number of occurrences, demonstrating a focus on making the most of the resource of public sector information. Literature has analyzed economics (P. A. Johnson et al., 2017; Tornhildur Jetzek et al., 2012; Ahmadi Zeleti et al., 2016), technical, institutional (T.-M. Yang et al., 2015; Cranefield et al., 2014), political and policy (Nugroho et al., 2015; Thorhildur Jetzek et al., 2013) factors that influence the value chain of open data, suggesting that theoretical benefits have not been seen as cities' stakeholders expected (Carrara, Vollers, et al., 2017b; Janssen et al., 2012; Conradie and Choenni, 2014; Barry and Bannister, 2014; Attard, Orlandi, Scerri, et al., 2015; Cranefield et al., 2014).

Regarding the overlay of initiatives between national and local efforts, both working to improve their Open Government's efforts, in many cases datasets are offered on several websites (T.-M. Yang et al., 2015) in a fragmented way, which is in some cases difficult to find. Adequate metadata are also necessary to improve data reuse (Janssen et al., 2012).

3.2.4 Barriers to Open Government Data Reuse

Much published work is related to open data and desired benefits that this trend might bring to governments and its stakeholders. There are several authors (Barry and Bannister, 2014; Conradie and Choenni, 2014; S. Martin and Foulonneau, 2013; Janssen et al., 2012; Beno et al., 2017; Cranefield et al., 2014; Access Info Europe, 2010) who have worked on open data barriers from different perspectives; most of the work done has been focused on national governments, OGD initiatives applied for data producers, integrators, or suppliers. Beyond promoting a sustainable reuse of Open Government data in cities, a constant and circular reuse should be considered in OGD initiatives.

In Janssen et al., 2012, a set of benefits, adoption barriers, and five myths of open data initiatives are defined; most of them are still present in current initiatives. For instance, myth number five is about open data and the incorrect interpretation that will result as Open Government. In Section 3.2.1 it was explained that releasing open data is only the first stage in getting the expected benefits of Open Government, especially collaboration and participation. The process can only start when the published data is used.Janssen et al. also suggested a set of adoption barriers from a national government perspective; however, at the local level, barriers, data user communities, and even the expected benefits may vary. Likely, national and local levels are both pursuing the improvement of accessibility, legal issues, and technical integration concerns, but the contact with data users could be easier at the city level. Factors such as reuse, feedback, channels to influence, and integration requirements create a solid way to work towards the benefits mentioned.

Barry and Bannister, 2014 also worked on the definition of open data barriers when the data is published from a data integrator perspective. They took Ireland as a use case, creating a detailed comparison among the current literature about open data barriers, and proposed a new barrier schema as a taxonomy of release barriers from senior managers in this country.

T.-M. Yang et al., 2015 suggested factors that could reduce the possible impact of published data, using several authorities in Taiwan as a use case. Thus, their work presents those factors as barriers from data producer perspective and at the national level. Another related paper about barriers—but at a local level—was published by Conradie and Choenni, 2014. They found that the ways in which data is stored, obtained, and used by local departments are crucial indicators of open data release. Conradie and Choenni, 2014 suggest taking small incremental steps to explore and learn about the data release, avoiding releasing data for political or internal purposes.

In the literature, there are also some reports created by the European Commission and its project EDP; the initial and related report taken into account is the reuse of open data (Carrara, Vollers, et al., 2017b) from a business perspective. This report presents a study of several companies—most of them from the private sector—around their business model built using open data. It lists a set of factors that European countries or corporations should consider to promote the reuse of open data. Internal and external barriers that do not allow the standardization and automatization of open data are defined, and at the same time, some recommendations for the public and private sectors are illustrated.

Another report related to the last one by the European Commission is the fifth analytic report of the EDP [2] Carrara, Vollers, et al., 2017a, where barriers are seen as a core of the problem to reusing open data, basically from two perspectives: data producers or suppliers and data consumers. However, this report is based on the same findings as the reuse and maturity level report Carrara, Nieuwenhuis, et al., 2016 that the European Commission studied as well. The description of the barriers are listed according to the national level in the EU28+ countries of Europe and their open data initiatives.

[2] https://www.europeandataportal.eu, accessed 14 July, 2017.

3.2.5 Geographic data reuse barriers and the importance of data users' perspective

The above-mentioned report (Carrara, Vollers, et al., 2017a) showed an insight that is relevant to this research—the role of the geospatial domain in the open data movement. Presenting some barriers from a data producer perspective, the authors discuss why geospatial data plays a major role in an open data strategy for any country.

Many of the identified barriers to improving the reuse level of OGD has been already tackled from the geographical community several years ago before the open data movement has started being recognized by public administrations and research field. In cities but especially in countries issues like standardization, accessibility or integration of several data sources has been a constant headache for many geographic institutions. Since 1993, the term SDI was coined by the U.S. National Research Council to define a framework of technologies, policies, and institutional arrangements working together to facilitate the creation, sharing, and use of geospatial data and related information resources across an information-sharing community (ESRI, 2010; Douglas D. Nebert, 2014). Such a framework can be implemented at local, national, regional or even international levels to allow different stakeholders have the effective and easy access to official, high data quality, and standard geographic information. Taking into account the important role of cities, Harvey and Tulloch, 2006 presented a typology of local-government data sharing arrangements in the US in times where the local SDI was moving to a second generation. Harvey and Tulloch, 2006 suggested that political, institutional and economic factors need be considered in local governments to guarantee the effectiveness of the sharing-data process and likewise a continuous reuse of geodata in cites.

Janssen et al., 2012 suggested the creation of open data infrastructure as a possible way to guarantee a constant support around all political, institutional and even technical issues that are involved in the sharing data process. At the same time, current local SDI projects have a significant challenge regarding the way that geographic data user communities are using and re-using the available data, leading both projects with a common problem, which could be tacked working together. Both in local open data initiatives and local SDI, the role of data user is fundamental, a better understanding of their needs or requirements could be the key factor to refine the current initiatives and find the way to be more effective.

There are few authors consider a data user's perspective at the local level. In (Zuiderwijk, Janssen, and C. Davis, 2014) work a particular emphasis on the components of the open data ecosystem where users' pathways reveal the direction of how open data can be used, then the initiative can use this direction to move towards data users' requirements. Based on the work's conclusions, three aspects are especially salient: 1) More and clear information related to license or terms of use. Data re-users get confused more often than data producers think (despite the fact that the terms of use are included in most of the open data portals). Additionally, most of them are difficult to read. 2) More statistical and geographic context. This means that raw data is important and is considered as a requirement to consider published data as open (Sunlight Foundation, 2010); however, it is necessary to include statistical and spatial relationships to guide users to understand what this data is about. Including comparison with other regions, or neighborhoods with different geographic features, a comparison during the time or even with the inclusion of basic statistics, published data can reach more users' attention offering an enriched perspective, than only a list of downloads. Finally, 3) Feedback for both data providers and data users; providing ways to discuss, both sides can learn and enhance the value of available data. This component is likely one of the most forgotten resource in current open data initiatives, where the feedback resource is limited to email contact or a questionnaire to end-users to express some issue. Only few open data portals have a proper systematic way to discuss issues, use cases, best practices or suggestions from end-users, and also show updated data or features to their community.

3.2.6 Summary

To summarize, previous work—using interviews, surveys, workshops, or sets of references—has identified a set of barriers mostly from a data producer point of view, where national authorities are having the main role of open data initiatives. Regarding the reuse of open data, there is not too much work done; we found only four related references, none of them have considered the possible potential of geographic data, or the role that cities can play. The number of articles that examine reuse obstacles from a data user's perspective is also limited. As a contribution of this work to address this gap, we presented a taxonomy of barriers experienced by data users in four cities.

3.3 Research Method

Many authors (Zuiderwijk, Janssen, and C. Davis, 2014; Ubaldi, 2013; Barry and Bannister, 2014; Conradie and Choenni, 2014; Open Data Institute, 2015) have mentioned that the potential value of open data is in its use. The re-usability and discoverability levels of open data at local levels are critical factors to truly make an impact through the city stakeholders. The main research question addressed in this article is: *what barriers prevent open data reuse by data consumers?*. This research took place from August 2016 until May 2017 based on multiple use cases and a combination of structured online survey and hands-on activities (i.e., participatory workshops). The research covers data consumers' barriers from three angles: a literature review, a structured online survey (what people say), and outcomes from a set of participatory workshops (what people do). The identified barriers from these three angles are summarized using a taxonomy. This taxonomy presents six obstacles to the reuse of open geographic data in cities. It can be used to inform data authorities about weaknesses of current city open data systems, thereby enabling them to design better and more effective strategies to improve the reuse of their data. This taxonomy is presented in Section 3.4.4.

This research took four use cases, with local authorities in the three principal cities in Colombia (Bogotá, Medellín, Cali) and the third main city of Spain (València). Initially this research studied the current status of their open data initiatives, considering that cities have different Open Government data approaches (Beno et al., 2017; Williamson and Eisen, 2016). To enrich the discussion and reduce a possible bias of the findings considering only one city, the selected cities have distinct progress and perspectives from legal, technical, institutional, political, and awareness points of view (see Table 3.1). Beno et al., 2017 worked in the delimitation of barriers to use open data in Austria at a national level, and claimed that "*caution must be applied as the findings might not be transferable to other countries*", because there may be differences in terms of maturity of their open data "*culture*" and the datasets that official authorities offer. The available datasets in each city have an important role to understand possible frictions to use or reuse the data in each city. All selected cities have their own data portals. Valencia and Medellin have a central portal called "Transparency and open data portal [3]" and "OpenData Alcaldía de Medellín [4]" respectively, with considerable number of web services, mostly are geo-services, related to several city domains like mobility, education, environmental, urban planning, demographic and culture. In general topics that each local authority considered relevant to users and the city. Another aspect that also contribute to the diversity of the selected cities is the current role of the local authorities contacted in terms of open data "*culture*" in each city. Initially, both Bogotá and Cali were contacted by the local SDIs, whose principal objective is to facilitate the production of and access to geographic information in the city, thus placing the importance of open geographic data considerably high. On the other hand, in Medellín and València the authorities contacted were the City Halls, where the open data initiative is assessed and created in terms of Open Government; therefore, geographic information is taken as another type of data, and its relevance is moderately less than in Bogotá and Cali.

[3]http://gobiernoabierto.valencia.es/en/data/, accessed November 11, 2017.
[4]https://geomedellin-m-medellin.opendata.arcgis.com/, accessed November 11, 2017

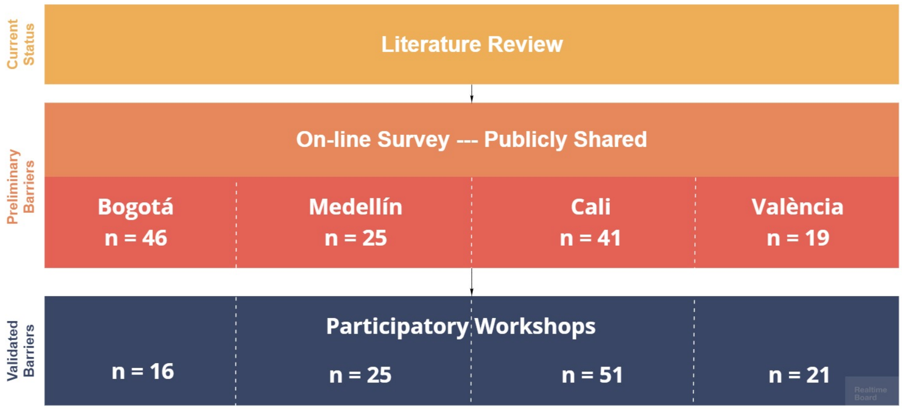


Fig. 3.1: Workflow used. Literature review, identifying initial barriers and perceptions, then a publicly shared online survey (n = 195 valid responses). Using the preliminary results, four cities were selected to reply to the initial survey, and the identified barriers were contrasted with the participatory workshop in each city.

The combination of local official authorities, data user communities, and open data experts allow this research to take a bottom-top view of the open data schema to understand what the data users' requirements are and their contributions to improving the reuse of open geographic data in cities. As it was mentioned by Nugroho et al., 2015, a better relationship between local data authorities and data user communities stimulates the provision of data and increases the involvement of data users.

The local authorities contacted were, Bogotá SDI (Infraestructura de Datos Espaciales para el Distrito Capital, IDECA [5]), Cali SDI (Infraestructura de Datos Espaciales de Santiago de Cali, IDESC [6]), City Hall of Medellín [7], and City Hall of València [8].

The next two subsections provide some background information about the online survey and the conducted participatory workshops. Figure 3.1 displays an overview of the steps that this research took to collect the barriers identified and understand each open data initiative in the selected cities.

3.3.1 Literature Review

A literature review of open data barriers was conducted by collecting journals, conference papers, and governmental or non-governmental reports in several databases: Science Direct (eight papers related), Scopus (four papers related), and Emerald Insights (eight papers related). The words used to find related articles were, *'barriers*

[5] https://www.ideca.gov.co/, accessed July 11, 2017.
[6] http://www.cali.gov.co/planeacion/publicaciones/3560/idesc/, accessed July 14,2017.
[7] https://www.medellin.gov.co/irj/portal/medellin, accessed July 14, 2017.
[8] http://gobiernoabierto.valencia.es/en/, accessed July 14, 2017

in open data', *'barriers in Open Government data'* and *'barriers in Open Government'*. Only papers that addressed barriers, challenges, issues to reuse, adoption, and releasing data were taken into account. Additionally, use was made of the cited references in papers where barriers were identified in order to enrich the discussion and literature review. The number of articles was filtered by year, choosing only articles from the last five years (2012 to present) in order to have a current approach, and only journals related to governments, open data, geography and economics were taken into account. The literature review was classified in two groups: barriers from data producers' and users' perspectives. In total, 12 relevant papers were selected and related to barriers to reuse. The relevance of those papers was determined by scanning and manually reviewing their title and abstract. These related papers can be found in Tables 3.2 and 3.3 in Section 3.4.

3.3.2 Online Survey

Taking into consideration the potential data users' barriers obtained in the literature review (Section 3.3.1), and considering that citizens access to data through the official open data portals, an online survey was designed with the public Google Forms web application. The survey aim was to know the barriers, errors, or problems that users have encountered while using cities' open data portals and its shared datasets, especially geographic data web services. The questionnaire was released in three different languages (Spanish, English, and Portuguese) to gather more responses from several cities. The survey was a modular form with seven sections, including general information about the respondents (working country, city and age), their work (employment role and industry), perception of open data, possible barriers faced, most-used features in well-known cities' open data portals, and finally method(s) used to find open data in a city—especially geographic data. The survey took about five minutes to complete, and was anonymous (i.e., no information about the name of the participant or email was collected). Participation in the survey was voluntary, and it was not necessary to answer all questions. The appendix 3.6.2 presents the questions formulated in the survey. For this research, only questions related to reuse barriers and most used features in cities' open data portals were included in the analysis (see section 3.4.2).

The survey was launched in August 2016 and remained active until December 2016. The survey was shared in several ways: 1) Through social networks (e.g., Facebook, Twitter, Linkedin), 2) E-mail lists, and 3) Several open data and smart cities events during spring–winter 2016 (e.g., International Open Data Conference

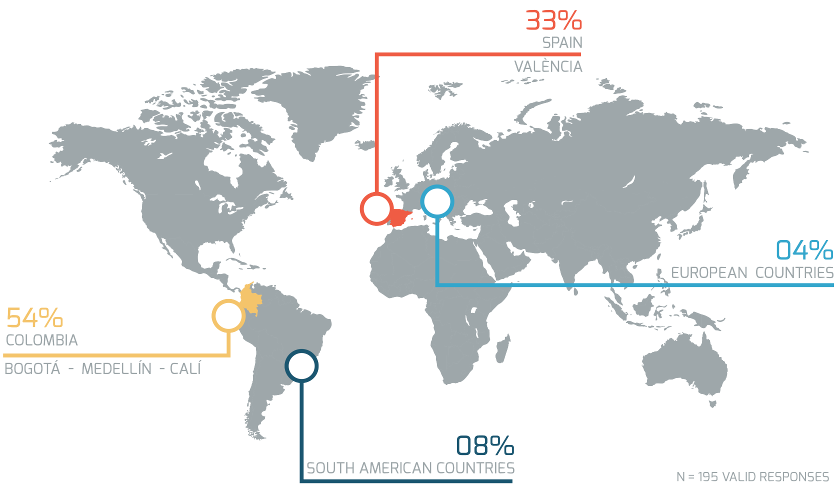


Fig. 3.2: Online survey responses

2016[9], Open Cities Summit 2016[10], Inspire 2016[11], Geo Mundus Conference 2016[12], Data Latam 2016[13], Esri User Conference 2016[14], Esri Spain User Conference[15], and Esri Colombian User Conference 2016[16]). The survey received replies from data users from cities in South America and Europe, but especially cities in Colombia and Spain (see Figure 3.2).

Overall, a total of 195 participants completed the survey. However, some of them did not completely answer the questions; therefore, some questions have a smaller sample. Only responses that were fully completed were considered. Concerning the employment role (n = 195), 25% (48) of participants saw themselves as geographic data analysts, and 19% (n = 37) as part of academia (e.g., professor, researcher, or student). It could be argued that the high prevalence of participants with a geographical background and from academia was due to the way that survey was promoted with university colleagues that helped to distribute the survey and organizations that work with geographic data. Regarding managers and project leaders, about 18% of participants (a third of respondents) were part of this group of open data users. About 17% (n=33) saw themselves in multiple roles, as developer and analyst at the same time. Finally over 21% of participants were developers of any type of application, exclusive geographic developers, or had a different role (see Figure 3.3).

[9] http://opendatacon.org/, accessed July 14, 2017.
[10] http://opencitiessummit.org/, accessed July 14, 2017.
[11] http://inspire.ec.europa.eu/, accessed July 14, 2017.
[12] http://geomundus.org, accessed July 14, 2017.
[13] http://www.datalatam.com/, accessed July 14, 2017.
[14] http://www.esri.com/about/events/uc, accessed July 14, 2017.
[15] http://conferencia.esri.es/, accessed July 14, 2017.
[16] http://esri.co/esri/, accessed July 14, 2017.

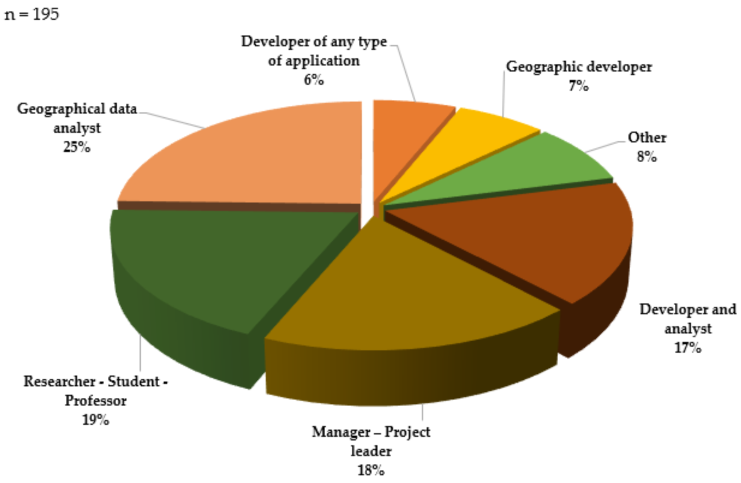


Fig. 3.3: Employment role of respondents.

3.3.3 Participatory Workshops

Having responses from an online survey is not sufficient to understand the whole picture of open data users' issues regarding use or reuse. To shed some additional light on users' barriers, participatory workshops were conducted with participants from different backgrounds. The participatory workshops were called Open Data for Open Cities. Figure 3.4 shows that the participants in these participatory activities were developers, entrepreneurs, analysts, journalists, professors, researchers, open data experts, or data authorities who also consider themselves as data users (Zuiderwijk, Janssen, and C. Davis, 2014). During this stage the workshops aim was to observe, confirm the mentioned barriers in the survey, allowing to data users to express their concerns to effectively use or reuse of the available datasets in each city through the open data portals that they consider relevant for their external application or analysis. Bringing together the data user profiles that have been working in the same city give this research a broad view of the current data user barriers at a local level. To consider the cities with the most collaboration (see Figure 3.2), this research has chosen the aforementioned cities (see Section 3.3) for the workshops. Likewise, two more workshops were conducted in Castellón de la Plana, Spain and Wageningen, The Netherlands with students of a Master's Geographic Information Science (GIScience) (33 participants) and open data experts (11 participants) in order to have better insight into barriers faced by geographic data users.

The participatory workshops lasted approximately four hours, split into two-hour sessions. The initial session was about finding suitable city open data in the official or well-know open data portal, using the main data domains defined by the contacted data authorities (e.g. mobility, education, urban planning, air pollution, crime, and others) depending on the priority of each city (see Table 3.1). Participants were

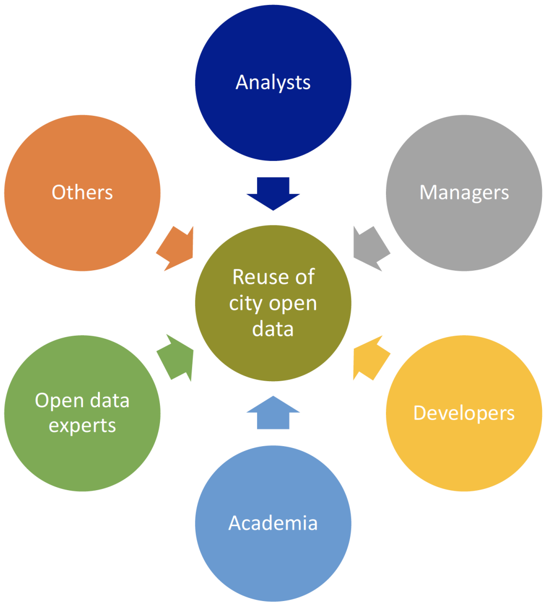


Fig. 3.4: Workshop participant roles. There were 56 from academia, 49 analysts, 20 managers or project leaders, 11 developers, 3 Open data experts and 16 counted as others that include politicians, journalists, entrepreneurs and others roles. In total there were 155 workshop participants.

required to create groups with three members at most, then choose one category which they found more interesting or related to their work. Once the groups were created and the category was selected, participants were required to think about a general idea, analysis, or application that included published data of each city. During the mentioned first session participants were looking and evaluating datasets and its properties to use or reuse for external projects. During this activity, participants were able to bring their own laptops or use computers provided by the organizers with access to the Internet. Details of this research were not included at the beginning of the activity in order to reduce a possible bias presenting the initially identified barriers in literature or the online survey. Participants were able to use any search method that they considered appropriate to find the open data in each city (e.g., search engines like Google, official open data portals, or any web portal). The second session was a discussion, where participants could express all the found barriers to reuse, their requirements, or common issues when they need to include open data in their work. Local authorities were also part of this discussion, but only obstacles from a data user point of view were collected.

Table 3.1 illustrates the general aspects of each selected city, such as population, context, current data policy regulation, main data thematics, terms of use, and identified users. In addition to the workshops mentioned above, there were two participatory workshops in different cities with other kinds of participants.

During the workshop, a follow-up questionnaire was administered regarding methods used to find data, barriers found, and users' suggestions to overcome the obstacles found. Participants' personal information such as name, email, and organization were collected in analog form, but not in a way that would identify them personally. This information was used to share the workshop report with participants to explore the results and insights collected. Participation in those workshops was also voluntary.

3.4 Findings

In the previous section, the method used in this research was presented as involving three sources: the literature, the online survey, and the participatory workshops. This section elaborates on findings, taking the same sources into consideration. Section 3.4.1 presents findings from the literature review, Section 3.4.2 shows the results of the online survey, and Section 3.4.3 synthesizes what people did and discussed during the participatory workshops in selected cities.

3.4.1 Findings from Literature

Much of the documentation that this research has reviewed referred to three aspects. 1) Benefits of Open Government implementations, through several countries, explaining the ways to reach the economic, social, or political benefits of releasing government data. 2) Regarding barriers, challenges, or issues in the literature, there are mainly two categories of open data barriers: from data producer or data user perspectives. 3) Most of the published papers discuss national governments, but the local governments are briefly mentioned in addition to the possible barriers from the data user point of view. The key findings from previous work are mentioned next.

S. Martin and Foulonneau, 2013 demonstrated through local and national cases that the sustainability of the open data initiatives needs to be considered regarding risks, challenges, and limitations, having in mind the evolution of the stakeholders involved (re-users, data creators, and national aggregators). Related to the role of users in open data systems, Janssen et al., 2012 suggest that feedback and insights from this point of view must be considered in order to continuously improve. Janssen et al. also established a list of adoption barriers of open data, presenting "Use and Participation" as part of those obstacles in the open data implementation process. Barry and Bannister, 2014 and Conradie and Choenni, 2014 consider the process of releasing open data as the center of attention in Open Government initiatives; they examined the barriers to open data release at national and local levels from the perspective of senior managers and six local public sector organizations.

Tab. 3.1: General aspects of selected cities data authorities. IDECA: Infraestructura de Datos Espaciales para el Distrito Capital; IDESC: Infraestructura de datos espaciales de Santiago de Cali; SDI: spatial data infrastructure.

	Bogotá	**Medellín**	**Cali**	**València**
Country	Colombia	Colombia	Colombia	Spain
Population	8.080.734 inhabitants (2017)	2.508.452 inhabitants (2017)	2.420.013 inhabitants (2017)	790.201 inhabitants (2016)
Context	The most populated and capital of Colombia	The second most populated city of Colombia	The third most populated city of Colombia	The third most populated city of Spain
Autority(ies) Contacted	SDI Bogotá (IDECA)	Medellín City Hall and Ruta N	Cali City Hall and SDI Cali (IDESC)	València City Hall and Las Naves
Main Open Data Theme of Interest	Urban Planning, Economic Development, and Infrastructure	Security, Environment and Urban Planning around a sustainable smart city strategy	Mobility, Security, and Health	Environment, Transport, Society, and Wellbeing are the themes more used and consulted of the Open Data catalogue
License or Terms of Use of Open Data	IDECA license	License Attribution-Share Alike 4.0 International	No open data license, only IDESC web site terms of use	All the data sets offered by the City of València, unless otherwise indicated, are published under the terms of the Creative Commons license - Recognition (CC-By 4.0)
Open Data Portal or Official Portal	IDECA website	GeoMedellin Website	IDESC website	València Open data website
Current Engagement Activities	Strategies implementation to facilitate the discovery, use, and reuse of available open data.	Engagement activities with the community and identified users. Creation of the platform of open data, dynamic visualizations, and analysis with the data of the different dependencies of the Mayor's Office of Medellín.	Create channels of communication with citizen initiatives related to open data in the city. Promote the publication of open data of utility by the agencies of the Mayor of Cali. Promotional events for the open data available in Cali.	The position of the City Council in relation to Open Government is that the technologies serve for the citizens to have more knowledge of municipal action and to make possible participation and collaboration with the management of the city; actively listen to citizens in social networks or any other media. They also work on the creation and application of standards as well as the use of transmedia to bring important issues to citizens.
Developer Companies Identified as Open Data Users	A few companies identified. Note that this identification is not done periodically.	There was one company identified	There were three companies identified	The policy of the City Council in Open Government, does not see as relevant to collect data of entities or individuals who have used the datasets.
Universities or Colleges Identified as Open Data Users	There were several universities identified	There are several universities identified	There were several universities identified	Public Valencian universities collaborate with the city council in organizing activities and events on open data
Internal and Official Authorities Identified as Open Data users	There are 73 local entities integrated and identified	City Hall, Metropolitan and regional authority	Utilities, Transportation, Urban planing and Environmental, and Economical authorities	Representatives of the regional government have collaborated in some of the events of the Open Government Chair with the Polytechnic University of València, and both policies—local and regional.
Urban Observatories or Analysis Groups Identified as Open Data Users	Several urban observatories were identified	Only one Urban observatory was identified	Several urban observatories were identified	N/A
Others Identified Open Data Users	Several cities stakeholders considered relevant	Several cities stakeholders considered relevant	Several cities stakeholders considered relevant	N/A

H.-J. Wang and Lo, 2016 examined three factors that influence the adoption of OGD, where perceived benefits of OGD are more significant than other determinants of OGD; however, looking into the perceived barriers in this work, the participants mentioned data findability, personal privacy, data layout, and licenses as potential barriers of OGD adoption in their organizations. However, not only the official governments have been consulted, in (Schmidt, Gemeinholzer, Treloar, et al., 2016), global environmental data research and data infrastructure communities were considered in a survey to highlight users' perceptions in terms of open data, and also barriers to share data. The survey revealed that "paying for data", "varying degrees of data quality in different datasets", and "varying standards in how data is gathered" are seen as the most significant burdens. Attard, Orlandi, Scerri, et al., 2015 presented a systematic review of OGD initiatives describing 15 challenges where citizen participation is an essential factor to promote innovation among developers and other stakeholders. However, a number of barriers prevent public participation—most of them are included as cultural challenges in this work.

In terms of the private sector or organizations that have the skill to transform open data in a new bunch of innovative services is likewise a relevant group of users considered in the literature. Although Beno et al., 2017 mentioned that the barriers faced by the private sector have not been sufficiently studied, the EDP project performed a study with 76 organizations across Europe (Carrara, Vollers, et al., 2017b) to understand how they use open data and what business models have been developed based on the reuse of the available data, finding that there is a mismatch between the available data sets that public organizations are releasing and the data sets that are most reused. Meanwhile, another report also from the European Data Portal project (Carrara, Vollers, et al., 2017a) presents a set of barriers faced by open data suppliers and users considering the study above in EU28+ countries. For open data publishers, the most frequently encountered obstacles are financial and legal; however, for re-users of open data, lack of awareness and low availability are the barriers most mentioned in this report. An important remark of this report is that geographical data is counted as a technical barrier; according to Carrara, Vollers, et al., 2017a, a significant part of all information used and published by public administrations and exchanged with citizens has a spatial component. Thus, aspects such as different standards, level of geographic knowledge, lack of metadata, and even file size are significant barriers that prevent users and publishers from efficiently working with geospatial data.

Notwithstanding that the benefits of open and government data have been mentioned in most of the literature, there is also some work that has been done analyzing determinants of the success or failure of open data projects, especially involving government authorities. T.-M. Yang et al., 2015 illustrated through four perspectives of the impact of open data initiatives in Taiwan that legislation and policy have the most significant impact. Additionally, Keefe et al., 2013 used a case study of an e-Government project to explore the key factors of an open data project's success. Revealing that the development of a management and measurement framework of all the objectives and aims can bring some success, at the same time the lack of clarity about aims and specific objectives from the side of partners could affect the project development. In Bargh et al. work, the definition of Semi-Open Data paradigm is presented to define and frame initiatives and efforts that publish data but do not entirely accomplish the open data requirements. The authors presented a method to assess the level of implementation of the semi-open data in organizations, acknowledge their effort and guide them to reach the open data requirements. In fact, public agencies like Great Britain's Ordnance Survey from geospatial services sector, got realistic economic benefits partially releasing data, developing a mixed- cost model, with some free data and also some paid data (Young and Verhulst, 2016).

To review the barriers found in the literature and categorize what barriers belong to the data producer's perspective and what barriers belong to the data user's point of view, Table 3.2 illustrates authors, types of barriers, and the geographic context that proves that most of the work done has not considered the local level. Additionally, table 3.3 presents the references where data users' barriers were included. Finally, because most of the obstacles cited were not mentioned in the same way and there was no generic categorization found, Table 3.4 summarizes the number of occurrences to determine what barriers have been most analyzed. There are five relevant findings listed regarding the literature, as follows:

1. Seven relevant categories of barriers considering the data producer's point of view were most mentioned in the literature:

    - Technical
    - Organizational
    - Legal and Policy
    - Data quality
    - Financial issues
    - Cultural
    - Use and Participation

2. It seems that Use and Participation barriers are still not significant barriers; only two authors mentioned the user perception and active participation as an important issue to release or use open data.

3. Regarding the previously mentioned barriers experienced by data users, the categories that were not included are as follows:

   - **Standardization**: Included as another category where fragmentation of data, lack of interoperability, and many standards in how data is gathered are seen as issues from data re-users.

   - **Accessibility**: It is seen as heterogeneity of formats and lack of access to re-users.

   - **Discoverability**: Defined as how easy it is to find the data that is required. Related to other barriers such as standardization of data quality (metadata) but categorized as a remaining challenge by users.

4. Categories such as legal, financial, and technical were also mentioned from a data user point of view, but were less cited.

5. Data quality is still a significant burden from data producer and user perspectives.

3.4.2 Findings from the Online Survey

The participants were asked several questions. However, for this article, we have considered questions related to barriers regarding the reuse of open geodata in cities. The first question was: *From your experience with cities' open data portals, what do you consider to be barriers when using those portals?*. Using a Likert scale (Nemoto and Beglar, 2014) with three options (Major barrier, Moderate barrier, Not a barrier) respondents provided their option regarding barriers listed (see Figure 3.5). Overall, the top five obstacles considered by respondents as the most significant obstacles for the whole sample are **lack of update on published data** with 68.04% (Update data) and **low integration of data sources** with 53.09% (Standardization). Barriers related to Accessibility such as **low relevance to access for re-users** and **Published data is hard to access** with 47.94% and 47.42%, respectively. Finally, there was Discoverability barriers related to **time spent searching for data** with 43% (see Figure 3.5).

Tab. 3.2: This table represents the type of barriers to release data considering national or local use cases of open data initiatives, mentioned by each author. Due to there is no standard classification, barriers columns illustrate the barriers mentioned in each work, and at the same time the geographic context used for the use case. Note that mostly the national level is considered.

Author(s)	**Barriers**							**Geographic context**
T.-M. Yang et al., 2015	Technological	Organizational	Legal and policy					New York State
Janssen et al., 2012	Institutional	Task complexity	Use and Participation	Legislation	Information quality	Technical		The Netherlands
S. Martin and Foulonneau, 2013	Governance	Economic issues	Licenses and legal frameworks	Data characteristics	Metadata	Access	Skills	Rennes, France, Berlin, Germany, and UK
Barry and Bannister, 2014	Economic	Technical	Cultural	Legal	Administrative	Risk related		Ireland
Conradie and Choenni, 2014	Fear of false conclusions	Financial effects	Opaque ownership and unknown data locations	Priority (i.e., local government has more important things to do first)				Rotterdam
H.-J. Wang and Lo, 2016	Data findbility and collection	Data layout and format selection	Personal privacy	Data licensing	Data Description			Taiwan
Attard, Orlandi, Scerri, et al., 2015	Technical	Policy/Legal	Economic/Financial Budget	Cultural				N/A
Schmidt, Gemeinholzer, Treloar, et al., 2016	Desire to publish results before releasing data	Legal constraints	Loss of credit or recognition	Misinterpretation or misuse	Loss of control over intellectual property	Organizational constraints		N/A
Carrara, Vollers, et al., 2017b	Poor quality Open Data	A lack of standardization or heterogeneity	Difficulties in obtaining the data with the right information (metadata) for the purpose of its usability					European National level
Carrara, Vollers, et al., 2017a	Political	Legal	Technical	Financial	Others			European National level

Tab. 3.3: This table represent the mentioned barriers by some authors to release and reuse open data, considering the perspective of data users.

Author(s)	**Barriers**					
Carrara, Vollers, et al., 2017b	Low quality of Open Data	Lack of standardization	Availablity of open data, poor discoverability	Incorrect metadata		
Carrara, Vollers, et al., 2017a	Little awareness	Low availability	Legal	Technical	Financial	
Zuiderwijk, Janssen, Choenni, et al., 2012	Fragmentation of data	Lack of access to data	Lack of interoperability	Difficulties in processing the data		
Janev et al., 2014	Lack of standard procedures for querying government portals	The low quality of metadata	Low reliability and incompleteness of public datasets	The heterogeneity of formats used to publish open data		
Schmidt, Gemeinholzer, Treloar, et al., 2016	Paying for data	Varying degrees of data quality in different datasets	Varying standard in how data has been gathered	Varying data formats		

Tab. 3.4: The highlighted rows correspond to data users' barriers mentioned in the literature. The remaining rows were barriers mentioned as data producers' barriers.

Category	Occurrences
Data quality	5
Standardization	5
Accessibility	3
Awareness (cultural)	2
Technical	2
Financial	2
Discoverability	1
Legal and policy	1

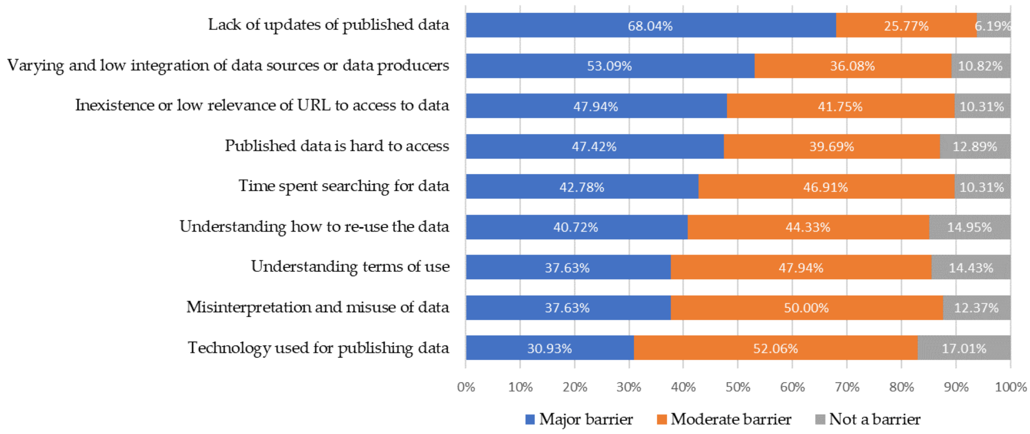


Fig. 3.5: Barriers mentioned by respondents. Lack of updates of published data, varying and low integration of data sources, and low accessibility were considered as major barriers (n = 195).

We now turn to the top five barriers mentioned by data users in the selected cities. Table 3.5 shows that **Lack of updated data** and **low integration among data producers** are the major barriers mentioned by data users in each city except Bogotá, where *time spent finding data* was the second major burden. A possible explanation is that data users in Bogotá (46 respondents, 23%) did not mention integration as a problem, possibly due to the existence and continuous progress of their local SDI (IDECA), which integrates more than 73 local entities (see Table 3.5). Misunderstanding about the reuse of available data and the terms of use were also relevant burdens chosen by respondents in all cities. Although in the whole sample those barriers are not considered within the top five concerns, the users of cities show a significant concern with understanding how the data can be used, and under what terms of use they are available. Finally, access to data through URL to establish a direct connection to available data in external applications or analysis processes (probably to get updated data) was chosen as another relevant barrier for data users in Bogotá, Medellín, and València.

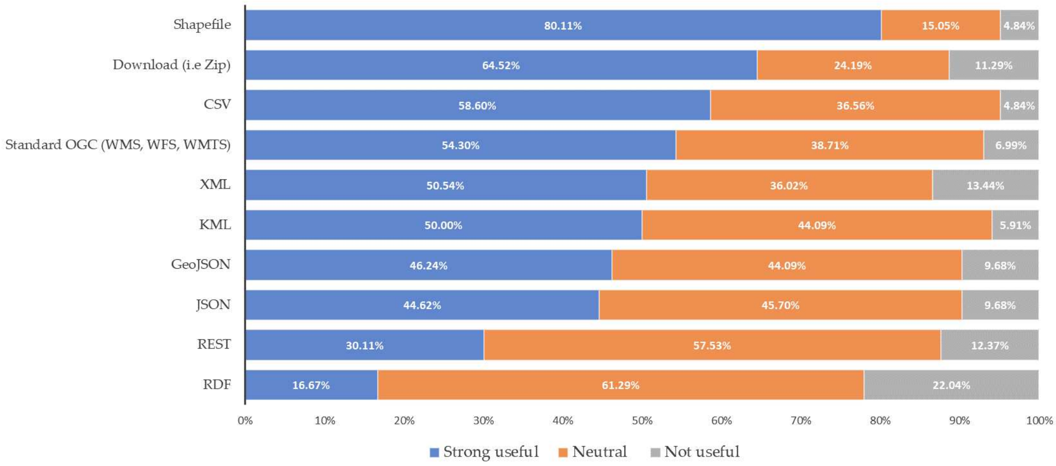


Fig. 3.6: Formats or services mentioned by respondents as most useful for their work. Shapefile, .Zip and CSV are considered strong useful (n = 186).

Regarding the low relevance of URL to access data, we also gathered users' opinions about the format or service they consider most useful for their work. This was achieved through the question *what format do you consider most useful for your work?* (see Appendix 3.6.2) in the survey. We found that for 186 respondents, the shapefile (80.11%) is the most useful format, secondly the downloadable formats like .zip (64.52%) and CSV (58.60%) in third place. This can explain that despite the effort in open data initiatives to promote formats like RDF or access through services like REST or JSON, users still consider most useful having the data in their own computers and manipulate as they want. This result may be due to the fact that in our sample 25% were geographical data analysts (probably a cultural aspect could have had an influence, since shapefile is a well-known format by this community, see Figure 3.3). Others typical geospatial services like OGC (WMS, WFS, WMTS), KML and GeoJSON were mentioned by the participants, but had a lower percentage of occurence (54.30%, 50.0%, and 46.24% respectively). The surprising finding was that typical machine-readable formats like RDF, REST even JSON has been mentioned as less useful for our respondents (see figure 3.6).

The third considered question related to barriers was: *From your experience, which was the most common error/barrier you have faced (not have faced) when searching or using data from city open data portals?*. It was an open question, and respondents were able to enter barriers from their own standpoint. The aim of this question was to identify any barriers that were not categorized or included in the question mentioned above but which are still an issue from the data users' point of view. This question was answered by only 164 people. Some participants' answers were not related to barriers or were challenging to interpret, and were excluded from the analysis, leaving a total of 151 valid responses for this question. Since most of the replies were in Spanish, it was necessary to translate to English, then group by categories and summarize the occurrences along the replies.

Table 3.6 illustrates the number of occurrences and the frequency of all barriers mentioned by data users, clustered by the categories as stated earlier. Currency is disclosed as the most mentioned category, related to available data, but not updated, 24% (36 occurrences) were reported for data users. This means that users not only expect a vast list of data from data providers, but the possibility of having access to current data is also a constant user requirement. Barriers related to categories such as Usability (15%, 22 occurrences), Data Quality (14%, 21 occurrences), and Standardization (13%, 20 occurrences) are also described by users as the most common errors when the available data is being used or searched. It was surprising that in this question Legal and Policy (3%, 5 occurrences) and Awareness (3%, 5 occurrences) were categories with fewer occurrences. It could be argued that current cities' open data portals have unclear and complicated licensing schema (where sometimes it is better not to use the available data to avoid any legal trouble, as also mentioned by Beno et al., 2017).

Taking into account the responses to both questions, Table 3.7 summarizes the most mentioned categories. Barriers related to **Currency** and **Usability** are two significant obstacles that are not considered in the literature (see Section 3.4.1); however, in this section they are validated as one of the main requirements from a data user point of view.

3.4.3 Findings from Participatory Workshops

During this activity, over 113 data users in selected cities (see Figure 3.1) discussed the data reuse and filled out over 46 follow-up questionnaires, where we asked participants about found reuse barriers and suggestions to overcome them. Since most of the replies in the questionnaires were in Spanish, it was also necessary to translate to English. Data users mentioned over 60 barriers grouped and filtered by six categories mentioned above during this activity. Table 3.8 groups these issues described in the selected cites; **Accessibility**, **Usability**, **Data Quality**, and **Currency** were the most frequently pointed out categories.

The lack of a relationship (direct or indirect) among the available datasets, defined as **non-existent geographic or statistical context**, was expressed as one the aspects to improve the usability and discoverability by data users, most of them economic analysts in the city of Medellín, geographical analysts and professors in urban planning in the city of València, and entrepreneurs who were looking for open geographic data to establish a new way to understand the education rates and

Tab. 3.5: Top five of barriers mentioned by data users along the online survey first question, for the entire sample and also group by each selected city.

Category	Barriers most mentioned in online survey	Percentage	n
Entire survey	Lack of updates of published data	68%	195
	Varying and low integration of data sources or data producers	53%	
	Nonexistence or low relevance of URL to access to data	48%	
	Published data is hard to access	47%	
	Time spent searching for data	43%	
Bogotá	Lack of updates of published data	74%	46
	Time spent searching for data	54%	
	Understanding terms of use	52%	
	Nonexistence or low relevance of URL to access to data	48%	
	Published data is hard to access	46%	
Medellín	Varying and low integration of data sources or data producers	68%	25
	Lack of updates of published data	64%	
	Nonexistence or low relevance of URL to access to data	60%	
	Time spent searching for data	44%	
	Misinterpretation and misuse of data	44%	
Cali	Lack of updates of published data	71%	41
	Misinterpretation and misuse of data	71%	
	Varying and low integration of data sources or data producers	54%	
	Published data is hard to access	54%	
	Understanding terms of use	46%	
València	Understanding terms of use	68%	19
	Lack of updates of published data	63%	
	Varying and low integration of data sources or data producers	53%	
	Misinterpretation and misuse of data	47%	
	Nonexistence or low relevance of URL to access to data	37%	

Tab. 3.6: Number of occurrences of the mentioned barriers by data users in the open question regarding the most common error/barrier when searching or using data from cities' open data portals.

Barrier category	Occurrences	Percentage
Currency	36	24%
Usability	22	15%
Data Quality	21	14%
Standardization	20	13%
Accessibility	16	11%
Technical	16	11%
Discoverability	10	7%
Legal and Policy	5	3%
Awareness	5	3%

Tab. 3.7: Summary of most mentioned category barriers by data users along the used questions in the online survey.

Category	Example of barrier
Currency	Lack of updates of published data
Accessibility	Varying and low integration of data producers. Nonexistence or low relevance of URL to access to data.
Discoverability	Published data is hard to access. Time spent searching for data
Usability	Misinterpretation and misuse of data
Data Quality	Data catalogs with poor descriptions
Standardization	Many formats, difficulty in searching the data

their relationship with cultural indicators in city of Bogotá. In terms of accessibility barriers, two points of view have been described: user accessibility (in terms of an analyst, for whom a download option is necessary to have full control of the datasets) and re-user accessibility (in terms of developers or data enrichers, where automatic and machine access is the most relevant way to connect for their applications) Ubaldi, 2013. Barriers related to this category were mentioned in all cities, but having most of the mentions in Cali, where GIScience master students cited the need to download the data in a suitable format to develop analysis processes concerning mobility and safety issues inside the city. Analysts have claimed, for instance, the following: "*there is no download option*", "*lack of mobility data*", "*data only for visualization but not able to download*", and "*many data related to events in the city but not suitable for analysis*".

Other accessibility barriers were mentioned by data users in València, Medellín and Bogotá; the data download option was sometimes complicated and included web log-in. Often the available data was not in a suitable format to reuse (e.g., PDF). Having data in pdf format not only restricts the automatic extraction that results in low reuse level, it is also considered as poor open data (The World Wide Web, 2017; Carrara, Vollers, et al., 2017a). Regarding data quality, "*gaps of data*", "*duplication of data*", "*no-clear metadata*", and "*no spatial resolution for local analysis*" were mentioned by the journalists and analysts in each city—especially in València, where the generalization level of available data (data at regional or national scales not suitable for local analysis—e.g., air pollution). The level of updated metadata was also considered by participants as an obstacle to understanding how the published data was gathered. Technical issues which were less mentioned but also cited by developers complaining that there is not enough information to understand how to use or apply the development resource. The multi-language option in some portals is not entirely supported according to València data users. Lastly, regarding terms of use, Bogotá's users mentioned a misunderstanding over the policy of available data.

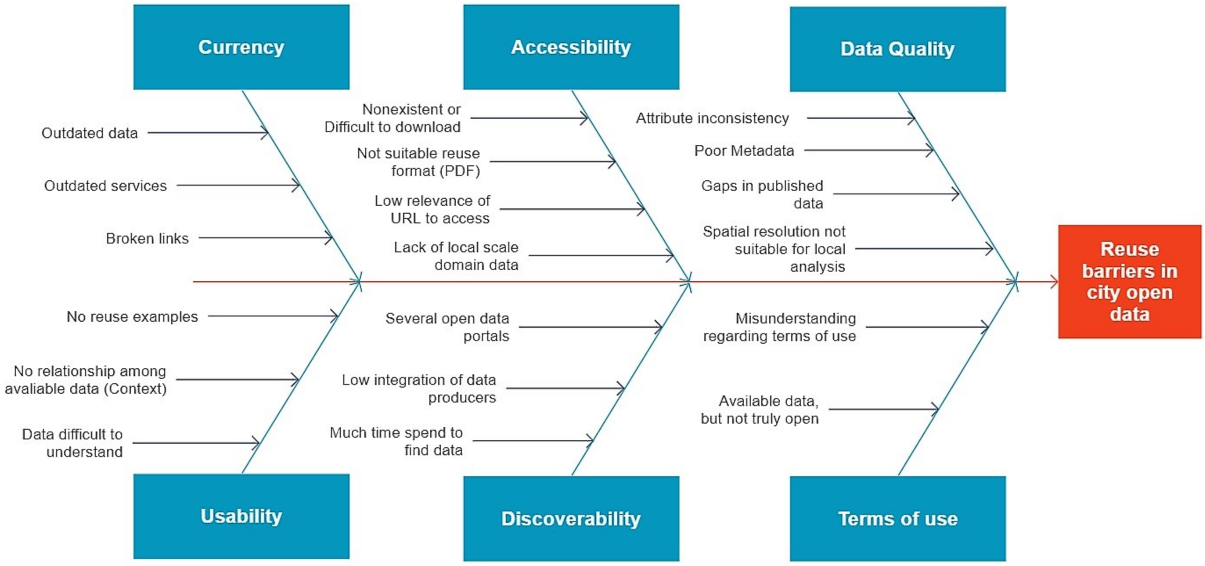


Fig. 3.7: Fishbone diagram of barriers identified from a data user point of view

A frequent issue mentioned by entrepreneurs and managers throughout the workshops was related to commercial use allowed in published data. This activity found a lack of clear terms of reuse in selected cities; some of them have created a specific license to use their data (e.g., IDECA in Bogotá), and other cities only have open data portal terms of use or they do not have a clear reuse policy. In general, after the set of participatory workshops and the interaction with data users from several backgrounds, Accessibility, Usability, and Currency categories have been reinforced as constant concerns from a data user point of view. Terms of use were less mentioned by data users, though this does not mean that licenses of available data are well-defined for re-users (i.e., developers). Many data users did not consider the available data "fully reusable" once the data was found in the cities' open data portals.

3.4.4 Data Users' Barriers Taxonomy

A total of six categories have been identified and validated during this research considering the data users' barriers from literature review, using a online survey, and validating through a set of participatory workshops; each category corresponds to barriers mentioned and identified from a data user's perspective. Figure 3.7 is a fishbone diagram that represents the categories and barriers that prevent the reuse of open geographic data in cities, based on the opinions from data users of selected cities.

- **Currency**: The lack of updated data in the local open data initiatives was considered by data users as the major barrier to reusing the available data.

Tab. 3.8: Most mentioned barriers by data users in selected cities in the participatory workshops.

Category	Cali	València	Medellín	Bogotá
Usability	Data difficult to understand	No suitable for reuse data format No applications to validate the reuse of data No relationship among the datasets available Reduced usability	Misunderstanding of available data No categories for available data No relationship among available datasets	No relationship among published datasets No apparent usability of available datasets There are no examples of reuse
Accessibility	No download option Official data web sites have no data Lack of data for transportation Lack of accessibility More marketing of current initiatives Information related to events, but no data related Data only for visualization, not downloadable option	Only one dataset for education No transportation data is available Lack of important attributes Reduced discoverability, to find data it was necessary to spend a great deal of time	No downloaded option No georeferenced data available Lack of accessibility for some datasets Data in PDF format	Available data in PDF format
Data Quality	No metadata Gaps in available data No georeferenced data No raw data, the available data is processed No metadata is related to the data source Processed data	Not enough metadata Generalization of data, only for regional or national approach, Not local level	No suitable format for open data	Duplication of data Attribute inconsistency Gaps in published data No updated metadata Generalization of data, Nor for local reuse Published data not georeferenced
Technical		No API documentation or examples Language issues among datasets No advanced searching options to find datasets JSON file with issues	No advanced search option	Some web sites based on Flash technology User authentication for some portals
Legal and Policy				Misunderstanding regarding terms of use License not clear Available data, but no open The terms of reuse are not clear Lot of available data, but not truly open
Currency	Not up to date data	Some datasets are not up to date No up to date apps in official websites	Not up to date data	Data not up to date

Outdated data and services or broken links were mentioned in both online survey and workshops as most disappointing when analysts, entrepreneurs, geospatial developers, journalists, and other data users need to include data in their processes or external applications. Having updated data is a common requirement for all kind of open data, regarding geodata, data users mentioned currency also due to the difference among the available data by paid versus the available data accessible through geo-portals. Considering that there is much work to do to get full accessibility to updated data. A possible precedent associated to this issue if the way that some geographical authorities found having a mixed-open data model, releasing only a certain among of data but keeping the most updated as a premium service (Young and Verhulst, 2016).

- **Accessibility**: Although all selected cities have their own data portal initiatives, with several available data sets, accessibility barriers were mentioned over and over again by the data user communities. The most mentioned obstacles were the nonexistent or difficult way to download data for users that need full access to make a local analysis 3.4.2. As well as the low relevance of the developers' resources for re-users that need to link the published data in external applications. However, the URL access was not the only concern; in cases where the API resources were included, the lack of documentation and guidelines to use was also cited by re-users. Ultimately, there was a lack of datasets with specific geographic component (e.g., *air quality*, *local mobility*, *education*, and *urbanization*) that was not accessible through current cities open data portals.

- **Data Quality**: This category is a large topic and was mentioned by the literature review (see Section 3.4.1) and is included in the empirical analysis that this research carries out. However, the criteria of data quality from a data user point of view could be more specific. Based on the findings of this study, the lack of metadata (especially for geographic data) was one of the major barriers mentioned by re-users. Attribute-inconsistent or gaps in published data is also a relevant feature to improve. According to data users, the possibility of predicting published data that are not complete or data which has specific characteristics (e.g., local reference system) might help them to save time. Generalization of data was cited for many users when they found relevant data which was not appropriate for local analysis or development. As an example of this issue, users mentioned an environmental use case that could be considers as an accessibility issue—the air quality data found in most of the selected cities have a regional or national scale. As another example data users in Valencia mentioned that education rates were published only in a regional or national scale which not contribute at all to analyze the local issues. Once cities become involved in open data initiatives, they need to consider extracting, processing, and integrating the correct information for the city's needs, not only integrating any open data from several national or local departments with any local propose.

- **Usability**: Further barriers—especially in the participatory workshops—were related to the lack of reuse examples. Many city portals limit their actions to publishing data, but there are no examples or use cases that users can use as a guideline to understand how the data is applied or how it could be integrated with other applications. Based on the data user's opinions, many open data portals are a vast list of data, but there is no context to understand how data could be relevant to the city. Likewise, besides the data category, there is no relationship among the available services. This lack of context creates a misunderstanding of data and misuses about how data can be applied or reused.

- **Discoverability**: This research identified that although all selected cities have an ongoing open data project, when users need to find the required data they search in several websites but not in the local open data initiative. Using search engines (e.g., Google, Yahoo, Bing) or in the best case the open data national initiative websites, when users were asked to find specific data such as bike routes in their city, they encountered several issues in obtaining the required data. In some occasions, users went to the data authorities' website to find the current open data initiative, but most of them did not have the expected emphasis on the initiative. It seems that the lack of open data centralization could be a relevant usability barrier from data users' point of view. Another mentioned obstacle was the low integration between city departments regarding the data release process—especially in Cali and València. Data users claimed that the existence of several city department websites—sometimes all of them offering a different kind of data about the same topic—could confuse and reduce the reliability of the releasing process. This minor integration could result in a significant amount of time required to find relevant or useful data.

- **Terms of Use**: The least-pronounced but still a common category barrier among three data sources used in this research was legal and policy concerns. Many data user communities manifested a significant misunderstanding about the terms of use or reuse of available data. Most of the open data policies around cities depend on national legal implementation; many countries have been involved in their own open data policy, and the transition to the local level could affect the way that the published data is being reused. Currently, to have a successful national open data initiative, cities have a determinant role to play in this value chain (Carrara, Vollers, et al., 2017a). Having a consistent, clear, and integrated open data policy could attend to re-users to understand what kind of use is allowed and how they should include the published data in their external process or applications. Regarding terms of use in cities' open data, portals are not clear and easy to read, and the reliability to reuse could be affected. As was mentioned by Beno et al., 2017, potential users may feel misled when they find that available data have legal restrictions. Some entrepreneurs in the participatory workshop in Bogotá referred to the need to include whether commercial use is included or not to avoid future legal issues. This research notes that many of the terms of use available in cities' open data portals are related to websites or portals rather than data per se. Having specific terms of reuse and use for published data might avoid any misunderstanding.

3.5 Discussion

Previous studies have considered open data initiatives at the national level as the scale to shape the possible benefits and implementation obstacles of open data when it is reused, considering the data release process as the core of the open data systems. This research takes another perspective, where data user needs are the basis for improving the reuse of open data. The study includes open geographic data as a type of data, local level as the scale of study, and data users with different backgrounds playing the main role. Section 3.5.1 summarizes barriers mentioned by data users from four cities with open initiatives with different approaches. Section 3.5.2 presents some remarks about the role of local data user communities and how data authorities are facing similar issues regarding license, identification process of data users, and their needs and current user engagement strategy.

3.5.1 Summary of Barriers

As mentioned earlier, most of the barriers to open data reuse from the literature were determined from the data producer perspective. In Section 3.4.1, we found that most of the authors directed their efforts towards analyzing the possible benefits, adoption barriers, implementation limitations, and determinants to having successful or failed open data initiatives. None of the references mentioned in that section has considered the role geographic data could play in the strategies of local open data initiatives to tackle OGD challenges. At the same time, the context used in the work illustrated that national efforts and the data releasing process have an important role.

In Table 3.3, we listed some work done taking the data user viewpoint into consideration and presenting possible obstacles that could prevent taking the full advantage of open data: **Discoverability**, **Accessibility**, and **Standardization** were the categories less identified. Most of the barriers related to these categories were confirmed during our online survey (see Section 3.4.2). Beyond these, we extended the barriers mentioned above, and found that **Currency**, **Usability**, and **Data quality** are additional, relevant concerns of data user communities when the open geodata is being searched or reused at a local level. These barriers were highlighted in the participatory workshops where geographic data was the most requested kind of data by users, however, also was the most criticized along the activity. Out of date web services, lack or gaps of metadata, data available without any quality control, lack of standardization of the reference systems (some services even had custom reference systems) contribute to making the task of reusing the data more difficult.

Most of the discussion in the literature is centered on accessibility issues, and indeed most of the official organizations at national or local levels take the data release process as the primary task. Data users are currently demanding to have not only accessibility (Ubaldi, 2013; Carrara, Vollers, et al., 2017b)—they want to go beyond access. According to our findings (presented in Section 3.4.2), a constant concern in data user communities is the currency of published data. The "*lack of updated published data*" was selected as a significant burden for 68.04% of 195 participants in our survey. Furthermore, "*misinterpretation and misuse of data*" were also considered by data users as an obstacle to the efficient reuse of published data. Data catalogs with large lists of data with neither statistical nor geographical relationship or context may confuse data users and make them spend too much time searching for the relevant data.

Two of eight OGD principles are related the format that data is released and the way that data should be open to public in a machine-readable format which is also non-proprietary. This research found for the sample considered (see Section 3.3.2) the shapefile as most useful format (see Figure 3.6). The respondents consider typical geographic services like OGC services, KML, GeoJSON as more valuable than the promoted open data formats like RDF. A possible reason to explain this result could be data ambiguity existing in local open government initiatives, where format like RDF have an inadequate description, and also in the geographic community do not have a significant representation or use in the analysis process.

In (Beno et al., 2017), the lack of harmonization between portals was considered a severe burden that makes data users confused about similar available data in different portals. This research has confirmed this finding and group under the usability category barriers mentioned by data users (in all selected cities) such as "*data difficult to understand*", "*no relationship among published data*", or "*no applications to validate the usability of available data*" (see Table 3.8). The quality of data is also a constant burden for data user communities—in particular for data users included in this research. Although this category is already considered in the literature from a data producer perspective, it is still an aspect of improving an open data chain Carrara, Nieuwenhuis, et al., 2016. According to our survey and workshop participants, having data with issues like "*no metadata*", "*published data not geo-referenced*", or "*not enough or clear metadata*" considerably reduce the data source reliability, and thus the open data initiative effectiveness. Data which is not machine-readable (e.g., PDF) was another barrier mentioned in our workshops in selected cities—especially in Cali and Medellín . At the same time, issues like "*data only for visualization*" or "*no download option*" were mentioned by users that require the full control of data for local analysis. As a concrete example of this situation in València and Bogotá, the local road layers were required to create a mobility analysis. However, data users cited that there was no option to download and only visualization was possible through geo-portals.

To conclude, we revisit the research question presented in Section 3.3 (What barriers prevent open data reuse by data consumers?), and summarize the discussion in this paper with the following observations. We identified and explored 19 barriers, categorized them into six categories (see Figure 3.4.4). We identified the most mentioned concerns and requirements from data users in four cities—particularly those that work daily with open geographic data. Currency was the most mentioned concern by data users from different backgrounds. Accessibility and Data quality were also highly mentioned during this research. Usability, Discoverability, and Terms of use were also included in this taxonomy of reuse barriers, having the low integration of city departments, misunderstanding of terms of use, and no geographical or statistical relationship as constant issues faced by data users in selected cities.

3.5.2 The Role of Cities and Their Data User Communities

The open data chain (European Commision, 2013) is presented by the European Commission in its strategy as an interaction between official departments and open data stakeholders. Carrara, Vollers, et al., 2017b illustrated how raw data is transformed into economic value considering the creation of data until the aggregated services. At the same time, this report categorized the roles of open data stakeholders into four types of actors: Suppliers and Aggregators in charge of the creation and aggregation process, and Developers and Enrichers generating analysis and a new bunch of data services or products. Ubaldi, 2013 presented a similar scheme, but included one additional step, named "final data use" as the last stage to promote the sustainability of the public data creation process. Correspondingly, Ubaldi also suggested the identification of an "ecosystem of users" that responds to specific user demands to promote the creation of value. In terms of open data ecosystems, Zuiderwijk, Janssen, and C. Davis, 2014 proposed the essential elements of a multi-dimensional system where the feedback from data users in one of the key elements. Likewise, Janssen et al., 2012 suggest that open data systems must consider the data users' feedback, mentioning that "*there is no insight into users' perspective and users' needs*".

Data user communities and their feedback is becoming more important in the current open data value chain, but the geographic context where those users are involved is also important. Indeed, cities have a relevant role to play here. During this research, several cities were considered to compare the current actions of local authorities in charge of leading the open data initiative and their data user communities. In Table 3.1, several aspects regarding open data in those cities are compared. The initial element was about what open data thematic is of their interest; all cities have mentioned thematics like mobility, urban planning, economic development, or security. Medellín mentioned that data urban planning around a sustainable and smart city strategy are of interest to them. This is an interesting claim. According to Carrara, Vollers, et al., 2017a, open data could enable the reinforcement or implementation of a smart city initiative, as a more "connected" city and the development of new services related to sensors around the city could result in an important amount of data that users can use to enhance the quality of life in the city.

In relation to terms of reuse, Medellín and València have adopted a creative commons license (Attribution 4.0 International CC-BY 4.0) for their published data (see Table 3.1); however, it seems that this does not guarantee the prevention of any misunderstanding from a data user point of view (approximately 68% chose this as one of the major and moderate barriers in this city; see Figure 3.5). Cali does not have any defined open data terms of use, but the local authority follows and has a coordination mechanism in place at the national level. Likewise, 46% of their respondents mentioned "the understanding of terms of use" as a barrier. Only Bogotá—which has a local authority in charge of the open data initiative and at the same time is the Local SDI (IDECA)—has their own license (IDECA License), a kind of barrier for their users that get confused when they need to understand what use or reuse is allowed. Fifty-two percent of their participants in our survey chose "*understanding terms of use*" as the third barrier to use of the open data portal in Bogotá.

According to the open data value chain (European Commision, 2013; Carrara, Vollers, et al., 2017a), developers have an important role in any open data initiatives. At the same time, they have the skills to enrich and transform the available services into new kinds of innovate services or applications that show the real potential of open data Attard, Orlandi, Scerri, et al., 2015. Thirty percent of our respondents were developers (see Figure 3.3). However, València does not consider it relevant to collect any entities or organizations who have used the available datasets. In Medellín, Bogotá, and Cali, the identification of those stakeholders is quite poor (see Table 3.1). Nonetheless, València, Medellín, and Bogotá users have chosen the low relevance of the URL access to data as a major obstacle and in the set of workshops, *"API documentation", "JSON files with issues"*, and others were among the technical barriers most mentioned.

Finally, the internal departments in each city were also compared in this research. We found that the barrier "varying and low integration of data sources or data producers" not chosen as a major burden only in Bogotá. In other selected cities, this obstacle had an important percentage (68% for Medellín, 54% for Cali, and 53% for València). A possible explanation of this result could be the work-done of IDECA, who is a well-known authority among their data users (especially who work with geographic data), and the integration of the spatial information of more than 70 local entities. Although Cali also has a local SDI and was the authority contacted, this SDI is in an initial phase and Cali data users are only getting used to knowing what IDESC is doing and what kind of data it is publishing.

3.6 Limitations and Final Recommendations

This last section discusses possible ways of to using the observations made during this research. Limitations of the research are presented in Section 3.6.1 before the article ends with a set of recommendations in Section 3.6.2.

3.6.1 Limitations

During this research, we identified a set of barriers (see Section 3.4.4) from a data users point of view. Using three data sources, we aimed to identify what obstacles data users in cities face when they are looking for data, but especially when they want to reuse and incorporate the available data from a city in their projects, analysis, or external applications. In Section 3.2.2, we illustrated the consideration of geographic data in open and government initiatives due to its relevance to the reuse of available data according to the Reuse of Open Data report of European Data Portal (Carrara, Vollers, et al., 2017b). We gathered opinions, requirements, and barriers to the reuse of open data in cities through participatory workshops, and contacted over 100 people from different backgrounds.

Most of our respondents and participants had a geographical background or had worked with spatial data. Therefore, there are possible limitations that need to be acknowledged. A possible bias of the identified barriers could be that they are not applicable to other open data users from backgrounds like journalists, analysts, or developers who work with any kind of data, but might be interested in open data. We have delimited the barriers according to the respondents and participants of four cities, especially in Spanish language, with the data user communities of each city. We thus encourage the use of those conclusions with caution, as these barriers might not apply to other cities.

During this research, we found that there are some significant differences among the open data initiatives led by data authorities which are in charge of the spatial data integration in the city and entities who consider the open data initiative as another project of Open Government. For example, Bogotá has IDECA, who is the data authority that is currently leading the open data initiative while at the same time it is the local SDI. In Medellín or València, the open data leadership is in the charge of city halls like "Alcaldia de Medellín" or "Ayuntamiento de València". An explanation of the differences is barely possible at this point, since this necessitates information about the open data agendas and working processes of the different institutions represented. An extension of this study could thus investigate the actual interplay between the strategy of the local SDI/open data initiatives and the way that data is being released, searched, and used.

3.6.2 Recommendations

Based on the findings of this research and the data users' opinions collected during the participatory workshops, there are some suggestions that local data authorities might apply to engage and integrate their data communities into current open data initiatives.

Identifying data user groups: We have noted that most of the local authorities still need to clearly identify their data user groups. Some universities have been contacted and are working in some activities (e.g., hackathons or workshops) along the open data strategy. Other cities have identified development companies or organizations that continuously work with open data; however, the identification of those users, their needs, or their requirements are not part of the strategy. All data authorities have an interest in engaging more users and adjusting their strategy to data users' requirements (see Table 3.1). Data users should be integrated during the whole open data initiative, not just included as the last step of the strategy. The current research has listed barriers which inform data authorities about aspects to focus on while working towards higher integration of users' wishes in their strategies.

Continuous services tracking: Other suggestions that were also mentioned in the literature (Conradie and Choenni, 2014; Zuiderwijk, Janssen, and C. Davis, 2014) is related to the analysis and continued tracking of the available services. The accessibility and data quality concerns mentioned by data users in participatory workshops (see Table 3.8) might be tackled by understanding what the top five most requested services are, what services users want to download, and what services need more accurate and complete metadata. In general, this continuous tracking might yield an improvement of the published services.

Notification of further released data: In cities like València and Bogotá, data users have mentioned the need to know through an automatic service what services or new data have been released. Data producers can put more efforts to including notifications or alerts regarding the state of available services—especially services that have been identified as the most frequently used. At the same time, syndications like Really Simple Syndication (RSS) can also be used for future services or data that will be included as part of open city data.

Clear and straightforward terms of use or license: During the survey and set of workshops the terms of use were mentioned for data users as one of the obstacles to reusing the current data in cities. This barrier reduced the reliability of the open data in the selected cities. We consider that creating a simple and specific set of terms with natural language will help to reduce any misunderstanding regarding the utilization allowed of the available data.

More examples or basic reuse kit: Regarding usability barriers, data users cited that the lack of examples and basic guidelines to use and enrichment of the available data have a negative impact on the reuse level. Creating guidelines (as suggested in Degbelo, Bhattacharya, et al., 2016) to reuse and explore the essential technical elements as part of the local open data initiative might have a positive effect and reduce the misunderstandings of published data. This research suggests that data producers should not limit their work to the provision of an extended list of available datasets. Creating a basic reuse kit that includes, for example, a guideline to downloading, connecting, enriching, and displaying released data could help newcomers and other re-users to understand how the city's open datasets could be used in a meaningful way.

From a research point of view, we have analyzed different local open data initiatives in two countries, we found that some cities the local authorities that lead the open data movement is the local SDI, framing the data user engagement based on the SDI approach, where geographic data and standardization issues are the priority task. However in cities where the open data initiative is leading by open government offices inside city hall, the strategy and the way that data is released could have different impact in data users communities. Comparing current open data strategies and SDIs projects along cities, there are similarities between the two approaches. For instance, SDIs had to face standardization barriers in the past; through geo-viewers they also wanted to tackle accessibility issues, and they also faced barriers related to providing high data quality services. In this sense, we suggest that more research explores the role of local SDIs in open data times so that lessons learned from years of work on SDIs could flow into current open data projects.

Acknowledgments: The authors of this research gratefully acknowledge funding from the European Commission within the Marie Skłodowska-Curie Actions, International Training Networks (ITN), European Joint Doctorates (EJD). The funding period is January 1, 2015 - December 31, 2018, Grant Agreement number 642332 — GEO-C — H2020-MSCA-ITN-2014. Sergio Trilles has been funded by the postdoctoral programme Vali+d (GVA) (grant number APOSTD/2016/058). We are grateful to Medellín, Cali and València city Halls (Alcaldía de Medellín, Alcaldía de Cali and Ayuntamiento de València), IDESC, IDECA and Las Naves (València Innovation Authority) who shared their experience with open data initiatives and support in this research. The authors thank Roberto Henriques from NOVA University in Lisbon who assisted in proofreading of the article.

Author Contributions: The introduction, literature review and findings sections were developed jointly by A.D and F.B. Survey, participatory workshops, cities selection, authorities and data users communities were designed and performed by F.B. J.H., A.D and F.B. conceived jointly the idea and research method. Discussion and recommendations were developed jointly by A.D and F.B. S.T contributed in the creation of tables and format of the paper.

Appendix: Online survey questions

This appendix illustrates the questions and sections included in the online survey that was publicly shared. The following format was used to guide respondents through the survey's sections.

1. **Personal information:** Tell us a little about yourself. We will not share or publish this information.

   a) Which country are you currently working?: Open Question.

   b) Which city/cities are you working or using geographical data?: Open Question.

c) How old are you?: Open Question.

2. **Your work**: In this section we are interested in aspects of your work and your experience level in the sector or industry to which you belong or have belonged to in the past. You can mention the elements that are the most relevant.

   a) What is your employment role?: Multiple choice: Geographical apps developer, Geographical data analyst, Developer and analyst, Data Science analyst, Manager – Project leader, Researcher - Student - Teacher, Other.

   b) In which industry do you work?: Multiple choice: Local Government, National Government, Education, Non- profit, Media, Startup – Entrepreneurship, Business, Other.

   c) How much experience do you have in the industry?: Multiple choice: Less than 1 year, 2 to 6 years, 7 to 10 years, 11 to 20 years, More than 20 years.

3. **City Open Data** It is important for us to know your opinion about open data available in the cities. In particular geographic data. In this section we will ask you about your reasons for use this data and your knowledge of those current initiatives.

   a) Please indicate the level of importance for each option when using city open data?: Multiple choice grid, with Very important, Neutral and Not important as choices: Geographic information accessibility, High-quality geographic information, Scalability and ease of project maintenance, City innovation improvement, Transparency and collaboration improvement, Economic benefits for the city, Academic and research improvement

   b) Do you know or use the cities' open data portals?: Multiple choice with yes or not as choices.

4. **Cities' open data portals** Please provide specifics on data portals, adding a URL where possible. If your previous answer was Yes, please specify which city open data portals you know or have used

5. **Barriers and features**: We would like to know the barriers, errors, and problems that you have encountered while using cities' open data portals. Also, we would like to know the features and aspects that you consider positive and that should be kept within these initiatives.

a) Which functionalities do you think are not useful in city open data portals?: Open Question.

b) From your experience with city open data portals, what do you consider to be barriers when using those portals? Multiple choice grid with Not a barrier, Moderate barrier and Major barrier as choices: Published data is hard to access, Misinterpretation and misuse of data, Time spent searching for data, Understanding how to re-use the data, Understanding terms of use, Nonexistence or low relevance of URL to access to data, Technology used for publishing data, Varying and low integration of data sources or data producers, Lack of updates of published data.

c) From your experience, which was the most common error/barrier you have faced (not have faced) when searching or using data from city open data portals? Open Question.

d) Which of following do you think are the most needed features of city open data portals? Multiple choice grid with Highly necessary, Neither necessary nor unnecessary and Unnecessary as choices: Filters for advanced search, URL to Access data, URL to Access data, Data Categories, Table view and graphs, Terms of use and re-use, Details on how the data has been produced, Viewers and interface to explore the data, Feedback from other users.

e) Which of following functionalities, is your frequency of use in cities' open data portals? Multiple choice grid with Every time, Occasionally/Sometimes, and Never as choices: Filters for advanced search, Access data URL, Data Categories, Table view and graphs, Terms of use and re-use, How the data has been produced?, Viewers and interface to explore the data, Viewers and interface to explore the data, Feedback from others users.

6. **City open data portals usability** We'd like to know about the level of use of city open data portals and the available geographic data. In this section, we will ask your frequency of use and we want to determine the usability level of those portals.

   a) When you need to use city geographical information which portals do you normally use? Multiple choice grid with Often, Sometimes and Not used as choices: Government data portals. (National), Government data portals. (City-Local), Private repositories, Pay or collect data, International repositories, Other.

b) Indicate your agreement level regarding these statements on current city open data portals: Multiple choice grid with Agree, Neither agree or disagree and Disagree as choices. I would like to use these portals frequently, I found the portals unnecessarily complex, These portals were easy to use, I would need the support of a technical person to be able to use the portals, I found the various functions in the portals were well integrated, There was too much inconsistency in the portals, I would imagine that most people would learn to use the portals very quickly, I found the portals very cumbersome to use, I felt very confident using the portals, I needed to learn a lot of things before I could get going with the portals.

7. **Searching for geographical data** We'd like to know which criteria and formats you use when searching and choosing geographical data.

   a) Tell us about your data quality criteria when choosing available data in city open data portals? Multiple choice grid with Desirable, Neutral and Undesirable as choices. Accuracy: data/metadata record correctly described, Completeness: the number of completed fields in a data/metadata record, Consistency: discrepancy between data published and entire data catalogs, Currency: data or metadata is up date, Technical accessibility, Openness.

   b) Which of the following are main features that you consider when choosing available data in city open data portals. Multiple choice grid with Definitely consider, Might or might not consider and Would not consider as choices. Data quality, how data was produced, Geometry (Point, Lines, Polygons, raster, other), Lack of information (Incomplete fields), Terms of use and re-use, Technology used for the publication process, Creation/Publication date, Author (Public agency, Private), Cost, Openness.

   c) What of the following output formats do you consider most useful for your work? Multiple choice grid with Strong useful, Neutral and Not useful as choices. KML, OGC Standard (WMS, WFS, WMTS), REST, CSV, Shapefile, GeoJSON, JSON, RDF, XML, Download files (i.e Zip).

   d) If you had the chance to improve city open data portals, which are the improvements/features or tools will you would add and why? Open Question.

e) In your industry, how do you think we might increase the usage of geographical data on current city open data portals? Open Question.

Part II

User Information

4

Designing Semantic Application Programming Interfaces for Open Government Data

This chapter was published in the JeDEM - eJournal of eDemocracy and Open Government as *Degbelo, A., Trilles, S., Kray, C., Bhattacharya, D., Schiestel, N., Wissing, J. and Granell, C. (2016) 'Designing semantic application programming interfaces for open government data', JeDEM - eJournal of eDemocracy and Open Government, 8(2), pp. 21–58. Note: Some figures in this chapter are best viewed in the online version available at* `https://doi.org/10.29379/jedem.v8i2.420`.

Abstract. *Many countries currently maintain a national data catalog, which provides access to the available datasets - sometimes via an Application Programming Interface (API). These APIs play a crucial role in realizing the benefits of open data as they are the means by which data is discovered and accessed by applications that make use of it. This article proposes semantic APIs as a way of improving access to open data. A semantic API helps to retrieve datasets according to their type (e.g., sensor, climate, finance), and facilitates reasoning about and learning from data. The article examines categories of open datasets from 40 European open data catalogs to gather some insights into types of datasets which should be considered while building semantic APIs for open government data. The results show that the probability of inter-country agreement between open data catalogs is less than 30 percent, and that few categories stand out as candidates for a transnational semantic API. They stress the need for coordination - at the local, regional, and national level - between data providers of Germany, France, Spain, and the United Kingdom.*

4.1 Introduction

Various factors have contributed towards the increased availability of Open Data, including national and international legislation, requests for transparency and hopes for enabling new services. Cities are a particular 'hotbed' for producing data and consuming it: for example, a lot of sensor data is produced in a smart city (see Hancke et al., 2012; Lecue et al., 2012), and numerous apps have been developed that make use of city data (see M. Lee et al., 2016). While many have pointed out the high potential for open data in terms of better participation, increased transparency, new services and better use of ressources (e.g., Fechner and Kray, 2014; Hartog et al., 2014; Janssen et al., 2012; Masip-Bruin et al., 2013; Ojo et al., 2015, there are still several challenges that need to be tackled. These include managing the vast amount (and high bandwidth) of data being produced (Chen et al., 2014), the heterogeneity of the data (Janssen et al., 2012; Masip-Bruin et al., 2013; D'Aquin et al., 2014), data quality and recency (Janssen et al., 2012; Masip-Bruin et al., 2013), coordination mechanisms at the technical and political levels (M. Lee et al., 2016), as well as privacy issues (Janssen et al., 2012; Chen et al., 2014), to name but a few.

One important challenge when working with open data is to enable machines (or applications) to re-use it. This aspect is not considered in many definitions of open data. For example, a commonly used definition says that open[1] data is data which is freely available and shareable online, without charge (The World Wide Web Foundation, 2015). While this definition ensures that *humans* can access and inspect the data, it says little about the use by *machines*. There are two key aspects to consider while enabling open data reuse by machines: the provision of open data in a structured (a.k.a. machine-readable) data format such as Comma Separated Values (CSV), Extensible Markup Language (XML) or Resource Description Framework (RDF); and the querying of the data by machines. The first aspect is addressed by the five stars deployment scheme for Linked Data introduced in (Berners-Lee, 2006). In essence, the greater the degree of structure of the data, the easier its further processing by machines. Regarding the second aspect, the common way to enable open data querying over the Web for machines is to provide an Application Programming Interface (API) that describes the functions an application can execute to access an open data repository, the parameters that are expected and the kind of results returned. APIs were mentioned as one of the 12 critical factors for success of open data initiatives in (Susha et al., 2015). This notwithstanding, the recent edition of the Open Data Barometer pointed out that "*More elaborated APIs that facilitate access to data are still very rare among government data*" (The World Wide Web Foundation, 2015). This article aims at initiating a discussion on the required components of such APIs. More specifically, the work focuses on the types of data categories that more elaborated APIs should offer, a topic which has received little attention in the literature so far.

[1]The world 'open' is associated with many different connotations (for a recent discussion see Pomerantz and Peek, 2016). For the purposes of this article, the definition of open data as "data which is freely available and shareable online, without charge" (The World Wide Web Foundation, 2015) is adopted.

APIs form an essential component of the World Wide Web. For instance, ProgrammableWeb.com, the "*Web's defacto journal of the API economy*"[2] lists more than 15,000 Internet-based APIs as of June 30, 2016. One of the main motivations for building APIs is to improve programmer's productivity by enabling code reuse instead of code writing from scratch (see Stylos and Myers, 2007). Note that improving programmers' productivity and enabling data access to machines are, in the context of API development, two sides of the same coin. Machines or software agents use APIs to autonomously access data stored in external repositories, but it is the programmer who tells these machines or software agents which APIs to use, and how they can best access the data. Programming is an important aspect of API development, but it is not the most important of it. As Henning, 2009 pointed out, "*an API is not about programming, data structures, or algorithms - an API is a user interface*". Thus designing an API boils essentially down to providing a useful interface by which machines can access resources (e.g., datasets). API design is equally providing a useful interface by which programmers can access resources for the purpose of application development. API design is thus, like any other design problem in the context of information sharing, best viewed as a "human-machine-human" conversation problem (see Scheider and Kuhn, 2015) for a detailed discussion of the "human-machine-human" perspective on information sharing).

The salient peculiarity of API design is the "*stakes of getting the design right in the first place*" (Myers and Stylos, 2016). Since APIs are used by many applications once they are developed, any change in their interface incurs thousands of broken apps, and therefore considerable loss of time and money[3]. Stylos and Myers, 2007 list three different stakeholders of an API: **API designers** (whose goals are to maximize the adoption of an API and minimize its support costs), **API users** (willing to write error-free programs, and use APIs that many other programmers use), and **consumers of products built with the API**. The work in this article is mostly relevant to API designers and users (i.e., programmers). Citizens (as consumers of Apps built with these APIs) will benefit indirectly from semantic APIs for exiting government data.

RESTful APIs are one of the major types of APIs nowadays, covering about 60% of the API market[4]. In a nutshell, a RESTful API provides the opportunity to retrieve resources via the methods from the Hypertext Transfer Protocol (HTTP). Richardson and Amundsen, 2013 recommend using the HTTP methods GET, POST, PUT, DELETE, and PATCH for Web API development. The first step of designing a RESTful API is, following (Richardson and Amundsen, 2013), to list *semantic descriptors*. Semantic descriptors are all the pieces of information that API users might want to get out of the API, or put into the API: they are the data items (a.k.a.

[2]See `http://www.programmableweb.com/about` (last accessed: June 30, 2016).
[3]For an anecdote showing undesirable consequences of minor changes in an API, see Henning, 2009).
[4]As of June 30, 2016, ProgrammableWeb.com lists about 9,500 RESTful APIs.

informational resources) that the API should return. Depending of the application scenario, these semantic descriptors can be grouped together and organized into hierarchies. An example of semantic descriptor for an API returning a list of books is 'books'. A RESTful API with a base url 'http://mylibrary.com' could:

- Return all books in the library via GET `http://mylibrary.com/books`
- Return the book with the ISBN 1098-6596 via GET `http://mylibrary.com/books/1098-6596`
- Add a new book with the ISBN 1098-6596 to the library catalog via

  POST http://mylibrary.com/books/addbook/1098-6596
- Delete this book from the catalog via DELETE `http://mylibrary.com/books/1098-6596`

The work reported in the next sections aims at providing an empirical basis for the choice of semantic descriptors for APIs for open government data. The research question motivating the work is: what are the recurrent *types of open datasets* relevant in an open government context? The assumption is that providing an answer to this question is key to the development of new application programming interfaces, which will ease data access to programmers and reduce barriers to data re-use. Imagine for example programmers accessing environmental datasets from two catalogs CAT1 (belonging to City 1) and CAT2 (belonging to City 2) using:

- GET `http://cat1.city1.com/dataset/environment`;
- GET `http://cat2.city2.com/dataset/environment`.

Such a situation would be a great improvement over the current state of affairs where datasets are accessed via cryptic items' identification numbers[5] (IDs). First, it is more user-friendly to interact with APIs which return items according to their types (e.g., climate, finance), rather than their IDs. APIs which return data items according to their types are termed *semantic APIs* in this paper. Since RESTful APIs require the definition of semantic descriptors, there are an appropriate architectural style for the technical implementation of semantic APIs. Second, naming schemes re-used consistently by many API designers will create an environment where programmers'

[5] For instance, an example url to retrieve a table in Comma Separated Value (CSV) from a ckan catalog is `http://giv-oct.uni-muenster.de:5000/api/action/datastore_search?resource_id=a774f073-ba31-44e3-8edb-ed0fca79c216&limit=5`.

learning struggles for the re-use of datasets from different open data catalogs in their application could be drastically reduced[6]. Third, semantic APIs contribute to greater transparency. As Michener and Bersch, 2013 pointed out, transparency has two dimensions, namely visibility and inferability. Visibility means that the information is (i) reasonably complete and (ii) found with relative ease; inferability refers to the degree to which the information at hand can be used to draw accurate inference. A semantic API increases information visibility (i.e., it makes available the *types of data* which are used while building city applications).

4.2 A survey of existing categories for open government data

Many open data catalogs classify the open datasets they provide according to categories (the term 'theme' is occasionally used to denote these categories). The goal of this section is to survey these categories across different European countries and extract some recurrent patterns (if any). 40 open data catalogs from four different countries - Germany, Spain, France and the UK - are assessed. These countries were chosen partly because the authors of this paper are native speakers of these languages. In addition, the four countries are currently among the top 10 European countries which are most-ready for open data initiatives according to the Open Data Barometer[7]. The steps followed in collecting this data were as follows:

- **Step 1**: go through the catalog, and list the categories offered by each of them;
- **Step 2**: translate the categories in the target language. The target language used for this paper is English (all researchers understand it), but it is conceivable that other target languages (e.g., German, French, Spanish) could have been used for the same purpose;
- **Step 3**: harmonize the terms' translation across the four countries;

[6] It is a well-known fact that "*programmers at all levels, from novices to experts, repeatedly spend significant time learning new APIs*" (Myers and Stylos, 2016); and "*until [...] standards are more universal, coders must write numerous interfaces for each city and maintain them individually*" (M. Lee et al., 2016). In the current context, a programmer willing to use datasets from two different catalogs would need to go through two learning phases to become familiar with the naming policies of their different APIs. Learning of the APIs' interfaces (and maintenance of the Apps built with these interfaces) may never be entirely removed, but it could be reduced to the strict minimum if similar naming policies were adopted while developing semantic APIs for current open data catalogs.

[7] The UK ranks 1st in Europe (1st worldwide), France ranks 2nd in Europe (2nd Worldwide), Germany ranks 7th in Europe (11th worldwide), and Spain ranks 8th in Europe (12th worldwide) according to (The World Wide Web Foundation, 2015). 'Readiness', in (The World Wide Web Foundation, 2015), means the degree of preparation for, as well as the policies in place to support open data initiatives.

- **Step 4**: generate descriptive statistics about the dataset.

The data collection took place from June 20th to July 6th 2016. Steps 1 and 2 were carried out independently by the first, second and fourth author of the paper[8]. Conjunctions meaning 'AND' in the original languages were left out during the translation because semantic descriptors for APIs should be single words. Step 3 (harmonization) is necessary because of the possibility of translating certain terms differently in English. For instance, the terms 'labour' and 'job' were both present in the dataset after the researchers performed the initial translation. After the harmonization, only the term 'job' has been kept in the dataset to facilitate the comparison of the results across countries. The choice of 'job' rather than 'labour' is, of course, a matter of personal preference, and does not influence the validity of the final conclusions. In addition, Step 3 was useful to prepare the dataset in a format useful for further processing. For example 'urban planning' was transformed into 'urbanplanning' so that it is treated as a single word (and thus data category). Words with hyphens (e.g., E-Administration, procès-verbaux) were also converted into single words during this step. Step 3 was performed through discussions between the first, second, and fourth author. Finally, Step 4 has consisted in counting the frequencies of the different data categories[9], and is meant to help answer two questions:

- What data categories *must* semantic API designers consider? The answer to this question are data categories which appear in *all* catalogs;

- What data categories *could* semantic API designers consider? The answer to this question are all data categories obtained from our data collection (or put differently, data categories which appear *at least once* in the dataset). Table 4.1 presents all catalogs surveyed, their spatial granularities (i.e., whether they catalog datasets for a city, a region, or the whole country), as well as their URLs.

Tab. 4.1: Open data catalogs surveyed

Catalog name	Granularity	URL
	GERMANY	
OffeneDaten.de	country	`https://offenedaten.de/`
GovData	country	`https://www.govdata.de/web/guest/daten`
Open Data Berlin	city	`http://daten.berlin.de/`

[8] Only three were needed during these steps because the first author speaks French as native language, and German fluently. This researcher has also collected the data for Germany.

[9] The online tool Online-Utility.org (http://www.online-utility.org/, last accessed: July 5th, 2016) was used to count the frequencies of the different data categories.

Open Data Köln	city	http://www.offenedaten-koeln.de/dataset
Open Data HRO	city	http://www.opendata-hro.de/group
Open Data München	city	https://www.opengov-muenchen.de/dataset
Open Data ULM	city	http://daten.ulm.de/datenkatalog/offene_daten
Open Government Data Portal Rheinland-Pfalz	region	http://daten.rlp.de/group
Open NRW	region	https://open.nrw/de/dat_kat
Transparenzportal Hamburg	city	http://transparenz.hamburg.de/
	SPAIN	
Datos Abiertos JCYL	region	http://www.datosabiertos.jcyl.es/
Datos Abiertos Junta de Andalucía	region	http://www.juntadeandalucia.es/datosabiertos/portal.html
Datos Abiertos Madrid	city	http://datos.madrid.es/
Open data Ajuntament de Valencia	city	http://gobiernoabierto.valencia.es/
Open data Aragon	region	http://opendata.aragon.es/
OpenDataBCN	city	http://opendata.bcn.cat/opendata/
Open Data Euskadi	region	http://opendata.euskadi.eus/w79-home/eu/
Open data Gobierno de Canarias	region	opendata.gobiernodecanarias.org/
Open Data Navarra	region	www.gobiernoabierto.navarra.es/es/open-data
Portal Open Data Xunta de Galicia	region	abertos.xunta.gal/
	FRANCE	
data.gouv.fr	country	http://www.data.gouv.fr/fr/datasets/
Data GrandLyon	region	http://data.grandlyon.com/
Montpellier Territoire Numérique	city	http://opendata.montpelliernumerique.fr/Les-donnees
Nantes Ouverture des Données	city	http://data.nantes.fr/
Open Data Nice Côte d’Azur	region	http://opendata.nicecotedazur.org/site/news
Open Data Bordeaux	city	http://opendata.bordeaux.fr/catalogue-des-donnees
Open PACA	region	http://opendata.regionpaca.fr/donnees.html?no_cache=1

ParisData	city	http://opendata.paris.fr/page/home/
Rennes métropole en accès libre	city	http://www.data.rennes-metropole.fr/les-donnees/catalogue/
Toulouse Métropole Data	city	https://data.toulouse-metropole.fr/page/home/
	UNITED KINGDOM	
Birmingham DataFactory	city	https://data.birmingham.gov.uk/dataset
Bournemouth Data Stream	city	http://bournemouthdata.io/
Data.gov.uk	country	https://data.gov.uk/
Data- Liverpool City Council	city	http://liverpool.gov.uk/council/key-statistics-and-data/data/
Edinburgh Open Data Portal	city	http://edinburghopendata.info/
Leeds Data Mill	city	http://leedsdatamill.org/
London Datastore	region	http://data.london.gov.uk/
Open Data Bristol	city	https://opendata.bristol.gov.uk/
OpenDataNI	region	https://www.opendatani.gov.uk/
Sheffield City Council Open Data	city	https://data.sheffield.gov.uk/

Appendices A1, B1, C1, and D1 presents all catalogs' data categories as well as their translations into English. Figure 4.1 presents the categories' respective frequencies for Germany. The figure shows 48 distinct categories[10]. The figure shows also that five terms are used in all catalogs surveyed, namely: *culture, elections, education, sport,* and *economy*. Figure 4.2 shows example categories for Spanish open data catalogs. The word frequency count for Spanish open data catalogs yields 65 distinct categories. Contrary to the German case, none of the categories shown appear in all catalogs. The categories *culture, leisure, economy, education, health, environment, transport, tourism* and *employment* seem the most popular, with 8 of 10 of the catalogs surveyed proposing them for the access of open data. The French catalogs surveyed present 78 distinct categories, some of which are shown in Figure 4.3. *Culture* is both the most popular, and the only category which appears in all French catalogs surveyed. Figure 4.4 presents the different categories from the UK's catalogs, as well as their respective frequencies. 59 distinct categories were

[10]Actually, the categories 'tax' and 'taxes' could have been merged into one single category, reducing this number to 47. However, both categories were kept distinct, because of the different spellings (i.e., 'Steuern', and 'Steuer') in the original German open data catalogs.

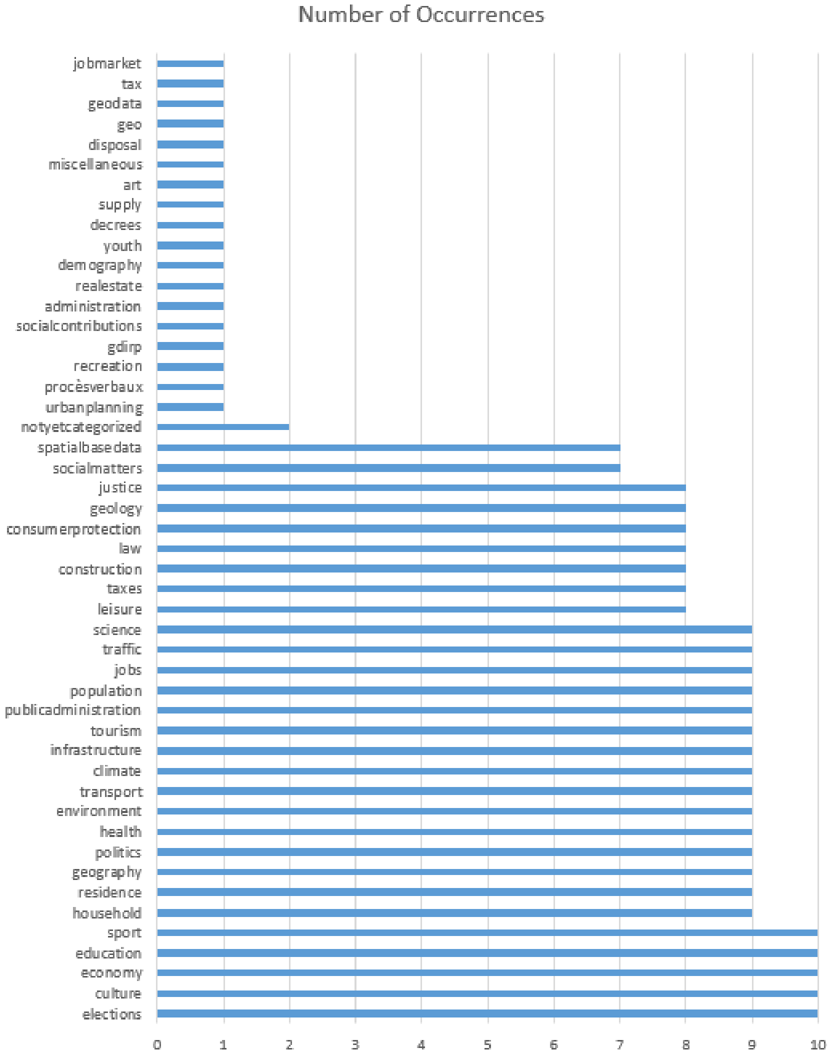


Fig. 4.1: Categories of German open data catalogs and their frequencies

obtained[11] as the result of the words frequencies count. The categories of *education* and *health* are the most popular among the UK catalogs with 9 occurrences each.

4.3 Towards semantic APIs for open government data

Section 4.2 has surveyed 40 European open data catalogs and the different categories they provide for open data access. This process yielded 171 distinct words, which were used by providers of open datalogs to offer access to their data. This section discusses in detail how the categories gathered in the previous section can inform the choice of core categories for semantic APIs for cities. The section also briefly

[11]The categories 'art' and 'arts' could have been merged into one single category. However, both categories were kept distinct, because of the different spellings (i.e., 'art', and 'arts') in the original open data catalogs from the UK.

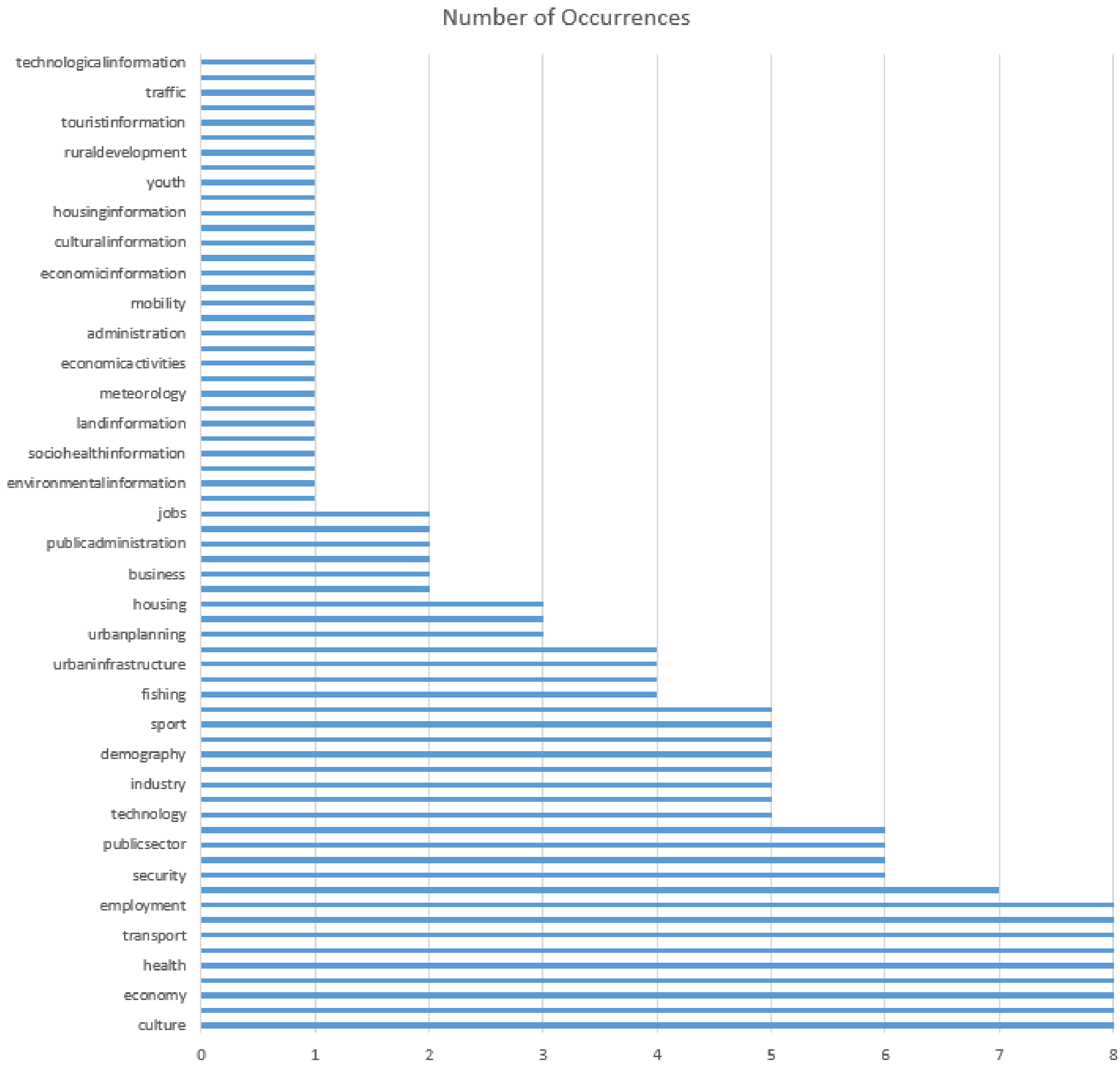


Fig. 4.2: Categories of Spanish open data catalogs and their frequencies

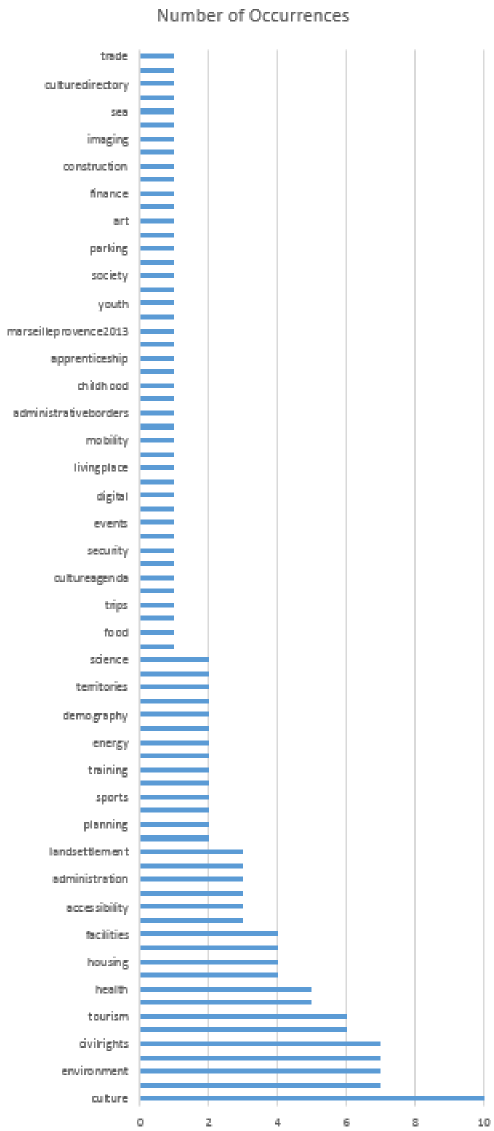


Fig. 4.3: Categories of French open data catalogs and their frequencies

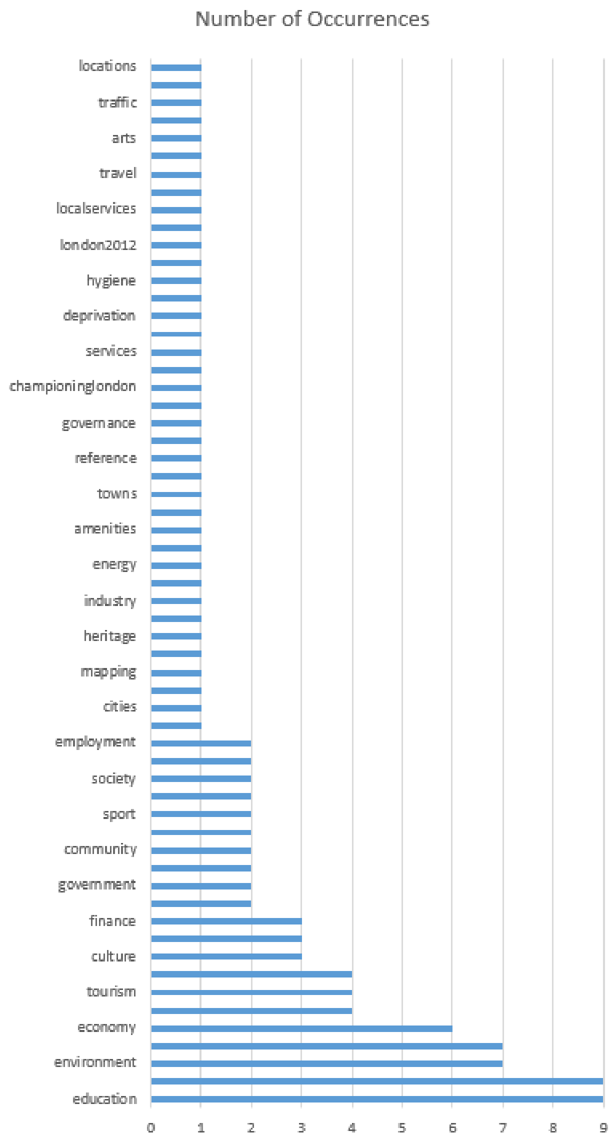


Fig. 4.4: Categories of the UK's open data catalogs as well as their frequencies

touches upon the different technical components useful to implement a semantic API for open government data.

4.3.1 Categories for a semantic API for open government data

The main motivation behind this work is to shed some light on the recurrent *types of open datasets* relevant in an open government context. Beyond informing the design of semantic APIs, the recurrent types of open datasets are an indicator of the *topics of interests in the respective countries*, the types of questions data publishers *assume* users will ask, and ultimately the types of questions citizens *can* ask. It is worth mentioning that some categories from the surveyed catalogs (appendices A1, B1, C1, and D1) could have been translated differently from how they were in this article. For example, 'Wohnen' was translated as 'residence', but could have also been translated as 'habitation' or 'home' (which are appropriate synonyms for the word retrieved from the Merriam-Webster dictionary[12]); 'Arbeit' was translated as 'jobs' ('Arbeitsmarkt' was translated as 'job market'), but the world 'labour' could have been used in lieu of 'jobs'; 'Stadtplanung' could have been translated as 'city planning' instead of 'urban planning'; and so on. This limitation is an inherent limitation of all studies which will endeavour to compare different categories offered by open data providers in Europe. The harmonization stage (Step 3, Section 2) has ensured that the translation remained consistent both within tables, and across countries. Two questions were mentioned in Section 2, namely: what data categories *must* semantic API designers consider? And what data categories *could* semantic API designers consider? The first question is referred to as Q1, and the second as Q2 in the rest of the paper. Both are now considered in turn.

Categories for national semantic APIs

Appendices A2, B2, C2, and D2 present the inter-catalog agreements for the different catalogs surveyed in Germany, Spain, France, and the UK respectively. The values for inter-catalog agreement were computed using the Jaccard index, i.e., the size of the intersection of two sets divided by the size of their unions. The indices show some great differences between the countries examined: the average inter-catalog agreement for German catalogs is 0.70 (standard deviation: 0.21); this value drops to 0.34 (standard deviation: 0.31) for Spanish open data catalogs; the mean inter-catalog agreement for French open data catalogs is 0.18 (standard deviation: 0.08); and the average inter-catalog agreement for catalogs from the UK surveyed is 0.19 (standard deviation: 0.09) [13]. The differences between the average inter-catalog

[12]See `http://www.merriam-webster.com/dictionary/residence` (last accessed: July 1st, 2016).

[13]The Jaccard indices and the statistical values were calculated using two open source libraries, namely `https://github.com/ecto/jaccard` and `https://github.com/simple-statistics/simple-statistics` respectively.

agreements in the countries surveyed indicates that the level of harmonization of terms used in open data catalogs accross these countries is, at the moment, quite disparate.

The averages of inter-catalog agreements in each country were also computed taking into account the spatial granularities (see appendices A2, B2, C2, and D2), and are as follows:

- Germany: 0.57 (city), no average at the regional level because there were only two catalogs available at this level, no average at the national level computed because there were only two catalogs in the datasets for the national level[14];
- Spain: 0.17 (city), 0.37 (region), no average at the national level computed because there is no catalog in the dataset for the national level;
- France: 0.16 (city), 0.22 (region), no average at the national level computed because there is only one catalog in the dataset for the national level;
- UK: 0.17 (city), no average at the regional level because there were only two catalogs available at this level, no average at the national level computed because there is only one catalog in the dataset for the national level.

These values lead to the following observations: within each of the countries surveyed, the instances of open data catalogs for the national level are too few to draw some useful conclusion; and the inter-catalog agreements at the city level and at the regional level are in general quite low. The former observation is not surprising because there may not be many institutions in a single country which can take an inventory of open government data across a whole country. The latter observation suggests that within each of the country, more effort - at both the local and the regional levels - is needed to harmonize the categories offered by open data providers.

Five terms appeared in all catalogs for open data in Germany surveyed: *culture, elections, education, sport*, and *economy*. Their very high rate of occurrence suggests that they seem inevitable in the German open data landscape, and that designers of semantic APIs for German cities should include them in their own APIs. Put differently, a possible answer to Q1 is *culture, elections, education, sport*, and *economy*. As regards data categories API designers could consider as eligible categories (i.e., Q2), there are different ways of providing an answer:

[14]One needs at least three catalogs to get a meaningful value for the mean inter-catalog agreements.

- Include all categories *already used* by other open data catalog publishers (i.e., the 48 terms from Figure 4.1);

- Set a threshold T ($0 \leq T < 1$) that categories to be included should surpass. T denotes here the frequency of appearance in existing open data catalogs. The choice of T will necessarily involve some degree of conventionality and arbitrariness, but a similar value of T across all open data catalogs makes transnational comparison possible. In the rest of this work, the illustrative value of $T = 0.75$ is chosen, that is, the answer to Q2 is limited to categories which appear in at least 75% of the surveyed catalogs. In the case of Germany, 27 categories fulfill this requirement. These are: *culture, elections, education, sport, economy, population, transport, jobs, geography, household, traffic, health, environment, residence, science, climate, tourism, publicadministration, infrastructure, politics, justice, construction, taxes, consumerprotection, leisure, law,* and *geology.*

With respect to Spanish open data catalogs, no term seems so popular that it can be deemed as inevitable (Q1). However, the data collected suggests that designers of semantic APIs for Spanish cities could consider the following nine terms (threshold of appearance $T = 0.75$): *culture, transport, education, health, environment, leisure, tourism, economy* and *employment*. *Culture* is the only term appearing in all French open data catalogs surveyed. It imposes thus itself as a category of semantic APIs for French cities (Q1). However (and contrary to Germany and Spain), no other term appears in at least 75% of the catalogs surveyed to be suggested as a possible answer to Q2. As to open data catalogs in the UK, no term appears in all of the catalogs surveyed (Q1). Nevertheless, *health* and *transport* appear in at least 75% of the catalogs surveyed, and could be considered while designing semantic APIs for the UK's cities (Q2).

Categories for bi-national semantic APIs

What if semantic API designers want to provide APIs for a bi-national audience? Relevant categories for this task are the *intersection* of the two countries' sets of categories. With respect to Q1 (i.e., terms which appear in all catalogs surveyed), and Q2 (i.e., terms which appear in at least 75% of the catalogs surveyed), the following categories are possible answers:

- Germany-Spain: Q1 (no category found); Q2 (*culture, economy, education, health, environment, transport, tourism, leisure* and *sport*);

- Germany-France: Q1 (*culture*); Q2 (*culture, economy, environment, transport, education* and *tourism*);
- Germany-UK: Q1 (no category found); Q2 (*education, health, economy, environment* and *transport*);
- Spain-France: Q1 (no category found); Q2 (*culture, economy, environment* and *transport*);
- Spain-UK: Q1 (no category found); Q2 (*education, health, environment* and *transport*);
- France-UK: Q1 (no category found); Q2 (no category found).

Table 4.2 presents the values of the Jaccard indices between the different countries. The Jaccard indices show the inter-country agreement between the categories offered by the open data providers. This inter-country agreement is again quite low, oscillating between 0.14 and 0.25. Below are the values of the inter-catalog agreements for open data catalogs at the same spatial granularity. They indicate that this low inter-catalog agreement is more or less homogenously present at all levels (local, regional and national).

- Germany-Spain: 0.26 (city), 0.23 (region), no value computed because there is no Spanish catalog in the dataset for the national level;
- Germany-France: 0.20 (city), 0.18 (region), 0.23 (country);
- Germany-UK: 0.17 (city), 0.20 (region), 0.14 (country);
- Spain-France: 0.25 (city), 0.20 (region), no value computed because there is no Spanish catalog in the dataset for the national level;
- Spain-UK: 0.24 (city), 0.18 (region), no value computed because there is no Spanish catalog in the dataset for the national level;
- France-UK: 0.19 (city), 0.14 (region), 0.2 (country).

Tab. 4.2: Jaccard indices showing inter-country agreement between data categories offered by open data providers

Country (A)	Jaccard Index (A, France)	Jaccard Index (A, Germany)	Jaccard Index (A, Spain)	Jaccard Index (A, UK)

France	1	0.17	0.25	0.17
Germany	0.17	1	0.24	0.14
Spain	0.25	0.24	1	0.18
UK	0.17	0.14	0.18	1

Categories for a transnational semantic API

There are various ways of obtaining a list of useful categories to consider while designing semantic APIs at a transnational level. One way is to adopt a minimalistic approach, i.e., only categories from the national level appearing in all countries should be considered. This approach yields no category for a transnational semantic API (i.e., no answer for Q1). A variant of the minimalist approach is to consider categories which appear in at least three of the four countries to be relevant. In that case, *culture, health* and *transport* would be good candidate categories for a transnational semantic API (Q1).

An alternative to the minimalist approach would be a maximalist approach, i.e., each of the category from the national level should be included at the transnational level. This results in the following list of 27 categories (mostly inherited from German open data catalogs) for a transnational semantic API: *culture, elections, education, sport, economy, population, transport, jobs, geography, household, traffic, health, environment, residence, science, climate, tourism, publicadministration, infrastructure, politics, justice, construction, taxes, consumerprotection, leisure, law,* and *geology*. These categories are also possible answers to Q2.

A third way of generating categories for a transnational semantic API is to compute descriptive statistics using all terms from all catalogs surveyed (appendices A1, B1, C1 and D1) altogether. Figure 4.5 presents the list of terms resulting from this approach (only the 20 terms which occurred the most are shown). If the threshold of appearance is set to 0.75 (i.e., T= 0.75), the candidate list of terms obtained (Q2) is *education, culture, economy, health, environment* and *transport*.

Discussion

Section 4.3.1 has examined possible answers to the two questions motivating this work: what data categories *must* semantic API designers consider? And what data categories *could* semantic API designers consider?. There are a couple of insights which can be summarized from these sections. First, the sections illustrate that an empirical approach to generate terms for semantic APIs is applicable, and may be used to generate terms for semantic APIs at the local, regional and national levels. However, these sections also illustrate that the answers obtained are strongly

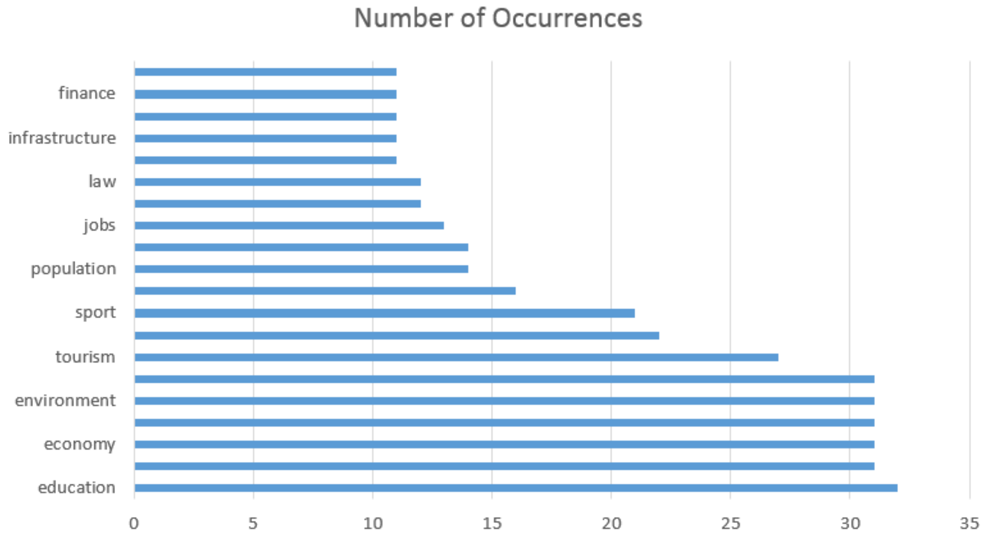


Fig. 4.5: Candidate categories for a transnational semantic API

dependent on the approach taken. Since any of the approach mentioned necessarily involves some degree of conventionality and arbitrariness, any empirical approach to generate categories for semantic APIs should make explicit what the underlying parameters (e.g., maximalist vs minimalist approach, threshold of appearance) are to facilitate traceability.

Second, previous sections have presented inter-catalog agreements from different perspectives. The following conclusions can be drawn based on the values obtained:

- The four countries examined are non-homogeneous with respect to level of harmonization of terms used in their open data catalogs;
- Within each of the countries, the inter-catalog agreements at the city level and at the regional level are in general quite low;
- There is also a low inter-catalog agreement between terms used in a country, and terms used in another (in average less than 30%). In other words, the probability that a category chosen by a data provider in one European Country will also be chosen by a data provider in another European country is somewhere less than 30 percent.

The consistently low values obtained for inter-catalog agreements remind of the 'vocabulary problem', i.e., the low probability that two people use the same term to refer to a specific object in computer applications. The solution to this problem proposed in (Furnas et al., 1987) is unlimited aliasing, i.e., the provision of many alternative words to users so that they can get what they want from large and complex systems. Unlimited aliasing is not entirely suitable for the case of semantic API design, since designers can only choose one term as entry point for their data items. A possible solution to the problem (perhaps the only one?) is the *coordination of efforts* between different data providers. Coordination, in this case, would involve some commitment from different data providers (local, regional, national) to use a set of terms, with agreed upon definitions, in their catalogs. Finally, Susha et al., 2015 identified a set of 12 critical factors for the success of open data initiatives. One of these factors is to "Integrate metadata schemas and *federated controlled vocabularies for properly categorizing information*" (emphasis added). Susha et al.'s study derived the factor from two workshops conducted with a number of experts. The consistently low values for inter-catalog agreements, obtained from an empirical survey of existing open data catalog categories, confirm the need to implement this success factor in current open government initiatives from a different perspective. Moreover, the low values for inter-catalog agreements suggest that *federated controlled vocabularies for properly categorizing information* is necessary both a local, regional and national level in the four European countries surveyed.

There are also a couple of limitations of the study worth mentioning: one limitation is that the results are dependent on the quality of the translated terms in English. As mentioned in Section 4.2, some of the terms could have been translated differently yielding slightly different results. These effects were minimized through Step 3 and Step 4 of the method. In addition, since no information is available on how the categories were chosen by the open data provider (e.g., based on institutional mandate, or simply because it fits best their existing data, etc.), there are some limits to the explanatory power of this study (i.e., why some of the difference are observed).

4.3.2 Technical components of a semantic API

As mentioned in Section 4.1, one of the benefits of semantic APIs is to increase transparency. This happens because semantic APIs improve information visibility (i.e., they make available the *types of data* which are used while building applications with open government data). The implementation of semantic APIs necessitates some technical considerations which are discussed in this section. Six main components are needed to realize them: a metadata-management component, a registration component, a logging layer, a semantic layer, a connector, and the databases. All components introduced are illustrated in Figure 4.6. Their role is described below:

- **Registration component**: this component is helpful to register any developer who wants to use a semantic API to build city applications. After a successful registration, a developer receives an access-token[15] which will be used to identify the app making calls to the semantic API. The ability to know *who is making an API call* is one feature of semantic APIs;

- **Logging layer**: this component records all events related to the semantic API, for example, the ids of the applications which request a certain type of dataset (the applications can be automatically identified using via the access-token), the types of data requested (e.g., culture, health or transport), the number of applications accessing a certain data, and the frequency of API calls. The log files generated by the logging layer can be formatted using a well-known open format such as the common log format presented in (World Wide Web Consortium, 1995). The logging layer generates information about *what is happening* with the API, and *when*.

- **Metadata-management component**: the role of this component is to establish all mappings between the requests of the users, and the databases relevant to process these requests. It is the 'brain' of the semantic API because it stores all relevant conceptual relationships for the functioning of the API. This component is built and maintained by the API provider (e.g., an institution such as a city council). The metadata-management component would specify for example that 'health' in English is equivalent to 'Gesundheit' (in German) and that any request related to health is also a request about 'Gesundheit'. The metadata-management component can also specify hierarchical relationships between concepts (e.g., a request about health is a request about all items with a direct relation to the 'health' concept, and more specific health concepts such as 'health insurance' and 'preventive care').

- **Semantic layer**: this component provides a bridge between the request of the user (or software agent) and the metadata-management component; it retrieves the databases (and all concepts) to look for based on the user requests, and forwards this information to the connector layer.

- **Connector-layer**: The queries performed on the databases, as well as the query languages (e.g., SQL, SPARQL, interfaces of a RESTful API) needed to return data items are dependent on the user request and the database being queried. This layer stores therefore the different queries needed to retrieve specific datasets from the databases.

[15] See `https://msdn.microsoft.com/en-us/library/Aa374909.aspx` (last accessed: July 7th, 2016) for a short introduction to access-tokens.

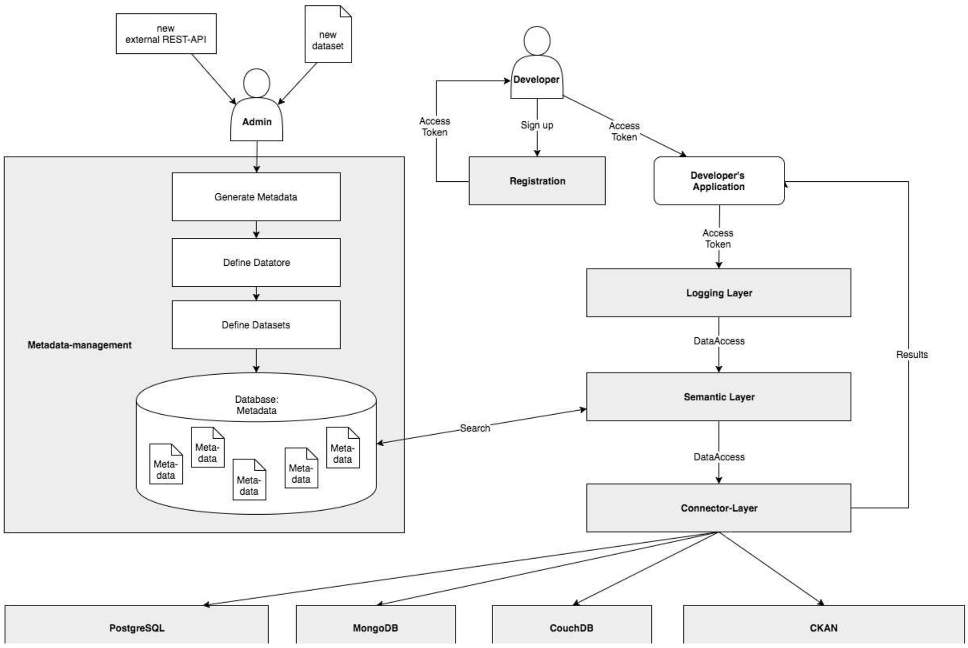


Fig. 4.6: Generic components of a semantic API

- **Databases**: they store the data which can be of any type: relational data (PostgreSQL), document-oriented data (MongoDB, CouchDB), data coming from CKAN-based platforms, graph-based data (stored in triple stores such as Parliament, Fuseki or Virtuoso), or even data coming from other semantic APIs.

The JavaScript development environment Node.js[16] is currently used to implement these components. The main reason for choosing Node.js is its portability; all Node.js applications (irrespective of their functionality) can be run using two commands, namely 'npm install' followed by 'npm start'. That is, any city council could install and run the API with relative ease (i.e., only two commands). Semantic APIs is an essential component for the realization of the vision of the Open City Toolkit described in (Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016). Figure 4.7 illustrates one practical use of a semantic API, namely generate information about applications in a city which access some types of datasets, and also types of datasets which are often requested in a city. The example on the figure as well as the documentation of the features of the API implemented can be accessed from `https://github.com/geo-c/OCT-Core`.

[16]See `https://nodejs.org/en/` (last accessed: July 7th 2016) for further information about Node.js.

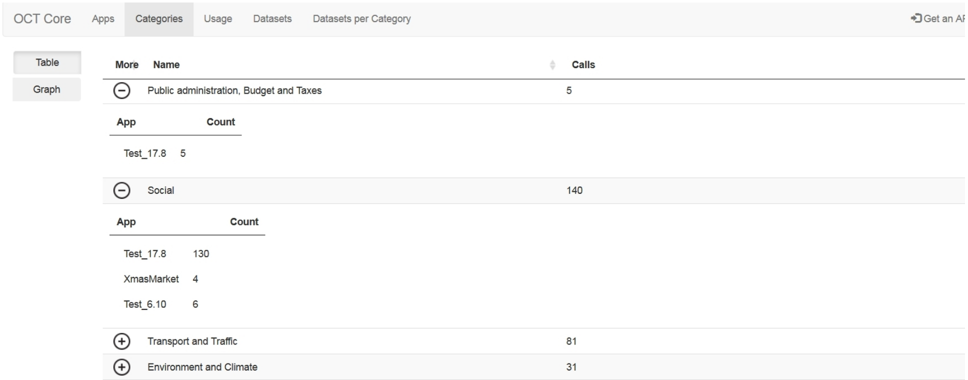


Fig. 4.7: An example of practical use of semantic API - information can be produced about applications which access some types of datasets in a city. Datasets of type 'social' seem to be the most requested in this example; the app Test_17.8 re-uses dataset which are related to the categories 'Public administration, Budget and Taxes', and 'Social'

4.4 Related work

To the best of the authors' knowledge, there has not been any attempt to classify the different categories of open data catalogs to inform the design of semantic APIs in the literature. Example classifications of city data appear in (Lecue et al., 2012; Bischof et al., 2014), yet they were not generated based on an empirical consideration of data categories provided by current open data catalogs. Attard, Orlandi, Scerri, et al., 2015 provided a systematic survey of open (government) data initiatives but did not specifically look at categories offered by open data catalogs.

From a practical perspective, designers of semantic APIs could resort also to categories provided in (The World Wide Web Foundation, 2015), though these categories were not proposed to this end. The categories found in (The World Wide Web Foundation, 2015) are: *maps, land, statistics, budgets, spending, companies, legislation, transport, trade, heath, education, crime, environment, elections*, and *contracts*. Since these categories were already used to assess the progress of 92 countries with respect to their adoption of open data, they could also be considered while providing access to open data. Nevertheless there is no information as to the reason why these specific categories were chosen to perform the assessment in (The World Wide Web Foundation, 2015), and further iterations may find alternative classifications in this article.

A spanish standard called UNE 178301:2015 (Aenor, 2015) was elaborated by a group of spanish smart cities. UNE 178301:20 defines a set of indicators divided into five categories: political, organizational, technical, legal and economy. Only, one category of this standard appears in Figure 4.2, namely 'economy'. Also, this standard defines the metrics to quantify the level of the open data in Spanish cities.

Another possible source of categories for the design of semantic APIs is the European Data Portal which "harvests the metadata of Public Sector Information available on public data portals across European countries"[17]. Categories for data access offered by this portal are: *agriculture, fisheries, forestry, foods, energy, regions, cities, transport, economy, finance, internationalissues, government, publicsector, justice, legalsystem, publicsafety, environment, education, culture, sport, health, population, society, science*, and *technology*. A good sign is that 11 of these 25 categories appear in Figure 4.5. There are chances that the team developing the European Data Portal has asked itself questions similar to those asked in this paper, though there is no information of how the categories were created.

The CitySDK Linked Data API[18] was developed to provide unified and direct access to "open" data, with an interface for writing data. It was designed to work closely with other open source projects such as OpenTripPlanner, OpenTripPlanner Analyst, Open311, GTFS, and OpenStreetMap, where one query about one object provides results from multiple datasets, annotated using semantic web technologies. CitySDK provides a web service offering integrated and direct access to open data from government, commercial and crowd sources identically. The web service is adopted by six European cities. The CitySDK Linked Data API makes data available by collecting data or web services from different sources, describing the data, linking the data to reference datasets when applicable (viz. Cadastre/OSM), offers the data as a unified service to other applications (API), also allowing the applications to annotate and enrich the data. Independent of file format, refresh rate or granularity open data is easily accessible for commercial use, research and software developers. The research in this paper considered the European cities' open data held by CitySDK platform and incorporated the strength of semantics links for those data sets to yield the result for the openness of the respective city/country.

The data API from Data.gov.uk is RESTful, and may be considered the most advanced implementation of semantic APIs in the current open data landscape. However only two categories are available at the moment of this writing, namely *health* and *transport*[19]. Though the API offers the opportunity to retrieve datasets according to these categories, its documentation says nothing about registration & logging capabilities which are the pre-requisite for increased transparency as regards the use of data sources in a city context. The categories obtained in this work as well as the technical discussion in Section 4.3.2 provide a solid ground for making this API more sophisticated at the technical level, and adding new topics to it.

[17]See `http://www.europeandataportal.eu/en/what-we-do` (last accessed: July 8th, 2016).
[18]`http://www.citysdk.eu/mobility/` (last accessed: November 11th, 2016).
[19]See `https://data.gov.uk/data/api/` (last accessed: July 8th, 2016).

4.5 Conclusion

As the recent edition of the Open Data Barometer (The World Wide Web Foundation, 2015) has pointed out, open data is entering the mainstream but more elaborated APIs that facilitate access to data are still very rare among government data. This work has proposed that semantic APIs could be such 'elaborated APIs', and presented their technical components. The work pointed also out that the REST architectural style is an adequate paradigm for the implementation of semantic APIs. As semantic APIs rely on data categories to make data items available to both programmers and machines, this paper has looked into data categories relevant for semantic APIs designers in European countries. The article has surveyed 40 European data catalogues from four countries (France, Germany, Spain, and the United Kingdom) and observed the recurrent data categories offered by open data providers in these countries. The results show great disparities between the countries surveyed, but suggest that *culture*, *health* and *transport* would be good candidate categories for a transnational semantic API. The results also show that the probability of inter-country agreement between open data catalogs is less than 30 percent. This suggests that effort is needed with respect to coordination among countries so that semantic APIs built in one country have greater chances of adoption by other countries. Any study like the one presented in this paper is dependent upon the quality of the translation of terms between the languages (which is by definition never perfect) and the translator. This aspect puts some limits on the generalizability of the results. Nonetheless, the merit of this work has been to provide a set of categories based on the current practice to inform the design of semantic APIs. The results obtained stress the need for *federated controlled vocabularies for properly categorizing information* at both a local, regional and national level in the four European countries surveyed.

The data categories surveyed reflect a data provider perspective of the current open data landscape. That is, they give an indication of the types of datasets that open data providers assume citizens will look for. A useful complement to this study could look at the most requested data categories (e.g., number of downloads) of open data catalogs to get an understanding of what citizens actually often look for[20]. Another direction for future work would be a large-scale survey asking programmers across the four countries assessed the types of datasets they would retrieve in case they were provided with semantic APIs for their cities. Finally, future work could also, in addition to the aspects discussed in this paper, have a closer look at other usability factors (e.g., complexity, documentation, error handling: for a complete list, see Zibran et al., 2011) which will favor API adoption in the open data landscape.

[20] At the moment of this writing most of the catalogs surveyed in the paper do not provide this information.

Acknowledgements: The authors gratefully acknowledge funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, http://www.geo-c.eu/). Carlos Granell has been funded by the Ramón y Cajal Programme (grant number RYC-2014-16913). Sergio Trilles has been funded by the postdoctoral programme Vali+d (GVA) (grant number APOSTD/2016/058). We also thank the two anonymous reviewers for their constructive comments.

Appendix A1: 10 German Open Data Catalogs with their respective data categories

Catalog name	Publisher	Data Categories (German)	Data Categories (English)
OffeneDaten.de	Open Knowledge Foundation Germany	Geographie, Geologie und Geobasisdaten, Bildung und Wissenschaft, Umwelt und Klima, Soziales, Infrastruktur, Bauen und Wohnen, Bevölkerung, Öffentliche Verwaltung, Haushalt und Steuern, Verbraucherschutz, Wirtschaft und Arbeit, Transport und Verkehr, Kultur, Freizeit, Sport und Tourismus, Gesundheit, Politik und Wahlen, Noch nicht kategorisiert, Gesetze und Justiz	geography, geology, spatialbasedata, education, science, environment, climate, infrastructure, construction, residence, population, publicadministration, household, taxes, consumerprotection, economy, jobs, transport, traffic, culture, leisure, sport, tourism, health, politics, elections, notyetcategorized, law, justice
GovData	Finanzbehörde Hamburg, Feschäfts- und Koordinierungsstelle GovData	Bevölkerung, Bildung und Wissenschaft, Geographie, Geologie und Geobasisdaten, Gesetze und Justiz, Gesundheit, Infrastruktur, Bauen und Wohnen, Kultur, Freizeit, Sport und Tourismus, Politik und Wahlen, Soziales, Transport und Verkehr, Umwelt und Klima, Verbraucherschutz, Öffentliche Verwaltung, Haushalt und Steuern, Wirtschaft und Arbeit	Population, education, science, geography, geology, spatialbasedata, law, justice, health, infrastructure, construction, residence, culture, leisure, sport, tourism, politics, elections, socialmatters, transport, traffic, environment, climate, consumerprotection, publicadministration, household, taxes, economy, jobs
Open Data Berlin	Senatsverwaltung für Wirtschaft, Technologie, und Forschung	Arbeitsmarkt, Bildung, Demographie, Geographie und Stadtplanung, Gesundheit, Jugend, Kunst und Kultur, Öffenliche Verwaltung, Haushalt, und Steuern, Protokolle und Beschlüsse, Sonstiges, Sozialleistungen, Sport und Erholung, Tourismus, Umwelt und Klima, Ver- und Entsorgung, Verbraucherschutz, Verkehr, Wahlen, Wirtschaft, Wohnen und Immobilien	Jobmarket, education, demography, geography, urbanplanning, health, youth, art, culture, publicadministration, household, taxes, procèsverbaux, decrees, miscellaneous, sport, socialcontributions, recreation, tourism, environment, climate, supply, disposal, consumerprotection, traffic, elections, economy, residence, realestate

Open Data Köln	Stadt Köln	Geo, Bevölkerung, Politik und Wahlen, Transport und Verkehr, Umwelt und Klima, Verwaltung, Haushalt und Steuern, Kultur, Freizeit, Sport und Tourismus, Soziales, Bildung und Wissenschaft, Infrastruktur, Bauen und Wohnen, Gesundheit, Gesetzte und Justiz, Wirtschaft und Arbeit	Geo, population, politics, elections, transport, traffic, environment, climate, administration, household, taxes, culture, leisure, sport, tourism, socialmatters, education, science, infrastructure, construction, residence, health, law, justice, economy, jobs
Open Data HRO	Hansesdadt Rostock	Bevölkerung, Bildung und Wissenschaft, Geographie, Geologie und Geobasisdaten, Gesetze und Justiz, Gesundheit, Infrastruktur, Bauen und Wohnen, Kultur, Freizeit, Sport und Tourismus, Politik und Wahlen, Soziales, Transport und Verkehr, Umwelt und Klima, Verbraucherschutz, Öffentliche Verwaltung, Haushalt und Steuern, Wirtschaft und Arbeit	Population, education, science, geography, geology, spatialbasedata, law, justice, health, infrastructure, construction, residence, culture, leisure, sport, tourism, politics, elections, socialmatters, transport, traffic, environment, climate, consumerprotection, publicadministration, household, taxes, economy, jobs
Open Data München	Landeshauptstadt München	Bevölkerung, Wirtschaft und Arbeit, Geographie, Geologie und Geobasisdaten, Transport und Verkehr, Kultur, Freizeit, Sport und Tourismus, Soziales, Politik und Wahlen, Infrastruktur, Bauen und Wohnen, Bildung und Wissenschaft, Öffentliche Verwaltung, Haushalt und Steuern, Gesundheit	Population, economy, jobs, geography, geology, spatialbasedata, transport, traffic, culture, leisure, sport, tourism, socialmatters, politics, elections, infrastructure, construction, residence, education, science, publicadministration, household, taxes, health
Open Data ULM		Geographie, Geologie und Geobasisdaten, Bildung und Wissenschaft, Umwelt und Klima, Soziales, Infrastruktur, Bauen und Wohnen, Bevölkerung, Öffentliche Verwaltung, Haushalt und Steuer, Verbraucherschutz, Wirtschaft und Arbeit, Transport und Verkehr, Kultur, Freizeit, Sport und Tourismus, Gesundheit, Politik und Wahl, Gesetze und Justiz	geography, geology, spatialbasedata, education, science, environment, climate, infrastructure, construction, residence, population, publicadministration, household, tax, consumerprotection, economy, jobs, transport, traffic, culture, leisure, sport, tourism, health, politics, elections, notyetcategorized, law, justice

Open Government Data Portal Rheinland-Pfalz	Ministerium des Innern, für Sport und Infrastruktur des Landes Rheinland-Pfalz	Bevölkerung, Bildung und Wissenschaft, GDI-RP, Geographie, Geologie und Geobasisdaten, Gesundheit, Infrastruktur, Bauen und Wohnen, Gesetze und Justiz, Kultur, Freizeit, Sport und Tourismus, Politik und Wahlen, Soziales, Transport und Verkehr, Umwelt und Klima, Verbraucherschutz, Öffentliche Verwaltung, Haushalt und Steuern, Wirtschaft und Arbeit	Population, education, science, GDIRP, geography, geology, spatialbasedata, health, infrastructure, construction, residence, law, justice, culture, leisure, sport, tourism, politics, elections, socialmatters, transport, traffic, environment, climate, consumerprotection, publicadministration, household, taxes, economy, jobs
Open NRW	Innenministerium NRW	Bevölkerung, Bildung und Wissenschaft, Geographie, Geologie, und Geobasisdaten, Gesetze und Justiz, Infrastruktur, Bauen und Wohnen, Kultur, Freizeit, Sport und Tourismus, Öffentliche Verwaltung, Haushalt und Steuern, Politik und Wahlen, Soziales, Transport und Verkehr, Umwelt und Klima, Verbraucherschutz, Wirtschaft und Arbeit	Population, education, science, geography, geology, spatialbasedata, law, justice, infrastructure, construction, residence, culture, leisure, sport, tourism, publicadministration, household, taxes, politics, elections, socialmatters, transport, traffic, environment, climate, consumerprotection, economy, jobs
Transparenzportal Hamburg	Freie und Hansestadt Hamburg	Bevölkerung, Bildung und Wissenschaft, Geographie, Geologie und Geodaten, Gesetze und Justiz, Gesundheit, Infrastruktur, Kultur und Sport, Politik und Wahlen, Soziales, Transport, Umwelt und Klima, Verbraucherschutz, Öffentliche Verwaltung, Wirtschaft und Arbeit	Population, education, science, geography, geology, geodata, law, justice, health, infrastructure, culture, sport, politics, elections, socialmatters, transport, environment, climate, consumerprotection, publicadministration, economy, jobs

Appendix B1: 10 Spanish Open Data Catalogs with their respective data categories

Catalog name	Publisher	Data Categories (Spanish)	Data Categories (English)
Datos Abiertos JCYL	Junta de Castilla y León	Ciencia tecnología, comercio, cultura ocio, demografía, deporte, economía, educación, empleo, energía, hacienda, industria, legislación justicia, medio ambiente, medio rural pesca, salud, sector público, seguridad, sociedad bienestar, transporte, turismo, urbanismo infraestructuras, vivienda	Science, technology, commerce, leisure, culture, demography, sport, economy, education, employment, energy, finance, industry, law, justice, environment, ruralenvironment, fishing, health, publicsector, security, socialwelfare, transport, tourism, urbaninfrastructure, livingplace
Datos Abiertos Junta de Andalucía	Junta de Andalucía	Ciencia tecnología, comercio, cultura ocio, demografía, deporte, economía, educación, empleo, energía, hacienda, industria, legislación justicia, medio ambiente, medio rural pesca, salud, sector público, seguridad, sociedad bienestar, transporte, turismo, urbanismo infraestructuras, vivienda	Science, technology, commerce, leisure, culture, demography, sport, economy, education, employment, energy, finance, industry, law, justice, environment, ruralenvironment, fishing, health, publicsector, security, socialwelfare, transport, tourism, urbaninfrastructure, livingplace
Datos Abiertos Madrid	Ayuntamiento de Madrid	Ciencia tecnología, comercio, cultura ocio, demografía, deporte, economía, educación, empleo, energía, hacienda, industria, legislación justicia, medio ambiente, medio rural pesca, salud, sector público, seguridad, sociedad bienestar, transporte, turismo, urbanismo infraestructuras, vivienda	Science, technology, commerce, leisure, culture, demography, sport, economy, education, employment, energy, finance, industry, law, justice, environment, ruralenvironment, fishing, health, publicsector, security, socialwelfare, transport, tourism, urbaninfrastructure, livingplace
Open data Ajuntament de Valencia	Ayuntamiento de Valencia	Medio ambiente, sociedad y bienestar, transporte, urbanismo e infraestructuras, salud, turismo, cultura y ocio, sector público, comercio, economía, hacienda, ciencia y tecnología, educación, seguridad y vivienda	Environment, socialwelfare, transport, urbanplanning, infrastructure, health, tourism, culture, commerce, publicsector, trade, economy, finance, science, technology, education, security, housing

Open data Aragon	Gobierno de Aragón	Ciencia tecnología, comercio, cultura ocio, demografía, deporte, economía, educación, empleo, energía, hacienda, industria, legislación justicia, medio ambiente, medio rural pesca, salud, sector público, seguridad, sociedad bienestar, transporte, turismo, urbanismo infraestructuras, vivienda	Science, technology, commerce, leisure, culture, demography, sport, economy, education, employment, energy, finance, industry, law, justice, environment, ruralenvironment, fishing, health, publicsector, security, socialwelfare, transport, tourism, urbaninfrastructure, livingplace
OpenDataBCN	Ajuntament de Barcelona	Territorio, población, Ciudad y servicios, Economía y empresa y Administración	Territory, population, city, services, economy, business, administration
Open Data Euskadi	Gobierno Vasco	Actividades económicas, Administración Pública, Asuntos Sociales, Cultura, Euskera, Educación, Medio Ambiente, Justicia, Meteorología, Ocio y Turismo, Salud, Seguridad e Interior, Transporte y movilidad, Trabajo y Empleo, Urbanismo y territorio, Vivienda	Economicactivities, publicadministration, socialmatters, culture, basquelanguage, education, environment, justice, meteorology, leisure, tourism, health, security, localgovernment, transport, mobility, jobs, employment, urbanplanning, territory, housing
Open data Gobierno de Canarias	Gobierno de Canarias	Sociedad y bienestar, Sector Público, Medio rural, Empleo, Demografía, Urbanismo e infraestructuras, Turismo, Educación, Salud, Economía, Transporte, Medio Ambiente, Hacienda, Cultura y ocio	Socialwelfare, ruralenvironment, publicsector, employment, demography, urbanism, infrastructure, tourism, education, health, economy, transport, environment, finance, culture, leisure
Open Data Navarra	Gobierno de Navarra	Administración electrónica, Administración pública, Ámbito local, Asuntos sociales, Deporte, Desarrollo rural, Economía y finanzas, Educación, Energía, Estadística, Formación, Industria, Justicia, Juventud, Medio ambiente, Salud, Territorio y urbanismo, Trabajo y Empleo, Tráfico, Transporte, Turismo, ocio y cultura, Vivienda	eAdministration, publicadministration, localgovernment, socialmatters, sport, ruraldevelopment, economy, finance, education, energy, statistics, training, industry, justice, youth, environment, health, territory, urbanplanning, jobs, employment, traffic, transport, tourism, leisure, culture, housing

Portal Open Data Xunta de Galicia	Xunta de Galicia	Información medioambiental, información geográfica, información turística, información cultural, deportiva y de ocio, información sobre transporte, información territorial y de vivienda, información administrativa y legal, información socio-sanitaria, información económica, empresarial y de empleo, información científico-tecnológica	Environmentalinformation, geographicinformation, touristinformation, culturalinformation, sports, leisure, transportinformation, landinformation, housinginformation, administrativeinformation, legalinformation, sociohealthinformation, economicinformation, business, employment, scientificinformation, technologicalinformation

Appendix A2: Inter-catalog agreement for the 10 German catalogs surveyed

	OffeneDaten.de	GovData	Open Data Berlin	Open Data Köln	Open Data HRO	Open Data München	Open Data ULM	OGDPR	Open NRW	TH
OffeneDaten.de	1	0.9	0.38	0.72	0.93	0.77	0.93	0.90	0.9	0.65
GovData	0.9	1	0.39	0.74	0.97	0.79	0.84	0.93	0.93	0.67
Open Data Berlin	0.38	0.39	1	0.31	0.38	0.33	0.35	0.37	0.36	0.28
Open Data Köln	0.72	0.74	0.31	1	0.77	0.67	0.67	0.75	0.74	0.55
Open Data HRO	0.93	0.97	0.38	0.77	1	0.83	0.87	0.97	0.97	0.7
Open Data München	0.77	0.79	0.33	0.67	0.83	1	0.71	0.8	0.79	0.53
Open Data ULM	0.93	0.84	0.35	0.67	0.87	0.71	1	0.84	0.84	0.65
Open Data Rheinland-Pfalz (OGDPR)	0.90	0.93	0.37	0.75	0.97	0.8	0.84	1	0.93	0.68
Open NRW	0.9	0.93	0.36	0.74	0.97	0.79	0.84	0.93	1	0.67
Transparenzportal Hamburg (TH)	0.65	0.67	0.28	0.55	0.7	0.53	0.65	0.68	0.67	1

Mean (all)[a]: 0.70; Standard deviation: 0.21; Min: 0.28; Max: 0.97; Mode: 0.67
Mean (city): 0.57; Standard deviation: 0.19; Min: 0.28; Max: 0.87; Mode: 0.67
Mean (region)/Standard deviation/Min/Max/Mode: N/A

Mean (country)/Standard deviation/Min/Max/Mode: N/A

[a]The values in the table above are rounded to the second decimal place to ease readability, but the values for the descriptive statistics were computed based on non-rounded values of the jaccard indices.

Appendix C1: 10 French Open Data Catalogs with their respective data categories

Catalog name	Publisher	Data Categories (French)	Data Categories (English)
data.gouv.fr	Etalab	Agriculture et alimentation, culture, économie et emploi, éducation et recherche, international et Europe, Logement, développement durable et énergie, santé et social, société, territoires, transports, tourisme	Agriculture, food, culture, economy, jobs, education, science, international, Europe, housing, sustainabledevelopment, energy, health, socialmatters, society, territories, transport, tourism
Data Grand-Lyon	Métropole de Lyon	Transport, imagerie, citoyenneté, services, culture, localisation, limites administratives, économie, environnement, occupation du sol, urbanisme, équipements, accessibilité, démographie	Transport, imaging, civilrights, services, culture, localization, administrative-borders, economy, environment, landuse, urbanplanning, facilities, accessibility, demography
Montpellier Territoire Numérique	Ville de Montpellier	Environnement, patrimoine/tourisme, économie, urbanisme, arts & culture, numérique, équipements, localisation, santé, politique publique & démocratie, démographie, transport, éducation, vie associative, sports & loisirs, proximité, habitat & aménagement	Environment, heritage, tourism, economy, urbanplanning, art, culture, digital, facilities, localization, health, publicpolicy, democracy, demography, transport, education, communitylife, sports, leisure, proximity, accommodation, planning
Nantes Ouverture des Données	Ville de Nantes	Citoyenneté/Institution, mobilité, santé/social, culture/tourisme, territoires, éducation/formation, environnement, économie, urbanisme, logement, jeunesse	Civilrights, institutions, mobility, health, socialmatters, culture, tourism, territories, education, training, environment, economy, urbanplanning, housing, youth
Open Data Nice Côte d'Azur	Métropole Nice Côte d'Azur	Accessibilité, administration électronique, aménagement du territoire, citoyenneté, culture, économie, éducation, environnement, événementiel, loisirs, santé, sécurité, sport, tourisme, transport	Accessibility, eGovernment, landsettlement, civilrights, culture, economy, education, environment, events, leisure, health, security, sport, tourism, transport
Open Data Bordeaux	Mairie de Bordeaux	Cadre de vie, citoyenneté et administration, culture, sports et loisirs	Livingplace, civilrights, administration, culture, sports, leisure

Open PACA	Région Provence-Alpes-Côte d'Azur	Administration-marchés publics, agriculture, aménagement du territoire, citoyenneté-démocratie, culture-patrimoine, économie-emploi, éducation-recherche, environnement-énergie, équipement collectif, finances, fonds institutionnels, formation-apprentissage, information-TIC, international-Europe-Bassin méditerranéen, Marseille-Provence 2013, Mer-Littoral, réseau de distribution, santé-social-sport, secteur public, tourisme, transports, urbanisme	Administration, public-contracts, agriculture, landsettlement, civilrights, democracy, culture, heritage, economy, jobs, education, science, environment, energy, publicfacilities, finances, institutional-funds, training, apprenticeship, information, TIC, international, Europe, Mediterraneanbasin, MarseilleProvence2013, sea, coastline, distributionnetwork, health, socialmatters, sport, publicsector, tourism, transport, urbanplanning
ParisData	Mairie de Paris	Services, déplacements, urbanisme, citoyens, culture, environnement, administration, finances, commerces	Services, trips, urbanplanning, citizens, culture, environment, administration, finances, trade
Rennes métropole en accès libre	Service Innovation Numérique, Hôtel de Rennes Métropole	Accessibilité, citoyenneté, culture, culture:agenda, culture:annuaire, culture:statistiques, données budgétaires, environment, equipements, logement, référentiel géographique, sports et loisirs, stationnement, transports	Accessibility, civilrights, culture, cultureagenda, culturedirectory, culturestatistics, budget, geographicreferenceframe, facilities, housing, sport, leisure, parking, transport
Toulouse Métropole Data	Mairie de Toulouse	Citoyenneté, culture, transport, finance, statistiques, sport, aménagement du territoire, urbanisme, bâtiments, equipements, logement, environnement, enfance, patrimoine, services, aménagement, tourisme, economie	Civilrights, culture, transport, finance, statistics, sport, landsettlement, urbanplanning, construction, facilities, housing, environment, childhood, heritage, services, planning, tourism, economy

Appendix B2: Inter-catalog agreement for the 10 Spanish catalogs surveyed

	D.A. JCYL	D.A. Andalucia	D.A. Madrid	Open Data Valencia	Open Data Aragon	Open Data BCN	Open Data Euskadi	Open Data Canarias	Open Data Navarra	Open Data Galicia
D.A. JCYL	1	1	1	0.47	1	0.03	0.27	0.5	0.36	0.05
D.A. Andalucia	1	1	1	0.47	1	0.03	0.27	0.5	0.36	0.05
D.A. Madrid	1	1	1	0.47	1	0.03	0.27	0.5	0.36	0.05
Open Data Valencia	0.47	0.47	0.47	1	0.47	0.04	0.3	0.48	0.29	0
Open Data Aragon	1	1	1	0.47	1	0.03	0.27	0.5	0.36	0.05
Open Data BCN	0.03	0.03	0.03	0.04	0.03	1	0.04	0.27	0.5	0.36
Open Data Euskadi	0.27	0.27	0.27	0.3	0.27	0.04	1	0.28	0.5	0.06
Open Data Canarias	0.5	0.5	0.5	0.48	0.5	0.27	0.28	1	0.30	0.06
Open Data Navarra	0.36	0.36	0.36	0.29	0.36	0.5	0.5	0.30	1	0.04
Open Data Galicia	0.05	0.05	0.05	0	0.05	0.36	0.06	0.06	0.04	1

Mean (all)[a]: 0.34; Standard deviation: 0.31; Min: 0; Max: 1; Mode: 1
Mean (city): 0.17; Standard deviation: 0.20; Min: 0.03; Max: 0.47; Mode: 0.03
Mean (region): 0.37; Standard deviation: 0.30; Min: 0.05; Max: 1; Mode: 0.05

Mean (country)/Standard deviation/Min/Max/Mode: N/A

[a] The values in the table above are rounded to the second decimal place to ease readability, but the values for the descriptive statistics were computed based on non-rounded values of the jaccard indices.

Appendix D1: 10 Open Data Catalogs from the UK with their respective data categories

Catalog name	Publisher	Data Categories (Catalogue)	Data Categories (API)
Birmingham DataFactory	Birmingham City Council	Travel and transport, council business, your local area, locations, environment, education	Travel, transport, councilbusiness, yourlocalarea, locations, environment, education
Bournemouth Data Stream	Bournemouth Borough Council	Tourism and population, traffic and geography, amenities, services and buildings, health and hygiene, finance	Tourism, population, traffic, geography, amenities, services, buildings, health, hygiene, finance
Data.gov.uk	UK Government	Environment, towns & cities, mapping, government, society, health, government spending, education, business & economy, transport	Environment, towns, cities, mapping, government, society, health, governmentspending, education, business, economy, transport
Data-Liverpool City Council	Liverpool City Council	Economy, population, education and skills, health, deprivation, labour market, housing, crime	Economy, population, education, skills, health, deprivation, labourmarket, housing, crime
Edinburgh Open Data Portal	City Council Edinburgh	Environment, health, education, transport, tourism, leisure, community	Environment, health, education, transport, tourism, leisure, community
Leeds Data Mill	Leeds City Council	Local services, transport, education, housing, health, business and economy, art and culture, geospatial, licenses, tourism, sport, transparency	Localservices, transport, education, housing, health, business, economy, art, culture, geospatial, licenses, tourism, sport, transparency
London Datastore	Greater London Authority	Demographics, employment and skills, transparency, environment, housing, health, transport, business and economy, education, planning, crime and community safety, young people, sport, art and culture, championing london, london 2012	Demographics, employment, skills, transparency, environment, housing, health, transport, business, economy, education, planning, crime, communitysafety, youngpeople, sport, art, culture, championinglondon, london2012
Open Data Bristol	Bristol City Council	Community, education, energy, environment, finance, government, health, internet of things, land use, mobility, reference, safety	Community, education, energy, environment, finance, government, health, internetofthings, landuse, mobility, reference, safety
OpenDataNI	The Open Data Team	Property & land, population & society, transport, health, finance, environment & agriculture, economy, industry & employment, tourism, leisure, culture & arts, education	Property, land, population, society, transport, health, finance, environment, agriculture, economy, industry, employment, tourism, leisure, culture, arts, education

Sheffield City Council Open Data	Sheffield City Council	Economy, education, environment, governance, health, heritage, housing, population, transport	Economy, education, environment, governance, health, heritage, housing, population, transport

Appendix C2: Inter-catalog agreement for the 10 French catalogs surveyed

	data.gouv.fr	Data Grand-Lyon	Montpellier T.N.	Nantes O.D.D.	Open Data Nice	Open Data Bor-deaux	Open PACA	ParisData	Rennes M.E.A.L	Toulouse M.D.
data.gouv.fr	1	0.10	0.18	0.32	0.22	0.04	0.32	0.04	0.10	0.16
Data GrandLyon	0.10	1	0.29	0.21	0.26	0.11	0.14	0.21	0.22	0.33
Montpellier T.N.	0.18	0.29	1	0.23	0.28	0.12	0.21	0.10	0.13	0.29
Nantes O.D.D.	0.32	0.21	0.23	1	0.30	0.10	0.25	0.14	0.12	0.27
Open Data Nice	0.22	0.26	0.28	0.30	1	0.17	0.25	0.09	0.26	0.32
Open Data Bor-deaux	0.04	0.11	0.12	0.10	0.17	1	0.08	0.15	0.17	0.09
Open PACA	0.32	0.14	0.21	0.25	0.25	0.08	1	0.13	0.09	0.23
ParisData	0.04	0.21	0.10	0.14	0.09	0.15	0.13	1	0.05	0.17
Rennes M.E.A.L	0.10	0.22	0.13	0.12	0.26	0.17	0.09	0.05	1	0.23
Toulouse M.D.	0.16	0.33	0.29	0.27	0.32	0.09	0.23	0.17	0.23	1

Mean (all)[a]: 0.18; Standard deviation: 0.08; Min: 0.04; Max: 0.33; Mode: 0.09
Mean (city): 0.16; Standard deviation: 0.07; Min: 0.05; Max: 0.29; Mode: 0.05
Mean (region)/Standard deviation/Min/Max/Mode: N/A
Mean (country)/Standard deviation/Min/Max/Mode: N/A

[a] The values in the table above are rounded to the second decimal place to ease readability, but the values for the descriptive statistics were computed based on non-rounded values of the jaccard indices.

Appendix D2: Inter-catalog agreement for the 10 catalogs from the UK surveyed

	Birmingham D.F	Bournemouth D.S.	Data.gov.uk	Data Liverpool	Edinburgh O.D.P.	Leeds Data Mill	London Datastore	Open Data Bristol	OpenDataNI	Sheffield Open Data
Birmingham D.F	1	0	0.19	0.07	0.27	0.11	0.13	0.12	0.14	0.23
Bournemouth D.S.	0	1	0.05	0.12	0.13	0.09	0.03	0.1	0.17	0.12
Data.gov.uk	0.19	0.05	1	0.17	0.27	0.24	0.23	0.2	0.26	0.31
Data Liverpool	0.07	0.12	0.17	1	0.14	0.21	0.26	0.10	0.18	0.38
Edinburgh O.D.P.	0.27	0.13	0.27	0.14	1	0.24	0.17	0.27	0.33	0.33
Leeds Data Mill	0.11	0.09	0.24	0.21	0.24	1	0.42	0.08	0.24	0.28
London Datastore	0.13	0.03	0.23	0.26	0.17	0.42	1	0.10	0.23	0.26
Open Data Bristol	0.12	0.1	0.2	0.10	0.27	0.08	0.10	1	0.16	0.17
OpenDataNI	0.14	0.17	0.26	0.18	0.33	0.24	0.23	0.16	1	0.3
Sheffield Open Data	0.23	0.12	0.31	0.38	0.33	0.28	0.26	0.17	0.3	1

Mean (all)[a]: 0.19; Standard deviation: 0.09; Min: 0; Max: 0.42; Mode: 0.12
Mean (city): 0.17; Standard deviation: 0.09; Min: 0; Max: 0.38; Mode: 0.12
Mean (region): 0.23; Standard deviation: 0; Min/Max/Mode: N/A
Mean (country)/Standard deviation/Min/Max/Mode: N/A

[a]The values in the table above are rounded to the second decimal place to ease readability, but the values for the descriptive statistics were computed based on non-rounded values of the jaccard indices.

5

Increasing Transparency through Web Maps

This chapter was published in the Proceedings of the 4th International Smart City Workshop (AW4city 2018) as *Degbelo, A. and Kauppinen, T. (2018) 'Increasing transparency through web maps', in Champin, P.-A., Gandon, F. L., Lalmas, M., and Ipeirotis, P. G. (eds) Companion of Proceedings of the Web Conference 2018 - WWW '18. Lyon, France: ACM Press, pp. 899–904. doi: 10.1145/3184558.3191515.*

Abstract. *Recent years have witnessed progress of public institutions in making their datasets available online, free of charge, for re-use. This notwithstanding, there is still a long way to go to put the power of data in the hands of citizens. This article suggests that transparency in the context of open government can be increased through web maps featuring: i) Application Programming Interfaces (APIs) which support app and data usage tracking; and (ii) 'transparency badges' which inform the users about the presence/absence of extra, useful contextual information. Eight examples of web maps are introduced as proof of concept for the idea. Designing and implementing these web maps has reminded of the need of interactive guidelines to help non-experts select vocabularies, and datasets to link to. The ideas presented are relevant to making existing open data more user friendly (and ultimately more usable).*

5.1 Introduction

The topic of smart cities has attracted growing interest from research, industry and local governments. Many definitions exist (for a review, see Yin et al., 2015), reflecting the plurality of perspectives in the context. Within this article, smart city is defined after Yin et al., 2015 as "a systematic integration of technological infrastructures that relies on advanced data processing, with the goals of making city governance more efficient, citizens happier, businesses more prosperous and the environment more sustainable". Citizen participation (*i.e.,* getting citizens to timely voice their opinions and wishes) is a key aspect of making city governance more efficient and citizens happier. Indeed, as Milakovich, 2010 noted, "Citizen participation provides a source of special insight, information, knowledge, and experience, which contributes to the soundness of government solutions to public problems". Improved citizen participation, in turn, requires greater transparency as citizens must know (or be made known) what is happening in their city and how they can best contribute to it, in order to effectively participate. There are several dimensions of transparency discussed in (Johannessen and Berntzen, 2018), but in this work we focus on what Johannessen and Berntzen called benchmarking transparency, *i.e.,* the availability of open data (*e.g.,* results from user surveys, demographic information), which citizens and interested parties can use to get a better idea of what is happening within government entities.

Despite a greater availability of open datasets, there is, as the second edition of the Open Data Barometer pointed out "still a long way to go to put the power of data in the hands of citizens" (`http://opendatabarometer.org/2ndEdition/`, last accessed: January 31, 2018). Visualising or geovisualizing open data seems the next logical step to put open data in the hands of citizens. Brunetti, Auer, and García, 2012 formalised the whole process of getting from a raw dataset to a visualisation as a framework called the Linked Data Visualisation Model (LDVM). LODVisualization (Brunetti, Auer, and García, 2012) and LinkedPipes Visualisation (Klímek et al., 2016) are two examples of tools which support LDVM. The current work differs from these two in mainly two ways: (i) a deliberate focus on geographic data preparation, visualisation and interaction (while the two works aforementioned take a more generic approach towards visualisation of open data on the web); and (ii) an account for the transformation from non-RDF data sources to RDF (which the two other tools did not intend to address). The main contributions of this paper are twofold.

First, we present a set of web maps to enable greater transparency in society. These web maps are part the Open City Toolkit, described in (Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016) as a platform to "to deliver services based on open data that are useful for citizens, businesses and governing bodies alike". As such, they are *one way of realising* Dadzie and Pietriga's aspiration (regarding work on Linked Data visualisation) expressed as follows: *We look forward to a* ***growing library of shared knowledge*** *and* ***visualisation-driven tools that break down technological barriers****, promoting instead richer exploration and intuitive, insightful analysis of users' personal context, myriad, shared situations and complex problems captured in Linked Data, and enable end users to draw confident conclusions about data and situations and add value to their everyday, knowledge driven tasks* (Dadzie and Pietriga, 2017, emphasis added). All the web maps focus primarily on the presentation of the information, essentially hiding the technicalities of Linked Data (*e.g.,* RDF Syntax) to the users. Second, the paper provides a critical analysis of these tools using concepts from cartographic representation and interaction. The analysis produces intrinsic, descriptive knowledge about the web maps (with no claim of generalisability to all web maps), and ends with some lessons learned about representation and interaction with geographic data on the web. In addition, as Çöltekin et al., 2017 recently reminded, "we do not know enough about in which domains geovisualization can be (potentially) of use". The set of web maps provided are artefacts illustrating one possible application domain of geovisualisations, namely enabling greater transparency in the society.

5.2 Background

Kamaruddin and Md Noor, 2017 identified four components of citizen-centricity which are used as a starting point in this paper (*i.e.,* openness, transparency, responsiveness and participation). In line with Michener and Bersch, 2013, transparency is viewed here as having two dimensions: visibility and inferability. The visibility dimension refers to the extent to which information is complete and easily located; the inferability dimension points to the degree to which information can be used to draw accurate conclusions. Conceptually, a map can be viewed as a *geometric structure* (Peuquet, 1988), a *graphical image* (Peuquet, 1988) or a *set of statements made by an author at a point in time* (Degbelo, 2017). Taking the viewpoint of *maps as statements* as a starting point, web maps are helpful to enable greater transparency in that they can make *value* more visible and inferable. Value of what? Of activities, processes and products pertaining to the public sphere. Why value? Because getting and keeping citizens interested in the participating in public decisions relies upon an appropriate communication of the value of their participation. *Value*, as used here, is in line with Benington's definition of 'public value', and encompasses "ecological, political, social, and cultural dimensions of value" (or simply said, all that adds value to the public sphere). The remainder of the article will not discuss all possible (and numerous) dimensions of values in the context of public sphere. Instead, it focuses on web maps which enable greater transparency by making the value of open (government) data more visible and inferable. Value of open data has many dimensions (*i.e.,* technical, economical, social, cultural, and political) which were discussed in (Attard, Orlandi, and Auer, 2016). The value creation assessment framework of (Attard, Orlandi, and Auer, 2016) lists no less than 19 (mostly technical) aspects which should be considered when evaluating the potential of an open government data initiative to enable value creation. Three of these 19 aspects were addressed in the current work:

- *Data usability*: datasets in formats such as Comma Separated Values (CSV), Portable Document Format (PDF) or Resource Description Framework (RDF) are not necessarily citizen-friendly. Visualising them is a way of adding to their value;

- *Background context*: linking datasets to related datasets (or simply making more specific their semantics through a conversion into RDF) does add value to existing datasets;
- *Rate of reuse*: providing information about the re-use rate of some datasets is a way of unveiling their actual *social value*.

The next section presents a set of web maps adding value to existing open datasets by realising these three aspects.

5.3 Research approach

This work follows a two-step approach. In the first step, a set of web maps built to enable greater transparency in society are presented, along with the technical features needed to implement them. In the second step, these web maps are critically analysed to bring forth visual variables and interaction primitives relevant for maps enabling greater transparency.

5.3.1 Generating the web maps

As mentioned above, the main purpose of the web maps is to enable greater transparency in society. As discussed in (Degbelo, 2017), two techniques are particularly suitable for this goal, namely Linked Data and visualisation. Linked Data increases transparency for machines, and visualisations do so for humans. To increase transparency, 36 students (divided into groups of three to six members) were asked to take existing open data, transform it into linked open data, and geovisualise it. The students were part of two classes organized in a blended learning fashion at two consequent years (one class took place with 19 people in the Winter term 2015/2017, and the second took place with 17 people in the Winter term 2016/2017). In the first class, open data from Münster was used as raw data; in the second class, participants were asked to work with open data of their choice. They were all non-familiar with Linked Data, and had various degrees of familiarity with web technologies (like HTML5, CSS, JavaScript or Node.js). The apps based on existing open data, and built as part of the practical work within the classes are: **Crime Mapper (A1)**: a web app for citizens & tourists to get a better overview of the crimes in Greater London; **Münster Households (A2)**: an interactive map for citizens & city councils to see households data from Münster between 2010 and 2014; **Münster Migration (A3)**:

an interactive map for citizens & city councils to go through migration statistics from Münster between 2010 and 2014; **Münster Population (A4)**: an interactive map for citizens & city councils to browse population data from Münster between 2010 and 2014; **Münster Social Insurance (A5)**: an interactive map for citizens & city councils to get an idea about the number of employees subject to social insurance contributions in Münster between 2010 and 2014; **Münster Unemployment (A6)**: an interactive map for citizens & city councils to explore unemployment data from Münster between 2010 and 2014; **Referendum Map Münster (A7)**: an interactive map for citizens & city councils to see results of the 2016 referendum regarding opening shops in the Münster city center; and **Wildlife Columbia (A8)**: a web app for policy makers & researchers to see information about protected natural reserves in Columbia, and species that inhabit these reserves.

Besides increasing data usability and providing background context about the datasets intrinsically, a novel feature of the web maps is the provision of information of the rate of open data usage. Technically, all web maps use the semantic API from (Degbelo, Trilles, Kray, et al., 2016) which enables app and dataset usage tracking, resulting in greater transparency. Degbelo, Trilles, Kray, et al., 2016 suggested that APIs which return data items according to their types - what they called semantic APIs - would lead to greater transparency (for developers) in an open government context, and identified recurrent categories of open datasets based on a survey of 40 European open data catalogues. Each of the web maps using the semantic API gets a 'transparency badge' (see Figure 5.1, bottom left corner), which indicates their support for dataset usage tracking. By clicking on this badge, the user is redirected to a dashboard-like platform which provides information about all applications available, the open datasets needed for their functioning, and their access rates of these datasets (see Figure 5.2). The information potential of users regarding what is happening with open datasets (*i.e.,* how these are used in one or many apps) is thereby increased. One can also visualise most demanded datasets using the 'Datasets' tab (see Figure 5.2). The transparency badge is mainly useful here to inform about rate of dataset usage. Yet, its conceptual scope should not be limited to this. One could envision further useful information provided to citizens after a click on a transparency badge. Example of relevant information in the context of open data visualisation include (the list is far from exhaustive):

- *source datasets of the visualisation*: according to the survey from Graves and Hendler, 2013, this is a most desired information by participants;
- *trustworthiness of the visualisation, and of the dataset*: as Tim Berners Lee recently reminded Tim Berners-Lee, 2017 "It's too easy for misinformation to spread on the web". The transparency badge could, for example, say whether the data (and/or its visualisation) has been verified by a public institution;

- *hints about data completeness*: participants from Beno et al., 2017 mentioned data incompleteness as one of the most severe barriers to open data adoption. Informing about data completeness may not solve the issue, but is already a way forward;

- *hints about data currency*: the lack of updates of published open data appears at the top of the list of participants from Benitez-Paez, Degbelo, et al., 2018 when it comes to major barriers to open data re-use. Here also, informing about data updating policies does not solve the issue, but can, at least, help citizens know what to expect;

- *licensing information about the dataset, and the visualisation*: this is mostly relevant to developers interested in re-use;

- *purpose of the data and the visualisation*: why the dataset has been collected, and why the visualisation has been created;

- *adoption examples*: how the dataset has been adopted elsewhere, and how it has been used in that (or these) scenario.

The final list of the transparency badge's informational items may be decided by its provider. This being said, experience from the food industry (where nutrition facts labels for packaged foods have proven simple and informative to consumers) suggests that standardisation of the informational items of a transparency badge (*e.g.,* through the W3C) could be helpful for the web as a whole at some point. The source codes of all web maps is available on GitHub (`https://github.com/geo-c`). Short demos can be accessed on Youtube (`https://goo.gl/73nxvv`). The apps were built using open source technologies. Examples of libraries used include Leaflet (open source map), Bootstrap (responsive web design), HighCharts, Charts.js, D3.js, C3.js and CanvasJS (histograms generation), Chroma.js (colour manipulation), and IntroJS (short intro to the main functionalities). Parliament and Virtuoso were used as triple stores. Vocabularies used while producing the RDF datasets include: geosparql (`http://www.opengis.net/ont/geosparql`), dc (`http://purl.org/dc/elements/1.1/`), dbpedia-ont (`http://dbpedia.org/ontology/`), geonames (`http://www.geonames.org/ontology/`), geo (`https://www.w3.org/2003/01/geo/`), time (`www.w3.org/2006/time#`) and datacube (`https://www.w3.org/TR/vocab-data-cube/`), to name but a few. Custom terms were created per application domain (*i.e.,* population, migration, referendum, and so forth) to meet their respective needs.

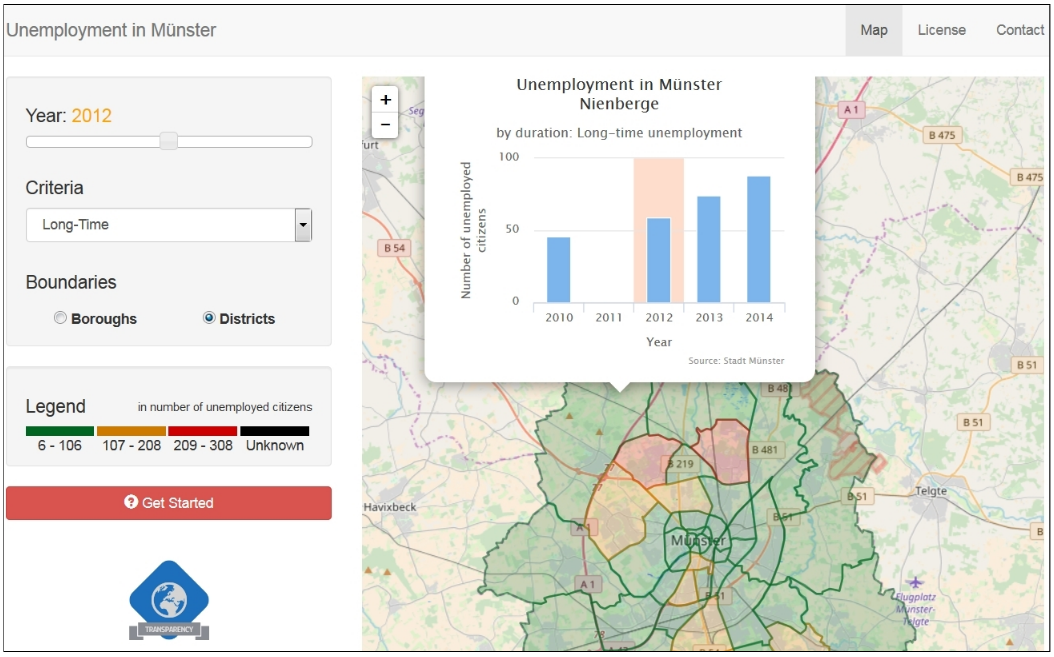
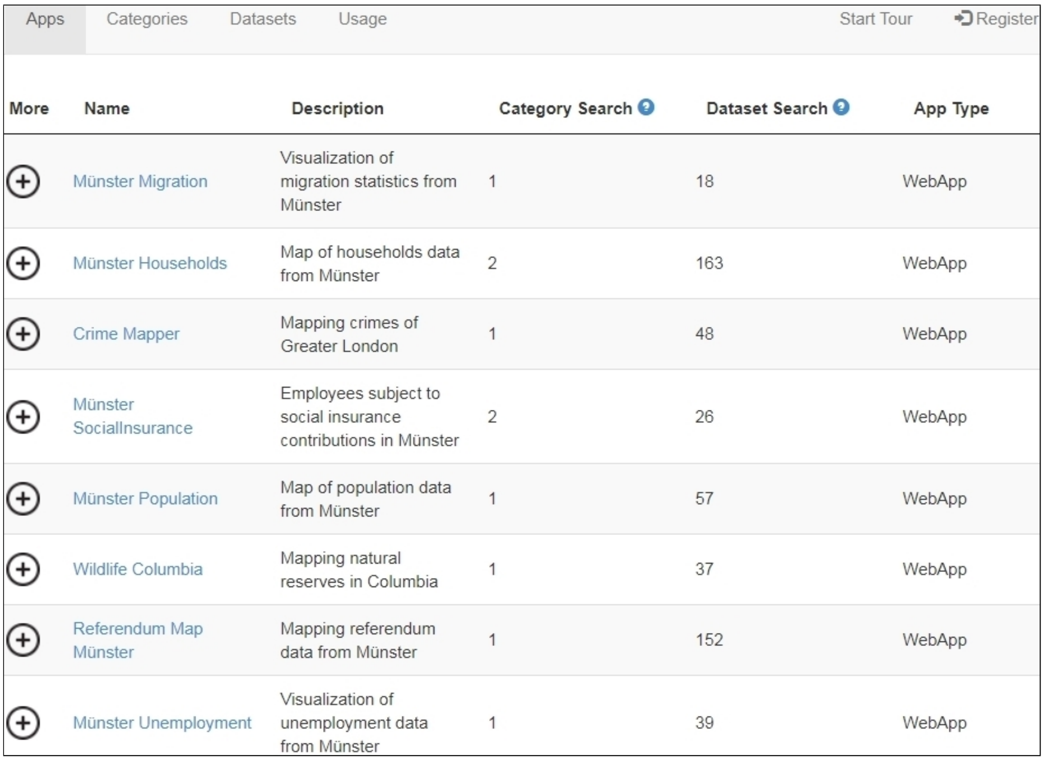


Fig. 5.1: Münster Unemployment - Application with visualises open data from Münster as a web map. The transparency badge signals greater transparency support (*i.e.*, the presence of extra, useful contextual information) for users of the visualisation.



More	Name	Description	Category Search	Dataset Search	App Type
⊕	Münster Migration	Visualization of migration statistics from Münster	1	18	WebApp
⊕	Münster Households	Map of households data from Münster	2	163	WebApp
⊕	Crime Mapper	Mapping crimes of Greater London	1	48	WebApp
⊕	Münster SocialInsurance	Employees subject to social insurance contributions in Münster	2	26	WebApp
⊕	Münster Population	Map of population data from Münster	1	57	WebApp
⊕	Wildlife Columbia	Mapping natural reserves in Columbia	1	37	WebApp
⊕	Referendum Map Münster	Mapping referendum data from Münster	1	152	WebApp
⊕	Münster Unemployment	Visualization of unemployment data from Münster	1	39	WebApp

Fig. 5.2: Dashboard-like visualisation of available applications as well as their access rates to existing applications.

5.3.2 Analysis

There are many dimensions along which the maps could be analysed. Since all maps are Linked Data visualisations, the dimensions presented in (Dadzie and Pietriga, 2017) are a possible choice. However, since little is known on how geovisualisations apply to different domains (see Section 10.1), the work instead adopted a framework for analysis which will help see how theoretical concepts from interactive maps have beenf applied to the task of enabling greater transparency. Of particular relevance are the concepts of 'visual variables' summarised in (Roth, 2017), and of 'interaction operators' from (Roth, 2013a).

Visual variables are one basic building block of a map or a visualisation. They describe "the graphic dimensions across which a map or other visualization can be varied to encode information" (Roth, 2017). Many visual variables were suggested over the course of the years, and Roth, 2017 synthesised them into a list of 12: location (*i.e.,* position of a map symbol with respect to a coordinate frame); size (*i.e.,* amount of space occupied by a map symbol); shape (*i.e.,* outline of the map symbol); orientation (*i.e.,* rotation of the map symbol from "normal"); colour hue (*i.e.,* dominant wavelength of the map symbol on the visible spectrum); colour value (*i.e.,* relative amount of energy emitted or reflected by the map symbol); texture (*i.e.,* coarseness of the fill pattern within the map symbol); colour saturation (*i.e.,* intensity of the colour of the map symbol); arrangement (*i.e.,* layout of graphic marks constituting a symbol); crispness (*i.e.,* sharpness of the boundary of the map symbol); resolution (*i.e.,* spatial precision at which the map symbol is displayed), and transparency (*i.e.,* amount of graphic blending between a map symbol and underlying map symbols).

Next to visual variables, interaction primitives are another basic building block of a map. Roth, 2013a brought forth an empirically-derived taxonomy of interaction primitives. According to this taxonomy, there are three primitive *interaction goals* (procure, predict, and prescribe), and five primitive *interaction objectives* (identify, compare, rank, associate, and delineate). In addition, the taxonomy comes up with a distinction between *enabling interaction operators* (import, export, save, edit, and annotate) and *work interaction operators* (reexpress, arrange, sequence, resymbolize, overlay, reproject, pan, zoom, filter, search, retrieve, and calculate). Work operators accomplish the desired objective, while enabling operators are useful to prepare for (or clean up) from work operators. Finally, the taxonomy lists three types of *interaction operands* related to the search target (space-alone, attributes-in-space, and space-in-time). There are two further interaction operands related to the search level (elementary and general). The reader is referred to (Roth, 2013a) for a full description of the taxonomy. The assessment of the web maps using this taxonomy is summarized in Table 5.1. The broad interaction goal enabled by all web maps

Tab. 5.1: Features of the web maps*

	NT	NV (CV)	Open Dataset Used	Visual Variables	Operands // Objectives Supported	Work Operators
A1	485,331	10 (1)	crime data	colour hue	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve, resymbolize, overlay, calculate
A2	31,162	21 (1)	households count data	colour hue, colour value	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve
A3	1,391	4 (1)	migration data	-	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve, resymbolize, overlay
A4	6,668	9 (1)	population data	colour hue, colour value	space-alone, attribute-in-space, space-in-time // identify, compare	pan, zoom, retrieve, resymbolize, overlay, calculate
A5	1,869	8 (1)	social insurance contribution data	colour value	space-alone, attribute-in-space, space-in-time // identify, compare	pan, zoom, retrieve, calculate
A6	4,399	6 (1)	unemployment data	colour hue	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve, overlay
A7	1,327	5 (1)	election data	colour hue	space-alone, attribute-in-space // identify	pan, zoom, retrieve, overlay
A8	2,432	16 (1)	endangered species data, deforestation data, social index data	colour hue	space-alone, attribute-in-space // identify	pan, zoom, retrieve, resymbolize, overlay, search

*Table Legend - NT: Number of triples; NV: Number of vocabularies used; CV: Custom vocabulary.

is *procure* (*i.e.,* enable retrieval of information about a geographic phenomenon represented, as opposed to make predictions about future states of the phenomenon at hand). Apart from A3 which proposes interaction to *export data*, none of the apps used enabling operators. The visual variable of *colour saturation* was left out of the analysis, because assessing it with the human eye is error-prone.

5.4 Discussion

This section briefly presents lessons learned from the building process, and subsequent analysis of the web maps.

Lessons learned on visual variables: of the 12 visual variables listed in Roth, 2017, colour is the only one which has recurrently been used across the various web maps. This reminds that effective colour selection will be key in enabling greater citizen-centricity on the web. ColorBrewer.org Harrower and Brewer, 2003 was a tool proposed in the early 2000s to help map makers choose effective colour schemes for thematic maps. Though it was tested for a variety of display types (*e.g.,* LCD, CRT) and widely used, the emergence of new display types (*e.g.,* AMOLED or Retina) suggests the need for new brewers which take into account advances in cartographic research, displays types and colour theory to assist developers in selecting most effective colour schemes while making their maps.

Lessons learned on interaction primitives: the question of how to best systematically document (Linked Data) visualisations in the context of the Semantic Web is important, but still open. In essence, this systematic documentation is important to synthesise gained knowledge across various visualisations (and use cases). As discussed in Section 10.3, the dimensions used in (Dadzie and Pietriga, 2017) (which were derived from general visualisation design guidelines and best practices) are an option. This work chose instead Roth's taxonomy of interaction primitives, which is specific to map interaction, and was derived from an empirical study. The taxonomy has proven quite usable while characterising the maps. This taxonomy may also be used in the future to stress interaction aspects of map visualisations on the web. The main lesson learned is that finding a definite answer to the question *how does objective restricts the space of possible map interaction operators on the Web?* might need a further specialisation of the objective primitives from Roth, 2013a as current primitive terms (*e.g.,* identify, compare) do not lead to a conclusive answer.

Lessons learned on Linked Data vocabularies: Table 5.1 shows that (perhaps unsurprisingly) all web maps used a custom vocabulary, in addition to existing ones. Though there is an increasing number of vocabularies indexed by LOV (Vandenbussche et al., 2016), having to define one's own terms to fully cover the use case at hand might remain the rule rather than the exception for some time. The process of designing and implementing the eight web maps (Section 10.3) has reminded that finding datasets other than Dbpedia to link to is still a challenge, and finding vocabularies to re-use remains challenging for non-experts. Interactive guidelines assisting them for the two tasks could help tackle these issues.

The way forward - Enhancing citizen centricity with web maps: as we have argued, citizens can benefit when web maps present diverse phenomena about their surroundings in a comprehensible way. It is possible to make use of a variety of open datasets, often linkable together, to create rich, visually communicated messages as web maps. However, the core value—*citizen centricity*—is enabled by transparency and openness of the web maps approach. When citizens can take a look at both web maps and their transparency badges, they are at the centre using and benefiting from information. Evidencing the source of used datasets and visualisations, metrics for their trustworthiness and completeness all contribute to creating value well beyond just having open data online. Further, seeing how many other people (including authorities) are interacting with information with the same web maps can support creating trust for information. A data-driven approach is not only to inform citizens but can also lead to "citizen-led urban innovations" (Annika Wolff et al., 2015) and create opportunities to react—with evidence—on issues emerging in local communities.

Limitations: though we argue in this paper for the use of web maps to enable greater transparency, it must be admitted that maps have their own learning curve. In fact, each communication medium has its own advantages and disadvantages. For instance, PDF files may be easier to generate from various sources such as text editors, but tables and other structured contents are challenging to parse; CSV or RDF files are machine processable, but without proper information visualization tools challenging to communicate to people. A systematic comparison of these different ways of making datasets available to the wider public (using *e.g.,* the evaluation model from L. Wang et al., 2005) could help better understand their respective merits in the context of open government.

5.5 Conclusion and outlook

In this paper we suggested that citizen-centricity of open data initiatives could be increased via web maps. This will make the value of activities, processes and products pertaining to the public sphere more visible. Open data is one of these products. We suggested that web maps featuring 'transparency badges' can be used to make their value more visible, thus increasing transparency. We presented eight example web maps to illustrate the idea, and documented lessons learned while designing and implementing them. An immediate research direction for future work is the understanding of citizens' wishes regarding the information to be provided by transparency badges. Shedding light on this can happen through a large-scale citizen survey, or via partnerships with city councils which have already made their data open. For instance, one could get statistics from city councils regarding actual users of these open datasets, select some of these users via purposive sampling, and interview them to understand what they actually need, and why they need it. Finally, it has become clear during the course of our work that a systematic evaluation of different communication mediums (*e.g.,* when, and for which citizen groups do PDF or web maps perform best regarding information provision?) would be useful to advance citizen-centricity of open data initiatives.

Acknowledgments: Auriol Degbelo gratefully acknowledges funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, `http://www.geo-c.eu/`). Tomi Kauppinen gratefully acknowledges Aalto Online Learning (A!OLE), a strategic initiative of the Aalto University, the project which he serves as a project leader. We thank Jonas Wissing for designing the logo of the transparency badge, and participants of the seminars "An Introduction to Javascript and Linked Data" (WiSe 2015/2016), and "Geoinformation in Society" (WiSe 2016/2017) at the Institute for Geoinformatics, University of Münster for implementing the web maps mentioned in this article.

6 Tell Me How my Open Data is Re-used: Increasing Transparency through the Open City Toolkit

This chapter was published as chapter in the book Open Data | Open Cities: Collaborative Cities in the Information Era as *Degbelo, A., Granell, C., Trilles, S., Bhattacharya, D. and Wissing, J. (2019) 'Tell me how my open data is re-used: increasing transparency through the Open City Toolkit', in Hawken, S., Han, H., and Pettit, C. (eds) Open Data | Open Cities: Collaborative Cities in the Information Era. Palgrave Macmillan.*

Abstract. *The open data movement has been gaining momentum in the recent years, with increasingly many public institutions making their data freely accessible. Despite much data being already open (and more to come), finding information about the actual usage of these open datasets is still a challenge. This chapter introduces two tools of the Open City Toolkit (OCT) which tackle this issue: a tool to increase transparency, and interactive guidelines. Interviews with city council employees confirmed the utility of the transparency tool. Both tools can be used by city councils (for planning purposes), and users interested to know more about the value of current open datasets (for information purposes).*

6.1 Introduction

The open data movement has been gaining momentum in the recent years, with increasingly many public institutions making their data freely accessible. Open data is improving government around the world, empowering citizens, creating new economic opportunities, and solving big public problems (see Young and Verhulst, 2016). As of May 2018, the Open Data Inception (opendatainception.io) [1] lists no less than 2,600 open data portals all around the world; the US Open Data Portal[2]

[1] `https://opendatainception.io/` (last accessed: May 19, 2018).
[2] `https://www.data.gov/` (last accessed: May 15, 2018).

lists about 190,000 datasets available; the European Union Open Data Portal[3] offers about 12,000 datasets; the data portal of the Australian Government[4] contains about 57,000 datasets; and the UK's open data portal provides about 45,000 datasets to browse through. These figures are indicative of the amplitude of the open data movement. The term 'open' may have different interpretations (for a recent review, see Pomerantz and Peek, 2016), but is used in this chapter to denote data "that anyone can freely access, use, modify, and share for any purpose"[5].

Open data has also attracted a significant amount of scholarly attention in recent years. A detailed presentation of open data ecosystems in Europe was done by S. Schade et al., 2015. Attard, Orlandi, Scerri, et al., 2015 provided a systematic survey of open (government) data initiatives with a detailed description of processes within the open government data lifecycle. Taking Chile as a case study, Gonzalez-Zapata and Heeks, 2015 identified two main types of stakeholders of open government data: primary stakeholders (i.e., politicians, public officials, public sector practitioners, international organizations) and secondary stakeholders (i.e., civil society activists, funding donors, ICT providers, academics). Susha et al., 2015 organized workshops with experts from the field of open government and open data to identify factors influencing the success or failure of open data initiatives. They provided a list of 47 success factors for open data publication and 18 success factors for open data use. Hartog et al., 2014 interviewed different types of stakeholders (e.g., civil servants, data source holders, and policy makers) to uncover the 'readiness' for open data of two governmental bodies: the municipality of The Hague, and the province of South-Holland. Citizens' motivations to participate was the subject of (Wijnhoven et al., 2015), where the authors found that strong belief that their suggestions will be applied correctly, perception of fun, and ideology (i.e., the person's attitude towards civic duties) are key factors of citizen engagement in open government projects. Additional work in the context of open government data has looked into open government portals' support for transparency and political accountability (Lourenço, 2015), openness and maturity indices for e-government (Veljković et al., 2014), a measurement framework to quantitatively assess the quality open government data (Vetrò et al., 2016), and visualization tools for open government data (Graves and Hendler, 2013), to name but a few.

Despite much attention of the scholarly community, many datasets being already open and more to come, finding information about the **actual usage** of these open datasets is still a challenge. Platforms such as CKAN offer a plugin (i.e., the *stats* extension[6]) to retrieve summary statistics about the most viewed datasets. This is

[3] `http://data.europa.eu/euodp/en/home` (last accessed: May 15, 2018).
[4] `http://data.gov.au/` (last accessed: May 15, 2018).
[5] `http://opendefinition.org/` (last accessed: May 15, 2018).
[6] `http://docs.ckan.org/en/ckan-2.7.3/maintaining/tracking.html` (last accessed: May 15, 2018).

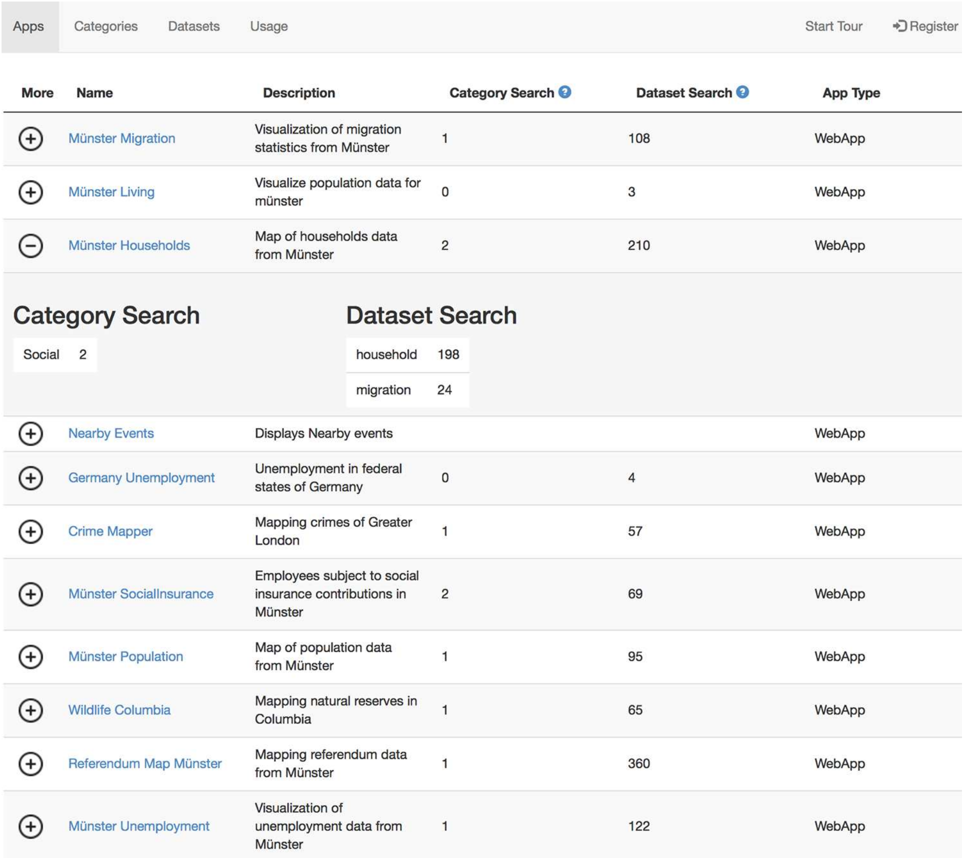
valuable information, but there is still a need for techniques, which enable re-use tracking beyond dataset views. Rate of re-use was mentioned in (Attard, Orlandi, and Auer, 2016) as one of the aspects of open data value creation not sufficiently addressed at the moment. Benitez-Paez, Degbelo, et al., 2018 found the lack of re-use examples to be one of the issues encountered by users while navigating open data portals. Having more information about the re-use of open datasets is critical to unveil their true value as: "[o]pen data on its own has little intrinsic value; the value is created by its use" (Janssen et al., 2012). Open data re-use information is also necessary for effective planning in the city context. For instance, it provides public institutions with a better idea of the types of datasets that are highly demanded (and by whom), and helps them prioritize the types of datasets to curate or regularly update.

This chapter introduces two software tools intended to advance the state of the art on open (government) data re-use: a tool to increase transparency, and interactive guidelines. The tools tackle the re-use problem at two levels: automatic re-use tracking (the former) and re-use documentation (the latter). Both tools are part of the Open City Toolkit (OCT), a collection of datasets, tools, services, specifications and guidelines to deliver services based on open data that are useful for citizens, businesses and governing bodies (Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016). The OCT combines technology-driven and citizen-centric strategies. It purports, as indicated in (Degbelo, Bhattacharya, et al., 2016), to address the lack of integrated and open collections of software components to realize smart cities.

6.2 OCT transparency tool

The OCT transparency tool is useful to answer the questions: what are datasets available in my city? How often are these datasets used? And which apps use these datasets? An essential technical means of realizing this is the use of semantic Application Programming Interfaces (APIs). The design of semantic APIs and their different layers were discussed in detail in (Degbelo, Trilles, Kray, et al., 2016). The main features of the OCT transparency module are:

- App registration: each developer (individual or organization) can register its app by getting an API key. This API key is used later to identify apps which access some datasets;
- Dataset registration: through this functionality, developers can register their own dataset to the OCT transparency module, so as to make it visible to other users (e.g., citizens, city councils, companies, developers);



More	Name	Description	Category Search	Dataset Search	App Type
⊕	Münster Migration	Visualization of migration statistics from Münster	1	108	WebApp
⊕	Münster Living	Visualize population data for münster	0	3	WebApp
⊖	Münster Households	Map of households data from Münster	2	210	WebApp
⊕	Nearby Events	Displays Nearby events			WebApp
⊕	Germany Unemployment	Unemployment in federal states of Germany	0	4	WebApp
⊕	Crime Mapper	Mapping crimes of Greater London	1	57	WebApp
⊕	Münster SocialInsurance	Employees subject to social insurance contributions in Münster	2	69	WebApp
⊕	Münster Population	Map of population data from Münster	1	95	WebApp
⊕	Wildlife Columbia	Mapping natural reserves in Columbia	1	65	WebApp
⊕	Referendum Map Münster	Mapping referendum data from Münster	1	360	WebApp
⊕	Münster Unemployment	Visualization of unemployment data from Münster	1	122	WebApp

Fig. 6.1: Dashboard visualization about datasets usage provided by the OCT transparency tool

- Logging: this functionality involves recording all activities related to an app (i.e., topics of datasets accessed, frequency of access, spatial locations from which the datasets are accessed).

As a proof of concept for the idea, eight web applications were created based on existing open government data (e.g., population, migration and referendum data). The applications and the process of their creation were presented in (Degbelo and Kauppinen, 2018). The datasets used are available on Zenodo (`https://doi.org/10.5281/zenodo.293201`). Figure 6.1 presents a dashboard visualization illustrating information about dataset usage provided by the OCT transparency tool. The tool also informs about the places from which an app has accessed datasets, and places from which datasets were called (see Figure 6.2). It is a dashboard in the sense of (Matheus et al., 2018) who define dashboards as "the visualization of a consolidated set data for a certain purpose, which enables to see what is happening and to initiate actions". The next subsections report on some tests about the usability, usefulness and scalability of the tool.

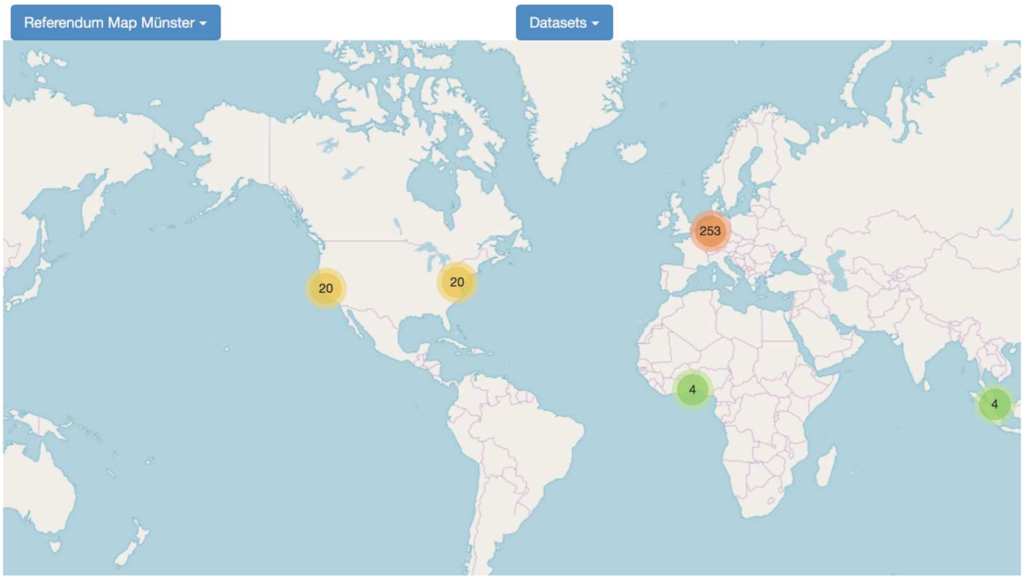


Fig. 6.2: Example visualization of spatial locations from which one specific app (i.e., Referendum Map Münster) is accessed

6.2.1 Usability

Two rounds of usability tests were conducted in February 2017 and October 2017. Each of the round involved seven people, leading to a total of 14 usability test participants. The usability tests were summative (see Lewis, 2014, for a definition of summative usability), focusing on efficiency and effectiveness. In the first round, students were asked to register an app and a dataset, and provide informal feedback about their experience doing so. Their feedback was integrated in the development of the second version of the transparency tool, which was used during the second round of tests. In this second round, participants were asked to register their app, register a dataset, and build their first OCT app. The three tasks were completed successfully by all seven participants in less than 30 minutes (see Figure 6.3). The SUS (System Usability Scale) score for the participants was 67.14, which means (following the scale introduced in Bangor et al., 2008; Bangor et al., 2009), that the participants rated the usability of the OCT transparency module as "good". Using SUS as usability questionnaire is suitable in this case, because previous work (Brooke, 2013; Sauro, 2013; Tullis and Stetson, 2004) pointed out that it produces acceptable results even with a small number of participants.

6.2.2 Utility

Eight semi-structured interviews of employees from two different city councils (Lisbon and Münster) were conducted in October 2017[7]. The purpose of the interviews

[7]The ideal number of participants for interviews is purpose-dependent (see e.g., Guest et al., 2006), but a common range is between 8-15 participants (Lopez and Whitehead, 2013). When doing

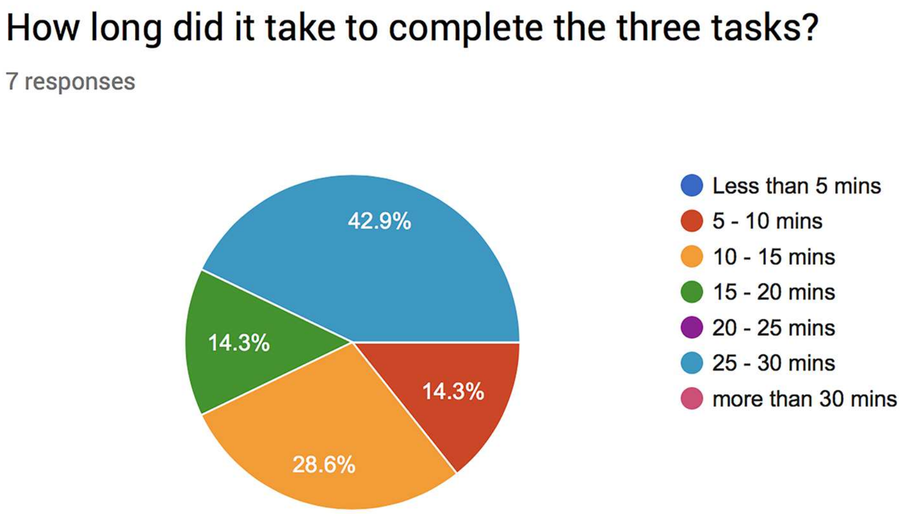


Fig. 6.3: Registering an app, a dataset, and building one's first OCT app can be done within 30 minutes

was to 1) gather insights from people working on (or having worked with open data) about the importance of open data re-use for city councils; and 2) collect some feedback from domain experts on the OCT transparency tool. Participants were recruited through snowball sampling (for a description of the sampling method, see Lazar et al., 2010). The OCT transparency tool was used as probe during the interviews.

The interview protocol was adapted from (Roth, 2009), and included an introductory question, five key questions, and an ending question. The first three key questions were: 1) How important is information about open data re-use for your institution? 2) What are you currently doing to collect information about open data re-use? 3) What issues do you face while collecting information about open data re-use? 4) In your opinion, what could be the benefits of the module for open data re-use? And 5) In your opinion, what could be the limitations of the module for open data re-use? Questions 4) and 5) were asked after showing an introductory video of 90 seconds about the tool.

The interviews lasted in average about 30 minutes. Table 6.1 reports on the results of questions 4) and 5), which are directly related to the transparency tool. The table illustrates that the participants, overall, saw more pros than cons. The pros often mentioned included: feedback to the city council about popularity of datasets, and an easier discoverability of datasets. Cons often reported included meta(data) maintenance (existing apps and datasets must be registered again on

qualitative research, "the 'richness' of data collected is far more important than the number of participants" (Lopez and Whitehead, 2013).

the tool to be made visible), as well as the current lack of quality checks by the tool. There is no guarantee that data saturation[8] has been reached with the sample of eight participants, that is, that the eight interviewees have listed all possible pros and cons pertaining to the tool. This notwithstanding, their feedback is useful: the pros mentioned validate the utility of the tool, while the cons point at areas where work is still needed in order to facilitate its adoption in the city context.

Tab. 6.1: Interviewees' feedback about the OCT transparency tool

Participant ID	Current Role	Advantages Mentioned	Limitations Mentioned
#1	Head of department	a) Data publishers can get some feedback about most popular datasets and categories b) Knowledge about most popular categories can inform about the types of datasets to make open	Data and metadata maintenance
#2	Project manager	a) Knowledge about datasets, which the city council does not need to publish b) Knowledge about most popular categories can inform about the types of datasets to make open c) Asking new questions (e.g., why someone access datasets from a place?)	Maintenance
#3	Team leader	a) Facilitate discoverability of datasets b) Show politicians that open data is the way to go	None
#4	Manager open data portal	See datasets and apps which are used	a) Module currently lacks information quality checks b) Module currently lacks verification of data
#5	Technical lead	Helps understand data use	a) No verification b) Coherency of the data
#6	Head of division	Easier discoverability of datasets	Module currently lacks notifications to users about crashes, and data additions
#7	Head of library	Easy to gather statistics about the data that is being used	No answer
#8	Geologist	Information about data usage	No idea

[8] See (Fusch and Ness, 2015) for a brief introduction to data saturation.

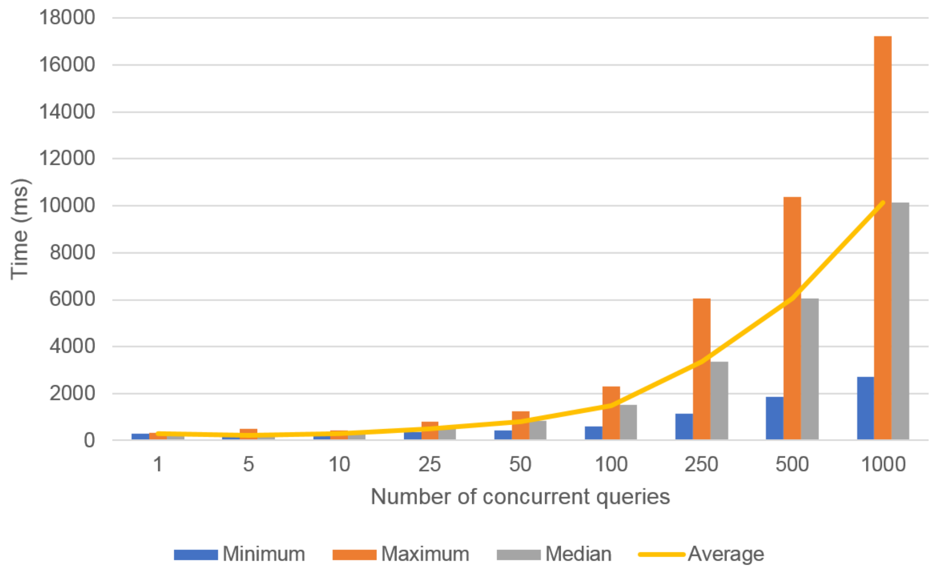


Fig. 6.4: The OCT Transparency Tool in reaction to growing instances of concurrent requests

6.2.3 Scalability

Several tests were conducted to assess the performance of the platform under a growing number of requests. The tests measured the response time of simultaneous database accesses on the system. Each test involved a group of queries to the endpoint of the API. Each group of query was executed five times, and the response time was averaged over the five executions. The tests simulated 1, 5, 10, 25, 50, 100, 250, 500 and 1000 concurrent uses respectively. The data packets retrieved was kept constant (7KB) during the whole test sessions. As Figure 4 suggests the scalability of the platform is better-than-linear. The code source of the application is available on GitHub (`https://github.com/geo-c/OCT-Core`).

6.3 OCT interactive guidelines

While the OCT transparency tool enables the monitoring of cities' open datasets usage, the OCT interactive guidelines tool deals with the following question: what can I do with all these datasets? The main audiences of both tools are different: decision makers, data publishers and managers are typically the focus of the OCT transparency tool. Of course, citizens can also seek for the impact of open datasets. Yet, they often measure the usefulness and impact of open datasets through a different lens, namely 'how can open data sets enhance their daily activities'? Both tools are complementary, addressing open data usage and open-data-based innovations (though the interpretation of these two terms may vary depending on the target stakeholders).

6.3.1 The need for guidelines

We can explain the role and purpose of the OCT interactive guidelines by borrowing an analogy from MIT researcher Cesar Hidalgo, who compared open data sites and supermarkets[9]: "Imagine shopping in a supermarket where every item is stored in boxes that look exactly the same. Some are filled with cereal, others with apples, and others with shampoo. Shopping would be an absolute nightmare!" Hidalgo argued that most, if not all, open data sites are organised like arrays of "brown boxes" in supermarkets, i.e., arrays of links to public datasets that quite often are published as they were collected. This way, most of these sites look like they are only addressing a small portion of the whole population: those with technical skills (programmers, researchers, etc.) or professionals (e.g., data-driven journalists, civic agents, etc.), i.e., those few specialists who are able to handle and transform datasets to tell stories to the rest of people. The OCT has a CKAN-based module (not introduced in this chapter, see Degbelo, Bhattacharya, et al., 2016) which is not that far off this strategy; research resources are registered and made publicly available as endpoints that can be queried via well-documented data access and retrieval APIs. The expected stakeholders of the CKAN-based OCT module are other scientists, researchers and civic hackers/programmers who feel comfortable (programmatically) handling open data and coding.

If we do not consider the tech elite, which is the remaining 95% of the population (Kankaraš, et al., 2016), open data sites become difficult to understand (see e.g., Benitez-Paez, Degbelo, et al., 2018; Beno et al., 2017). Returning to Hidalgo's analogy of the supermarket, imagine you (citizen) are asking for "cannelloni" in the food section and the clerk delivers you a bag with all the raw ingredients to cook them yourself. Like most of the open data sites, open data is delivered in the way in which it was collected. Next, you look again at the clerk and order cannelloni "ready to be eaten", because you do not have time or do not know to cook them. Like most open data sites, open data is not delivered in the way it can best be used and/or understood. Rather, open datasets are often delivered with no clue on how to process them, manage them, or, even worse, whether they can be useful for citizens at all. In sum, citizens demand "ready-to-consume, easy-to-understand products" rather than raw ingredients like open datasets. Sometimes these products take the form of apps, or can be expressed as interactive guidelines. The OCT interactive guidelines tool seeks to make city problems and subsequent actions understandable to citizens.

[9] What's Wrong with Open-Data Sites--and How We Can Fix Them, by Cesar Hidalgo. `https://blogs.scientificamerican.com/guest-blog/what-s-wrong-with-open-data-sites-and-how-we-can-fix-them/` (last accessed: Oct 4, 2017).

Most open data sites do not deliver elaborated stories that emerge from the combination of their contained open datasets. However, most people are looking for stories ("cannelloni") that can be easily comprehended ("eaten"). In case people want to know the details (e.g., raw ingredients to cook cannelloni themselves), they can directly download or access data sets through the corresponding data access API. What we pursue here is the design and creation of "stories" that bring together, behind the scenes various datasets and other types of resources and transform them into interactive city guidelines, which help a large portion of the society to understand their benefits and impacts regardless the complexity of the details.

The OCT interactive guidelines tool is intended to help city stakeholders walk through a story. On one hand, the term "guidelines" is seen as narratives that refer to problem-solution patterns by presenting challenges, benefits and impacts in an understandable manner, i.e., everyone may share and refer to when talking to others. Problems may be of diverse nature, such as social, mobility, environment, and cultural; solutions may involve a combination of datasets, code, apps, services and any other relevant resource that helps to sort out the current problem. On the other hand, the qualifier "interactive" underlines the ability of users to dynamically explore (to certain degree) the guideline through a set of blocks for different purposes such as graphs, plots creation, maps visualisation, custom JavaScript code, p5 code (a sort of JavaScript wrapper for processing), and the inclusion of text and markdown formats. We intentionally avoid static guidelines, as in the form of tutorials or paper-based posters, to let stakeholders engage dynamically with the content of the visual narratives.

6.3.2 Conceptual architecture

Figure 6.5 shows the conceptual architecture to materialise the OCT interactive guidelines tool. Designers of stories are one type of users. These could be for instance researchers, data journalists, or data publishers: they use "storytelling" formats for creating visual and interactive narratives of how smart city solutions are being installed and deployed in cities. That is, they design and tell stories based on external or own datasets and other research resources. The tool provides an edit mode to create and easily update each story and publish it into the catalogue of guidelines (Figure 6.6). Citizens are the second type users. They can pick a guideline from the catalogue and explore it through interactive elements at their disposal. For example, via interactive plots, charts and maps, and through on the fly annotations as a way to provide feedback about the story being visualised (this feature will be released shortly). The source code of the OCT interactive guidelines tool is also available on GitHub (`https://github.com/geo-c/OCT-Guidelines`).

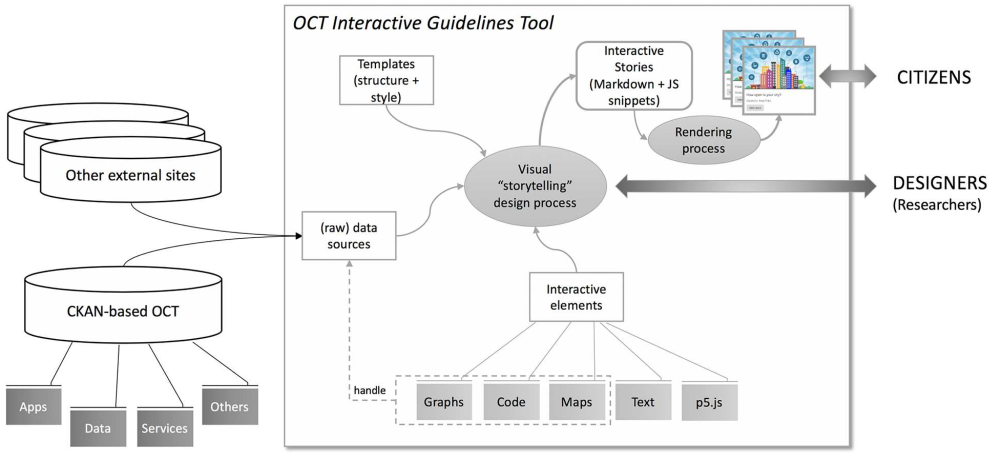


Fig. 6.5: Architecture of the OCT interactive guidelines tool

Technically, each guideline is stored as a markdown file (like a regular text file). Markdown tags specify the sections of a guideline, and keep information about the author, last update, title and a list of data sources used. To ease the process of creating guidelines, the tool provides a range of templates with predefined sections and style. Each guideline is exportable as a regular markdown file for offline edition, which may be uploaded again later on. Besides, a guideline contains a collection of interactive elements (codified as JavaScript snippets). Currently, the supported elements are graphs and plots, custom JavaScript code, maps and text; adding annotations is part of an ongoing work.

These interactive elements are able to handle data sources, both deployed in a CKAN-based instance platform and published elsewhere. For example, a graph can take as input a data source available in the OCT catalogue, permitting users to interact with the graph, and thereby with the associated data source. Furthermore, any resource registered in the OCT catalogue (`http://giv-oct.uni-muenster.de:5000/`) is potentially an input source for interactive guidelines by only specifying its access point (e.g., URL). Moreover, interactive guidelines can be registered in the OCT catalogue as any other public and open resource. This way, the OCT interactive guidelines tool augments the capabilities of the OCT catalogue, to deliver not only datasets, but stories to a wide range of stakeholders. On the down side, designing compelling, understandable, and thought-provoking guidelines requires authors with proven communication and design skills so that the intended messages are effectively transmitted to the public.

Examples of these interactive guidelines are available at `http://elcano.init.uji.es/guidelines`. At the moment of this writing, there are ten guidelines; some of them are like "tutorials" to guide user creators to use and combine the different available blocks. Others intent to help stakeholders to solve particular problems or

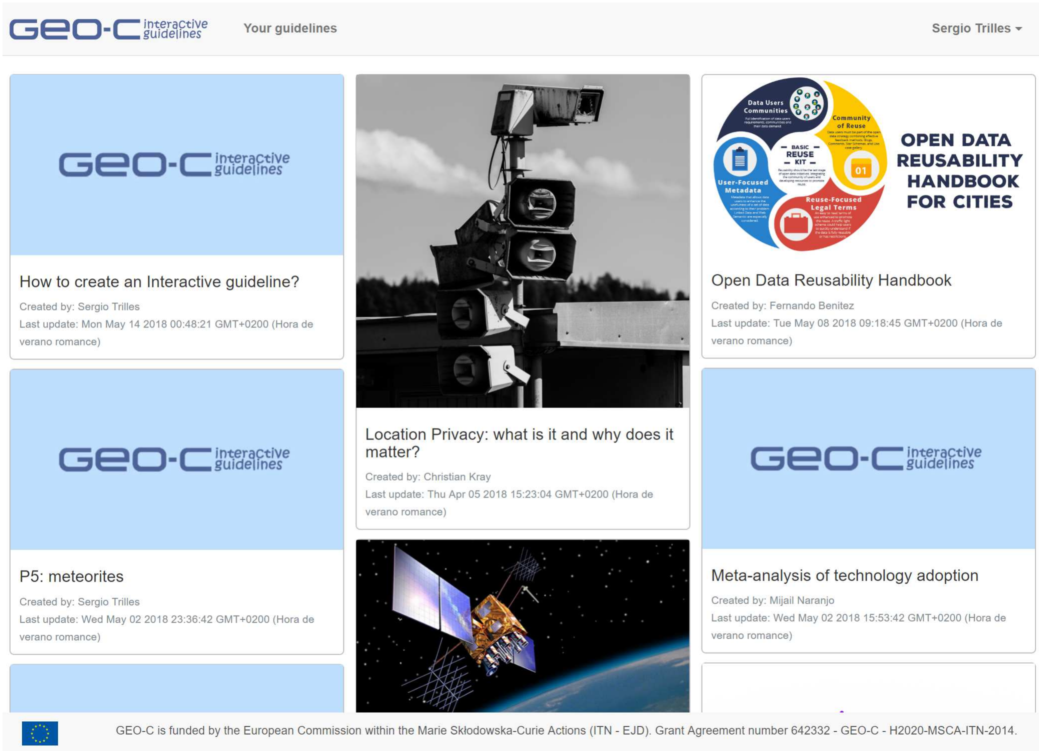


Fig. 6.6: Catalogue of interactive guidelines showing examples of OCT interactive guidelines

show a list of suggestions to follow. For instance, the guideline "Location Privacy: what is it and why does it matter?" attempts to communicate to citizens the importance and implications of sharing the location using smartphones. The "Checklist and tools when publishing open data" guideline tries to explain what an open data needs to earn some stars of Tim Berners Lee's five star model (Berners-Lee, 2006). To achieve that, some blocks such as text and p5.js blocks are used. The latter blocks are particularly useful to provide interactivity to the guidelines (e.g., they help to generate buttons, where users can click and see what each star category means).

In sum, the interactive guidelines do not specifically target technologically savvy people such as open data advocates and programmers. These guidelines aim to inform people about problems that matter in their cities, making them understandable, and presenting potential solutions. Interactive guidelines, when designed as effective narratives, can raise awareness about certain problems that matter to citizens, even to the point to persuade and reframe thinking. In this sense, interactive guidelines could have an educational footprint in the long run.

6.4 Conclusion

There is an increasing amount of open datasets available through open government portals, but still much work to be done to inform about the actual usage of these open datasets. This chapter has introduced two software components to enable progress on this issue: the Open City Toolkit transparency module, as well as interactive guidelines. The former aims at informing about the rate of open data re-use and the latter purports to communicate innovations ensuant on open data. The chapter has presented the key ideas behind the two components, and (evolving) prototypical implementations illustrating them. The work introduced is relevant to open data publishers and citizens at large.

Immediate directions for future work, based on the feedback from the participants, include further improving the usability of the tool, and devising means to automatically check the quality of the datasets. Developing metrics for (subjective) aspects of data quality such as 'fitness for purpose', 'trustworthiness' or 'understandability' is a challenge, but other aspects of quality such as 'dataset availability' or 'dataset currency' are easier to assess, and can be implemented. Automatically recommending (possibly) relevant datasets to new apps registering on the OCT transparency tool seems also a promising direction for future work.

Acknowledgements: Comments from three anonymous reviewers have helped improve the clarity of the article. The authors gratefully acknowledge funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, `http://www.geo-c.eu/`). Carlos Granell is funded by the Ramón y Cajal Programme (grant number RYC-2014-16913). Sergio Trilles is funded by the postdoctoral programme Vali+d (GVA) (grant number APOSTD/2016/058). We thank participants of the course "Geoinformation in Society" who volunteered to do the usability tests and employees from the city councils of Münster and Lisbon for their feedback. Finally, we thank Christian Kray and Marco Painho for the support provided during the course of this work.

7

Linked Data and Visualization: Two Sides of the Transparency Coin

This chapter was published in the Proceedings of the 3rd International Workshop on Smart Cities and Urban Analytics (UrbanGIS 2017) as *Degbelo, A. (2017) 'Linked data and visualization: two sides of the transparency coin', in Vo, H. T. and Howe, B. (eds) Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS'17. Los Angeles, California, USA: ACM Press, pp. 1–8. doi: 10.1145/3152178.3152191.*

Abstract. *Transparency is an important element of smart cities, and ongoing work is exploring the use of available open data to maximize it. This position paper argues that Linked Data and visualization play similar roles, for different agents, in this context. Linked Data increases transparency for machines, while visualization increases transparency for humans. The work also proposes a quantitative approach to the evaluation of visualization insights which rests on two premises: (i) visualizations could be modelled as a set of statements made by authors at some point in time, and (ii) statements made by experts could be used as ground truth while evaluating how much insights are effectively conveyed by visualizations on the Web. Drawing on the linked data rating scheme of Tim Berners-Lee, the paper proposes a five-stars rating scheme for visualizations on the Web. The ideas suggested are relevant to the development of techniques to automatically assess the transparency level of existing visualizations on the Web.*

7.1 Introduction

Smart cities are attracting growing interest from various stakeholders. As for research, Adegboyega Ojo et al., 2016 recently reported an increase of 200 % in publication volume for smart cities research since 2009. As for local governments, a survey conducted in the US in 2016 by the International City/County Management Association (in cooperation with the Smart Cities Council) revealed that more than

50% of the respondents identified smart cities activities as being of medium to high priority to them[1]. Several industry players (e.g., IBM[2], Microsoft[3], Siemens[4]) have their own smart city solutions, while various toolkits are currently on the market to make cities smarter (see Degbelo, Bhattacharya, et al., 2016). There are various possible ways of defining "smart cities". In this work, the term is used to denote "a system integration of technological infrastructure that relies on advanced data processing with the goals of making city governance more efficient, citizens happier, businesses more prosperous and the environment more sustainable" (Yin et al., 2015).

Citizen participation has been acknowledged in the literature as a key component of smart cities (e.g., Cucciniello et al., 2016; Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016; Johannessen and Berntzen, 2018). Inputs from citizens indeed have the potential of providing "information, knowledge, and experience, which contributes to the soundness of government solutions to public problems" (Milakovich, 2010). *Transparency* is an important component of citizen participation. For instance, Johannessen and Berntzen stressed that "A key aspect of participation is that of transparency" Johannessen and Berntzen, 2018; Kim and J. Lee, 2017 reported on some positive correlations between citizen engagement and transparency; and Attard, Orlandi, Scerri, et al., 2015 pointed out that transparency is one major motivation of open government data initiatives, and can help citizens establish a trusting relationship with the government.

Linked Data and visualization are two important technologies in the context of smart cities. As to the former, Dadzie and Pietriga, 2017 indicated that Linked Data provides a structured source of knowledge for both research and practical applications, and has been adopted as a practice in various open data initiatives. Mathieu D'Aquin et al., 2015 highlighted the relevance of Linked Data to deal with the diversity of smart city data. Regarding the latter, there is a growing amount of visualizations on the Web targeting city use cases (for examples of collections, see `https://goo.gl/PFGqua` and `https://goo.gl/QCHFZM`). This confirms Dykes et al., 2010's early hunch that visualizations have a key role to play as one strives to make sense of city data that are being collected. An early discussion of the potential role of Geographic Information Systems for smart cities pointed out that "Geo-visualization is the most essential instrument of representing geographic data and analysis results and exploring potential interesting findings" (Tao, 2013).

[1] See `https://goo.gl/SCYjUH` (last accessed: September 06, 2017) for the full report.
[2] `https://goo.gl/xe1niw` (last accessed: September 06, 2017).
[3] `https://goo.gl/45vL6w` (last accessed: September 06, 2017).
[4] `https://goo.gl/q86QrA` (last accessed: September 06, 2017).

This position paper argues that (as far as transparency and smart city are concerned), Linked Data and visualization play a similar role for different audiences. Linked Data, due to its structured format, facilitates open data consumption for machines; visualization facilitates open data consumption for humans. This implies that (as far as transparency and smart city are concerned), it is worthwhile to consider areas where research on Linked Data, and research on visualization can learn from one another. This is what the article will try to do. In particular, the paper will discuss how assessing the effectiveness of visualizations with respect to insight generation (an open issue at the moment, see e.g., Çöltekin et al., 2017; Isenberg et al., 2013; Roth, Çöltekin, et al., 2017; Jarke J. van Wijk, 2013) may benefit from work already done in Linked Data.

The arguments are exposed in six sections. Section 7.2 presents the framework for transparency considered, and the contributions of Linked Data to transparency. Section 7.3 discusses visualizations as enablers of transparency. Section 7.4 briefly touches upon the challenges of assessing the effectiveness of Linked Data and visualizations regarding transparency enablement. Section 7.5 brings forth an approach to quantitatively assess the insightfulness of visualizations, and elaborates on its advantages as well as drawbacks. Section 7.6 proposes a rating scheme for visualizations on the Web (as a first step towards making them more findable), and Section 10.7 concludes the article.

7.2 Linked Data and Transparency

Transparency is important in the smart city context, but there are various kinds of it. For instance, Johannessen and Berntzen, 2018 suggested six categories of transparency: document transparency (i.e., access to government documents), meeting transparency (i.e., access to meetings of public bodies including their agenda and minutes), process transparency (i.e., explanation of processes leading to government decisions including when and how citizens may have their say), benchmarking transparency (i.e., access to data about the performance of public institutions), decision-maker transparency (i.e., access to information about who the decision-makers are and what conflicting interests they may have) and disclosure transparency (i.e., the right to ask written or oral questions related to information not in documents or meeting agendas). Heald, 2006 mentioned some additional types of transparency including transparency onwards (i.e., the right of rulers to observe activities of subordinate agents), transparency downwards (i.e., the right of ruled to observe activities of their rulers), transparency outwards (i.e., an agent can observe what is going on outside the organization), transparency inwards (i.e., those outside can observe what is happening inside the organization), transparency in retrospect (i.e., an organization reports at periodic intervals about its activities) and transparency in real-time (i.e., continuous surveillance of the organization). Not present in the previous lists, but also relevant to the smart city context is algorithmic transparency (Brauneis and Goodman, 2018), i.e., the extent to which parameters of predictive algorithms shaping local government actions are made known to the public. Though all these aspects of transparency are important for cities, the remainder of this work focuses only on transparency in the specific context of open government data (which is akin to 'benchmarking transparency' mentioned above).

Michener and Bersch, 2013 presented transparency as a *continuum* and proposed two dimensions of transparency: visibility and inferability. Visibility means that the information is (i) reasonably complete and (ii) found with relative ease. Inferability refers to the degree to which the information at hand can be used to draw accurate conclusions. According to Michener and Bersch, 2013, properties of visibility are intrinsic to the information, whereas inferability is contingent on the receptive capacity of the target audience. Viewing transparency as a continuum implies that existing (legally or technically) open data might be associated with various levels of it (i.e., both visibility and inferability).

Transparency of open datasets can be increased through the *explicit linking* of these datasets to related datasets, as well as applications which consume them. Creating these explicit links boosts transparency in at least two ways. First, linking expands the completeness of information, and thereby its visibility. Second, creating explicit links between open datasets and applications which re-use them is a way of unveiling their use, and making their *value* more apparent. As Janssen et al., 2012 indicated "open data has no value in itself; it only becomes valuable when used".

Linked Data is machine-readable data, and therefore primarily suitable for consumption by machines. As a result, making typed links between resources explicit contributes to making the data more transparent for software agents. Since Linked Data is part of the Big Data landscape (see Hitzler and Krzysztof Janowicz, 2013), the arguments presented in this section hold also for Big Data in the context of smart cities. Some examples of the use of Linked Data to increase transparency are found in (Futia et al., 2017; M. Martin et al., 2014; Mora-Rodriguez et al., 2017). Futia et al., 2017 report on using Linked Data principles to reduce fragmentation (i.e., increase visibility) of information in Italian procurement data. M. Martin et al., 2014 present an RDF (Resource Description Framework) version of the Financial Transparency System data from the European Commission, and point out that Linked Data leads to an increased financial transparency of EU project funding. Mora-Rodriguez et al., 2017 used a combination of XBRL (Extensible Business Reporting Language) and Linked Data, to make corporate data more transparent.

7.3 Visualization and Transparency

Visualization plays also an important role in the context of open data re-use and smart city. Segel and Heer, 2010 discussed that visualization of data has a storytelling potential. In addition, data visualization "makes it possible for researchers, analysts, engineers, and the lay audience to obtain insight in these data in an efficient and effective way" (Jarke J van Wijk, 2005). Visualization can be considered from three different viewpoints presented in (Jarke J van Wijk, 2005): technology, art, or science. In the current article, visualization is viewed as a *technology to communicate a story*.

Scheider, Jones, et al., 2014 suggested to model the content of a map as the set of assertions which can be extracted by looking at it. That is, geovisualizations (and more generally visualizations) could be seen as a set of RDF statements made by authors with a certain reputation at some point in time (see Kuhn, Kauppinen, et al., 2014). The consequence of this view is that visualizations are also enablers of transparency. They make (some) statements visible to the consumer, namely those that the author of the visualization includes in her narrative. The number of statements that the consumer actually notices while using a visualization depends on many factors, including her own experience and the degree of interactivity provided. Visualizations stimulate visual thinking, and are therefore primarily suitable for human agents. To echo Shneiderman, 1996, "The attraction of visual displays ... is that they make use of the remarkable human perceptual ability for visual information". Therefore, making (selected) facts much more prominent via a visualization leads primarily to an increase in visibility and/or inferability for humans agents.

7.4 Assessing Effectiveness Regarding Transparency

From the previous two sections, Linked Data and Visualization contribute to increase transparency for machines and humans respectively. They are two sides of the same coin. The question now is how effective both are with respect to transparency. Why care? Because increasing transparency means making *more insights visible* and *inferable*. In particular, visualizations which increase transparency are key to effective information of users. Effective information of users, in turn, is necessary if visualizations are to be potent for problem-solving. As Griffin et al., 2017 pointed out: "Because maps are used to solve problems that underlie the sustainability of life on Earth (e.g., climate change, water resource allocation, declines in biodiversity, etc.), understanding how maps are *insightful* is more important than ever" [original emphasis].

Measuring the effectiveness of two linked datasets with respect to transparency is a matter of comparing how many statements (i.e., triples) each dataset makes visible with respect to a given phenomenon. A reasoner could also be used to check the respective performance of the two datasets with respect to inferability (i.e., how many additional statements they enable). Given a dataset D and a topic T, producing algorithms to answer the question *How many triples of D are about T?* is currently a challenge.

Measuring the effectiveness of two visualizations with respect to transparency faces also a challenge, but of a different kind. The main issue here is that of specifying what 'insight' is. As Jarke J. van Wijk, 2013 pointed out, insight is "ill-defined and hard to measure". Chang et al., 2009 suggested a distinction between spontaneous insight (i.e., a moment of enlightenment) and knowledge-building insights (i.e., an advance in knowledge or a piece of information). As Chang *et al.* indicated, "cognitive scientists have successfully identified the neural patterns of the spontaneous insight phenomenon and can now observe and measure the insight process". The ideas suggested in the following focus on measuring knowledge-building insights made visible to user by a visualization. Discussing the inferability aspect of these knowledge-building insights is left for future work.

7.5 Measuring knowledge-building insights of visualizations

Visualizations could be approached based on the following premises:

- P1: a visualization is a set of statements made by authors with a certain reputation at some point in time.
- P2: finding an absolute ground truth for a visualization may be unattainable, but statements made by experts can be used as ground truth for evaluation purposes.

P1 was already discussed in Section 7.3. Regarding P2, an expert denotes the creator of the visualization, or a user who has an accumulated knowledge about it through prolonged use. P2 adapts the idea well known in GIScience (see e.g., Dowman, 1999) to use the best available data as 'truth' against which the accuracy of other datasets can be assessed. The effectiveness of visualizations regarding the number of insights actually made visible can be assessed through the administration of questionnaires both pre- and post-interaction. Linked Data formulates statements as triples, and this happens to be the simplest form in which statements can be made in natural language (see Kuhn, Kauppinen, et al., 2014). As a result, one could take advantage of the simplicity of triples while formulating statements to assess the insighfulness of visualizations. An additional aspect of the triple syntax which makes them attractive in this context is their inherent structuredness. As J. Johnson, 2010 indicated, information is easier for people to scan and understand when presented in a terse and structured way. It's a research question in itself to identify which of the two questionnaires (i.e., triple-based or natural language based) is easier to scan and understand for participants.

Illustrative example: Consider the visualization of the unemployment rate in Münster between 2010 and 2014 shown in Figure 7.1. There are few statements which can be listed to assess this visualization - see Figure 7.3 (the triple-based version is shown on Figure 7.4). Statement 1 is an *identification statement*: it is related to one characteristic of Münster. Statement 2 is a *comparison statement*: it expresses relationships pertaining to the unemployment phenomenon in Münster. Identification and comparison were identified in (N. Andrienko, G. Andrienko, and Gatalsky, 2003) as two elementary cognitive operations for geovisualizations, and apply to city visualizations at large. Statement 3 is a *spatial statement* and may be useful while assessing the spatial learning effects of the visualization. Statement 4 points at *a low-level fact* whereas statement 5 is a about a *higher-level fact* (i.e., a general trend of the dataset). These examples show the significance of the approach proposed here, and its capacity to cope with various flavors of knowledge-building insights in the context of smart cities.

Pros & Cons: There are few advantages of the approach introduced here to measure the effectiveness of visualizations with respect to insight communication. First (and as shown above), the approach is *flexible* and can account for various examples of insights produced by city visualizations. The fact that the questionnaire provides the "I don't know" option helps to establish a user-dependent baseline, and accounts for the (likely) diverse background knowledge of users interacting with city visualizations. Second, the approach is *quantitative*. North, 2006 suggested a qualitative way of measuring visualization insight where users verbalize insights in a think-aloud protocol. The ideas brought forth here provide a useful complement. For example, one could map participants' answers to the set {-1, 0, 1} (where -1 denotes an incorrect answer, 0 is given for "I don't know", and 1 is assigned to a correct answer) and sum the scores. Variants of this rating are conceivable (e.g., give weights to different questions, introduce and assign a score to deceptive questions). Developing ratings which are ecologically valid is an open research question, and would necessitate collaboration with other disciplines, in particular, with researchers working in the field of psychology. Third, accounting for *multiple truths* is possible. Viewing 'truth' as what the visualization expert says allows to account for the fact that different stakeholders may have different priorities/interests when assessing visualization effectiveness. For example, a researcher on spatial learning may focus only on spatial statements to see how much the user knows after interacting with the visualization. A data journalist may design a questionnaire to assess how many (and which) high-level facts the user retains after interacting with her visualization. Fourth, a *semi-automatic generation of the questionnaires* may be possible, in some cases. The visualization from Figure 7.1 is based on unemployment data made open by the Münster City Council as PDF (Portable Document Format) files[5]. The data from the Münster City Council was converted into Linked Data and visualized during a seminar at the Institute for Geoinformatics. The map-based interface visualizes a total of 4399 triples. Listing 7.1 presents an excerpt of the triples. If RDF triples about the dataset visualized are available, one could randomly generate a fixed number of questions from the pool of RDF triples after an interaction session. This necessitates tools to automate the *RdfToNaturalLanguage* translation process (a less demanding task than the reverse operation of *NaturalLanguageToRdf* translation). Using triple-based questionnaires (such as the one shown on Figure 7.4) is also an option worth exploring when the visualization is built on top of RDF data (i.e., triples). With respect to drawbacks, one objection that can be raised about the questionnaire-based approach is that it does not help to account for the temporally elusive aspect of insight - "insight triggered by a given visualization may occur hours, days, or even weeks after the actual interaction with the visualization" (Carpendale, 2008). There are two possible answers to this. First, the questionnaire-based approach is suggested to assess knowledge-building insight, not spontaneous insight. Insight occurring long after interaction with the visualization is spontaneous insight. Second, the questionnaire-based approach is intended to primarily account for the tacit understandability of visualizations. This tacit understandability, in turn, is critical for visualizations that matter. As A. C. Robinson, Demšar, et al., 2017 recently pointed out in the context of geovisualizations, "Maps that matter are those that pique interest, are *tacitly understandable* and are relevant to our society" [emphasis added].

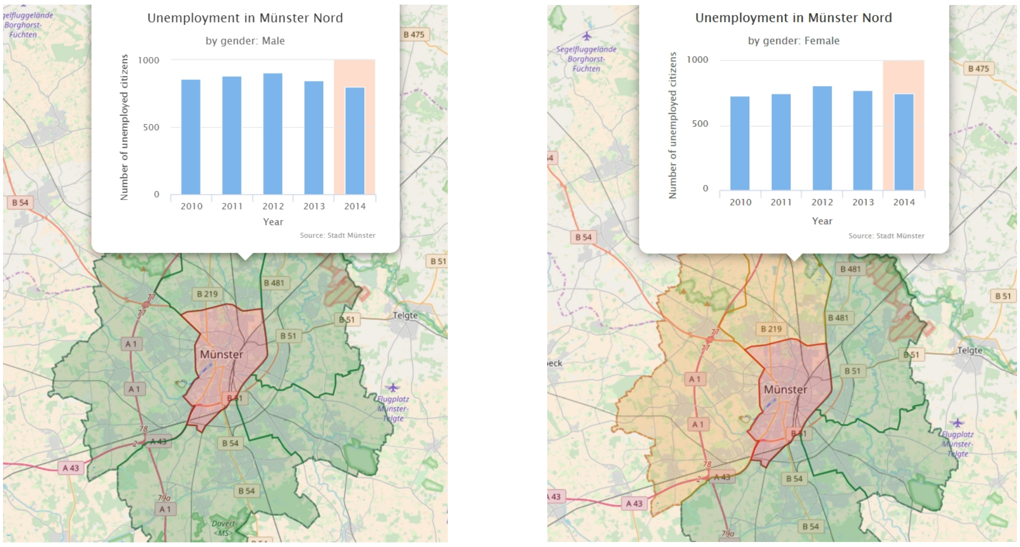
Listing 7.1: Example statements about unemployment in Münster as RDF.

```
@prefix dbpedia: <http://dbpedia.org/ontology/>.
@prefix dc:  <http://purl.org/dc/elements/1.1/>.
@prefix lodcom: <http://vocab.lodcom.de/>.
@prefix sparel: <http://data.ordnancesurvey.co.uk/ontology/spatialrelations/>.
@prefix xsd: <http://www.w3.org/TR/xmlschema-2/>.

# Unemployment in Muenster Nord (2010-2014)
lodcom:nord dc:description "Female Amount of Unemployed in Nord Borough From 2010 to 2014"@en;
        dc:description "Female Anzahl der Arbeitslose in Nord Stadtbezirk Von 2010 bis 2014"@de;
        lodcom:hasFemaleUnemployment2010 "733"^^xsd:integer;
        lodcom:hasFemaleUnemployment2011 "748"^^xsd:integer;
        lodcom:hasFemaleUnemployment2012 "810"^^xsd:integer;
        lodcom:hasFemaleUnemployment2013 "774"^^xsd:integer;
        lodcom:hasFemaleUnemployment2014 "750"^^xsd:integer;
        lodcom:TypeofCityDivision dbpedia:borough;
        sparel:contains lodcom:coerde;
        sparel:contains lodcom:kinderhaus-ost;
        sparel:contains lodcom:kinderaus-west;
        sparel:contains lodcom:sprakel.
```

[5] See http://www.stadt-muenster.de/stadtentwicklung/zahlen-daten-fakten.html (last accessed: September 18, 2017).

Fig. 7.1: A visualization of unemployment rates in Münster.



(a) Rates for males

(b) Rates for females

Fig. 7.3: Evaluating the insightfulness of visualizations. Having questionnaires administered both pre-interaction and post-interaction can help assess the effectiveness of the visualization regarding the number of insights actually made visible.

Which of the following statements are correct?

1. Münster has four administrative districts
   Yes ☐ No ☐ I don't know ☐

2. The number of male unemployed in 2010 is lower than the number of female unemployed in 2010
   Yes ☐ No ☐ I don't know ☐

3. Telgte is located north of Münster
   Yes ☐ No ☐ I don't know ☐

4. Münster North has 1500 unemployed female in 2010
   Yes ☐ No ☐ I don't know ☐

5. Overall, the number of unemployed male in a year is always greater than the number of unemployed female in the same year
   Yes ☐ No ☐ I don't know ☐

Fig. 7.4: Evaluating the insightfulness of visualizations. Questions could be formulated taking advantage of the triple syntax. If the visualizations are built on top of RDF datasets, the process of generating the questions might be automated (to some extent).

Which of the following statements are correct?

1. Münster hasNumberOfAdministrativeDisctricts 4
   Yes ☐ No ☐ I don't know ☐

2. NumberOfMaleUnemployedIn2010 isLowerThan NumberOfFemaleUnemployedIn2010
   Yes ☐ No ☐ I don't know ☐

3. Telgte isLocatedNorthOf Münster
   Yes ☐ No ☐ I don't know ☐

4. MünsterNorth hasNumberOfUnemployedFemaleIn2010 1500
   Yes ☐ No ☐ I don't know ☐

5. NumberOfMaleUnemployed isAlwaysGreatherThan NumberOfFemaleUnemployed
   Yes ☐ No ☐ I don't know ☐

7.6 A Rating Scheme for Visualizations on the Web

Since Linked Data and visualization share some commonalities with respect to transparency, and the former has a star rating scheme[6], it may be worthwhile to explore the idea of a rating scheme for visualizations on the Web. Two reasons motivate this. First, the five stars rating scheme for Linked Data is arguably crude, but it has the advantage that background programs can use it to check existing datasets within minutes (for an example, see the portal of Data.gov.uk[7]). Second, while there is useful ongoing work about benchmarking RDF data (for a recent discussion, see Marx et al., 2016), there are relatively little discussions about how to proceed with the growing amount of visualizations of datasets on the Web. If both a technique for ranking RDF datasets and a technique for ranking visualizations would be available soon, the conjecture made in this article is that a layman would go for the latter first. Graves and Hendler have provided early insights into the wishes of laymen regarding visualizations of open data in (Graves and Hendler, 2013)[8]). They pointed out that there is a "real interest" from the people surveyed to *create*, *reuse* and *explore* visualizations of open data.

[6] `https://www.w3.org/DesignIssues/LinkedData.html` (last accessed: July 21, 2017).

[7] See `http://guidance.data.gov.uk/five_stars_of_openness.html` (last accessed: July 21, 2017).

[8] (Graves and Hendler, 2013) is one of the very few surveys which dealt with users' perceptions about the role of visualization in the context of open data so far. Surveys about open data in general are available see e.g., Schmidt, Gemeinholzer, and Treloar, 2016; John B. Horrigan and Rainie, 2015, but explorations of users' perceptions about visualizations in the context of open data publication and consumption have been less common.

The first and foremost question when it comes to ranking is *what to reward*. The five data rating scheme of Sir Tim Berners-Lee rewards at least four things: the use of an open license, the use of open standards, the use of a structured (i.e., machine readable) format, and linking. There are a plethora of aspects to reward when it comes to visualization: ease of use, effectiveness, degree of interactivity (the more interactivity, the more possibility for the consumer to explore), to name but a few. Not all of these however would be amenable to machine processing. The next paragraph attempts to take advantage of the fact that many visualizations on the Web have *are*, in essence, HTML (Hypertext Markup Language) pages. The goal is to encourage visualization authors to make their visualizations more open (through the supply of structured metadata about them), and to reward the efforts they would put in doing this.

Looking at the aspects rewarded by the five stars Linked Data rating scheme above, and adapting them for visualizations on the Web:

- The first star could reward the *machine-readability* aspect. For instance, a visualization for which the source code can be consulted as HTML would get a star (a mere .png or .jpeg image would not);

- The use of JSON-LD to provide an additional description about the visualization earns the visualization author a further star. JSON-LD[9] has the status of W3C recommendation since 2014 and is an *open standard* for the annotation of content in Web-based programming environments. Recent statistics from the WebDataCommons suggest that JSON-LD is quickly gaining popularity. As of October 2016, JSON-LD ranked second behind microdata as most used method of embedding structured data in HTML[10];

- The use of an *open license* (e.g., Creative Commons) could be warrant a third star (note that the license is only required to be open, and needs not be Creative Commons). Open License is used here in line with the Open Definition to denote a license "which grants permission to access, re-use and redistribute a work with few or no restrictions"[11];

- The use of *open source libraries* could be rewarded by a fourth star. Example of open source libraries include Leaflet (Map), jQuery UI (User Interaction), Intro.js (Documentation). The use of open source libraries should be particularly encouraged on the Web because the survey data from (Graves and Hendler,

[9] `https://www.w3.org/TR/json-ld/;https://json-ld.org/` (last accessed: September 19, 2017).
[10] See `http://webdatacommons.org/structureddata/2016-10/stats/stats.html` (last accessed: September 19, 2017).
[11] `http://opendefinition.org/guide/` (last accessed: September 19, 2017).

Fig. 7.5: JSON-LD description of the visualization of unemployment rate in Münster.

```
17 <script type="application/ld+json">
18 {
19   "@context": "http://schema.org",
20   "@type": "WebApplication",
21   "applicationCategory": "visualization",
22   "name": "Unemployment in Münster",
23   "description": "Visualization of unemployment rates in Münster between 2010 and 2014",
24   "creator": [
25   {  "@type": "Person", "name": "Ana Maria Bustamante" },
26   {  "@type": "Person", "name": "Lukas Lohoff"},
27   {  "@type": "Person", "name": "Jahangir Fahad"},
28   {  "@type": "Person", "name": "Matthias Mohr" } ],
29   "artMedium": "digital",
30   "url": "http://giv-oct.uni-muenster.de/ijald/g1/",
31   "spatialCoverage":{"@type": "Place", "name": "Münster", "sameAs":"http://dbpedia.org/resource/Münster"},
32   "temporalCoverage": "2010/2014",
33   "license": "http://www.apache.org/licenses/",
34   "isBasedOn": [
35     {
36       "@type": "WebApplication",
37       "name": "Bootstrap 3.3.6",
38       "url": "https://maxcdn.bootstrapcdn.com/bootstrap/3.3.6/js/bootstrap.min.js"
39     },
40      {
41       "@type": "WebApplication",
42       "name": "Leaflet 0.7.7",
43       "url": "http://cdn.leafletjs.com/leaflet/v0.7.7/leaflet.js"
44     }
45   ],
46   "supportingData": {"@type": "DataFeed", "name": "RDF Data about unemployment in Münster",
47   "url": "https://github.com/LukasLohoff/js-linked-data-1/blob/master/data/Unemployment%20Triples%20Muenster%202011-2014.ttl"},
48   "availableOnDevice": "Browser, Mobile Phone, Tablet"
49 }
50 </script>
```

2013) suggests that users are keen on (a) modifying the available visualization a little bit, and (b) creating a similar visualization, but using their own data. Open source libraries are a key enabler of these two tasks;

- Finally, *explicitly linking* to the data source (s) visualized could earn a fifth star. This is in line with the wishes of stakeholders surveyed in (Graves and Hendler, 2013) to have some information about the origin of the dataset visualized.

Illustrative example: Figure 7.5 presents the annotation of the visualization from Figure 7.1 in JSON-LD. The annotation can be retrieved from the source code of the visualization available at `http://giv-oct.uni-muenster.de/ijald/g1/`. Concepts and relationships from Schema.org (see `http://schema.org/docs/full.html`) were used for the annotation. From the *<script type="application/ld+json">..</script>* statement (Lines 17 and 50 of Figure 7.5) one can infer both machine-readability and the use of an open standard to describe the visualization, granting it the first two stars. Line 20 of the figure states that the application is a WebApplication, and line 21 further specifies that it is a visualization. Line 33 mentions that the visualization is licensed under the terms of the Apache License (which is open source), earning the visualization a third star. The "isBasedOn" property from Schema.org helps to describe the libraries which were used when creating the visualization. Since both jQuery and Leaflet are open source, the visualization has four stars. Finally, the "supportingData" property from the Schema.org vocabulary enables the specification of the original data on top of which the visualization has been built, leading to a five stars visualization.

Pros & Cons: There are three main advantages of the approach suggested above. First (and as the example has illustrated), the resources to implement it are already available. The example has indeed shown that Schema.org (which is already used by applications from Google, Microsoft, Pinterest, Yandex) could be easily extended to document visualizations in a more telling way for their future search. Second, since Schema.org provides some properties to describe spatio-temporal aspects of an entity, spatio-temporal search of visualization would become possible. The already existing GeoJSON-LD vocabulary[12] make an additional case for the extensibility of the approach to cope with geographic visualizations. Third, JSON-LD (instead of other formats such as Microdata or RDFa) offers the advantage of a clean separation between HTML code and structured documentation of the code, making maintenance of these documents for large websites potentially easier.

As to the drawbacks of the approach, developing heuristics to reliably assign the third and fourth star is a challenge. In particular, more work is needed to find out what to do best when the visualization builds on both closed-source and open-source materials. This is worthy challenge though. From a user's point of view, having software agents crawling the Web (or part of it), and performing a preliminary information aggregation task using the five stars presented above, could be valuable while looking for visualizations for their open data. The second disadvantage of the rating scheme is that it rewards *some* aspects, to the detriment of others. For example, the provision of spatio-temporal metadata is not rewarded at the moment. This drawback is acknowledged, but would be inherent to any rating scheme of the sort introduced here. The final decision of which aspects to reward, and which not, will need concerted effort of the research community as a whole.

7.7 Conclusion

Transparency is an important component of citizen participation and smart cities. This paper has argued that both Linked Data and visualizations are enablers of transparency. They are two sides of the same coin: the former increases transparency primarily for machines, while the latter does so essentially for humans. The article then went on to point out current challenges regarding assessing the effectiveness of Linked Data, and visualizations with respect to transparency enablement. In addition, the paper explored how assessing the effectiveness of visualizations with respect to insight generation could be approached, and take advantage of Linked Data (whenever available). It suggested to view insight as a set of statements made by an author, and to use statements made by experts as ground truth while evaluating the insightfulness of visualizations. The work also proposed a rating scheme for Linked Data visualization which is derived from the well-known five stars open data

[12] See `http://geojson.org/geojson-ld/` (last accessed: September 19, 2017).

scheme, and can be used to make visualizations more transparent with respect to their terms of use (i.e., license), tweaking options (i.e., use of open source libraries), and provenance (explicit linking to the source dataset). The ideas brought forward in this article are useful to advance research and practice of both Linked Data and visualization through the development of techniques to automatically assess the transparency level of visualizations, and their effectiveness in making insights visible to the end user.

Acknowledgments: The author gratefully acknowledges funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, `http://www.geo-c.eu/`). The visualization used as an illustration in the paper was created by (in alphabetical order) Ana Maria Bustamante, Lukas Lohoff, Jahangir Fahad and Matthias Mohr during the Seminar "An Introduction to JavaScript and Linked Data" at the Institute for Geoinformatics, University of Muenster in 2015.

8

A Comparison of Geovisualizations and Data Tables for Transparency Enablement in the Open Government Data Landscape

This chapter is currently under review in the International Journal of Electronic Government Research as *Degbelo, A., Wissing, J. and Kauppinen, T. (2019) 'A comparison of geovisualizations and data tables for transparency enablement in the open government data landscape', International Journal of Electronic Government Research.*[1]

Abstract. *Recent years have witnessed progress of public institutions in making their datasets available online, free of charge, for re-use. There has been however limited studies which assess the actual effectiveness of different communication media in making key facts visible to citizens. This article analysed and systematically compared two representations which are relevant in the context of open government data: geovisualizations and data tables. An empirical user study (N=16) revealed that both types of representations have their strengths: geovisualizations make spatial knowledge and the attractiveness of open government data more visible, while data tables are more adequate for the communication of numerical data. The ideas presented are relevant to open data publishers interested in strategies to effectively put the hidden knowledge in current open government datasets into the hands of citizens.*

8.1 Introduction

The topic of smart cities has attracted growing interest from research, industry and local governments. Many definitions exist, reflecting the plurality of perspectives in the context. Within this article, a smart city is defined after Yin et al., 2015 as "a systematic integration of technological infrastructures that relies on advanced data processing, with the goals of making city governance more efficient, citizens

[1]This article is an extended version of (Degbelo and Kauppinen, 2018), submitted to the special issue on "Enhancing Citizen Centricity with Web Applications in the Smart City Era" - organized as a follow-up to the 4th AW4city workshop where (Degbelo and Kauppinen, 2018) was presented.

happier, businesses more prosperous and the environment more sustainable". Citizen participation (i.e. getting citizens to timely voice their opinions and wishes) is a key aspect of making city governance more efficient and citizens happier. Indeed, as Milakovich, 2010 noted, "Citizen participation provides a source of special insight, information, knowledge, and experience, which contributes to the soundness of government solutions to public problems". Improved citizen participation, in turn, requires greater transparency as citizens must know (or be made known) what is happening in their city and how they can best contribute to it, in order to effectively participate. As indicated by previous work (e.g. Janssen, Charalabidis and Zuiderwijk, 2012; Ubaldi, 2013; Hossain, Dwivedi and Rana, 2016), open government data is a key enabler of transparency. There are several dimensions of transparency discussed in (Johannessen and Berntzen, 2018), but in this work the focus is on what Johannessen and Berntzen, 2018 called 'benchmarking transparency', i.e. the availability of open data (e.g. results from user surveys, demographic information), which citizens and interested parties can use to get a better idea of what is happening within government entities.

Despite a greater availability of open datasets, there is "still a long way to go to put the power of data in the hands of citizens" (The World Wide Web Foundation, 2015). Visualising - or geovisualizing - open data seems the next logical step to put open data in the hands of citizens. Brunetti, Auer, and García, 2012; Brunetti, Auer, García, et al., 2013 formalised the whole process of getting from a raw dataset to a visualisation as a framework called the Linked Data Visualisation Model (LDVM). LODVisualization (Brunetti, Auer, and García, 2012) and LinkedPipes Visualisation (Klímek et al., 2016) are two examples of tools which support LDVM. The current work differs from these two in mainly two ways: (i) a deliberate focus on geographic data preparation, visualisation and interaction (while the two works aforementioned take a more generic approach towards visualisation of open data on the web); and (ii) an account for the transformation from non-RDF data sources to RDF (which the two other tools did not intend to address). The main contributions of this article are twofold:

- An empirical investigation of the merits of table-based and geovisualization-based representations for information search in the context of open government data (OGD). Given that geovisualization creation on top of open government data necessitates human effort, empirical investigations of this sort are needed to increase our understanding of when making that extra investment is sensible, and when not;

- An articulation, based on existing theoretical work and data collected from participants, of the distinguishing characteristics of interactive maps and interactive data tables. The value of this characterization lies in a greater understanding of the strengths and limitations of both types of representations when used as communication media.

Background is presented in Section 8.2, before the introduction of some illustrative geovisualizations developed to increase transparency in the context of OGD (Section 8.3). Section 8.4 presents a controlled experiment done with 16 participants to assess the impact of both types of representations on transparency enablement. Section 8.5 discusses the implications of the results obtained as well as the overall limitations of the work, and Section 8.6 concludes the article.

8.2 Background and related work

Kamaruddin and Md Noor, 2017 identified four components of citizen-centricity which are used as a starting point in this paper: openness, transparency, responsiveness and participation. In line with (Michener and Bersch, 2013), transparency is viewed here as having two dimensions: visibility and inferability. The visibility dimension refers to the extent to which information is complete and easily located; the inferability dimension points to the degree to which information can be used to draw accurate conclusions. Conceptually, a map can be viewed as a *geometric structure* (Peuquet, 1988), *a graphical image* (Peuquet, 1988) or a *set of statements made by an author at a point in time* (Degbelo, 2017). Taking the viewpoint of *maps as statements* as a starting point, web maps are helpful to enable greater transparency in that they can make *value* more visible and inferable. Value of what? Of activities, processes and products pertaining to the public sphere. Why value? Because getting and keeping citizens interested in the participating in public decisions relies upon an appropriate communication of the *value of their participation*. Value, as used here, is in line with Benington, 2009's definition of 'public value', and encompasses "ecological, political, social, and cultural dimensions of value" (or simply said, all that adds value to the public sphere). The remainder of the article will not discuss all possible (and numerous) dimensions of values in the context of public sphere. Instead, it focuses on geovisualizations which enable greater transparency by making the value of open (government) data more visible and inferable. Value of open data has many dimensions (i.e. technical, economic, social, cultural, and political) which were discussed in (Attard, Orlandi, and Auer, 2016). The value creation assessment framework of (Attard, Orlandi, and Auer, 2016) lists no less than 19 (mostly technical) aspects which should be considered when evaluating the potential of an open government data initiative to enable value creation. Three of these 19 aspects were addressed in the current work (see Section 3):

- Data usability: datasets in formats such as Comma Separated Values (CSV), Portable Document Format (PDF) or Resource Description Framework (RDF) are not necessarily citizen-friendly. Visualising them is a way of adding to their value;

- Background context: linking datasets to related datasets (or simply making more specific their semantics through a conversion into RDF) adds value to existing datasets by reducing ambiguities regarding their interpretation;

- Rate of reuse: providing information about the re-use rate of some datasets is a way of unveiling their actual *social value*.

In sum, transparency is key for increased citizen-centricity on the road towards smarter cities. Geovisualizations of open government data can act as transparency enablers in that they help make the *value* of exiting data more visible. The next two subsections briefly review related work on open government data as well as geovisualizations in the context of smart cities. It is worth mentioning here that there is a strong conceptual overlap between city data and open government data. For instance, Ubaldi, 2013 defined open government data to include meteorological data, social data (e.g. statistics on employment, health, population, public administration) and transportation data. All these types of datasets *are* also city datasets by virtue of the fact that their geographical component is tied to (one or many) cities. Put differently, georeferenced open government data is very likely city data.

8.2.1 Open government data

Open Government and Open Government Data have attracted significant attention from research in the recent years. Criado et al., 2018 found transparency and participation to be strongly tied to open government in their review of an international literature covering the period of 2011-2015. Other reviews of the literature have pointed out that OGD includes a wide range of topics, both technological and non-technological ones (Charalabidis et al., 2016), that most common approaches to OGD currently include data portals, data catalogues, and services (Attard, Orlandi, Scerri, et al., 2015), and that potential users of open government data include developers, activists, non-governmental organizations and citizens (Safarov et al., 2017). Factors influencing citizens' participation in open government projects include the perceived enjoyment of the project, the extent to which they believe they can actually change their environment, and their attitude towards civic duties (see Wijnhoven et al., 2015).

Thorhildur Jetzek et al., 2013 pointed out that OGD has the potential to increase social welfare through the generation of economic and social value. Along the same lines, Geiger and von Lucke (2012) indicated that OGD comes with several opportunities such as the modernizing of public administrations in an increasingly open world, the strengthening of an active citizenship, and innovations for citizens and public administrations. Despite these promises, a number of studies (e.g. Zuiderwijk, Janssen, Choenni, et al., 2012; Beno et al., 2017; Benitez-Paez, Degbelo, et al., 2018) have indicated some obstacles on the roads towards reaping these benefits. Solutions to facilitate OGD re-use include frameworks, ontologies as well as tools. Benitez-Paez, Comber, et al., 2018 proposed a framework to improve re-use of open geodata in cities. The framework included four components: user-focused metadata, community of re-use, data users' identification, as well as re-use focused legal terms. Another framework proposed in previous work is the 'Linked Government Data publishing pipeline' (Maali, Cyganiak and Peristeras, 2012), which is based on Google Refine, and aims at enabling data consumers to convert government data of their choice into linked data. With respect to ontologies, Muñoz-Soro et al., 2016 developed PPROC, an ontology to support the semantic description of public procurement processes and contracts. Mockus and Palmirani, 2017 presented the OGDL4M ontology, a collection of terms for the description of legal rules, copyright and database rights in the context of OGD. Regarding tools, Futia et al., 2017 presented a linked data driven approach (as well as a system) to integrate Italian procurement datasets and enhance information coherence in open Italian public procurement datasets. Matheus et al., 2018 argued that dashboards can improve transparency and accountability, and presented two dashboards re-using open government datasets from the city of Rio de Janeiro, Brazil. Degbelo, Granell, Trilles, Bhattacharya, and Wissing, 2019 presented the OCT Transparency Tool, a prototype which aims at increasing transparency by informing about open dataset usage in applications, places from which an app has accessed datasets, and places from which datasets were called. The IES CITIES platform (López-de-Ipiña et al., 2013; Aguilera et al., 2017) takes advantage of exiting open government data, and combines it with urban data generated by sensors as well as user-generated data, to offer a variety of services pertinent to urban life.

8.2.2 Geovisualizations for open government data and smarter cities

A geovisualization can be defined as the 'mapping of geographic information to visuals' (definition adapted from Murray, 2013). In line with (J. Roberts, 2008), geovisualizations can be of one of seven types: Maps/Cartograms, Networks, Charts/Graphs, Tables, Symbols, Diagrams and Pictures. That is, any map is a geovisualization, but a geovisualization need not be a map. Geovisualization is a form of information processing, a compelling form of rhetorical communication, and is more a process of creating than a process of revealing spatial knowledge (see Dodge et al., 2008).

The importance of geovisualizations for the realization of the vision of smarter cities has been acknowledged in previous work. As Dykes et al., 2010 indicated, the quantity, complexity and heterogeneity of the city datasets pose a series of research challenges, and geovisualizations can play a vital role in making sense of these datasets. Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016 listed six citizen-centric challenges for smart cities (i.e. engagement of citizens, the improvement of citizens' data literacy, the pairing of quantitative and qualitative data, the need for open standards, the development of personal services, and the development of persuasive interfaces), and pointed out that maps are a helpful tool in addressing all six challenges. Fechner and Kray, 2014 and Marzouki et al., 2017 argued that geovisualizations can facilitate citizen engagement. In particular, Fechner and Kray, 2014 proposed that maps can be useful in public online dialog platforms, and presented 'Dialog Map', an interactive map which displays engagement opportunities for sustainability projects and open issues. If the conjectures on geovisualizations and citizen engagement in the literature paint a positive future, one must not forget that involving citizens comes with its' own bunch of issues. For instance, Ballatore, 2014 listed a number of issues in collaboratively-generated digital cartographic artefacts (e.g. intentional or unintentional defacement), and these issues are likely to resurface, should geovisualizations be adopted as medium during participatory processes in cities.

Geovisualizations have been used as a tool to support the analysis of criminal activity (Roth, Ross, et al., 2015; Godwin and Stasko, 2017), urban changes over a period of 250 years (Tucci et al., 2010), public health (A. C. Robinson, Roth, et al., 2011) and urban emissions (Ahlers et al., 2018), to name but a few. Recent work has also begun to explore the use of social-media generated datasets for a greater understanding of city processes. Godwin, Y. Wang, et al., 2017 used geotagged tweets to construct representations of neighbourhood topics as typographic maps. A. C. Robinson, 2018 presented a framework to evaluate the design of maps that reach rapid popularity via social media dissemination. Graves and Hendler, 2013; Graves and Hendler, 2014 provided insights into users' wishes in the context of visualization of open government data. They reported for instance that users find it important to (i) know where the data used in a visualization comes from, (ii) know how the data was processed to yield a visualization, and (iii) be given the possibility of modifying existing visualizations a bit, and (iv) bring in their own data, and have a visualization which they have seen on the Web be re-created for them.

Collections of (geo) visualizations for city data on the Web have begun to emerge. Examples include DataMade (`https://datamade.us/`), CityViz (`https://`

cityfutures.be.unsw.edu.au/cityviz/). Visualizing Cities (https://cityvis.io/), CityLab (https://www.citylab.com/) and Data-Smart City Solutions (https://datasmart.ash.harvard.edu/)[2]. A drawback of number of these geovisualizations is that re-using them in different contexts is still a challenge. Degbelo and Kray, 2018 suggested that increasing the intelligence of these geovisualizations could help mitigate that issue.

8.2.3 Summary

This brief summary of previous work illustrates that transparency is an important topic for OGD, and that solutions are coming forth to help mitigate open data re-use issues. Geovisualizations are crucial for OGD. Their importance has been recognized for the broader vision of smarter cities, and they are key too for OGD, since georeferenced OGD is city data. Despite the use of geovisualizations in several smart cities use cases, there is still a need for empirical investigations clarifying the actual role of geovisualizations in enabling (or not) transparency, a topic of importance for both OGD and smart cities visions. This gap is addressed in the remainder of the paper (see Section 4).

8.3 Example web maps enabling greater transparency

As discussed in (Degbelo, 2017), two techniques are particularly suitable to enable greater transparency in the context of open government, namely Linked Data and visualisation. Linked Data increases transparency for machines, and visualisations do so for humans. To illustrate the idea, 36 students (divided into groups of three to six members) were asked to take existing open data, transform it into linked open data, and geovisualise it. The students were part of two classes organized at two consequent years (one class took place with 19 people in the winter term 2015/2017, and the second took place with 17 people in the winter term 2016/2017).

We designed both classes around the idea of blended learning, thus combining activities online with those at the classroom. We shared readings and visualization examples online. This was done in a flipped (or inverted) classroom fashion (see e.g. Mason et al., 2013). Each group also presented their progress in online sessions aired between University of Münster and Aalto University (in Finland). We used classroom sessions for agile co-creation of sketches of data models and visualizations by groups. Different phases of the works by students were presented via gallery walk for getting feedback from other groups, and to support improving their own work.

[2] All links were last accessed on November 20, 2018.

For class designs, we prioritized active learning methods over passive learning ones for both online and classroom.

In the first class, open data from Münster was used as raw data; in the second class, participants were asked to work with open data of their choice. They were all non-familiar with Linked Data, and had various degrees of familiarity with web technologies (like HTML5, CSS, JavaScript or Node.js). The apps based on existing open data, and built as part of the practical work within the classes are: **Crime Mapper** (A1): a web app for citizens and tourists to get a better overview of the crimes in Greater London; **Münster Households** (A2): an interactive map for citizens and city councils to see households data from Münster between 2010 and 2014; **Münster Migration** (A3): an interactive map for citizens and city councils to go through migration statistics from Münster between 2010 and 2014; **Münster Population** (A4): an interactive map for citizens and city councils to browse population data from Münster between 2010 and 2014; **Münster Social Insurance** (A5): an interactive map for citizens and city councils to get an idea about the number of employees subject to social insurance contributions in Münster between 2010 and 2014; **Münster Unemployment** (A6): an interactive map for citizens and city councils to explore unemployment data from Münster between 2010 and 2014; **Referendum Map Münster** (A7): an interactive map for citizens and city councils to see results of the 2016 referendum regarding opening shops in the Münster city centre; and **Wildlife Columbia** (A8): a web app for policy makers and researchers to see information about protected natural reserves in Columbia, and species that inhabit these reserves.

Besides increasing data usability and providing background context about the datasets intrinsically, a novel feature of the web maps is the provision of information of the rate of open data usage. Technically, all web maps use the semantic API from (Degbelo, Trilles, Kray, et al., 2016) which enables app and dataset usage tracking, resulting in greater transparency. Degbelo, Trilles, Kray, et al., 2016 suggested that APIs which return data items according to their types - what they called semantic APIs - would lead to greater transparency (for developers) in an open government context, and identified recurrent categories of open datasets based on a survey of 40 European open data catalogues. Each of the web maps using the semantic API gets a 'transparency badge' (see Figure 8.1, bottom left corner), which indicates their support for dataset usage tracking. By clicking on this badge, the user is redirected to a dashboard-like platform which provides information about all applications available, the open datasets needed for their functioning, and their access rates of these datasets (see Figure 8.2). The information potential of users regarding what is happening with open datasets (i.e. how these are used in one or many apps) is thereby increased. One can also visualise most demanded datasets using the 'Datasets' tab (see Figure 8.2).

The transparency badge is mainly useful here to inform about rate of dataset usage. Yet, its conceptual scope should not be limited to this. One could envision further useful information provided to citizens after a click on a transparency badge. Example of relevant information in the context of open data visualisation include (the list is far from exhaustive):

- *Estimated interaction time to get most out of the visualization*: this estimate could be provided by the creator of the visualization and her experience using it, or computed based on feedback provided by past users of the visualization;

- *Source datasets of the visualisation*: according to the survey from (Graves and Hendler, 2013; Graves and Hendler, 2014), this is a most desired information by participants;

- *Trustworthiness of the visualisation, and of the dataset*: as Tim Berners Lee recently reminded, "It's too easy for misinformation to spread on the web" (Tim Berners-Lee, 2017). The transparency badge could, for example, say whether the data (and/or its visualisation) has been verified by a public institution;

- *Hints about data completeness*: participants from (Beno et al., 2017) mentioned data incompleteness as one of the most severe barriers to open data adoption. Informing about data completeness may not solve the issue, but is already a way forward;

- *Hints about data currency*: the lack of updates of published open data appears at the top of the list of participants from (Benitez-Paez, Degbelo, et al., 2018) when it comes to major barriers to open data re-use. Here also, informing about data updating policies does not solve the issue, but can, at least, help citizens know what to expect;

- *Licensing information about the dataset, and the visualisation*: this is mostly relevant to developers interested in re-use;

- *Purpose of the data and the visualisation*: why the dataset has been collected, and why the visualisation has been created;

- *Adoption examples*: how the dataset has been adopted elsewhere, and how it has been used in that (or these) scenario.

Tab. 8.1: Features of the web maps from (Degbelo and Kauppinen, 2018) - NT: Number of triples; NV: Number of vocabularies used; CV: custom vocabulary.

	NT	NV (CV)	Open Dataset Used	Visual Variables	Operands // Objectives Supported	Work Operators
A1	485,331	10 (1)	crime data	colour hue	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve, resymbolize, overlay, calculate
A2	31,162	21 (1)	households count data	colour hue, colour value	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve
A3	1,391	4 (1)	migration data	-	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve, resymbolize, overlay
A4	6,668	9 (1)	population data	colour hue, colour value	space-alone, attribute-in-space, space-in-time // identify, compare	pan, zoom, retrieve, resymbolize, overlay, calculate
A5	1,869	8 (1)	social insurance contribution data	colour value	space-alone, attribute-in-space, space-in-time // identify, compare	pan, zoom, retrieve, calculate
A6	4,399	6 (1)	unemployment data	colour hue	space-alone, attribute-in-space, space-in-time // identify	pan, zoom, retrieve, overlay
A7	1,327	5 (1)	election data	colour hue	space-alone, attribute-in-space // identify	pan, zoom, retrieve, overlay
A8	2,432	16 (1)	endangered species data, deforestation data, social index data	colour hue	space-alone, attribute-in-space // identify	pan, zoom, retrieve, resymbolize, overlay, search

The final list of the transparency badge's informational items may be decided by its provider. This being said, experience from the food industry (where nutrition facts labels for packaged foods have proven simple and informative to consumers) suggests that standardisation of the informational items of a transparency badge (e.g. through the W3C) could be helpful for the web as a whole at some point.

The detailed analysis of the web maps using the taxonomy of interaction primitives from (Roth, 2013a) and current visual variables (Roth, 2017) was presented in (Degbelo and Kauppinen, 2018). Table 8.1 summarizes the results. As Griffin, 2017 indicated, the use of 'colour saturation' as visual variable in its own right is uncommon. In addition, assessing 'colour saturation' with the human eye is error-prone. Thus, 'colour saturation' was not included in the analysis. The visual variable of 'location' is present in all visualizations, and therefore not mentioned in the table. Finally, the analysis took only the map component of the geovisualizations into account (i.e. other components such as histograms, when present, were not included).

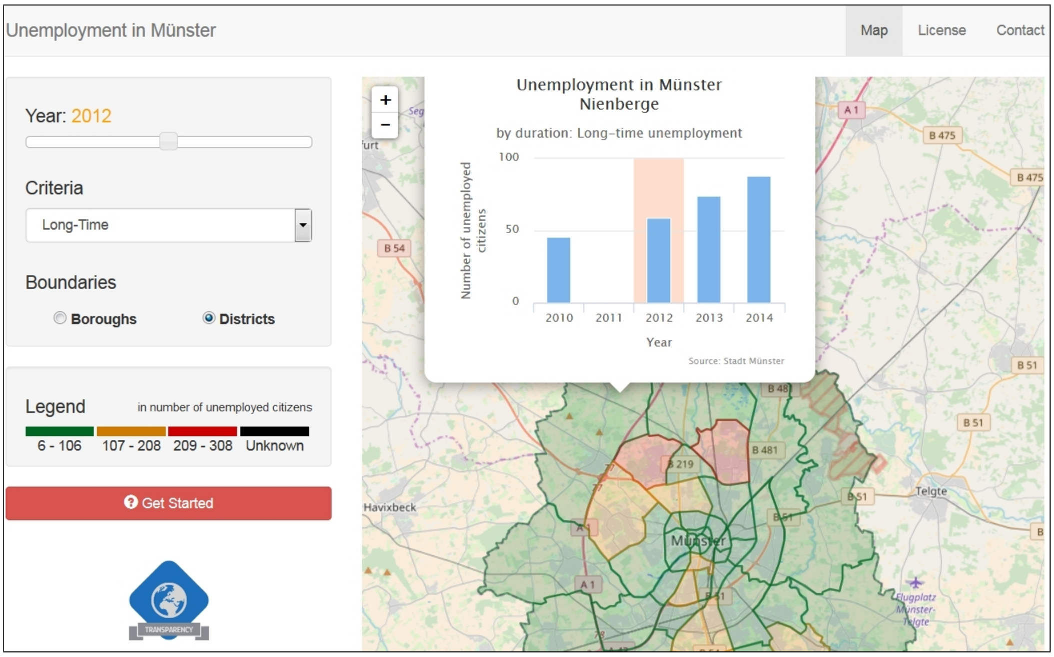


Fig. 8.1: Münster Unemployment - Application with visualises open data from Münster as a web map. The transparency badge signals greater transparency support (i.e. the presence of extra, useful contextual information) for users of the visualisation.

Apps | Categories | Datasets | Usage | Start Tour | Register

More	Name	Description	Category Search	Dataset Search	App Type
(+)	Münster Migration	Visualization of migration statistics from Münster	1	18	WebApp
(+)	Münster Households	Map of households data from Münster	2	163	WebApp
(+)	Crime Mapper	Mapping crimes of Greater London	1	48	WebApp
(+)	Münster SocialInsurance	Employees subject to social insurance contributions in Münster	2	26	WebApp
(+)	Münster Population	Map of population data from Münster	1	57	WebApp
(+)	Wildlife Columbia	Mapping natural reserves in Columbia	1	37	WebApp
(+)	Referendum Map Münster	Mapping referendum data from Münster	1	152	WebApp
(+)	Münster Unemployment	Visualization of unemployment data from Münster	1	39	WebApp

Fig. 8.2: Dashboard-like visualisation of available applications as well as their access rates to existing applications.

8.4 Comparing geovisualizations and data tables for the purpose of transparency enablement

Given the web maps generated in the previous section, the question is whether they truly enable transparency as advanced by previous work (e.g. Degbelo, 2017). A controlled experiment was performed to collect some empirical evidence for this question. The experiment used the geovisualizations A4 and A6 from Table 8.1 and their respective table-based data sources from the city council of Münster[3]. The study has simulated information search by citizens in the context of open government, and rested on the following key assumptions:

- A table-based rendering of some data, and a geovisualization-based rendering of the same data are, in fact, two *representations* of the same data. Representation is used here in the sense of (Mocnik and Fairbairn, 2018): 'Representations are substitutes for, and transformations of, reality and real-world phenomena: they are layers between our understanding and the real world, intended to be used as surrogates for experiencing the real world';

- A source dataset and its geovisualization can be considered informationally equivalent in the sense of (Larkin and Simon, 1987): 'Two representations are informationally equivalent if all of the information in the one is also inferable from the other, and vice versa';

- A representation A is said to enable greater transparency than another representation B, if and only if A provides faster access to some information than B;

- 'Information is the reply to a question' (Bertin, 1981, page 11).

8.4.1 The spirits of map-based and table-based representations

A spirit of a representation is "the important set of ideas and inspirations that lie behind (and, significantly, are often less obvious than) the concrete machinery used to implement the representation" (R. Davis et al., 1993). Table 8.2 presents an overview of most important underlying assumptions of geovisualization-based and table-based representations. The ideas presented in Table 8.2 are adapted from Mocnik and Fairbairn, 2018, who provided a detailed comparison of map-based and text-based representations for the purpose of storytelling. It is worth mentioning

[3] `https://www.stadt-muenster.de/stadtentwicklung/zahlen-daten-fakten.html` (last accessed: November 25, 2018).

here that 'table-based representation' denotes the presentation of a dataset presented in the form of a table in a PDF file (as is often the case of open data in the context of open government). Aspects touched upon in Table 8.2 address characteristics of geovisualizations and tables when they play (R. Davis et al., 1993)'s fifth role of knowledge representations, that is, when they are used as media of human expression and communication. Since the experiment described later primarily focused on the map component of the geovisualizations, the brief comparison that follows only exposes the core similarities and differences between data tables and maps.

- Nature: Following (Bertin, 1981; Ermilov et al., 2013), a table is a matrix, where data items are ordered according to some characteristics; a map on the contrary is a spatially ordered network (see Bertin, 1981; Bertin, 1983). The nodes of this network are points, and the edges are connecting lines between two points;

- Number of dimensions: It follows from their nature that both tables and maps have two dimensions. Since the context here is that of digital representations, these two dimensions are those of the screen;

- Interactivity: both types of representations are interactive, albeit with some noticeable differences. Interactive maps, which are most common in the digital age, are a 'dialogue between a human and a map mediated through a computing device' (Roth, 2013b). Data tables in a PDF file are primarily static, but afford interactivity through the use of the search function. Interactivity in the context of PDF data tables will arguably remain low; in contrast, interactive maps support a range of interaction possibilities from 'low' to 'high' as the cartography cube (see for example Roth, 2013b) suggests;

- Dependencies: As mentioned in (Bertin, 1981), 'any graphic construction originates with a data table'. That is, a map is necessarily dependent on another type of representation, namely the one from which it derives; a data table, on the contrary, can stand on its own. Thus, a data table is a *standalone* representation, while the map is a *dependent* representation;

- Level of geographic detail: as discussed in (Mocnik and Fairbairn, 2018), geographic details cannot be added to maps *ad infinitum*. A map usually has a fixed number of predefined levels of details (e.g. a fixed number of zoom levels), but this is typically not the case for data tables, where the use of text enables boundless levels of geographic details (at least in theory);

- Modelling of spatial relations: a map shows relative locations, spatial hierarchies, spatial patterns and spatial arrangements, representing thereby spatial relations implicitly (Mocnik and Fairbairn, 2018). Since these spatial relationships are exactly similar to spatial relationships between objects in the real world, spatial incoherencies are hardly possible. Data-tables cannot implicitly impose spatial relations (these can then be inferred through a closer scrutiny of the table), and this implies that spatial incoherencies/contradictions are possible.

Tab. 8.2: Similarities and differences between maps and data tables when used as media of human expression and communication.

Aspect	Map-based representation	Table-based representation
Nature	Spatially ordered network	Matrix
Number of dimensions	Two-dimensional	Two-dimensional
Interactivity	Low to High	Low
Dependency	Dependent on a data table	Standalone
Level of geographic detail	Restricted	Boundless
Modelling of spatial relations	Spatial relations implicitly represented; incoherencies hardly possible	Spatial relations need to be explicitly represented; incoherencies possible

8.4.2 User study

The main research question examined during the study is 'what are differences, if any, between geovisualization-based and table-based representation for transparency enablement in the OGD context'? Data to answer the question was gathered during a user study, in which participants were asked to complete a set of information search tasks (see Table 8.3).

Tasks: The six information search tasks were defined based on previous work Roth, 2013a; Roth and MacEachren, 2016. Roth, 2013a presented an empirically derived taxonomy of interaction primitives. The primitives were arranged across the dimensions of *operand* (physical/virtual object with which the user interacts), *goal* (ill-defined task motivating the use of a visualization), *objective* (well-defined task supporting the goal), and *operator* (action supporting the goal). Roth identified five primitive objectives (identify, compare, rank, associate and delineate) and three primitive operands (space-alone, attributes-in-space, space-in-time), and used objective-operand pairings to define and test benchmark tasks for geovisual analytics in (Roth and MacEachren, 2016). To minimize the effect of participants' fatigue, the focus was only on the 'identify' and 'compare' objective primitives in this study. Definitions of both objective and operands for the study, adapted from (Roth, 2013a), are:

- Identify objective: examine an individual data element;
- Compare objective: determine the similarity and difference between two data elements;
- Space-alone operand: interactions with the geographic component of the visualization only;
- Attributes-in-space operand: interactions with the mapped attributes to understand how characteristics or qualities of geographic phenomena vary in space;
- Space-in-time operand: interactions with the temporal component of the map to understand how geographic phenomena change over time.

Tab. 8.3: Information search tasks given to the participants during the study.

Dataset	**Objective**	**Information Search Questions**		
		Space-alone	**Attribute-in-space**	**Space-in-Time**
Population	*Identify*	In which district is the city district of 'Berg Fidel' located? [T1]	How many people live in the city district of 'Berg Fidel'? [T2]	What was the population of the city district of 'Berg Fidel' in 2013? [T3]
	Compare	Among 'Sentrup', 'Aaseestadt' and 'Geist' which one is the northernmost city district? [T4]	Among 'Nord' and 'West' which district has the larger population? [T5]	Has the population in 'Mitte' increased between 2013 and 2014? [T6]
Employment	*Identify*	In which district is the city district of 'Herz-Jesu' located? [T1]	How many employees are in the city district of 'Herz-Jesu'? [T2]	How many employees of the city district of 'Herz-Jesu' were in 2013? [T3]
	Compare	Among 'Handorf', 'Geist' and 'Rumphorst' which one is the northernmost city district? [T4]	Among 'Südost' and 'Hiltrup' which district has the most employees? [T5]	Has the number of employees in 'Mitte' increased between 2013 and 2014? [T6]

Procedure: A within-group design, where participants were exposed to both representations was used. To minimize learning effects, the order of exposure to the representations was counterbalanced, and the participants were randomly assigned to one of four groups: Group1 (A4, D6), Group2 (A6, D4), Group3 (D4, A6) and Group4 (D6, A4). First, participants were welcomed and provided general background information about themselves using a background questionnaire. Second, they performed the six information search tasks. After the completion of these tasks, they were asked to list pros and cons for each form of representation, and rate each of the two representations using the scale introduced in (Lohse et al., 1994). All interaction tasks were recorded using Camtasia 9. Mouse movement and keyboard input were recorded using RUI-Recording User Input[4]. The participants took in average 35 minutes to complete the tasks and fill-in the questionnaires. The study was approved by the institutional ethics board at the Institute for Geoinformatics and ran from July 24th till August 2nd, 2018. Screenshots of the representations used during the study are provided in Appendix.

Participants: 17 participants (seven female), from diverse backgrounds, took part in the study. One participant took much more time (requiring much assistance during the study) and his data was removed from the analysis. Figure 8.3 shows background information about the remaining 16 participants (9 male, 7 female).

As the figure illustrates, the participants were mostly young, but had a quite heterogeneous background with respect to profession, use of OGD, interaction with geovisualization, place of living, and familiarity with the city districts of Münster.

8.4.3 Results

Accuracy: Table 8.4 presents the overall accuracy rates obtained for both representations. Accuracy here denotes the proportion of correct answers provided by the participants, and as one can see, the accuracy rates were high, and similar in both conditions. An analysis of almost 1200 usability tasks showed that the average task-completion rate is 78% (Sauro, 2011). The fact that accuracies (and thus task-completion rates) were high in both conditions (and beyond that benchmark value) suggests that both representations are usable (i.e. appropriate) for the given OGD information search tasks.

Tab. 8.4: Accuracy per task and representation (in percentage).

	T1	**T2**	**T3**	**T4**	**T5**	**T6**	*Average overall accuracy*
Geovizualization1 (A4)	100	100	100	87.5	87.5	100	95.8
Geovisualization2 (A6)	87.5	87.5	87.5	87.5	100	100	91.7

[4] `http://acs.ist.psu.edu/projects/RUI/` (last accessed: November 27, 2018).

Tab. 8.4: Accuracy per task and representation (in percentage).

	T1	**T2**	**T3**	**T4**	**T5**	**T6**	*Average overall accuracy*
Average accuracy per Task	*93.75*	*93.75*	*93.75*	*87.5*	*93.75*	*100*	
Table1 (D4)	87.5	100	100	87.5	100	100	95.8
Table2 (D6)	75	100	100	87.5	87.5	100	91.7
Average accuracy per Task	*81.25*	*100*	*100*	*87.5*	*93.75*	*100*	

Time-based efficiency: time-based efficiency refers to the time taken by the participants to complete the information search tasks. Table 8.5 presents the average durations per tasks and Figure 8.4 informs about the variations per task and participants. R and lme4[5] were used to perform a linear model analysis of the relationship between 'representation' and 'time' taken to find answers to the questions. First, to test for the interdependencies between the two variables of representation and task, a linear mixed model, with fixed effects being representation and task, and subjects (i.e. participants) as random effect was used (see below).

InfoSearch.model = lmer (time ~ representation*task + (1|subject), data = ...)

The interaction between representation and task was significant ($\chi^2(15) = 31.605$, p=0.0072), indicating that the time taken for info search in the context of OGD is *a factor of both task and representation*. Next, six linear models (one per task) of time as a function of representation were built, to shed some light on the within-representation differences when it comes to task performance. The models were not significant for T1, T2, T3, and T5. On the contrary they were significant for T4 ($F(1, 26) = 4.65$, p=0.04) and T6 ($F(1, 29) = 9.648$, p=0.004). In particular, participants using the geovisualizations were 59% faster for T4 (95% confidence interval [3%, 115%]), while participants using the data tables were 46% faster during T6 (95% confidence interval [16%, 76%]).

Tab. 8.5: Time-based efficiency per task and representation in seconds (values in the table are rounded to the first decimal place).

	T1	**T2**	**T3**	**T4**	**T5**	**T6**	*Average overall efficiency*
Geovizualization1 (A4)	37.3	15.6	10.3	40.0	18.5	18.0	23.3
Geovisualization2 (A6)	63.8	13.1	8.7	76.9	51.7	16.7	38.5
Average efficiency per Task	*50.6*	*14.4*	*9.5*	*58.4*	*35.1*	*17.4*	
Table1 (D4)	32.4	14.7	9.9	155.6	25.2	11.5	41.6
Table2 (D6)	42.4	17.3	16.1	128.6	22.8	12.4	39.9
Average efficiency per Task	*37.8*	*16.0*	*13.0*	*142.1*	*24.0*	*11.9*	

[5] `https://github.com/lme4/lme4` (last accessed: November 28, 2018).

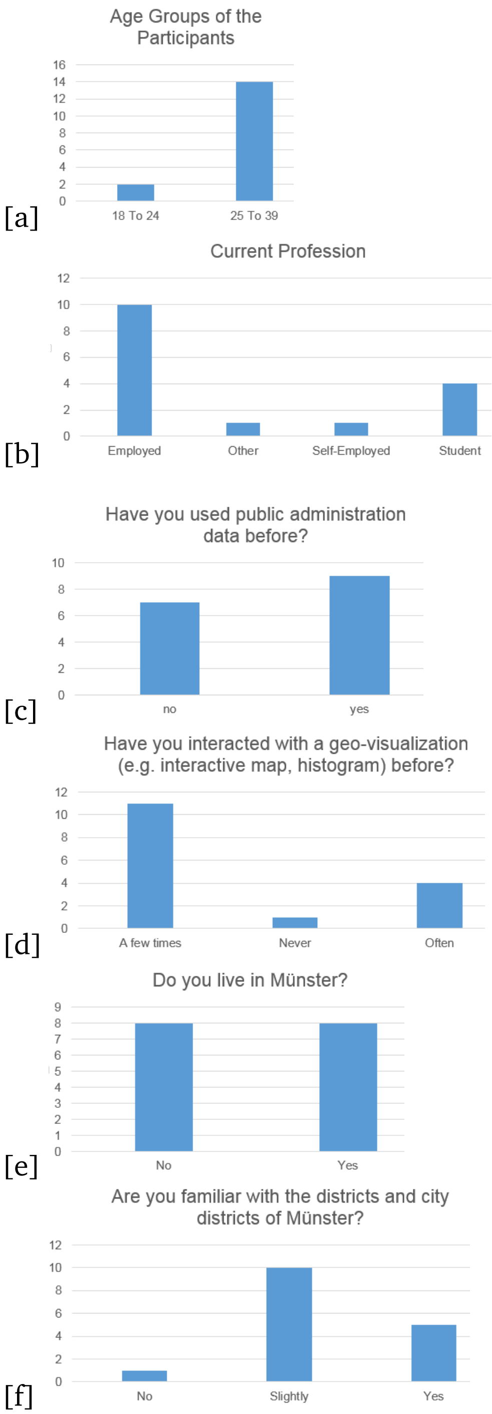


Fig. 8.3: Background information about the participants.

User-ratings of the two representations: Lohse et al., 1994 used 10 scales to ask users to rate 11 different types of representations. All representations were static, that is, there is still a need of replicating their study to unveil properties of representations in the digital age, which are mostly interactive. Since the 10 scales they used in their work were independent, they were re-used to assess the differences between the two types of representations examined in this work. Table 8.6 presents the results. ICC (intra-class-correlation) values were computed using the icc package in R[6], to get an idea of the percentage of agreements between the two groups of participants (ours and those from (Lohse et al., 1994)). ICC is one of the most commonly-used statistics for assessing inter-rater agreements for ordinal, interval, and ratio variables, and is thus appropriate for this study (the 10 scales are on an ordinal level). For the ICC computation, the unit of analysis is 'average' (the ratings are averaged over the participants in both studies respectively), and the type of model is 'two-way' (the same subjects rated all representations) in line with recommendations from previous work (Hallgren, 2012). The ICC values were computed for 'absolute agreement' and were 0.431 (confidence interval [-0.805, 0.849]) for the geovisualizations and 0.679 (confidence interval [-0.275, 0.92]) for the data tables.

A linear mixed model analysis of the relationship between representation and ratings was done to assess the differences between the two representations from the participants' point of view. 10 models (one per rating dimension) were built, with representation as fixed effect, and the participant as random effect. The models were significant on five dimensions (each of these dimensions were originally defined in (Lohse *et al.*, 1994) as the type of knowledge conveyed by the representation): spatial (χ^2 (1) = 55.014, $p < 0.001$), attractive (χ^2 (1) = 18.624, $p < 0.001$), part-whole (χ^2 (1) = 5.006, p=0.026), numerical (χ^2 (1) = 40.502, $p < 0.001$) and dynamic (χ^2 (1) = 97.494, $p < 0.001$). The models were not significant on the other five dimensions (i.e. temporal knowledge, understandability, degree of abstraction, expression of continuous relationships, and amount of information).

User-feedback about the two representations: Table 8.7 summarizes the feedback provided by the users on the pros and cons of both types of representations. The clear layout of the geovisualization and its easing of the comparison of districts were most frequently mentioned by the users as positive features. On the contrary, the users pointed out that the structure of the geovisualization is less clear (i.e. not as predictable as that of the table), and reported the need to use the mouse to see some map labels as negative. Positive features most often mentioned for the table include the easy comparison of attribute values and the clear structure, while negative features most frequently listed were the absence of spatial information

[6] `https://www.rdocumentation.org/packages/irr/versions/0.84/topics/icc` (last accessed: November 28, 2018).

Tab. 8.6: User ratings reported in Lohse et al. (maps/tables), and those from the participants in the study (geovisualizations/table-based representations).

	1 spatial 9 nonspatial	1 nontemporal 9 temporal	1 hard 9 easy	1 concrete 9 abstract	1 continuous 9 discrete	1 attractive 9 unattractive	1 parts 9 whole	1 nonnumeric 9 numeric	1 static 9 dynamic	1 a lot 9 a little
Maps	1.9	2.0	7.5	3.6	4.3	3.9	4.5	3.8	2.4	2.7
Geovisualizations	1.9	6.8	7.7	3.3	3.8	2.3	6.0	4.8	7.9	3.4
Tables	7.1	1.8	5.1	5.1	5.4	5.2	2.5	8.0	2.4	2.8
Table-based representation	7.8	7.4	7.4	2.3	3.9	5.4	4.5	8.6	1.1	2.8

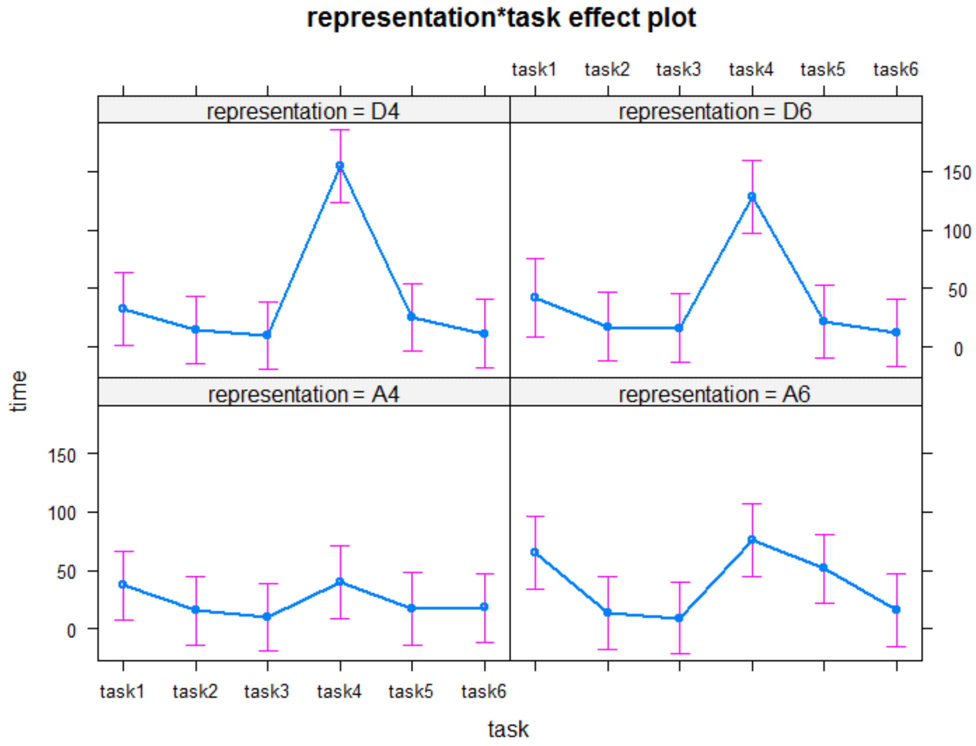


Fig. 8.4: Variations of the times taken by the participants, per task and representation.

(both the locations of the districts and the spatial relationships between districts) and the ordering of districts/city districts (which was not deemed most intuitive).

Tab. 8.7: Participants' positive and negative feedback about the two representations, with their respective frequencies.

Geovisualizations		**Data Table**	
Positive	*Negative*	*Positive*	*Negative*
Clear layout (8)	Structure less clear (4)	Comparison of attribute information is easy (7)	No spatial information (8)
Easy comparison of the districts (5)	Needs mouse hover to display names (3)	Clear structure (6)	Arrangement/Ordering of city districts and districts (7)
Appealing visuals (4)	Searching takes more time when one does not have local knowledge about the locations of the districts (2)	All information at one look (6)	Tendencies are not evident on a first look (2)
User friendly / Intuitive (4)	Precise numbers not immediately visible (2)	Italic and bold emphasis (3)	No interaction possible to sort the data (1)

Tab. 8.7: Participants' positive and negative feedback about the two representations, with their respective frequencies.

Geovisualizations		Data Table	
Selection of specific information possible (3)	Search function missing (2)	Easy to understand (2)	Boring presentation (1)
Tendencies more accessible (2)	Comparison of attribute information more laborious (2)	Uniform data (1)	The amount of numbers is overwhelming (1)
Interactivity (1)	Bubble map does not help if attribute values are similar (1)	Search function of PDF file (1)	
	Comparison of years missing (1)		
	Scrolling of the map necessary to see more (1)		

8.5 Discussion

From the preceding section and the accuracy values obtained, both geovisualization and table-based representations are suitable for information search in the OGD landscape. Given that both representations performed differently when it comes to the time needed to find information, one can conclude that they make different types of information more visible. Both types of representation seem to present no difference when it comes to space-alone, attributes-in-space, and space-in-time questions at the elementary level (i.e. identify questions in Table 8.3). Geovisualizations seem to be more appropriate for space-alone questions which a comparison objective, while data tables seem to be more adequate for space-in-time comparison questions. That the tables performed better here is a bit surprising since one component of the geovisualizations was histograms, which accounted for temporal variations in the data (see Appendix). None of the two representations could offer a significant gain in time when it comes to attributes-in-space comparison questions. In sum, geovisualizations make space-alone comparison knowledge more visible, while data tables make space-in-time comparison information more visible to users [Takeaway1].

User-ratings: The previous systematic comparison between (static) geovisualizations and (static) data tables from (Lohse et al., 1994) is valuable, but the icc values obtained when correlating the static and interactive representations (0.7 for table, and 0.4 for geovisualization) are an indication that the conclusions reached by Lohse and colleagues need to be re-examined in the light of recent technological developments [Takeaway2]. An icc of 0.7 indicates 'good' agreement, while an icc of 0.4 means 'fair' agreement (Hallgren, 2012). In addition, the ratings themselves are valuable for research on effective media for data communication. In particular the relevance of such ratings is that they are a way of elucidating the *mental representations* users intuitively associate with a certain type of representation.

Looking at the user ratings, one can see that each representation has its areas of strengths and weaknesses. According to the users' assessments [Takeaway3], geovisualizations seem, *intrinsically*, to make spatial knowledge, the attractiveness of open government data, dynamic knowledge and holistic knowledge more visible. On the contrary, data tables seem to make numerical knowledge more visible. The user ratings are in general consistent with the time-based measurements, and the time-based measurements actually help to refine some of the claims. Taking into account the observations from the previous paragraph, one can conjecture [Takeaway4] that geovisualizations would make holistic knowledge more visible if that knowledge is of type space-alone comparison. They may not be more successful in making holistic knowledge more visible than interactive tables, if the type of knowledge is space-in-time comparison.

Implications for open data publishing: The user ratings can be used to formulate general recommendations to open data publishers[7]. For instance, Table 8.6 suggests that spatial data should be presented as geovisualizations to citizens (people intuitively associate the problem of communicating spatial knowledge with geovisualizations) and that numerical data should better be presented as table. A recommendation to open data publishers that can be drawn from this [Takeaway5] is that geovisualizations representing numerical data should offer an alternative view of the data as table to users. Numerical data involving spatial information should, whenever possible, be accompanied by geovisualizations. Interestingly, many of the participants mentioned the absence of spatial information as one of the drawback of the data table (see Table 8.7), when in fact the spatial information was *in* the table. That is, data tables may not simply fail to convey spatial information, they can also *obscure* the very fact that spatial information is present.

Another recommendation to open data publishers that can be extracted from the table [Takeaway6] is that publishers can take advantage of geovisualizations to stress the *attractiveness* of open data to the general public. In addition (and consistent with intuition), greater understandability, appropriate degree of abstraction, and adequate amount of information seem much less intrinsic to either type of representation (see Section 8.4.3). Open data publishers should thus devise strategies to maximize these features on a case-by-case basis.

[7] As done earlier for example in (Lohse et al., 1994).

Methodical aspects of representation comparison: As indicated by the cognitive fit theory (Vessey, 1991), matching (a) problem representation to mental representation and (b) mental representation to task, could predict the performance of information presentation formats on specific tasks. 'Problem representation' denotes the way the information is presented to the user (i.e. geovisualization or data table), while 'task' refers to the specific task the user has to perform (i.e. in this case, information finding); 'mental representation', according to (Vessey, 1991), is the way the problem is represented in human working memory. The user ratings are primarily useful for a better understanding of (a) and they can help to predict the performance of information presentation formats on tasks. Since the user ratings touch upon eleven dimensions, they enable a much higher number of predictions than the spatial-symbolic dichotomy suggested by Vessey (i.e. graphs are expected to perform better than tables on spatial tasks, tables expected to perform better than graphs on symbolic tasks).

User ratings need to be *complemented by empirical investigations* of the sort done in this work, based on empirically derived taxonomies as the one proposed in (Roth, 2013a), to get a complete picture of the merits of a representation. For instance, the ratings suggest that users intuitively associate geovisualizations (more than they do with data tables) to the communication of holistic spatial knowledge, yet, space-in-time comparison information was retrieved faster through data tables. That is [Takeaway7], the framework UserRatings+BenchmarkTasks is useful to gather fine-grained insights on different types of representations, on the roads towards general theories of media effectiveness in the OGD context. As discussed in (Whetten, 1989), a complete theory has four components: the 'what' (i.e. relevant concepts), the 'how' (i.e. relationships between the concepts), the 'why' (i.e. underlying factors justifying the relationships between concepts) and the 'who, where and when' (i.e. boundaries of generalizability of the theoretical propositions). In this work, the UserRatings have helped to partially formulate the 'what, how and why' (e.g. geovisualizations make holistic spatial knowledge more visible to users than data tables, because they provide a better fit to their mental representations). The BenchmarkTasks helped to specify the sensitivity to context (e.g. geovisualizations may be more successful in making holistic knowledge more visible than interactive tables, if the type of knowledge is space-alone comparison). The UserRatings+BenchmarkTasks framework used in this work is thus a promising technique for a further investigation of strengths and weaknesses of media in the OGD context (and beyond).

Interaction as an important dimension of graph vs table comparison: previous work (Coll et al., 1994) has compared the relative efficiency of tables and graphs (i.e. bar charts) and arrived at the conclusion that the use of data in a graph form is superior than the use of data in table form, when the task involves the retrieval of relational information. The use of tables is more efficient when the task involves the retrieval of specific value (Coll et al., 1994). The results show that this conclusion does not extrapolate entirely to the case of interactive geovisualization vs interactive data tables. A possible reason could be that interaction as a dimension was left out of the study aforementioned. Research on the overall effect of interactivity on cognition in the Web is still lacking consensus (see F. Yang and Shen, 2017), but interactivity is a dimension which can potentially influence information search results. Put differently [Takeaway8], findings in the graph vs table literature should be refined in light of the recent developments in the information era (and the ensuant possibilities for interactivity), with a controlled assessment of the impact of interactivity. Evidence for this need in the current work[8] is the substantial difference between user ratings from (Lohse et al., 1994) and those provide by users in the current work, on the 'dynamic knowledge' dimension. As Table 8.6 shows, there was no difference perceived by users regarding the two representations back in 1994, but there is a gulf between the two when interaction is added (users rated geovisualization as significantly more 'dynamic' than data tables).

Limitations: L. Wang et al., 2005 pointed out that information seeking performance is a co-result of citizens' characteristics, information task attributes, and the web-based application. The relatively homogeneous group with respect to age (i.e. young people), limits the generalizability of the results to the whole population of citizens. Additional evidence is needed (which more diverse age groups, and a higher number of participants) before a definitive statement can be made on the observations made in this work. In addition, the study has only covered two of the six map interaction goals proposed by Roth, 2013a, and it is likely that the results vary if new types of tasks (i.e. rank, associate or delineate) are included. Contrasting the findings of subsequent studies on these four types of tasks with findings from the existing literature (which mostly touched on 'associate' tasks, e.g. Smelcer and Carmel, 1997; Dennis and Carte, 1998) would be necessary to formulate general conclusions on the respective properties of both types of media.

8.6 Conclusion

This article has presented a synthesis of the distinguishing characteristics between geovisualizations and data tables for the purpose of greater transparency enablement in the context of open government data (OGD). Transparency was defined as making information more visible and the article has assessed the capacities of both types of representation in making six types of information visible: space-alone identify, attributes-in-space identify, space-in-time identify, space-alone compare,

[8] Detailed information about the background of participants from (Lohse et al., 1994) is not available, and the work has assumed throughout that a comparison of the aggregated rating scores across all participants is meaningful.

attributes-in-space compare, and space-in-time compare. A user study with 16 participants led to the observation that both types of representations do no exhibit significant differences on four of the types of information (i.e. space-alone identify, attributes-in-space identify, space-in-time identify and attributes-in-space compare). On the contrary, geovisualizations seem to make space-alone compare information more visible, while the tables make space-in-time compare information more visible to users. The empirical data collected can be used by open data publishers to decide on when to go for one representation or the other, depending on the information search tasks they intend to primarily support.

Acknowledgements: Auriol Degbelo gratefully acknowledges funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, http://www.geo-c.eu/). Tomi Kauppinen gratefully acknowledges Aalto Online Learning (A!OLE), a strategic initiative of the Aalto University, the project which he serves as a project leader. We thank participants of the seminars "An Introduction to JavaScript and Linked Data" (WiSe 2015/2016), and "Geoinformation in Society" (WiSe 2016/2017) at the Institute for Geoinformatics, University of Münster for implementing the web maps mentioned in this article.

Appendix: Screenshots of the different representations used during the study.

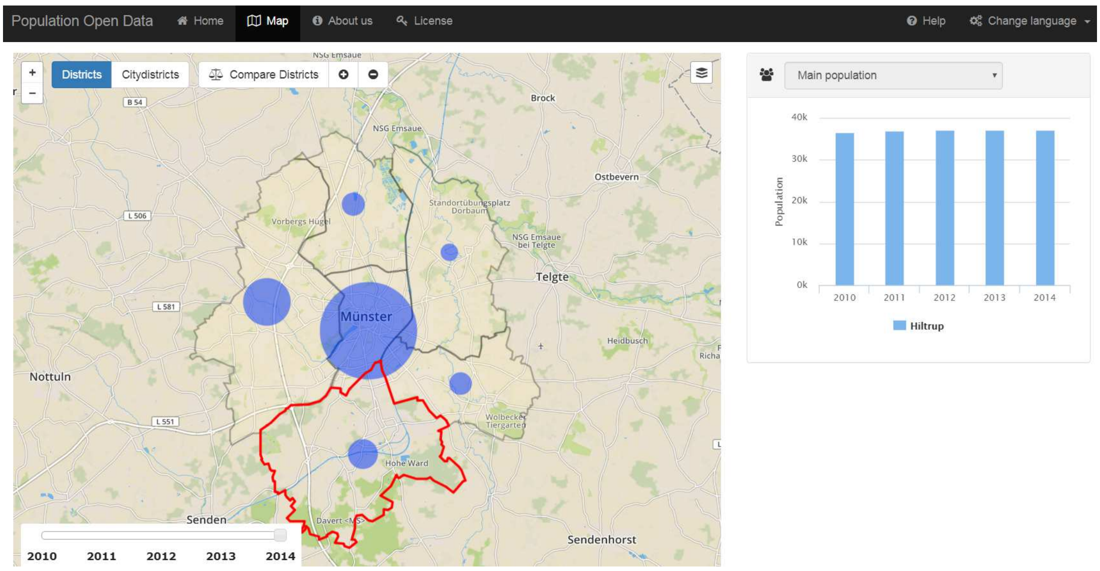


Fig. 8.5: A4

Bevölkerung am Ort der Hauptwohnung 2010 – 2014

Stadtteil *Teilbereich* **Stadtbezirk**	Bevölkerung am Ort der Hauptwohnung in Münster am 31.12.				
	2010	**2011**	**2012**	**2013**	**2014**
			Anzahl		
11 Aegidii	1 286	1 389	1 401	1 421	1 411
12 Überwasser	1 166	1 322	1 315	1 333	1 324
13 Dom	2 026	2 155	2 139	2 190	2 151
14 Buddenturm	2 119	2 286	2 294	2 367	2 394
15 Martini	1 200	1 310	1 346	1 370	1 364
Altstadt	*7 797*	*8 462*	*8 495*	*8 681*	*8 644*
21 Pluggendorf	3 715	4 159	4 227	4 309	4 447
22 Josef	7 497	8 123	8 193	8 537	8 575
23 Bahnhof	962	1 065	1 082	1 178	1 234
24 Hansaplatz	6 255	6 696	6 768	6 762	6 853
25 Mauritz-West	5 641	6 018	6 011	6 116	6 199
26 Schlachthof	4 710	4 993	5 045	5 164	5 215
27 Kreuz	11 319	12 139	12 245	12 307	12 337
28 Neutor	3 509	4 180	4 168	4 164	4 247
29 Schloss	1 947	2 144	2 168	2 198	2 173
Innenstadtring	*45 555*	*49 517*	*49 907*	*50 735*	*51 280*
31 Aaseestadt	5 223	5 385	5 447	5 464	5 415
32 Geist	8 494	8 713	8 741	8 774	8 895
33 Schützenhof	6 790	7 298	7 362	7 532	7 606
34 Düesberg	6 720	6 858	6 911	6 968	6 989
Mitte-Süd	*27 227*	*28 254*	*28 461*	*28 738*	*28 905*
43 Hafen	928	1 043	1 023	1 049	1 039
44 Herz-Jesu	4 846	5 038	5 079	5 070	5 136
45 Mauritz-Mitte	9 683	9 939	9 922	10 113	10 126
46 Rumphorst	7 684	7 828	7 910	8 000	8 047
47 Uppenberg	7 744	8 257	8 397	8 496	8 523
Mitte-Nordost	*30 885*	*32 105*	*32 331*	*32 728*	*32 871*
Mitte	**111 464**	**118 338**	**119 194**	**120 882**	**121 700**
51 Gievenbeck	17 926	19 778	20 348	20 790	20 928
52 Sentrup	5 924	6 829	7 110	7 283	7 506
54 Mecklenbeck	9 142	9 412	8 720	8 855	9 166
56 Albachten	5 688	5 820	5 954	6 094	6 217
57 Roxel	8 246	8 502	8 686	8 866	8 993
58 Nienberge	6 394	6 605	6 631	6 699	6 677
West	**53 320**	**56 946**	**57 449**	**58 587**	**59 487**

Fig. 8.6: D4

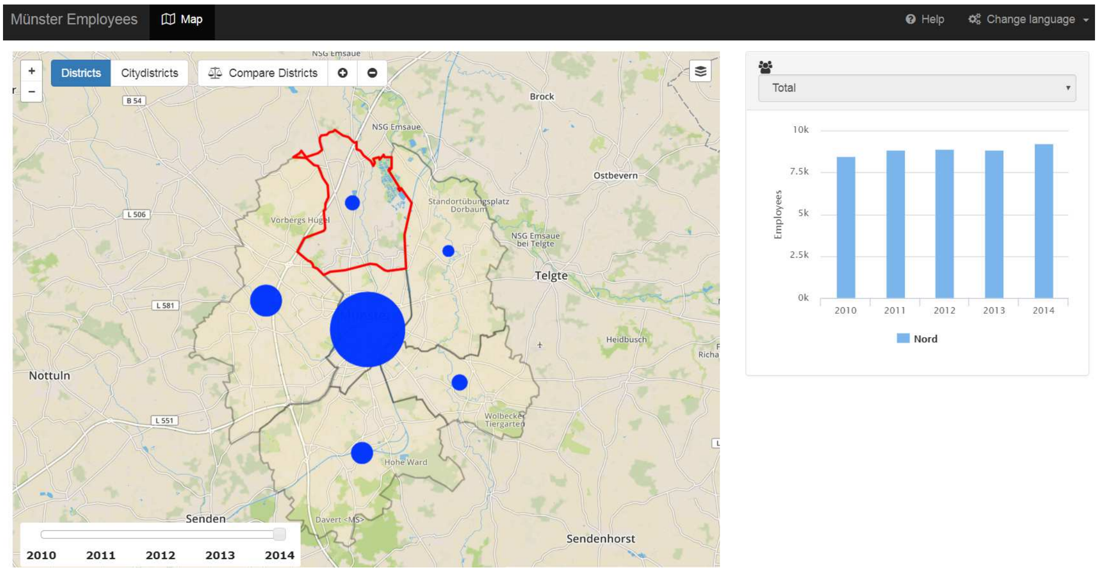


Fig. 8.7: A6

Sozialversicherungspflichtig Beschäftigte

Stadtteil *Teilbereich* **Stadtbezirk**	**Sozialversicherungspflichtig Beschäftigte am WOHNORT Münster am 31.12.**				
	2010	**2011**	**2012**	**2013**	**2014**
	Anzahl				
11 Aegidii	436	455	455	464	466
12 Überwasser	429	471	446	445	450
13 Dom	686	742	789	783	811
14 Buddenturm	672	722	727	747	783
15 Martini	444	462	493	511	483
Altstadt	*2 667*	*2 852*	*2 910*	*2 950*	*2 993*
21 Pluggendorf	1 418	1 519	1 596	1 654	1 675
22 Josef	3 287	3 470	3 554	3 685	3 773
23 Bahnhof	396	430	461	466	501
24 Hansaplatz	2 842	2 983	3 052	3 042	3 063
25 Mauritz-West	2 283	2 448	2 460	2 518	2 616
26 Schlachthof	1 731	1 878	1 912	1 929	1 946
27 Kreuz	4 234	4 467	4 526	4 538	4 661
28 Neutor	1 285	1 417	1 424	1 454	1 468
29 Schloss	637	684	694	686	684
Innenstadtring	*18 113*	*19 296*	*19 679*	*19 972*	*20 387*
31 Aaseestadt	1 684	1 714	1 728	1 798	1 875
32 Geist	3 112	3 221	3 296	3 341	3 481
33 Schützenhof	2 833	3 082	3 135	3 168	3 354
34 Düesberg	2 317	2 385	2 468	2 537	2 591
Mitte-Süd	*9 946*	*10 402*	*10 627*	*10 844*	*11 301*
43 Hafen	395	423	469	468	470
44 Herz-Jesu	1 627	1 743	1 776	1 791	1 900
45 Mauritz-Mitte	3 554	3 695	3 775	3 845	3 949
46 Rumphorst	2 666	2 725	2 792	2 902	2 946
47 Uppenberg	2 632	2 805	2 823	2 882	3 009
Mitte-Nordost	*10 874*	*11 391*	*11 635*	*11 888*	*12 274*
Mitte	**41 600**	**43 941**	**44 851**	**45 654**	**46 955**
51 Gievenbeck	5 777	6 190	6 373	6 474	6 643
52 Sentrup	1 703	1 907	2 003	2 081	2 194
54 Mecklenbeck	3 094	3 209	3 160	3 191	3 266
56 Albachten	1 900	1 932	2 045	2 094	2 240
57 Roxel	2 771	2 960	3 066	3 161	3 273
58 Nienberge	1 951	2 052	2 091	2 112	2 134
West	**17 196**	**18 250**	**18 738**	**19 113**	**19 750**
61 Coerde	2 976	3 131	3 115	3 062	3 274
62 Kinderhaus-Ost	1 561	1 597	1 634	1 683	1 734

Fig. 8.8: D6

Part III

User Empowerment

9 Intelligent Geovisualizations for Open Government Data

This chapter was published in the Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL 2018) as *Degbelo, A. and Kray, C. (2018) 'Intelligent geovisualizations for open government data (vision paper)', in Banaei-Kashani, F., Hoel, E. G., Güting, R. H., Tamassia, R., and Xiong, L. (eds) 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Seattle, Washington, USA: ACM Press, pp. 77–80. doi: 10.1145/3274895.3274940.*

Abstract. *Open government datasets (OGD) have been flooding the Web in recent years. Geovisualisations are the natural way of making sense of them, and have been gradually coming out. However, one key problem is the lack of flexibility of these visualizations, which severely limits their re-use in new scenarios. This article therefore proposes to increase the intelligence of existing geovisualisations by incorporating five features, to make better use of OGD: (i) automatic geographic data type recognition, (ii) generation of geovisualisation designs, (iii) monitoring of users' understanding of geographic facts, (iv) self-optimization, and (v) user activity recognition. In addition to benefiting users of OGD, realizing these features presents rich scientific challenges and opportunities for Geovisualization research, the OGD landscape (and beyond).*

9.1 Motivation and Background

Open data laws, political movements and other drivers have led to the increasing availability of public and governmental data. Open Data Inception (`opendatainception.io`) lists no less than 2600 open data portals around the world, covering a broad range of topics such as education, weather, population, environment and heath, to name just a few. Some of these portals (*e.g.*, the European Open Data Portal, `https://www.europeandataportal.eu/` or the OpenDataSoft's data network, `https://data.opendatasoft.com/`) provide widgets to visualize these datasets. Since in many cases, data is linked to spatial concepts (such as zip codes, districts or even coordinates) the relevancy of geovisualizations in this context has increased as well. This trend has, however, also highlighted some key issues in this context. In particular, the (geo-)visualizations for different datasets vary, and if untrained users try to generate them from raw datasets there are few if any safeguards to prevent the selection of inappropriate visualizations. These issues make it difficult for citizens to understand the meaning of, and to compare different datasets (*e.g.,* from the same governmental body or from different ones).

There are at least three reasons why more versatile (or flexible) visualizations will have a positive impact on the Open Government Data (OGD) landscape: 1) there is a need (confirmed in a previous survey, see (Graves and Hendler, 2013)) of a portion of the OGD consumers' population not just to consume existing visualizations, but to bring in their own data and visualize it as they like; 2) the expertise to produce a visualization from a dataset is not always available; and 3) even if the expertise (*e.g.,* programming skills) is there, the time to create a new visualization is not always available.

This article presents a vision of intelligent geovisualizations for open government data in order to inspire future research that will make OGD more useful and accessible to a broad audience. The vision covers five main aspects: (i) automatic geographic data type recognition, (ii) generation of geovisualization designs, (iii) monitoring of users' understanding of geographic facts, (iv) self-optimization, and (v) user activity recognition. Intelligence is defined after Albus, 1991 as *the ability of a system to act appropriately in an uncertain environment* (emphasis added). For each aspect, we first introduce the 'uncertainties' the geovisualization needs to cope with, followed by existing work on which future endeavours can build upon to realize intelligent OGD geovisualizations.

We use the term 'Geovisualization' in line with J. Roberts, 2008 to broadly include interactive maps, network graphs, charts/graphs, tables, symbols, diagrams and pictures. Geovisualizations can act as catalysts for citizen engagement in the OGD landscape (Fechner and Kray, 2014). In keeping with Albus, 1991, intelligence requires at least the ability to sense the environment, make decisions, and control action. The sort of intelligence to be attained in machines does not have to simulate human intelligence see Taylor, 2009. 'Open Government Data' is an inherently ambiguous term see *e.g.,* Yu and D. G. Robinson, 2012, but is used here to denote public sector data which is freely available for re-use. Following Roth, 2013b, a user (*i.e.,* an open government data consumer) is modelled as having three characteristics: expertise, ability, and motivation. The anticipated main beneficiary of this vision has low expertise in programming and geovisualization, possibly low/medium spatial thinking abilities, and desires to consume open geographic data out of curiosity. Though the features of an intelligent geovisualization are all introduced at once

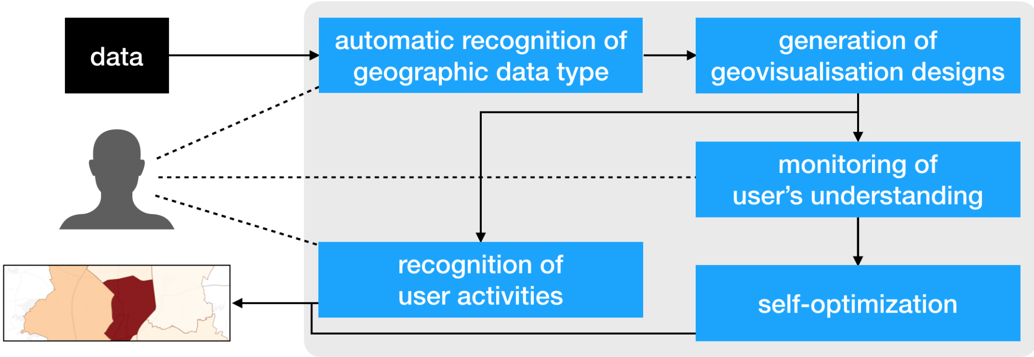


Fig. 9.1: From open government data to intelligent geovisualizations. Dotted arrows suggest implicit or explicit input from the user in the process of realizing the feature.

here (see Figure 9.1), it is not required that a single geovisualization has them all. Arguably, any geovisualization which supports any of the features presented next, exhibits some *sort of* intelligence. The more features supported, the more intelligent the geovisualization.

9.2 Recognize geographic data types

If an untrained user is to be able to easily get a useful visualization for an open geographic dataset, a critical feature of the next generation of geovisualizations would be the ability to automatically recognize the *semantic type* of the geodataset at hand. This task is fraught with much uncertainty for the computer (and the untrained data analyst) because there are diverse semantic types for spatial datasets. Ferreira et al., 2014 identified time series, trajectories and coverage as basic types of observations; Scheider, Gräler, et al., 2016 proposed fields, inverted fields, lattices, events, and objects; Stevens's distinction between nominal, ordinal, interval and ratio data is also relevant in the context of spatial information. The uncertainties are further aggravated when one considers that there is a many-to-many relationship between semantic (*e.g.,* trajectory) and syntactic types (*e.g.,* GeoJSON, Comma Separated Value, or Resource Description Framework) of geodatasets. We have recently begun experimenting with automatic recognition of Stevens's four types of datasets, taking a GeoJSON dataset as a starting point. One lesson learned is that automatic semantic geodata type recognition is a human-computation problem (in the sense of Ahn, 2009). Hence, the approach of algorithmically-guided user interaction proposed recently in (Dijk and Alexander Wolff, 2017) is one promising way of approaching this issue.

9.3 Propose geovisualization designs

'Insight is formed through interaction' J. Roberts, 2008, and untrained users are interested mostly in the information/insight they can extract from a dataset (not in the raw dataset per se). That is, after having identified the type of spatial data at hand, an OGD geovisualization system should propose meaningful visualization designs pertaining to the current dataset. Here also, there are uncertainties, not so much on the computer side, but more on the side of the untrained data analyst. Indeed, pitfalls and best practices of geovisualizations may be well-known to experts, but unknown to non-experts. For instance, novice map makers have the tendency to overload a single map with too much detail, while it is usually better to err on the side of simplicity and to produce two or three maps, each focused on a single topic (Kent and Klosterman, 2000). Count data (*e.g.,* number of births in a country) is better represented through a 'bigger-means-more' coding in a dot-array map; representing count data using the 'darker-means-more' rule in a choropleth map is misleading (Monmonier, 2005). In his seminal work, Bertin, 1983 suggested eight 'visual variables' (*i.e.,* graphical dimensions across which a map/visualization can be varied to encode information) which could be useful while producing automatic designs. Rules prescribing the use of a visual variable given Stevens's data types were recently summarized in (Roth, 2017): for instance, color, hue, orientation and shape are useful for encoding nominal information; ordinal information is better encoded using other visual variables such as color value, crispness and transparency.

While these heuristics have been documented in previous work, a consolidated set of guidelines for the design of interactive maps is yet to be produced (see Roth, 2013b), and the same holds true for other types of geovisualizations. Automatically suggesting designs for datasets would need a democratization of the heuristics listed above, and of further best practices of geovisualizations. Bresciani and Eppler, 2015 lists common pitfalls of information visualization in general. Their list (and in particular the designer-induced cognitive pitfalls) can be translated into constraints for the automation process so that no design encodes input data in a way that gives rise to a "pitfall".

Rawgraphs (Mauri et al., 2017) provides useful features for the automatic visualization of spreadsheet data, and could be modified for the case of geographic data. The SIGSPATIAL community has done some work on automatic generation of road networks from GPS traces (see *e.g.,* Cao and Krumm, 2009; Duran et al., 2016), which addresses one specific type of geovisualization (*i.e.,* network), and might inform future approaches that automatically produce more complex geovisualizations (*e.g.,* interactive maps and timelines). Zavala-Romero et al.'s work on generating web GIS without programming knowledge (Zavala-Romero et al.,

2014) is in line with the idea outlined here, but only implements a small subset of it (*i.e.,* their tool automatically builds web GIS interfaces to visualize NetCDF data). Existing work on map labelling (*e.g.,* Barth et al., 2016; Peng et al., 2014), or computing aggregation of large point datasets Beilschmidt et al., 2017 is also relevant in this context. Insights from ongoing work on meaningful spatial prediction and aggregation Stasch et al., 2014b would also help tackle the issue presented here.

9.4 Monitor understanding

Once users have a geovisualization design (or several geovisualization designs) for their dataset, the next big question is whether they *truly* understand it. If we are to realize Shneiderman's vision of designing computer-based tools which 'enable more people to be more creative more often' (Shneiderman, 2007), this question is of great importance. Here the uncertainties, primarily on the side of the computer, lie in getting it to determine whether the geovisualization is understandable for the current user. In a classical geovisualization maker/user setting, the geovisualization maker can informally ask the user (or observe her) to assess the extent to which her goal (*i.e.,* transmit geographic insight) is attained (or not). Getting to the point where the geovisualization (reliably) collects this feedback by itself will require years of sustained research effort.

One approach to tackle this issue is the use of computer-generated questionnaires. Using these questionnaires to monitor understandability could drive an 'anticipation feedback loop' (Jameson and Wahlster, 1982), where the computer modifies and adapts designs until the users' interpretation corresponds to the one by the computer system. The questionnaire-based insight measurement approach proposed in (Degbelo, 2017), and the idea of automatically generating insight-questions from the source dataset(s) of a geovisualization, could also be beneficial at this point. The underlying assumption behind this idea is that a source dataset and its geovisualization can be considered informationally equivalent in the sense of Larkin and Simon, 1987. In terms of 'metaphors we compute by' Videla, 2017, computer-generated questionnaires approach the monitoring issue by gathering *information about users' understanding from their self-reports*. Another, equally relevant metaphor worth exploring is that of *information about users from behavioral observations, i.e.,* explore the extent to which bread crumbs of their requests, keyboard inputs, mouse clicks or other relevant user activities can be used to make reliable inferences about their current level of understanding.

9.5 Optimize geovisualization designs

Once users have provided feedback on how useful the geovisualization is when trying to understand the data at hand, there is the question whether there could be a geovisualization doing an even better job at supporting them. Since they are not aware of all the intricacies of the geovisualization generated, the onus of answering this question is on the computer. That is, the uncertainties arise here mainly on the side of the computer.

Insights from work on computational user interface design (for a review of approaches, see Oulasvirta, 2017) could be useful while addressing this issue. The main question here is that of the objective function for the optimization. Here, the suggestion is that *understanding* on the side of the user is used as objective function (so that the geovisualization ultimately enables people to effectively act intelligently). In the context of Open Government Data, understanding as an objective function is critical: unless there is an (understandable) visualization, knowledge is not really put into the hands of the data consumer (*i.e.,* one of the main goals of opening up government data is *not yet achieved*). Empirical guidelines on factors which facilitate the generation of geographic insight during interaction are scarce (Roth, 2013b), and more work is needed in this area to (i) identify these factors, and (ii) produce algorithms which maximize understanding as an objective function of computer-generated geovisualizations.

Constraint-based visualization systems, which consider both expressiveness and perceptual effectiveness while automatically generating visualizations, are relevant in this context. Expressiveness here refers to the ability of a visualization to convey all facts in the data, and only facts in the data, while effectiveness denotes the visualization's ability to convey facts in a way that they are readily perceivable by the end users (see Mackinlay, 1986). Existing systems (*e.g.,* Voyager (Wongsuphasawat et al., 2016) or Draco (Moritz et al., 2018)) have been used to produce bar/line charts and scatterplots, and more work will be needed in the future to expand them to more complex interactive geovisualizations.

9.6 Recognize user activities

A fifth, useful feature of the next generation of geovisualizations is to automatically determine what users are trying to do, and support them well in doing that. For example (and as suggested in G. Andrienko and N. Andrienko, 1998), multiple different presentations of the same data should be proposed during the data exploration phase. Shortcuts to save (or share) the geovisualization could be hidden by default (reducing the number of symbols to render on the interface), and displayed when it makes most sense (*i.e.,* towards the end of the interaction session). If basic interaction scenarios for a given geovisualization are specified (like in N. Andrienko and G. Andrienko, 2001), intelligent guidance could try to fit users' activities to one of the most plausible scenario, and use that knowledge to suggest functionalities to try out next at key stages of the users' navigation. The uncertainties here (mostly on the side of the computer) lie in pinpointing exactly what users are trying to achieve. This is an underexplored area in geovisualization research, and Fabrikant's early call to 'formulate domain independent visualization tasks that are generic enough to be effectively shared amongst a heterogeneous user community' Fabrikant, 2001 is as topical as it was 20 years ago.

Taxonomies of interaction tasks have begun to emerge, and will be useful in addressing this issue. Roth, 2013a suggested three broad interaction goals (*i.e.,* procure, predict and prescribe) and five more specific objectives (*i.e.,* identify, compare, rank, associate, and delineate) for geovisualizations based on an empirical user study. Roth's taxonomy proved usable while characterizing web maps, but still needs refinement so that answering the question of 'which interaction operators matter for which interaction objective' becomes possible (see Degbelo and Kauppinen, 2018). Kiefer, Giannopoulos, and Raubal, 2013 considered six main map-based activities during their work: free exploration, (global) search, route planning, focused search, line following, and polygon comparison. Though they did not provide the rationale for choosing only these six types of activities, their study using eye movement characteristics to automatically detect map users' activities ended with an overall accuracy of 78%, providing evidence for the pertinence of these six types. Building upon the works mentioned here will be crucial for the implementation of algorithms, which enable automatic user activity recognition for intelligent geovisualizations.

9.7 Further remarks

Besides the five kinds of uncertainties listed above, a number of further features are desirable for the success of the vision. Notably, the geovisualization should tell about it's decision tree (and allow users to modify it) - or put differently comply with the principles of 'algorithmic transparency' (Brauneis and Goodman, 2018). Algorithmic transparency (which very likely would require visualizations so that untrained professionals can benefit from it) would be helpful to build trust between users and the geovisualization - and indirectly between users and the (public) institution (or community) using it as a means to make its data more understandable. Furthermore, multi-device portability (*e.g.,* building on the early idea of interface plasticity, see Thevenin and Coutaz, 1999) and multi-modality support (*e.g.,* account for speech and hand gesture) could be convenient add-ons, though not discussed explicitly here.

Finally it's worth mentioning few key differences between this vision, and the occurrence of the term in previous articles. Yingjie et al., 2001 proposed a related discussion on adaptive geovisualizations, focusing mainly on architectural aspects. Instead, the focus of this work is deliberately functional, and the paper has brought forth five *user requirements* of intelligent geovisualizations. Using the verb form to formulate the title of each of the previous sections is intentional, and helps to highlight this ('the system should recognize the geographic type of my current dataset', 'propose (to me) meaningful geovisualization designs', and so on). Furthermore, the article used the existing literature to point out where we currently stand, and to outline ways forward (highlighting that much still needs to be done to get where we want to be). N. Andrienko and G. Andrienko, 2007 proposed a system which can be viewed as an early implementation of the 'generation of geovisualization designs' feature mentioned above. Their distinction between domain-dependent and domain-independent components of an intelligent visualization is useful, and this article has outlined steps currently needed to further advance research on the latter type of component, pointing out that recent progress in geo-ontologies, information visualization research, computational user interface design, and the science of cartographic interaction offer new prospects for tackling the issue of automatically picking the right geovisualization for a given task.

9.8 Conclusion

Open Government Data (OGD) has been flooding the Web in recent years, and geovisualizations are gradually coming out to help make sense of them. This article has suggested increasing the intelligence of these geovisualizations to enhance their flexibility (and ability to be re-used in several scenarios). The vision covers five main aspects: automatic geographic data type recognition, generation of geovisualization designs, monitoring of users' understanding of geographic facts, self-optimization, and user activity recognition. In modelling intelligence as the ability to act appropriately in an uncertain environment, the work provides a clear criterion to distinguish non-intelligent geovisualizations from intelligent ones (or less intelligent geovisualizations from more intelligent ones). The vision has taken OGD as main scenario, but intelligent geovisualizations would be beneficial in a number of other areas, notably data journalism (recognize spatial data types & propose geovisualization designs), E-participation (monitor understanding), computer-assisted spatial learning (optimise geovisualization designs) and pedestrian/car navigation (recognize user activities). In sum, intelligent geovisualizations are essential to make OGD understandable and offer many opportunities for future research.

Acknowledgements: Funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, http://www.geo-c.eu/) is gratefully acknowledged. We also thank the anonymous reviewers for their feedback.

10 A Semi-Automatic Approach for Thematic Web Map Creation

This chapter is currently under review in the Cartography and Geographic Information Science Journal as *Degbelo, A., Sarfraz, S. and Kray, C. (2019) 'A semi-automatic approach for thematic web map creation', Cartography and Geographic Information Science.*

Abstract. *Open Government holds the promise of increasing transparency by providing citizens with datasets about city processes, and open data portals have been emerging all over the world as mines of open geographic datasets. Thematic web maps are key in making sense of these open geographic datasets. Current thematic web maps are created by people having programming and/or cartographic expertise, and are not designed in such a way that they can be easily re-used with new geographic datasets. As a result, the obstacle to overcome by non-experts willing to adapt them to new scenarios is high. To lower the hurdle, this article introduces a semi-automatic approach for the creation of thematic web maps, by and for users with no prior training in Cartography. The approach relies on the mapping between Steven's data types and Bertin's visual variables to suggest (meaningful) thematic map visualizations for a given input geographic dataset. It was implemented as a web prototype in AngularJS and evaluated with 19 participants. Results from the user study suggest that despite few challenges in correctly finding out Steven's data types, participants managed to successfully create web maps and were able to correctly identify geographic facts. The prototype and the insights gathered during the user study are relevant to make cartographic products more accessible to a broader population, and make open geographic data in the context of open government more usable.*

10.1 Introduction

Open government data (OGD) has been increasingly available, and the issues of more accessible government data for society have attracted the interest of many researchers. Recent systematic reviews have revealed that OGD has many uses such as decision-making and innovation (Safarov et al., 2017) and that the OGD research domain is versatile (Charalabidis et al., 2016). Previous work has also pointed out that the most direct impact of OGD is greater access to information (Attard, Orlandi, and Auer, 2016), and that there is a general consensus in the literature that OGD increases government transparency (Hossain et al., 2016). Despite these promises, there are still a number of issues preventing users to take full advantage of OGD as documented in (Benitez-Paez, Degbelo, et al., 2018; Zuiderwijk, Janssen, Choenni, et al., 2012). Zuiderwijk, Janssen, Choenni, et al., 2012 identified 118 issues preventing the full adoption of open data including usability and understandability barriers (e.g. expert advice needed to understand the data, data not understandable to the general public, data not visualized or lack of support/help/training for the use of the data). Likewise, Benitez-Paez, Degbelo, et al., 2018 reported that usability barriers (e.g. data difficult to understand, no applications to validate the usability of available data) are preventing users to take full advantage of current open government geodata. A premise of this work is that geovisualizations can help to mitigate these issues.

The role of visualizations in making OGD more usable has been stressed in a number of articles within GIScience (e.g. Degbelo and Kauppinen, 2018; Degbelo, 2017; Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016) and outside the field (e.g. Graves and Hendler, 2013; Graves and Hendler, 2014). In particular, interactive maps have been identified as potentially increasing transparency (Degbelo and Kauppinen, 2018; Degbelo, 2017), catalyzing citizen engagement (Fechner and Kray, 2014), and critical in addressing issues related citizens' data literacy (Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016). Graves and Hendler, 2013; Graves and Hendler, 2014 provided an empirical survey aiming at collecting users' perceptions towards the use of visualization tools for OGD. They found that both experts and non-experts value the possibility of bringing in their own data, and creating a visualization for it.

In their online survey of heterogeneous users, Poplin et al., 2017 found that the creation of maps still tends to represent a challenge to many people. Following Roth, 2013b, there are four broad strategies to assist these users: (i) the introduction of two modes (i.e. regular vs expert mode) in current online mapping applications; (ii) the development of richer and more effective training materials for the novices; (iii) the recourse to expert systems (i.e. encode knowledge from cartography experts in systems which then provide appropriate interaction solutions depending on the context), or (v) the provision of map brewers (i.e. cartographic design systems which recommend appropriate design solutions to users and enable them to choose from that subset). The approach taken in this work falls under the fourth category, and this article intends to advance the state of the art on web map creation along two complementary axes: (i) enable novice users to re-use an existing thematic map visualization to explore their own dataset; and (ii) offer novice users suggestions of (meaningful) thematic map visualizations which can be created based on an input geographic dataset. Throughout the work, a novice user is defined as a person who has limited expertise in Cartography. The main contributions are threefold:

- A snapshot of most common thematic maps currently used for urban data visualizations, based on a survey of 40 visualizations taken from a wide range of online sources;
- A semi-automatic approach to help novice users create a wide variety of thematic maps. This approach relies on the identification of Stevens's level of measurement of the attribute information, and the use of Bertin's visual variables;
- An open-source, web-based prototype as proof of concept for the approach, and insights from an evaluation of the prototype with 19 participants.

The approach and the prototype developed as proof of concept to illustrate its feasibility are introduced in Section 10.3. The user study conducted to evaluate the approach is presented in Section 10.4, followed by the results obtained (Section 10.5), and a discussion of their implications in Section 10.6. Section 10.7 concludes the article and mentions possible directions for future work.

10.2 Background

This section briefly reviews existing work on online maps and adaptive maps. In general, the number of online maps has increased dramatically in the recent years, but using and creating these maps is still a challenge for a portion of the population. In addition, there is still a need for an adaptive web mapping system to leverage existing open geodatasets.

10.2.1 Online maps

The number of online web maps has exploded in recent years, but as remarked by Plewe, 2007; Tsou, 2011, web mapping research seems to have not been as dynamic as web map production. Current studies have touched on topics such as technologies for web map development, users' perceptions and the social dissemination of maps in the digital age. Roth, Donohue, et al., 2014 explored technologies currently available for web mapping. They identified four types of technologies (i.e. frameworks, open libraries, closed application programming interfaces and tile rendering services), and reported that the large majority use JavaScript as the base programming language. Early work (Harrower and Brewer, 2003; Brewer, 2003) has produced ColorBrewer, a web tool to select color schemes for thematic maps. ColorBrewer is now used in a number of web mapping libraries including Mapbox[1], D3.js[2] and Leaflet.js[3]. Poplin et al., 2017 presented a user-centered analysis of the Worldmap platform, and reported three most challenging activities for users, namely: finding information about data in WorldMap; adding one's own data to WorldMap and use it; and creating one's own map using data in WorldMap. From a more user-centric perspective, Poplin, 2015 looked into issues related to the interaction with GIS-based interactive online maps. She observed that users have difficulties with the map drawing mode, and conjectured based on the observations in her study that "not-GIS skilled" users would have issues dealing with online interactive maps. Another study looking into users' perception was presented in (Schnur et al., 2018). The authors proposed a measure for assessing map complexity and reported that image clutter plays a greater role in human perception of web map complexity that symbol variability. Finally, A. C. Robinson, 2018 looked closely into social dissemination of web maps, provided a framework for defining and analyzing viral maps (i.e. maps that by some measure have been designed and shared via social media, and have been viewed by a large audience). He proposed ten dimensions along with viral maps can be analyzed, namely: purpose, audience, format, scale, location, visual variables and symbology, projection, marginalia, message context and social engagement.

10.2.2 Toolkits for geovisual analytics

Several toolkits have been proposed to support users in completing geovisual analytics tasks. As indicated in (A. C. Robinson, 2017), a key difference between geovisual analytics and geovisualization is a stronger focus of geovisual analytics on *analytical reasoning* (i.e. the process of examining information in order to find patterns within that information). The support for analytical reasoning in geovisual analytics typically comes in the form of computational methods that are used to detect patterns and/or predict future outcomes (see A. C. Robinson, 2017). The current work is primarily concerned with geovisualization support (i.e. producing a meaningful visualization for a given dataset). The computational analysis of the geographic datasets provided as input is thus kept to the minimum for the time being. Nonetheless, work on geovisual analytics will be briefly reviewed here, as the approach and the tool developed could be extended later to support geovisual analytics tasks.

[1] See `https://www.mapbox.com/colorbrewer-carto/` (last accessed: November 11, 2018).
[2] See `https://github.com/d3/d3-scale-chromatic` (last accessed: November 11, 2018).
[3] See `https://leafletjs.com/examples/choropleth/` (last accessed: November 11, 2018).

The GAV Toolkit (Van Ho et al., 2012) is a collection of Flash applications built through a combination of visualization components and linking modules that control selection, filtering, colour and animation. Datasets in GAV are loaded through a set of 'common data loaders' with a Graphical User Interface. SensePlace3 (Pezanowski et al., 2018; Savelyev and MacEachren, 2018) intends to help analysts extract insights out of geospatial social media data. SensePlace3 supports both the analysis of explicit geographical information provided by Twitter, and implicit geographical information derived from tweets by means of natural language processing and geocoding of place mentions. STempo (A. C. Robinson, Peuquet, et al., 2017) is a toolkit which includes coordinated-view geovisualization components to support visual exploration and analysis of event data. The Geoviz Toolkit (Hardisty and A. C. Robinson, 2011) is a geovisualization environment which helps users create and modify custom palettes of coordinated exploratory and analytical tools.

A distinguishing criterion of these toolkits is their targeted audience. GAV seems to take a one-size-fits-all approach: it has been evaluated with 'domain experts' from the fields of statistics, meteorology and networking, and using some of its features do not require 'IT expertise'. According to Pezanowski et al., 2018, SensePlace3 would be most useful for social scientists and nearly as useful for crisis managers and journalists. STempo was evaluated using geography and social science experts, while the Geoviz Toolkit targeted explicitly users who do not have programming expertise. The approach introduced next targets users with low Cartography *and* programming expertise.

10.2.3 Adaptive Maps

An adaptive map is "able to change its own characteristics automatically according to the user's needs" (L. T. Sarjakoski and T. Sarjakoski, 2008). Adaptive maps have a number of application areas including personalized services for tourists (L. T. Sarjakoski and T. Sarjakoski, 2008), navigational applications (L. T. Sarjakoski and T. Sarjakoski, 2008), location-based services (Reichenbacher, 2017), mobile guide systems (Reichenbacher, 2017), and mobile geospatial web services (Reichenbacher, 2017). Components of an adaptive geovisualization system were discussed in (Yingjie et al., 2001). An example of an early system endowed with adaptive functionalities for map creation is the Descartes system (G. Andrienko and N. Andrienko, 1998; G. Andrienko and N. Andrienko, 1999; N. Andrienko and G. Andrienko, 2001). Descartes leveraged knowledge of map design to offer automatic presentation of data on maps as well as facilities to interactively manipulate these maps. Descartes also provided some intelligent guidance to users (see N. Andrienko and G. Andrienko, 2001), to cope with the issue that users need training to make best use of the tool, yet are reluctant to read online help in form of text. More recently, Gould and Mackaness, 2016 have formalized cartographic knowledge for road mapping, encoding it into an ontology, with the end goal of supporting the automation of the cartographic design process. Taking a more user-centric perspective, Kiefer, Giannopoulos, Anagnostopoulos, et al., 2017 reported that users preferred having an adaptation than not having one, and regarding the adaptation type, preferred toggable adaptation (the system offers a trigger for the adaptation to the user) over revertible adaption (automatic system adaptation without user control).

Descartes was innovative for its time, but provides little support for the addition of external open datasets by users. RawGraphs (Mauri et al., 2017), a recent adaptive system, supports the exploration of existing open datasets, but still lacks dedicated functionalities for producing web maps. The issue of providing an adaptive system, which creates meaningful thematic maps and supports the addition of external open geographic datasets by novice users, is tackled in the current work.

10.3 Research Method

As mentioned in Section 10.1, this work intends to advance the state of the art along two complementary axes: (i) enable novice users to re-use an existing thematic map visualization to explore their own dataset; and (ii) offer novice users suggestions of (meaningful) thematic map visualizations which can be created based on an input geographic dataset. Realizing these two features necessitates a recognition of the type of the dataset at hand, and a set of rules to propose visualizations relevant to this dataset. In a recent vision paper on intelligent geovisualizations for open government data, Degbelo and Kray, 2018 discussed five features of the next generation of the next generation of geovisualizations: (i) automatic geographic data type recognition, (ii) generation of geovisualisation designs, (iii) monitoring of users' understanding of geographic facts, (iv) self-optimization, and (v) user activity recognition. Degbelo and Kray, 2018 also indicated that "automatic geographic data type recognition" is a human-computation problem. A literature review by Quinn and Bederson, 2011 indicated that there is a consensus as to what constitutes a human-computation problem: (a) the problem fits the general paradigm of computation (and thus might someday be solvable by computers); and (b) the human participation is directed by the computational system or process. In essence, (a) and (b) point at the need for *collaboration* of both human and computers towards the completion of a human-computation task. Hence, a semi-automatic approach is adopted in the current work. It relies on the *user* specifying the type of dataset which is at hand (e.g. nominal, interval or ratio), and the *computer* (using existing heuristics mapping between dataset type and visualization type) to generate a set of relevant thematic maps for a given dataset. The two key theoretical concepts involved here are those of *scale of measurement* from (Stevens, 1946) and *visual* *variables* suggested originally by Bertin, 1983 (and reviewed recently in Roth, 2017). Stevens's four scales of measurement (nominal, ordinal, interval, ratio) have a strong similarity to Bertin's levels of organization of information components (qualitative, ordered, quantitative-interval, quantitative-ratio). The two concepts of information component organization and visual variables have proven remarkably robust to changes in technology for maps and related graphics (see MacEachren, 2018), and provide therefore a solid theoretical underpinning for the approach of this work.

The work followed three main steps: first, conduct a survey to get a better understanding of the *use* of thematic maps in the context of urban data visualization; second build a prototype called *AdaptiveMaps*, which realizes the semi-automatic approach to thematic map creation described just above; and third evaluate the prototype to get feedback on its usability and usefulness, as well as general qualitative feedback from novice users (i.e. users without training in Cartography). The rest of the current section presents results of the map survey as well as the inner workings of *AdaptiveMaps*. Section 10.4 presents the results of the user study as well as the feedback gathered from the participants.

10.3.1 Recurrent thematic maps for urban data visualization

The purpose of the survey was to identify the kind of thematic maps most frequently used in the context of urban data visualization 40 different thematic map visualizations were collected using multiple online resources including GIS blogs, social network pages, and several other resources. Some of the GIS blogs include GISGeography[4], Geoawesomeness[5] and CARTO Blog[6]. Social network pages included ESRI Story Maps[7] and other sources include DataMade[8], Visualizing Cities[9], CityLab[10] and Data-Smart City Solutions[11]. At the beginning of the survey, there was no strict limit on the maximum number of maps to be surveyed. The surveyed ended at 40 maps because *data saturation*[12] was already reached at this point. Table 10.1 presents the results of the survey. The categorization of the kind of maps surveyed was done mainly by the second author. To control for inter-rater reliability the first author did the same categorization using 38 maps (two maps were no longer accessible online shortly after the end of the study). The inter-rater reliability was Kappa = 0.79 (p <

[4] `https://gisgeography.com/` (last accessed: October 14, 2018).
[5] `http://geoawesomeness.com/` (last accessed: October 14, 2018).
[6] `https://carto.com/blog/` (last accessed: October 14, 2018).
[7] `https://storymaps.arcgis.com/` (last accessed: October 14, 2018).
[8] `https://datamade.us/` (last accessed: October 14, 2018).
[9] `https://cityvis.io/` (last accessed: October 14, 2018).
[10] `https://www.citylab.com/` (last accessed: October 14, 2018).
[11] `https://datasmart.ash.harvard.edu/` (last accessed: October 14, 2018).
[12] Data saturation is primarily a matter of identifying redundancy in the data (see e.g. Saunders et al., 2018), and means here that a point was reached where the same type of maps were encountered over and over again during the survey.

Map Type	Frequency	Geometry	Data scales	Visual Variables
Choropleth Map	16	Polygon	Nominal, Ratio	Color Hue, Color Value
Graduated Circle Map	12	Point/Polygon	Nominal, Ratio	Color Hue/Value, Size
Dot Map	8	Point	Nominal, Interval	Color Hue, Color Value
Heat Map	3	Point	Ratio	Color Hue, Color Value
Isochrone Map	2	Polygon	Ratio	Color Hue, Color Value
Pie Chart Map	2	Point/Polygon	Ratio	Color Hue, Size

Tab. 10.1: Outcome of the short survey of thematic map visualizations (the total does not sum up to 40 because some maps may be classified as belonging to more than one category).

0.05), 95% Confidence Interval [0.63, 0.94]. Kappa values between 0.61 and 0.8 can be regarded as 'substantial' (Landis and Koch, 1977). The visual variable of location is present in all visualizations (see e.g. Bertin, 1983; Roth, 2017), and is therefore not explicitly mentioned in the table. In addition, there are three possible variables associated with color: color hue, color value and color saturation. Color saturation was not explicitly considered because its assessment with the human eye only may be error prone (also "its use as visual variable in its own right is uncommon" (Griffin, 2017)). The full list of the maps surveyed, along with their brief description and the definition of the types of maps is presented in Appendix 10.7.

10.3.2 Prototype: AdaptiveMaps

As Table 10.1 suggests, the generation of many thematic maps currently available on the web can be reduced to a manageable number of dimensions. They use either point or polygon as geometry, mostly depict nominal or ratio data[13], and rely on three visual variables, namely color hue, color value and size. Based on these results, an algorithm was developed to produce example thematic maps. It relies on two inputs: type of geometry and type of data. At the moment, the algorithm accounts for five of the six map types identified. Generating isochrone maps has been deferred to future work since one would need an additional input from the user, namely that the polygons are isochrones.

The requirements specification of the prototype are presented in Appendix 10.7. At the moment of this writing, all functional and non-functional requirements with high priority have been implemented in a tool called AdaptiveMaps. Providing the tool as a web application was one of the key objectives of this work. The application supports the uploading of datasets in GeoJON formats. A number

[13]Interval data was only present once in the maps surveyed, see Appendix 10.7.

Algorithm 1: Algorithm to generate the thematic maps.

```
Data: Geometry type, Data scale type
Result: Maps which can help meaningfully visualize the current data

if geometry = polygon then
    if datascale = nominal OR datascale = ratio then
        choropleth map;
    end
    ;
    if datascale = ratio then
        pie chart map;
    end
end
;
if geometry = point then
    if datascale = nominal OR datascale = ratio then
        graduated circle map;
    end
    ;
    if datascale = nominal OR datascale = interval OR datascale = ratio then
        dot map;
    end
    ;
    heat map;
end
```

of technologies were reviewed and the Angular framework[14] was selected for the implementation task. Angular was selected mainly because it supports the reusability of "components". This makes application development much easier for applications, which require to reuse a parts (a.k.a. reusable components) of a user interface multiple. Angular also enables to create applications using the Model-View-Controller (MVC) and Model-View-ViewModel (MVVM) design patterns. The main advantage of using these patterns is that application development becomes much more flexible. Leaflet, along with some of its plugins such as leaflet-dvf (0.3.1), leaflet-omnivore (0.3.4), and leaflet.heat (0.2.0), was used as library supporting the creation of maps. The open source code of the application is available on GitHub (`https://github.com/saadsarfrazz/AdaptiveMaps`).

AdaptiveMaps was inspired by RawGraphs (Mauri et al., 2017). It is implemented as a client-side application, does not send any data to a server, and does all the processing locally in the user's browser. It does not follow the map creation process used by many GIS software, but instead introduces a semi-automatic approach to create thematic map visualizations. The semi-automation recommends the correct outcome visualizations, provided the input data scales are provided correctly by the user. The default behavior of AdaptiveMaps does not allow to create a wrong visualization, because all numerical data are classified as interval data in the first instance. This has the drawback that the user may miss some visualizations which can only be created with ratio data (e.g. pie chart map), but the choice of prioritizing quality of results over number of visualizations was made keeping in mind that novice users might not always get the right classification of data scales. Finally, a user participating in Graves and Hendler's survey on open government data visualizations indicated: "I want a tool where I can create a visualization in, say, no more than 6 clicks". This was taken into account while designing the interaction, and the tool enables the creation of visualizations in 5-6 clicks. Figure 10.1 shows a screenshot of the application.

10.4 User study

The purpose of the study was to evaluate the applicability of the semi-automatic approach for thematic map creation by novice users. This was done through a controlled experiment in which participants were asked to perform a variety of tasks (eight in total) in three different activities. Four dependent variables were measured: *efficiency* (time taken by a participant to complete a task), *effectiveness* (participant's success rate in creating valid visualizations), *understandability* (ability of a participant to get a right answer to spatial questions after a visualization has been created), and *usability* (assessed through the System Usability Scale, see (Brooke,

[14]`https://angularjs.org/` (last accessed: November 1, 2018).

Fig. 10.1: AdaptiveMaps: The image shows the types of visualizations which can be created by the tool. The green checkbox below a visualization indicates that the visualization can be created for given dataset.

2013; Sauro, 2013; Lewis, 2018)). The study was approved by the institutional ethics board.

10.4.1 Procedure

Participants started the experiment by filling in a form to provide background information about themselves. Afterwards, they were shown a short demonstration video of the tool. The video is available on Youtube at `https://www.youtube.com/watch?v=CWVpYJK071I`. The video covered the following points: (1) upload process of new datasets; (2) data scale mapping and how one can identify data-scales; (3) what happens if data scales are not properly mapped; (4) which visualizations can be created for the uploaded dataset (checkbox on visualization); and (5) how the mapping between data dimensions and visual variables works. Afterwards the participants were invited to perform a series of visualization creation tasks. Participants' interaction with the tool were recorded using Open Broadcaster Software[15]. Once they were done with the visualization creation tasks, they were asked to report on the overall usability of the system using the system usability scale (SUS), and provide some qualitative feedback on the difficulties experienced as well as possible suggestions/recommendations regarding the tool. The qualitative feedback was audio-recorded. Participants were then rewarded 10 Euros as compensation for their time and discharged.

[15] `https://obsproject.com/` (last accessed: November 1, 2018).

10.4.2 Activities and tasks

As said above, the study involved three activities. An activity consisted of several visualization creation tasks, which can be completed by uploading the same data. All visualizations to be created in an activity were always of same type (e.g. a choropleth map or a dot map). Hence, each activity involved creating only one map type, but using different data attributes. Activity 1 and 2 intended to simulate the situation where novice users know the type of visualization they want to create (e.g. they have seen it somewhere on the web and like it), and try to use the system to build a similar visualization to the one they have already seen. During Activity 1, participants were shown a dot map displaying Airbnb data of Berlin (Germany)[16], and asked to create a 'similar map' using Airbnb data of Antwerp (Belgium). During Activity 2, the participants were shown a choropleth map about the 2017 German elections[17], and given referendum data from Scotland to create a 'similar map'. Activity 3 had a different goal than the previous two activities, namely to simulate the situation where the user has a dataset, but no idea of the geovisualization(s) which could be appropriate to explore it. Accordingly, no visual hint was provided to the participants in this activity. They were asked to create a visualization based on noise data from Hamburg (Germany), so that they can answer a given spatial question. Activity 1 had four tasks, Activity 2 had three tasks, and Activity 3 had one task. All tasks were visualization creation tasks (the user had to create a visualization in such a way that it enables her to provide an answer to a given spatial question). A detailed description of the tasks within each activity is presented in Appendix 10.7.

10.4.3 Pilot study

A pilot study was performed before formally starting the study to find out possible loopholes in the execution of a study. Five participants participated in the pilot study. The major takeaway from the pilot study was that some of the participants did not know the general cardinal directions concept on a map e.g., which direction is north or east direction on the map? This could have affected the results of the study substantially, given that some of the 'understandability questions' required basic knowledge about cardinal directions (see Appendix 10.7). To cope with this, a picture about cardinal directions was added to the instructions (see Appendix 10.7.3). The rationale for this was that the focus of the study was not to measure the acquaintance of novice users with cardinal directions concepts in general, but rather their understanding of basic spatial facts depicted on the map.

[16] `http://insideairbnb.com/berlin/` (last accessed: November 1, 2018).
[17] `https://www.theguardian.com/world/ng-interactive/2017/sep/24/german-elections-2017-latest-results-live-merkel-bundestag-afd` (last accessed: November 1, 2018).

10.4.4 Participants

A total of 20 users participated in the study, seventy percent of which were male and thirty percent were female. The participants had between 20 and 29 years of age. Fifty percent of the participants have completed bachelor studies, whereas thirty percent had completed high school. Five percent of the participants have completed the master studies whereas ten percent had a doctoral degree. The participants were from a variety of academic background including information systems, law, medicine, arts and history, linguistics and engineering. All participants were non-native English speakers and not all of the participants had English as language of the instruction in their studies. Hence, their ability to speak and comprehend everything (the study was conducted in English) varied a lot. Most participants were students at the university of Münster. Below are some further key figures about the participants:

Forty percent of the participants answered that they do not feel confident in reading maps whereas twenty-five percent of people agreed that they feel confident in reading maps. Thirty-five percent were not so sure and answered as neutral (Figure 10.2a);

Sixty-five percent of participants strongly agreed that they have never used a geodata to create a map visualization whereas thirty-five percent of the participants had some experience in creating a visualization using geodata (Figure 10.2b);

Hundred percent of the participants had not studied cartography in university or in college. Hence all the participants have no theoretical background in cartography Figure 10.2c);

Ninety-five percents of participants did not know the difference between a choropleth map and a graduated circle map. Five percent of the participants were not so sure and answered as neutral;

Ninety percent of the people did not know about any tool that can be used to create thematic map visualizations like a choropleth map or dot map. Ten percent knew about a tool which can be used to create thematic map visualizations. Tools that they reported to know were “D3JS" and “Kibana" Figure 10.2d); .

Finally, participants were also asked about their experience in the creation of visualizations in general. Forty-five percent had some experience in creating data visualizations using tools like Excel sheets whereas fifty percent didn’t.

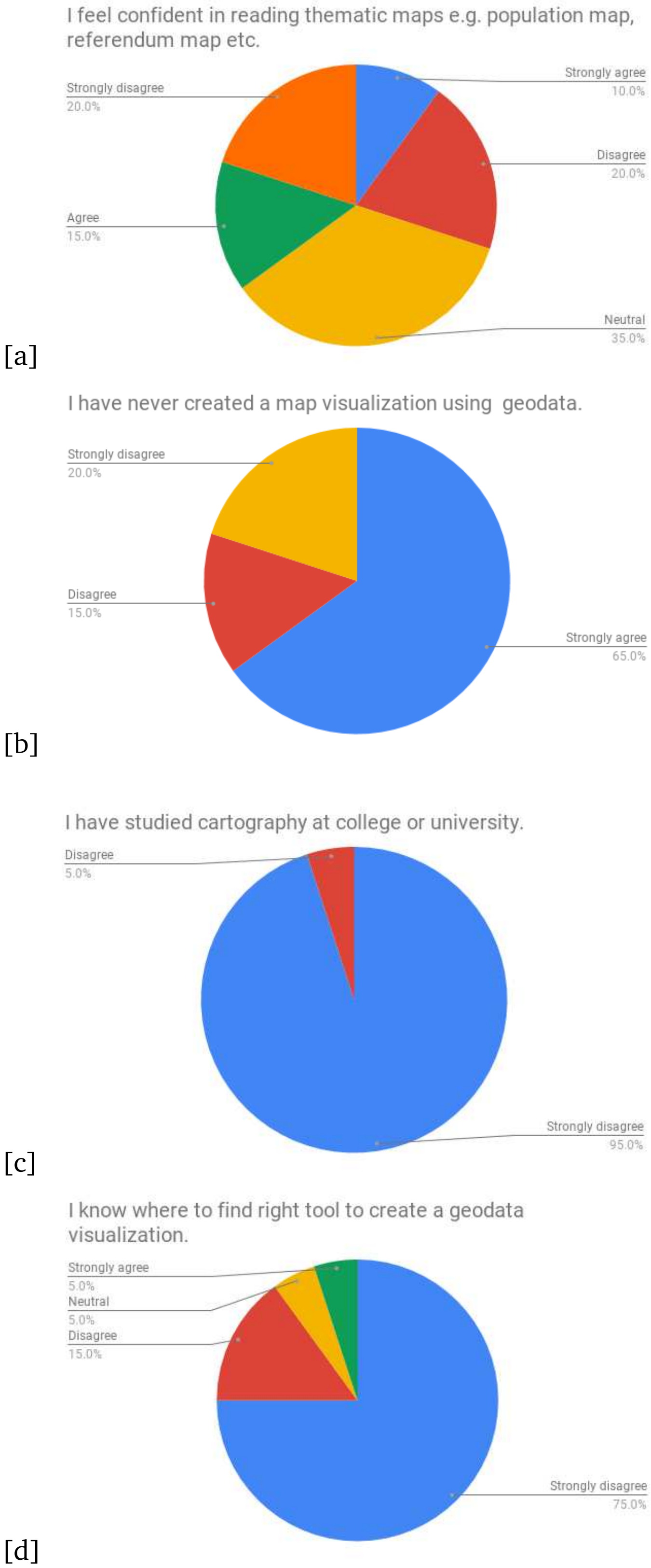


Fig. 10.2: Background of the participants.

10.5 Results of the user study

As mentioned above, 20 participants took part in the user study. One of the video recording files got corrupted during the experiment and could not be recovered. Therefore the data from that participant was dropped from the analysis. The results presented in this section summarize the insights got from analyzing data about/from the remaining 19 participants.

10.5.1 Effectiveness of AdaptiveMaps

Most of the participants managed to create web maps during the activities. The average success rate over all the tree activities was 90%. Figure 10.3 shows the distribution of these percentages across each of the tasks. The success rate was reduced during Activity 3 when participants were no longer provided visual hints, but was still relatively high in that activity (78%).

At the beginning of each activity, the users needed to select the data scales of the attributes to be visualized. The average assignment accuracy were 61% (Activity 1), 44% (Activity 2), and 58% (Activity 3) respectively. As the datasets had many attributes for which the user should inform about the type of data scale, accuracy here denotes the number of attributes for which the data scale was correct, in proportion to the total number of attributes. Figure 10.4 shows the distribution of accuracy values per activities, and across the number of participants (e.g. five participants did not manage to select any data scale correctly in Activity 1; six participants had a accuracy rate between 30 and 40 percent during Activity 3; and so on). Here, the key takeaway is that overall, the performance of the users varied quite widely with respect to the data scale assignment task.

Finally, R (R Core Team, 2016) and lme4[18] were used to perform a linear mixed effects analysis of the relationship between activity and accuracy. 'Activity' was used as fixed effect, and 'participants' were included as random effect in the model. No significance relationship was found between type of activity and scale assignment accuracy.

10.5.2 Efficiency of AdaptiveMaps

Participants took in average about four minutes (228 seconds) in Activity 1, about three minutes (166 seconds) in Activity 2, and about a minuted and a half (78 seconds) in Activity 3 to create the first visualization. Of these, a non-negligible

[18] `https://github.com/lme4/lme4/` (last accessed: November 2nd, 2018).

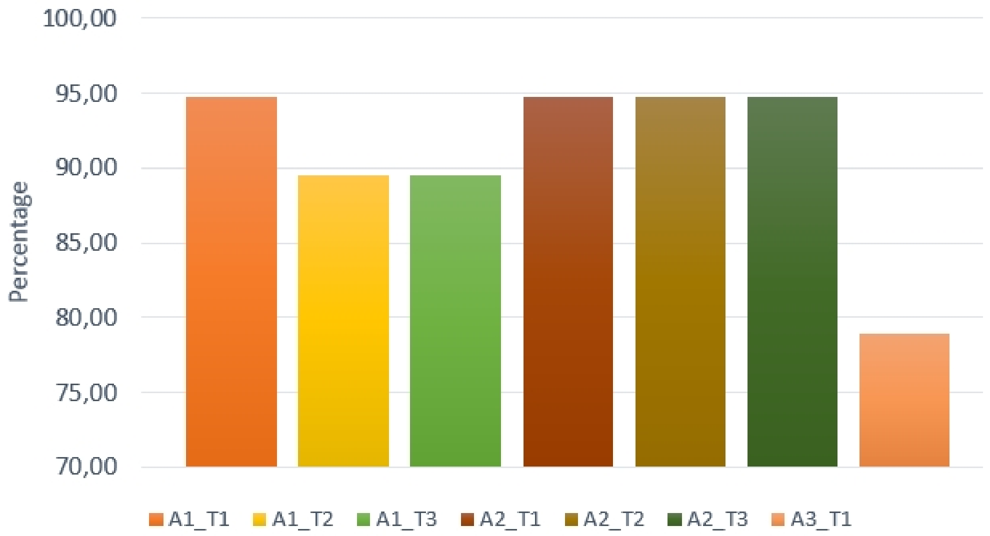


Fig. 10.3: Percentage of success in creating the web maps - AX indicates the number of the activity (Activity 1, 2 or 3), and QX indicates the question within a specific activity.

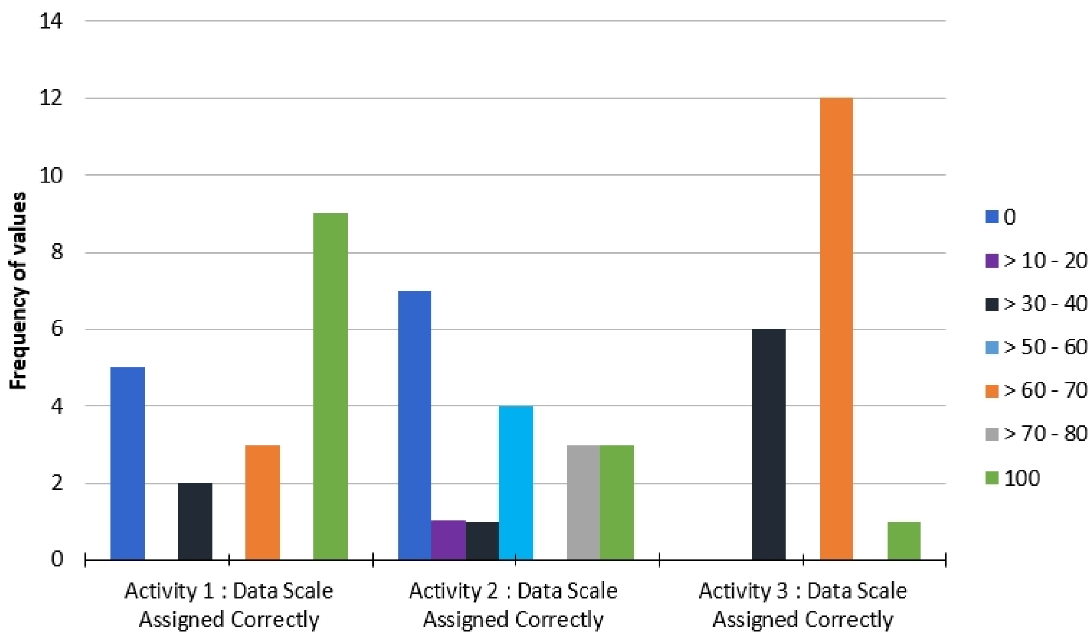


Fig. 10.4: Data scale assignment accuracy over the three activities.

amount of time went into figuring out the appropriate data scale for the attributes. An average time of 140s (62% of the total), 57s (34%) and 46s (59%) were taken for data scale assignment during Activity 1, Activity 2 and Activity 3 respectively.

R and lme4 were also used here to perform a linear mixed effects analysis of the relationship between activity and data scale assignment time. 'Activity' was used as fixed effect, and 'participants' were included as random effect in the model. The model was significant (χ^2 (2) = 29.696 $p<0.001$), confirming learning effects from the participants between the different activities (the time taken by participants for data scale assignment for each activity was reduced significantly as the experiment progressed).

Finally, the frequency of upload of datasets during the experiment was analyzed. Activities 1, 2 and 3 had an average data upload frequency of 1.4, 3.2 and 2.1 respectively. The variation of these frequencies may be attributed to the nature of the activities. Activity 1 required a participant to create a dot map. A dot map can be created by using both interval and ratio data scales. Therefore even if a participant did not classify the data scales correctly, she was still able to create visualizations. Activity 3 (creation of a heatmap) had a similar setup. However, Activity 2 required a participant to create a choropleth map using ratio data scales. Since interval data scale is assigned by default in AdaptiveMaps, the participants who assigned the wrong data scale (possibly because of lower attention or difficulty to understand the difference between interval and ratio) had to upload the data multiple times to find out what was expected from them in this activity. Thus participants having a higher upload frequency in second activity suggests a poor understanding of data scales. No all participants had issue though. Some of the users managed to upload the data only once, with high accuracy, as Figure 10.5 illustrates. Pearson correlation coefficients indicate no linear relationship between data scale assignment accuracy and data upload frequency in Activity 1 (r = 0.07), a moderate negative correlation in Activity 2 (r = -0.5), and a moderate positive correlation in Activity 3 (r = 0.46). A linear mixed effects analysis of the relationship between activity and data upload frequency revealed a significant effect of activity on data upload frequency (χ^2 (2) = 18.82 $p<0.001$).

10.5.3 Overall usability of AdaptiveMaps

The average SUS value for the AdaptiveMaps based on participants' filling in of the form was 68. 68 is the average score for many products tested Sauro, 2013, and means (on the scale provided by Bangor et al., 2008; Bangor et al., 2009) that the participants rated the prototype as "Good". Figure 10.6 presents the distribution of the SUS values accross participants. To ease readability and interpretation, the

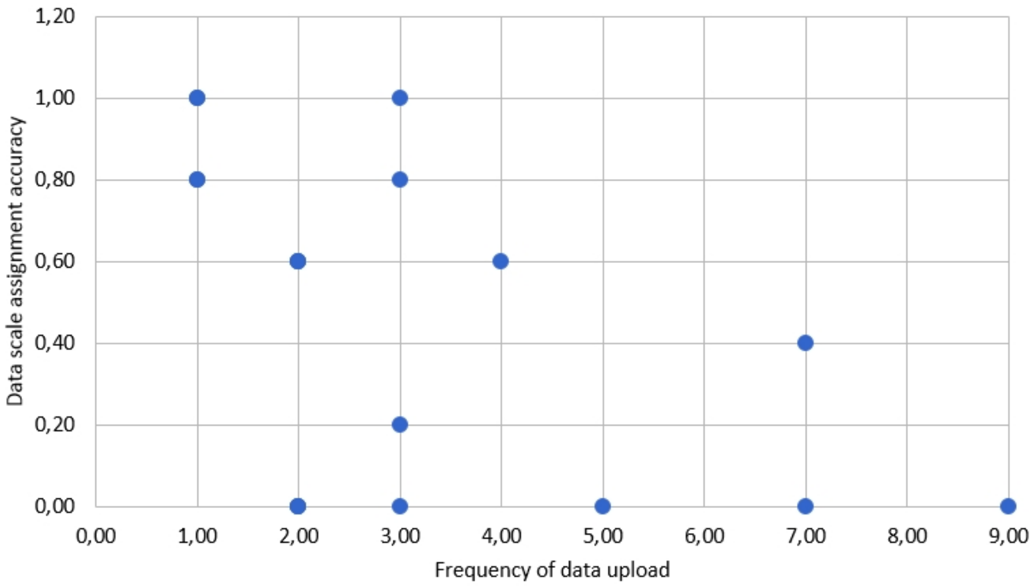


Fig. 10.5: Number of times the dataset was uploaded and the corresponding accuracies achieved by participants in Activity 2.

distribution has been created using the thresholds for SUS acceptability ranges proposed in (Bangor et al., 2008; Bangor et al., 2009): for instance products having a score between 0 and 25 have a rating of 'worst imaginable', a score between 26 and 39 suggests that the product is "Poor", a score between 40 and 52 indicates that the usability of the product is "OK" (... and so on, see Bangor et al., 2008; Bangor et al., 2009, for the full details).

10.5.4 Understandability of the visualizations

An important aspect of spatial visualizations for novice users is their ability to comprehend what is being visualized in the map. A thematic map visualization can be useless if a novice user does not understand it. In order to evaluate the users' understandability, participants were asked to answer some simple questions regarding spatial patterns in the visualizations. All questions were spatial in nature, except Q1, Activity 1 (see Appendix 10.7). Figure 10.7 shows the performance of participants with respect to answering the questions. Most participants seemed, not only to have been able to create visualizations with the tool, but also to have been 'empowered' to 'consume' geographic facts.

10.5.5 The impact of previous visualization experience

As Ottley et al., 2015 indicated "very little is known how users actually use visualizations to solve problems and even less is known about how individual differences

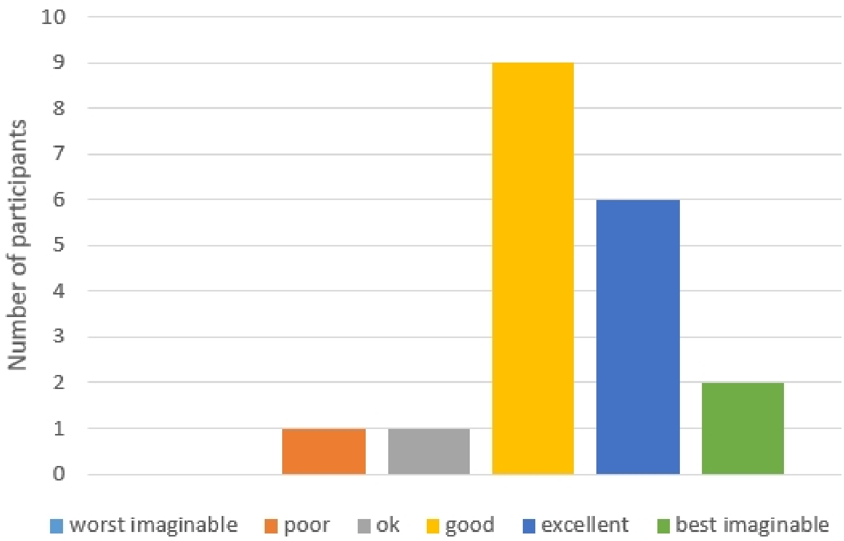


Fig. 10.6: Histogram of SUS ratings from the participants, mapped to the adjectives from (Bangor et al., 2008; Bangor et al., 2009).

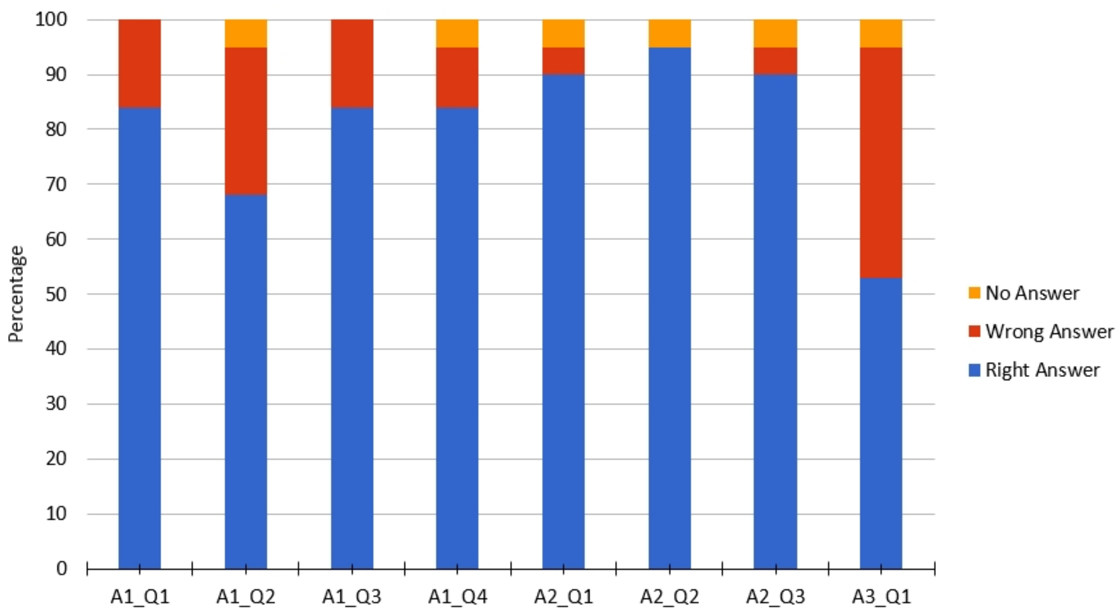


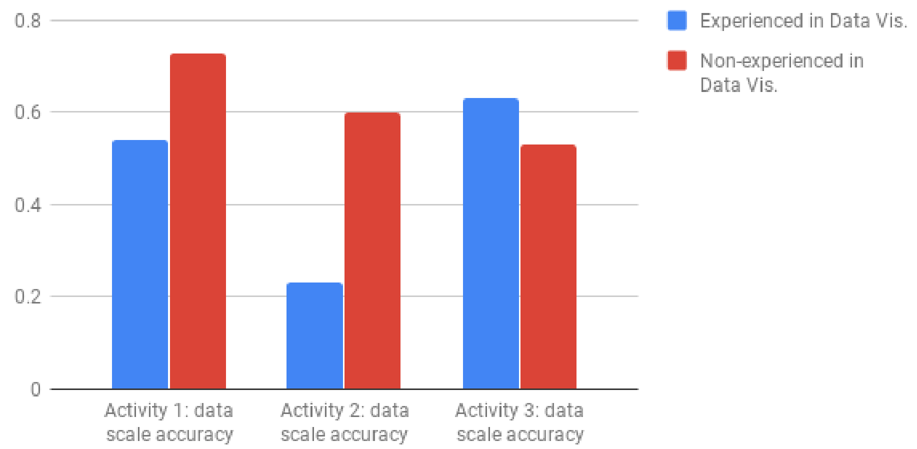
Fig. 10.7: Performance of participants in answering spatial questions - AX indicates the number of the activity (Activity 1, 2 or 3), and QX indicates the question within a specific activity.

affect these problem-solving strategies". This is true not only for visualizations in general, but for interactive maps and geovisualizations also. To shed some light on possible within-group differences between experienced and non-experienced users in visualization, a linear mixed effects analysis of the relationship between participant background and time to create the first visualization, participant background and data upload frequency, as well as participant background and data scale assignment accuracy, was performed. Data from participants who did not create a visualization in at least one of the activities was removed from the analysis. The intercepts for subjects were entered as random effects, and the models without interaction and with interaction were compared (for a detailed description of linear mixed models analysis, see for example Winter, 2013). The interactions between participant background and time to create the first visualization were not significant. However, participant background had a significant effect on data upload frequency (surprisingly, the non-experienced had a much lower upload frequency, χ^2 (2) = 9.90 p=0.007). Likewise, participant background had a significant effect on data scale accuracy assignment (surprisingly, the non-experienced had a much higher accuracy, χ^2 (2) = 6.42 p=0.04). Figure 10.8 illustrates that overall, the non-experienced seemed to have been more comfortable with the tasks than the experienced.

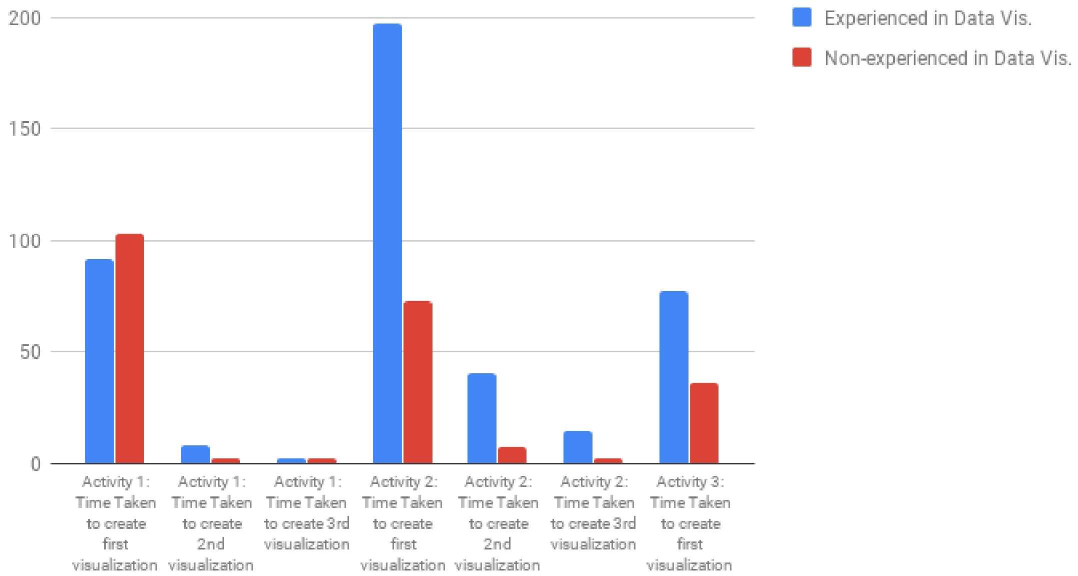
10.5.6 Participants' qualitative feedback

Participants were given the opportunity to voice their difficulties, and make some suggestions to improved the tool during the semi-structured interviews which took place right after the tasks. Most reported that rightly assigning the data scales was their main challenge. Example comments include "distinguishing between interval and ratio data was hard" (P1), "it is little bit frustrating when you have to recognize the type of data" (P6), "interval data and ratio data is somehow very confusing to me" (P2), "I am little bit confused regarding assigning the data scales. It could be beneficial if there are some concrete examples. I was confused, should I use interval or ratio" (P17), and "I was not sure how to use AdaptiveMaps and which scale and how to put items because I really did not understand. But the other part was OK" (P18).

The recommendations of the participants revolved around that aspect too. Several participants suggested to provide extra information regarding data attributes when assigning data scales. They felt that the name of an attribute is not enough to know about data. Some meta information or some additional hint (in addition to the current description already available in the tool, see Appendix 10.7.3) should be provided regarding data being assigned data scales.



[a]



[b]

Fig. 10.8: Average data assignment accuracy (a), and average time taken by each of the group to create the visualizations (b).

10.5.7 Further lessons learned on the approach

As Figure 10.3 showed, the participants' success rate in creating the visualization was very high overall. Participants were able to create visualizations with a success rate of greater than 90 percent in Activities 1 and 2. However, during the subsequent analysis of the video recordings as well as during the audio interviews, it was found that participants were in fact using trial and error strategy to create the visualizations. P17's comment during the semi-structured interview nicely summarized this. She mentioned that he was a little bit confused with the data scale concept, and when asked about the researcher how she decided about which scale to go with, replied: "I was guessing the data scales. At one point I chose interval but I did not work out, then I simply upload the data again, chose ratio and then it was OK" (P17). It has been reported in the literature that "learners ... typically start with trial-and-error approaches; they will consequently detect certain regularities concerning successful problem-solving steps, and will then develop some heuristics to proceed in a more goal-oriented fashion" (Freksa and Schultheis, 2015). The trial-and-error strategy seems to have been relied upon heavily in the study, but the fact that many participants did struggle in the third activity suggests that they may not have had enough time to develop more goal-oriented heuristics to spatial question answering with AdaptiveMaps.

10.6 Discussion

Overall, the results presented in the previous section suggest that a semi-automatic approach for thematic map creation by novice users is workable, and worth further consideration in the research on geovisualization for open data re-use. All participants in the study had no training in cartography, and the visualization creation tasks required no coding, suggesting that the approach enables users - without programming expertise and expertise in Cartography - to generate thematic web maps on top of open government data. Participants found AdaptiveMaps use-able despite few challenges, and managed not only to find their way to successfully create visualizations with it, but also answer spatial questions with it. The fact that the success rates were much higher in Activity 1 and 2 than in Activity 3 delineate the boundaries within which the approach offers higher benefits to novice users. The approach is most useful when users have seen a visualization somewhere on the web, want to re-create that same visualization for their own dataset, and achieve this goal without programming the visualizations themselves. The approach may thus be helpful to data consumers who, like many other people in (Graves and Hendler, 2014), say "I have other data and I want to create a visualization similar to this one with it".

A semi-automatic approach of the sort followed in this work may be termed 'data selection as cartography' and is a needed complement to the paradigm of 'code as cartography' (Bostock and Davies, 2013). Bostock and Davies, 2013 argued that there is no substitute for writing code, and stressed that the creation of maps through code greatly increases the cartographer's ability for self-expression. The D3 visualization library has been designed to meet that need (i.e. increased self-expression), but requires a number of programming technologies to be mastered (see a list in Murray, 2013)), and comes thus with an arguably steep learning curve. This will possibly remain so for any tool requiring coding (i.e. programming) to produce versatile visualizations. In 'data selection as cartography', leveraging cartographic knowledge and encoding it in tools which provide the users with *meaningful defaults for the dataset at hand* is the key objective. This inevitably comes with limitations in terms of choice and self-expression, but empowers users with little expertise in programming and cartography to create visualizations and answer spatial questions as AdaptiveMaps has made possible.

Good defaulting has been helpful: attributes with values as string in the geojson dataset have been automatically classified as nominal data, while numerical attributes were suggested to be either interval or ratio data, based on the survey of thematic maps conducted. The fact that interval or ordinal data appeared quite rarely during the survey was a bit unexpected, and may be a peculiarity of the sample used (e.g. visualizing the ranks of political parties after an election is an example of ordinal-level data that could be relevant for city contexts), or an indication that some types of datasets are more relevant during city discourse than others. Also surprising was the fact that people with less experience in visualization seemed to have been more comfortable with the tool overall. It may have been due to overconfidence on the side of the experienced ('It's just another visualization tool, so that should be easy'), or traced backed to the personality traits of the participants. For instance, Ottley et al., 2015 pointed out that the personality trait of 'locus of control' has been shown to consistently correlate with performance when using visualizations. They distinguished two types of people: *internals*, which tend to believe that events are influenced by their own actions, and *externals*, which are more likely to blame outside factors such as luck. Their study pointed out that internals perform better when a visualization allows them to explore the data freely and doesn't impose a strategy. Externals, on the contrary, are more efficient when a visualization provides guided/restricted exploration. There are reasons to hypothesize that the experienced users were internals and the non-experienced were externals (the tool basically supports guided exploration), but since the background questionnaire did not specifically collected information about the participants' locus of control, the why of non-experienced users performing better needs some further exploration in future work.

10.6.1 Practical significance

Graves and Hendler, 2014 derived five profiles of OGD users, based on a set of interviews: government data providers, government data consumers, researchers/journalists, civil programmers and common citizens. The approach and prototype presented above are primarily relevant to users with little expertise in programming and as such could ease the lives of government data providers, government data consumers, researchers/journalists and common citizens. Note that for the latter group, Graves and Hendler, 2014 indicated in their work that "it would not be reasonable to try to engage Common Citizens in the creation and reuse of visualizations based on OGD", since social participation is "reduced to a group of people who are really interested in a theme". The argument of Graves and Hendler is shared, but common citizens are still listed a potential beneficiaries of the approach and prototype since nothing actually prevents a user from this group to try to create a visualization in his spare time (e.g. out of sheer curiosity). In fact, since after all every person is a citizen, the term "population at large" might be better to indicate the people referred to as Common citizen by Graves and Hendler, namely anyone from the society which "may consume data via an application, a visualization or a report, but it is not directly involved with OGD". With respect to the framework for costs of open government data provision proposed in (P. A. Johnson et al., 2017), creating a geovisualization to facilitate the understanding of open data may be classified as a *direct cost* for open data providers. Geovisualizations with good defaulting will help reduce this direct cost, and the work presented in the article makes small steps towards that goal. Finally, as aptly remarked by Meng, 2018: "The role of the cartographer has evolved from map maker to maker of design tools. Map users have evolved from passive geoinformation receivers to co-creators of online maps". The approach presented in this work (along with its materialization as a tool), arguably fits this paradigm change with new roles for both stakeholders (i.e. cartographers and map users) in the digital age.

10.6.2 Cardinal directions and web map development

As mentioned in Section 10.4.3, one key observation of the pilot study was that many participants were much less knowledgeable in the concepts of cardinal directions as we assumed, i.e. participants could not easily identify cardinal directions such as north or east. This observation suggests three main things. First, that experiments involving participants with non-cartographic training and dealing with the understanding of geographic facts should explicitly watch out for their knowledge of cardinal directions *during a pilot study*. Not having done so, would have had severely (and negatively) impacted the results of the current study as mentioned above. Second, this observation points at the need for more empirical studies to assess the understandability of cardinal references on online interactive maps for the population at large. As names/annotations enable users to orient themselves on a map see e.g. Poplin, 2015, the main practical outcome of these investigations could be to find out when (and for whom) more names/annotations are needed during the interaction with online maps. Third, adding an image (see Appendix 10.7.3) to the instructions provided, showing general cardinal directions seemed to have alleviated the issue and enable the experiment run smoothly. Given that current online maps (e.g. Google Maps, Bing Maps, OpenStreetMap) provide no symbol for cardinal directions recognition (probably assuming like we did that this is a 'common sense' knowledge), it would be interesting to question this assumption, and investigate the impact of adding symbol(s) to communicate such directions on spatial learning and user experience for novice users of online maps.

10.6.3 Limitations

Despite positive user feedback validating the whole approach, the prototype is still at an early stage, and needs further improvement through additional development and usability testing rounds. Devising new strategies to facilitate the data scale selection process by users was the key takeaway from the experiment, and could be the focus of future improvements of the tool. Using the naming scheme of 'qualitative', 'quantitative-interval' and 'quantitative-ratio' (as for example in (Bertin, 1983)) and getting feedback from users could be a way of moving forward. In addition, participants made some recommendations during the semi-structured interview which indicate areas for further improvements of the tool. These include (i) providing filters on legends such that numerical data can be searched; and (ii) creating multiple maps at once, and overlaying the results.

Another limitation of the work may be found in the very fact that the whole appraoch starts with the four options provided by Stevens's levels of measurements. As discussed in (Chrisman, 1998), Stevens's four levels are a bit simplistic to account for the richness of geographic information. Chrisman, 1998 proposed to extend the four levels of measurement to ten levels to better reflect nuances in types of geographic information. Thus, using items from Chrisman's ten levels of measurement as a starting point for the framework could also be an option. Here, one should keep in mind that if choosing between three options has already proven challenging to some users, expanding the number of options may be even more challenging. That is, the ideal number of starting options (i.e. that which strikes an optimum balance between user-friendliness and thematic map creation options) remains to be identified empirically.

For now, the prototype only enables the creation of visualizations, but as already aptly noticed by Graves and Hendler, 2014, "the process of creating a visualization is [usually] followed by the subsequent action of sharing it". Thus, providing mechanisms to share newly created thematic maps on Social Media (e.g. Twitter, Facebook) could be a useful add-on.

10.7 Conclusion

This article has presented a semi-automatic approach for the creation of web maps. The approach rests on the user informing about the type of dataset at hand (i.e. nominal, interval or ratio), and the system generating meaningful interactive thematic maps for the given dataset. Despite some difficulties identifying the correct data scales, 19 participants without training in Cartography were able to create various thematic maps, and answer spatial questions after the maps have been created. The semi-automatic approach for web map creation proposed in this work thus holds promise on the roads towards (i) a better democratization of cartographic knowledge, and (ii) lowering usability barriers of existing open georeferenced to a broader population.

Observations from the experimenter and comments from participants during the user study have pointed at the necessity of strategies to facilitate data-scale identification by novice users, and this could drive immediate extensions of this work. On the long run, expanding the functionality range of the current prototype to include collaborative features, the creation of isochrones, and scalability (e.g. in order to cope with big geodata), raises interesting questions for further work.

Acknowledgements: Funding from the European Union through the GEO-C project (H2020-MSCA-ITN-2014, Grant Agreement Number 642332, `http://www.geo-c.eu/`) is gratefully acknowledged.

Appendix: 40 thematic map visualizations surveyed

Map	Description	Map Type	Data Scale
Chicago's Million Dollar Blocks (Cooper and Lugalia-Hollon, 2016)	Visualization of amount of money spent for incarnation of people per block in chicago	Choropleth Map	Ratio
Large lots (LISC Chicago and DataMade, 2018)	Displays map of lots which are sold, available or pending for sale.	Choropleth Map	Nominal
Lärmkarte Berlin 2018 (Tröger et al., 2018)	Visualization of noise surrounding buildings in berlin	Dot Map	Interval

Towards a Comparative Science of Cities (The SENSEable City Lab and Ericsson, 2014)	Visualization of spatial and temporal mobile phone activities data using maps and charts.	Choropleth Map	Ratio
In Berlin. For Berlin (Tröger et al., 2014)	Visualization of population distribution with respect to native and and non-native residents in Berlin	Choropleth Map	Ratio
Anti-Eviction Mapping Project (Anti-Eviction Mapping Project, 2018)	Vis. of "No-cause Evictions" in San Jose, California	Choropleth Map and Graduated Circle Map	Ratio
Urban Data Platform (European Commission, 2018)	Vis. of different kind of data including population density for many cities in Europe.	Graduated Circle Map	Ratio
America's growing news deserts (Columbia Journalism Review, 2017)	Vis. of US counties with daily newspapers outlets (America's news deserts)	Choropleth Map	Ratio
Participatory budgeting (Participatory Budgeting Project (PBP), 2009)	Vis. of participatory budgeting in various cities in the world. Size of the circle represents a number of participants and color represent (most recent) amount allocated	Graduated Circle Map	Ratio
Uganda Refugees (UNHCR, 2018)	Point location of refugees settlements in Uganda	Dot Map	Nominal
Inside Airbnb: Adding data to the debate (Cox, 2018)	Visualization of Airbnb room with respect to types. Additional information include income per month, nights/year etc.	Dot Map	Nominal
MuseumStat: Reach of museums in communities in the United States (Drexel University and the Institute of Museum and Library Services, 2018)	Visualization of museum locations in US.	Dot Map	Nominal
Trash City (jhaddadin, 2017)	Dot Map displaying rodent and pest complaints in the city.	Dot Map	Nominal
Population in different French muncipalities (Emc3, 2017c)	Map displaying population in different French municipalities	Graduated Circle Map	Ratio
A share of the university degree, BTS-DUT 2014 (Emc3, 2017e)	Map showing students of age greater than 25 who have not attended university in different municipalities in France.	Choropleth Map, Pie Chart Map	Ratio
Inhabitants (Emc3, 2017b)	Number of inhabitants living per kilometer square in French municipalities	Choropleth Map	Ratio
Distribution of population by level of education (Emc3, 2017a)	Visualization displaying a distribution of population by level of education in France	Graduated Circle Map	Ratio and Nominal

Presidential election outcomes (Emc3, 2017d)	Map showing 2012 presidential elections and how people voted yes and no in different regions	Graduated Circle Map	Ratio and Nominal
2018 Winter Olympics - Team USA Hometowns (ESRI, 2018)	Map displaying the resident state of US-Athletes participating in olympics 2018	Dot Map	Nominal
Celebrating Lost Loved Ones - National Safety Council (National Safety Council, 2018)	Map showing address of deceased people (died as result of opioid epidemic)	Graduated Circle Map	Ratio
Population Bubble Map (MangoMap Limited, 2017b)	Map showing location and population of capital of each US state	Graduated Circle Map	Ratio
Customer Map (MangoMap Limited, 2017a)	Heatmap based generated based on customers' addresses	Heat Map	Ratio
Sales Territories (MangoMap Limited, 2017c)	Map showing different sales territories	Choropleth Map	Nominal
Socio-Economic Data (MangoMap Limited, 2017d)	3 Different map layer showing median house income, unemployment rate and population density	Choropleth Maps	Ratio
How Much: Cost of living in different US countries (Howmuch.net, 2018a)	Map showing the true cost of living in different US cities based on money left after living at a particular place after one year	Choropleth Map	Ratio
The Working Class Can't Afford the American Dream (Howmuch.net, 2018c)	Map showing how much money is left at the end of the year for typical American working class family	Graduated Circle Map	Ratio
The Rising Costs of Sending Your Kids to a Private School (Howmuch.net, 2018b)	Map showing average high school tuition cost for US states	Choropleth Map	Ratio
Power Generation by Primary Type (Chakrabarti, 2016)	Map showing the location of electricity resources in the US and how much energy is produced by each source in different layers	Graduated Circle Map	Ratio
Building-level energy performance (Office of Sustainability (OOS), 2016)	Building-level energy performance of Philadelphia's largest commercial and multifamily buildings	Graduated Circle Map	Ratio
Geographic and temporal visualization of historical events (Azavea, 2018)	Spatio-temporal visualization of US emancipations.	Dot Map and Heat Map	Nominal and Ratio
Police Complaints (Azavea East, 2018)	Map of Complaints Against Philadelphia Police from 2013 to 2017	Choropleth Map	Ratio
PhillyTreeMap (Azavea, 2014)	Map showing trees and empty planting sites	Dot Map	Nominal
Global Heatmap (Strava, Inc., 2018)	Heat map, visualization showing people's physical activities around the globe	Heat Map	Ratio
Canada's Election Results (CartoVista, Inc., 2018a)	Map visualization of Canadian federal election results for year 2015	Pie Chart Map	Ratio

Quebec Demographics in 2016 (CartoVista, Inc., 2018b)	Map displaying quebec demographics in 2016	Choropleth Map	Ratio
McDonald's vs. Burger King Accessibility (Targomo GmbH, 2018b)	Visualization of how many customers can reach McDonald's (or Burger King) outlet given the travel type (car, train) and travel time (5 min, 15)	Isochrone Map	Ratio
How long does it take (Targomo GmbH, 2018a)	Visualizing Isochrone-map of travel access locations using a bike from single point	Isochrone Map	Ratio
Residential Broadband Providers (Federal Communications Commission, 2018b)	Map displaying number of fixed residential broadband providers	Choropleth Map	Ratio
Residential Fixed Internet Access Service Connections per 1000 Households by Census Tract (Federal Communications Commission, 2018a)	Map shows the number of residential fixed Internet access service connections per 1,000 households based on December 2016 Form 477 broadband subscribership data	Graduated Circle Map	Ratio
Residential Broadband Providers (Federal Communications Commission, 2018c)	Map displaying number of actions taken against pirate radio	Choropleth Map	Ratio

Anti-Eviction Mapping Project (2018). Anti-eviction mapping project. https://www.antievictionmap.com/san-jose-nocause-evictions/.

Azavea (2014). Phillytreemap. https://www.opentreemap.org/phillytreemap/map/.

Azavea (2018). Geographic and temporal visualization of historical events. https://www.azavea.com/projects/visualizing-emancipation/.

Azavea East (2018). Map of complaints against Philadelphia police from 2013 to 2017. https://www.azavea.com/blog/2018/01/19/map-of-complaints-against-philadelphia-police-from-2013-to-2017/.

CartoVista, Inc. (2018a). 2015 Canadian federal election results. https://cartovista.com/demos-listing/canada-federal-election-results-2015.

CartoVista, Inc. (2018b). Quebec demographics in 2016. https://cartovista.com/demos-listing/quebec-demographics-2016.

Chakrabarti, M. (2016). Power generation by primary type. https://mishmashmaps.carto.com/viz/4da91c12-7bbd-11e6-bc9f-0e233c30368f/public_map.

Columbia Journalism Review (2017). America's growing news deserts. https://cchisholm.carto.com/builder/d74181a0-3a7e-11e7-a896-0e233c30368f/embed.

Cooper, D. and Lugalia-Hollon, R. (2016). Chicago's million dollar blocks. http://chicagosmilliondollarblocks.com.

Cox, M. (2018). Inside airbnb: Adding data to the debate. http://insideairbnb.com/barcelona/.

Drexel University and the Institute of Museum and Library Services (2018). Museumstat: Reach of museums in communities in the united states. http://museumstat.org/.

Emc3, Y. L. (2017a). Distribution of population by level of education. https://france-decouverte.geoclip.fr/?lang=en#c=indicator&i=diplomeh.camembert_scol16p&i2=diplomeh.pt_pxx_a2nscol16p_sup&s=2014&s2=2014&view=map26.

Emc3, Y. L. (2017b). Inhabitants. https://france-decouverte.geoclip.fr/?lang=en#bbox=-370845,6483404,1359381,1472663&c=indicator&i=pop.dens&s=2015.

Emc3, Y. L. (2017c). Population in different french municipalities. https://france-decouverte.geoclip.fr/?lang=en#c=indicator&i=pop.pxx_pop&s=2015&view=map12.

Emc3, Y. L. (2017d). Presidential election outcomes. https://france-decouverte.geoclip.fr/?lang=en#c=indicator&i=xd_pres2012_t2.solde_hollande_sarkozy.

Emc3, Y. L. (2017e). Share of university degree, bts-dut 2014. https://france-decouverte.geoclip.fr/?lang=en#bbox=-207313,6381651,564399,611432&c=indicator&i=diplomeh.pt_pxx_a2nscol16p_sup&s=2014.

ESRI (2018). 2018 winter olympics - team usa hometowns. https://esrimedia.maps.arcgis.com/apps/......MapSeries/index.html?appid=c31dfaab1a8e4ab8badbc1a8bb9a8276&adbsc=social2136681&adbid=963124912222224384&adbpl=tw&adbpr=615300761.

European Commission (2018). Urban data platform. http://urban.jrc.ec.europa.eu.

Federal Communications Commission (2018a). FCC enforcement actions against pirate radio by location. https://www.fcc.gov/reports-research/maps/fcc-enforcement-actions-against-pirate-radio-location/.

Federal Communications Commission (2018b). Number of fixed residential broadband providers. https://broadbandmap.fcc.gov/#/.

Federal Communications Commission (2018c). Residential fixed internet access service connections per 1000 households by census tract. https://www.fcc.gov/reports-research/maps/residential-fixed......-internet-access-service-connections-per-1000-households-by-census-tract-dec-2016/.

Howmuch.net (2018a). How much: Cost of living in different us countries. https://howmuch.net/cost-of-living/ny/new-york.

Howmuch.net (2018b). The rising costs of sending your kids to a private school. https://howmuch.net/articles/average-private-school-tuition-cost.

Howmuch.net (2018c). The working class can't afford the american dream. https://howmuch.net/articles/where-the-working-class-can-afford-to-live.

jhaddadin (2017). Trash city. https://jhaddadin.github.io/trashcity/rodentpestmap.html.

LISC Chicago and DataMade (2018). Large lots. https://www.largelots.org/.

MangoMap Limited (2017a). Customer map. https://mangomap.com/mango-business/maps/57039/customer-map#.

MangoMap Limited (2017b). Population bubble map. https://mangomap.com/examples/maps/51682/bubble-map.

MangoMap Limited (2017c). Sales territories. https://mangomap.com/mango-business/maps/......57030/sales-territories#.

MangoMap Limited (2017d). Socio-economic data. https://mangomap.com/mango-business/maps/57040/socio-economic-data#.

National Safety Council (2018). Celebrating lost loved ones - national safety council. https://losttoopioids.nsc.org/index.html.

Office of Sustainability (OOS) (2016). Building-level energy performance of philadelphia's largest commercial and multifamily buildings. http://visualization.phillybuildingbenchmarking.com/#!/map.

Participatory Budgeting Project (PBP) (2009). Participatory budgeting. https://pbpmaps.carto.com/builder/45180c5a-29d4-11e7-8c77-0e05a8b3e3d7/embed.

Strava, Inc. (2018). Global heatmap. https://www.strava.com/heatmap. Targomo GmbH (2018a). How long does it take. http://sykledit.no/. Targomo GmbH (2018b). Mcdonald's vs. burger king accessibility. https://apps.route360.net/fastfood/#/.

The SENSEable City Lab and Ericsson (2014). Towards a comparative science of cities: Mobile traffic records in new york, london and hong kong. http://www.manycities.org/.

Tröger, J., Klack, M., André, P., and David, W. (2014). Wie der s-bahn-ring die stadt teilt. https://interaktiv.morgenpost.de/zugezogene-in-berlin/.

Tröger, J., Möller, C., Klack, M., David, W., and André, P. (2018). Lärmkarte berlin 2018 so laut ist es vor ihrer haustür. https://interaktiv.morgenpost.de/laermkarte-berlin/.

UNHCR (2018). Uganda refugees. https://ugandarefugees.org/analysis/settlements/.

Appendix: Requirement specifications of AdaptiveMaps

No.	Requirement	Priority	Requirement Type
1	App. shall allow the user to upload data in geojson format	High	Functional
2	App. shall disable the visualizations which can't be generated for a given uploaded dataset	High	Functional
3	App. shall allow the user to specify manually which attribute to be used to create a given visualization	High	Functional
4	App. shall allow the user to view all visualizations that can be created using this tool	High	Functional
5	App. shall allow the user to reuse uploaded data to create another visualization without uploading the data again	High	Functional
6	App. shall be available on the web	High	Functional
7	App. shall allow the user to save the visualization to view it later	Low	Functional
8	App. shall allow the user to create unique links to share the online visualization with others	Low	Functional
9	App. shall allow the user to export the visualization in a suitable format	Low	Functional
10	App. shall allow the user to explore data without extensive delays	High	Non-functional

Appendix: Tasks and questions

The study had three activities and each of the activity involved visualization creation tasks to answer a given spatial question. That is, participant were given an instruction of the form "please create a visualization to answer the following question.....". Each of the question to be answered was a multiple-choice question. Answering each question required the use of AdaptiveMaps and mapping/assigning data dimensions to visual variables to create a variety of map visualizations.

10.7.1 Activity-1

Participants were shown a visualization (see `http://insideairbnb.com/berlin/`) displaying Airbnb data of Berlin (Germany). They were then asked to do the following: "Please create a similar visualization as above for Antwerp City to answer the following question".

1. What are the most occurring room types in the city?

    a) Private room

    b) Shared room

    c) Entire home/apt

    d) I do not know

2. Which areas have more "Entire home/apt" than "private room"?

    a) East of City

    b) North-West of city

    c) South of City

    d) I do not know

3. Which places offer a minimum 16 number of beds?

    a) Almost outer parts/Outskirts of city

    b) Near City Center

c) None of the places offer 16 beds

d) I do not know

4. Which places have the highest number of reviews (number_of_reviews)?

a) South of city

b) Near center and west of city

c) East and South of city

d) I do not know

10.7.2 Activity-2

Participants were shown a visualization displaying the 2017 German elections results (see `https://www.theguardian.com/world/ng-interactive/2017/sep/24/german-elections-2017-latest-results-live-merkel-bundestag-afd`). They were then asked to do the following: "Please create a similar visualization as above for Scotland to answer the following question".

1. Which geographic areas have the highest turnout?

a) Areas in North of the Country

b) Areas in East of the Country

c) Areas in Center And West of the Country

d) I do not know

2. Which geographic areas have the maximum win difference?

a) Areas in North and South of the Country

b) Areas in West of the Country

c) Areas in East of the country

d) I do not know

3. How are the people who voted yes are geographically distributed?

   a) East of the country

   b) In the center and north of country

   c) West of the country

   d) I do not know

10.7.3 Activity-3

Participants were not provided any visual hint regarding how a final visualization may look like. They were given a point dataset with noise intensities values at different locations in Hamburg (Germany). They were then asked to do the following: “Please select an appropriate visualization to answer the following question”.

1. Where is the point with highest noise accuracy?

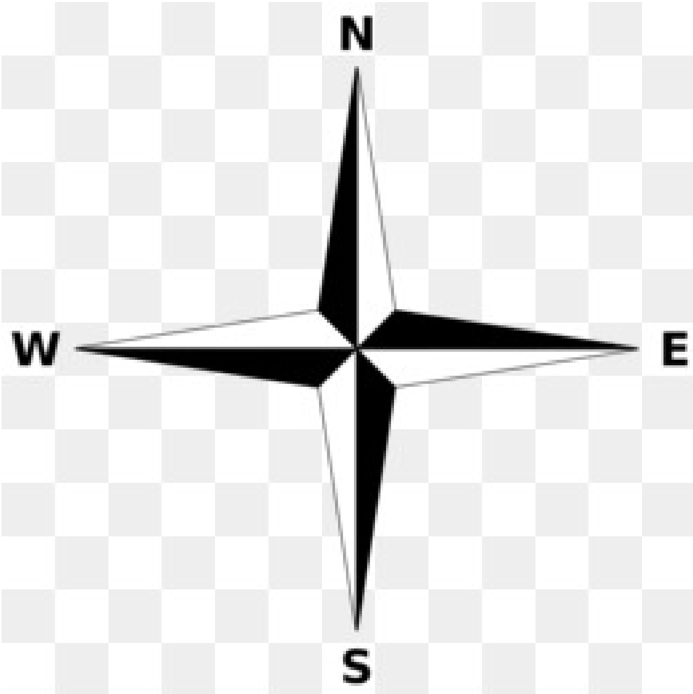
   a) East of the City

   b) In the center of the City

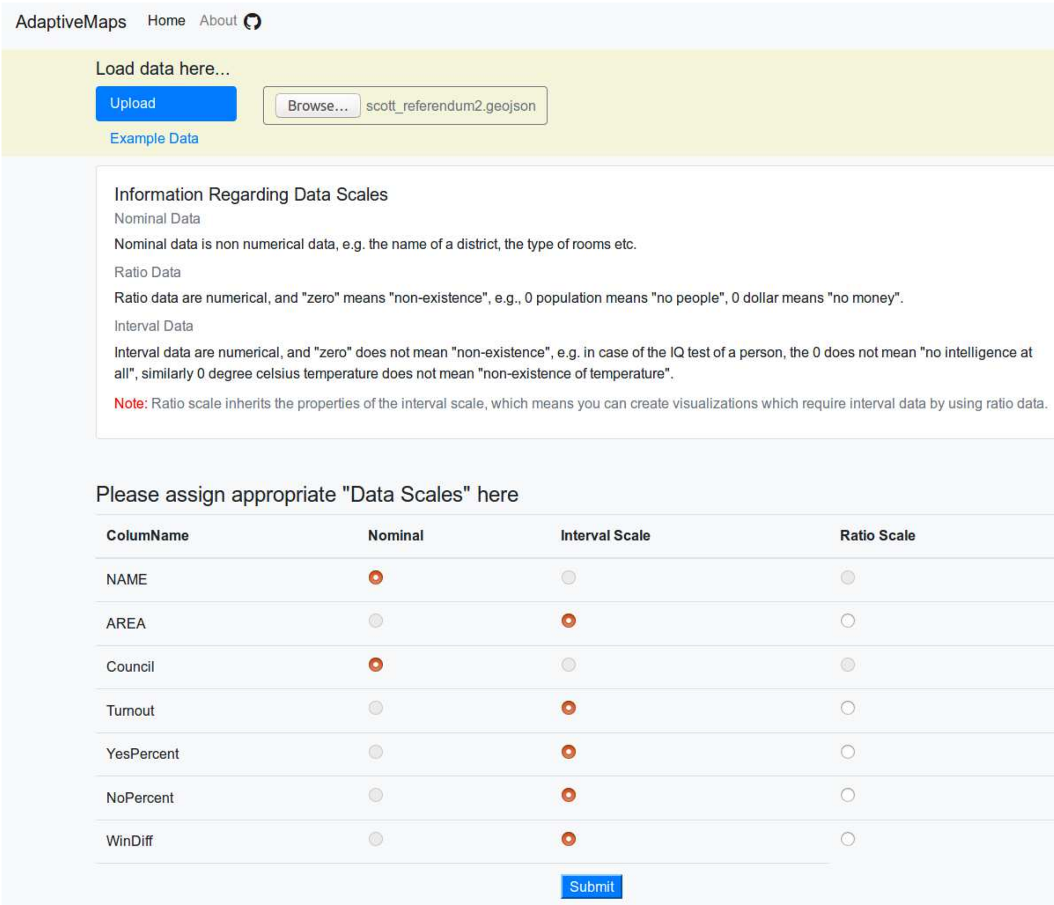
   c) North of the City

   d) I do not know

Appendix: Picture added to the instructions to help participants identify cardinal directions



Appendix: Instructions provided at the beginning of the tool (next to the demonstration video) to help users differentiate between data scales. These were kept as concise as possible to enable a quick start to the participants.



11 Discussion

Table 1.2 has presented the main contributions of the thesis. The purpose of this chapter is to comment on their generalizability, expose general limitations of the thesis, and point out directions for future research on users needs, user information, and user empowerment in the OGD context.

11.1 Generalizability

11.1.1 User needs

Chapter 2 presented six important citizen-centric challenges on the road towards open cities. Since the chapter was written with the mindset that "empowering citizens to take full advantage of available open data, is a promising way to foster innovation, creativity, and citizens-centric solutions for smart cities" (Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016), the challenges presented point in fact at needs of users to take full advantage of OGD. The needs listed were: deep participation (i.e. enabling data users to *cooperate with* governing institutions), data literate citizenry (i.e. necessary skills to understand OGD), pairing quantitative and qualitative data (i.e. methods to facilitate the analysis of diverse types of datasets), adoption of open standards (i.e. choice of data formats which are convenient to a wide spectrum of consumers), personal services (i.e. customized and focused services based on OGD[1]), and persuasive interfaces (i.e. user interfaces which encourage data consumers to change their behavior).

Chapter 3 has further presented 18 empirically-derived barriers[2] of OGD users. The users in the study were mainly researchers and developers (following the typology introduced in Chapter 1). All these barriers are a further specification of

[1] These services should also take into account privacy risks as discussed in (Degbelo, Granell, Trilles, Bhattacharya, Casteleyn, et al., 2016).

[2] Figure 3.7 lists 19 barriers, but the 'lack of higher spatial resolution data' appears twice on the figure. This reflects two slightly different perspectives on the same issue: accessibility perspective (i.e. the user rates an open data portal as unable to provide her with higher spatial resolution data) and data quality perspective (i.e. the user considers lower spatial resolution data as unfit for her purpose of local trends analysis).

the broad category of *deep participation needs* mentioned above, and point mostly at *informational needs* recapped in Table 11.1. The needs mentioned in the table are a rephrasing of the barriers presented on Figure 3.7.

There is no evidence in this work to claim that these needs apply to other groups of OGD users identified by Safarov et al. (i.e. citizens, NGOs, businesses, journalists), nor is there some evidence that these could apply to OGD initiatives in countries other than Columbia and Spain. Reliable OGD statistics to extract meaningful similarities between the two countries (i.e. go beyond the fact that both are Spanish-speaking countries) are also scarce. For instance, Colombia had a score of 64% on the Global Open Data Index in 2015, and scores 52% on the 4th edition of the Open Data Barometer (4th edition)[3]. Spain has a score of 73% on the Open Data Barometer (ODB), but scored 58% on the Global Open Data Index (GODI) in 2015. The ODB index reflects "how governments are publishing and using open data for accountability, innovation and social impact"[4], while GODI measures the state of the "openness of specific government datasets according to the Open Definition"[5]. The different purposes of the indices, and the absence of well-defined categories for countries according to their (a) OGD impact level or (b) OGD maturity states prevents any meaningful generalization on these dimensions.

Zuiderwijk, Janssen, Choenni, et al., 2012 reported on barriers to open data use, which were compiled from four workshops (totalling 71 participants) on the topic. Among the needs identified in this work, those which were also mentioned - implicitly or explicitly - in these workshops are marked with an asterisk (*) in Table 11.1. These are (the phrasings used in Zuiderwijk, Janssen, Choenni, et al.'s work are in parentheses): (i) need of contextual information about the data (No information about the provenance [context] of data); (ii) need of explanatory information about the data; (Metadata are provided in a language that the user does not understand); (iii) need of data at a 'local-level' spatial granularity (Data are not collected, especially detailed data); (iv) need of more metadata (The current metadata provision is insufficient. Especially contextual metadata is lacking); (v) need of complete datasets (Datasets are not complete); (vi) need of easy-to-understand terms of use (No explanation of the applied licences for open data); (vii) need of re-use example (No support and/or help and/or training for the use of the data is provided).

These seven needs listed in the previous paragraph seem thus potentially relevant beyond the context of this work (geographically speaking and user-wise).

[3] See `http://2015.index.okfn.org/place/`, and `https://opendatabarometer.org/4thedition/?_year=2016&indicator=ODB` (last accessed: December 24, 2018).
[4] See `https://opendatabarometer.org/4thedition/?_year=2016&indicator=ODB` (last accessed: December 24, 2018).
[5] See `http://2015.index.okfn.org/about/` (last accessed: December 24, 2018).

Tab. 11.1: Empirically-derived OGD user needs*

Category	User needs
Currency	• Need of recent data • Need of up-to-date services • Need of working (web) links
Usability	• Need of re-use examples* • Need of contextual information about the data* • Need of explanatory information about the data*
Accessibility	• Need of an easy way to download datasets • Need of data in other formats than PDF • Need of a better documentation of developers' resources • Need of data at a 'local-level' spatial granularity*
Discoverability	• Need of a more centralized access-point to datasets • Need of a better integration between different local data publishers • Need of a better organization of datasets on portals, so that their findability is eased/sped up
Data quality	• Need of consistent data attributes • Need of more metadata* • Need of complete datasets (i.e. without gaps)* • Need of data at a 'local-level' spatial granularity*
Terms of use	• Need of easy-to-understand terms of use* • Need of less restrictions on available open data

*An asterisk (*) indicates that the need was also mentioned in the workshops from (Zuiderwijk, Janssen, Choenni, et al., 2012).

11.1.2 User information

Chapters 4, 5 and 6 have presented the OCT Transparency tool, which was designed to address the current lack of monitoring possibilities for OGD usage. Gomes and Soares, 2014 proposed the measure of "dataset-to-application conversion rate"[6] (DTA) in their work to provide an impression of the level of usage of available OGD. A key drawback of DTA is that it does not inform about the relative importance of the different datasets (e.g. DTA obscures the fact that many apps can have access to one dataset). Knowing the relative importance of these different datasets is, in turn, critical to improve our understanding of OGD usage over time. The OCT transparency tool helps to address this gap. Figure 6.2 showed spatial locations from which an app connected to the OCT transparency tool (and thereby its datasets) has been assessed, demonstrating that the tool is helpful to address the *need of data at a 'local-level' spatial granularity* mentioned in Table 11.1. The OCT transparency tool is available as a Docker container on GitHub to facilitate its re-installation by interested users[7]. Finally, the categories of datasets identified in Chapter 4 are an indicator of the topics of interest in France, Germany, Spain and the United Kingdom (back in 2016), the types of questions OGD publishers assume users will ask, and ultimately the types of questions OGD users can ask. The categories are informative, but need to be revisited from time to time as the topics of interest in the countries surveyed may evolve over time. The fact that the probability of inter-country agreement between open data catalogs was found to be less than 30%, suggests that the categories identified may not apply to all OGD portals in the world (and it may be unlikely to find categories which universally apply anytime soon, unless concerted efforts at a transnational level are put into place to standardize existing OGD categories).

The second theme discussed under 'user information' in the thesis was the effectiveness of different representations in enabling transparency (Chapters 7 and 8). Linked Data and Geovisualization increase transparency as discussed in Chapter 7, but the claim is only valid when *raw data* is taken as baseline. Chapter 8 has helped to precisify the role of geovisualization in transparency enablement: in comparison to table-based representations, geovisualizations make the attractiveness of OGD more visible on the one hand, and holistic knowledge more visible on the other hand (if that knowledge is of type space-in-time comparison).

11.1.3 User empowerment

Chapter 9 has presented five requirements of intelligent geovisualizations (i.e. more flexible geovisualizations enabling Non-Cartographers to take advantage of OGD).

[6] The formula used in their work was $\frac{Number\ of applications\ developed}{Number\ of\ datasets} * 100$.

[7] `https://github.com/geo-c/OCT-Core`.

The ideas discussed are relevant to many other research areas beyond OGD as discussed in the chapter. Lessons learned while implementing intelligent geovisualizations for OGD could also be relevant while realizing the recent vision of intelligent systems for the geosciences by Gil et al., 2019. In particular, Gil et al.'s call for intelligent user interaction presents nice synergistic opportunities for the two visions.

Chapter 10 is an early step towards realizing the vision presented in chapter 9. The chapter provided a promising way of implementing the two features of automatic geographic data type recognition, and generation of geovisualisation designs. All participants in the study had no training in cartography, and the visualization creation tasks required no coding, suggesting that the approach introduced in Chapter 10 can be useful to empower OGD users - without programming expertise and expertise in Cartography - to generate geovisualizations on top of OGD.

11.2 Limitations

The OCT Transparency Tool (Chapter 6) and AdaptiveMaps (Chapter 10) have provided examples of how georeferenced OGD can be re-used. Figure 3.6 gives some initial evidence that georeferenced open data is useful to OGD users. Nevertheless, the work has remained silent throughout on how much OGD is actually georeferenced. The rationale for that is the current lack of statistics of OGD (a consequence of OGD research being still in its early days). In the absence of more accurate information, the work relied on the assumption that 60% of existing OGD is georeferenced, in line with the existing (and at the moment, only trustable) finding that 60% of existing information could be georeferenced (Hahmann and Burghardt, 2013).

Another limitation of the work is that the two tools presented earlier (Chapters 6 and 10) are not finished products: they are still prototypes which would benefit from further rounds of usability improvements. Furthermore, they have not been deployed outside the lab, suggesting that the ecological validity[8] of the conclusions presented in Section 11.1.1 still needs to be demonstrated. Finally, though both tools have shown promise, there is still a need to develop mechanisms to ensure their adoption. Arguably, adoption by OGD research and practice cannot be guaranteed at this point, but irrespective of their subsequent adoption, a merit of these tools is that they have brought to light new possibilities regarding OGD usage.

[8]Ecological validity denotes the extent to which findings can be generalized to real-life settings.

11.3 Outlook

Several issues from Table 1.1 were not touched upon during this work (e.g. data ambiguity, data fragmentation, lack of search support, understandability of licences) and could be taken up in future studies. The rest of this section presents possible extensions of the contributions made during the thesis.

11.3.1 User needs

The user needs discussed in Section 11.1.1 mainly touch on the topic of participation (as mentioned above) and transparency (since they point at information which should be made visible according to the data consumers). Collaboration, the third key pillar of OGD, has not been addressed in this thesis, and future work could investigate users' desiderata in that area. These user wishes could be gathered through self-report from OGD users (e.g. during participants' interviews) and/or derived from motivations reported in the literature about why people volunteer to participate in OGD projects (for an example, see Wijnhoven et al., 2015)[9].

11.3.2 User information

Metadata generation

Table 11.1 highlights the need for (consistent) metadata to catalyze OGD adoption on the user side. The key opportunity presented by this need is that of *automatic metadata generation*. Indeed, as Craglia, Goodchild, et al., 2008 noticed, metadata are "useful to make visible the wealth of information resources already existing and [...] allow the opportunity for re-use". Yet, "separating the metadata from the data they refer to poses challenges of synchronization in the event of change" (Craglia, Goodchild, et al., 2008). Automating metadata generation can help mitigate that issue. Automatic metadata generation (AMG) has been discussed so far in the broader context of digital curation (e.g. Dorbeva et al., 2013), but the potential and value of AMG for georeferenced OGD is largely untapped. If metadata generation cannot be fully automated, exploring the extent to which semi-automatic approaches[10] or human-computation[11] approaches to metadata generation could improve metadata provision in the OGD landscape presents also interesting possibilities for future research.

[9] The assumption here is that what motivates people to participate points indirectly at some of their wishes or needs while participating.
[10] See (Park and Brenza, 2015) for a recent survey of semi-automatic tools for metadata generation.
[11] See (Quinn and Bederson, 2011) for a survey of human-computation approaches.

Properties of communication media in the OGD landscape

Chapter 8 presented a comparison of geovisualizations and data tables for the purpose of information communication (or transparency enablement) in the OGD landscape. As acknowledged in the chapter, only two of interactive primitive objectives were considered, and more work is needed to assess the performances of these different types of representations for objectives such as rank, associate and delineate. Another type of representation used in the context of OGD and warrant more attention is dashboards. Dashboards can be of many types discussed in (Sarikaya et al., 2019), and empirical assessments of their performance in the context of OGD are still needed. Animations too could be useful to communicate OGD insights to data consumers: they are useful to show how data changes and evolves over time, and can be helpful to show other non-temporal dimensions such as age or income (see Chevalier et al., 2016). Overall, OGD research would benefit from empirically-derived insights on the respective merits of different representations for communication. If visualizations are considered, these empirical assessments could take into account the visual literacy of the users explicitly (using for example, techniques such as the one presented in (Boy et al., 2014)).

Recommender systems for OGD

The increasing availability of OGD invites new research on methods to make OGD discoverable. Investing efforts in improving metadata generation as discussed above is one approach, but is mostly suitable to the scenario of information search. An alternative to information needs satisfaction through information search, is recommendations. As discussed by Garcia-Molina et al., 2011, a *search mechanism* takes a query that describes the user's current interests as input and returns objects that match the query; a *recommendation mechanism*, by contrast, does not use explicit query. It rather analyzes the user's context, and the user profile (if available) to present to the user objects that may be of interest. There are several recommendation strategies discussed in (Jannach et al., 2016), and recommendations in the context of OGD could address a single user (e.g. an OGD developer) or a group of interested stakeholders (e.g. a group of city councils with similar preferences). Few researchers have begun experimenting with recommendation strategies suitable for different OGD (see for e.g. Cantador et al., 2018), but recommender systems for OGD need much more attention from OGD research overall. In particular, recommender systems for OGD should provide users with an understandable representation of how the system represents them, and allow users to control the recommendation process (as proposed in Calero Valdez et al., 2016; Spano and Boratto, 2019).

11.3.3 User empowerment

Geovisualizations for user empowerment

Chapter 10 has experimented with adaptive visualizations which help users create their own interactive maps on top of OGD. Nevertheless, the vision of intelligent geovisualization for OGD needs sustained years of efforts to become reality. In addition to research aiming at making geovisualizations more flexible, and democratizing geovizualization skills, user empowerment needs also geovisualizations that enable *reporting* (for example, enable OGD data consumers to report back to data publishers that some datasets tied to a certain location is inaccurate, or inconsistent). These would help solve the issue of 'lack of support to improve already opened datasets' identified by previous work, and mentioned in Table 1.1.

smart APIs and semantic APIs

The API contributions of this thesis has so far been mentioned under the these 'User Information' (Chapter 4) because the semantic API was used as a key component of the OCT transparency tool. Nonetheless, APIs could be seen more generally as a critical component of users empowerment (i.e. enabling data consumers to build their own products on top of OGD). In the context of OGD, they can play the three roles of APIs elicitated in (de Souza and Redmiles, 2009): act as contracts between OGD data publishers and consumers (i.e. telling OGD consumers what to expect from publishers), as communication mechanisms between publishers and consumers, and as boundaries (i.e. set the frame of communication between consumers and publishers). More work is needed to bring semantic APIs (i.e. APIs which enable retrieval of data according to their types) into the mainstream. More specifically, there is a need for *de facto* taxonomies of common geographic data types in OGD, and common kinds of attribute information requested by users. Next to semantic APIs which were the focus of this work, smart APIs (i.e. APIs with rich semantic annotations) have also been proposed in previous work (Zaveri et al., 2017) as an improvement over existing APIs. The main challenge addressed by smart APIs is API findability, and as the OGD ecosystem grows, smart APIs could be valuable for OGD too.

Interactive guidelines

Text is currently the main form to document products related to OGD (e.g. projects and applications). Data consumers wish to have re-use examples (see Table 11.1), but text may not be the most effective (and attractive) way to present those examples. Interactive guidelines (briefly discussed in Chapter 6) could be a novel way of documenting insights from past data usage projects, and provide data consumers with an engaging way to start their own data re-use work. As the name suggests, they are guidelines (i.e. they walk the user through a story), and they are interactive (i.e. designed to provide different outputs depending on the actions of the user).

Since interactive guidelines intend to support data consumers find their way to re-use OGD, they could have the following elements: story, problem, solution, interaction, documentation and scaling up. A *story* in this context refers to a problem-solution pattern. The *problem* is a task that is currently challenging to solve, whereas a *solution* involves a combination of OGD and software to solve the problem (i.e. improve the current situation). *Interaction*, in this context, denotes user interface elements which enable the user to move through different aspects of the story. Interaction involves testing a specific OGD re-use product and visualizing the dataset to readily get an idea of how to solve the problem by oneself. *Documentation* of impact of the solution and it's *scaling up* can give the user an idea of the portability of interactive guidelines (i.e. what to reasonably expect when it comes to using the solution in another context). The necessity of supplementing current OGD platforms with data-driven storytelling features has been acknowledged in previous work (Brolcháin et al., 2017), but clarity hasn't been reached yet as to how to best realize this. Designing, implementing and evaluating interactive guidelines for this task, could be a way of making progress in that area.

12 Conclusion

Many countries worldwide have been jumping on the open data bandwagon, opening up their datasets, and generating a wealth of open government data (OGD) ready to be exploited. There are still a number of issues preventing data consumers to take full advantage of OGD, and this works has aimed at advancing OGD research in three areas: user needs, user information, and user empowerment. The work has identified current user needs (e.g. need of re-use examples, need of more metadata, and need of data at a higher spatial granularity), provided an open-source tool to help data publishers monitor OGD data usage, brought forth strengths and weaknesses of geovisualizationsss and data tables for information provision in the OGD landscape, and introduced an approach to web map creation for users without training in cartography.

These contributions solve some of the current issues preventing OGD re-use, but OGD research is still in its infancy and more work is needed to advance OGD scholarship. Opportunities for future research on OGD are rich and exciting, with numerous promises for interdisciplinary research and society.

Bibliography

Abras, Chadia, Diane Maloney-krichmar, and Jenny Preece (2004). „User-centered design“. In: *Encyclopedia of Human-Computer Interaction*. Ed. by William Sims Bainbridge. Thousand Oaks: Sage Publications (cit. on p. 5).

Access Info Europe (2010). „Beyond Access: Open Government Data and the ’Right to Reuse’“. In: *October* September, p. 103 (cit. on p. 58).

Aenor (2015). *UNE 178301:2015. Ciudades inteligentes. Datos abiertos*. Tech. rep. (cit. on p. 120).

Afful-Dadzie, Eric and Anthony Afful-Dadzie (2017). „Open Government Data in Africa: A preference elicitation analysis of media practitioners“. In: *Government Information Quarterly* (cit. on p. 55).

Aguilera, Unai, Oscar Peña, Oscar Belmonte, and Diego López-de-Ipiña (2017). „Citizen-centric data services for smarter cities“. In: *Future Generation Computer Systems* 76, pp. 234–247 (cit. on p. 183).

Ahlers, Dirk, Frank Alexander Kraemer, Anders Eivind Braten, et al. (2018). „Analysis and visualization of urban emission measurements in smart cities“. In: *Proceedings of the 21st International Conference on Extending Database Technology (EDBT)*. Ed. by Michael H. Böhlen, Reinhard Pichler, Norman May, et al. Vienna, Austria: OpenProceedings.org, pp. 698–701 (cit. on p. 184).

Ahmadi Zeleti, Fatemeh, Adegboyega Ojo, and Edward Curry (2016). „Exploring the economic value of open government data“. In: *Government Information Quarterly* 33.3, pp. 535–551 (cit. on pp. 52, 58).

Ahn, Luis von (2009). „Human computation“. In: *Proceedings of the 46th Annual Design Automation Conference on ZZZ - DAC ’09*, pp. 418–419 (cit. on p. 211).

Akyürek, Haydar, Cora Scholl, Regina Stodden, Tobias Siebenlist, and Agnes Mainka (2018). „Maturity and usability of open data in North Rhine-Westphalia“. In: *Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age - dgo’ 18*. Ed. by Marijn Janssen, Soon Ae Chun, and Vishanth Weerakkody. Delft, The Netherlands: ACM Press, pp. 1–10 (cit. on pp. 3, 4).

Albus, James S (1991). „Outline for a theory of intelligence“. In: *IEEE Trans. on Systems, Man, and Cybernetics* 21.3, pp. 473–509 (cit. on p. 210).

Andrienko, Gennady and Natalia Andrienko (1998). „Intelligent visualization and dynamic manipulation: two complementary instruments to support data exploration with GIS“. In: *Proceedings of AVI’98*, pp. 66–75 (cit. on pp. 214, 223).

Andrienko, Gennady and Natalia Andrienko (1999). „Interactive maps for visual data exploration". In: *International Journal of Geographical Information Science* 13.4, pp. 355–374 (cit. on p. 223).

Andrienko, Natalia and Gennady Andrienko (2001). „Intelligent support for geographic data analysis and decision making in the web". In: *Journal of Geographic Information and Decision Analysis* 5.2, pp. 115–128 (cit. on pp. 215, 223).

– (2007). „Intelligent visualisation and information presentation for civil crisis management". In: *Transactions in GIS* 11.6, pp. 889–909 (cit. on p. 216).

Andrienko, Natalia, Gennady Andrienko, and Peter Gatalsky (2003). „Exploratory spatio-temporal visualization: an analytical review". In: *Journal of Visual Languages & Computing* 14.6, pp. 503–541 (cit. on p. 171).

Angelidou, M (2014). „Smart city policies: a spatial approach". In: *Cities* 41, S3–S11 (cit. on p. 23).

Arnstein, S (1969). „A ladder of citizen participation". In: *Journal of the American Institute of Planners* 35.4, pp. 216–224 (cit. on p. 5).

Attard, Judie, Fabrizio Orlandi, and Sören Auer (2016). „Value creation on open government data". In: *2016 49th Hawaii International Conference on System Sciences*. Koloa, Hawaii, USA: IEEE, pp. 2605–2614 (cit. on pp. 141, 153, 181, 220).

Attard, Judie, Fabrizio Orlandi, Simon Scerri, and Sören Auer (2015). „A systematic review of open government data initiatives". In: *Government Information Quarterly* 32.4, pp. 399–418 (cit. on pp. 3, 6, 7, 52, 58, 70, 73, 87, 120, 152, 166, 182).

Atzori, L, A Iera, G Morabito, and M Nitti (2012). „The Social Internet of Things (SIoT) - When Social Networks Meet the Internet of Things: Concept, Architecture and Network Characterization". In: *Comput. Netw.* 56.16, pp. 3594–3608 (cit. on pp. 22, 47).

Ballatore, Andrea (2014). „Defacing the map: cartographic vandalism in the digital commons". In: *The Cartographic Journal* 51.3, pp. 214–224 (cit. on p. 184).

Bangor, Aaron, Philip Kortum, and James Miller (2008). „An empirical evaluation of the system usability scale". In: *International Journal of Human-Computer Interaction* 24.6, pp. 574–594 (cit. on pp. 155, 235–237).

– (2009). „Determining what individual SUS Scores mean: Adding an adjective rating scale". In: *Journal of Usability Studies* 4.3, pp. 114–123 (cit. on pp. 155, 235–237).

Barba, C.T., M.A. Mateos, P.R. Soto, A.M. Mezher, and M.A. Igartua (2012). „Smart city for VANETs using warning messages, traffic statistics and intelligent traffic lights". In: *In Proceedings fo the IEEE Intelligent Vehicles Symposium, Alcala de Henares, Spain*, pp. 902–907 (cit. on p. 41).

Bargh, Mortaza S, Sunil Choenni, and Ronald Meijer (2016). „Meeting Open Data Halfway". In: *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance - ICEGOV '15-16*, pp. 199–206 (cit. on p. 71).

Barry, Emily and Frank Bannister (2014). „Barriers to open data release: A view from the top". In: *Information Polity* 19.1,2, pp. 129–152 (cit. on pp. 5, 58, 59, 61, 68, 73).

Barth, Lukas, Benjamin Niedermann, Martin Nöllenburg, and Darren Strash (2016). „Temporal map labeling: a new unified framework with experiments". In: *Proceedings of SIGSPATIAL 2016*, pp. 1–10 (cit. on p. 213).

Bartoschek, Thomas, Gerald Pape, Christian Kray, Jim Jones, and Tomi Kauppinen (2014). „Gestural interaction with spatiotemporal linked open data“. In: *OSGeo J.* 13.1, pp. 60–67 (cit. on pp. 40, 41).

Batty, M (1997). „Cellular automata and urban form: a primer“. In: *J. Am. Planning Assoc.* 63.2, pp. 266–274 (cit. on p. 36).

– (2013a). „Big data, smart cities and city planning“. In: *Dialogues in Hum. Geogr.* 3.3, pp. 274–279 (cit. on p. 40).

– (2013b). *The new science of cities.* Cambridge, MA, USA: MIT Press (cit. on pp. 21, 23).

Batty, M, K W Axhausen, F Giannotti, et al. (2012). „Smart cities of the future“. In: *Eur. Phys. J. Spec. Top.* 214, pp. 481–518 (cit. on p. 27).

Beilschmidt, Christian, Thomas Fober, Michael Mattig, and Bernhard Seeger (2017). „A linear-time algorithm for the aggregation and visualization of big spatial point data“. In: *SIGSPATIAL 2017*, 73:1–73:4 (cit. on p. 213).

Benington, John (2009). „Creating the public in order to create public value?“ In: *International Journal of Public Administration* 32.3-4, pp. 232–249 (cit. on pp. 141, 181).

Benitez-Paez, Fernando, Alexis Comber, Sergio Trilles, and Joaquin Huerta (2018). „Creating a conceptual framework to improve the re-usability of open geographic data in cities“. In: *Transactions in GIS* 22.3, pp. 806–822 (cit. on p. 183).

Benitez-Paez, Fernando, Auriol Degbelo, Sergio Trilles, and Joaquín Huerta (2018). „Roadblocks hindering the reuse of open geodata in Colombia and Spain: a data user's perspective“. In: *ISPRS International Journal of Geo-Information* 7.1, p. 6 (cit. on pp. 144, 153, 159, 183, 187, 220).

Beno, Martin, Kathrin Figl, Jurgen Umbrich, and Axel Polleres (2017). „Open data hopes and fears: determining the barriers of open data“. In: *2017 Conference for E-Democracy and Open Government (CeDEM)*. Ed. by Peter Parycek and Noella Edelmann. Krems, Austria: IEEE, pp. 69–81 (cit. on pp. 6, 7, 53, 58, 62, 70, 77, 84, 85, 144, 159, 183, 187).

Berners-Lee, T (2006). *Linked Data - Design issues* (cit. on pp. 100, 162).

Berners-Lee, Tim (2017). *Tim Berners-Lee: I invented the web. Here are three things we need to change to save it.* (Last accessed: January 18, 2018) (cit. on pp. 143, 187).

Bertin, Jacques (1981). *Graphics and graphic information processing (Translated by Willam J. Berg and Paul Scott)*. Berlin: Walter de Gruyer, p. 273 (cit. on pp. 190, 191).

– (1983). *Semiology of graphics: diagrams, networks, maps (Translated by William J. Berg)*. The University of Wisconsin Press, Madison (cit. on pp. 191, 212, 221, 225, 226, 243).

Bischof, Stefan, Athanasios Karapantelakis, Amit Sheth, Alessandra Mileo, and Payam Barnaghi (2014). „Semantic modelling of smart city data“. In: *W3C Workshop on the Web of Things Enablers and services for an open Web of Devices*. Berlin, Germany, pp. 1–5 (cit. on p. 120).

Bostock, Michael and Jason Davies (2013). „Code as cartography“. In: *The Cartographic Journal* 50.2, pp. 129–135 (cit. on p. 241).

Boulos Kamel, M N and N M Al-Shorbaji (2014). „On the Internet of Things, smart cities and the WHO Healthy Cities“. In: *Int. J. Health Geogr.* 13.10 (cit. on p. 22).

Boulos Kamel, M N, A D Tsouros, and A Holopainen (2015). „Social, innovative and smart cities are happy and resilient: insights from the WHO EURO 2014 International Healthy Cities Conference“. In: *Int. J. Health Geogr.* 14.1 (cit. on p. 21).

Boy, Jeremy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete (2014). „A principled way of assessing visualization literacy“. In: *IEEE Transactions on Visualization and Computer Graphics* 20.12, pp. 1963–1972 (cit. on p. 263).

Branchi, P, C Fernández-Valdivielso, and I Matias (2014). „Analysis matrix for smart cities“. In: *Future Internet* 6.1, pp. 61–75 (cit. on pp. 22, 23).

Brauer, B, M Eisel, and L M Kolbe (2015). „The state of the art in smart city research - a literature analysis on green IS solutions to foster environmental sustainability“. In: *In Proceedings of the Pacific Asia Conference on Information Systems, Singapore, Singapore* (cit. on pp. 24, 25, 41).

Brauneis, Robert and Ellen P. Goodman (2018). „Algorithmic transparency for the smart city“. In: *Yale Journal of Law & Technology* (cit. on pp. 168, 215).

Bresciani, Sabrina and Martin J. Eppler (2015). „The pitfalls of visual representations“. In: *SAGE Open* 5.4, pp. 1–14 (cit. on p. 212).

Brewer, Cynthia A. (2003). „A transition in improving maps: the ColorBrewer example“. In: *Cartography and Geographic Information Science* 30.2, pp. 159–162 (cit. on p. 222).

Brolcháin, Niall Ó, Lukasz Porwol, Adegboyega Ojo, et al. (2017). „Extending open data platforms with storytelling features“. In: *Proceedings of the 18th Annual International Conference on Digital Government Research - dg.o’ 17*. Ed. by Charles C. Hinnant and Adegboyega Ojo. Staten Island, NY, USA: ACM Press, pp. 48–53 (cit. on p. 265).

Brooke, John (2013). „SUS: A Retrospective“. In: *Journal of Usability Studies* 8.2, pp. 29–40 (cit. on pp. 155, 228).

Bröring, A, J Echterhoff, S Jirka, et al. (2011). „New generation sensor web enablement“. In: *Sensors* 11.3, pp. 2652–2699 (cit. on p. 38).

Brunetti, Josep Maria, Sören Auer, and Roberto García (2012). „The linked data visualization model“. In: *Proceedings of the ISWC 2012 Posters & Demonstrations Track*. Ed. by Birte Glimm and David Huynh. Boston, Massachusetts, USA: CEUR-WS.org (cit. on pp. 140, 180).

Brunetti, Josep Maria, Sören Auer, Roberto García, Jakub Klímek, and Martin Nečaský (2013). „Formal linked data visualization model“. In: *Proceedings of International Conference on Information Integration and Web-based Applications & Services - IIWAS ’13*. Ed. by Edgar R. Weippl, Maria Indrawan-Santiago, Matthias Steinbauer, Gabriele Kotsis, and Ismail Khalil. Vienna, Austria: ACM Press, pp. 309–318 (cit. on p. 180).

Bugs, G, Carlos Granell, O Fonts, Joaquín Huerta, and Marco Painho (2010). „An assessment of Public Participation GIS and Web 2.0 technologies in urban planning practice in Canela, Brazil“. In: *Cities* 27.3, pp. 172–181 (cit. on p. 34).

Calero Valdez, André, Martina Ziefle, and Katrien Verbert (2016). „HCI for recommender systems: the past, the present and the future“. In: *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys ’16*. Ed. by Shilad Sen, Werner Geyer, Jill Freyne, and Pablo Castells. Boston, Massachusetts, USA: ACM Press, pp. 123–126 (cit. on p. 263).

Camara, G, M Egenhofer, K Ferreira, et al. (2014). „Fields as a generic data type for big spatial data“. In: *In Proceedigns fo the 8th International Conference on Geographic Information Science*. Ed. by M Duckham, E Pebesma, K Stewart, and A U Frank. Vienna, Austria: Springer International Publishing, pp. 159–172 (cit. on p. 37).

Cantador, Iván, María E. Cortés-Cediel, Miriam Fernández, and Harith Alani (2018). „What's going on in my city? Recommender systems and electronic participatory budgeting“. In: *Proceedings of the 12th ACM Conference on Recommender Systems - RecSys '18*. Ed. by Sole Pera, Michael D. Ekstrand, Xavier Amatriain, and John O'Donovan. Vancouver, British Columbia, Canada: ACM Press, pp. 219–223 (cit. on p. 263).

Cao, Lili and John Krumm (2009). „From GPS traces to a routable road map“. In: *Proceedings of SIGSPATIAL 2009*, pp. 3–12 (cit. on p. 212).

Caragliu, A, C Del Bo, and P Nijkamp (2011). „Smart Cities in Europe“. In: *J. Urban Technol.* 18, pp. 65–82 (cit. on p. 27).

Carpendale, Sheelagh (2008). „Evaluating information visualizations“. In: *Information Visualization*. Ed. by Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 19–45 (cit. on p. 172).

Carrara, Wendy, Wae San Chan, Sander Fischer, and Eva van Steenbergen (2015). *Creating value through open data - A study on the impact of re-use of public data resources*. Tech. rep. Luxembourg, Luxembourg: European Commission, p. 112 (cit. on p. 4).

Carrara, Wendy, Margriet Nieuwenhuis, and Heleen Vollers (2016). *Open Data Maturity in Europe 2016*. Tech. rep. European Data Portal (cit. on pp. 59, 86).

Carrara, Wendy, Heleen Vollers, and Jorn Berends (2017a). *Analytical Report 5: Barriers in working with Open Data*. Tech. rep. European Data Portal, p. 37 (cit. on pp. 57, 59, 60, 70, 73, 74, 79, 83, 87).

– (2017b). *Re-Using Open Data*. Tech. rep. European Data Portal, p. 106 (cit. on pp. 56–59, 70, 73, 74, 85, 86, 88).

Celino, I and S Kotoulas (2013). „Smart Cities“. In: *IEEE Internet Comput.* 17.6, pp. 8–11 (cit. on p. 21).

Chang, Remco, Caroline Ziemkiewicz, Tera Marie Green, and William Ribarsky (2009). „Defining insight for visual analytics“. In: *IEEE Computer Graphics and Applications* 29.2, pp. 14–17 (cit. on p. 170).

Charalabidis, Yannis, Charalampos Alexopoulos, and Euripidis Loukis (2016). „A taxonomy of open government data research areas and topics“. In: *Journal of Organizational Computing and Electronic Commerce* 26.1-2, pp. 41–63 (cit. on pp. 5, 8, 12–14, 182, 220).

Chaudhuri, G and K Clarke (2013). „The SLEUTH land use change model: A review“. In: *Int. J. Environ. Resour. Res.* 1.1, pp. 88–105 (cit. on p. 36).

Chen, M, S Mao, and Y Liu (2014). „Big data: A survey“. In: *Mob. Netw. Appl.* 19.2, pp. 171–209 (cit. on pp. 22, 29, 100).

Chevalier, Fanny, Nathalie Henry Riche, Catherine Plaisant, Amira Chalbi, and Christophe Hurter (2016). „Animations 25 Years later: new roles and opportunities“. In: *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '16*. Ed. by Paolo Buono, Rosa Lanzilotti, Maristella Matera, and Maria Francesca Costabile. Bari, Italy: ACM Press, pp. 280–287 (cit. on p. 263).

Chrisman, Nicholas R. (1998). „Rethinking levels of measurement for Cartography“. In: *Cartography and Geographic Information Systems* 25.4, pp. 231–242 (cit. on p. 243).

Clarke, K and L Gaydos (1998). „Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore“. In: *Int. J. Geogr. Inf. Sci.* 12.7, pp. 699–714 (cit. on p. 36).

Clarke, K, N Gazulis, C Dietzel, and N Goldstein (2007). „A decade of SLEUTHing: lessons learned from applications of a cellular automaton land use change model“. In: *Classics from IJGIS - Twenty Years of the International Journal of Geographical Information Science and Systems*. Ed. by P F Fisher. CRC Press, pp. 413–427 (cit. on p. 36).

Coleman, D J, Y Georgiadou, and J Labonte (2009). „Volunteered geographic information: The nature and motivation of produsers“. In: *Int. J. Spat. Data Infrastruct. Res.* 4, pp. 332–358 (cit. on p. 34).

Coll, Richard A., Joan H. Coll, and Ganesh Thakur (1994). „Graphs and tables: a four-factor experiment“. In: *Communications of the ACM* 37.4, pp. 76–86 (cit. on pp. 202, 203).

Çöltekin, Arzu, Susanne Bleisch, Gennady Andrienko, and Jason Dykes (2017). „Persistent challenges in geovisualization – a community perspective“. In: *Int. Journal of Cartography*, pp. 1–25 (cit. on pp. 141, 167).

Commission of the European Communities the Council, the European Economic and Social Committee, and the Committee of the Regions (2014). *COM (2014) 442: Towards a thriving data-driven economy*. http://eur-lex.europa.eu/procedure/EN/1042141 (accessed on 11 January 2016) (cit. on p. 46).

Connected Smart Cities Network (2015). *Open and Agile Smart Cities*. http://connectedsmartcities.eu/ (accessed on 1 October 2015) (cit. on p. 24).

Conradie, Peter and Sunil Choenni (2014). „On the barriers for local government releasing open data“. In: *Government Information Quarterly* 31.SUPPL.1 (cit. on pp. 58, 59, 61, 68, 73, 90).

Craglia, Max, M F Goodchild, Alessandro Annoni, et al. (2008). „Next-generation Digital Earth: a position paper from the Vespucci Initiative for the advancement of geographic information science“. In: *International Journal of Spatial Data Infrastructures Research* 3, pp. 146–167 (cit. on p. 262).

Craglia, Max and Carlos Granell (2014). *Citizen Science and Smart Cities*. Luxembourg: Publications Office of the European Union (cit. on pp. 28, 29, 34).

Cranefield, Jocelyn, Oliver Robertson, and Gillian Oliver (2014). „Value in the mash: Exploring the benefits, barriers and enablers of open data apps“. In: *ECIS 2014 Proceedings - 22nd European Conference on Information Systems*, pp. 1–15 (cit. on p. 58).

Criado, J. Ignacio, Edgar Alejandro Ruvalcaba-Gómez, and Rafael Valenzuela-Mendoza (2018). „Revisiting the open government phenomenon. A meta-analysis of the international literature“. In: *JeDEM - eJournal of eDemocracy and Open Government* 10.1, pp. 50–81 (cit. on pp. 4, 182).

Cucciniello, Maria, Nicola Belle, Greta Nasi, and Marco Mena (2016). „Smart cities and transparency. Does smartness influence transparency?“ In: *49th Hawaii International Conference on System Sciences (HICSS 2016)*. Ed. by Tung X. Bui and Ralph H. Sprague Jr. Koloa, Hawaii, USA: IEEE Computer Society, pp. 2944–2952 (cit. on p. 166).

Dadzie, Aba-Sah and Emmanuel Pietriga (2017). „Visualisation of Linked Data - Reprise“. In: *Semantic Web* 8.1. Ed. by Aba-Sah Dadzie and Emmanuel Pietriga, pp. 1–21 (cit. on pp. 140, 146, 148, 166).

Dameri, R P and C Rosenthal-Sabroux (2014). *Smart City: How to Create Public and Economic Value with High Technology in Urban Space*. Cham, Switzerland: Springer (cit. on p. 38).

Damiani, ML (2014). „Location privacy models in mobile applications: conceptual view and research directions“. In: *Geoinformatica* 18, pp. 819–842 (cit. on p. 47).

Daniel, S and M Doran (2013). „geoSmartCity: geomatics contribution to the Smart City“. In: *In Proceedings of the 14th Annual International Conference on Digital Government Research, Quebec, Canada*, pp. 65–71 (cit. on pp. 25, 39).

D'Aquin, M, A Adamou, E Daga, et al. (2014). „Dealing with diversity in a smart-city datahub“. In: *Proceedings of the Fifth Workshop on Semantics for Smarter Cities (S4SC 2014)*. Ed. by T Omitola, J Breslin, and P Barnaghi. Riva del Garda, Italy: CEUR-WS.org, pp. 68–82 (cit. on pp. 20, 100).

D'Aquin, Mathieu, John Davies, and Enrico Motta (2015). „Smart cities' data: challenges and opportunities for semantic technologies“. In: *IEEE Internet Computing* 19.6, pp. 66–70 (cit. on p. 166).

Davis, Randall, Howard Shrobe, and Peter Szolovits (1993). „What is a knowledge representation?“ In: *AI Magazine* 14.1, p. 17 (cit. on pp. 190, 191).

de Souza, Cleidson R. B. and David F. Redmiles (2009). „On the roles of APIs in the coordination of collaborative software development“. In: *Computer Supported Cooperative Work (CSCW)* 18.5-6, pp. 445–475 (cit. on p. 264).

Degbelo, Auriol (2013). „Modelling the spatial and temporal resolution of a sensor observation“. In: *GI_Forum 2013 - Creating the GISociety, Salzburg, Austria*. Ed. by T Jekel, A Car, J Strobl, and G Griesebner. Herbert Wichmann Verlag, pp. 71–80 (cit. on p. 36).

– (2015). *Spatial and temporal resolution of sensor observations*. IOS Press, p. 206 (cit. on p. 36).

– (2017). „Linked data and visualization: two sides of the transparency coin“. In: *Proceedings of UrbanGIS'17*. Los Angeles, California, USA: ACM Press, pp. 1–8 (cit. on pp. 141, 142, 181, 185, 190, 213, 220).

Degbelo, Auriol, Devanjan Bhattacharya, Carlos Granell, and Sergio Trilles (2016). „Toolkits for smarter cities: a brief assessment“. In: *UCAmI 2016 - 10th International Conference on Ubiquitous Computing & Ambient Intelligence*. Ed. by R García, P Caballero-Gil, M Burmester, and A Quesada-Arencibia. Las Palmas, Gran Canaria, Spain: Springer International Publishing, pp. 431–436 (cit. on pp. 90, 153, 159, 166).

Degbelo, Auriol, Giorgio De Felice, and Jae Hee Lee (2013). „Integration of quantitative and qualitative geographic data for disaster response: research challenges“. In: *Poster Session at GI_Forum 2013*. Salzburg, Austria (cit. on p. 29).

Degbelo, Auriol, Carlos Granell, Sergio Trilles, Devanjan Bhattacharya, Sven Casteleyn, et al. (2016). „Opening up smart cities: citizen-centric challenges and opportunities from GIScience“. In: *ISPRS International Journal of Geo-Information* 5.2, p. 16 (cit. on pp. 52, 119, 140, 153, 166, 184, 220, 257).

Degbelo, Auriol, Carlos Granell, Sergio Trilles, Devanjan Bhattacharya, and Jonas Wissing (2019). „Tell me how my open data is re-used: increasing transparency through the Open City Toolkit“. In: *Open Data | Open Cities: Collaborative Cities in the Information Era*. Ed. by Scott Hawken, Hoon Han, and Chris Pettit. Palgrave Macmillan (cit. on p. 183).

Degbelo, Auriol and Tomi Kauppinen (2018). „Increasing transparency through web maps“. In: *Companion of Proceedings of the Web Conference 2018 - WWW '18*. Ed. by Pierre-Antoine Champin, Fabien L. Gandon, Mounia Lalmas, and Panagiotis G Ipeirotis. Lyon, France: ACM Press, pp. 899–904 (cit. on pp. 154, 179, 188, 215, 220).

Degbelo, Auriol and Christian Kray (2018). „Intelligent geovisualizations for open government data (vision paper)“. In: *26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Ed. by Farnoush Banaei-Kashani, Erik G. Hoel, Ralf Hartmut Güting, Roberto Tamassia, and Li Xiong. Seattle, Washington, USA: ACM Press, pp. 77–80 (cit. on pp. 185, 224).

Degbelo, Auriol, Sergio Trilles, and Devanjan Bhattacharya (2016). *Categories of datasets offered by current open data portals [Dataset]* (cit. on p. 13).

Degbelo, Auriol, Sergio Trilles, Christian Kray, et al. (2016). „Designing semantic APIs for open government data“. In: *eJournal of Democracy (JeDEM)* 8.2, pp. 21–58 (cit. on pp. 52, 143, 153, 186).

Dennis, Alan R. and Traci A. Carte (1998). „Using Geographical Information Systems for decision making: extending cognitive fit theory to map-based presentations“. In: *Information Systems Research* 9.2, pp. 194–203 (cit. on p. 203).

Department for Business, Innovation & Skills (2013a). *Smart cities, background paper*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/246019/bis-13-1209-smart-cities-background-paper-digital.pdf (accessed on 13 October 2015) (cit. on pp. 21, 23).

– (2013b). *Smart cities international case studies: global innovators*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/249397/bis-13-1216-global-innovators-international-smart-cities.pdf (accessed on 29 October 2015) (cit. on p. 33).

Deterding, S, D Dixon, R Khaled, and L Nacke (2011). „From game design elements to gamefulness: defining gamification“. In: *In Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland*, pp. 9–15 (cit. on p. 47).

Devaraju, A, S Jirka, R Kunkel, and J Sorg (2015). „Q-SOS—A Sensor Observation Service for Accessing Quality Descriptions of Environmental Data“. In: *ISPRS Int. J. Geo-Inf.* 4, pp. 1346–1365 (cit. on p. 38).

Diaz, Laura, Albert Remke, Tomi Kauppinen, et al. (2012). „Future SDI - Impulses from Geoinformatics research and IT trends“. In: *International Journal of Spatial Data Infrastructures Research* 7, pp. 378–410 (cit. on p. 52).

Dijk, Thomas van and Alexander Wolff (2017). „Algorithmically-guided user interaction“. In: *Proceedings of ACM SIGSPATIAL 2017*. SIGSPATIAL'17, 11:1–11:4 (cit. on p. 211).

Dodge, Martin, Mary Mcderby, and Martin Turner (2008). „The power of geographical visualizations“. In: *Geographic Visualization: Concepts, Tools and Applications*. Ed. by Martin Dodge, Mary Mcderby, and Martin Turner, pp. 1–10 (cit. on p. 184).

Domingo, Alberto, Boris Bellalta, Manuel Palacin, Miquel Oliver, and Esteve Almirall (2013). „Public open sensor data: Revolutionizing smart cities“. In: *IEEE Technol. and Soc. Mag.* 32.4, pp. 50–56 (cit. on p. 33).

Donovang-Kuhlisch, M, M Heindrichs-Krusch, W Stoll, et al. (2011). *D2.1 Project Vision. European Platform for Intelligent Cities. No: 270895*. http://www.epic-cities.eu/sites/default/files/documents/D3.1B%20Technical%20Integration%20Architecture%20Plan.pdf (accessed on 1 October 2015) (cit. on p. 23).

Dorbeva, Milena, Yunhyong Kim, and Seamus Ross (2013). „Automated metadata generation“. In: *DCC Digital Curation Manual*. Ed. by Davidson Joy, Seamus Ross, and Michael Day, p. 35 (cit. on p. 262).

Douglas D. Nebert (2014). „Developing Spatial Data Infrastructures: The SDI Cookbook“. In: pp. 1–128 (cit. on pp. 52, 60).

Dowman, I J (1999). „Encoding and validating data from maps and images“. In: *Geographical information systems: principles and technical issues*. Ed. by P A Longley, D J Maguire, M F Goodchild, and D W Rhind. 2nd. New York: John Wiley and Sons. Chap. 31, pp. 437–450 (cit. on p. 170).

Duckham, M (2010). „Moving forward: location privacy and location awareness“. In: *In Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS, San Jose, California, USA*, pp. 1–3 (cit. on p. 39).

Duckham, M and L Kulik (2006). „Location privacy and location-aware computing“. In: *Dynamic and Mobile GIS: Investigating Changes in Space and Time*. Ed. by R Billen, E Joao, and D Forrest. CRC Press, pp. 35–51 (cit. on p. 31).

Duran, David, Vera Sacristán, and Rodrigo Silveira (2016). „Map construction algorithms: an evaluation through hiking data“. In: *Proceedings of MobiGIS'16*. Burlingame, California, USA, pp. 74–83 (cit. on p. 212).

Dykes, Jason, Gennady Andrienko, Natalia Andrienko, Volker Paelke, and Jochen Schiewe (2010). „Editorial – GeoVisualization and the Digital City“. In: *Computers, Environment and Urban Systems* 34.6, pp. 443–451 (cit. on pp. 166, 184).

Eberhardt, André and Milene Selbach Silveira (2018). „Show me the data! A systematic mapping on open government data visualization“. In: *Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age - dgo' 18*. Ed. by Marijn Janssen, Soon Ae Chun, and Vishanth Weerakkody. Delft, The Netherlands: ACM Press, pp. 1–10 (cit. on pp. 3, 8).

Ermilov, Ivan, Sören Auer, and Claus Stadler (2013). „User-driven semantic mapping of tabular data“. In: *Proceedings of the 9th International Conference on Semantic Systems - I-SEMANTICS '13*. Ed. by Marta Sabou, Eva Blomqvist, Tommaso Di Noia, Harald Sack, and Tassilo Pellegrini. Graz, Austria: ACM Press, pp. 105–112 (cit. on p. 191).

ESRI (2010). „Spatial Data Infrastructure A Collaborative Network“. In: *Esri Journal*, p. 8 (cit. on p. 60).

European Commision (2013). *Elements of a data value chain strategy | Digital Single Market*. Tech. rep. (cit. on pp. 86, 87).

Fabrikant, Sara I (2001). „Building task-ontologies for geovisualization“. In: *ICA Pre-Conference Workshop on Geovisualization on the Web* (cit. on p. 215).

Fan, J, F Han, and H Liu (2014). „Challenges of big data analysis“. In: *Natl. Sci. Rev.* 1.2, pp. 293–314 (cit. on p. 29).

Fechner, Thore and Christian Kray (2014). „Georeferenced open data and augmented interactive geo-visualizations as catalysts for citizen engagement“. In: *eJ. of eDemocr. Open Gov.* 6.1, pp. 14–35 (cit. on pp. 34, 35, 100, 184, 210, 220).

Ferreira, K R, G Câmara, and A M V Monteiro (2014). „An algebra for spatiotemporal data: from observations to events“. In: *Trans. GIS* 18, pp. 253–269 (cit. on pp. 37, 211).

Fioretti, M (2012). *Open Data, Open Society: A research project about openness of public data in EU local administration.* `http://www.dime-eu.org/files/active/0/ODOS_report_1.pdf` (accessed on 1 October 2015) (cit. on p. 33).

Forum Virium Helsinki (2014). *What is CitySDK.* `http://www.citysdk.eu/about-the-project/what-is-citysdk/` (accessed on 1 October 2015) (cit. on p. 24).

Frank, A (2003). „Ontology for spatio-temporal databases“. In: *Spatio-Temporal Databases: The CHOROCHRONOS Approach.* Ed. by T Sellis, M Koubarakis, A U Frank, et al. Springer-Verlag Berlin Heidelberg, pp. 9–77 (cit. on p. 36).

Freksa, Christian and Holger Schultheis (2015). „Three ways of using space“. In: *Space in Mind: Concepts for Spatial Learning and Education.* Ed. by Daniel Montello, Karl Grossner, and Donald Janelle. MIT press. Chap. 2, pp. 31–48 (cit. on p. 240).

Furnas, G W, T K Landauer, L M Gomez, and S T Dumais (1987). „The vocabulary problem in human-system communication“. In: *Communications of the ACM* 30.11, pp. 964–971 (cit. on p. 116).

Fusch, Patricia I. and Lawrence R. Ness (2015). „Are we there yet? Data saturation in qualitative research“. In: *The Qualitative Report* 20.9, pp. 1408–1416. arXiv: `arXiv:1403.5723v1` (cit. on p. 157).

Futia, Giuseppe, Alessio Melandri, Antonio Vetrò, Federico Morando, and Juan Carlos De Martin (2017). „Removing barriers to transparency: A case study on the use of semantic technologies to tackle procurement data inconsistency“. In: *The Semantic Web: 14th International Conference, ESWC 2017.* Ed. by Eva Blomqvist, Diana Maynard, Aldo Gangemi, et al. Portoroz, Slovenia, pp. 623–637 (cit. on pp. 168, 183).

Garcia-Molina, Hector, Georgia Koutrika, and Aditya Parameswaran (2011). „Information seeking: convergence of search, recommendations, and advertising“. In: *Communications of the ACM* 54.11, p. 121 (cit. on p. 263).

Gil, Yolanda, Suzanne A Pierce, Hassan Babaie, et al. (2019). „Intelligent systems for geosciences: an essential research agenda“. In: *Communications of the ACM* 62.1, pp. 76–84 (cit. on p. 261).

Glasmeier, A and S Christopherson (2015). „Thinking about smart cities“. In: *Camb. J. Reg. Econo. Soc.* 8, pp. 3–12 (cit. on p. 27).

Godwin, Alex and John T Stasko (2017). „Nodes, paths, and edges: using mental maps to augment crime data analysis in urban spaces“. In: *Eurographics Conference on Visualization, EuroVis 2017, Short Papers.* Ed. by Barbora Kozlíková, Tobias Schreck, and Thomas Wischgoll. Barcelona, Spain: Eurographics Association, pp. 19–23 (cit. on p. 184).

Godwin, Alex, Yongxin Wang, and John T Stasko (2017). „TypoTweet Maps: characterizing urban areas through typographic social media visualization“. In: *Eurographics Conference on Visualization, EuroVis 2017, Short Papers*. Ed. by Barbora Kozlíková, Tobias Schreck, and Thomas Wischgoll. Barcelona, Spain: Eurographics Association, pp. 25–29 (cit. on p. 184).

Gomes, Álvaro and Delfina Soares (2014). „Open government data initiatives in Europe: northern versus southern countries analysis“. In: *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance - ICEGOV '14*. Ed. by Elsa Estevez, Marijn Janssen, and Luís Soares Barbosa. Guimarães, Portugal: ACM Press, pp. 342–350 (cit. on p. 260).

González, A, B Villazón-Terrazas, and J M Gómez (2014). „A linked data lifecycle for smart cities in Spain“. In: *In Proceedings of the Fifth Workshop on Semantics for Smarter Cities, Riva del Garda, Italy*, pp. 9–14 (cit. on p. 20).

Gonzalez-Zapata, Felipe and Richard Heeks (2015). „The multiple meanings of open government data: Understanding different stakeholders and their perspectives“. In: *Government Information Quarterly* 32.4, pp. 441–452 (cit. on pp. 54, 152).

Goodchild, M F (1998). „The geolibrary“. In: *Innovations In GIS 5*. Ed. by Steve Carver. Taylor & Francis, pp. 59–68 (cit. on p. 9).

– (2007). „Citizens as sensors: the world of volunteered geography“. In: *GeoJournal* 69.4, pp. 211–221 (cit. on p. 34).

– (2010). „Twenty years of progress: GIScience in 2010“. In: *J. Spat. Inf. Sci.* 1, pp. 3–20 (cit. on pp. 25, 32, 35).

Gould, Nicholas and William Mackaness (2016). „From taxonomies to ontologies: formalizing generalization knowledge for on-demand mapping“. In: *Cartography and Geographic Information Science* 43.3, pp. 208–222 (cit. on p. 223).

Granell, Carlos, D Havlik, Sven Schade, et al. (2016). „Future Internet technologies for environmental applications“. In: *Environ. Model. Softw.* 78, pp. 1–15 (cit. on p. 27).

Grant, Maria J. and Andrew Booth (2009). „A typology of reviews: an analysis of 14 review types and associated methodologies“. In: *Health Information & Libraries Journal* 26.2, pp. 91–108 (cit. on p. 9).

Graves, Alvaro and James Hendler (2013). „Visualization tools for open government data“. In: *Proceedings of the 14th Annual International Conference on Digital Government Research - dg.o '13*. Ed. by Sehl Mellouli, Luis F. Luna-Reyes, and Jing Zhang. Quebec, Canada: ACM Press, p. 136 (cit. on pp. 143, 152, 174–176, 184, 187, 210, 220).

– (2014). „A study on the use of visualizations for Open Government Data“. In: *Information Polity* 19, pp. 73–91 (cit. on pp. 184, 187, 220, 228, 240, 242, 244).

Griffin, Amy L. (2017). „Color theory“. In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Oxford, UK: John Wiley & Sons, Ltd, pp. 1–14 (cit. on pp. 188, 226).

Griffin, Amy L., Anthony C. Robinson, and Robert E. Roth (2017). „Envisioning the future of cartographic research“. In: *International Journal of Cartography*, pp. 1–8 (cit. on p. 169).

Gruen, A (2013). „SMART Cities: The need for spatial intelligence“. In: *Geo-Spat. Inf. Sci.* 16.1, pp. 3–6 (cit. on p. 32).

Guest, Greg, Arwen Bunce, and Laura Johnson (2006). „How many interviews are enough? An experiment with data saturation and variability". In: *Field Methods* 18.1, pp. 59–82 (cit. on p. 155).

Gurstein, Michael B. (2011). „Open data: Empowering the empowered or effective data use for everyone?" In: *First Monday* 16.2 (cit. on p. 7).

Hahmann, Stefan and Dirk Burghardt (2013). „How much information is geospatially referenced? Networks and cognition". In: *International Journal of Geographical Information Science* 27.6, pp. 1171–1189 (cit. on p. 261).

Hallgren, Kevin A (2012). „Computing inter-rater reliability for observational data: an overview and tutorial". In: *Tutorials in quantitative methods for psychology* 8.1, p. 23 (cit. on pp. 197, 201).

Hancke, Gerhard, Bruno Silva, and Gerhard Hancke Jr. (2012). „The Role of Advanced Sensing in Smart Cities". In: *Sensors* 13.1, pp. 393–425 (cit. on pp. 38, 100).

Hardisty, Frank and Anthony C. Robinson (2011). „The GeoViz toolkit: using component-oriented coordination methods for geographic visualization and analysis". In: *International Journal of Geographical Information Science* 25.2, pp. 191–210 (cit. on p. 223).

Harrie, L, H Stigmar, and M Djordjevic (2015). „Analytical estimation of map readability". In: *ISPRS Int. J. Geo-Inf.* 4.2, pp. 418–446 (cit. on p. 35).

Harrower, Mark and Cynthia A. Brewer (2003). „ColorBrewer.org: an online tool for selecting colour schemes for maps". In: *The Cartographic Journal* 40.1, pp. 27–37 (cit. on pp. 147, 222).

Hartog, Martijn, Bert Mulder, Bart Spée, Ed Visser, and Antoine Gribnau (2014). „Open data within governmental organisations". In: *eJournal of eDemocracy and Open Government* 6.1, pp. 49–61 (cit. on pp. 100, 152).

Harvey, F and D. Tulloch (2006). „Local-government data sharing: Evaluating the foundations of spatial data infrastructures". In: *International Journal of Geographical Information Science* 20.7, pp. 743–768 (cit. on p. 60).

Heald, David (2006). „Varieties of transparency". In: *Transparency: The Key to Better Governance?* Ed. by Christopher Hood and David Heald. British Academy, pp. 24–43 (cit. on p. 167).

Henning, Michi (2009). „API design matters". In: *Communications of the ACM* 52.5, pp. 25–36 (cit. on p. 101).

Hespanhol, Luke, Martin Tomitsch, Ian McArthur, et al. (2016). „Situated interfaces for engaging citizens on the go". In: *Interactions* 23.1, pp. 40–45 (cit. on p. 32).

Hinz, Matthias and Ralf Bill (2018a). „Ein zentraler Einstiegspunkt für die Suche nach offenen Geodaten im deutschsprachigen Raum". In: *AGIT – Journal für Angewandte Geoinformatik*, pp. 298–307 (cit. on pp. 3, 4).

– (2018b). „Mapping the landscape of open geodata". In: *Geospatial Technologies for All : short papers, posters and poster abstracts of the 21th AGILE Conference on Geographic Information Science (AGILE 2018)*. Ed. by Ali Mansourian, Petter Pilesjö, Lars Harrie, and Ron van Lammeren. Lund, Sweden: Online Proceedings (cit. on pp. 3, 4).

Hitzler, Pascal and Krzysztof Janowicz (2013). „Linked Data, Big Data, and the 4th Paradigm". In: *Semantic Web* 4.3, pp. 233–235 (cit. on p. 168).

Hoevenaars (2015). *The biggest challenges to develop a successful smart city project*. http://www.smart-circle.org/smartcity/blog/biggest-challenges-develop-successful-smart-city-project/ (accessed on 29 October 2015) (cit. on p. 27).

Hollands, RG (2008). „Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?" In: *City* 12.3, pp. 303–320 (cit. on p. 27).

Hossain, Mohammad Alamgir, Yogesh K. Dwivedi, and Nripendra P. Rana (2016). „State-of-the-art in open data research: Insights from existing literature and a research agenda". In: *Journal of Organizational Computing and Electronic Commerce* 26.1-2, pp. 14–40 (cit. on pp. 3, 4, 7, 220).

Hunnius, Sirko, Bernhard Krieger, and Tino Schuppan (2014). „Providing, guarding, shielding: open government data in Spain and Germany". In: *European Group for Public Administration Annual Conference*. Speyer, Germany (cit. on p. 3).

Isenberg, Tobias, Petra Isenberg, Jian Chen, Michael Sedlmair, and Torsten Moller (2013). „A systematic review on the practice of evaluating visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12, pp. 2818–2827 (cit. on p. 167).

Jameson, Anthony and Wolfgang Wahlster (1982). „User modelling in anaphora generation: ellipsis and definite description". In: *Proceedings of the 5th European Conference on Artificial Intelligence*. Paris, France, pp. 222–227 (cit. on p. 213).

Janev, Valentina, Vuk Mijović, Dejan Paunović, and Uroš Milošević (2014). „Modeling, Fusion and Exploration of Regional Statistics and Indicators with Linked Data Tools". In: Springer, Cham, pp. 208–221 (cit. on p. 74).

Jannach, Dietmar, Paul Resnick, Alexander Tuzhilin, and Markus Zanker (2016). „Recommender systems - beyond matrix completion". In: *Communications of the ACM* 59.11, pp. 94–102 (cit. on p. 263).

Janowicz, K (2012). „Observation-driven geo-ontology engineering". In: *Trans. GIS* 16.3, pp. 351–374 (cit. on p. 36).

Janssen, Marijn, Yannis Charalabidis, and Anneke Zuiderwijk (2012). „Benefits, adoption barriers and myths of open data and open government". In: *Information Systems Management* 29.4, pp. 258–268 (cit. on pp. 4, 6, 7, 20, 26, 28, 31, 45, 52, 53, 58, 60, 68, 73, 87, 100, 153, 168).

Jensen, Ryan R. and J. Matthew Shumway (2010). „Sampling our world". In: *Research Methods in Geography: A Critical Introduction*. Ed. by Basil Gomez and John Paul Jones III. John Wiley & Sons, pp. 77–90 (cit. on p. 10).

Jetzek, Thorhildur, Michel Avital, and Niels Bjorn-Andersen (2013). „Generating Value from Open Government Data". In: *ICIS 2013 Proceedings - International Conference on Information Systems*. July 2016, pp. 1737–1756 (cit. on pp. 58, 183).

Jetzek, Tornhildur, Michel Avital, and Niles Bjorn-Andersen (2012). „The Value of Open Government Data". In: *Openarchive.Cbs.Dk* 2002, p. 12 (cit. on p. 58).

Johannessen, Marius Rohde and Lasse Berntzen (2018). „The transparent smart city". In: *Smart Technologies for Smart Governments*. Ed. by Manuel Pedro Rodríguez Bolívar. Springer, pp. 67–94 (cit. on pp. 140, 166, 167, 180).

John B. Horrigan and Lee Rainie (2015). „Americans' views on open government data". In: *Pew Research Center* April. (last accessed: October 17, 2017) (cit. on pp. 53, 174).

Johnson, Jeff (2010). *Designing with the mind in mind*. Elsevier (cit. on p. 171).

Johnson, Peter A., Renee Sieber, Teresa Scassa, Monica Stephens, and Pamela Robinson (2017). „The cost(s) of geospatial open data“. In: *Transactions in GIS* 21.3, pp. 434–445 (cit. on pp. 58, 242).

Kamaruddin, Kamalia and Nor Laila Md Noor (2017). „Citizen-centric demand model for transformational government systems“. In: *21st Pacific Asia Conference on Information Systems (PACIS 2017)*. Langkawi, Malaysia (cit. on pp. 141, 181).

Keefe, Terry, Andrea Bikfalvi, Martin Beer, and Josep la Rosa (2013). „A Case Study Analysis of Factors Determining Success or Failure for Participants in Collaborative Innovation Projects in e-Government“. In: *Proceedings of the 13Th European Conference on Egovernment*, pp. 276–282 (cit. on p. 71).

Kehoe, M, M Cosgrove, S D Gennaro, et al. (2011). *Smarter Cities Series: A Foundation for Understanding IBM Smarter Cities*. http://www.redbooks.ibm.com/redpapers/pdfs/redp4733.pdf (accessed on 1 October 2015) (cit. on p. 23).

Kent, Robert B. and Richard E. Klosterman (2000). „GIS and mapping“. In: *Journal of the American Planning Association* 66.2, pp. 189–198 (cit. on p. 212).

Kiefer, Peter, Ioannis Giannopoulos, Vasileios Anagnostopoulos, Johannes Schöning, and Martin Raubal (2017). „Controllability matters: The user experience of adaptive maps“. In: *GeoInformatica* 21.3, pp. 619–641 (cit. on p. 224).

Kiefer, Peter, Ioannis Giannopoulos, and Martin Raubal (2013). „Using eye movements to recognize activities on cartographic maps“. In: *Proceedings of ACM SIGSPATIAL 2013*, pp. 488–491 (cit. on p. 215).

Kim, Soonhee and Jooho Lee (2017). „Citizen participation and transparency in local government: Do participation channels and policy making phases matter?“ In: *50th Hawaii International Conference on System Sciences - HICSS 2017*. Hilton Waikoloa Village, Hawaii, USA: AIS Electronic Library (AISeL) (cit. on p. 166).

Klímek, Jakub, Jiří Helmich, and Martin Nečaský (2016). „LinkedPipes visualization: simple useful linked data visualization use cases“. In: *ESWC 2016 Satellite Events -Revised Selected Papers*. Ed. by Harald Sack, Giuseppe Rizzo, Nadine Steinmetz, et al. Springer, Cham, pp. 112–117 (cit. on pp. 140, 180).

Kogan, N (2014). „Exploratory research on success factors and challenges of Smart City Projects“. MA thesis. Department of Business Administration, Graduate School, Kyung Hee University, Seoul, Korea (cit. on p. 28).

Kraak, M (2006). „Why maps matter in GIScience“. In: *Cartogr. J.* 43.1, p. 82 (cit. on p. 35).

Kucera, Jan and Dusan Chlapek (2014). „Benefits and Risks of Open Government Data“. In: *J. Syst. Integr.* 5.1, pp. 30–41 (cit. on p. 26).

Kučera, Jan, Dušan Chlapek, and Martin Nečaský (2013). „Open Government Data Catalogs : Current Approaches and Quality Perspective“. In: *Technology-Enabled Innovation for Democracy, Government and Governance* 8061, pp. 152–166 (cit. on pp. 52, 54).

Kuhn, Werner (1996). „Handling data spatially: spatializating user interfaces“. In: *In Proceedings of the 7th International Symposium on Spatial Data Handling, Delft, The Netherlands*. Ed. by M Kraak and M Molenaar. Vol. 2, 13B.1–13B.23 (cit. on p. 40).

– (2009). „A functional ontology of observation and measurement“. In: *In Proceedings of the Third International Conference on GeoSpatial Semantics*. Ed. by K Janowicz, M Raubal, and S Levashkin. Mexico City, Mexico: Springer Berlin Heidelberg, pp. 26–43 (cit. on p. 36).

– (2013). „Cognitive and linguistic ideas in geographic information semantics“. In: *Cognitive and Linguistic Aspects of Geographic Space*. Ed. by M Raubal, D M Mark, and A U Frank. Springer-Verlag Berlin Heidelberg, pp. 159–174 (cit. on p. 37).

Kuhn, Werner and Andrew Frank (1991). „A formalization of metaphors and image-schemas in user interfaces“. In: *Cognitive and linguistic aspects of geographic space*. Ed. by A U Frank and D M Mark. Kluwer Academic Publishers, pp. 419–434 (cit. on p. 40).

Kuhn, Werner, Tomi Kauppinen, and Krzysztof Janowicz (2014). „Linked Data - A paradigm shift for Geographic Information Science“. In: *Geographic Information Science - Eighth International Conference*. Ed. by M Duckham, E Pebesma, K Stewart, and A U Frank. Vienna, Austria: Springer International Publishing, pp. 173–186 (cit. on pp. 169, 170).

Landis, J R and G G Koch (1977). „The measurement of observer agreement for categorical data“. In: *Biometrics* 33.1, pp. 159–174 (cit. on p. 226).

Larkin, Jill H. and Herbert A. Simon (1987). „Why a diagram is (sometimes) worth ten thousand words“. In: *Cognitive Science* 11.1, pp. 65–100 (cit. on pp. 190, 213).

Lathrop, D and L Ruma (2010). *Open government: Collaboration, transparency, and participation in practice*. Sebastopol, CA, USA: O'Reilly Media, Inc. (cit. on pp. 28, 33).

Lazar, Jonathan, Jinjuan Heidi Jh Feng, and Harry Hochheiser (2010). *Research methods in human-computer interaction*. John Wiley & Sons, p. 426 (cit. on pp. 9–11, 156).

Lecue, F, S Kotoulas, and P Mac Aonghusa (2012). „Capturing the pulse of cities: A robust stream data reasoning approach“. In: *The AAAI 2012 Workshop on Semantic Cities*. Ed. by B Srivastava, F Lecue, and A Joshi. Toronto, Ontario, Canada: AAAI Press, pp. 9–14 (cit. on pp. 100, 120).

Lee, Jae-Nam, Juyeon Ham, and Byounggu Choi (2016). „Effect of government data openness on a knowledge-based economy“. In: *Procedia Computer Science* 91, pp. 158–167 (cit. on p. 54).

Lee, Melissa, Esteve Almirall, and Jonathan Wareham (2016). „Open data and civic apps: First-generation failures, second-generation improvements“. In: *Communications of the ACM* 59.1, pp. 82–89 (cit. on pp. 26, 100, 103).

Lewis, James R. (2006). „Sample sizes for usability tests: mostly math, not magic“. In: *interactions* 13.6, p. 29 (cit. on p. 10).

– (2014). „Usability: lessons learned ... and yet to be learned“. In: *International Journal of Human-Computer Interaction* 30.9, pp. 663–684 (cit. on pp. 11, 155).

– (2018). „The System Usability Scale: Past, Present, and Future“. In: *International Journal of Human–Computer Interaction* 34.7, pp. 577–590 (cit. on p. 229).

Li, W, L Li, M F Goodchild, and L Anselin (2013). „A Geospatial Cyberinfrastructure for Urban Economic Analysis and Spatial Decision-Making“. In: *ISPRS Int. J. Geo-Inf.* 2, pp. 413–431 (cit. on p. 38).

Liang, Steve, Chih-Yuan Huang, and Tania Khalafbeigi (2015). *OGC SensorThing API*. `http://ogc-iot.github.io/ogc-iot-api` (accessed on 3 November 2015) (cit. on p. 38).

Lohse, Gerald L., Kevin Biolsi, Neff Walker, and Henry H. Rueter (1994). „A classification of visual representations“. In: *Communications of the ACM* 37.12, pp. 36–49 (cit. on pp. 194, 197, 198, 200, 201, 203).

Lopez, Violeta and Dean Whitehead (2013). „Sampling data and data collection in qualitative research“. In: *Nursing and Midwifery Research: Methods and Critical Appraisal for Evidence-based Practice*. Ed. by Zevia Schneider, Dean Whitehead, Geri LoBiondo-Wood, and Judith Haber. Elsevier Health Sciences, London, pp. 124–140 (cit. on pp. 155, 156).

López-de-Ipiña, Diego, Sacha Vanhecke, Oscar Peña, Tom De Nies, and Erik Mannens (2013). „Citizen-centric linked data apps for smart cities“. In: *Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction (UCAmI 2013)*. Ed. by Gabriel Urzaiz, Sergio F Ochoa, José Bravo, Liming Luke Chen, and Jonice Oliveira. Guanacaste, Costa Rica: Springer International Publishing, pp. 70–77 (cit. on p. 183).

Lourenço, Rui Pedro (2015). „An analysis of open government portals: A perspective of transparency for accountability“. In: *Government Information Quarterly* 32.3, pp. 323–332 (cit. on pp. 53, 152).

Lyon, D (2014). „Surveillance, Snowden, and big data: capacities, consequences, critique“. In: *Big Data Soc.* 1.2, pp. 1–13 (cit. on p. 31).

MacEachren, Alan M. (2018). „(re)Considering Bertin in the age of big data and visual analytics“. In: *Cartography and Geographic Information Science*, pp. 1–18 (cit. on p. 225).

Mackinlay, Jock (1986). „Automating the design of graphical presentations of relational information“. In: *ACM Transactions on Graphics* 5.2, pp. 110–141 (cit. on p. 214).

Maisonneuve, Nicolas, Matthias Stevens, Maria E Niessen, and Luc Steels (2009). „NoiseTube: Measuring and mapping noise pollution with mobile phones“. In: *Information Technologies in Environmental Engineering*. Springer, pp. 215–228 (cit. on p. 34).

Martella, Roberta, Christian Kray, and Eliseo Clementini (2015). „A gamification framework for volunteered geographic information“. In: *In Proceedings of the International AGILE 2015 Conference - Geographic Information Science as an Enabler of Smarter Cities and Communities, Lisbon, Portugal*, pp. 73–89 (cit. on p. 34).

Martin, Michael, Claus Stadler, Philipp Frischmuth, and Jens Lehmann (2014). „Increasing the financial transparency of European Commission project funding“. In: *Semantic Web* 5.2, pp. 157–164 (cit. on p. 168).

Martin, Sébastien and Muriel Foulonneau (2013). „Open Data: Barriers, Risks and Opportunities“. In: pp. 301–309 (cit. on pp. 58, 68, 73).

Martin, Sébastien, Muriel Foulonneau, Slim Turki, and Madjid Ihadjadene (2013a). „Open data: barriers, risks and opportunities“. In: *In Proceedings of the 13th European Conference on eGovernment: ECEG, Como, Italy*, pp. 301–309 (cit. on p. 26).

– (2013b). „Risk Analysis to Overcome Barriers to Open Data“. In: *Electronic J. of e-Gov.* 11.1, pp. 348–359 (cit. on p. 26).

Marx, Edgard, Amrapali Zaveri, Mofeed Mohammed, et al. (2016). „DBtrends: Publishing and benchmarking RDF ranking functions“. In: *Proceedings of the 2nd International Workshop on Summarizing and Presenting Entities and Ontologies (SumPre 2016)*. Ed. by Andreas Thalhammer, Gong Cheng, and Kalpa Gunaratna. Anissaras, Greece: CEUR-WS.org (cit. on p. 174).

Marzouki, A, F Lafrance, S Daniel, and S Mellouli (2017). „The relevance of geovisualization in citizen participation processes“. In: *Proceedings of the 18th Annual International Conference on Digital Government Research*. Ed. by Charles C. Hinnant and Adegboyega Ojo. dg.o ’17. Staten Island, NY, USA: ACM, pp. 397–406 (cit. on p. 184).

Masip-Bruin, X, Guang-Jie R, R Serral-Gracia, and M Yannuzzi (2013). „Unlocking the value of open data with a process-based information platform“. In: *In Proceedings of 15th IEEE Conference on Business Informatics, Vienna, Austria*, pp. 331–337 (cit. on pp. 20, 26, 100).

Mason, G. S., T. R. Shuman, and K. E. Cook (2013). „Comparing the Effectiveness of an Inverted Classroom to a Traditional Classroom in an Upper-Division Engineering Course“. In: *IEEE Transactions on Education* 56.4, pp. 430–435 (cit. on p. 185).

Matheus, Ricardo, Marijn Janssen, and Devender Maheshwari (2018). „Data science empowering the public: data-driven dashboards for transparent and accountable decision-making in smart cities“. In: *Government Information Quarterly* (cit. on pp. 154, 183).

Mauri, Michele, Tommaso Elli, Giorgio Caviglia, Giorgio Uboldi, and Matteo Azzi (2017). „RAWGraphs: a visualisation platform to create open outputs“. In: *Proceedings of CHItaly ’17*. Cagliari, Italy, pp. 1–5 (cit. on pp. 212, 224, 228).

McGrath, B and STA Pickett (2011). „The metacity: A conceptual framework for integrating ecology and urban design“. In: *Challenges* 2.4, pp. 55–72 (cit. on p. 23).

Meng, Liqiu (2018). „Four persistent research questions in Cartography“. In: *Journal of the Croatian Cartographic Society* 17.29 (cit. on p. 242).

Menychtas, A, P Kranas, M Donovang-Kuhlisch, M Heindrichs-Krusch, et al. (2011). *D2.3 Online Service Delivery Baseline and Technical Requirements Report. European Platform for Intelligent Cities. No: 270895*. `http://www.epic-cities.eu/sites/default/files/documents/D2.3B%20Online%20Service%20Delivery%20Baseline%20and%20technical%20requirements%20report.pdf` (accessed on 1 October 2015) (cit. on p. 23).

Menychtas, A, P Kranas, M Donovang-Kuhlisch, U Schade, et al. (2011). *D2.1 Project Vision. European Platform for Intelligent Cities. No: 270895*. `http://epic-cities.eu/sites/default/files/documents/D2.1%20Project%20Vision.pdf` (accessed on 1 October 2015) (cit. on p. 23).

MH Government, Cabinet Office, London, United Kingdom (2012). *Open Data White Paper-Unleashing the potential*. `https://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf` (accessed on 11 January 2016) (cit. on p. 33).

Michener, G and K Bersch (2013). „Identifying transparency“. In: *Information Polity* 18.3, pp. 233–242 (cit. on pp. 5, 43, 103, 141, 168, 181).

Milakovich, Michael E. (2010). „The internet and increased citizen participation in government“. In: *eJournal of Democracy (JeDEM)* 2.1, pp. 1–9 (cit. on pp. 140, 166, 180).

Miller, H and M F Goodchild (2015). „Data-driven geography“. In: *GeoJournal* 80.4, pp. 449–461 (cit. on p. 40).

Mitton, Nathalie, Symeon Papavassiliou, Antonio Puliafito, and KishorS Trivedi (2012). „Combining Cloud and sensors in a smart city environment“. In: *EURASIP J. Wirel. Commun. Netw.* 2012.1, 247 (cit. on p. 38).

Mockus, Martynas and Monica Palmirani (2017). „Legal ontology for open government data mashups“. In: *2017 Conference for E-Democracy and Open Government (CeDEM)*. Ed. by Peter Parycek and Noella Edelmann. Krems, Austria: IEEE, pp. 113–124 (cit. on p. 183).

Mocnik, Franz-Benjamin and David Fairbairn (2018). „Maps telling stories?“ In: *The Cartographic Journal* 55.1, pp. 36–57 (cit. on pp. 190–192).

Monmonier, Mark (2005). „Lying with maps“. In: *Statistical Science* 20.3, pp. 215–222 (cit. on p. 212).

Mora-Rodriguez, Maria, Ghislain Auguste Atemezing, and Chris Preist (2017). „Adopting semantic technologies for effective corporate transparency“. In: *The Semantic Web: 14th International Conference, ESWC 2017*. Ed. by Eva Blomqvist, Diana Maynard, Aldo Gangemi, et al. Portoroz, Slovenia, pp. 655–670 (cit. on p. 168).

Moritz, Dominik, Chenglong Wang, Greg L. Nelson, et al. (2018). „Formalizing visualization design knowledge as constraints: actionable and extensible models in Draco“. In: *IEEE TVCG* (cit. on p. 214).

Muñoz-Soro, José Félix, Guillermo EstebAn, Oscar Corcho, and Francisco Serón (2016). „PPROC, an ontology for transparency in public procurement“. In: *Semantic Web* 7.3, pp. 295–309 (cit. on p. 183).

Murray, Scott (2013). *Interactive data visualization for the Web: an introduction to designing with D3*. 2nd. O'Reilly Media, Inc. (cit. on pp. 183, 241).

Myers, Brad and Jeffrey Stylos (2016). „Improving API usability“. In: *Communications of the ACM* 59.6, pp. 62–69 (cit. on pp. 101, 103).

Nam, T and T A Pardo (2011). „Conceptualizing smart city with dimensions of technology, people, and institutions“. In: *In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, College Park, Maryland, USA*. Ed. by J Bertot, K Nahon, S A Chun, L Luna-Reyes, and V Atluri. ACM, pp. 282–291 (cit. on pp. 20–22).

Naphade, M, G Banavar, C Harrison, J Paraszczak, and R Morris (2011). „Smarter Cities and Their Innovation Challenges“. In: *IEEE Comput.* 44.6, pp. 32–39 (cit. on p. 27).

Nemoto, Tomoko and David Beglar (2014). *Likert-Scale Questionnaires* (cit. on p. 72).

North, Chris (2006). „Toward measuring visualization insight“. In: *IEEE Computer Graphics and Applications* 26.3, pp. 6–9 (cit. on p. 171).

Nugroho, Rininta Putri, Anneke Zuiderwijk, Marijn Janssen, and Martin de Jong (2015). „A comparison of national open data policies: lessons learned“. In: *Transforming Government: People, Process and Policy* 9.3 (cit. on pp. 58, 63).

OECD (2018). *Open government data - OECD*. *`http://www.oecd.org/gov/digital-government/open-government-data.htm` (last accessed: December 18, 2018).* (Cit. on p. 3).

Ojo, A, E Curry, and F A Zeleti (2015). „A tale of open data innovations in five smart cities“. In: *In Proceedings of the 48th Hawaii International Conference on System Sciences, Kauai, Hawaii*, pp. 2326–2335 (cit. on pp. 4, 20, 100).

Ojo, Adegboyega, Zamira Dzhusupova, and Edward Curry (2016). „Exploring the nature of the smart cities research landscape“. In: *Smarter as the New Urban Agenda: A Comprehensive View of the 21st Century City*. Ed. by Ramon J Gil-Garcia, A Theresa Pardo, and Taewoo Nam. Springer International Publishing, pp. 23–47 (cit. on p. 165).

Open Cities (n.d.). `http://www.opencitiesproject.org/` (accessed on 16 October 2015) (cit. on p. 24).

Open Cities Consortium (2011). *Project Open Cities*. `http://www.opencities.net/content/project` (accessed on 1 October 2015) (cit. on p. 23).

Open Data Institute (2015). *The Value of Open Data* (cit. on p. 61).

Open311 Organisation (2014). *A collaborative model and open standard for civic issue tracking*. `http://www.open311.org/learn/` (accessed on 1 October 2015) (cit. on p. 24).

Ottley, Alvitta, Huahai Yang, and Remco Chang (2015). „Personality as a predictor of user strategy: how locus of control affects search strategies on tree visualizations“. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. Ed. by Bo Begole, Jinwoo Kim, Kori Inkpen, and Woontack Woo. Seoul, Korea: ACM Press, pp. 3251–3254 (cit. on pp. 236, 241).

Oulasvirta, Antti (2017). „User interface design with combinatorial optimization“. In: *Computer* 50.1, pp. 40–47 (cit. on p. 214).

Park, Jung-ran and Andrew Brenza (2015). „Evaluation of semi-automatic metadata generation tools: A survey of the current state of the art“. In: *Information Technology and Libraries* 34.3 (cit. on p. 262).

Pearce, J. (2009). „Regression, linear and nonlinear“. In: *International Encyclopedia of Human Geography*. Elsevier, pp. 302–308 (cit. on p. 11).

Peng, Shangfu, Hanan Samet, and Marco Adelfio (2014). „Viewing streaming spatially-referenced data at interactive rates“. In: *ACM SIGSPATIAL 2014*, pp. 409–412 (cit. on p. 213).

Percivall, G (2015). *OGC smart cities spatial information framework*. `http://www.opengeospatial.org/pressroom/pressreleases/2181` (accessed on 11 January 2016) (cit. on pp. 38, 41).

Peuquet, Donna J. (1988). „Representations of geographic space: toward a conceptual synthesis“. In: *Annals of the Association of American Geographers* 78.3, pp. 375–394 (cit. on pp. 141, 181).

Pezanowski, Scott, Alan M MacEachren, Alexander Savelyev, and Anthony C Robinson (2018). „SensePlace3: a geovisual framework to analyze place–time–attribute information in social media“. In: *Cartography and Geographic Information Science* 45.5, pp. 420–437 (cit. on p. 223).

Plewe, Brandon (2007). „Web cartography in the United States“. In: *Cartography and Geographic Information Science* 34.2, pp. 133–136 (cit. on p. 221).

Pomerantz, Jeffrey and Robin Peek (2016). „Fifty shades of open“. In: *First Monday* 21.5 (cit. on pp. 51, 100, 152).

Poplin, Alenka (2015). „How user-friendly are online interactive maps? Survey based on experiments with heterogeneous users“. In: *Cartography and Geographic Information Science* 42.4, pp. 358–376 (cit. on pp. 222, 243).

Poplin, Alenka, Wendy Guan, and Ben Lewis (2017). „Online survey of heterogeneous users and their usage of the interactive mapping platform WorldMap“. In: *The Cartographic Journal* 54.3, pp. 214–232 (cit. on pp. 220, 222).

Quinn, Alexander J. and Benjamin B. Bederson (2011). „Human computation: a survey and taxonomy of a growing field“. In: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. Ed. by Desney S. Tan, Saleema Amershi, Bo Begole, Wendy A. Kellogg, and Manas Tungare. Vancouver, British Columbia, Canada: ACM Press, p. 1403 (cit. on pp. 224, 262).

R Core Team (2016). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria (cit. on p. 233).

Ratti, Carlo and Anthony Townsend (2011). „Harnessing residents' electronic devices will yield truly smart cities“. In: *Sci. Am.* 305.1, pp. 42–48 (cit. on p. 33).

Reed, C, K Buehler, and L McKee (2015). „OGC Consensus: How Successful Standards Are Made“. In: *ISPRS Int. J. Geo-Inf.* 4, pp. 1693–1706 (cit. on p. 38).

Reichenbacher, Tumasch (2017). „Mobile usage and adaptive visualization“. In: *Encyclopedia of GIS*. Ed. by Shashi Shekhar, Hui Xiong, and Xun Zhou. 2nd. Springer, Cham, pp. 1289–1294 (cit. on p. 223).

Richardson, Leonard and Mike Amundsen (2013). *RESTful web APIs*. O'Reilly Media, Inc., p. 379. eprint: `arXiv:1011.1669v3` (cit. on p. 101).

Roberts, Jonathan (2008). „Coordinated multiple views for exploratory geovisualization“. In: *Geographic Visualization*. John Wiley & Sons, Ltd, pp. 25–48 (cit. on pp. 183, 210, 212).

Roberts, Tony (2012). *The problem with Open Data*. `http://www.computerweekly.com/opinion/The-problem-with-Open-Data` (accessed on 05 January 2016) (cit. on p. 26).

Robinson, Anthony C. (2017). „Geovisual analytics“. In: *Geographic Information Science & Technology Body of Knowledge* 2017.Q3 (cit. on p. 222).

– (2018). „Elements of viral cartography“. In: *Cartography and Geographic Information Science*, pp. 1–18 (cit. on pp. 184, 222).

Robinson, Anthony C., Urška Demšar, Antoni B. Moore, et al. (2017). „Geospatial big data and cartography: research challenges and opportunities for making maps that matter“. In: *International Journal of Cartography*, pp. 1–29 (cit. on p. 172).

Robinson, Anthony C., Donna J. Peuquet, Scott Pezanowski, Franklin A. Hardisty, and Brian Swedberg (2017). „Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data“. In: *Cartography and Geographic Information Science* 44.3, pp. 216–228 (cit. on p. 223).

Robinson, Anthony C., Robert E. Roth, and Alan M. MacEachren (2011). „Designing a web-based learning portal for geographic visualization and analysis in public health“. In: *Health Informatics Journal* 17.3, pp. 191–208. eprint: `NIHMS150003` (cit. on p. 184).

Roche, S (2014). „Geographic Information Science I: why does a smart city need to be spatially enabled?“ In: *Pro. Hum. Geogr.* 38.5, pp. 703–711 (cit. on pp. 25, 39).

Rodger, Lea (2015). *Smart Cities: technology challenges*. `http://urbanopus.net/smart-cities-technology-challenges/` (accessed on 26 October 2015) (cit. on pp. 25, 27).

Roth, Robert E. (2013a). „An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization“. In: *IEEE Trans. Vis. Comput. Graphics* 19.12, pp. 2356–2365 (cit. on pp. 146, 148, 188, 192, 202, 203, 215).

– (2013b). „Interactive maps: What we know and what we need to know“. In: *JOSIS* 6, pp. 59–115 (cit. on pp. 191, 210, 212, 214, 220).

– (2017). „Visual variables“. In: *International Encyclopedia of Geography: People, the Earth, Environment and Technology*. Oxford, UK: John Wiley & Sons, Ltd, pp. 1–11 (cit. on pp. 146, 147, 188, 212, 225, 226).

Roth, Robert E., Arzu Çöltekin, Luciene Delazari, et al. (2017). „User studies in cartography: opportunities for empirical research on interactive maps and visualizations“. In: *International Journal of Cartography*, pp. 1–29 (cit. on p. 167).

Roth, Robert E., Richard G. Donohue, Carl M. Sack, Timothy R. Wallace, and Tanya M A Buckingham (2014). „A process for keeping pace with evolving web mapping technologies“. In: *Cartographic Perspectives* 2014.78, pp. 25–52 (cit. on p. 221).

Roth, Robert E. and Alan M. MacEachren (2016). „Geovisual analytics and the science of interaction: an empirical interaction study“. In: *Cartography and Geographic Information Science* 43.1, pp. 30–54 (cit. on p. 192).

Roth, Robert E., Kevin Ross, and Alan M. MacEachren (2015). „User-centered design for interactive maps: A case study in crime analysis“. In: *ISPRS International Journal of Geo-Information* 4.1, pp. 262–301 (cit. on p. 184).

Safarov, Igbal, Albert Meijer, and Stephan Grimmelikhuijsen (2017). „Utilization of open government data: A systematic literature review of types, conditions, effects and users“. In: *Information Polity* 22.1, pp. 1–24 (cit. on pp. 3, 4, 182, 220, 258).

Sarikaya, Alper, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher (2019). „What do we talk about when we talk about dashboards?“ In: *IEEE Transactions on Visualization and Computer Graphics* 25.1, pp. 682–692 (cit. on p. 263).

Sarjakoski, L. Tiina and Tapani Sarjakoski (2008). „User interfaces and adaptive maps“. In: *Encyclopedia of GIS*. Springer, Boston, MA, pp. 1205–1212 (cit. on p. 223).

Saunders, Benjamin, Julius Sim, Tom Kingstone, et al. (2018). „Saturation in qualitative research: exploring its conceptualization and operationalization“. In: *Quality & Quantity* 52.4, pp. 1893–1907 (cit. on p. 225).

Sauro, Jeff (2011). *What is a good task completion rate?* (Cit. on p. 194).

– (2013). *10 things to know about the system usability scale (SUS)* (cit. on pp. 155, 229, 235).

Savelyev, Alexander and Alan M. MacEachren (2018). „Augmenting geovisual analytics of social media data with heterogeneous information network mining—Cognitive plausibility assessment“. In: *PLOS ONE* 13.12. Ed. by Chaowei Yang, e0206906 (cit. on p. 223).

Schade, Sven, Carlos Granell, and Andrea Perego (2015). „Coupling Public Sector information and Public-Funded Research Data in Europe: A Vision of an open Data ecosystem“. In: *Information and Communication Technologies in Public Administration: Innovations from Developed Countries*. Ed. by C.G. Reddick and L. Anthopoulos. Boca Raton, FL, USA: CRC Press, pp. 195–275 (cit. on pp. 33, 152).

Schaffers, H, N Komninos, M Pallot, et al. (2011). „Smart Cities and the Future Internet: Towards Cooperation Frameworks for Open Innovation“. In: *Future Internet Assembly*. Ed. by J Domingue, A Galis, A Gavras, et al. Springer, pp. 431–446 (cit. on p. 27).

Scheider, Simon, Auriol Degbelo, Werner Kuhn, and Holger Przibytzin (2014). „Content and context description - How linked spatio-temporal data enables novel information services for libraries“. In: *GIS.Science* 4, pp. 138–149 (cit. on p. 31).

Scheider, Simon, Benedikt Gräler, Edzer Pebesma, and Christoph Stasch (2016). „Modeling spatiotemporal information generation“. In: *Int. J. Geogr. Inf. Sci.* 30.10, pp. 1980–2008 (cit. on p. 211).

Scheider, Simon, Jim Jones, Albert Sanchez, and Carsten Keßler (2014). „Encoding and querying historic map content“. In: *The 17th AGILE International Conference on Geographic Information Science - Connecting a Digital Europe Through Location and Place*. Ed. by Joaquín Huerta, Sven Schade, and Carlos Granell. Castellón, Spain, pp. 251–273 (cit. on p. 169).

Scheider, Simon and Werner Kuhn (2015). „How to talk to each other via computers - Semantic interoperability as conceptual imitation“. In: *Applications of Conceptual Spaces*. Ed. by Frank Zenker and Peter Gärdenfors. Springer International Publishing, pp. 97–122 (cit. on p. 101).

Schmidt, Birgit, Birgit Gemeinholzer, and Andrew Treloar (2016). „Open data in global environmental research: The belmont forum's open data survey“. In: *PLOS ONE* 11.1. Ed. by Robert Guralnick, e0146695 (cit. on p. 174).

Schmidt, Birgit, Birgit Gemeinholzer, Andrew Treloar, et al. (2016). „Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey“. In: *PLOS ONE* 11.1. Ed. by Robert Guralnick, e0146695 (cit. on pp. 53, 70, 73, 74).

Schnur, Susan, Kenan Bektaş, and Arzu Çöltekin (2018). „Measured and perceived visual complexity: a comparative study among three online map providers“. In: *Cartography and Geographic Information Science* 45.3, pp. 238–254 (cit. on p. 222).

Secor, Anna J (2010). „Social surveys, interviews, and focus groups“. In: *Research Methods in Geography: A Critical Introduction*. Ed. by Basil Gomez and John Paul Jones III. John Wiley & Sons, pp. 194–205 (cit. on pp. 9, 10).

See, L, S Fritz, C Perger, et al. (2015). „Harnessing the power of volunteers, the internet and Google Earth to collect and validate global spatial information using Geo-Wiki“. In: *Technol. Forecast. Soc. Chang.* 98, pp. 324–335 (cit. on p. 35).

Segel, Edward and Jeffrey Heer (2010). „Narrative visualization: Telling stories with data“. In: *IEEE Transactions on Visualization and Computer Graphics* 16.6, pp. 1139–1148 (cit. on p. 169).

Shadbolt, Nigel, Kieron O'Hara, Tim Berners-Lee, et al. (2012). „Linked open government data: lessons from Data.gov.uk“. In: *IEEE Intelligent Systems* 27.3, pp. 16–24 (cit. on pp. 7, 8).

Shneiderman, Ben (1996). „The eyes have it: a task by data type taxonomy for information visualizations“. In: *Proceedings of the IEEE Symposium on Visual Languages*. Boulder, Colorado, USA: IEEE Computer Society Press, pp. 336–343 (cit. on p. 169).

– (2007). „Creativity support tools: accelerating discovery and innovation“. In: *Communications of the ACM* 50.12, pp. 20–32 (cit. on p. 213).

Silva, W M da, A Alvaro, G Tomas, et al. (2013). „Smart cities software architectures: a survey“. In: *In Proceedings of the 28th Annual ACM Symposium on Applied Computing, Coimbra, Portugal*, pp. 1722–1727 (cit. on p. 22).

SmartCitiesWeek (2015). *Accelerating Smart City Deployments: Challenges, Competitions and Collaborations from Smart Cities Around the Globe and Across the Region*. http://www.smartcitiesweek.com/Pages/Detail/17254 (accessed on 26 October 2015) (cit. on p. 27).

Smelcer, John B. and Erran Carmel (1997). „The effectiveness of different representations for managerial problem solving: comparing tables and maps“. In: *Decision Sciences* 28.2, pp. 391–420 (cit. on p. 203).

Solove, D (2006). „A taxonomy of privacy“. In: *Law Rev.* 154.3, pp. 477–564 (cit. on p. 31).

Spano, Lucio Davide and Ludovico Boratto (2019). „Advances in computer-human interaction for recommender systems (AdCHIReS)“. In: *International Journal of Human-Computer Studies* 121, pp. 1–3 (cit. on p. 263).

Stasch, Christoph, Simon Scheider, Edzer Pebesma, and Werner Kuhn (2014a). „Meaningful spatial prediction and aggregation“. In: *Environ. Model. Softw.* 51, pp. 149–165 (cit. on p. 37).

– (2014b). „Meaningful spatial prediction and aggregation“. In: *Environmental Modelling & Software* 51, pp. 149–165 (cit. on p. 213).

Steinitz, Carl (2012). *A Framework for Geodesign: Changing Geography by Design*. Redlands, CA, USA: ESRI Press (cit. on p. 25).

Stevens, S S (1946). „On the theory of scales of measurement“. In: *Science* 103.2684, pp. 677–680 (cit. on pp. 211, 212, 221, 224, 225, 243).

Stylos, Jeffrey and Brad Myers (2007). „Mapping the space of API design decisions“. In: *Proceedings of the IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC 2007)*. Coeur d'Alène, Idaho, USA: IEEE, pp. 50–57 (cit. on p. 101).

Sunlight Foundation (2010). *Ten Principles For Opening Up Government Information : Sunlight Foundation* (cit. on p. 61).

Susha, Iryna, Anneke Zuiderwijk, Yannis Charalabidis, Peter Parycek, and Marijn Janssen (2015). „Critical factors for open data publication and use: A comparison of city-level, regional, and transnational cases“. In: *eJournal of eDemocracy and Open Government* 7.2, pp. 94–115 (cit. on pp. 100, 117, 152).

Tamayo, Alain, Carlos Granell, and Joaquín Huerta (2012). „Using SWE Standards for Ubiquitous Environmental Sensing: A Performance Analysis“. In: *Sensors* 12.12, pp. 12026–12051 (cit. on p. 38).

Tamayo, Alain, Joaquín Huerta, Carlos Granell, Laura Díaz, and Ricardo Quirós (2009). „gvSOS: A new client for OGC Sensor Observation Service interface standard“. In: *Trans. GIS* 13.s1, pp. 46–61 (cit. on p. 38).

Tao, Wang (2013). „Interdisciplinary urban GIS for smart cities: advancements and opportunities“. In: *Geo-spatial Information Science* 16.1, pp. 25–34 (cit. on p. 166).

Taylor, Alex S. (2009). „Machine intelligence“. In: *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, pp. 2109–2118 (cit. on p. 210).

The Open Definition (2016). *The Open Definition* (cit. on pp. 51, 54).

The World Wide Web (2017). „Open Data Barometer Global Report Fourth Edition“. In: p. 36 (cit. on p. 79).

The World Wide Web Foundation (2015). *Open Data Barometer - Global Report*. 3rd. The World Wide Web Foundation, pp. 1–47 (cit. on pp. 51, 52, 100, 103, 120, 122, 180).

Thevenin, David and Joëlle Coutaz (1999). „Plasticity of user interfaces: framework and research agenda“. In: *Proceedings of INTERACT '99*, pp. 110–117 (cit. on p. 215).

Thurmond, Veronica A. (2001). „The point of triangulation“. In: *Journal of Nursing Scholarship* 33.3, pp. 253–258 (cit. on p. 11).

Tomitsch, Martin, Cara Wrigley, Madeleine Borthwick, et al. (2018). *Design. Think. Make. Break. Repeat. A Handbook of Methods*. BIS Publishers, p. 208 (cit. on pp. 9, 10).

Townsend, Anthony (2013). *Smart cities: big data, civic hackers, and the quest for a new utopia*. New York, NY, USA: WW Norton & Company (cit. on p. 28).

Trilles, Sergio, Alejandro Lujan, Oscar Belmonte, et al. (2015). „SEnviro: A Sensorized Platform Proposal Using Open Hardware and Open Standards“. In: *Sensors* 15.3, pp. 5555–5582 (cit. on p. 38).

Tsou, Ming-Hsiang (2011). „Revisiting web cartography in the United States: the rise of user-centered design“. In: *Cartography and Geographic Information Science* 38.3, pp. 250–257 (cit. on p. 221).

Tucci, Michele, Alberto Giordano, and Rocco W. Ronza (2010). „Using spatial analysis and geovisualization to reveal urban changes: Milan, Italy, 1737–2005“. In: *Cartographica: The International Journal for Geographic Information and Geovisualization* 45.1, pp. 47–63 (cit. on p. 184).

Tullis, Thomas and Jacqueline Stetson (2004). „A comparison of questionnaires for assessing website usability“. In: *Proceedings of the Usability Professionals Association (UPA) 2004 Conference*, pp. 7–11 (cit. on p. 155).

Tversky, Amos and Daniel Kahneman (1992). „Advances in prospect theory: Cumulative representation of uncertainty“. In: *J. of Risk and Uncertain.* 5.4, pp. 297–323 (cit. on p. 32).

Ubaldi, Barbara (2013). „Open government data: towards empirical analysis of open government data initiatives“. In: 22 (cit. on pp. 4, 52, 54, 56, 61, 79, 85, 86, 182).

Umpfenbach, K (2014). *Influences on consumer behavior-Policy implications beyond nudging*. http://ec.europa.eu/environment/enveco/economics_policy/pdf/Behaviour%20Policy%20Brief.pdf (accessed on 11 January 2016) (cit. on p. 32).

Usländer, T, A J Berre, Carlos Granell, et al. (2013). „The future internet enablement of the environment information space“. In: *Environmental Software Systems. Fostering Information Sharing, Neusiedl am See, Austria*. Springer, pp. 109–120 (cit. on p. 24).

Uttal, D H, D I Miller, and N S Newcombe (2013). „Exploring and enhancing spatial thinking links to achievement in science, technology, engineering, and mathematics?“ In: *Curr. Dir. Psychol. Sci.* 22.5, pp. 367–373 (cit. on p. 35).

Van den Bergh, J and S Viaene (2015). „Key Challenges for the Smart City: Turning Ambition into Reality“. In: *In Proceedings of the 48th International Conference on System Sciences, Hawaii*, pp. 2385–2394 (cit. on pp. 27, 28).

Van Ho, Q., P. Lundblad, T. Astrom, and M. Jern (2012). „A web-enabled visualization toolkit for geovisual analytics“. In: *Information Visualization* 11.1, pp. 22–42 (cit. on p. 223).

Vandenbussche, Pierre-Yves, Ghislain A. Atemezing, María Poveda-Villalón, and Bernard Vatant (2016). „Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web“. In: *Semantic Web* 8.3. Ed. by Michel Dumontier, pp. 437–452 (cit. on p. 148).

Veljković, Nataša, Sanja Bogdanović-Dinić, and Leonid Stoimenov (2014). „Benchmarking open government: An open data perspective“. In: *Government Information Quarterly* 31.2, pp. 278–290 (cit. on pp. 52, 152).

Vermesan, O and P Friess (2014). *Internet of Things Applications-From Research and Innovation to Market Deployment*. Aalborg, Denmark: The River Publishers (cit. on p. 38).

Vessey, Iris (1991). „Cognitive fit: A theory-based analysis of the graphs versus tables literature“. In: *Decision Sciences* 22.2, pp. 219–240 (cit. on p. 202).

Vetrò, Antonio, Lorenzo Canova, Marco Torchiano, et al. (2016). „Open data quality measurement framework: Definition and application to Open Government Data“. In: *Government Information Quarterly* 33.2, pp. 325–337 (cit. on pp. 52, 152).

Videla, Alvaro (2017). „Metaphors we compute by“. In: *Communications of the ACM* 60.10 (cit. on p. 213).

Vinod Kumar, T M (2014). *Geographic Information System for Smart Cities*. Copal Publishing Group (cit. on p. 32).

Visser, Sent and John Paul Jones III (2010). „Measurement and interpretation“. In: *Research Methods in Geography: A Critical Introduction*. Ed. by Basil Gomez and John Paul Jones III. John Wiley & Sons, pp. 41–59 (cit. on p. 10).

Vlacheas, P, R Giaffreda, V Stavroulaki, et al. (2013). „Enabling smart cities through a cognitive management framework for the internet of things“. In: *IEEE Commun. Mag.* 51.6, pp. 102–111 (cit. on p. 22).

Vredenburg, Karel, Ji-Ye Mao, Paul W. Smith, and Tom Carey (2002). „A survey of user-centered design practice“. In: *Proceedings of the SIGCHI conference on Human factors in computing systems Changing our world, changing ourselves - CHI ’02*. Ed. by Dennis R. Wixon. 1. Minneapolis, Minnesota, USA: ACM Press, p. 471 (cit. on p. 5).

Wakabayashi, Y and T Ishikawa (2011). „Spatial thinking in geographic information science: a review of past studies and prospects for the future“. In: *Proced. Soc. Behav. Sci.* 21, pp. 304–313 (cit. on p. 35).

Wang, Fahui (2005). *Geographic Information Systems and Crime Analysis*. Hershey, PA, USA: Idea Group Inc (cit. on p. 41).

Wang, Hui-Ju and Jin Lo (2016). „Adoption of open government data among government agencies“. In: *Government Information Quarterly* 33.1, pp. 80–88 (cit. on pp. 55, 70, 73).

Wang, Lili, Stuart Bretschneider, and Jon Gant (2005). „Evaluating web-based e-government services with a citizen-centric approach“. In: *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. Big Island, Hawai, USA: IEEE Computer Society, 129b (cit. on pp. 149, 203).

Whetten, David A. (1989). „What constitutes a theoretical contribution?“ In: *Academy of Management Review* 14.4, pp. 490–495 (cit. on p. 202).

Wijk, Jarke J van (2005). „The value of visualization“. In: *IEEE Visualization 2005*. Minneapolis, Minnesota, USA: IEEE, pp. 79–86 (cit. on p. 169).

Wijk, Jarke J. van (2013). „Evaluation: A challenge for visual analytics“. In: *Computer* 46.7, pp. 56–60 (cit. on pp. 167, 170).

Wijnhoven, Fons, Michel Ehrenhard, and Johannes Kuhn (2015). „Open government objectives and participation motivations“. In: *Government Information Quarterly* 32.1, pp. 30–42 (cit. on pp. 52, 55, 152, 182, 262).

Williamson, Vanessa and Norman Eisen (2016). „The Impact of Open Government: Assessing the Evidence“. In: *Brookings Institution* Developmen.December, pp. 1–30 (cit. on pp. 55, 62).

Winter, Bodo (2013). *Linear models and linear mixed effects models in R with linguistic applications*. Tech. rep. (cit. on pp. 11, 238).

Wobbrock, Jacob O. and Julie A. Kientz (2016). „Research contributions in human-computer interaction“. In: *Interactions* 23.3, pp. 38–44 (cit. on p. 12).

Wolff, Annika, Daniel Gooch, Umar Mir, Jose Cavero, and Gerd Kortuem (2015). „Removing barriers for citizen participation to urban innovation“. In: *Digital Cities 9 Workshop* (cit. on p. 149).

Wongsuphasawat, Kanit, Dominik Moritz, Anushka Anand, et al. (2016). „Voyager: exploratory analysis via faceted browsing of visualization recommendations“. In: *IEEE TVCG* 22.1, pp. 649–658 (cit. on p. 214).

World Wide Web Consortium (1995). *Logging Control In W3C httpd* (cit. on p. 118).

Yan, An and Nicholas Weber (2018). „Mining open government data used in scientific research“. In: *Transforming Digital Worlds - 13th International Conference (iConference 2018)*. Ed. by Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett. Sheffield, United Kingdom: Springer International Publishing AG, pp. 303–313 (cit. on p. 5).

Yang, Fan and Fuyuan Shen (2017). „Effects of web interactivity - a meta analysis“. In: *Communication Research*, p. 009365021770074 (cit. on p. 203).

Yang, Tung-Mou, Jin Lo, and Jing Shiang (2015). „To open or not to open? Determinants of open government data“. In: *Journal of Information Science* 41.5, pp. 596–612 (cit. on pp. 56–59, 70, 73).

Yin, ChuanTao, Zhang Xiong, Hui Chen, et al. (2015). „A literature survey on smart cities“. In: *Science China Information Sciences* 58.10, pp. 1–18 (cit. on pp. 20, 21, 24, 139, 166, 179).

Yingjie, Wang, Liu Yue, Chen Xiaogang, Chen Yufen, and Liqiu Meng (2001). „Adaptive geovisualization: an approach towards the design of intelligent geovisualization systems“. In: *Journal of Geographical Sciences* 11.S1, pp. 1–8 (cit. on pp. 216, 223).

Young, Andrew and Stefaan Verhulst (2016). *The global impact of open data - key findings from detailed case studies around the world.* 1st. O'Reilly Media, Inc, p. 452 (cit. on pp. 57, 71, 82, 151).

Yu, Harlan and David G. Robinson (2012). „The new ambiguity of 'open government"'. In: *UCLA Law Review* 59.178 (cit. on pp. 4, 54, 57, 210).

Zaman, A U and S Lehmann (2011). „Challenges and opportunities in transforming a city into a "zero waste city"". In: *Challenges* 2.4, pp. 73–93 (cit. on p. 23).

Zanella, A, N Bui, A Castellani, L Vangelista, and M Zorzi (2014). „Internet of things for smart cities". In: *IEEE IoT Journal* 1.1, pp. 22–32 (cit. on p. 22).

Zavala-Romero, Olmo, Eric P. Chassignet, Jorge Zavala-Hidalgo, et al. (2014). „OWGIS 2.0: open source Java application that builds web GIS interfaces for desktop and mobile devices". In: *ACM SIGSPATIAL 2014*, pp. 311–320 (cit. on p. 212).

Zaveri, Amrapali, Shima Dastgheib, Chunlei Wu, et al. (2017). „smartAPI: towards a more intelligent network of web APIs". In: *The Semantic Web - 14th International Conference (ESWC 2017)*. Ed. by Eva Blomqvist, Diana Maynard, Aldo Gangemi, et al. Portorož, Slovenia: Springer International Publishing, pp. 154–169 (cit. on p. 264).

Zibran, Minhaz F., Farjana Z. Eishita, and Chanchal K. Roy (2011). „Useful, but usable? Factors affecting the usability of APIs". In: *18th Working Conference on Reverse Engineering (WCRE 2011)*. Ed. by Martin Pinzger, Denys Poshyvanyk, and Jim Buckley. Lero, Limerick, Ireland: IEEE Computer Society, pp. 151–155 (cit. on p. 122).

Zuiderwijk, Anneke, Natalie Helbig, J.Ramón Gil-Garcia, and Marijn Janssen (2014). „Special issue on innovation through open data - A review of the state-of-the-art and an emerging research agenda: guest editors' introduction". In: *Journal of theoretical and applied electronic commerce research* 9.2, pp. 1–2 (cit. on p. 8).

Zuiderwijk, Anneke and Marijn Janssen (2012). „A comparison of open data policies and their implementation in two dutch ministries". In: *Proceedings of the 13th Annual International Conference on Digital Government Research (dg.o' 12)*. Ed. by John Carlo Bertot, Luis F. Luna-Reyes, and Sehl Mellouli. College Park, Maryland, USA: ACM, pp. 84–89 (cit. on pp. 6, 7).

– (2014). „The negative effects of open government data-investigating the dark side of open data". In: *In Proceedings of the 15th Annual International Conference on Digital Government Research, ACM,* pp. 147–152 (cit. on p. 26).

Zuiderwijk, Anneke, Marijn Janssen, Sunil Choenni, Ronald Meijer, and Roexsana Sheikh Alibaks (2012). „Socio-technical impediments of open data". In: *Electronic Journal of e-Government* 10.2, pp. 156–172 (cit. on pp. 6, 7, 74, 183, 220, 258, 259).

Zuiderwijk, Anneke, Marijn Janssen, and Chris Davis (2014). „Innovation with open data: Essential elements of open data ecosystems". In: *Information Polity* 19.1-2, pp. 17–33 (cit. on pp. 60, 61, 66, 87, 90).

Zuiderwijk, Anneke, Marijn Janssen, and Iryna Susha (2016). „Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators". In: *Journal of Organizational Computing and Electronic Commerce* 26.1-2, pp. 116–146 (cit. on pp. 6, 7).