Det åpne geodata-økosystemet

Atle Frenvik Sveen

Vitenskapelig bedømt (refereed) artikkel

Atle Frenvik Sveen: The Open Geospatial Data Ecosystem

KART OG PLAN, Vol. 77, pp. 108–120, POB 5003, NO-1432 Ås, ISSN 0047-3278

Open Governmental Data, Linked Open Data, Open Government, Volunteered Geographic Information, Participatory GIS, and Free and Open Source Software are all parts of The Open Geospatial Data Ecosystem. How do these data types shape what we define as Open Geospatial Data; Open Data of a geospatial nature? While all these areas are well described in the literature, there is a lack of a formal definition and exploration of the concept of Open Geospatial Data as a whole. A review of current research, case-studies, and real-world examples, such as OpenStreetMap, reveal some common features; governments are a large source of open data due to their historical role and as a result of political pressure on making data public, and the large role volunteers play both in collecting and managing open data and in developing open source tools. This article provides a common base for discussion. Open Geospatial data will be even more important as it matures and more governments and corporations release and use open data.

Keywords: Open Geospatial Data, Open Data, VGI, PPGIS, OpenStreetMap

Atle Frenvik Sveen, PhD-student, Norwegian University of Science and Technology, Department of Civil and Environmental Engineering, NO-7491 Trondheim

1. Introduction

Over the last decades there has been an increased focus on *open*, both in terms of *software* and *data*. The software industry was transformed by the concept of Free and Open Source Software (FOSS)[1] and in recent years Open *Data* has become increasingly important. However, where does geospatial data fit into this picture?

Geospatial data inhabits some unique characteristics that justifies being a separate group of data with its own definition. Specialized data formats, applications and the close ties to GISience (Goodchild, 2010) are some examples of these characteristics. Currently, *Open Geospatial Data* has no clear definition in literature, but this article aims to provide one.

Open Geospatial Data is simply Open Data of a geospatial nature. This nature is also shaping the environment where Open Geospatial Data is created, managed and used. We refer to this as the *Open Geospatial Data Ecosystem*. Politics, economics, law, software development, crowdsourcing, and data management are all niches in this ecosystem. While all these aspects have been examined separately, there is a need for a combined overview. The search here for a combined overview is structured in four paths: First, we explore the relation of Open Geospatial Data to the more general term of Open Data. Then we examine why governments are a major source of Open Data and see that geospatial data plays a major role among these. Third, we investigate the domain of Volunteered Geographic Information (VGI). In the fourth and final path, we look at Free and Open Source Software and the subset of Free and Open Source Software for Geospatial (FOSS4G) and explain how it relates to Open Geospatial Data.

By combing these paths, it is possible to survey the open geospatial ecosystem and provide the reader with a map that ensures safe navigation. While maps do not traditionally predict the future, the current trends allow us to suggest possible future

1. While the terms "Open Source" and "Free Software" (or "Free/Libre Software") carries slightly different meanings we use the term FOSS to cover the concept in general, noting the distinctions where necessary.

trajectories for this area in terms of both real-life applications and further research topics.

2. What is Open Data?

Open Data is, to describe it with an oxymoron, data that is open. However, what does *open* mean? And what is data? In their use, the terms often carry different meaning. There is, as such, a need to discuss their semantics. Open Data is a well-established term, defined in great length in the Open Definition. The most succinct version is "Open data and content can be freely used, modified, and shared by anyone for any purpose" (Open Knowledge International, 2016). To establish a better understanding of this, we will examine the full Open Definition in more detail. The Open Definition is divided in two major sections: *open works* and *open licenses*. The first section covers open works, and states that a work must satisfy four requirements to be considered to be open works:

1. Open License or Status: The work must be either in the public domain *or* provided under an open license.
2. The work must be provided as a whole and at no more than a reasonable one-time reproduction cost, and should be downloadable via the Internet without charge. Any additional information necessary for license compliance (such as names of contributors required for compliance with attribution requirements) must also accompany the work.
3. Machine Readability: The work must be provided in a form readily processable by a computer and where the individual elements of the work can be easily accessed and modified.
4. Open Format: The work must be provided in an open format. An open format is one which places no restrictions, monetary or otherwise, upon its use and can be fully processed with at least one FOSS tool.

This definition uses the terms data, content, and knowledge interchangeably, referring to them as *a work*. While there are established distinctions of these three terms (Rowley, 2007), in the context of defining Open Data (or open works) it makes sense to group them. We will however continue to use the term *data*.

The access requirement should be rather self-explanatory. For a work to be open, it must be accessible. An interesting and often overlooked aspect is the fact that the definition does not require the data to be available free of charge. In fact, costs related to the reproduction of the data can be charged. This is reminiscent of the practice of charging for postage and floppy disks for early FOSS software (Levy, 2001; Free Software Foundation, 2016). However, just as digital distribution through the Internet has rendered this an outdated practice; the Open Definition encourages digital distribution free of charge. In the Internet age access to data seems trivial, but as Conradie and Choenni (2014) found governmental bodies are often unsure as to where their data actually resides, complicating its release as Open Data.

The third and fourth requirements are closely related, but they cover different aspects. For data to be machine readable, the data has to be digital in the first place and stored in some format. The Open Definition does not impose any format requirements, meaning that both binary and text-based formats are accepted. One can argue that scanned documents are machine-readable using Optical Character Recognition-software, but such data does not allow for easy modification, and thus does not fulfill this requirement. What the machine readability requirement does not cover is the software needed to read or process the data. Data can be machine readable, but require specialized software to be read. The open format requirement serves to assure that there is no cost related to reading the data. This is done by making sure that there are existing FOSS tools that can read and process the data. Open formats that are governed by a standard is one way to ensure this, but the wording also allow for *de-facto* standards (Belleflamme, 2002) such as the well-known (in the geospatial domain at least) ESRI Shapefile format, a proprietary format from ESRI (ESRI, 1998) that is supported by several FOSS tools and considered a *de-facto* standard for geospatial vector data.

These requirements serve to ensure that the data can be accessed and modified. However, the idea that anyone can freely modify and use data raises concerns about trust. A data owners control of reliability and validity of the data may decrease due to a lack of proper interpretation by the user (Meijer *et al.*, 2014). Outright malicious redistribution, in which a third party modifies open data in order to gain advantages or cause damage, is also a concern (Geiger and Lucke, 2012).

The second section of the Open Definition deals with open licenses. The legalities concerning public domain and open licenses are complex fields in intellectual property law. A full description is beyond the scope of this article, but the reader is referred to Huang (2009) for a more thorough study. Herein, a general description will be given.

The term Open License may seem like an oxymoron. We usually encounter the term license in regards to software, where the license restricts how a piece of software can be used and shared. The FOSS movement reaction to this was to reverse the concept. The license, and thus the copyright, is used to grant the user more freedom, not less. These open software licenses are known as "copyleft" licenses, in reference to copyright licences. The freedoms these licenses grant is the right to use, modify, and redistribute computer source code for any purpose (Frantsvog, 2012). A wide range of open licences exists for software, the main difference being the distinction between "viral" licences that requires all modifications to be published using the same licence, and more relaxed licences that does not require this (Vetter, 2004). Since these licenses all are targeted at source code specifically they cannot be used for data. This means that specialized licences have to be created for Open Data. However, the open license section of the Open Definition closely resembles the definition of FOSS software licenses (Open Source Initiative, 2007). The term public domain is, as Huang argues, complex to define in terms of law, but a layman's definition is that a work in public domain is a work where the copyright is either expired, forfeited, or does not apply. The main difference between Public Domain data and data with an open license is thus the use of copyright; in public domain copyright is absent, while open licenses uses copyright to ensure openness.

Data that satisfies the four requirements of section one can be considered Open Data, but in general, there is a tendency to view data that is freely available on the Internet as Open Data. Other definitions of Open Data closely resembles the Open Definition in intention, if not in wording (Malamud *et al.*, 2007; Eaves, 2009).

3. Governments as a source of Open Data

Although The Open Data definition is not concerned about the source of the data, current literature seems to imply a strong connection between Open Data and government. This link is established through focus on the idea that Open Data acts as a conduit for citizen engagement with government (Sieber and Johnson, 2015), the importance of Open Data for public policy development (Janssen *et al.*, 2012), and the close linkage between Open Data and the Freedom of Information Movement and E-Government concept (Sieber and Johnson, 2015). The focus on governmental accountability and policy influence through Open Data is termed *Open Government* (Janssen *et al.*, 2012).

The focus on governments releasing their data (Public Sector Information (PSI) or Open Government Data (OGD)) is not unjustified. The motivations for this practice includes the ambition to increase democratic control and political participation, to foster service and product innovation, and the strengthening of law enforcement (Huijboom and Broek, 2011). The political interest in the topic is exemplified through initiatives such as EU Directive 2003/98/EC (Cox and Alemanno, 2003), the Obama administration's *Open Government Initiative* (Ginsberg, 2011), and similar initiatives in countries such as South Korea, Australia, and Chile (Yang and Kankanhalli, 2013). A McKinsey report from 2013 estimates that Open Data can enable $3 trillion in annual value, *e.g.* by creating more jobs (Manyika *et al.*, 2013). While several countries have performed case studies of Open Data, an assessment of the realized impacts and benefits of Open Data has not yet been performed (Koski and Tutkimuslaitos, 2015).

Another reason for the close linkage between governmental data and open data is the fact that public bodies are among the largest creators and collectors of data in many different domains (Janssen, 2011) and have a long history of collecting information for many purposes (Sieber and Johnson, 2015). The nature of this activity means that it is funded by public money. This makes a compelling case for the release of the results as Open Data. After all, why should taxpayers pay to get access to data that is produced using their tax money?

While the above discussion underlines the rationale for a close link between Open Data and government, this does not imply that all Open Data originates from governments. Data collected or produced by private citizens or corporations that adhere to the Open Definition should also be considered to be Open Data. The online encyclopedia; Wikipedia is created by volunteers through a crowdsourcing approach and available under the Creative Commons license (Wikipedia, 2017). Wikipedia is a prime example of Open Data generated by private citizens. The recent Movement initiative by the transportation company Uber promises to release anonymized data from over 2 billion trips using a Creative Commons license (Uber Technologies Inc., 2017) and is a good example of Open Data originating from a private corporation. These examples serve to illustrate that while Open Governmental Data is a large and important subset of Open Data it is not the only component in the ecosystem.

Another, closely linked, topic to Open Data is *Linked Data* or Linked Open Data (LOD). The main idea here is that data is published on the Internet in a linked and machine readable way. The links between data enable exploration and the connection of data in new and interesting ways. Technically, these links are recorded using Resource Description Framework (RDF) links (Bizer *et al.*, 2008). A premise for this linking of data is that the data is available on the web, making the connection to Open Data straightforward. Some argue that the linked data principles should be applied to Open Governmental Data in order to foster commercial re-use because unstructured and non-semantic data inhibits complex, machine-driven use and exploration of the data (Alani *et al.*, 2007).

4. Defining Open Geospatial Data

The discussion of Open Governmental Data focused on the division of Open Data by the *origin* of the data. A division can also be made based on the *type* of data. Geospatial data is one such type of data, leading to the definition of *Open Geospatial Data*: *Open Data of a geospatial nature*. Open Geospatial Data can often act as a "glue" to link other open datasets together (Correndo *et al.*, 2010), but it should be treated as a field of its own. As noted, Open Geospatial Data shares several traits with its superset Open Data. There are also a range of characteristics, issues, and problems specific to Open Geospatial Data. In the following section, we will explore these in detail.

4.1 Governments as a source of Open Geospatial Data

As with Open Data in general, a lot of the Open Geospatial Data available is released by governments. The production and management of geospatial data is a task that in many countries is considered to be a task for the public sector. Cadastral management, environmental agencies, infrastructure projects, and land development are just some examples of public sectors in need of geospatial data. Thus, many countries operate a national mapping agency (NMA), providing up-to-date maps and geospatial data, often in collaboration with administrative regions, other national agencies, and foreign states. Much of this data falls under the pressure for opening up public sector data, becoming Open Geospatial Data, or more specifically Open Governmental Geospatial Data (OGGD).

Geospatial data is among the largest and most important categories of Open Data released by governments. One example is found by examining the Global Open Data Index. This is an effort to track the openness of data from 122 countries around the world. Data is gathered from and categorized by seven categories (Open Knowledge International, 2017). Of these seven categories, there are three, which can be considered geospatial datasets: *National Map*, *Location Datasets*, and *Land Ownership*. These categories have an average openness percentage of 34, 22, and 18 %, respectively.

Another example is a survey of Spanish businesses that use Open Data to provide products or services. The survey found that in 2014, geospatial information was the most important sector, making up 35 % of the businesses (National Observatory of Telecommunications and the Information Society, 2014).

A third example can be made by examining some of the geospatial datasets that are released as OGGD around the world. One of the oldest, and perhaps best known, examples is the TIGER dataset from the US Census Bureau. This is a nation-wide, digital topographical dataset at the 1:100 000 scale based on digitization of paper maps from US Geological Survey and earlier digital map data from the Census Bureau (Marx, 1986; Sperling, 1995). This dataset was completed for the 1990 US census, but has been updated in the following years. The TIGER dataset was born as a collaboration between several US governmental organizations, and the data is shared with other organizations. Interestingly, the US Census Bureau also shares its data with the public "at the cost of dissemination" as a public resource (Sperling, 1995). This makes the TIGER dataset an early example of OGGD. Since the TIGER dataset is designed for internal purposes, the TIGER data was provided as extracts in the TIGER/Line format (Carbaugh and Marx, 1990). In recent years, the format was changed to the more familiar ESRI Shapefile format, and the data is now available for download on the Internet.

The release of the TIGER dataset is mainly credited to the US policy of keeping government data open as a consequence of *Public Domain* law (US Copyright Office, 2012). The case of OGGD in Denmark illustrates another aspect of Open Data. Denmark made almost all their geospatial data available as Open Data in 2013, expecting "a positive effect on the national economy, and that it will create growth and a more efficient public sector" (Deloitte, 2014), following the example of the UK, which released their topographical map data at a scale of 1:10 000 in 2010 (Arthur, 2010). Finland has plans to "open up all important data resources collected and maintained by the public administration by 2020", its digital topographic data being opened in 2012 (Koski and Tutkimuslaitos, 2015). In Norway, geospatial data has traditionally been exempt from the principle of free access to public data, but in 2013, a process of opening up datasets started, with plans to release all geospatial data in the following years, citing both transparency and economic benefits. An analysis estimates the economic benefit to be in the range 32 – 174 MNOK (3 – 18 million EUR) (Vennemo *et al.*, 2014). Table 1 shows an overview of open topographical datasets in the countries discussed above.

Table 1: Open topographical datasets in selected countries.

Country	Year released	Scale	Format(s)
USA	1990 [a]	1:100 000	ESRI Shapefile format, Geodatabase
UK	2010	1:10 000	GML, ESRI Shapefile format
Finland	2012	1:10 000	GML, ESRI Shapefile format, MIF, MAAGIS/XL
Denmark	2013	1:30 000 – 1:500 [b]	ESRI Shapefile format, Tabfile
Norway	2013	1:50 000	Filegeodatabase, SOSI, PostGIS-dump

a. Exact year uncertain.
b. This scale range is based on the similarity between the Danish FOT2007-data and the Norwegian FKB-data.

An important aspect of OGGD is the coverage and structure of the data. Governments typically create and maintain data sets that cover their own country, or parts of it. In addition to this, each country usually operates with their own data schemas, file formats, and semantics. This means that a user that wants to use data spanning multiple countries typically have to hunt for OGD from different governments, dealing with lacking data due to differing policies of data sharing, and handle the task of combining data from disparate sources. NMAs usually operate their own *Geoportals* (Maguire and Longley, 2005) that aims to let users discover their geodata. However, these portals use different layouts and concepts, and does necessarily not contain the same categories of data. As evident from the Global Open Data Index the types of data each country shares are not consistent. When data is obtained from multiple sources, the data has to be combined, which in many cases is a non-trivial task (Peng, 2005). Several attempts have been made to solve these issues and provide a "Geospatial One-Stop", the most ambitious being the European INSPIRE project (Goodchild *et al.*, 2007). The OpenStreetMap project, which will be discussed in the next section, also gathers geospatial data from several governmental sources, and thus have the potential to serve as a source of global geodata with a common format.

4.2 Volunteered Geographic Information and Participatory GIS

Another considerable source of Open Geospatial Data is Volunteered Geographic Information (VGI), a term coined by Goodchild (2007), considering it "a special case of the more general Web phenomenon of user-generated content". Goodchild notes that while this phenomenon may be traced back to the 16th century, the true enabler has been technological advancements in the late 20th and early 21st centuries. Web 2.0, broadband Internet, consumer GPS receivers, and the advancements in computer graphics are listed as enablers for this movement. While not all VGIs are Open Data according to the Open Data Definition, much of the VGIs comes with an Open Data license. Strobl and Nazarkulova (2014) argues that such open licenses is the only practical business model for VGIs due to the large contributor base. A prime example here is OpenStreetMap (OSM), a collaborative effort to create and maintain a database of world-spanning map data and make it available to the public using an open license (Haklay and Weber, 2008).

An interesting aspect of VGI is the sometimes blurred line between OGGD and VGI. Johnson (2017) explores different methods for "direct editing of government spatial data", the main question being how non-experts can contribute to the geospatial data that governments rely on. This idea is an application of the E-government concept applied to geospatial data, and is generally referred to in the literature as Participatory GIS (PGIS) or Public Participation GIS (PPGIS). Dunn (2007) argues that a clear definition of the term PGIS is even more elusive than the definition of GIS, but points to the integration of local and indigenous knowledge with "expert data". In particular, Dunn points to the inclusion of "the public" and marginalized groups.

While the distinction between VGI and PGIS/PPGIS may seem elusive, a rule of thumb for differentiating between them is that VGI is focusing on the creation or gathering of geospatial data while PGIS/PPGIS is more concerned about the use. In VGI the focus is on the actual creation or gathering of data, not on the ultimate use. Data is collected by non-experts on a volunteer basis. The use of the gathered data can vary from individual to individual and case to case, much like Free and Open Source Software. This is in contrast to PGIS/PPGIS which is more concerned with the eventual use of the data, using it as a means to influence political processes, closely resembling the Open Government philosophy.

An important note is that neither of these concepts require the gathered data to be Open Data. One example is map feedback solutions operated by commercial companies. The Google Map Maker project, now being merged into Google Maps, TomTom, and Here (formerly Navteq) are all examples of companies that employ VGI to let their users contribute feedback without releasing the result as Open Data (Google, 2017; HERE, 2017; TomTom, 2017). The primary concern of these users is to improve the map service they use, while the concept of Open Data may not even be relevant.

Several National Mapping Agencies (NMAs) are also applying this VGI strategy of user reports to improve their maps, often piping the reports through a review process. Olteanu-Raimond et al. (2016) points out that the practice of responding to error reports from users is by no means a new practice. The crucial difference is between an *ad hoc* process in which such reports are received through mail or telephone *versus* a more formal and digital process. When NMAs or other governmental bodies employ VGI strategies the gathered data may end up as Open Data if the legislation enforces a policy of Open Data. The blurred line between VGI and PGIS/PPGIS is also apparent in this situation, as citizens may well use these error reporting tools as a means to influence policies and legislation, not just to improve the map data.

Error reporting is just one example of how NMAs employ and explore the use of VGI and PGIS/PPGIS. A survey of 23 European NMAs found that 12 out of 23 agencies used VGI for tasks such as change detection, error reports, collection of new content, vernacular place names, and photo interpretation. Another six agencies had plans for using VGI (Olteanu-Raimond *et al.*, 2016).

Just as NMAs are employing VGI (and to some extent PGIS/PPGIS) the opposite is also true. OGGD canbe used to improve or "bootstrap" VGI projects by providing a baseline on which users can improve. OpenStreetMap used the US TIGER dataset as a tool for "bootstrapping" their coverage in the US by means of an automatic import to the OSM database in 2007/2008 (OSM Wiki, 2016b). While this import did improve the OSM coverage and served as a starting ground for further edits, it also caused a lot of problems and such large-scale, automated imports to OSM is now discouraged. The most promising alternative to these automated imports is known as the micro-tasking method. When applying micro-tasking to an OSM import, the source dataset is divided into smaller parts that are reviewed by volunteers in a crowdsourcing manner. This allows for fine-grained control and the possibility of merging geometries and attributes (Erichsen, 2016). A number of open datasets from various sources have been imported to OSM using either large-scale imports, smaller automated imports, or the micro-tasking method (OSM Wiki, 2016a).

This practice of imports exemplifies the blurred line between VGI and OGGD. It is also worth noting that the process works the other way as well, with NMAs applying VGI as discussed above, and the recent Change Within project from Mapbox, OSM, and New York City. The aim of Change Within is to monitor changes to OSM in New York. Any such changes will trigger a notification to the authorities. This enables the authorities to incorporate important changes in their city to their geospatial database (Johnson, 2017).

The symbiosis between VGI and OGGD works great for geospatial data where there is some kind of "ground truth"; where a user can edit the governmental data due to observed changes or poor initial quality. However, not all geospatial data has this "ground truth". Addresses is one example. In their essence, addresses are just conventions and rely on people agreeing to a given source of truth, usually the government in some form. For users to maintain a separate database of addresses and updating it based on observations in the field is a near impossible task. While street signs and house numbers are usually physically present, the truth is the data in the governmental database. Fortunately, addresses are increasingly released as Open Data, and the project OpenAddresses.io is an effort to solve the issue of collecting addresses from several countries by creating automated scripts that downloads, converts, and stores the data in a uniform way. This methodology is currently in its infancy but could provide a feasible route to importing geospatial data that is not directly surveyable. Addresses is the prime example, but land plots, voting districts, postal zones, and borders are all candidates for such an automated harvesting of OGGD.

4.3 Geospatial Data Formats

Most fields related to information gathering, storage, and distribution have a plethora of data formats to choose from. This variety is caused by a number of factors, such as different computer systems, types of data, history, and personal preference. Thus, the Open Definition just requires data to be available in a machine readable and open format, without further specification. A working draft from the W3C on publishing open governmental data mentions XML, RDF, and CSV, but also points out that "raw data is more likely to be produced using formats customized to the specific data, the tools used, or industry standards" (Bennett and Harvey, 2009). While geospatial data to some extent can be expressed as XML, RDF, or CSV it should be considered "raw data" and use customized open formats.

But what are these customized open formats? Geography Markup Language (GML) is an "XML application that provides a grammar and base vocabulary for describing geographic data", maintained by the Open Geospatial Consortium (OGC) (Burggraf, 2003). This should make it an obvious choice for publishing Open Geospatial Data. However, a survey of Open Geospatial Data in 10 cities around the world found that data is published in a variety of formats, ranging from 1 to 12 formats per city. ESRI Shapefiles, GeoJSON, and KML were common formats, but GML was also used (Seto and Sekimoto, 2015). Table 1 lists the formats used by the counties we examined in a previous section, and includes some national formats as well. Regardless of the distribution in these particular surveys, what is clear that there is no standard format used for sharing Open Geospatial Data. The reasons for this are many. One is that the data varies. Vector data formats cannot be used for sharing raster data, and while plain text formats work fine for relatively small datasets binary formats may be needed for larger datasets. In addition, there is a cost related to converting formats used internally in an organization. Another complexity, specific to geospatial data, is the coordinate system used. Power users might not have any problems transforming spatial data, but data in an unfamiliar coordinate system may pose a great obstacle for a novice.

4.4 Open Geospatial Data and licensing issues

The case of OSM can serve as an example of the juridical complexities concerning licenses of Open Data in general, and Open Geospatial Data in particular. OSM data was initially licensed using a variant of the Creative Commons (CC) license (specifically CC-BY-SA 2.0). After a community vote in 2012 the license was changed to the new Open Database License (ODbL) (Open Data Commons, 2009). The reason was that the CC license is based on copyright law and clearly protects items such as text and photographs. Data and map data, on the other hand, has a much less certain legal grounding in copyright law, especially in the US. The ODbL was created with databases such as OSM in mind, and should cover their use more clearly. The details that separate the different licenses and their applications is a complex area, and while the OSM example clearly shows that the CC license is not an ideal license for Open Data, several researchers seem to equate the two (such as Strobl and Nazarkulova, 2014).

The OSM license case highlights some important aspects of open geospatial data licensing. First is the fact that national laws differ. This means that licenses for global data, created by people from all over the world, and mixed with OGGD from different nations can be a legal nightmare. Second, copyright law regarding geospatial data and maps exhibits some peculiarities on its own. A printed map and a database of geospatial data is not the same in terms of law. The existence of a road is a (non-copyrightable) fact, while the depiction of the road on a map can be considered an artistic work, due to the cartographers' interpretation and "exercise of skill and judgement" (Judge and Scassa, 2010).

5. Free and Open Source Software

The Open Data Definition defines three basic freedoms of data: to use, to modify, and to share the data by anyone for any purpose. This definition is derived from the Open Source Definition (Open Source Initiative, 2007) with the main difference being the type of material it applies to; digital *data* for Open Data, computer software *source code* for Open Source. While these share a great deal of traits, there are differences that warrants separate definitions and licenses. One aspect is maturity; the concept of Free and Open Source Software is nearly as old as the software industry (Bretthauer, 2002), while Open Data is a fairly new phenomenon, compared to the era of digital data. Another aspect is the format itself; source code is by its nature available in digital form[2], while examples of data being stored as printed tables in a physical archive is rather common (Conradie and Choenni, 2014). A third differentiator is perhaps the producers; production of software (and thus source code) requires a certain degree of competence in programming, while data is being generated by a diverse set of professions, ranging from researchers, surveyors, and statisticians to plain number punching or automated sensors.

While FOSS is, strictly speaking, all software that adheres to the three freedoms the term has come to include other connotations as well. Chief among these is the open source development process described as the "Bazaar style". This process is characterized by "part-time hacking by several thousand developers scattered all over the planet, connected only by the tenuous strands of the Internet" (Raymond, 2001). Another aspect is the Free Software ideology championed by Richard Stallman; software is a commodity that should be free, and to use proprietary software (or non-free software) is wrong (Elliott and Scacchi, 2005; Williams, 2011). However, this latter aspect is not a belief shared by all who engage in FOSS, and is perhaps the main differentiator between the terms "Free Software" and "Open Source".

Free and Open Source software clearly has some defining traits besides the three freedoms. Does the field of Open Data exhibit links to other areas in the same way? We have already established that governmental data is a large source of Open Data, and that governments partly release their data due to an "Open Government" strategy or ideology. Thus, for some actors, Open Data can be considered a means to achieve the goal of social justice and governmental accountability. The Linked Data movement also has close ties to Open Data, making this another optional (but not defining) trait of Open Data.

The term FOSS4G is used to describe FOSS tools and projects that deal with geospatial data (Steiniger and Hunter, 2013). While not all FOSS4G developers are engaged in Open Geospatial Data or VGI, there is a large overlap between the two groups. This overlap can be seen both in terms of methods used, funding models, and tools shared.

The open source development process with users working part-time on tasks of their own choice, from locations around the world, and communicating via the Internet is a familiar process to OpenStreetMap . Research has shown that this large pool of contributors is beneficial for the quality of the data. This resembles the FOSS concept of Linus' Law: "Given enough eyeballs, all bugs are shallow". In other words, the fact that several developers inspect and use the source code serves as a quality assurance measure (Haklay *et al.*, 2010).

Another area where FOSS and VGI coincide is the funding model. The popular image of FOSS developers is that they are individual "hackers" working on projects on their spare time to "scratch an itch" (Raymond, 2001). In reality it is increasingly more common that developers work on FOSS as part of their paid job (Mockus *et al.*, 2002; Roberts *et al.*, 2006). This is also an emerging trend in VGI, again exemplified by OSM. While volunteers perform the largest share of the data creation and import to OSM, companies such as Mapbox (which use OSM for commercial purposes) employ a data team fixing errors and adding data to OSM. The "Mapbox Data Team" consists of about 40 persons (OSM Wiki, 2017), presumably working with OSM either fulltime or as part of their paid job. In addition to this general and ongoing effort, Mapbox and other commercial firms also have contributed to more focused data import such as the New York buildings import in 2014 (Barth, 2014) and the Los Angeles buildings import in 2016 (Schleuss *et al.*, 2016). These imports were carried out using the micro-tasking method, the import teams used both existing FOSS tools and created new ones.

The creation and re-purposing of existing FOSS tools to manage VGI projects is also an interesting aspect. McConchie (2015) use the term Hacker Cartography to describe the use of FOSS and repurposed tools and data to collaboratively create and curate crowdsourced geographic data. The ideas behind this movement is reminiscent of the "hacker ethics" still influential in the FOSS commu-

2. There are some exceptions to this rule, but today we consider punch-cards and printed source code to be more historical artifacts and curiosities than anything else.

nity (Levy, 2001). Hacker cartographers does not only create and manage geospatial data, they create "tools that can be used by others". These tools are often released as FOSS. One such tool is the Mapnik map renderer, used for generating map tiles from vector data. This tool was originally created as a FOSS4G component, and used early on by OSM, but has seen widespread use in several projects, both commercial and FOSS4G.

Thus, FOSS and Open Geospatial Data has several things in common. The most relevant is probably the fact that there is a feedback-loop between the two. First, a VGI project requires software. Then this software is created by FOSS4G-developers. The creation of such software might trigger the creation of new VGI projects again, and the loop continues.

6. Conclusions

Existing literature on the subject of Open Geospatial Data may seem scarce. As demonstrated in this article, this is not the case. VGI, PGIS/PPGIS, Hacker Cartography, and FOSS4G are all terms that are related to Open Geospatial Data and are discussed at length in the literature. Geospatial data is a major component of Open Data, a field that is studied in detail. What is missing from the body of knowledge is a more thorough understanding of the complete landscape of the Open Geospatial Ecosystem. The different components making up this ecosystem are all intertwined and need to be studied in a combined fashion.

What is clear is that several governments have already released Open Geospatial Data and that even more governmental institutions have plans to do so. As more governments release more data, the field will mature due to experience. What may at present be described as a somewhat idealistic and optimistic idea, will mature in the coming years. As more businesses are becoming dependent on Open Geospatial Data, the demands will increase. Release schedules, regularity, and, in general, data quality are all factors that may be overlooked in the early stages of opening data. A professionalization and knowledge of best practices of these aspects are needed for the field to support the large economy that several researchers and policy makers foresee. Demands for more standardization of formats, metadata, and catalogs will also increase, perhaps enforcing the Linked Data aspect of Open Data. Initiatives such as the EU INSPIRE directive is obviously also driving factors in this context.

Another consequence of the professionalization and increased use of Open Geospatial Data is the increased demand for more data. As end users start using released data they will discover that other, perhaps unreleased, datasets will enhance their products and solutions. This process can trigger a ripple effect in that releasing one data set will increase demand for opening even more data sets. Predicting what data sets this will be is difficult. A better approach is to develop good systems for the public to signal interest in specific data sets.

Another future trend is the continuing symbiosis between VGI and Open Governmental Geospatial Data. Imports to OSM is nothing new, but the process has improved from the massive TIGER-import to the micro-tasking methods used for the more recent building imports. Today, all imports to OSM are considered one-time imports. When the governmental data is updated there are no mechanisms in place for gracefully updating OSM, where the original governmental data may have been edited by volunteers. This is a problem that poses interesting challenges and we will likely see developments here in the future. The Change Within project is an interesting example of using OSM to enrich governmental data. With continuous imports to OSM where the data is enriched and quality controlled by citizens and then fed back to government, we may end up with completely blurred lines between VGI and Open Governmental Data.

Open Geospatial Data is an emerging and interesting topic. While it bears close resemblance to the general field of Open Data there are some key characteristics that distinctly differentiate. Most important is the close link to VGI and PGIS/PPGIS and the bridge to FOSS4G provided by hacker cartographers. OpenStreetMap is a mature platform that serves as the basis for several commercial mapping businesses, but it is also a live testbed for exploring issues related to Open Geospatial Data. Due to the complete openness of OSM, this is an unprecedented research opportunity that will continue to contribute data in the following years.

References

Alani, H. *et al.* (2007) 'Unlocking the potential of public sector information with semantic web technology', in *The Semantic Web*. Springer, Berlin, Heidelberg, pp. 708–721. doi: 10.1007/978-3-540-76298-0_51.

Arthur, C. (2010) *Ordnance Survey launches free downloadable maps*, *The Guardian*. Available at: https://www.theguardian.com/technology/2010/apr/01/ordnance-survey-maps-download-free (Accessed: 23 February 2017).

Barth, A. (2014) *Importing 1 million New York City buildings and addresses*. Available at: http://www.openstreetmap.org/user/lxbarth/diary/23588 (Accessed: 22 February 2017).

Belleflamme, P. (2002) 'Coordination on formal vs. de facto standards: a dynamic approach', *European Journal of Political Economy*, 18(1), pp. 153–176. doi: 10.1016/S0176-2680(01)00073-8.

Bennett, D. and Harvey, A. (2009) *Publishing Open Government Data*. Available at: https://www.w3.org/TR/gov-data/ (Accessed: 21 February 2017).

Bizer, C. *et al.* (2008) 'Linked data on the web', *WWW2008 Workshop on Linked Data on the Web*. New York, New York, USA: ACM Press, pp. 1265–1266. doi: 10.1145/1367497.1367760.

Bretthauer, D. (2002) 'Open source software: A history', *Information Technology and Libraries*. American Library Association, 21(1).

Burggraf, D. S. (2003) 'Geography Markup Language', *Data Science Journal*. CODATA, 2004(3–1), pp. 178–204. doi: 10.2481/dsj.5.178.

Carbaugh, L. W. and Marx, R. W. (1990) 'The TIGER system: A census bureau innovation serving data analysts', *Government Information Quarterly*, 7(3), pp. 285–306. doi: 10.1016/0740-624X(90)90026-K.

Conradie, P. and Choenni, S. (2014) 'On the barriers for local government releasing open data', *Government Information Quarterly*, 31(SUPPL.1), pp. 10–17. doi: 10.1016/j.giq.2014.01.003.

Correndo, G. *et al.* (2010) 'Geographical service: A compass for the web of data', in *Proceedings of the WWW2010 Workshop on Linked Data on the Web (LDOW 2010)*. Raleigh, USA.

Cox, P. and Alemanno, G. (2003) 'Directive 2003/98/EC of the european parliament and of the council of 17 november 2003 on the re-use of public sector information', *Official Journal of the European Union*, 46, pp. 1–156.

Deloitte (2014) *The impact of the open geographical data*. Available at: http://eng.gst.dk/media/gst/2626131/GST The impact of the open geographical data.pdf.

Dunn, C. E. (2007) 'Participatory GIS – a people's GIS?', *Progress in Human Geography*, 31(5), pp. 616–637. doi: 10.1177/0309132507081493.

Eaves, D. (2009) *The Three Laws of Open Government Data*. Available at: https://eaves.ca/2009/09/30/three-law-of-open-government-data/ (Accessed: 21 February 2017).

Elliott, M. S. and Scacchi, W. (2005) 'Free Software Development: Cooperation and Conflict in A Virtual Organizational Culture', in Koch, S. (ed.) *Free/open source software development*. Idea Group Publishing.

Erichsen, A. S. S. (2016) *Evaluation of the Micro-Tasking Method for Importing High-Detail Building Models to OpenStreetMap*. Trondheim.

ESRI (1998) *ESRI Shapefile Technical Description*. Available at: https://www.esri.com/library/whitepapers/pdfs/ shapefile.pdf.

Frantsvog, D. A. (2012) 'All Rights Reversed: A Study Of Copyleft, Open-Source, And Open-Content Licensing', *Contemporary Issues In Education Research*, 5(1), pp. 15–22.

Free Software Foundation (2016) *Selling Free Software*. Available at: https://www.gnu.org/philosophy/selling.html (Accessed: 21 February 2017).

Geiger, C. P. and Lucke, J. von (2012) 'Open Government and (Linked) (Open) (Government) (Data)', *E-Journal of Democracy*. Danube University of Krems, 4(2), pp. 265–278. Available at: http://jedem.org/index.php/jedem/article/view/143 (Accessed: 27 April 2017).

Ginsberg, W. R. (2011) *Obama Administation's Open Government Initiative: Issues for Congress*. Congressional Research Service. Available at: https://fas.org/sgp/crs/secrecy/R41361.pdf.

Goodchild, M. F. (2007) 'Citizens as sensors: The world of volunteered geography', *GeoJournal*, 69(4), pp. 211–221. doi: 10.1007/s10708-007-9111-y.

Goodchild, M. F. *et al.* (2007) 'Sharing Geographic Information: An Assessment of the Geospatial One-Stop', *Annals of the Association of American Geographers*. Taylor & Francis Group, 97(2), pp. 250–266. doi: 10.1111/j.1467-8306.2007.00534.x.

Goodchild, M. F. (2010) 'Twenty years of progress: GIScience in 2010', *Journal of Spatial Information Science*, 1(1), pp. 3–20. doi: 10.5311/JOSIS.2010.1.2.

Google (2017) *Google Map Maker graduates to Google Maps*. Available at: https://support.google.com/mapmaker/answer/7195127?hl=en (Accessed: 9 February 2017).

Haklay, M. (Muki) M. (Muki) *et al.* (2010) 'How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information', *The Cartographic Journal*. Taylor & Francis, 47(4), pp. 315–322. doi: 10.1179/000870410X12911304958827.

Haklay, M. and Weber, P. (2008) 'OpenStreetMap: User-Generated Street Maps', *IEEE Pervasive Computing*, 7(4), pp. 12–18. doi: 10.1109/MPRV.2008.80.

HERE (2017) *Welcome to Map Creator*. Available at: https://mapcreator.uservoice.com/knowledgebase/articles/260477-welcome-to-map-creator (Accessed: 20 April 2017).

Huang, H. (2009) 'On public domain in copyright law', *Frontiers of Law in China*. Springer, 4(2), pp. 178–195.

Huijboom, N. and Broek, T. Van Den (2011) 'Open data: an international comparison of strategies', *European Journal of ePractice*, 12(March/April 2011), pp. 1–13. doi: 1988-625X.

Janssen, K. (2011) 'The influence of the PSI directive on open government data: An overview of recent developments', *Government Information Quarterly*, 28(4), pp. 446–456. doi: 10.1016/j.giq.2011.01.004.

Janssen, M. *et al.* (2012) 'Benefits, Adoption Barriers and Myths of Open Data and Open Government', *Information Systems Management*. Taylor & Francis Group, 29(4), pp. 258–268. doi: 10.1080/10580530.2012.716740.

Johnson, P. A. (2017) 'Models of direct editing of government spatial data: challenges and constraints to the acceptance of contributed data', *Cartography and Geographic Information Science*. Taylor & Francis, 44(2), pp. 128–138. doi: 10.1080/15230406.2016.1176536.

Judge, E. F. and Scassa, T. (2010) 'Intellectual property and the licensing of Canadian government geospatial data: An examination of GeoConnections' recommendations for best practices and template licences', *Canadian Geographer*. Blackwell Publishing Inc, 54(3), pp. 366–374. doi: 10.1111/j.1541-0064.2010.00308.x.

Koski, H. and Tutkimuslaitos, E. (2015) 'The Impact of open data – a preliminary study', *The Research Institute of the Finnish Economy (ETLA)*, 15b.

Levy, S. (2001) *Hackers: Heroes of the computer revolution*. Penguin Books New York.

Maguire, D. J. and Longley, P. A. (2005) 'The emergence of geoportals and their role in spatial data infrastructures', *Computers, Environment and Urban Systems*, pp. 3–14. doi: 10.1016/j.compenvurbsys.2004.05.012.

Malamud, C. *et al.* (2007) *8 Principles of Open Government Data*. Available at: https://public.resource.org/8_principles.html (Accessed: 21 February 2017).

Manyika, J. *et al.* (2013) 'Open data: Unlocking innovation and performance with liquid information', *McKinsey Global Institute*.

Marx, R. W. (1986) 'The TIGER system: Automating the geographic structure of the United States census', *Government Publications Review*, 13(2), pp. 181–201. doi: 10.1016/0277-9390(86)90003-8.

McConchie, A. (2015) 'Hacker cartography: Crowdsourced geography, openstreetmap, and the hacker political imaginary', *ACME*, 14(3), pp. 874–898.

Meijer, R. *et al.* (2014) 'Reconciling contradictions of open data regarding transparency, privacy, security and trust', *Journal of Theoretical and Applied Electronic Commerce Research*. Universidad de Talca, 9(3), pp. 32–44. doi: 10.4067/S0718-18762014000300004.

Mockus, A. *et al.* (2002) 'Two Case Studies of Open Source Software Development: Apache and Mozilla', *ACM Transactions on Software Engineering and Methodology (TOSEM)*. ACM, 11(3), pp. 309–346. doi: 10.1145/567793.567795.

National Observatory of Telecommunications and the Information Society (2014) *Characterization Study of the Infomediary Sector 2014*. Available at: http://datos.gob.es/sites/default/files/doc/file/executive_summary_public_infomediary_sector_2014.pdf.

Olteanu-Raimond, A. M. *et al.* (2016) 'The Scale of VGI in Map Production: A Perspective on European National Mapping Agencies', *Transactions in GIS*, pp. 74–90. doi: 10.1111/tgis.12189.

Open Data Commons (2009) *Open Database License (ODbL) v1.0*. Available at: https://opendatacommons.org/licenses/odbl/1.0/ (Accessed: 20 April 2017).

Open Knowledge International (2016) *Open Definition 2.1*. Available at: http://opendefinition.org/od/2.1/en/ (Accessed: 12 January 2017).

Open Knowledge International (2017) *Global Open Data Index*. Available at: http://index.okfn.org (Accessed: 12 January 2017).

Open Source Initiative (2007) *The Open Source Definition*. Available at: https://opensource.org/osd (Accessed: 2 February 2017).

OSM Wiki (2016a) *Import / Catalogue*. Available at: http://wiki.openstreetmap.org/w/index.php?title=Import/Catalogue&oldid=1344831 (Accessed: 27 September 2016).

OSM Wiki (2016b) *Tiger*. Available at: http://wiki.openstreetmap.org/w/index.php?title=TIGER&oldid=1338304 (Accessed: 21 February 2017).

OSM Wiki (2017) *Mapbox*. Available at: http://wiki.openstreetmap.org/w/index.php?title=Mapbox&oldid=1422004 (Accessed: 22 February 2017).

Peng, Z. ? R. (2005) 'A proposed framework for feature? level geospatial data sharing: a case study for transportation network data', *International Journal of Geographical Information Science*. Taylor & Francis Group, 19(4), pp. 459–481. doi: 10.1080/13658810512331319127.

Raymond, E. S. (2001) 'The Cathedral and the Bazaar', in *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*. O'Reilly, pp. 19–65. Available at: https://people.eecs.berkeley.edu/~kubitron/cs162/hand-outs/cathedral-bazaar.pdf (Accessed: 27 September 2016).

Roberts, J. A. *et al.* (2006) 'Understanding the Motivations, Participation, and Performance of Open Source Software Developers: A Longitudinal Study of the Apache Projects', *Management Science*. INFORMS , 52(7), pp. 984–999. doi: 10.1287/mnsc.1060.0554.

Rowley, J. (2007) 'The wisdom hierarchy: representations of the DIKW hierarchy', *Journal of Information Science*. Sage PublicationsSage CA: Thousand Oaks, CA, 33(2), pp. 163–180. doi: 10.1177/0165551506070706.

Schleuss, J. *et al.* (2016) 'Let's get LA on the map!: The Los Angeles Building Import Case Study', in *State Of The Map US*. Seattle. Available at: http://stateofthemap.us/2016/lets-get-la-on-the-map/.

Seto, T. and Sekimoto, Y. (2015) 'Comparing the Distribution of Open Geospatial Information between the Cities of Japan and Other Countries', in Ferreira, J. and Goodspeed, R. (eds) *CUPUM 2015*. Cambridge, Massachusetts. Available at: http://web.mit.edu/cron/project/CUPUM2015/proceedings.

Sieber, R. E. and Johnson, P. A. (2015) 'Civic open data at a crossroads: Dominant models and current challenges', *Government Information Quarterly*, 32(3), pp. 308–315. doi: 10.1016/j.giq.2015.05.003.

Sperling, J. (1995) 'Development and Maintenance of the TIGER Database: Experiences in Spatial Data Sharing at the U.S. Bureau of the Census', *Sharing geographic information*, pp. 377–396. Available at: http://geospatial-solutions.com/tiger-database-historical-perspective/.

Steiniger, S. and Hunter, A. J. S. (2013) 'The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption', *Computers, Environment and Urban Systems*, 39, pp. 136–150. doi: 10.1016/j.compenvurbsys.2012.10.003.

Strobl, J. and Nazarkulova, A. (2014) 'Open Geospatial Data: New Opportunities for GIS and GIScience in Central Asia?', in *Proceedings of the Annual Central Asia GIS Conference GISCA*, pp. 55–64.

TomTom (2017) *About Mapshare Reporter*. Available at: https://uk.support.tomtom.com/app/answers/detail/a_id/14911 (Accessed: 20 April 2017).

Uber Technologies Inc. (2017) *Uber Movement FAQ*. Available at: https://movement.uber.com/faqs (Accessed: 12 January 2017).

US Copyright Office (2012) *Copyright Law of the United States and Related Laws Contained in Tıtle 17 of the United States Code*.

Vennemo, H. *et al.* (2014) *Verdien av gratis kart- og eiendomsdata*. Available at: https://www.regjeringen.no/globalassets/upload/kmd/plan/verdien-av-gratis-kart-og-eiendomsdata.pdf.

Vetter, G. R. (2004) 'Infectious open source software: Spreading incentives or promoting resistance', *Rutgers Law Journal*. HeinOnline, 36(53), pp. 53–162.

Wikipedia (2017) *Copyrights*. Available at: https://en.wikipedia.org/w/index.php?title=Wikipedia:Copyrights&oldid=743896666 (Accessed: 21 February 2017).

Williams, S. (2011) *Free as in Freedom: Richard Stallman's Crusade for Free Software*. O'Reilly Media, Inc.

Yang, Z. and Kankanhalli, A. (2013) 'Innovation in Government Services? : The Case of Open Data', *IFIP International Federation for Information Processing 2013*. Springer, Berlin, Heidelberg, pp. 644–651. doi: 10.1007/978-3-642-38862-0_47.