Threats Related to Open Geospatial Data in the Uncertain Geopolitical Environment

Jussi Nikander [1], Teemu Jama [1], Henrikki Tenkanen [1]

[1] Department of Built Environment, School of Engineering, Aalto University, Espoo, Finland - (jussi.nikander, teemu.jama, henrikki.tenkanen)@aalto.fi

Keywords: Information security, national security, open geospatial data, risk assessment

Abstract

During the last few decades, there has been increasing interest in open data, especially with regards to the public sector. Open data promotes transparency in the society, and the availability of data creates significant additional value. There are also potential threats associated with open data. After the Russian invasion of Ukraine, threats to society and national security related to open data have been given a lot more attention than before. Here, we conducted group interviews with a number of security experts in Finland to investigate potential threats related to open data and society. Based on the interviews, we created a number of threat scenarios that were further refined and discussed with experts. We recognized a large number of different potential malicious actors, ranging from hostile nations to individual criminals. The threats these actors might cause are varied, and for many, the best ways to mitigate the threat are not, in fact, related to open data themselves. In addition, changes to open datasets can have significant side effects that also need to be taken into account when considering how to manage the potential threats. Our study concludes that decisions to open new datasets, or to modify already existing ones, need to be made with care, and threat and risk assessment must always be weighted against benefits of publishing the data, as well as drawbacks of leaving the data unpublished.

1. Introduction

During the last few decades, there has been increasing interest in *open data*, especially with regards to the public sector. Public sector operates using taxpayer money, and therefore many feel that the data should be available for all without cost or limitations in how to use it. In addition, open data promotes transparency in the society, and the availability of data creates significant additional value for individuals, companies, and organizations who can easily use the data to create analyses, services, or business with it (Janssen et al., 2012). A strong belief in government transparency as one of the cornerstones of democracy spread in Europe and the United States during the 1990s. The background was influenced in particular by the end of the Cold War, which was followed by the establishment of the European Union (1992) in the wave of global democracy. As a central part of the EU's democratic objectives, the PSI Directive (Directive 2003/98/EC) was established that aimed to enable the wide and free re-use of public sector information. Similar developments also took place in the United States, where they developed a common infrastructure for the re-use of public sector information (Schrock, 2016). In 2009, the government-operated website data.gov was launched and in 2013 President Obama signed an executive order stating that open and machine-readable data production is the "default" for government data production. This led to the establishment of Open Data Project to give concrete expression to this policy in the United States (White House, 2013).

Finland, among many other European countries, has been a strong proponent for *open data* since the first years of the 21st century. Since 2010, a significant amount of public sector data has been published openly, and much of this data is *geospatial* by nature (Ahonen-Rainio et al., 2014). Accurate geospatial data with nation-wide coverage is highly valuable for many applications. It is used in logistics planning, environmental modeling, forestry planning, and in decision making on the municipal, national, and international scale, to list just a few application areas. In addition, such data is important in matters related to national security, as well as military applications.

In addition to the benefits of open data, there have always been *threats* associated with open geospatial data. When data is shared openly, it can be used by not only the actors with good intentions, but also criminals and other malicious actors, as well as hostile nations. Traditionally, threats related to open geospatial data have been divided into two broad categories: *threats to privacy* and *threats to society*. There are numerous national geospatial datasets that pose obvious threats to privacy, such as accurate census data on the level of individuals. Therefore, the public sector has developed mature best practices on how to handle privacy concerns, and there are also international guidelines to assess risks related to open data (Open Data Institute, 2022). For example, census or population registry data should never be published at an individual level, but the data should be aggregated to minimize the privacy risks.

After the Balkan wars of the 90s, the majority of Europe has been in a state of deep peace. Therefore, the potential *threats to society* related to open geospatial data have been given relatively little attention. Potential threats from other nation states have been sidelined by other concerns, and often dismissed as irrelevant due to increased European integration. This is true even in Finland, which never downsized her army or dismantled national preparedness organizations. However, the Russian invasion of Ukraine in February 2022 caused a rapid and radical change in the global geopolitical environment. In Finland, the potential threats to the nation became much more concrete, causing a radical shift in discussion about national security. Suddenly, voices about the threats caused by or related to open geospatial data were taken more seriously in the Finnish geospatial data ecosystem.

In this paper, we study the security concerns related to open geospatial data in Finland by conducting semi-structured interviews with experts from various Finnish organizations who are responsible in producing or using open geospatial data in their

operations, or who are experts in data-related threats to the Finnish geospatial ecosystem. The main research questions of the study are:

- What kinds of threats related to open geospatial data exist?
- How can the threat-related open geospatial data be mitigated and managed?

In this study, we focus on matters especially related to national security, which is currently an understudied area with significant relevance to society. We used the various national geospatial datasets maintained and produced by the National Land Survey of Finland (NLS), including e.g. the Finnish topographic database as an example. Even though our focus was on the data produced by NLS, many of our findings are applicable more generally, as our approach considers potential threats enabled by open geospatial data in general terms. Based on the interviews, we created a number of threat scenarios that were used as examples on what sorts of threats might be related to open geospatial data. The scenarios were then discussed and further refined with a number of experts on national security and the Finnish geospatial ecosystem.

2. Background

Before our project, open discussions regarding the need for threat assessment in the new geopolitical environment had already started within the Finnish geospatial ecosystem. This gave a useful basis for scoping our research, as well as provided an environment where the findings could be discussed.

2.1. Open data

In this work, we use the Open Data Institute (2023) definition for *open data*, meaning that data is considered open if it A) can be accessed by anyone, B) without requiring any sort of identification or registration, and C) is licensed using an open license, such as Creative Commons. Open Data Institute (ODI) characterizes open data in the *open data spectrum*, shown in Figure 1.

ODI roughly divides data into three categories according to how open the access to the data is: *closed data, shared data* and *open data.* Of these three categories, open data fulfills all three above mentioned requirements: open access, no need to identify access, and open license. In this work we follow this terminology, and thus use the term *shared data* for datasets that are freely available but do not fulfill all three requirements.

In Finland, the good availability of open data has led to widespread use of them, especially in the public sector. Open data is also used by the private sector, but there its effect is often indirect as the open data from the public sector is often not the primary means of providing added value to customers. Nevertheless, the Finnish geospatial data ecosystem would be considerably disrupted if open data availability would decrease significantly.

Due to changes in the geopolitical situation it is important to consider whether some of the currently open geospatial datasets would need to be modified, or even removed from open distribution in the future. The vast majority of the decisions to publish geospatial datasets as open data were done well before the year 2022. At the time, national security concerns were given secondary importance in data publication decisions. Therefore, the assumptions made regarding the *threats* and *risks* related to these datasets might no longer be true and need to be reassessed.

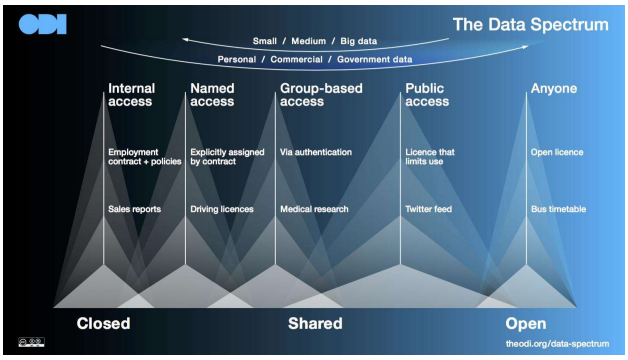

Figure 1. The Open Data Spectrum by Open data Institute. Originally created by ODI, used under CC-BY.

2.2. Threats

In this work we define *threat* as a potential unwanted event, action or development, where open geospatial data is involved. Similarly we define *risk* to be a combination of the probability of an unwanted event and the severity of the consequences of the said event. The higher the probability or the more severe the consequences, the bigger the risk.

There are many different ways to assess the threats or risks related to data. The Open Data Institute (2022), to give an example, divides risks into 4 categories: legal and regulatory, ethical, reputation and commercial risks. These categories are then divided further into more detailed questions that should be answered when an organization uses the ODI risk categories to map the risks related to a dataset. This sort of assessment is most likely the most useful for non-governmental organizations and companies. When risks and threats are assessed from governmental point of view, the potential threats are often related to either the privacy, health and wellbeing of citizens, or matters related to the orderly running of the nation. In addition to *threats to privacy* and *threats to society*, it is also possible to recognize a third category of threats: global threats such as climate change, large-scale warfare, or overpopulation (Ministry of Interior, 2023).

In Finland, the regional administration (cities, municipalities and regions) operates under the municipality law, which calls for openness both for the decisions and the decision making material (Eduskunta 2015). Thus, in order to keep a certain dataset closed, there should be an exceptional justification for this. Typically such justifications are based on threats the open sharing of the dataset might cause, either to privacy or to the society. The public sector in Finland is generally well aware of the threats to privacy that data may create, and therefore there are robust best practices available for assessing potential open publication of such datasets.

The guidelines according to which possible threats to society are assessed are typically related to national security. Therefore, the importance of these assessments has significantly increased after the beginning of the war in Ukraine. However, it should be noted that not all threats to society are related to national security or military matters, and not every threat is caused by malicious human actors. For example, the Finnish National Risk Assessment for 2023 contained 15 threat scenarios ranging from the use of military force to large-scale accidents or problems in

the logistics networks (Ministry of Interior, 2023). In this study, we mostly ignore the assessment of threats to privacy as they are out of scope for this paper. They are included in the assessment only when in relation to threats to the society.

3. Material and methods

We used semi-structured interviews as the primary research method in this research. We interviewed 20 individuals from 13 Finnish organizations. The majority of the interviewees were from public sector organizations and they represented organizations involved in areas such as national defense and emergency management, food production, rescue services, power and utilities, as well as personnel from appropriate ministries. After each interview the results were analyzed and new threats named and threat scenarios described, opinions and insight on threats related to specific datasets or open data as a concept were extracted. During the last few interviews, there were not many new insights to be gained. Thus, we concluded that we had reached the saturation point in terms of new information and no further interviews were needed.

The majority of the interviews were group interviews where 2-3 people were interviewed at the same time, but there were also some individual interviews. The interviews were arranged according to the speciality of the interviewees. Thus, for example, electrical network specialists were one interview, rescue services were one, as was food production, etc.

For all groups, there were 3 main questions presented to the participants. Each question covered a large topic, and thus the interviewers presented the participants with more detailed questions regarding the topics when required. Thus the initial questions were the starting point for a larger discussion. The three questions were:

1. Are open geospatial datasets related to the threats described in the Finnish national risk assessment?
2. Are there currently other relevant threats, where open geospatial data could contribute to the realization of the threat?
3. Are you aware of open geospatial datasets for which it would be prudent to reassess the open distribution of the data?

Based on the interview results, we identified relevant threats and wrote threat scenarios related to specific threads with specific open geospatial datasets. In addition, we identified information about how the quality of open geospatial data contributes to various threats. As a result, we were able to synthesize how open geospatial data contributes to various threats and what sort of threats are related to open geodata. The threat scenarios were then refined together with a panel of experts gathered specifically for the project. Members of the panel were geospatial data and national security experts from various Finnish organizations.

4. Results

The interviewees discussed many potential problems related to open geospatial data. One central conclusion was, in fact, that there are *no* fundamental problems related to open geospatial data. However, there are *potential* threats and risks related to every open geospatial dataset, and therefore appropriate risk assessment is crucial *before* publishing new open datasets. Removing data or changes in quality (accuracy) of already open datasets were considered useless, or even revealing in terms of security, which should be avoided. When national security is considered, it should be assumed that whatever data has been made openly available, malicious actors with sufficient resources have already acquired the data. Thus, if significant alterations are done to the data, malicious actors can compare previous versions of a dataset to new and modified ones and spot the differences.

In many geospatial datasets, there is a subset of data elements that contain information that may be sensitive. In addition, when data from two or more datasets are combined, it is possible to create new information that would be impossible to find from either source dataset alone. Therefore, for most geospatial datasets, the amount of potentially sensitive data is small, but it is difficult to make thorough assessment of potential threats.

In most of the threat scenarios discussed, open geospatial data could help malicious actors to plan and execute activities that cause harm. Based on our analysis, the threat related to a specific dataset most often did not directly target the publisher of the dataset, or affected the dataset itself. For example, detailed building data can be used to plan burglaries, and accurate road network and topographical data can be used to plan an armed invasion. Thus the target of the malicious activity is not the dataset itself, and the data is used as means to gain more information about these targets.

4.1. Data quality and usefulness

The usefulness of geospatial data, in this work defined as fitness of data for a specific purpose, is very closely related to the quality of the dataset. The data quality in this work is defined via the ISO spatial data quality standards. The current ISO standard is ISO 19157-1:2023 Geographic Information - Data Quality. Part of this discussion is, however, based on the older ISO quality standards 19113, 19114 and ISO/TS 19138. The reason for this is that the current Finnish national geospatial data quality standard is based on these older ISO standards (JUHTA, 2006). In the scope of this work, the differences are small enough not to matter.

The most important quality attributes related to potential risks and threats are *spatial precision* and *accuracy* and *attribute data quality*. Data is precise and accurate when the data location - data element coordinates and/or area covered by the data element - correspond to the location or area covered by the corresponding real world object or phenomenon. Attribute data in a spatial element is of high quality when its value describes the corresponding object or phenomenon in a manner that can be used to distinguish corresponding characteristics of the real world object or phenomenon from the same characteristic from other objects or phenomena. For example, the more detailed the classification of building types, and the better the building type attribute of building objects corresponds to the real-world buildings, the better the data quality is for this attribute.

When assessing data quality it must be taken into account that public sector actors in Finland have the responsibility to create high quality geospatial data. Thus, the attribute quality or spatial precision and accuracy of public sector datasets is typically very high, and public sector data often has guarantees on the data quality. Therefore, public sector data can be considered valuable for many purposes. However, the high data quality requirement may actually mean that the temporal quality of such data (i.e. how up-to-date it is) may suffer. In addition, datasets from the public sector actors that have national level responsibilities, such as the National Land Survey of Finland, provide valuable data since their datasets cover the whole country. This is in contrast to the municipalities, which may produce more precise or up-to-date

data, but the geographic coverage is typically limited to the area of the municipality.

In addition, easy availability of open data makes it useful for the private sector. It is rare to use public sector data as a core part of a business dataset. Instead, public sector data can be used to support the actual added value data and services, for example by providing a background map or reference data in the service. In addition, open data can also be used to easily test out new ideas and services. There is no need to buy data in order to build a quick prototype, if there are suitable open datasets available that can be used instead.

4.2. Data and threats

In the interviews, several different types of objects and phenomena were mentioned, which may contribute to different kinds of threats to society. Firstly, geospatial data often contains data that represents *critical infrastructure*, such as facilities and networks related to electricity, telecommunication, heat, or water management. Damage to critical infrastructure can cause widespread problems, and thus publication of data related to critical infrastructure requires careful consideration.

Geospatial datasets can include data about nationally or locally *important infrastructure*. Damage to important infrastructure is not as widespread as damage to critical infrastructure, but is vulnerable to similar kinds of threats and risk. A third category that can be found are *soft targets*, or locations where it is easy for malicious actors to create considerable effects. For example, schools can be targets of terror or revenge attacks. Geospatial data can also include *other* information that may be of interest to malicious actors. For example, out-of-the-way summer residences might be an interesting target for burglars outside the traditional summer holiday season.

Unfortunately, it can be clearly seen that the usefulness of geospatial datasets is not limited to just beneficial or wanted uses. Every dataset is *potentially* associated with a number of risks. Typically these risks are related to the data elements represented in the dataset somehow. For example, a dataset representing buildings can be used to find out summer residences, should the attribute data include this information. Then, the summer residences can be placed on a map and, for example, filtered by selecting only those residences that are sufficiently far away from other residential buildings to see locations that might be vulnerable outside the traditional summer holiday season.
However, it must also be taken into account that open geospatial data has tremendous benefits for many actors. Therefore, threat and risk assessment must always be weighted against benefits of publishing the data, as well as drawbacks of leaving the data unpublished. For example, the previous summer residence example describes a risk that should not be allowed to affect the publication of a national building dataset. Trying to hide the locations of summer residences is not a proper way to prevent potential burglaries. Instead, an owner who's worried about their property should avoid storing anything valuable at the summer residence and/or invest in good locks or an alarm system.

In general, proper risk assessment should always be one part of the process when publishing a geospatial dataset. There are many risk assessment frameworks, but in general the process consists of *recognizing*, *analyzing,* and *assessing the relevance* or risks related to a dataset. If the risks are sufficient, they need to be addressed and *managed* before the data can be published.

4.3. Risk assessment

There are different types of threats that are related to different datasets and to different real-world actors and organizations. Threats to privacy, for example, are primarily aimed at individual people, while threats to central critical infrastructure may have effects on the whole society. In risk assessment, the probability of an event and the possible consequences of it are assessed, and based on this, a risk level is established. The risks can be classified into different categories, shown in the risk assessment matrix in Figure 2.

Likehood of the event	Severity of the consequences		
	Minor	Moderate	Severe
Unlikely	Very low risk	Low risk	Moderate risk
Possible	Low risk	Moderate risk	High risk
Likely	Moderate risk	High risk	Intolerable risk

Figure 2. Risk assessment matrix.

For risk assessment, it is also important to consider how unique the data in question is. For example, modern earth observation technology and the existence of free global geospatial data products, such as Google Maps or OpenStreetMap, means that vast amounts of geospatial data that was once only available to nation states is now available also for many other actors. Furthermore, it means that it is significantly more difficult to keep things hidden than before. Therefore, many public sector geospatial datasets no longer contain information that would not be available from other data sources. For example, high-voltage electric lines tend to be aboveground. Thus, even if this information was removed or hidden in the data provided by the public sector, any actor with sufficient resources and technical expertise could find the information from satellite images, or even from global map data providers. For example, OpenStreetMap contains significant amounts of information regarding the electrical network. Thus hiding government data might not affect the risks related to the electrical network at all.

However, mere presence of the data in other data sources does not necessarily mean that the quality of the data is the same. For example, governmental sources may have more detailed attribute information regarding objects or phenomena than other open data sources. Such information may make the datasets much more useful for a range of purposes. Therefore, in addition to just comparing data elements to other existing data, their contents must also be compared. With this information, it is also easier to make assumptions about what different malicious actors could use the dataset for.

There are several ways the risks related to various datasets can be compared to one other. One way is to consider the *geographical extent* and *data quality* in datasets. The higher the extent and the quality, typically the bigger the potential risks. Sufficient geographical extent typically is required for a dataset to cause potential risks to the society, but even datasets with very limited extent may endanger the privacy of certain individuals. This is also depicted in Figure 3.

The Figure has two boxes, a yellow one that represents data that may endanger individual privacy and a red one that represents data that may endanger critical infrastructure and thus national security. In the Figure it is assumed that data of higher quality -

such as better resolution or more detailed attribute data - is typically required in order to endanger the privacy of an individual, than what is required in order to endanger critical infrastructure.

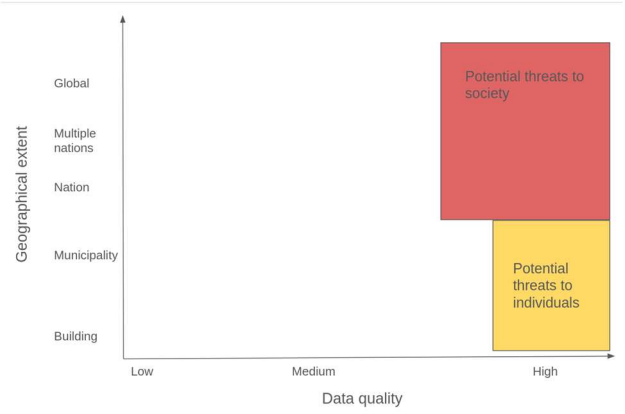

Figure 3. Data quality and extent compared to potential threats.

As can be seen in Figure 3, typically datasets that may contain sensitive data are of high quality. Those that may endanger national security are also typically of large geographical extent. There is already a lot of data, both commercial and governmental, that may fall into one or both of these categories. However, the majority of such datasets are not available as open data. In general, data that clearly may endanger either national security or individual privacy should not be published as open data.

4.4. Risk mitigation

Threats related to open geospatial data are very varied, and therefore the possible consequences and risk mitigation methods vary. The malicious actors recognized in this work range from individual people to nation states. Therefore, the resources and expertise the different malicious actors have available can be very different. A nation state can use a lot of resources, if they really consider some course of action sufficiently important, while individuals are significantly more limited. In general, the less resources and expertise a malicious actor has, the easier it is to hinder or prevent them. If the malicious actor has sufficient resources and will, the most that can be done by hiding existing datasets is to force them to either use more resources to find out what they want, or to switch their target to a less-protected one.
In general, risk mitigation methods related to open geospatial data are either related to *monitoring access* to the data, or *restricting access* to it. Such methods can be used in ways that prevent the dataset from being open data. For example, the data access can be set behind a user identification schema. However, it is also possible to use some methods while keeping the dataset open. For example, access monitoring can consist of logging all IP addresses that are used to access the data without adding any restrictions to the data use. Similarly, the access to certain dataset can be restricted by, for example, removing some attribute data elements, increasing the size of a pixel in a raster dataset, or removing some data elements completely from the open publication of the data.

However, many such methods will turn the data from open data to shared data. The results of such methods are often quite visible to the users. For example, access to a dataset can be monitored in more detail by requiring users to register to a service. This means the user needs to go through extra steps before being able to access the data. The type of registration can vary from requiring users to input their email address before being granted access to the use of a strong identification method before being granted access.

One important thing to remember is that no risk mitigation method is necessarily fool-proof. Many methods can prevent malicious actors with limited resources, such as activists, criminals or lone-wolf terrorists, from exploiting the data. On the other hand, with sufficient resources a malicious actor could gain access to even datasets that have no open access. However, the more stringent the security, the more expertise and resources is typically required to evade it. Furthermore, every time access to a dataset is restricted, the consequences of these restrictions to beneficial, legitimate use of the data should be carefully considered. Changes in data quality, like generalization of attribute data values or decreasing the resolution of raster datasets, do not affect only malicious actors, but may also prevent legitimate use of the data, decreasing the added value the publication of the data provides. In addition, for many risks, the best risk management methods may not be related to the geospatial datasets at all.

Finally, even if a single dataset does not contain any known risks, this does not prove that there are no risks related to the dataset. Data combination from two or more independent datasets may reveal information not present in any of the input datasets. That way it may be possible to find sensitive data, despite the data not being present in the input. Managing the threats related to such multiplier effects can be extremely difficult, as it is not possible to compare a dataset to all other relevant datasets.

5. Discussion

The results of the work are closely related to earlier threat assessment work done on a national level (Ministry of Interior, 2023). Our work includes several insights about how open geospatial data could be used to threaten critical infrastructure, important infrastructure, soft targets, as well as the privacy of individuals. Similarly, our results list potential *sources of threats*. Both the targets and the threats described in this work are not new. The threats are well-known and often discussed in work related national security. They are not unique to the geospatial ecosystem.

When considering the threats and mitigation strategies, it is crucial to remember the benefits of open data. Just the *possibility* to misuse a dataset is not a sufficient reason to try and limit the use of the data. Only if the threats are significant enough compared to the benefits gained from open data, should limitations to the data be considered.

Our study brings an important new aspect to the narratives around open geospatial data, as there is not much public discussion or research related to the potential threats caused by geospatial data. Furthermore, our study reveals that there is an urgent need for further developing the commensurate guidelines, such as the one by Open Data Institute (2022), and risk assessment frameworks for different stakeholders who produce geospatial data. These frameworks should have capabilities to recognize both the probabilities and the significance of the threats related to both scenarios for the dataset in hand: The threats related to the openness as well as the threats related to the opposite case, such as the loss of the potential societal safety gain generated through openness. Balanced consideration of these issues will play an increasingly central role as smart city development and AI technology take over industries in cities and in society. Uniform policies would help to consider the threats and risks more equally

related to opening and sharing new geospatial data from the perspective of national and societal security.

References

Ahonen-Rainio, P., Mäkelä, J., & Virrantaus, K. (2014). Menetelmä avoimen maastotiedon vaikuttavuuden arvioimiseksi (Method for assessing the impact of open geographic data). Maanmittauslaitos. https://www.maanmittauslaitos.fi/sites/maanmittauslaitos.fi/files/old/aalto_maastotietojen_vaikuttavuus_tutkimus2014.pdf

Eduskunta, 2015. Kuntalaki 10.4.2015/410 (the Municipality law 10.4.2015/410). https://www.finlex.fi/fi/laki/ajantasa/2015/20150410?search%5Btype%5D=pika&search%5Bpika%5D=kuntalaki (2 May 2024).

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

JUHTA, 2006. JHS 160 Paikkatiedon laadunhallinta (JHS 160, Spatial data quality management). https://geoforum.fi/jhs-160-paikkatiedon-laadunhallinta/ (2 May 2024).

Ministry of Interior, 2023. Kansallinen riskiarvio 2023 (National risk assessment 2023). Ministry of Interior. https://julkaisut.valtioneuvosto.fi/handle/10024/164627 (29 April 2024).

Open Data Institute. (2022). Assessing risk when sharing data: A guide. Open Data Institute. https://www.theodi.org/article/assessing-risk-when-sharing-data-a-guide/ (2 May 2024).

Open Data Institute. (2023). The Data Spectrum. https://www.theodi.org/about-the-odi/the-data-spectrum/ (2 May 2024).

Schrock, A. R. (2016). Civic hacking as data activism and advocacy: A history from publicity to open government data. New Media & Society, 18(4), 581-599. https://doi.org/10.1177/1461444816629469

White house, 2013. Introducing: project open data. https://obamawhitehouse.archives.gov/blog/2013/05/16/introducing-project-open-data (2 May 2024).