Jerkko Hietanen

THE STRATEGIES, MYTHS AND PRACTICES OF OPEN DATA
Analysis of access to open data in three countries

SUMMARY

The thesis examines the principles and practical measures for opening up public information resources. The study focuses on the national strategies and measures of three countries, Finland, Denmark and Estonia, in publishing open data. The principles of open data disclosure and distribution are examined using content analysis methods. The focus in the analysis is the formal open data strategy or public policy that the authorities have made. In addition to analyzing the actual text, the content analysis evaluates the goals and values behind the strategies. Furthermore, the use of myths related to open data is evaluated.

In the analysis of the national open data portals, the European Open Data Portal is also used to assess the publication of data. The information contained in the portals is compared with the criteria for publishing open data. Open data has included a wide variety of data in different countries. The publication of digital materials as open data has been analyzed in more detail in two subject areas (traffic information and environmental information).

The main findings of the study on open data strategies and practical dissemination concern, firstly, the differences between national open data strategies and, secondly, the differences in the distribution of open data and, in particular, the availability of linked open data. There are clear differences in open data strategies in open data management and in the utilization of open data standards. In Estonia, the role of the state is more central than in Finland or Denmark. In total, open data strategies in Finland, Denmark and Estonia have different content emphasis. There were clear differences in the use of open data myths in these countries. In the distribution of open data, the differences between Denmark, Finland and Estonia are significant. Denmark and Finland use advanced open data file formats, but in Estonia open data is mainly published as a simplified website.

The study found that the distribution of open data in Finland is more diverse and extensive than in Denmark or Estonia. Regarding the environmental data and traffic data that were analyzed, significantly more data are available in Finland than in the countries used in the comparison.

Keywords: open data, public information, authorities, national strategies, portals

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Table of contents

1 INTRODUCTION

Over the last ten years, open public information or open government data (OGD) has become a new theme in the public debate on the information society. In practice, open public information means opening up the information repositories of the national public administrations and enabling the use of those resources both citizens and businesses in an easily accessible form, mostly free of charge (Hossain 2016, 15-16). In the Nordic countries, there are certain historical factors which have had direct impact on the current policies on open government data. There have been regulations concerning the access to public sector information for over two centuries. The rules have applied to paper documents, but with the emerging digitalization, the emphasis is on the access to digital public data. In Finland, a significant policy reform has been made in 2011 in the Government's decision-in-principle concerning the availability of public administration digital data. According to the decision-in-principle, "the data must be openly accessible and re-usable under uniform, clear and equal conditions, in principle free of charge." (VN decision-in-principle of 3 March 2011) The discussion on the opening of public information resources had, of course, already taken place before the decision-in-principle (Poikola et al. 2010, 16-19).

The aim of this study is to examine the national policies concerning open data in three European countries and to find out how top-level policies have been implemented in practice in the distribution of open public information. As a point of departure in the study is the conclusion made previsously in the analysis of open data research (e.g. Zuiderwijn and Janssens 2014, 17), that the research has not focused on the comparison between the high-level open data policies or strategies and the grassroot level implementation of those policies. There seems to be scarce information and detailed facts on the relation of the strategies to practice.

A special issue is the criteria for releasing open data, as very different data sets have been included in the scope of open data in different countries. The criteria for linked open data are one subject to review. In addition to the five-star quality criteria for linked open data (Berners-Lee 2012), a broader seven-star quality criteria is under consideration (Hyvönen et al. 2014). In addition to Finland, the national guidelines and operating methods for open data dissemination in Denmark and Estonia have been selected for the study. The importance of digitalisation and open public information has long been emphasized in Denmark and Estonia, and both countries have made progress in introducing various electronic services in recent years. Therefore, they are an interesting point of compari-

son for Finland. In the European Union, the guidelines for open data have been set out in the Open Data Directives most recently in 2019 and in the European Data Strategy published in early 2020.

The main questions of the study are:

- What kind of strategic guidelines (strategies / action programs) have been published at national level on the principles and methods of open dissemination of information, and to what extent do the national strategies differ from each other or from the common guidelines of the European Union?

- What kind of open data is provided in these countries in the fields of environmental information and traffic information (the specific research question is to find out in what data format open data is available and whether the most advanced file formats are used, which enable versatile re-use).

The report is divided into six chapters, the first of which is an introduction. The second chapter discusses key concepts and the third chapter evaluates previous research and studies on the topic. The fourth chapter reviews the analytical methods used in the study and the research data. The fifth chapter deals with the comparison of strategies and practical data dissemination for data from three countries. The final chapter presents the results and conclusions of the study

The aim of the study is to evaluate the open data policies or frameworks in three European states and to find out how the high-level policies are reflected in the distribution of open public information. In addition, the focus is in the criteria of the publication of open data as a specific issue, since open data includes a wide variety of data sets in different countries.

The criteria for open data (linked open data) are one of the objects of the review. In addition to the Five Star Quality Criteria for Combined Open Knowledge (http://5stardata.info/en/), a broader seven-star quality criteria (Hyvönen et al. 2014) is discussed. In addition to Finland, the national policies and operating methods of the distribution of open data in Denmark and Estonia are studied. In Denmark and Estonia, the importance of open data has long been emphasized and both countries have advanced in the introduction of various electronic services in recent years. Therefore they are a useful reference point for Finland.

2   KEY CONCEPTS

2.1   The concept of open data

The discussion related to open data still involves vague and imprecise use of concepts. The concepts of Open Data and Linked Open Data are among the most unclear and most ambiguous concepts in the discussion concerning the dissemination of information and data. The concept of Open Data is fairly new, even if the first initiatives to share scientific data and to preserve the existing data were introduced in the 1950s. It combines two very popular words, open and data, in a way, which does not reflect the general use of those words. The concept "open" in everyday language refers to the possibility for everyone to have access to an object or to an entrance. Furthermore, it deals with something that is not closed or blocked. The dictionary definition of "data" refers to factual information that is used as a basis for reasoning or calculation. On the other hand, it deals with "information in numerical form that can be digitally transmitted or processed" (Merriam-Webster). In Finnish language the situation is even a bit more complex, because nowadays the recommendation, for example by YSO, the Finnish General Ontology, is to use the word "tieto", instead of "data" (YSO). Yet, in any definitions on information theory or Knowledge Management, "knowledge", "data" and "information" are three different concepts.

The key issue in this context is, whether Open Data is in practice open data in the true meaning of the concept. If it is information in numerical form and it should be accessible to everyone, is this the case in reality? Open Data is available only in machine-readable form, for example in excel spreadsheets, csv format, xml format or in the more comprehensive RDF (Resource Description Framework) format. To view this kind of open data, you need software which can interpret the source code. Actually this means that Open Data is far from being open, even if it is data. For any citizen, Open Data is totally unreadable.

Open Data has already been defined by various authors and organizations. One of the most accurate definitions has been formulated by the Open Knowledge Foundation, which in fact has made two slightly different versions (Open Knowledge Foundation, 2017). Firstly, it can be summed up in the statement that "open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)." In a shorter version "Open data and content can be freely used, modified, and shared by anyone for any purpose" (Open Knowledge Foundation, 2017). According to this definition Open Data can be reused and enriched

by outsiders. Finally, the original data can be part of a new package of data, which can be redistributed. The data itself can be in any format, paper, electronic or whatever. In the hierarchy of concepts, open data can be seen as a top-level concept, with sub-concepts like Linked Open Data (or Open Linked Data). According to the Finnish open data service, avoindata.fi, open data in a simplified manner means information that is freely available and recyclable under an open license. "Open data is information in digital form that is freely available to everyone for any use, as long as its original source is mentioned" (Avoindata.fi 2020).

According to the Open Government Data project, data can be considered open if it is made public in a way that complies with these principles (www.opengovdata.org 2007):

- Complete - All public data is made available

- Primary - Data is as collected at the source

- Timely - Data is made available as quickly as necessary to preserve the value of the data

- Accessible - Data is available to the largest possible number of users for an extensive range of purposes

- Machine processable - Data is systematically structured to enable automated processing

- Non-discriminatory - Data is available to anyone, with no requirement of registration

- Non-proprietary - Data is available in a format over which no organization or company has exclusive control

- License-free - Data is not subject to any copyright, patent or trade secret regulations

In the discussion on open data, also open datasets are discussed. Open dataset are data, which are public, do not infringe copyrights or intellectual property, are government-financed, fulfil open data standards and are computer-readable (Open Knowledge Foundation, 2017).

Linked Open Data (LOD) emphasizes, in addition to the basic definition of Open Data, the interlinking between data elements. As Attard et. al point out, the term Linked Data is used to refer to a collection of common practices for interlinking and connecting structured data on the Web. Linked

Data is published on the internet in a machine-readable format, with its meaning explicitly defined. It is also linked to and from external datasets. The potential of Linked Data is in the creation of the Web of Data (Semantic Web); a massive dataset replacing gradually decentralized and isolated data sources (Attard 2016).

Linked open data means information that has been opened in such a form that other systems can also utilize it. Linked open data can be fairly easily expanded, recycled and combined in different applications. The principles of producing and distributing open source information are:

• Citizens and businesses have the right to information (with tax-based financing)

• It is not advisable to reproduce data produced once and for all.

• By combining all data gets enriched

• By making the data easy to reach, another authority, company or citizen can make a useful application or service based on that information (Hyvönen 2018, 27).

Auer et al. see Linked Data as a paradigm and in their view the term Linked Data refers to a set of best practices for publishing and interlinking structured data on the Web (Auer 2014). This paradigm approach or viewing Linked Open Data as a set of best practices seems to be the most useful way to evaluate the complex framework of Linked Data, instead of viewing Linked Data merely as a standard of publishing structured data. Yet, the basis of all Linked Data discussions is the first contribution made by Tim Berners-Lee in the World-Wide Web consortium (W3C) (Berners Lee). The interlinking using the Linked Open Data tools can be local or global. The W3C has been developing the Resource Description Framework, which is actually a large group of specifications defining the principles of linking open data (Hyvönen 2018, 24).

In Finland, YSO Ontology included in Finto thesauri and ontology services has successfully completed a distinction between key concepts in Finnish ontology. Open data is, according to ontology, the upper level concept. Linked open data, in turn, replaces the terms used with open link information, open linked data, linked open data, and open link data (http://www.yso.fi/onto/yso/p26001). Open Data refers to data that is freely available and recyclable by an open access license. Linked data refers to information elements that are (mostly automatically) linked to each other. Linked data can be seen as a phase in the development of the internet, with individual links in linked servers, linked documents, linked data and linked open data.

2.2 Open Data Strategy

The Open Data strategy can be defined as a governmental policy on how public information resources are opened up and the extent to which data is provided to businesses and citizens to utilize. In several countries, specific Open Data strategies have been adopted. Yet, in a number of countries, the opening of governmental data in accordance with the open data standards is only seen as one element in the general digitalization of the society. The publication of open data follows a number of principles that can be summarized as follows (Open Government Data 2007):

- Data published as an open source is public information. This in practice means that the protection of personal data and privacy must be guaranteed when information is disclosed.

- The data must be in a machine-readable form so that it can handle different applications and information can be transmitted through different interfaces.

- The data can be used free of charge.

- The publisher (producer) of the information enables the reuse and further processing of data by clear terms of use

- The structure of the data is clearly described and metadata standards are met.

2.3 Five Star Criteria for Linked Open Data

Tim Berners-Lee, the founder of the internet's web environment (www), has developed five-star criteria for linked open data. Berners-Lee's criteria are hierarchical in nature, so the two-star level contains one-star-level features. The five-tier classification of Berners-Lee includes the following criteria:

*   Available online (in any file format) with an open license

**   Available in a constrained structure (for example, as an excel table, not just an image file)

***   Like ** but in system-independent format (eg CSV and not excel)

**** All above and addition: W3C Open Standards (RDF and SPARQL) which are used to identify information so that published information can be referenced

***** All of the above and in addition: Link your information to other information to include it in the correct context.

2.4   The Seven Star Criteria for Linked Open Data

Professor Eero Hyvönen (Aalto University, Finland) has developed a broader framework with Berners-Lee's criteria. The seven-star criteria are based on the first five criteria for Berners-Lee's rating. In addition to these, Hyvönen has released two new higher criteria and in fact, Hyvönen calls the model of seven-star-linked data service model. Stars are like bonuses to emphasize issues that are essential to the delivery of open data. The sixth star describes how well the data schemas (ie. describing what data are) are published. The seventh star is awarded when the data also meets the quality requirements of the staircase and when the production of the material is well documented.

2.5   Criteria for five-star Danish basic databases

Denmark has introduced a five-star rating for the quality classification of basic databases (Røyen 2014). The five-star rating is not hierarchically the same as Berners-Lee or Hyvönen's criteria, but rather a minimum requirement for disclosure of open data.

The five star rating in Denmark consists of the following components:

* basic databases must be of sufficient quality for most of the uses.

* ensuring the accuracy and timeliness of basic databases must be clearly responsive and effectively managed.

* basic datasets are semantically compatible and modeled according to "Basic Data Modeling Rules"

* basic databases are freely available and their re-use is not limited (does not apply to sensitive information), information is available through the information provider

* public sector and private use of basic databases and their use will be of great benefit (Røyen 2014).

2.6   On the background of the key concepts

The concepts of Open Data and Linked Open Data have a cultural-historical-political background. This is relevant especially in the open data repositories which are distributed by national governments. It is a common principle that the data which has been produced by the government, and which may be based on the data collected from citizens and enterprises, should be made available as open data. If government money has been spent on the data, it should be freely available. The data produced by the private sector is a different case. It is natural that the opening of the data is determined by the owner of the data.

The benefits of open data can be derived from the definition. If something is available freely, free-of-charge, in a format which can be utilized easily, it is evident that the data may promote innovation, new thinking and even new products or services. Yet, this does not occur automatically.

Here we reach the key chain of questions on the true nature of Open Data (this is my own formulation):

- who opens the data (citizen, researcher, enterprise, government, non-governmental organization, international organization)

- when the data is opened (is it already outdated or up-to-date)

- how the data is opened (which formats and channels)

- why the data is opened (the motivation)

- how the quality of data is guaranteed

- how the data is updated

2.7   On the myths related to the definition of the key concepts

One new approach to the topic of open data could be called myth analysis (Bong 2002). A myth is usually defined as a story based on traditions or legends, with no true connection to facts (Magoulick 2014). The core element of a myth is in its characteristics as something fictional or unprov-

en. Magoulick points out that often the use of myths results in providing meaning and new content to certain elements of culture and asserts processes of change (Magoulick 2015, 1)

Homburg has analysed the use of myths in public policy. He pays attention to the fact that myths can have a major role in the formulation of political argumentation and preferences. In the public discussion on open data, open data may easily be seen as a perfect choice in solving various problems and thereby the possible negative aspects and risks are not observed properly (Homburg 2005, 495)

Janssen et al. have discussed the myths related to Open Data. They have discerned five myths, which might be useful in the evaluation of the national open data policies or strategies, as well. The myths are the following:

1) Myth of automated benefits in the release of data

2) Myth of unlimited information delivery

3) Myth of public data as open data

4) Myth of utilization of data by citizens

5) Myth of open data as a path to open government

The first myth deals with the perspective that the release of open data results automatically in various benefits and positive outcome. The second myth is about the comprehensive approach to open data, emphasizing that all information should be made available without any restrictions or limitations. The third myth argues that open data is a simplistic process, merely making public data available. The fourth myth refers to general possibility of any citizen to utilize open data without any major difficulties. According to Janssen, there seems to be very strong emphasis on the data providers or producers. "Too much emphasis is given to the data supplier and only limited attention for the user. Lowering the threshold for use should also be part of the policy" (Janssen 2012, 264)

The fifth myth is about the belief that Open Data will contribute to the construction of open government (Janssen 2012). The analysis by Janssen indicates that the meaning of the concept "open" is culturally bound to certain assumptions and for that reason we have great difficulties in understanding the real meaning and proper impact of Open Data. One interesting question posed by Janssen is, whether it is cost-efficient to publish open data if nobody really uses the data that has been made available (Janssen 2012). This is something that has not been discussed in Finland until quite recently, when analyzing the impact of Open Data in the private sector (Avoimen datan

hyödyntäminen 2017). Another comment of importance and also related to the second myth of Janssen, has been made by Chernoff - "If open data is misunderstood as releasing any and all data to the public, people will become opposed to the concept due to their concerns about privacy." (Chernoff 2010).

In addition to the myths analyzed by Janssen et. al. (Janssen 2012), in the definition of the concept Open Data, two special aspects have to be taken into account. The first one are the restrictions imposed by the copyright. No citizen or enterprise can disseminate data, which is copyright protected, or there has to be some agreement with the copyright holder. Another aspect is the protection of privacy. At international level, there are many treaties on the protection of privacy and it is not legal to disseminate for example health records or bank account information of citizens. Therefore there are certain limits and background factors in the definition of open data and the same applies to linked open data. Open Data may sound very positive and productive as a concept, but it is neither open for anyone nor does it automatically result in increased transparency or open government.

In my own definition open data refers to data in electronic form which anyone can access, use, modify, and share for any purpose, taking into account privacy and copyright. To make a more comprehensive conceptualization, this definition can be widened with the eight criteria of the Open Government Data project.

2.8   Open data, re-use of public information and protection of privacy

Since 2003, the principles of open access to data have been influenced by the EU PSI Directive on the re-use of public sector information (Knuuti 2014, 52-53). The initial aim of the directive was to promote the utilization of data resources that were operated or owned by the public sector in the member states of the European Union. The directive was based on estimates that non-EU countries, in particular the United States and Japan, gained a competitive advantage by making extensive use of public sector data in their own business activities. The impact of the Directive was assessed in 2011 (European Commission 2011,). At the time, the conclusion was that the wording of the Directive for open access was too interpretative and the Directive had not significantly increased the re-use of public sector information (Attard et al. 2016, 6).

In 2013, the Directive was amended to make it as clear as possible for EU Member States to allow the re-use of all data, unless their availability is restricted or prevented by national legislation (Knuuti 2015, 55). However, the impact of the PSI Directive does not extend to national legislation

on public access to documents or the protection of privacy, thereby some public sector data (personal data and confidential data) will continue to be excluded from the scope of the Directive. The directive still has a clear impact on the opening of data, as it has considered the disclosure of open data to be a prominent topic of discussion. The European Commission has monitored the implementation of the directive on a separate website (Implementation 2017, 2). In 2019, a recast of the PSI Directive and its amendments, with a new title "directive 2019/1024 on open data and the re-use of public sector information", was published. EU Member States must transpose the directive nationally by 17.7.2021.

There have been different policies in the implementation of the PSI Directive in different EU Member States (Implementation 2017, 2). Several EU Member States (including Germany, Italy, Spain and Sweden) have separately enacted new legislation on the re-use of public sector information. A few EU countries have used combined measures, which include new legislation and changes to previous legislation (eg Austria, Slovenia and Denmark). Some EU countries have only adapted previous legislation on access to information (e.g. the Netherlands, Poland, France, Finland and Estonia). In Finland, the PSI Directive and its amendments have been implemented by the Act on Criteria for Charges Payable to the State (150/1992), the Act on the Openness of Government Activities (621/1999) and the Public Administration Information Management Act (906/2019). In Estonia, the PSI Directive and its amendments have been implemented by the Estonian Public Disclosure Act (2003) and its two amendments in 2012 and 2015 (Implementation 2017, p. 4).

There is still relatively little information on the economic implications of open access (Auer, S., Bryl, V., Tramp, S. 2014; European Data Portal 2020; Wood 2011). One extensive study on the utilization of open data (mass information) in Finnish companies has been published in Finland (Koski et al. 2017). It has been able to show that companies that make use of open data and mass data in their innovation activities bring new product innovations to the market relatively much more often than companies that make little use of open data. The study found that this finding was true, especially in medium-sized or large technology companies. (Koski et al. 2017, 23). In different countries, open data strategies and guidelines for the practical dissemination of information generally include a view on how information can be shared without violating the privacy or data protection of individuals (protection of personal data) (Simperl et al. 2016, 7). In most cases, however, privacy or personal data protection policies do not specify the material to which these policies apply. In Finland, emphasis has been placed on the protection of privacy since the guide-

lines for the principles of opening up public information resources (Ministry of Finance 2013). The policy set the goal that "by 2020, all relevant public data managed by public administrations will be available in machine-readable form, free of charge and under clear conditions of use, taking into account data protection, privacy and other legal restrictions." (Ministry of Finance 2013, 1). The criteria were benefits to the national economy, transparency and improved efficiency. The need for anonymisation of information has only been referred to in a recent study on the conditions for producing open data (Koski et al., 2017, 59). Of open data repositories, it is often difficult to discern what personal information they may contain. Especially authority decisions materials containing it would be useful for opening up sources of information, but they often include information on individuals that can be combined with information found in other materials. Such materials include, for example, business information, tax information, health information and social security information.

Big data can be considered a special challenge for privacy protection. Big data usually refers to huge, unorganized and distributed data masses for the processing of which new technologies are constantly being developed. The problems caused by big data are related to the fact that no one has accurate information about the details of the contents of the big data, and especially not about whether the data contains personal data (Hokkanen 2015, 12). One practical solution for expanding open data distribution has been considered to be anonymization or pseudonymization of information prior to open data distribution. However, anonymization or pseudonymization has its own problems, as automatic anonymization requires the use of advanced text filters. On the other hand, anonymisation can lead to a shrinking importance of the data if much more than person-related identifiers are removed from the data. In addition, anonymized or pseudonymized data can be restored back to an identifiable form by various technologies (Lubarsky 2017, 1)

3 LITERATURE REVIEW

In recent years, fairly few studies have been published on open knowledge strategies and practical distribution. As it has been pointed out by Zuiderwijk and Janssen, there is a lack of comprehensive and systematic studies with a comparative approach on open data (Zuiderwijk & Janssen 2014, 17). An essential element that has been missing in research is the comparison between the high level open data policies or strategies and the grassroot level implementation of those policies. There seems to be scarce information on the relation of the strategies to practice. Therefore is is quite natural that certain myths are still connected to open data. The high level strategies on open data have to be supported and reinforced by the measures taken by local communities and various government institutions.

The studies chosen for the literature review provide useful insights to the interlinking of the strategy level and implementation level. Here are the short summaries of Huijboom's and Van den Broek's comparative research into open-source information strategies, a description of the work carried out by Kassen and the European Commission's study on Open Data Maturity.

3.1 Huijboom and Van Den Broeks Research on Open Data Strategies

Noor Huijboom and Tijs Van den Broek have in their report ""Open Data: An International Comparison of Strategies" explored five state open data strategies (Huijboom & Van den Broek 2011). The targets included the US, UK, Australian, Danish and Spanish governments' strategies for publishing public sector information as open data as part of open administration. Huijboom and Van den Broek evaluate national strategies from four different points of view: firstly in terms of education and counseling, secondly in terms of voluntary measures, third in terms of funding and development, and fourth in terms of legislative and supervisory measures. The researchers concluded that in the five countries that were the subject of research, the weightings varied moderately. There were some special features in each country. For example, in Denmark, emphasis was placed on training and counseling (including various open-source knowledge platforms, workshops and training sessions) and voluntary activities (creating open-minded atmosphere with strategies and guidelines). In all five countries, open access to information and the creation of new applications based on open data were promoted, among other things, via application competitions (hackathons) and the public funding of open data portals. Similarly, in all five countries, legislative measures were used to en-

hance open access to information, the members of the EU benefited also from the implementation of the EU directive on public sector information.

The study also addressed key factors that facilitate or complicate the promotion of open access to information. The experiences and practical implementation of other countries were considered as important facilitating factors in the United States, Australia and the United Kingdom. In Denmark, the most important contributors to the discussion on open data were the citizens' initiatives and EU legislation (especially the directive on public sector information) (Huijboom & Van den Broek 2011). The main obstacles were the closed administrative cultures in all five countries, privacy legislation and the poor quality of information available. A major obstacle was also the lack of clarity as to the level of financial impact that the opening of data would have in practice.

3.2 Kassen's studies on the guidelines and implementation of open data

In recent years, Maxat Kassen has studied strategies and practices for publishing open data in several different countries, e.g. in Estonia, Finland and his home country Kazakhstan (Kassen 2018a and 2018b, Kassen 2019). In particular, Kassen has examined the role of the public sector as a publisher of open data and the role of the private sector and NGOs as lobbyists and facilitators of the opening of public information resources. In his research, Kassen has not addressed the opening of individual data repositories, but has focused on the dialogue between the public sector and other actors related to the publication of open data.

In the case of Estonia, Kassen considers the fact that the public sector has taken into account the needs of the private sector and non-governmental organizations in the preparation of open data policies to be an important advantage in opening up data resources. However, Kassen points out that the re-use of open data materials has not been much discussed in Estonia. Kassen analyzes the reasons behind the passive role of the Estonia government in releasing open data. He points out the facts, that the public sector has had a very passive role in the promotion of open data. Another factor has been the lack of financial incentives (from the government side) and the lack of any kind of fund-raising basis (Kassen 2019, 59). A third factor is the almost complete absence of the local business communities as financers of open data solutions. All these factors result in the paradox of advanced e-Government and scarce open data in one single country.

As regards Finland, Kassen estimates that many different actors - in addition to the public sector - have promoted awareness of the need to utilize open data. Kassen's key players have been e.g. Open Knowledge Finland, Open Ministry project and Open Eco-system Network (Kassen 2018a, 527-528). Kassen believes that effective re-use of knowledge requires a more active contribution from private development communities, which should be more clearly supported by the public sector (Kassen 2018a, 532).

3.3 European Commission study "Open Data Maturity"

In recent years, the European Commission has commissioned an annual EU-wide report on the maturity level of 'Open Data Maturity'. The report examines the availability of open data in all EU Member States, as well as in Switzerland, Norway and Liechtenstein. The review assesses two areas, the Open Data Reactivity and the maturity of the Data Portal Maturity. Five criteria are used to assess the capabilities of open data. First, the existence of an open data strategy or policy. The second criterion is the licensing guidelines and the third is the extent of national coordination. The fourth criterion concerns the extent of the use of the information and the fifth concerns the effects of open data. (Open Data Maturity 2019, p.6.)

According to the maturity survey, the policies supporting the distribution of open data and the use of open data increased significantly from 2016 to 2019, as did the maturity of open data portals (which refers to the diversification of portal functions) from 2016 by about 27%.

In the Open Data Maturity study, EU countries are divided into four groups. Leading the way are the trend setters, followed by fast-moving countries. They are followed by successor countries and beginners.
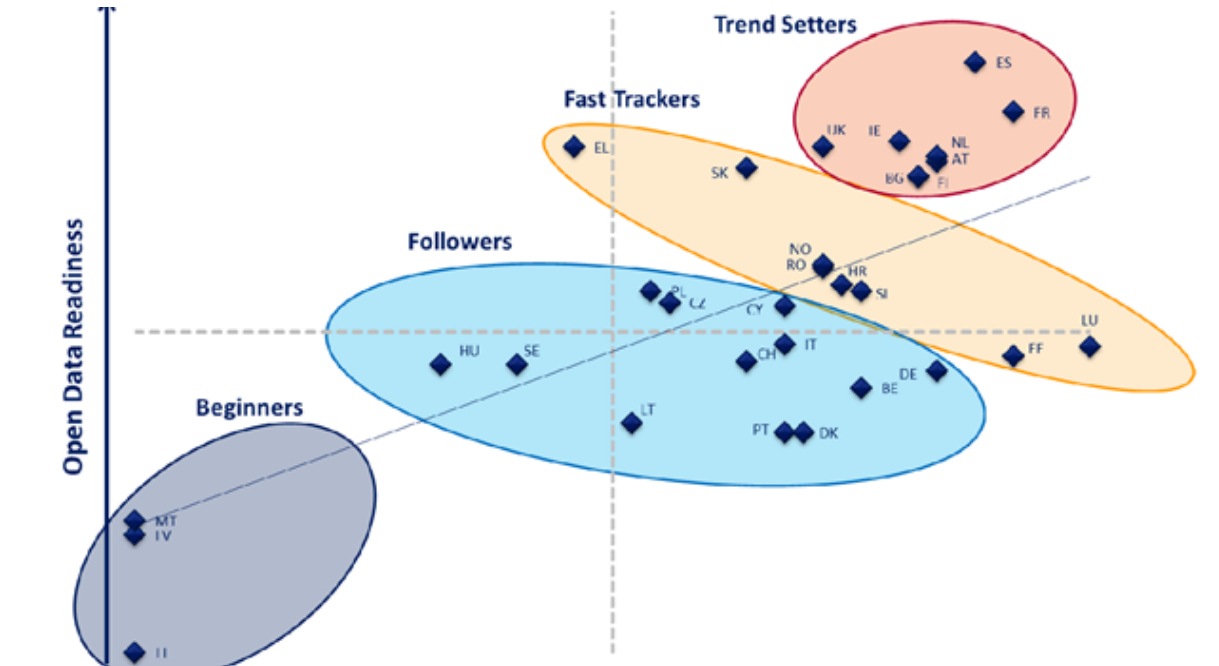
Figure 1: Open data maturity in EU countries 2016 (Open Data Maturity 2016, 64).
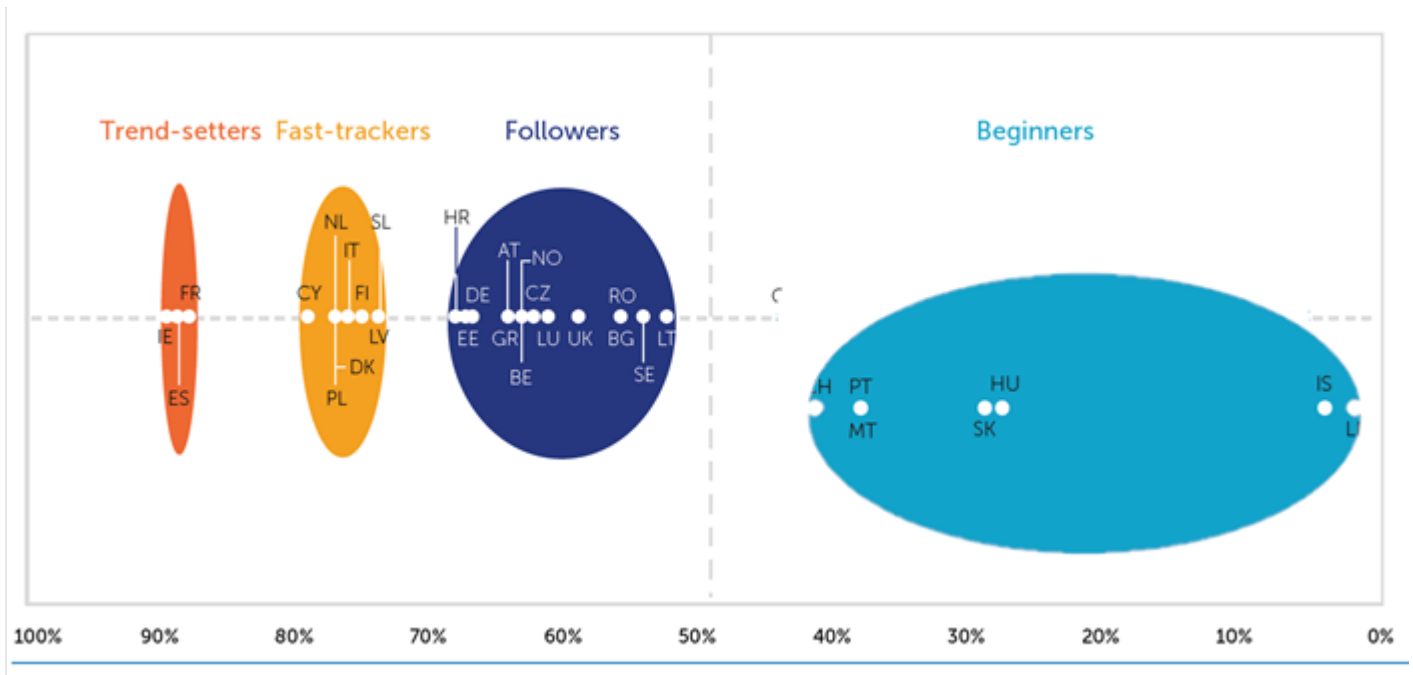

Figure 2: Open data maturity in EU countries 2019 (Open Data Maturity 2019, 72).

Between 2016 and 2019, the investments in the Open Data Maturity study have not changed significantly. However, there are only three countries in the trend-setting, Spain, Ireland and France. Finland has fallen from a trend setter to a fast-moving country. Austria has fallen even more clearly, from trend setters to successor countries. Rapidly advancing countries also include e.g. Italy, Latvia and Denmark. Sweden and Estonia have remained successors. Regarding the Open Data Maturity study, it it worth noting that the estimates are made by the national network of liaison officers and are not in themselves commensurate. The report also does not deal with national guidelines for the dissemination of open data, the utilization of open data standards or, for example, the provision of combined open data in the target countries.

3.4   Other studies on the topic

In 2015, the Organization for Economic Co-operation and Development (OECD) published a report on the development of electronic services in Finland and Estonia "Estonia and Finland - Fostering Strategic Capacity across Governments and Digital Services across the Borders". In addition to general e-government strategies, the report also discusses how open data is used in public administration to develop administrative services or to collect information about administrative customers. According to the report, only less than in Estonia only about 25% of authorities and in Finland about 40% of authorities use open data (OECD 2015, 139). However, the report states that the publication and utilization of open data is progressing in both Estonia and Finland. According to the OECD report, one of the obstacles remains that some public authorities are dependent on revenues from paid electronic services. Another obstacle is seen in the fact that the protection of privacy imposes restrictions on the provision of open data. The report focuses on the development of administration and the electrification of administration in Estonia and Finland. The report does not assess the extent to which companies use open data in their own research and development activities.

In 2019, the European Environment Agency (EEA) commissioned a study from the consulting firm PriceWaterhouseCoopers on the availability of open environmental data in Europe, Open data and e-Government - good practices for fostering environmental data sharing and dissemination (Price-WaterhouseCoopers 2019). The study has addressed the availability of open data in several EU

Member States and compiled good practices for the open dissemination of environmental data.

The study by Marijn Janssen et al. *Benefits, Adoption Barriers and Myths of Open Data and Open Government* provides a useful framework for the analysis of the open data policies and especially on the myths behind the official policies (Janssen et al. 2012). The presentation of the five key myths related to the open data is based on extensive research on the national open data policies and public statements.

In Finland, a working group of the Ministry of Transport and Communications published a report in 2013, Open Data on Transport and Communications, which deals with information resources in the transport and communications sector that would need to be published as open data (Ministry of Transport and Communications 2013). The report also includes a roadmap for opening up data resources, which has already been implemented in many respects.

4 RESEARCH METHODOLOGY AND RESEARCH MATERIAL

In the research on Open Data Strategies and the distribution of data, the key research method is content analysis. In the content analysis, the material is scanned by specifying the text content and looking for possible similarities and differences between the texts that are being compared. Content analysis is a text analysis that looks at already existing textual or modified material. The texts to be analyzed in my research are pre-textual documents published by the authorities and documents and reports on the availability of open access. Content analysis aims to create a comprehensible description of the phenomenon being investigated, which links the results to the wider context and other findings of the phenomenon. (Tuomi J., Sarajärvi 2002, 105). Krippendorff emphasizes that content analysis is not just a text-mediated message analysis (who communicates to whom, which instrument, which goals), but also needs to evaluate psychological aspects (why communicate in a certain way), institutional aspects (what are the values behind the message) and cultural aspects (which rituals or traditions are included in the message) (Krippendorff 2012). Particularly in the content analysis of open data strategies (which are the outputs of political decision-making), it is important to take into account the institutional and cultural aspects.

In the second part of my research, in the analysis of the delivery and availability of open data, I also use content analysis, combined with qualitative and quantitative analysis of the open data repositories.

In qualitative methods of content analysis, Bengtsson (Bengtsson 2016, 9) distinguishes between manifestation-level or manifesto-level analysis (manifest analysis) and latent analysis (latent analysis). The first examines the surface structure, i.e. what has been told in the text, and the second correspondingly examines the deep structure, i.e. what has been intended to be told. The surface structure and the text to be analyzed should also be able to look deeper than the surface and assess what the aims of the text have been. Bengtsson makes a process plan for qualitative content analysis. The first stage is decontextualisation of the texts to be analysed. In practice this means that the text is taken from the their original context and split into smaller units, sentences and meanings. The second stage is recontextualisation, referring to the necessary process of distancing, taking distance from the original text. The third stage in the path drawn by Bengtsson is categorisation, which refers to the identification of various themes and categories in the text. The final stage in the content analysis is logically the compilation and evaluation of the categorised materials (Bengtsson 2016,12).

A similar view is also pointed out by Mayring (2014, 39). Mayring argues that the content analysis must emphasize the communicative context of the material to be analyzed, i.e. the context to which the text in question relates. When analyzing the content, the origin of the text and the effects sought by the text must be taken into account. Content analysis as a method has also been criticized for the credibility of the interpretation of texts and the analytical process. Elo et al. have emphasized the importance that there has to be comprehensive description of the analysis and clarification on the use concepts. Otherwise it is practically impossible to evaluate the analytical process and to verify the outcome of research (Elo et al. 2014, 7).

In the study on the delivery and availability of open data, content analysis is used in combination with statistical methods. In the content analysis, the statement level (manifesto level) of the strategic guidelines is analyzed on the one hand and the latent level on the other. In addition, I compare the strategic guidelines to the myths related to open data. The objects of content analysis are the existing open data portals, the Finnish avoindata.fi portal, the Danish opendata.dk portal, the Estonian opendata.riik.ee portal and the European data portal (https://europeandataportal.eu/). From them, I compile information on the data in the subject areas I deal with. The method is a text-analytical and semi-statistical comparison of what data in the topic areas exist, in what form or format, what are the access rights, and an analytical qualitative assessment of how the data can be evaluated in accordance with the quality criteria of linked open data.

5 CONTENT ANALYSIS - OPEN DATA STRATEGIES AND DATA SETS

In the content analysis, the open data strategies and action programs are discussed first. After that, the focus will be on open access to information and, in particular, access to open data in the countries which are evaluated, namely Denmark, Estonia and Finland.

5.1 Content Analysis of Open Data - Denmark

The availability of basic databases is ensured in Denmark through a centralized distribution mechanism. In Denmark, the key development objective for the next few years is to develop the quality of basic databases to make better use of data. In Denmark, it is also worth exploring the potential for publishing wealth data and income data for individuals as open data (Digitaliseringsstrategi 2016).

In Denmark, e-Government strategies have been published since 2001. In strategies, the emphasis has changed over the years. The first strategies emphasized electronic transactions and the introduction of electronic payment methods. It was not until 2011 that guidelines for data re-use were included in the strategy. Since 2012, Denmark has implemented a Grunddata program (Horst 2014, 1). The opening up of Denmark's basic data resources has been justified in particular by boosting economic growth, promoting innovation and increasing employment. Since 2012, the basic data repositories have been opened e.g. real estate information, address information, business register information. In addition, e.g. opening up geo-data (map and soil data), water resources data and agricultural data resources.

The availability of basic data resources in Denmark has been ensured through a centralized distribution mechanism, the so-called data distribution service (Datafordeler service) www.datafordeler.dk. In Denmark, a key development goal for the next few years is to improve the quality of basic data resources so that data can be better utilized (Digitaliseringsstyrelsen 2016, 27).
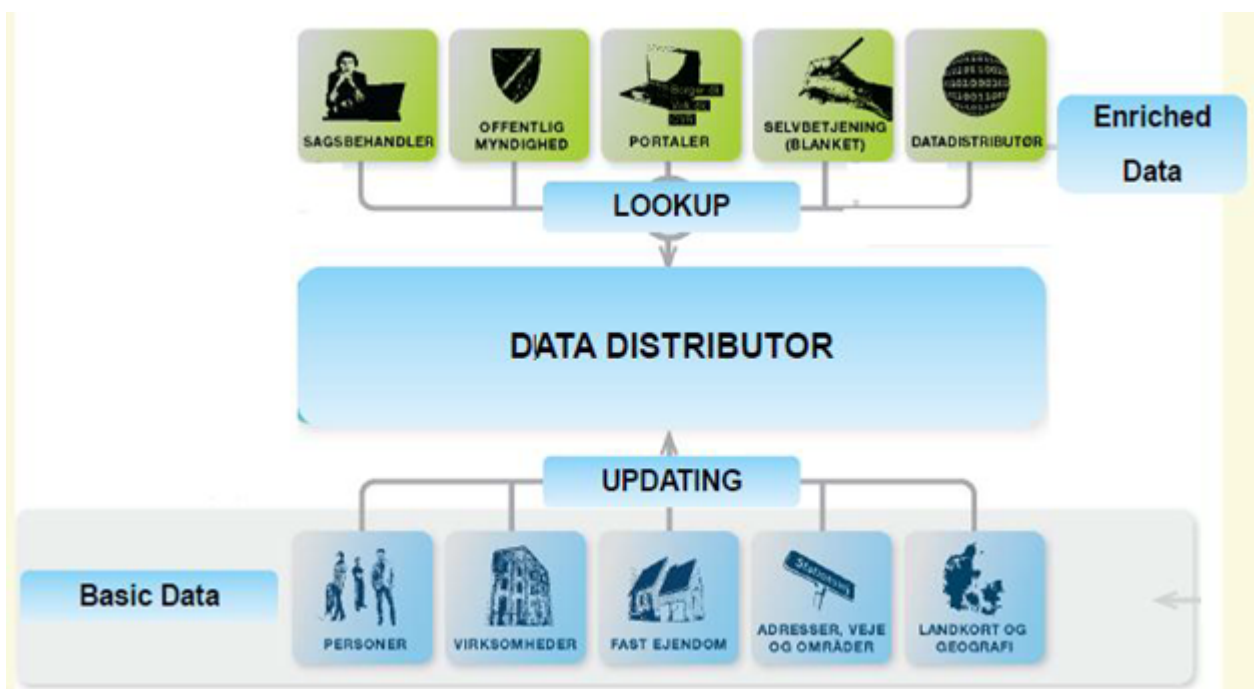
Figure 3. The distribution process of data in Denmark (Digitaliseringsstyrelsen 2016, 28)

The strategic guidelines for Danish open data are included in the digitalisation strategy for 2016-2020. The strategy emphasizes three main objectives. Firstly, digital information and services must be easy to use, easily accessible and of high quality. Secondly, digital materials in public administration must improve the conditions for economic growth. Thirdly, the digitization of data must take into account data security and data reliability (Digitaliseringsstyrelsen 2016, 5). The opening of basic data resources included in the digitization strategy emphasizes the five-star principles discussed in Chapter 2.5 above. The strategy on the open use of public sector information states: "In order to promote public sector data, including the commercial use of this type of data, a common public sector partnership will be set up consisting of Open Data DK, Danish regions and the Danish Business Authority (Erhvervstyrelsen). The Partnership will consult with companies and experts in order to establish a better overall picture of existing open data and promote efforts to ensure the access to more open data." (Digitaliseringsstyrelsen 2016, 40)

In Denmark, the guidelines for open data have been included not only in the digitization strategy and the basic data repository program, but also in other strategies, including the Digital Growth

Strategy (2018) and the National Artificial Intelligence Strategy (2019). The Digital Growth Strategy does not address open data issues in isolation, but merely states the implementation of an open knowledge partnership. The aim of the artificial intelligence strategy is to make public sector information available to artificial intelligence applications, and during 2020-2021 it is planned to select five data sets for artificial intelligence applications (Danish Ministry of Finance 2019, 3).

In the content analysis, the guidelines of the Danish open data strategy can be viewed at the statement level and the latent level (Digitaliseringsstyrelsen 2016, 40)

Statement level	Latent level
"The partnership will consult with businesses and experts to establish a better overview of existing open data and promote efforts to ensure access to more open data."	No specific requirements are set in advance for the publication of information as open data. There is no need to define new areas for the publication of open data.

Table 1: Content analysis at the level of statements and latent level in Denmark (Digitaliseringsstyrelsen 2016, 40)

As to the myths of open data, there are traces of some myths in the Danish strategic documents. The complete list of the myths by Janssen et al. (Janssen 2012) can be found in chapter 2.7.

1)   Myth of automated benefits in the release of data

"Easy and free access to data can give house buyers a more smooth case process, save time and money in the financial sector, and be of benefit to the entire society" (Horst 2014, 36)

2)   Myth of unlimited information delivery

"Denmark is setting the pace by unlocking a treasure trove of information. The release of so much of its core datasets used on a regular basis by both public administrations and private business is fantastic" (Digitaliseringsstyrelsen 2016, 9)

5)   Myth of open data as a path to open government

"State authorities can contribute to economic growth and to achieving higher degree of transparency in state administration by opening their data. It is the data that the authorities have already collected to carry out their tasks and which can create added value when reused by citizens and businesses" (Folketinget 2019, 3)

5.2 Content Analysis of Open Data - Estonia

In Estonia, general objectives for digitalisation are outlined in the Digital Agenda 2020 and the e-Estonia project (e-Estonia.com). Estonia has developed a number of public services (e.g. electronic trade register, population register, real estate register and patient register), but Estonia has not progressed at the same pace in publishing open data. This paradox has attracted attention in its studies Kassen (2019) and Margetts and Naumann (2017). According to Kassen, Estonia has not seen any need for a separate open data strategy, as public sector information has been widely available already before the open data debate was launched in Estonia (Kassen 2019, 572). Margetts and Naumann estimate that Estonia's digital by default and the strong role of the Government in digitalisation have overshadowed the culture of experimentation of open data and the possibilities for self-extracting data from open data data data sets (Margetts & Naumann 2017, 29).

The secondary role of open data is also visible as open data in the number of published data. As a whole, 433 data sets from Estonia are included in the European Data Portal, with 1871 data sets from Denmark and 5050 data sets from Finland (European Data Portal 2020). The Estonian open data portal opendata.riik.ee was opened in summer 2018 and contains 570 data sets. The material is therefore broader than the whole of the data provided through the European data portal.

A Green Paper on the publication and utilisation of open data was published in Estonia in 2014 and the Digital Agenda 2020 in 2014 (Estonian Digital Agenda 2020). The Green Paper contains several proposals for measures to improve the availability of open data. Of these, the first applications have been implemented or launched in 2015 –2020 as part of the e-Estonia development work. For example, Meieraha (the state budget monitoring tool) and Valimise Valvurid (the monitoring service for promises made by parliamentarians) have been implemented as open data services. In Estonia, the public disclosure objectives relate in particular to increasing transparency and transparency, improving access to information, reducing the number of requests for information in the administration, and utilising new technologies, such as the linked data, Big Data and the Internet of Things (IoT) (Open Data Portal of Estonia 2020, 1). The Digital Agenda 2020 guidelines briefly mention open data and linked open data – in the agenda the protection of personal data has been identified as a major challenge (Estonian Digital Agenda 2020, 13). Other challenges or barriers to release of open data in Estonia have been analyzed in detail by McBride et al. (McBride 2018). McBride argues – based on surveys and interviews with stakeholders – that the most crucial barriers in the open data environment, preventing the publication and re-use of open data are the lack of

knowledge and awareness of open data and its benefits, weak political will, high data integration costs and lack of perceived value of open government data compared to the costs of releasing open data (McBride 2018, 617).

The following principles are outlined in the Green Paper on Open Data in Estonia (Eesti Avaandmete 2016, 11):

a) equal access to information (several formats, without regional restrictions, etc.)

b) coverage of data (quality criterion)

c) integrity of the structure of the data

d) definition of the terms of use of the data

e) publication of the information as up-to-date

f) anonymisation of data (privacy protection)

g) taking into account secrecy provisions (no confidential material is published);

Some additional criteria have been added to the Green Paper on Open Data (Eesti Avaandmete 2016, 12), in which it is stated that in order to obtain easy access and potential added value, the user must be able to:

• browse and search for a collection of interesting public information available;

• download all or part of the discovery data through the search system immediately without the need for agreements or passwords that prevent immediate downloading;

• use the database as a whole freely and upload the database as a whole to the computer and use it in its own applications without a separate permit or (additional) payment.

In addition, according to the Green Paper, a public sector institution publishing data is not obliged to provide additional services to users of data, such as converting it into a format suitable for them, creating specific online services, translating material, etc.

The guidelines of the Green Paper can be examined in the content analysis at the level of statements and latent level (Eesti Avaandmente 2016, 12)

Statement level	Latent level
"When the information is published, a compromise must be sought between the following three objectives: • easy access to and comprehensibility of data for the data seeker and the data downloader; • simplicity of disclosure of information and minimisation of the publisher's labour costs; • protection of the privacy of persons whose data are included in the data."	Publication of information as open data is at the discretion of the authority and may be hampered by the complexity of the publication process and publication costs
"A public sector institution publishing information is not obliged to provide additional services to users of information, such as converting it into a format suitable for them"	There is no need to provide information in multiple file formats

Table 2: Content analysis at the level of statements and latent level in Estonia (Eesti Avaandmete 2016, 12)

As to the myths of open data, there are traces of some myths in the Estonian strategic documents (all the myths are listed in chapter 2.7.).

1)   Myth of automated benefits in the release of data

"Opening up public sector data allows the private and free sector to combine them with other data and create new value-added business services" (Estonian Green Book, Eesti Avaandmete 2016, 5)

3)   Myth of public data as open data

"In principle, all public information is available through information requirements. It can be informally called making available by passive means" (Estonian Green Book, Eesti Avaandmete 2016, 6)

5)   Myth of open data as a path to open government

"The publication of public information as open data increases transparency in the public sector"

(Estonian Green Book, Eesti Avaandmete 2016, 8)

5.3 Content Analysis of Open Data - Finland

In Finland, the principles and implementation of open data disclosure have been outlined in the open data programme in 2013-2015. In the start-up phase of the program, the Ministry of Finance compiled the principles for the disclosure of open data (Ministry of Finance 2013). According to the principle memory, the opening of data resources seeks benefits from three directions. First, the national economic benefits, which are materialized through innovation, new types of business and product development. Another aspect was to increase openness and transparency in society, which in turn increases opportunities for direct civic engagement and facilitates access to information for citizens and businesses. The third background factor was seen to be increasing the productivity and efficiency of the internal administration, as the various information resources become more visible and their utilization easier.

In the open data programme in 2013-2015, measures to open up data produced by the authorities were coordinated and implemented. A key part of the open data program has been the introduction of a national data portal and the development of its content. The open data portal www.avoindata.fi contains about 1,700 data sets. As principles for opening up data resources, the open data programme emphasized that each person owns his or her own information and that the protection of privacy must be taken into account when preparing for the delivery of data (Jaatinen, 2014, 18).

In addition, the open data programme compiled common practices for the opening of information resources to ensure the interoperability of data sets and information systems. The program mapped national data resources as part of data architecture work. The open data programme also defined open data interfaces and terms of use, the most visible guidance being the public administration's recommendation Open Data License JHS 189 (VM Avoimen tiedon loppuraportti 2015, 11).

The open data programme also addressed the pricing principles of data. The starting point is that the official information is as widely available as possible free of charge and the information is provided free of charge or at the cost of extracting the information. It is therefore not an absolute requirement of free access to data transfers.

The open data programme created also an inventory of the existing data repositories in Finland. The graph cleary indicates that a number of important data is not yet available as open data.

Figure 4: The Finnish data repositories (source: VM Avoimen tiedon ohjelman loppuraportti, 2015, 21)

In the content analysis, the guidelines of the open data programme can be examined at the statement level and the latency level (Ministry of Finance 2013, 2)

Statement level	Latent level
In the economic impact assessment, attention is paid to e.g. effects on the reduction of public premium income, investment needs and the social and economic benefits of the change in fees.	The publication of information as open data is at the discretion of the authority and may be slowed down by a decrease in fee income and the hard-to-assess benefits of a change in fees.
"The data resources will be opened in stages in as wide and diverse a range of administrative areas as possible, as the agencies' plans and capabilities are in place. Priority can be given to the assessed economic, social or administrative efficiency gains or the extent of the use or user base of the information. However, the demand for individual data contents is unpredictable, and the greatest economic and societal benefits and impacts come from combining different data resources."	There is no urgent need to specify the scope and timing of the opening of the data repositories. Prioritization of data resources is difficult to do when the demand for data is not precisely known.

Table 3: Content analysis at the level of statements and latent level in Finland (Ministry of Finance 2013, 2)

As to the myths of open data, there are traces of some myths in the Finnish strategic documents (all the myths are listed in chapter 2.7.).

1)   *Myth of automated benefits in the release of data*

> Open data increases the transparency of government activities and increases new opportunities for civic engagement and access to information (Valtiovarainministeriö 2013, 1)

4)   *Myth of easy utilization of data by all citizens*

"Anyone can benefit from open data. For example, you can explore topics that interest you with open data, use open data as material in research, download data for your own use, and use it for example in application development" (https://www.avoindata.fi/fi/opas/avoin-data-kansalaiselle)

5) *Myth of open data as a path to open government*

"Opening up data, i.e. making it available to everyone, promotes, for example, the openness and transparency of administration, and thus benefits all citizens. (https://www.avoindata.fi/fi/opas/avoin-data-kansalaiselle)

"Open data increases the transparency of official activities and increases the opportunities for civic participation and access to information" (Valtiovarainministeriö 2013, 1)

6 COMPARISON OF THE OPEN DATA STRATEGIES

6.1 Comparison of the criteria of access

This chapter compares the principles contained in the open data strategies of Denmark, Estonia and Finland, by comparing them with the general principles of the publication of open data and, for example, the principles compiled by the Finnish National Library (Chapter 2.4). The comparison takes into account that the strategic guidelines of the three countries under investigation are written in different ways and cannot easily be found in exactly the same formulations (see also Kassen 2018a and Kassen 2018b).

All three countries have outlined the principles of the protection and confidentiality of personal data in the same direction. Similarly, the principles governing the distribution and interfaces of information are similar. There are differences in the free-of-charge policies for information. In Estonia, policies concerning the free of charge of data (or the costs of dismantling data) are still under way, for example with regard to the provision of open spatial data, the trade register and the real estate register. In Denmark and Finland, spatial data sets are available free of charge, as are trade register data. There are also differences in the definitions of the structure of information. Denmark and Finland emphasise the use of standards, but in Estonia detailed policies are still to be drawn up.

Criterion	Denmark	Estonia	Finland
The information published as open data is public information (the protection of personal data and privacy are guaranteed when the information is published)	*	*	*
The data must be in a machine-readable format so that it can be processed by different applications and the data can be transmitted through different interfaces.	*	*	*
The data may be used free of charge or, at most, the costs of removing the data may be charged for	*	- (no defined policy)	*
The data publisher (producer)	*	*	*

enables the re-use and further processing of data under clear terms of use (Creative Commons or equivalent)			
The structure of the data is described clearly and in accordance with the metadata standards.	*	- (data mostly in plain text format)	*

Table 4: The implementation of the criteria for access to open data in Finland, Denmark and Estonia

An interesting comparison between the Danish basic data principles (5-stars model) and Finland basic data resource policy has been made by the Finnish Patent and Registration Office, the National Land Survey of Finland and Population Register Centre (Patentti- ja rekisterihallitus 2017) . In connection with the national interoperability study, the basic principles of the Danish 5-stars of basic data model were compared with the principles presented in the main documentation of the Finnish reference data architecture.

Criterion	Denmark	Finland
Quality	Principle 1: Basic Data must be of sufficient quality for the major uses of the data), the basic data must be of sufficient quality in most cases of use.	"In order to speed up and expand data exploitation, attention must be paid to basic data quality and quality control. The quality of data maintained by public administrations is trusted.
Accuracy and timeliness and responsibilities	Principle 2: Responsibility for keeping the Basic Data valid and up-to-date is clearly placed and efficiently handled), ensuring the accuracy and timeliness of basic data must be clearly responsible and effectively managed	The architecture lacks direct references to the accuracy. On timeliness, the aim is for the operational utilization of data to have the shortest possible time from the generation of the data to its introduction.
Interoperability	Principle 3: Basic Data is semantically coherent and modeled accordingly to the Model Rules of Basic Data), in the target level, the basic data are semantically compatible and modeled according to the basic data modelling rules	The wide availability of basic data is another important goal that requires interoperability between data resources and their interface services and data products. In addition to the operational use of data, the use of basic data in analytical processes must be taken into account.
Transparency and accessibility	Principle 4: Basic Data is available for free and with non-restrictive terms for reuse (does not apply to sensitive data)	The reference architecture for basic data resources lacks references to data transparency.

	through the Data Distributor) define transparency and availability as follows:	On access to data the recommendation is that services providing data from basic data resources are available in a way that the structure of the services is not restricting the use of data distribution
Broad use of data	Basic data is used widely in the public sector and elsewhere	Extensive use of basic information is another important goal

Table 5: The implementation of the Basic Data 5-stars criteria in Denmark and Finland

As a conclusion of the comparison between the Danish and Finnish approaches, it can be stated that many of the principles listed in the Danish 5-stars list are also included in the Finnish reference architecture. It is slightly surprising that the missing points in the Finnish basic data policy are correctness and transparency. In the Finnish reference architecture, openness is seen as an element on which there has to be an agreement between the data provider and the data user (Patentti- ja rekisterihallitus 2017)

6.2   Access to Open Data - Analysis of the Datasets

The content analysis of the open data content examines the data contained in the national data portals and the European data portal europeandataportal.eu (datasets). Databases have two thematic themes, namely traffic and environmental information. The data of the European Data Portal has also been compared to the data of national data portals, as the material of all national portals is not included on the European portal. Materials are analyzed with in the evaluation criteria presented in Chapter 1 (Berners-Lee Criteria, Hyvönen Criteria and Five Star Criteria in Denmark).

Denmark

The environmental Information

When looking for environmental information about the Danish Open Data Portal, the search term "miljø" found a total of 130 data sets. Most of these were produced by a single municipality (municipality of Vejle, open data such as parking spots, cycling routes, green areas and public buildings). Part of the data is available in map data in WMS (Web Map Service) and WFS (Web Feature Ser-

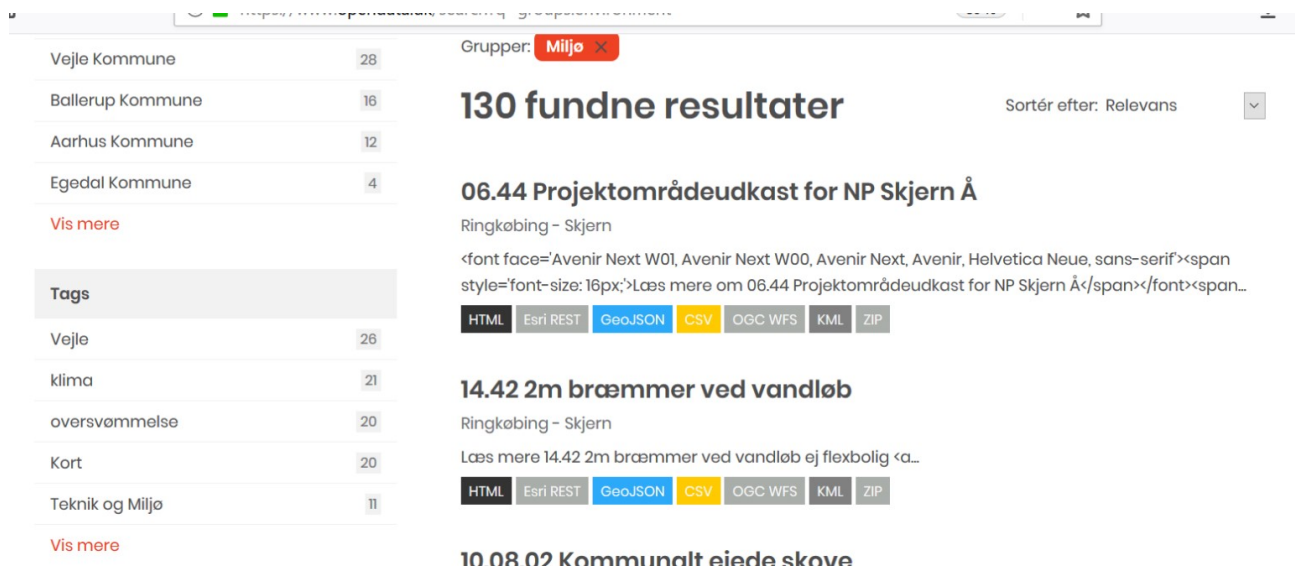vice) formats. A total of 68 data sets are distributed in open-source information in widely used JSON format.

Figure 5: User interface of the Danish opendata.dk portal (https://www.opendata.dk/)

Traffic information

When retrieving traffic data from the Danish Open Data Portal, the search term "transport" or "traffic" found a total of 70 data sets. The material consists of, for example, municipal information on traffic data, parking information areas, bicycle traffic information, traffic accident data, etc. 70 data sets in 25 have JSON format for open data distribution and 29 GeoJSON formats.

Estonia

The environmental Information

When looking for environmental information about the Estonian open data portal, the search term "keskkond" found a total of 23 data sets. These data sets include, for example, Estonia's water supply and information on the services of the Tallinn and Tartu cities (public sports facilities, toilets, public transport). Datasets are only available in html format on the website (veebileht). No standard format for open data is available.

Traffic information

When retrieving traffic data from the Estonian Open Data Portal, the search term "transport" found a total of 12 data sets. These data sets include, for example, the national address database, the Tallinn and Tartu service directories and the data on business licenses.

Datasets are available only in html form (veebileht) or xml format (business data). No standard format for open data is available.

Finland

The environmental Information

When searching for environmental information on the Finnish Open Data Portal, the search for "environment" found a total of 407 data sets. These datasets contain information on, for example, environmental permits, environmental reports, traffic noise, air quality, nature destinations, outdoor routes, etc.

About 20% of these data sets are merely excel files. Open data distribution formats are used in map data WFS and WMS. The JSON format is used in 13 data sets.

Traffic information

When retrieving traffic data from the Finnish Open Data Portal, the search term "traffic" found a total of 407 data sets. These datasets contain information on, for example, traffic places, traffic routes, traffic statistics, traffic lights, traffic accidents, etc.

About 15% of these data sets are merely excel files. Open data distribution formats are used in map data WFS and WMS. The JSON / REST format is used in 20 data sets.
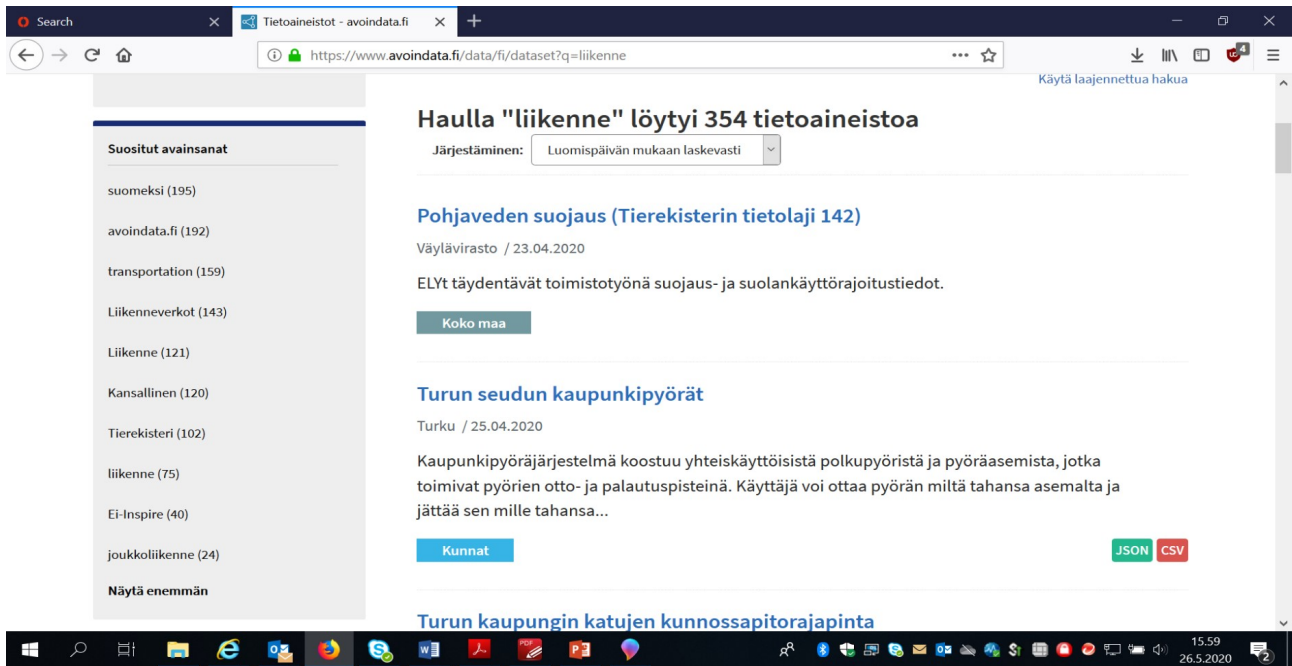
Figure 6: User interface of the Finnish open data portal www.avoindata.fi

6.3   Availability of open data - comparison

The availability of open data in Denmark, Estonia and Finland can be analyzed using Berners-Lee's five-star criteria, which are the most commonly used tool for assessing the quality of open data. The five-star criteria are usually associated with the file formats in use. At the one-star level, plain text files (txt, html or pdf) are used as the most common file formats. For example, excel files are used at the two-star level. At the three-star level, system-independent formats such as CSV or XML are used. At the four-star level, file formats and query languages based on semantic network standards are used, such as RDF, SPARQL, JSON or stand-alone applications (APIs) (Frosterus et al. 2011, 96-97; Hyvönen 2018, 115-116).

Using open data quality criteria, the situation can be illustrated as follows:

Criterion	Denmark		Estonia		Finland	
	Trans-port 216	Envi-ronment 130	Trans-port 12	Envi-ronment 23	Trans-port 354	Envi-ronment 469

*available on the web (in any data format txt, html, pdf, jpg) with open license	link 14 html 47 pdf 24 docx 3	link 15 html 19 pdf 9		html 14	pdf 12	pdf 29 html 14 zip 7
** available in machine-readable structured format (e.g. excel file and not as picture file)	excel 63 docx 3 ods 1 xyz 1	excel 21			excel 35	excel 44 docx 4
*** like ** but in system inde-pendent format 8e.g. CSV and not excel)	csv 109 kml 10 wfs 58 shp 66 wms 53 wmts 43	csv 45 shp 43 wfs 29 wms 23 kml 20 xyz 14 wmts 12 kmz 7	api 1 gtfs 1 xml 9	xml 8	csv 45 wms 37 wfs 30 shp 24 kml 17 geoser-vice 13 xml 11 graphql 9 tab 6	csv 27 wms 83 wfs 60 xml 48 shp 32 pxweb 31 pc-axis 30 tab 22 kml 9
**** in addition: W3C open standards are used (RDF ja SPARQL) to identify data, with possibility to direct references to data	GeoJSON 136	GeoJSON 57 JSON 11	JSON 1	JSON 1	GeoJSON 16 JSON 17	JSON 13 JSON/ REST 7

Table 6: Comparison of the data formats in the open data repositories of Denmark, Estonia and Finland

Based on the study, the differences between Denmark and Finland on the one hand and Estonia on the other are clear. In Denmark and Finland, a moderate amount of open data datasets are available

in a format that enables versatile utilization of the data (GeoJSON, JSON and JSON / REST related to geodata). In Denmark, up to half of the datasets are available in JSON format, and in Finland about 5% of the datasets are available in JSON format. In Estonia, most of the open data is available only in html or xml format and does not make use of open data metadata standards.

7 RESEARCH RESULTS AND CONCLUSIONS

The main findings of the study on open data strategies and practical dissemination concern, firstly, the differences between national open data strategies and, secondly, the differences in the distribution of open data and, in particular, the availability of linked open data. Of the national open data strategies, the Danish strategy has taken a broad view on open data (Digitaliseringsstrategi 2016), although the recommendations for action are narrow. The Danish strategy clearly emphasizes more than Finland or Estonia the importance of society's basic data resources (Grunddata) and the need to utilize them and to make them easily available. In Estonia, the release of open data has been seen as a relatively small part of the digitalisation of society. In Estonia, more emphasis has been placed on the rapid digitization of services, the digital by default approach and the development of citizens' digital skills. The Estonian guidelines strongly emphasize the anonymisation of information as a prerequisite for open data dissemination (Eesti Avaandmete 2016). In Finland, emphasis has been placed on the protection of privacy, but the need for anonymisation of information has only been referred to in a recent study on the conditions for producing open data (Koski et al. 2017, p.59). In Denmark and Finland, attention has been paid to the importance of the platforms and interfaces for the use of open data, but in Estonia the conditions of use have been dealt with mainly in terms of equal distribution of information.

There are key differences in open data strategies in matters concerning the provision of open data and the utilization of international open data standards. In Denmark and Finland, the guidelines for the opening and delivery of open data have already been fairly advanced, but in Estonia the guidelines have only been specified in more detail since the opening of the open data portal in 2018. Differences in open data management are also significant. In Estonia, the role of the state is more central than in Finland or Denmark. In Denmark, the open data portal is mainly the responsibility of municipalities.

As to the myths which prevail in the high-level discussion on open data, there are differences in the utilization of open data myths (which were discussed in chapter 2.7.) in these three countries. In Denmark, the most popular myth is the myth of automated benefits in the release of data. Another essential myth refers to the data without limits i.e. unlimited information delivery. The third myth in the Danish approach has seen open data as a direct path to open government. In Estonia there is strong emphasis on the myth of public data, describing all public data as open data. A second myth

in the Estonian documentations is the myth of automated benefits in the release of data. The third myth involves the perception that the release of open data is a path to open government. In Finland the most popular myths have been the myth of automated benefits in the release of data, the myth of easy utilization of data by all citizens and myth of open data as a path to open government.

As a conclusion it may be stated that in all three countries an equality sign has been made between open data and open government. In this comparison, Estonia is the only country that sees all public data as open data and Denmark is the only country that sees the potential in unlimited information delivery. It is interesting to notice that the myth of easy utilization of data by all citizens is utilized in Finland, but not in Estonia.

There are also significant differences between Denmark, Finland and Estonia in the use of open data distribution standards and the extent of open data. In Denmark and Finland, there are advanced open data file formats in common use (eg JSON / REST), but in Estonia, open data is mainly published on the website in html format. There are also clear differences in the volume of data published. In Denmark and Finland, cities in particular have published open data in a very diverse way. In Estonia, only the largest cities (Tallinn and Tartu) have started publishing open data. In Estonia, the wide supply of open data is only just beginning, as compared to the well-established practices in the production and distribution of open data in Denmark and Finland.

The findings of this study indicate that there remains a gap between the high-level statements on open data and the implementation of the open data policies at the grassroot level. It is fairly easy to prepared and publish open data strategies, but it is much more difficult to reach practical results. There is an urgent need to focus future research in the more detailed comparison of the national implementation of open data, to be able to get a more comprehensive view of open data in practice.

REFERENCES

Attard, J., Olandi, F. Auer, S. 2016. Value Creation on Open Government Data. IEEE Explore Digital Library. http://ieeexplore.ieee.org/document/7427509/ <Viewed 4.4.2020 >

Bengtsson, M. 2016. How to plan and perform a qualitative study using content analysis. Nursing-Plus Open Vol. 2, 2016, pp. 8-14.

Berners-Lee, T. 2012. Five-star deployment scheme for Open Data. http://5stardata.info/en/ <Viewed 8.4.2020 >

Bong, S., Debunking Myths in Qualitative Data Analysis. Forum Qualitative Sozialforschung 3(2) · April 2002, 1-16.

Chernoff, M. 2010. What "open data" means – and what it doesn't. http://opensource.com/government/10/12/what-"open-data"-means-and-what-it-doesn't

Danish Ministry of Finance. 2019. National Strategy of Artificial Intelligence. Copenhagen.

Eesti Avaandmete Portaal. 2016. Avaandmete loomise ja avaldamise juhend (Eesti Avaandmete 2016)

Digitaliseringsstyrelsen. 2016. Et Stærkere og Mere Trygt Digitalt Samfund. Den fællesoffentlige digitaliseringsstrategi 2016-2020. Digitaliseringsstyrelsen: København (Digitaliseringsstrategi 2016) https://digst.dk/strategier/digitaliseringsstrategien/ <Viewed 7.4.2020 >

Elo S., Kääriäinen M., Kanste O., Pölkki T., Utriainen T., Kyngäs H. 2014. Qualitative Content Analysis: A Focus on Trustworthiness. SAGE Open, January-March 2014: 1 –10.

Estonian Ministry of Economic Affairs and Communications. 2016. Digital Agenda 2020 for Estonia. Tallinn. (Estonian Digital Agenda 2020)

Euroopan komissio. 2011. Euroopan komission tiedonanto: Avoin data - Innovoinnin, kasvun ja läpinäkyvän hallinnon moottori. KOM(2011) 882 lopullinen. Bryssel 2011

Euroopan komissio. 2020. Komission tiedonanto Euroopan parlamentille, neuvostolle, Euroopan talous- ja sosiaalikomitealle ja alueiden komitealle. Euroopan datastrategia. COM/2020/66 final

Euroopan parlamentin ja neuvoston direktiivi 2003/98/EY, 17.11.2003 julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (EYVL 345, 31.12.2003, 90) (PSI-direktiivi)

Euroopan parlamentin ja neuvoston direktiivi (EU) 2019/1024, 20.6.2019, avoimesta datasta ja julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä (EUVL L 172, 26.6.2019, 56-83) (Avoin data –direktiivi)

European Data Portal 2020. The Economic Impact of Open Data. Opportunities for Value Creation in Europe. https://www.europeandataportal.eu/sites/default/files/the-economic-impact-of-open-data.pdf < Viewed 29.03.2020 >

Folketinget Rigsrevisionen. 2019. Beretning 12/2018 om åbne data. Folketinget: København (Folketinget 2019)

Frosterus, M., Hyvönen, E., Laitio, J. 2011. Creating and Publishing Semantic Metadata about Linked and Open Datasets , teoksessa: Wood, D. (toim.) Linking Government Data, Cham: Springer, ss. 95-112.

Homburg, V. 2005. The Myths of E-Government. Looking Beyond the Assumptions of a New and Better Government. Proceedings of the IRMA 2005 Conference "Managing Modern Organizations through Information Technology.

Horst, N.L., 2014. The Basic Data Programme – A Danish Infrastructure Model for Public Data. Perspektiv 24/2014, 36-48.

Hossain, M.A., Dwivedi, Y.K., Rana N.P. 2016. State of the Art in Open Data Research: insights from existing literature and a research agenda. Journal of Organizational Computing and Electronic Commerce, 26:1-2, 14-40.

Huijboom, N., Van den Broek, T. 2011. Open data: an international comparison of strategies. European Journal of E-Practice. No. 12, March/April 2011.

Hyvönen E., Tuominen J., Alonen M., Mäkelä E., 2014. Linked Data Finland: A 7-star Model and Platform for Publishing and Re-using Linked Datasets. Lecture Notes in Computer Science, vol . 8798. Cham: Springer.

Hyvönen E., 2018. Semanttinen Web. Linkitetyn avoimen datan käsikirja. Helsinki: Gaudeamus.

Implementation of the Public Sector Information Directive. European Commission. https://ec.europa.eu/digital-single-market/en/implementation-public-sector-information-directive (Implementation 2017). <Viewed 03.04.2020>

Jaatinen, T. 2014. Julkisen sektorin avoin data ja henkilötietojen suoja. Pro gradu –tutkielma, Helsingin yliopisto, Helsinki 2014.

Janssen, M., Charalabidis Y, Zuiderwijk, A. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management (ISM), vol. 29, no.4, 258-268

Kansalliskirjaston tiedon avoimen saatavuuden periaatteet. 2011. http://www.doria.fi/handle/10024/88839 (Kansalliskirjaston periaatteet 2011).

Kassen, M. 2018. Adopting and managing open data: Stakeholder perspectives, challenges and policy recommendations. Aslib Journal of Information Management, vol. 70, nro 5, 518-537. (Kassen 2018a)

Kassen, M. 2018. Open data and e-government – related or competing ecosystems: a paradox of open government and promise of civic engagement in Estonia. Information Technology for Development, 25 (3), 552−578. (Kassen 2018b)

Kassen, M. 2019. Open Data Politics: A Case Study on Estonia and Kazakhstan. Bern:Springer (Kassen 2019)

Knuuti, M. 2015. Avoin data ja PSI-direktiivi julkisuuden toteuttajina. Pro gradu tutkielma, Lapin yliopisto 2015.

Koski, H., Honkanen, M., Luukkonen J., Pajarinen, M., Ropponen T. Avoimen tiedon hyödyntäminen ja vaikuttavuus. Valtioneuvoston selvitys- ja tutkimustoiminta. Selvitys- ja tutkimustoiminnan julkaisusarja 40/2017. Helsinki 2017

Krippendorff, K., 2012. Content analysis: an introduction to methodology. New York: Sage.

Liikenteen ja viestinnän avoin tieto. Työryhmän raportti, LVM   Julkaisuja   10   /   2013. https://www.lvm.fi/-/liikenteen-ja-viestinnan-avoin-tieto-tyoryhman-raportti-811423   < Viewed 29.03.2020 >

Lohtander H., 2013. Avoin data ja semanttinen verkko – yhdessä kohti älykkäämpää internetiä . Pro gradu tutkielma. Tampereen yliopisto, informaatiotieteiden yksikkö.

Lubarsky, B. 2017. Re-Identification of "Anonymized" Data. Georgetown Law Technology Review 202 , April 2017.

Magoulick, M. 2015. What is Myth? https://faculty.gcsu.edu/custom-website/mary-magoulick/defmyth.htm

Margetts, H, Naumann A. 2017. Government as a Platform. What can Estonia show the world? Oxford Internet Institute. https://www.politics.ox.ac.uk/materials/publications/16061/government-as-a-platform.pdf

Mayring, P. 2014. Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173

McBride, K., Olesk M., Kalvet T., Krimmer R. 2018, Leader in e-Government, Laggard in Open Data: Exploring the Case of Estonia. Revue française d'administration publique no 167, 2018, 611-624

Merriam-Webster. Online Dictionary. https://www.merriam-webster.com/dictionary/ <Viewed 2.5.2020 >

OECD. 2015. Public Governance Reviews. Estonia and Finland - Fostering Strategic Capacity across Governments and Digital Services across Borders. Paris: OECD.

Open Data Denmark. http://www.opendata.dk   <Viewed 2.4.2020 >

Open Data Maturity in Europe 2019. Insights into the European state of play. European Commission, Cap Gemini Consulting. Brussels: European Commission.

Open Government Data. 2007. https://opengovdata.org/ <Viewed 02.04.2020>

Patentti- ja rekisterihallitus. 2017. PRH-MML-VRK Yhteentoimivuuden selvitystyö. https://www.avoindata.fi/data/fi/dataset/prh-mml-vrk-yhteentoimivuuden-selvitystyo

Poikola A., Kola, P., Hintikka, K. 2010. Julkinen data. Johdatus tietovarantojen avaamiseen. Liikenne- ja viestintäministeriö, Helsinki.

PriceWaterhouseCoopers. 2019. Open data and e-government good practices for fostering environmental information sharing and dissemination. Copenhagen: European Environment Agency.

Røyen, J.K., 2014, The Danish Basic Data Programme and the governance perspective. Copenhagen: Digitaliseringsstyrelsen,

Simperl, E., O'Hara, K, Gomer R. 2016. Open Data and Privacy. European Data Portal (https://www.europeandataportal.eu/sites/default/files/open_data_and_privacy_v1_final_clean.pdf)

Tuomi, J., Sarajärvi, A. 2002. Laadullinen tutkimus ja sisällönanalyysi. Helsinki: Tammi.

Valtioneuvosto. 2011. Valtioneuvoston periaatepäätös 03.03.2011 julkisen sektorin digitaalisten tietoaineistojen saatavuuden parantamisesta ja uudelleenkäytön edistämisestä. (VN-periaatepäätös 2011)   https://www.lvm.fi/-/julkisen-tietoaineiston-saatavuudesta-periaatepaatos-784401

Valtiovarainministeriö. 2013. Julkisen hallinnon tietovarantojen avaamisen periaatteet. Valtiovarainministeriön muistio VM041:00/2013.

Valtiovarainministeriö. 2016. Avoimen datan opas. https://www.avoindata.fi/fi/opas (VM Avoimen datan opas)

Valtiovarainministeriö. 2015. Avoimesta datasta innovatiiviseen tiedon hyödyntämiseen. Avoimen tiedon ohjelman 2013–2015 loppuraportti. Valtiovarainministeriön raportti 31/2015. Helsinki (VM Avoimen tiedon ohjelman loppuraportti 2015).

Valtiovarainministeriö. 2015. Avoimen tiedon vaikuttavuus. Valtiovarainministeriön julkaisuja 15a/2015. Helsinki. (VM Avoimen tiedon vaikuttavuus)

Zuiderwijk, A., Janssen, M. 2014. Open data policies, their implementation and impact: A framework for comparison. Government Information Quarterly 31 (2014), 17–29.