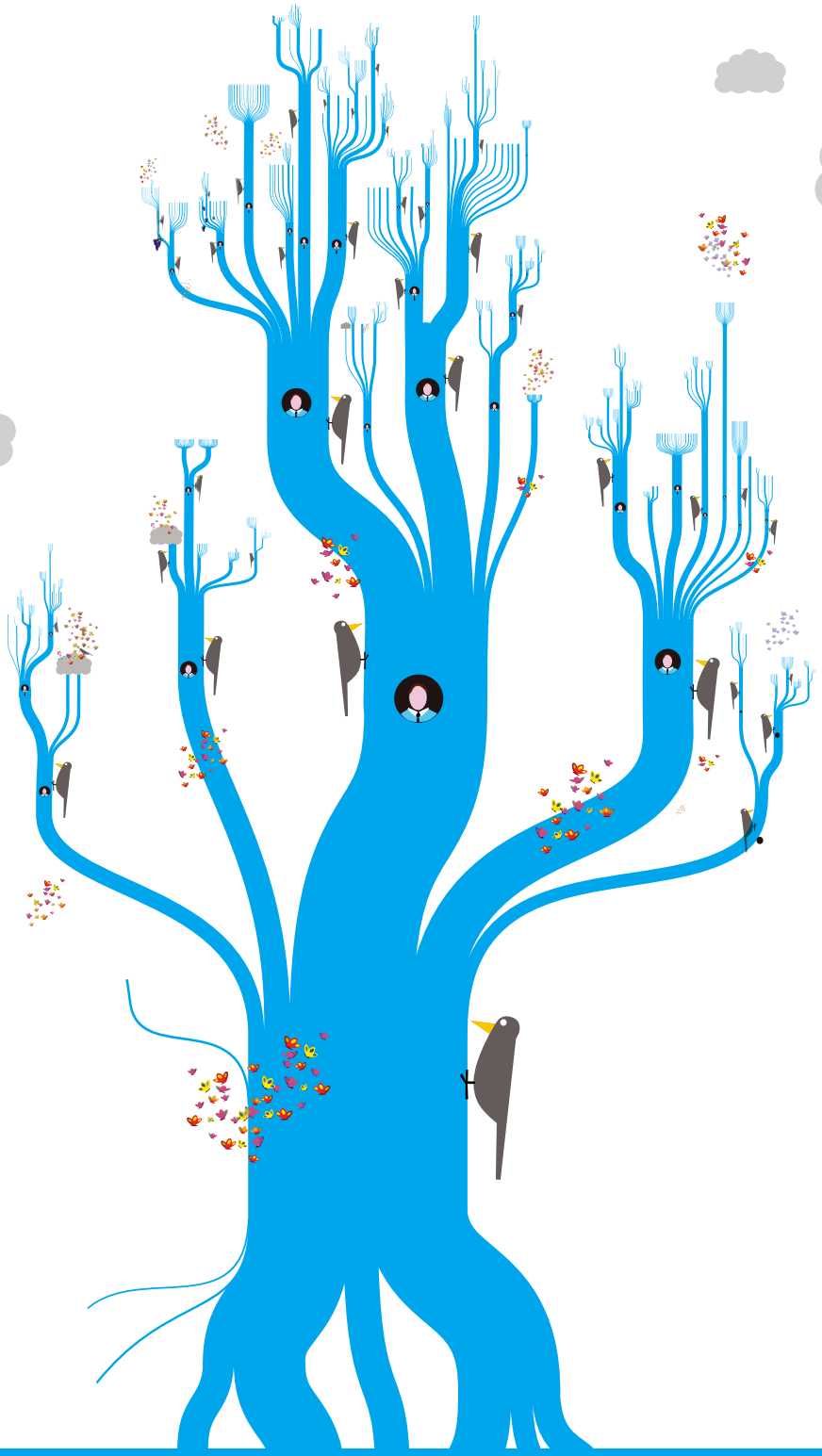
Ministry of Transport and Communications



Antti Poikola
Petri Kola
Kari A. Hintikka

Public data

an introduction to opening information resources

Case: **Tax tree**

Cover image: **Peter Tattersall – Tax tree**

Peter Tattersall was the winner of the idea-category in the 2009 Apps for Democracy Finland contest. Peter's idea is rather simple. Even for the professionals, the Budget of Finland is heavy reading. The budgetary materials have been available for some time, and the media has been trying to make it more visible.

The tax tree is an idea of an Internet service, one that would make the revenues and expenditures of an organisation ran by state, municipality or public administration even more visible. Revenues are the roots of the tree. They then become a part of the stem and finally they branch out as expenditures. The attained profits are represented as leafs and fruit. The thickness of the roots and branches are equivalent to the amount of sources of income and the items of expenditure.

In order to work, the tax tree and other similar applications need open data in a machine-readable format. The Netra service (www.netra.fi) is run by the State Treasury of Finland, and it provides information on the operation, resources and profitability of the state. However, Netra is designed to specifically serve the needs of the Finnish Government. It is the aim of this handbook to answer one simple question: what would we need to do to put the tax tree into practice in Finland?

Translation: **Emilia Norlamo**

Layout: **Kirmo Kivelä** and **Inari Savola**

ISBN 978-952-243-237-7 (print)

ISBN 978-952-243-238-4 (web publication http://www.lvm.fi)

Edita Prima Oy, Helsinki 2010

Public data

an introduction to opening information resources

Antti Poikola | Petri Kola | Kari A. Hintikka

Foreword

Public administration is in possession of data sets, the opening and free use of which are supported by international examples and social discussion.

The importance of the availability of public data in connection with productivity, competitiveness, and well-being has been emphasised by The Ubiquitous Information Society Board in Finland. It has been suggested that it should be made easier for the companies to access information in digital web services. In addition, Finland has been seen as a forerunner of open data societies in the future. The role of the data as a basic infrastructure is the central theme of the National Digital Agenda in Finland. Ministry-appointed working groups are trying to find solutions for improving the availability of public data. In March 2011, the Finnish Government accepted n agreement in principle, according to which data sets have to be openly available for everyone to reuse and marked with uniform and clear terms of use. Data transfers are viewed in light of their over-all benefits to national economy, which, as a principal rule, means non-chargeability. In addition, the European Union Commission has stated that the Member States need to take action to improve the utilisation of public data.

However, there is still rather little know-how to be applied into practice. Prior to this guidebook, there was no comprehensive guide available on the topic in Finland. This book aims to offer guidance on how to open data sets in a controlled way. Services and producing communities are emerging around the data sets. Introducing the building blocks of this ecosystem is in the centre of what this book offers. In the Finnish information society, people are beginning to see openness of data as an answer to societal and economic challenges. In addition to new opportunities for business, the openness of public data increases the sufficiency of the government and the participation of citizens. The benefits do not only include web services; the effects can be seen widely in our surroundings, for example in environmental planning.

At its best, open data creates a culture of doing things together, a culture that is enabled by the communal and technological development of the Internet. During this change the legal and administrative practices need to be updated. For now, we are only able to estimate the effects of open data but we cannot stand still and wait. It has been said that Finland has the possibility to become the leader among open information societies because of our high-quality public data sets and technical know-how. In order to become the leader, we need trial runs carried out open-mindedly.

The three authors of this guidebook are well capable of processing domestic and international expertise and have, therefore, impressively fulfilled the goal of this book; to let people know how to use data sets innovatively and productively. Therefore, it is the purpose of this book to support the producers and the users of data to create these practices.

Helsinki, April 20, 2011.
Taru Rastas,
Ministry of Transport and Communications

MATTI MATTILA / FLICR.COM | CREATIVE COMMONS

Abstract

Information resources produced and held by public administration have recently been widely discussed both in Finland and in the European Union. At the beginning of 2010, several working groups were formed in Finland and many opinions were heard; all asking the current legislation and methods to be revised.

The current definition of policy in Finland regarding PSI (Public Sector Information) is based on the Act on Criteria for Charges Payable to the State, which was passed in 1992. According to the Act in question, using data produced in the public administration is, more often than not, chargeable. The Act on Criteria for Charges Payable to the State was first passed when the Internet did not yet exist. Nowadays, during the Internet era, the costs for offering data are substantially lower than in the 1990s.

Free distribution of government held data would be overall profitable for Finnish businesses and civic activities. In addition, it would help make the government more effective. No accurate calculations are available on the topic yet. However, studies and reports show that currently most of the income from transferring data comes from within the government. According to our point of view, opening the public sector's data for free would be more profitable overall than the current system.

This book gives an overall picture of the process of opening the administration's data resources for free and open use for everyone. Opening government data has already been made part of operational policy and strategies in the US and Great Britain. In this handbook, we take a look at opening government data in a wider societal frame of reference.

Opening the data begins with evaluating the organisation's own information resources. This might be a long process, depending on the size and nature of the organisation. However, everything does not have to be done immediately. The opening process could proceed phase by phase, starting from the easier data and gradually moving on to more complex data sets. During the data inventory, organisations may come across data they had no knowledge of or did not know how to utilise.

During the inventory, organisations can create their own strategies and goals on how to utilise their data. Possible benefits include new ways to use the data, collaboration with new partners, or the development of the organisation's role. This guidebook is a toolbox, giving you the necessary tools to estimate the usability of the data. After the inventory, all data should be converted into machine-readable format. More and more often, data is applied to Internet and mobile applications. These applications offer extra value to the users by allowing them to access certain information without browsing the Internet. Finland has produced some high quality data resources but, in most cases, the information is published solely on PDF-format, which makes it harder to add value to the data.

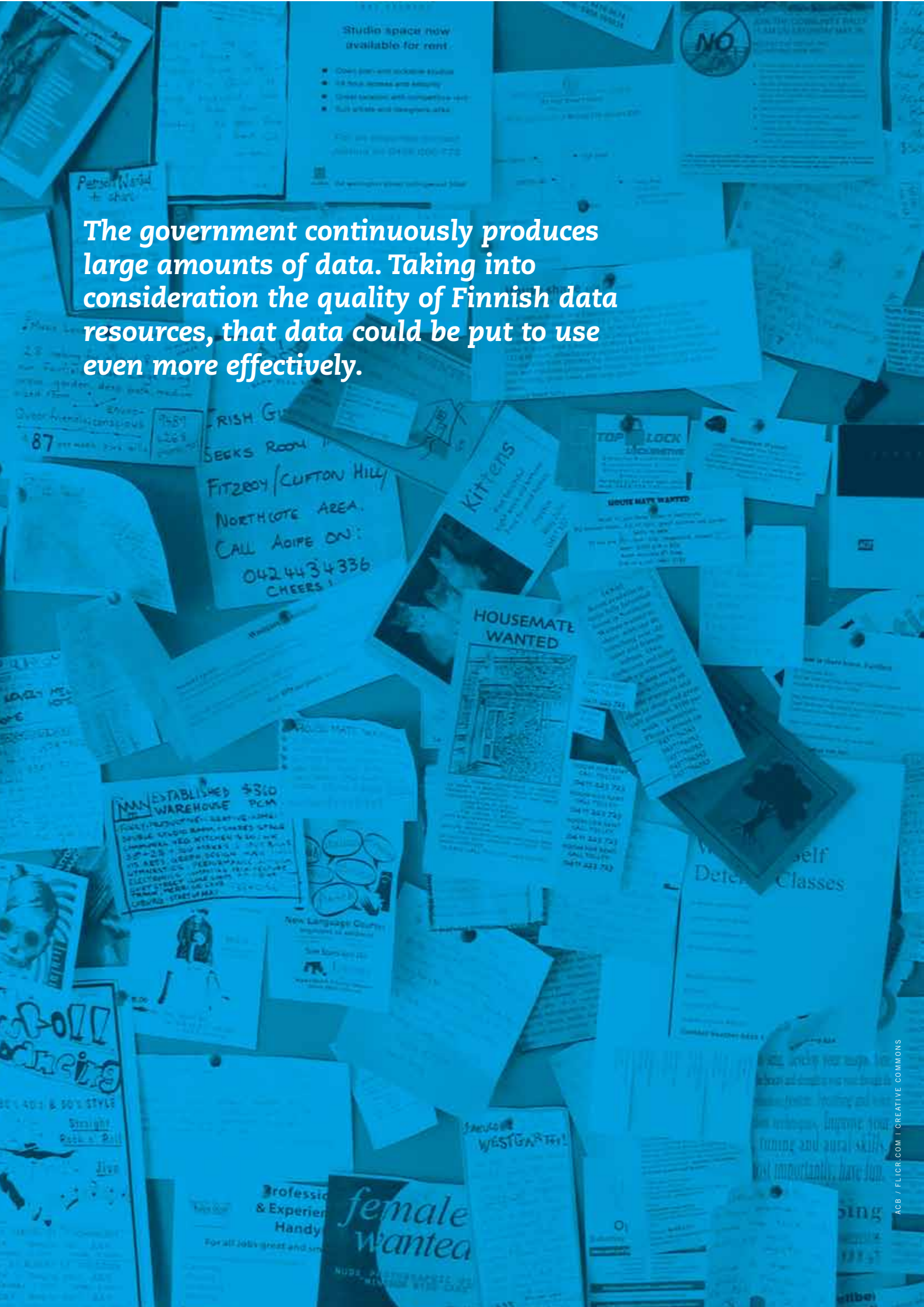
Many laws, Directives and recommendations need to be considered during the process of opening public data. These include, among others, the Freedom of Information Act, data protection legislation, Act on Criteria for Charges Payable to the State, the Copyright Act, international recommendations, competition legislation and EU Directives. None of the laws and regulations mentioned above prevent opening data. There are, however, parts in the regulations and the legislation that one should be familiar with to ensure a controlled opening process.

Opening data can be seen as an interactive process, mainly because often the best ways to use data are developed outside the organisation. In this book, we

compare the process of opening data to an ecosystem, where various actors offer data without reciprocity, in a way which benefits all parties involved. Another way of looking at open data is to see it as something infrastructural. Open data can be perceived as part of the infrastructure as both the enabler and the content, much like streets and electricity.

To date, there are no institutions coordinating the opening of data in Finland. In order to ensure coordinated progress that is as effortless to an organisation as possible, a clearing house of open public sector data could be set up in Finland. This clearing house would coordinate practical issues, offer guidance to the government, and solve problems, much like the Consumer Agency. In addition, a data catalogue could be developed in Finland, one that would list all public sector data.

Contents



The government continuously produces large amounts of data. Taking into consideration the quality of Finnish data resources, that data could be put to use even more effectively.

Introduction

The government produces, holds and administers wide information resources with great financial and societal value. At the moment, the only ones being able to use this raw data are the ones who have access to the information resources. According to estimates, only a small portion of available data ends up being reused. The development of communal and technological characteristics of the Internet opens up new possibilities for creating more open data policies. Currently, allowing the national data resources to be used free of charge, and a change in the current culture are the most important things to do to further develop the Finnish information society. These changes would bring new ways to face the current challenges. Free public data could offer the Finnish institutions an opportunity to renew themselves and learn the cooperation skills of the network era. It could also steer the development of Finnish society into a new culture of joint cooperation.

Recently, the idea of allowing private companies, research institutes and other interested parties to access public administration's data has been unanimously supported. The matter has been widely discussed, thanks to our national objectives and international examples. Therefore, it is our understanding that conversation on the principals of open data is no longer needed; the pressure is now on improving know-how and creating practices through guidance.

Open Data as a Source for Interdisciplinary Benefits

Until now, the conversation in Finland about the open and free government data has focused on benefiting businesses and national competitiveness. In the US, the leading country in market economy, the legislation regarding open data has for long been founded on the idea that all tax-paid data should be free and accessible to everyone. This point of view has led to financial activity and significant reuse of public data. In Finland, as in the rest of Europe, the basis has been quite different and the attitudes towards data resources have been rather protectionist. However, the preconditions for a new way of thinking keep growing continuously when open models keep developing, largely due to the Internet.

The authorities collect data primarily for their own use. The valuable, reusable data resources are a by-product of other activities. Many of the examples presented in this book came about outside the data producing organisation, in unforeseeable contexts' and joined with other data sources. Some of the examples were surprising to both the operators and the partners.

Opening public data resources contributes to at least three different goals: the transparency of democracy and administration, creation of new innovations and markets, and the increase of efficiency within the government. These goals are further explored in future chapters of this book. In addition to financial benefits, the new operational policy would strengthen our identity, in other words, it would force us to think about how we are going to live in the middle of the web era in the future, and how to utilise the data we produced together.

Open data and the machine-readable format of it help lighten the processes. However, this possibility has not been widely used. Cooperation can be further promoted through open communication and transparency. Free distribution of open data can prevent unintentional sub-optimisation of certain sectors. For example, government agencies, trying to minimise the expenses, keep their own list of customers instead of buying the service from Population Information System.

Content of the Guidebook

So far, no guidance or established practices on how to utilise open public data have existed in Finland. It is the aim of this book to fill that gap. This is a hands-on introduction first and foremost for the employees in the public administration, who want to further promote the opening of the data within their own organisations. Furthermore, this guide is useful to everyone interested in the topic.

Open public data, both as a theme and as a phenomenon, is a challenging area of examination. That is why it is both interesting and difficult to write a decent guide on the topic. In this book, we decided on the following order. At first, we present the topic and the concepts, practices and characteristics related to it. Next, the examination becomes more concrete as we take a look at the legal, financial and technical issues related to the distribution of open data. As a conclusion, the final chapter is about creating an infrastructure for open data.

One of the basic themes of the book examines opening government-produced data to everyone for free in a way which enables reusing the data in a machine-readable format. Using examples, this guidebook presents all the things one should keep in mind when opening data. In addition, we take a look at the co-operation between the government and the user communities. This cooperation supports the reuse of data and, therefore, adds to the societal and economic effectiveness.

In addition, this guidebook presents the best international practices adapted to match the Finnish culture and environment. This adaptation was done with one idea in mind; parts of the theories could be put to use immediately. A specific guidance to practical solutions is not easy because good practices are, for the most part, only now forming. We hope that this book will pave the way for first experiments and practice development.

The guidebook also examines the essential questions about technical issues related to opening data, evaluating the implementation and working together with communities. Due to the extent of the topic we offer an overall view on specific matters. Those interested in the subject most likely find the bibliography very helpful when searching for more information. Some interesting themes related to open data, such as data protection, IT architecture of the government, digital gap, usage of records and common discussion on copyrights have been left out.

The introduction and chapter 1 offer background information to those readers who are not yet familiar with the topic. Furthermore, the introduction presents some of the terminology used in this guidebook.

Chapter 1: *Extensive Use of Data as a Goal* analyses the organisation's attitudes towards co-operation, networking and other activities outside the organisation when the goal is to enable the emergence of new services based on open government data.

Chapters 2, 3, 4 and 5 help organisations shape their own viewpoints on open data. The content of the chapters include exemplifications and interviews. **Chapter 2**: *Organisation's Views on Openness* presents a suggestion for the process of making an inventory and opening data resources in an organisation. In this chapter, we also present some tools on how to estimate the usability of data regarding e.g. terms of use, machine-readability and other related matters. The following chapters take a closer look at the legal and technical aspects of the themes. **Chapter 3**: *Permission to republish and reuse* summarises the central laws and regulations. **Chapter 4**: *Financial Views on Open Data* pulls together recent discussions on how the production and distribution of government data is financed nowadays, what factors resulted to it and how it could be funded. **Chapter 5**: *Technical preparations* bring out the essential aspect in open data; its automated availability in machine-readable format.

Chapter 6: *Open data infrastructure* presents the national infrastructure of open data in Finland. The infrastructure includes data catalogues and cross-administrative actors who help other actors open and utilise their data resources.

The Concept of Data

In this guide, we mainly use the term data and its derivatives, such as data source (or source of data), data producer, data catalogue etc., to refer to raw material in a digital format.

Data is digitally stored information, such as documents, contract databases, transcripts of hearings and audio-visual recordings of events. Even though non-electronic resources, such as old paper documents, are not a part of open public sector data, it might be good to transform such resources into digital format.

This guidebook offers guidance on opening public data with the idea of the data as raw material that different actors can revise, combine, filter and process. The final outcome produced by one actor may function as raw material for another. For example, the up-to-date legislation on Finlex, the data bank of Finnish legislation, is an example of processed information. Should the information exist in machine-readable format, it could function as raw material for application developers.

By using the term data, we emphasise offering information in a digital, machine-readable format. For example, in the CE Convention on Cybercrime, data refers to "(...) any representation of facts, information or concepts in a form suitable for processing in a computer system, including a program suitable to cause a computer system to perform a function".

The understanding of the term data as raw material is closely related to the concept of ecosystem, which is further examined next.

Data as Part of the Ecosystem

The government continuously produces large amounts of data. Taking into consideration the quality of Finnish data resources, that data could be put to use even more effectively. Here, the topic is approached by sketching a Finnish ecosystem of open data. In previous discussions, the citizens, private sector and other organisations have only been seen as the final users of the data, instead of parties adding value to data resources. In addition to the government, the citizens, business life, organisations and research institutes should be regarded as important actors when processing open data. The methods, tools, principles and recommendations presented in this handbook are easily applicable to all organisations looking to improve the open use of their data resources.

The goal of this book is to promote the emergence of a functioning open data ecosystem in Finland. The ecosystem would greatly increase the utilisation of national data resources through open co-operation and technology. The word ecosystem, in general, refers to a functioning whole in a given area. One might talk about the ecosystem of a specific lake or a forest, which comprises of the dynamic interaction between the organic and inorganic environmental factors in the area. When discussing open data, the ecosystem is a multi-level and multi-dimensional entity where raw material, as far as distribution and developing are concerned, is the target of cooperation. Raw material is not a means for business; instead it is made easily accessible to all actors.

The resources previously needed to acquire data could be used for utilising it.

A large scale utilisation of data creates new services, research and information, some of which have commercial value. The process also promotes democracy and education and makes the everyday lives of people without any financial benefits. The increase in utilising data has a positive effect on producing data and continuously improves the quality and the usability of data resources. In the ecosystem model, government organisations, citizens and corporations are all both users and producers of data.

Nowadays, as the Internet and knowledge work become more and more common, the significance of production outside traditional business models and the monetary economy has become greater. This is manifested in new practices, such as Open Source, Wikipedia, and social media. In this guide, we see the collection, improvement, publication, and reuse of data as an entity and as interaction between different actors, not only as a business or trade. Ecosystem evokes an image of well-being of the entity and, on the other hand, fulfilling one's own needs through the richness and vitality of the ecosystem.

Another useful term to help the reader to fully understand the field is open data infrastructure. Open data infrastructure includes all organisations and systems operating with open data, in other words, the whole operational environment. This model is suitable for analysing the field of open data at state and municipal levels. For instance, Spatial Data Infrastructure Act well represents the understanding of data as a base material for infrastructure: Spatial data infrastructure refers to provided metadata, geodata and spatial data services, web services and web technologies, distribution of data, contracts regarding availability and use, as well as co-ordination and follow-up mechanisms. The role of the government should include producing infrastructure for everyone to use and therefore to function as an enabler for wider utilisation of data.

With the ecosystem we wish to highlight not only the technological systems and institutionalised organisations, but also the living, dynamically changing network of interaction. Individual citizens and government organisations are all part of this network. The concept of infrastructure is in the background throughout the entire book and we shall take a closer look at it in chapter 6: *The Infrastructure of Open Data*. Chapter 6 handles open data projects crossing the boundaries of organisations, such as the national data catalogue.

Many functional web services have emerged from the users' needs and perceptions.

1. Extensive Use of Data as a Goal

Opening data for others to use comprises of a lot more than merely adjusting and renewing technology. When planning the operation, one should keep users, methods and learning possibilities in mind right from the start. In this chapter, we take a look at the question of open data from people's point of view. We examine questions such as how the potential user groups of open data are met, what is user-driven innovation (TEM 2010) and what are the concrete uses for open data.

We presented some arguments for opening data already in the Introduction. The goal can be, for example, maximising the use of limited resources, better functioning civil society or economic growth. However, opening data will not automatically lead to the realisation of these goals. The goals cannot be met until the open data is used.

Organisations offering their own data to be distributed have several means to support the emergence of a truly useful service. By service, we refer to all the uses applications for government data, designed to maximise the benefits extracted from the data. Therefore, opening data is not a goal, instead it is a means to an end, and producing services is what we want to accomplish by opening data. The most common wish is that the process would bring about creative and interesting uses for data, services that are usable and meet the needs of the public. The emergence of business relying on these services is also desirable.

Nat Torkington, who worked in several data opening projects in New Zealand, wrote an article called Lessons learned from the Open Data front lines (Torkington 2010). In his article, Torkington encapsulates the meaning of community-based operations in the context of the emergence of new kinds of applications. The easiest way to convince people of the benefits of open data is by showing them a concrete example, one that people find useful and that aims to combine open data in new ways to produce a service.

Case: Apps for Democracy (Washington DC.)

«Apps for democracy contest produced more gains to the government of Washington DC than any other project.»

–Vivek Kundra, former Chief Technology Officer in Washington DC, current Chief Information Officer of the United States of America.

The data catalogue maintained by Washington DC (http://data.octo.dc.gov) was established already in 2006, and is perhaps the first, extensive public data catalogue. The catalogue contains hundreds of high-quality data sets, e.g. live data feeds on public transport, school ratings and regional demographics. However, not many noticeable applications emerged for a few years since the catalogue was published. In addition, the catalogue has been used mainly by the administration itself. Apps for Democracy contest was first introduced as an incentive to further the wider use of the catalogue.

Organising the contest cost the city 50,000 US dollars, out of which 20,000 were given out as prize money. The competition resulted in 47 functioning services, including mobile, Internet, Facebook and Twitter applications. According to calculations, producing these applications through traditional channels would have cost over 2 Million US dollars. A large part of the expenses would have consisted of internal project management and procurement procedures. It was estimated that, using conventional procedures, it would have taken over two years to provide citizens with this many applications. It now only took a couple of months. Free social media tools were used in organising and promoting the contest. Through social media, the target audience was reached efficiently and quickly.

(More in chapter 1.4.2. *Innovation Contest as an Incentive for Action*)

1.1 Where to Use Open Data

Internet applications are, for many, the first thing that comes to mind when thinking about possible ways to use open data. However, there are other ways to use it. Here, we will make a rough division of those uses: **1)** *mashups*, **2)** *education, research, and product development*, **3)** *automation of processes* and **4)** *crowdsourcing*. This is not an extensive list, but it can help you get started on thinking about the nature of your own organisation and the possibilities of produced data.

Mashups

At the time of writing this guide, the Internet applications designed to ease everyday life were by far the most common use for open data. The applications visualise, filter, and collect information from different sources to meet the needs of the final user.

Mashup is a generic term to describe public open data applications created in the business life or by citizens. These applications assemble available information from different sources in a way the designers want.

In 2005, an American journalist, Adrian Holovaty, got tired of calling the police station every morning to ask for crime reports. Once he discovered that the reports were automatically available, he combined them with Google Maps application, which had just been released. As a result, ChicagoCrimes service was born. In ChicagoCrimes, one can see the crimes committed, almost in real time. Tilannehuone.fi created a similar application in Finland. Their service combined the Emergency Response Centre's status reports with Google Maps. Tilannehuone.fi is a simple but interesting application; they have nearly 300,000 weekly visitors. Mashups can meet many needs, e.g. making everyday life a little easier, increasing the transparency of politics or, as in the case of Adrian Holovay, they can be used as tools for journalists.

Education, research and product development

In research organisations, easy access to data supports high-quality research and in education data can be used to demonstrate certain facts. The GapMinder service, developed by a professor Hans Rosling in Karolinska Institute in Stockholm, is a fine example of the power of visualisation. A pre-stage version of the software was developed when Rosling needed to show the students how the 1960s idea of global polarisation based on life expectancy and family size was no longer valid.

Research and product development offer a greater scale for mining, combining and visualising information than the mashups. In these cases, the aim is either to produce new information or optimise a certain large data set, and not merely to make life easier or further the transparency of governments. For example, optimisation models of a city's transportation system can be created based on traffic measurements, public transportation's user statistics and different regional statistics. Nowadays many organisations create these types of optimisations and simulations using their own data resources. Open data would enable the use of additional sources and the data sets of other organisations.

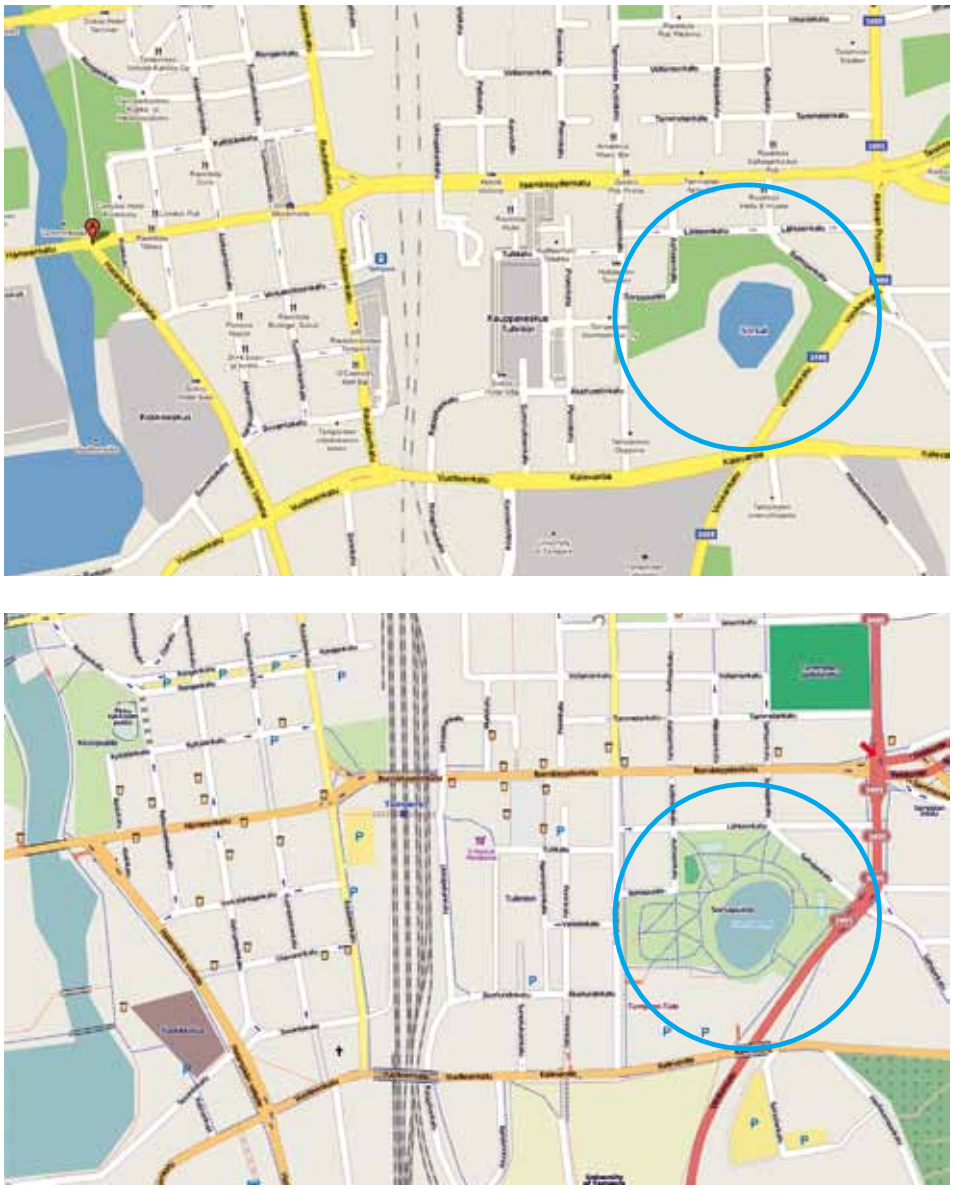
Automation

Data can be used for automation, where the data helps to guide a process, or make it easier. For example, filling in and updating address forms on web application can be made automatic using address and zip code data. In addition, heating and air conditioning systems could benefit from weather data and data on the capacity of power-distribution networks. This data could then help automate the systems in a way that could lower the consumption of electricity and the spikes in consumption would even out. This type of automation is yet to actualise, but having access to more data could speed the progress.

Crowdsourcing

Another benefit of open data is the improvement of quality and collective mustering of resources, as well as reducing the overlap of operations with a joint data resource. Jeff Howe (2006) came up with the term crowdsourcing to describe new ways of organising work, made possible by the Internet. The term simply means outsourcing the work to an anonymous crowd on the Internet. Crowdsourcing varies in form. Typically, one either searches for the best possible solution or alternatively data may be collected, classified, assorted, produced, and developed collectively.

The best known examples of crowdsourcing are Wikipedia and OpenStreetmap but, for some reason, it is often forgotten that citizens and companies could produce data for the public administration to use. An example of crowdsourcing utilised by the government in Finland is the pilot service Fillarikanava. This service is for cyclists, who can mark their cycling-related observations on a map. These observations are then saved as geodata, which is taken into consideration when planning improvement projects in the area. A slightly different application is the Times Educational Supplement (TES), created in Great Britain. TES is a platform for sharing and joint production for teaching materials. It has been estimated that TES will save 1 Billion GBP in teachers working hours in two years (UK 2009) after the teachers start sharing their materials to each other. This way not all teachers have to produce their own teaching material.



Picture 1.1: OpenStreetMap was produced collectively by volunteers. It is more accurate than Google Maps in, for instance, naming the parks and pavements in the city centre of Tampere. In Finland, the community has voluntarily drawn up the map and taken measurements with GPS devices. The OpenStreetMap community accepts geodata donations.

1.2 User-driven Innovation

The lifespan of services usually begins with identifying the needs. This is followed by ideation and testing and the interactive development, which is based on user experience and feedback. User-driven product development and innovation are models, where the developer of a service includes the final users to the ideation and development processes.

In the previous subsection, we looked at different ways to utilise open data. There is no one correct way to do it, and the same data can be used in many different ways. The best functioning ways to use open data can emerge from any of the users, such as non-governmental organisations, companies, individual programmers or innovative citizens, other organisations and so on. In all probability, the best ways to use data produced in an organisation are found somewhere else than within the producing organisation.

The application users' needs determine the functionality of an application. Many well-functioning web services were created to meet the needs of the creators and are based on their perceptions. In those cases, the users have found a service in the affordance, suitable specifically for them. There are, however, opposite examples: stylish and well-functioning services have been created but they attract little attention from the users. These failures could usually have been avoided by researching the users' needs and their know-how on using services.

Open ecosystem enables people in need of a certain service to start brainstorming, sketching and creating such a service. This is a great model for developing services designed to match actual needs. Not only the masses, but small groups and individual people as well, can create services matching their needs. On the other hand, it would benefit the provider of open data, if a large number of people took advantage of the provided data resources. For example, Amazon, the Internet bookstore, offers their massive purchase and transaction databases to programmers to use for free. The programmers can, therefore, utilise actual databases and get references for their careers, whereas Amazon gets tens of thousands of programmers to create new, innovative user interfaces for them. The application promotes the Internet coverage, business transactions and sales of Amazon.

The responsibility of ideation and the implementation stays in the organisation, unless the data is open. It is one of the tasks of the actors in the public administration to produce an overall picture of the most common needs, based on their data. Special needs, and those of smaller groups, are inevitably ignored in the process. This happened, for example, in the case of ChicagoCrimes. There was a clear need for this service but, in all likelihood, it would not have come up at the meetings held at the police station. Even if someone had come up with the idea, they probably would have considered crime reporters to be such a minor group, that the resources of the police would not have been directed to producing this service.

Taking user-driven product development to the extreme, is when the developers of the service are the final users of it. For instance, YouTube is said to have been born out of the users need to share video files (Cloud 2006). Furthermore, the first Wiki application emerged in 1995 when Ward Cunningham created the WikiWikiWeb platform for programmers to easily share information related to the source code with each other. Open public sector data enables people to further widen and adjust the public services to make them more suitable for them. For example, Mapumental service visualises both the time spent on commuting with public transportation and the prizes of apartments in the area. The Mapumental service was created to meet the needs of people who were looking for apartments. In Finland, software company Gemilo saw a market niche in companies wanting to easily track the public competitive biddings in the area.

1.3 Agile Methods of Opening Data

Subsequent to digitalisation and globalisation, the operational environments of organisations change rapidly. In fact, the changes are happening so fast, no organisation has the resources to handle it. To better manage the changes, a new approach to orientation is needed – agility. With operational environments constantly changing, there is little sense in making detailed plans far into the future. The further into the future the plans go, the thinner the assumptions they are based on, are.

Agile methods as a term have, for some time, been used in connection with software developing (agile software development). The logic of agile methods can be applied to other processes as well. Agility does not mean speeding up or stimulating everything, it refers more to constant steering. Essentially, decisions are made in light of constantly updating information, reacting to changing circumstances and taking advantage of things learned in the process. Getting everything done at once is not the goal; the progress can be made little by little through new iteration rounds.

In context with iteration and agility, tailored and functioning solutions as well as active communication between the developers are emphasised. One can proactively participate in on-going activities and thereby shape the operational environment. Not being able to plan far ahead does not complicate this type of operational culture. Iteration also means faster fixing of problems. Constant evaluation of operations makes fixing possible problems easier. Ultimately, iteration can be seen as interaction. An agile actor operates interactively with its surroundings, its ecosystem.

It is still too soon to estimate the global or national effects of opening government data. Among others, this guidebook will soon be out-dated. This is why agile methods offer great basis for opening government data. Agility is referred to in following chapters, introducing the data inventory and the technical solutions for publication. The main idea is to think small. The first services developed from opening data will pave the way for things to come.

1.4 Organisation as an Enabler

When practices change, so do the roles of organisations. This has become apparent already in e.g. the content industry due to the development of Internet. The music industry is not responsible for the revolution in record sales, Apple iTunes is. Commercial computer software lost part of their market share to Google and its free services. On a positive note, the change enables governments to act in new ways and take on new roles, some of which would have been far too expensive before the web era. These new roles include:

- Information provider
- Enabler
- Platform provider
- Facilitator / co-ordinator
- Online consultancy provider (Hintikka 2009)

In the Introduction, we presented the concept of open data as an ecosystem, where several actors have different roles and the data is seen as common, renewable raw material. In the open data ecosystem, the government acts as both the enabler, and the provider of information. Once data is offered to be used openly and free of charge, the development and specialisation of roles becomes possible.

One might think of open data as building the infrastructure of an information society. It enables the development of highly developed services and social innovations.

In 2008, the board of the Finnish National Broadcasting Company YLE added the enabler strategy to their general company strategy. The essential message of the enabler strategy is that an organisation no longer executes solely its own operations, but also offers others the possibility to activity.

The term government as a platform is used by Tim O'Reilly to describe the changing paradigm of government in the world of Internet technology and open data. The idea is that governments should focus on building infrastructure and thereby enable the development of sustainable private actor ecosystem. This is not a new way to operate within a government. For example, creating road networks has been seen as something naturally organised by the government (instead of the government organising the shipping of goods and/or people). Creating road networks has enabled a variety of private activities related to roads and transport. In the context of open data, this means that information might be better mediated by reducing the government's role in delivering the information. Currently, many government organisations see providing information through their own websites as more important, than the technical infrastructure of open data. This infrastructure would allow others to use the data.

Following Tim O'Reilly's trail of thought, we talk about the government as an enabler, which is slightly more than a mere platform. Being an enabler is active, whereas platform creates an image of something passive. Regarding open data, it is our understanding that it is well-grounded for the government to offer data in a reusable format and function as a platform but, in addition, actively urge companies, citizens and other instances to utilise the data

1.4.1 Open Communication with the Reusers

The open data projects are not random technical plans. It might be profitable to build these projects as part of the ecosystem and around the people and communities using and developing the data. The principles of user-driven innovation, presented earlier, can be applied to the process of opening information and creating a data interface. The data reusers are experts on how to offer data in a way that would promote active utilisation and the emergence of services. Organisations opening data can showcase interfaces, receive feed-back and get people to help with the development.

When creating public data, it is essential to make the data as accessible as possible and to reach the users interested in that particular data. Accessibility can be technically improved by linked data, optimising the results of search engines and adding the data to the largest data catalogues known to programmers. Traditional promotion is, however, equally important in making the wider audience aware of the existence of data. The first application to get coverage will promote the original data source. This would then help other applications to emerge faster. The sooner the users of data resources are included in the development process, the faster the opened data resources are utilised. The real challenge is to calculate the workload of participation and to efficiently organise the cooperation to make sure it gives more than it takes from the participants.

The programmers using and developing data are essential when talking about communication. The programmers may work in companies, research facilities or use their free time to program. Naturally, programming skills are not a requirement for being a part of an open data community. Everyone's opinion is needed,

regardless of programming skills, whether it is about the use of open data or how to fix concrete problems. However, the programmers have an important role; transforming data into information requires creating user interfaces, programmatic manipulation, visualisation, combining, or other illustration.

The preparations for opening data can be made, in part, publicly on the Internet. Dismounting to social media services is a natural way of getting in touch with the programming communities and other interested parties. Among others, the Finnish Innovation Fund Sitra has coordinated different community panels on the Internet. Expert-centred events that used to be enclosed have now been put on the Internet for anyone to see. Social media has been discovered by other organisations, as well. According to several studies, the government in Finland has been more active than the business life in using the Internet-provided possibilities. Internet presence allows new contacts that previously did not exist. For example, the Finnish police have operated in Facebook and other social media services, popular among the youth.

1.4.2 Innovation Contest as an Incentive for Action

In order for the services and applications to emerge, technical resources are needed. In addition, the communities, actors, programmers, researchers, companies, and organisations need to use raw data creatively. Organisations can support the utilisation of open data in many cost efficient ways, and thereby gather experience in cooperation with communities. For example, the authors of this guidebook took part in a workshop organised by the Helsinki Regional Transport Authority (HSL) for both the creators and users of their journey planner API (Application Programming Interface). Here, we will present, in greater detail, the innovation contests designed to further the utilisation of data. These innovation contests have, at least in early stages of forming an ecosystem, been proven efficient.

In the beginning of this chapter, we mentioned the Washington D.C. data catalogue and the *Apps for Democracy innovation contest*, designed to further the use of the catalogue. Both Vivek Kundra's statements and the calculations on the contest's benefits, provided by the Department of Data administration in Washington DC, are inspiring. Still, it cannot be claimed that the contest itself provided the city with 2 Million dollars in savings. Under normal circumstances, the city would not have developed most of the services that came about in context with the contest. In addition, the created services cannot be compared with the city's acquisitions as such, even though the services are useful to the citizens. However, the city did enable the emergence of these web services, gained positive publicity and experience in cooperating with communities. All this was achieved with reasonable expenses.

Many of the applications introduced in the contest still exist today, some have been further developed and businesses have been established around some of the applications. The most influential benefit of the contest, however, is the fact that it functioned as the foundation for the open data ecosystem in Washington D.C. Nowadays, the users are as aware of the Washington data catalogue as they are of GoogleMaps. As mentioned, the contest was a trigger for the open data ecosystem in Washington, which consists of data and active users.

The goal of the innovation contest was to provide citizens and companies with data and award the best applications with money and publicity. Similar innovation contests related to open data have been organised all over the world, e.g. the Pan-European Open Data Challenge in 2011 and the contests in Finland in 2009 and 2010.

The goal is to gain experience in open data distribution and ecosystem participation faster than would be possible though any renewal cycle of an IT-system.

2. Organisation's Views on Openness

The decisions on the free distribution of data resources have to do with financial matters, exercise of power, law and technology. Often, a wide network of actors participate which makes it time-consuming to reach joint decisions. There are many international examples of the strategic decisions on open data, brought in by nations, cities and organisations.

At the time of writing this guidebook, the attitudes towards opening government data in Finland had rapidly become positive. However, not many organisations have mentioned open data in their strategies or goals. The topic has become better known within the public administration through EU legislation (Inspire and PSI).

Open distribution of data supports the strategic goals of many organisations in the public administration. In most cases, offering data freely has to do with communication and data administration strategies. For example, Statistics Finland focuses on producing data. Therefore, offering open data is directly linked to its strategy. Tekes, the Finnish Funding Agency for Technology and Innovation, however, has their goal set on supporting Finnish business, which makes their relationship with data more indirect.

Organisations produce information resources. The purpose of openness is to find more ways to use those reserves, ways that could benefit not only the users but the organisations themselves. However, open data ecosystem does not function on strict reciprocity; the benefits come partly indirectly through the entity of the system. Building a name for an organisation and a rise in demand for its products are examples of indirect benefits.

Organisations within public administration are the initiators for opening data. However, in most organisations, the resources for opening data are still not sufficient. In this chapter, we shall take a look at organisations' roles in open data ecosystem. One should start up easy but with a clear goal set on making visible changes and creating new, real ways to use data. Once the data resources have been identified, it is not hard to choose the easiest ones to be opened first. These early efforts will help gather experience in opening data. These experiences help create a strategy to support the openness of data resources and the strategy will set guidelines on how to systematically open the records and interfaces.

The data usability indicators presented in this chapter (2.2.) and the suggested process for inventory and publication (2.3.) form the basis for the next chapters. The upcoming chapters go deeper in such themes as non-chargeability, legal and technical reusability and the accessibility of data. The data opening process should include a variety of know-how. In addition to competent personnel, experts on, at least, information, law, business, information technology, communication and management should be present.

2.1 Data Resources, Actors and Roles

Government organisations, companies, non-governmental organisations and citizens are all potential actors in the open data ecosystem. Public administration plays a significant role in producing data and, therefore, it is highly recommended that public administration takes part in the opening process. The best-known government data registers in Finland are population register and land register. Other well-known government data includes statistics, weather information, maps and spatial data. In Finland, the government data resources have not been charted to the full extent, but according to several estimates, the resources are remarkably large.

Case: Data is everywhere – the tree register

Public data is everywhere, if you know where to look. A practical example from New York Trees Near You service, created by Brett Camper, won the Best Application Honorable Mention in the New York BigApps competition in 2010. Trees Near You is a free iPhone application, giving information on more than 500,000 living trees in New York. This application combines the GPS coordinates from the phone, Wikipedia articles on wood species and street tree census data, offered publicly by the city. This application is a great example of all the available data one might never come to think of.

Government data can be structured based on producer organisations, content of the data sets or the assumed use. So far, spatial data (MMM 2005) and data produced in organisations within the Central Government (Kuronen 1998) have been the focus of analysis. In this guidebook, we apply the analysis on all government produced information that is legally public.

It is essential for the organisations to identify their data resources and give out information regarding those resources. In Finland, there is an on-going project for developing the government-level information architecture (VM 2009). In the context with this project, an introductory chart was created to demonstrate the overall field of Finnish data resources.

In this guidebook, we look at the government as an important data producer. However, from the ecosystems perspective, it does not matter who produces or uses the information. Currently, it is possible that private companies produce nearly as much data as the government organisations. Individual citizens are increasingly taking part in producing data and developing it into information and knowledge. Within the open data ecosystem, government organisations, citizens and companies are all not only the producers but also the users of information – in many cases, at the same time. The ecosystem sees organisations and private people as actors, interacting with each other to mould the conventions of ecosystem.

Many government processes, such as preparation of legislative proposals, involve communication and exchanging information between different organisations. Law-drafting in Finland, for example, may include using information from Statistics Finland or assessing the budget effects in cooperation with the ministry of Finance. It is a safe assumption, that all government organisations a) produce new data, b) process, c) handle and d) utilise data produced by some other party. Even though the field is wide and the producing organisations cannot be clearly identified, the actors' roles in relation to data can be pointed out quite accurately (table 2.1.).

Table 2.1. Examples of the actors' roles in relation to data.

Storer	collecting and saving raw material
Developer	managing and processing raw material
Aggregator	combining and editing data from different sources
Harmoniser	standardising and homogenising data from different sources (same terminology means the same thing)
Updater	updating information
Publisher	publishing data
Register keeper	administration of data resources
Application developer as the final user of data	utilising the data as part of the service
Interpreter as the final user of data	interpreting the data, e.g. researchers, companies or democracy activists
User of data-based services	an individual, company, or organisation using open data applications and interpretations

An organisation's data sets will be identified later in this chapter, when we discuss the inventory of an organisation's data resources. The above-mentioned separation of roles is meant as a tool to help with the inventory.

2.2 Indicators of Data Usability

Openness has quickly become a central term in the government terminology when discussing information systems and data resources. Openness may refer to open licences, technical interfaces, formats, metadata, harmonising data, the transparency of the government, peoples' possibilities to participate, the principle of openness, reusability and machine-readability. It is not uncommon for different actors to mean different things. Misunderstandings and excess generalisation are common, due to the fast development and wide scope of the field. What exactly is open public data?

The openness and the publicity of resources are not to be confused with each other. Open data refers to any data sets that are reusable by anyone without technical, legal, or financial restrictions. According to legislation, public data refers to information not included in the privacy policy – or classified for any other reason, such as national security. Naturally, any data set that is not public cannot be open, either. It might work the other way around but we still have a long way to go.

Technically easy utilisation of data does not automatically mean that the data resource is open. Especially when talking about "open interfaces", one should keep in mind that the reuse or redistribution of easily available data may be limited by Terms of use. Correspondingly, there might be cases where fully open data is not easy to utilise due to challenging file format, poor findability, or insufficient documentation.

The usability of data can be estimated using the following criteria: *Accessibility* **(2.2.1)**, *Completeness* **(2.2.2)**, *Equality of Terms of Use* **(2.3.3)**, *Timely and Original* **(2.2.4)**, *Legal and Free Reusability* **(2.2.5)**, *Non-Chargeability* **(2.2.6)**, *Machine-Readability* **(2.2.7)**, *Openness of the Format* **(2.2.8)** and *Good Documentation* **(2.2.9)**. Aiming for complete usability and utilisation according to all criteria is not cost-effective with all data sets. Often, the reusability of reources can be significantly improved by decisions that further some of the abovementioned criteria (e.g. using more permissive licences or offering the data free of charge).

2.2.1 Accessibility

Easy to use: The existence and location of the data are well known. The information and the licence terms allowing reuse are easily found on the Internet by both people and search engines.

Difficult to use: A data set only exists in an operative system of an organisation and no one outside the organisation has knowledge of it.

The Google Maps interface and the contents of Wikipedia are fine examples of data, the existence and usability of which are widely known. The availability of data resources can be improved by adding it to a well-kept data catalogue, optimising the metadata of the data sets for search engines, and publishing the data according to the linked data paradigm. Letting the potential re-users to know about the catalogue on the Internet, in publications, and in various events can further the general visibility of the data resource (see chapter 6: *The Infrastructure of Open Data*)

2.2.2 Completeness

Easy to use: Data, in its entirety, is free for downloading on the Internet. The accessibility and the potential use are not indirectly restricted by allowing access to only a certain part of the data set at a time.

Difficult to use: Only part of the entire data set is freely available and a separate contract is required to access the complete data set.

Typically, access to the complete information resource is restricted, intentionally or unintentionally, by only offering the data through a query interface and making it impossible to download the entire data set. If the data resource is available in its entirety, it is technically possible for anyone to start redistributing the data to themselves and others. Restricting the entirety may be a way to prevent copies. On the other hand, the restrictions prevent any use based on an extensive analysis and burden the query interface, which could be avoided by offering a copy of the data set (see chapter 5: *Technical Preparations*).

2.2.3 Equality of Terms of Use

Easy to use: The information is equally available to everyone to be used for all legal purposes. The users and purposes of use are all equal, as are other government operators, citizens, companies, and foreign actors.

Difficult to use: Access to the data resource in restrcited based on the user or the purpose of use. For example, data can be offered solely for research and product development, or for uncommercial purposes.

In practice, equality is achieved when everyone has access to the data set and no registration is required. In which case, anyone with standard licensing terms can use the data set. A licence does not prevent the use of the information in a certain range of use. In particular, commercial use is allowed because there are high hopes that the commercial actors will participate in the ecosystem. Equality means letting go of the anticipatory control. One is allowed to use data unskilfully and for political purposes (see chapter 3: *Permission to Publish and Reuse*)

2.2.4 Timely and Original

Easy to use: If the information does not violate an individual's right to privacy, it should be published in its original form. In addition, the data should be published as close to its original source as possible (preferably by the creator) and as quickly as is necessary to preserve the value of the content.

Difficult to use: The publication of the data set is not timely, not as accurate as the original, or the data is only published together with other data resources.

In addition to raw data, joint and processed forms of the data set can be placed available. In some cases, it is possible to publish data sets potentially in violation of privacy by making generalisations and lowering the level of accuracy. However, one has to be extremely careful with generalisations and turning the data anonymous (see chapter 5: *Technical Preparations*).

2.2.5 Legal and Free Reusability

Easy to use: There are standard terms of use, according to which the data set is free for reusing, regardless of the re-user or the purpose. Excluding the limitations regarding privacy or security, there are no legal limitations on the reuse of the data. The licence terms are clear and transparent, and support reuse. All copyrights have been waived and the waiver is clearly expressed in the Terms of use.

Difficult to use: The data set is protected by a licence limiting the reuse or a copyright, or the permission to reuse is not expressed clearly or at all.

Permissive licences include, for example, Creative Commons and Open Database licences. The copyrights of open public data should be waived using, for example, the Creative Commons Zero licence in order to avoid any obscurities later on the processing chain. Regarding the terms of use, the most common wish expressed during the conducted interviews was to find out who is using the data and for what purpose. Often, there were no reasons for restrictions but the creators wanted to know so that they could further develop their operations. The data can be monitored without signing separate contracts or restrictive terms of use by, for example, user registration on the Internet (see chapter 3: *Permission to publish and reuse*).

2.2.6 Non-Chargeability

Easy to use: The data is free of charge

Difficult to use: The data is offered for a charge and the profits from the sales are used for covering other expenses in the producing organisation.

Any data that is offered for a fee no larger than the marginal costs can be considered as open. Often, the costs of maintenance and producing a data set are many times higher than the charged marginal costs. Even a small fee may limit the use of the information because of the contracts. In fact, it is possible that most of the marginal costs consist of the bureaucracy related to billing. If that is the case, then there are no grounds for charging the marginal costs. If, for any particular reason, charges are collected, it should be possible to make the payment on the Internet and receive the data set immediately without burdening the authorities (see chapter 4: *Economic Viewpoints*)

2.2.7 Machine-Readability

Easy to use: The data resources have a permanent location on the Internet, allowing automated and programmatic access. The data is structured to enable automatic processing. The terms of use are machine-readable, they can be accepted on the Internet, and the data is received immediately without burdening the authorities.

Difficult to use: The data is published in a non-structured format, making it readable only for people (e.g. PDF documents and html pages).

As a ground rule for machine-readability, a capable programmer can, in a relatively short period of time, create a programme that automatically retrieves the data from the Internet, processes it in the machine's memory, and presents it in a new format, for example, on a screen of an iPhone. If a data set is not machine-readable at any stage, it is rather difficult to change it to such a format. Despite of that, many organisations offer data in a non-machine-readable format on their webpages. However, the same information can often be found elsewhere in a machine-readable format, which makes it a lot easier to publish the data (see chapter 5: *Technical Preparations*).

2.2.8 Openness of the Format

Easy to use: The data is available in a workable format and is easy to process. The definition and development of the format is not in possession of a single organisation.

Difficult to use: The data is only available in a format administered by a single company, and the utilisation of the data set requires software provided by that same company.

This can be achieved by offering the data in an open format. The definition of the format is freely and publicly available and the use is not restricted financially or otherwise. If possible, it might be worthwhile to offer the same data in several formats. Using open formats is not always realistic. For example, some spatial da-

ta systems use a producer-specific format, which means that switching to open formats would only be possible after a system renewal (see chapter 5: Technical Preparations).

2.2.9 Good Documentation

Easy to use: The data is clearly and comprehensively documented.

Difficult to use: The offered format's contents or use has not been described in a way that enables reuse.

The reusability of data can be significantly improved with metadata, documentation, user examples, and quality definitions. The only downside of good documentation is the work it takes. Regulation may notably slow down the publication of data sets. On the other hand, documentation can initially be done lightly and improved later. For example, including column headings in a Tab separated file is sufficient (see chapter 5: *Technical Preparations*).

2.3 Inventory and Publishing Processes of Data Resources

The biggest advantages of open data are realised through well-described data, distributed through reliable interfaces. However, opening data takes both time and resources, and everything cannot be built up at once. One should look for small actions that significantly increase the utility of data.

Here, we will present a model for opening data, one that guides gradually and through practical examples. The goal is to gain experience on open data distribution and ecosystem participation faster than would be possible though any renewal cycle of an IT-system. The model reveals which data has the most demand. In addition, it lets us know which data should be distributed through which interface. This is a speculative process, primarily meant to provoke discussion.

When opening data, one should avoid all presumptions on the use or potential users of the data. In some cases, there might be a hidden demand for certain information, only revealed once the data set has gotten publicity. For example, discussion on the spatial data legislation has made spatial data sets better known, and therefore increased the demand. In order to be able to prioritise, it is only logical to find out which data resources have the highest demand. Instead of images, it is better to lean on experiments, encourage data utilisation and see what happens. Even a single application that gains publicity, may lead to increased demand for certain data and, therefore, change the situation in a very short period of time.

Technically the publication of data resources is related to the development of an organisation's information system and the development of information architecture in particular. Publication of data can be seen as a support measure for developing information architecture, and not so much as a result of an architecture project. Instead of promising open interfaces in connection with the next system renewal, one should publish the results of the organisation's data inventory. After that, one could publish some interesting data sets as they are. The experiences gained through publishing data could function as an important feed to both harmonising the data set and developing general information systems.

Every phase of the process (picture 2.1) increases the organisation's knowl-

edge of their data resources and new data is published for the ecosystem to use. In addition, the organisation gathers more and more information on opening data during each phase. The phases of the process are **a)** *analysing the data set*, **b)** *publishing the information* and **c)** *learning from opening the data*.

Phase 1: Reporting the data

The announcement regarding information held by an organisation and the opening of data are done as quickly and lightly as possible. In this phase, the following questions need answering

1. What data sets are we in possession of?
2. Which of them are public?
3. How open are different data sets?

Phase 2: Publishing pilot material

Once your own data resources have been identified, it is not difficult to pick the ones that are the easiest to open. The following iteration round includes pilot materials that are technically, legally and economically the easiest to open. In this phase, the following questions need answering

4. Which data sets are easy to open?
5. Which terms of use should be applied?
6. Who, within the organisation, is in charge of the contents of data sets and technical systems?

Phase 3: Documentation and use cases

The amount of information being published can be increased phase by phase. However, one should also improve and clarify documentation for the re-users and collect information on use cases within the organisation. In this phase, the following questions need answering

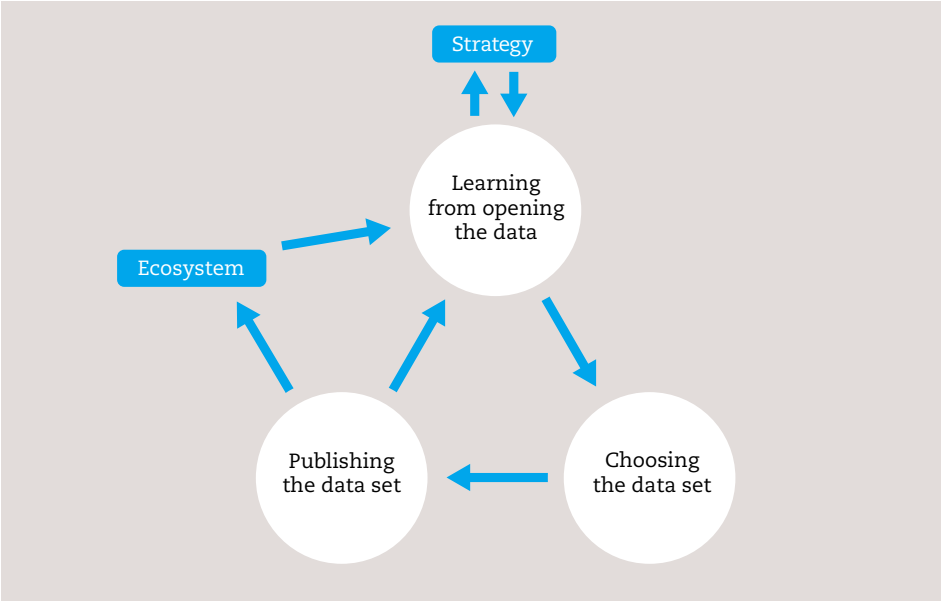
7. What is the contextual description of the data and how was it documented technically?
8. What kind of needs for information, user groups, and use cases are related to current use of different data sets?
9. Have any requests been made in or outside the organisation regarding the availability or usability of data?

Phase 4: Information architecture and terminology

The first pilots give the organisation enough experience to create a strategy for furthering the openness of data resources. This strategy would function as a guideline on how to systematically open registers and interfaces. After that, it is possible to create cross-organisational use of data resources. In this phase, the following questions need answering

10. What are the needs and possibilities for cross-organisational standardisation of data?
11. What type of system renewal is the improvement of the openness of data resources related to?
12. How could an organisation's data resources be better organised together with other actors?

As was mentioned previously, this is an indicative process and the execution of it may vary from organisation to organisation. However, the starting point is quite clear. Opening data always starts with identifying one's own data resources and evaluating the current state they are in. After the evaluation, it is possible to participate in the ecosystem of open data; first experimentally and later by including open data as a permanent part of working methods. One can start small, and experiences gathered along the way form a great basis for an organisation's own open data strategy.



Picture 2.1 Opening data is a gradual process, one that progresses in interaction with its users.

A) **Analysing the data**
- Phase 1: A list of all data in possession of the organisation
- Phase 2: Analysis of the data (technology, legislation, responsibilities)
- Phase 3: Interviews and use cases regarding the use of information resources
- Phase 4: Development plan for information architecture

B) **Publishing the data**
- Phase 1: Publishing the Information Asset Registry of the organisation
- Phase 2: Publishing the pilot material in its raw form
- Phase 3: Defining the interface and publishing documentation
- Phase 4: Creating a search service or a data portal

C) **Learning from opening data**
- Phase 1: Following the statistics on data use and downloads
- Phase 2: First applications, the word spreads
- Phase 3: Meeting with the users, for example in a user workshop
- Phase 4: Active use and development of the organisation's data resources in the ecosystem

2.3.1 Reporting the Data

A representative for an organisation says *«Tell us, what kind of information you want, and in which format and we will see what we can do about it»*. The re-users reply *«Tell us, what kind of information you have, and we will tell you what we want»*.

The very first data to be published should be a list of the data sets the organisation is in possession of, even if the data is not yet available. Many organisations may already have a list of their data resources, some have to start with an inventory.

The simplest way of putting together a list is to rely on one's memory and scan through the organisations webpages. A comprehensive list is achieved when more people take part in the mapping of data sets. For example, a productive workshop can, in just a few hours, put together a publishable list, containing essential information on the organisation's data. At this point, the aim is not to define all information on all data sets. Here are a few questions one should try to answer

- Name of the data set
- Short description of the content
- How open is the data, is it available for users and if so, how?

Once the list of public data resources has been posted on the net, it is time to send out messages to potential re-users, create contacts and gather opinions. An organisation can achieve this by, for example, contacting the developer community through social media or big events. Communication with the re-users should encourage further use of the data. If the decision on opening data has not been made and the organisation is only looking for ideas, they should let the re-users know that.

It is possible, that the users might ignore the organisation at first. In best case scenario, the mere publishing of a list may provoke discussion between the organisation and the users of the data. Finally, the organisation should gather and evaluate the feedback and experiences regarding opening the data. What was good? What was not? This is preparing for the next iteration phase.

2.3.2 Publishing the Pilot Material

Once the organisation's essential data resources have been identified, the next step is to separate the publishable information from the rest. Databases may sometimes contain both public and classified information. For example, the names of government employees are public information but, for the sake of national security, no information is given out about the fighter pilots. For the sake of fast opening of data, the publication of these types of unclear databases should be postponed until later. The data should be divided into public, non-public and unclear. At first, an organisation could concentrate on opening the public data.

An organisation should not clear the list of data sets they are not allowed to publish. If possible, information on the existence of those sets should be given. Actors operating outside the organisation find this type of information relevant in forming an overall picture of government data.

Next, a decision has to be made; who is going to operate as a mediator between the organisation and the re-users and how to address the issues that might come up. An organisation should prepare to deal with conflict situations and have answers to what to do with possible immoral use of their data.

At first, no interface for distributing data is needed. It would be easier for everyone, if the data users are not asked to register or order the data from the organisation. Instead, the organisation could let everyone know where to find the data sets. Distributing raw data "as it is" means that if the information is in a database format, the content of the database is printed on a Comma Separated file, which is then posted on a web server. If the data already exists in the files, it is enough to enable public availability of the files.

It is not the aim of initial raw data distribution to have someone immediately create systems based on it. Openly published data helps different quarters to explore the information content and make plans on how to utilise it. This is what creates publicity for the data in question. Once the data has provoked interest, the providing organisation can, in cooperation with the re-users, plan the best serving interfaces and other distribution methods.

The lack of resources within an organisation to plan the interfaces is not a reason for delaying the opening of data. Iteration is the key and small steps take you further. It is better to publish the data lightly but as a whole at first, and only then start thinking about creating possible interfaces. Final polishing and user interfaces are low on the list of priorities.

2.3.3 Documentation and the Use Cases

Viitteellisen datan avaamisprossessin mukaan on tehty jo kaksi täyttä kierrosta. Ensimmäisen tuloksena julkaistiin lista kaikista organisaation hallussa olevista julkisista aineistoista. Toisen tuloksena julkaistiin ensimmäiset helposti avattavat aineistot. Samalla organisaation omien tietoresurssien tuntemus on vahvistunut ja datan käyttäjäyhteisöihin on syntynyt ensimmäiset kontaktit. Tämän seuraavan *dokumentointi ja käyttötilanteet*-kierroksen aikana tehdään kattavampi datavarantojen inventaario, julkaistaan uusia aineistoja ja parannetaan niiden julkista dokumentaatiota.

Now, we have completed two full circles of the possible data opening process. As a result of the first one, a list of all data sets in possession of the organisation was published, and the second one led to the publishing of the first, easily opened data. During the process, the organisation has become more aware of their information resources and made contacts with user communities. During this next round, documentation and operating situations, we will make a more comprehensive inventory on the data resources, publish new data and improve their public documentation.

The inventory process of opening data is similar to one that might be carried out anyway, to analyse and reorganise data resources to be used within the organisation. Inventory helps to identify any needs for information the employees might have. This takes more effort than simply opening the data but it supports the operations of the organisation.

Analysing the mechanisms at hand, together with the data administration, gives a good overall picture of the organisation's data resources and the conventions related to them. However, it is possible that data administration is not familiar with all of the organisation's data sets and the conventions. The familiarity can be improved by interviewing different user groups and explaining the different situations within the organisation that require information. Listening to the users may help find new ways to utilise data which can then be intertwined as part of the data opening project.

During the publishing phase of this iteration circle, it is possible to develop the distribution of previously opened data by creating an open interface (see chapter 5.5.). Describing the access rights and the content, as well as technical documentation are also important targets for development.

2.3.4 Terminology and the Development of Information Architecture

A process that started out listing data resources may gradually culminate to information architecture development, system design and cross-organisational standardisation of data resources. Prior experience on utilising open data is of quintessence.

As during previous iteration rounds, again the documentation of resources is updated, the publication of data sets is improved and the lessons learned from opening data are included in the strategic work. The aim is to model the data processing and the needs in an organisation and to improve the joint use of data resources across organisational boundaries. By the publication phase of this iteration round, an organisation is in possession of high-quality data sets, which are harmonised to be jointly used with other organisations and are well-documented upon publishing.

In addition to web applications and databases, focus should be on frequently used documents, such as spread sheet files. The goal is not only to use data but to

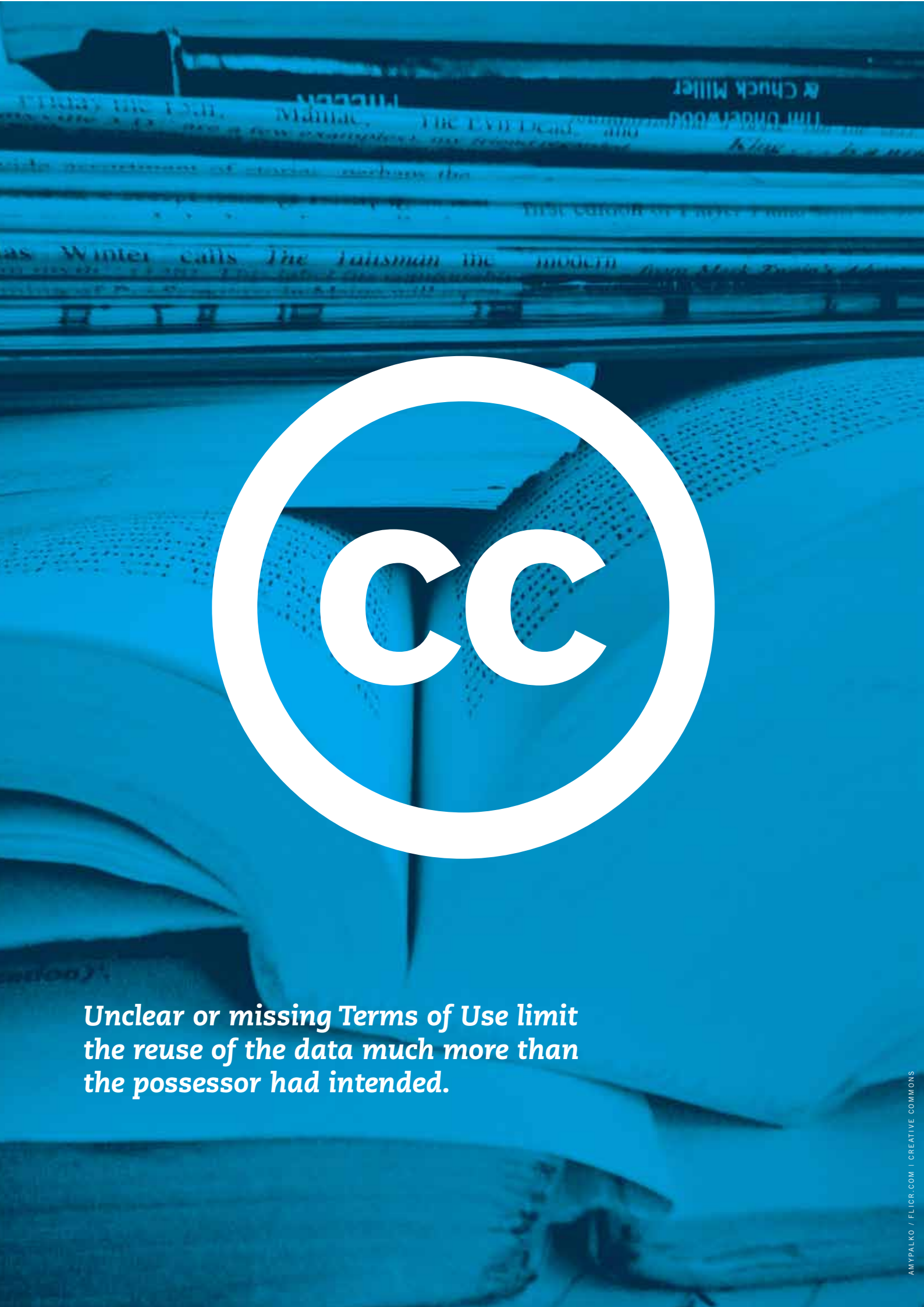
find targets for development. Organisations can ask the users what kind of needs they have, what, if any, problems they have encountered and what kind of needs they think other users may have. The goal is to find out if the information sources are well available and if the format of the data is acceptable. The format needs to be changed if there are any problems or technical issues to the utilisation.

Sometimes documents and files that seem to have no creators, or are filed inappropriately, can be found in the information architecture. These types of documents are usually a sign of a) a non-functioning policy that has been overlooked (e.g. a difficult practice), b) a procedure that has not been instructed well enough (lack of instructions) or c) an occurrence of a new need that was not anticipated (a new unit or a project team).

Many challenges regarding information architecture revolve around the same issue: information is produced inconsistently and in great quantities. The original tools were, in all likelihood, not designed to control current masses of information. The standardisation of data is a key issue, both within the organisations and in co-operation with other organisations. In worst case scenario, there is no knowledge of what information is saved on which system. Every actor looks at things from a perspective relevant to themselves, data is scattered in different systems, and the information is gathered and updated in several places.

The depiction of data sets and definition of methods are aimed to solve and prevent problems with compatibility, as well as create an overall view on data sets in different organisations. The goal is to standardise information architecture and find common ways of depicting information both within and between organisations.

Compiling of information architecture requires time, hard work and co-operation between the parties. The wider the intended view, the harder the work. An example of an information architecture project in Finland is the KuntaGML project. The project attempts to create standardised interfaces for spatial data, such as basic zoning information. Despite the hard work, a good architecture is rewarding; systems are easier to update, working hours are saved, the data resources are of better quality and, first and foremost, data can be reused even more widely.



Unclear or missing Terms of Use limit the reuse of the data much more than the possessor had intended.

AMYPALKO / FLICR.COM | CREATIVE COMMONS

3. Permission to Republish and Reuse

In this chapter, we will take a look at Finnish legislation and regulations, some of which have a bearing on the free distribution of open data in the web. We will examine the chargeability of data, access rights of data, competition legislation and EU regulations. The examined laws and regulations will be listed in table 3.1.

When planning to open data, one should find out which of the data, held by an organisation are public. This should be done, primarily, in accordance with data protection legislation and the Act on the Openness of Government Activities. If applying these laws and regulations produces a contradictory result, the decisions should be primarily based on the data protection legislation.

In many cases the privacy issues have been dealt with in another context, e.g. when presenting raw data in low resolution, making individuals unrecognisable. The Act on Criteria for Charges Payable to the State and, in some cases, the competition legislation predetermines the question of free and chargeable open data.

The goal is to increase the use of data sets. Therefore, it would make sense to give companies, citizens and organisations legal rights to reuse public data. In addition to publicity and chargeability, copyrights and licence policies should be taken into consideration. Public sector organisations can encourage the reuse of data by using permissive licences. Unclear or missing terms of use and heavy contract bureaucracy often limit the reuse of data much more than the holder of the data intended. Clear terms of use, adherent to standards, lessen the burden on both the producers and the users of the data.

Companies utilising open data in their businesses require contracts, or other guarantees, to ensure the continuity and level of service of a particular information source. Signing separate contracts and guaranteeing high service level through payments is not contradictory to the principal of free distribution of raw data. However, it is not the intention to artificially maintain two systems, one of which would be reliable and the other of low-quality. For example, GoogleMaps is a free application, but as volumes rise, one signs a contract of use with Google.

Table 3.1. Legislation regarding open government data in Finland

Freedom of information legislation	Freedom of Information legislation refers to the Act on the Openness of Government Activities (621/1999) and Decree on the Openness of Government Activities and on Good Practice in Information Management (1030/1999)
Data protection legislation	Data protection legislation refers primarily to Personal data Act (523/1999) but also to Act on the Protection of Privacy in Working Life (759/2004) and section 24 of the Criminal Code of Finland. Recent topics for discussion in Finland have included the Act on the Protection of Privacy in Electronic Communications (516/2004) and especially its reform from 2008, known in Finland as Lex Nokia.
Criteria for payable charges	Act on Criteria for Charges Payable to the State (150/1992)
International recommendations	OECD Recommendation of the Council concerning Access to Research Data from Public Funding (OECD 2006). OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information (OECD 2008)
Copyrights	Copyright Act (404/1961)
Competition legislation	The basis for competition legislation is the primary nature of financial competition compared to market regulation. The currently valid law is the Act on Competition Restrictions (480/1992)
EU Directives	The INSPIRE Directive aims at enhancing the use of spatial data, furthering the cooperation between authorities and creating diverse services for citizens. PSI (Public Sector Information) Directive (98/2003/EC) on the re-use of public sector information aims to increase the commercial utilisation of data.

3.1 Legislation linked to the publicity of data

The openness of government held data resources is determined by data protection legislation and the so called freedom of Information legislation, which governs the openness of government documents. Naturally, not all government held information is public, nor is it supposed to be. Respecting the privacy of individuals is central in maintaining the citizen's trust on authorities and reduces the risk of malpractice. Any data classified as delicate, for security reasons, for example, is also not public.

According to the Freedom of Information Act, everyone is entitled to receive information on any public government-held document or other record. All gov-

ernment-held documents are public, unless the publicity of them has specifically been limited by legislation. Giving out a public document is free of charge in many cases, e.g. when the document is electronically stored and sent to the receiver via e-mail, or when handing out the document is part of the obligation of the authority to consult and duty to provide advice and information of the authority. If a fee is collected, the costs of finding the document and omitting the classified information are not to be included in the fee.

Data protection is not about protecting information; it is about the citizens' rights to live without the fear of information regarding their private lives becoming public. The legislation protects privacy (such as information about an individual's financial status, health or political views) as well as other rights and freedoms, such as freedom of movement and freedom of assembly. For example, when the travel card system was first introduced in Finland, stampings on different bus lines were saved on the card. It was even possible to receive a print-out later to see where the holder of the travel card had been. This provoked intense discussions about confidentiality and the freedom of movement. Later, it was decided that the stampings would not stay on record.

The cornerstone of data protection legislation is the definition of personal data and the aim is to protect personal data from unjust use, potentially harmful to an individual. The definition on personal data is unequivocally challenging. The European Union has defined the term in a 26-page document (EU 2007). Personal data is easier to understand, if one thinks of anonymous data as an opposite for the term. Anonymous data is information not connectable to an individual. When opening data, it is easiest to start the publishing with anonymous data. The Personal Data Act contains specific regulations on registering and using personal data. Any data that enables the recognition of an individual is not to be made public without the permission of the individual in question. This, also, applies to data enabling indirect recognition. In some cases, personal data may be given out to trusted parties after signing a separate contract. Separate contracts can be drawn up to use personal data for, for example, research. When separate contracts are drawn up, it is vital to make them as transparent as possible, so that anyone could explore the reasons for such a contract.

The legislation regarding the publicity of data is quite straightforward and easy to construe. But, in some cases, it might be in order to limit the legal publicity of data sets, if, for instance, the accessibility of the information becomes drastically easier through Internet distribution. Currently, it is possible to obtain quite a lot of personal data, but it takes time and effort. If all personal data was on the Internet, merely a press of a button away, it might endanger the right for privacy.

For example, the National Land Survey of Finland was in a position where they had information about properties, their owners, lots and forests in the same service. The interpretation was that the name of the property's owner was not to be revealed in the same service, even though the owner's name was public information. The reason for this was that the service enabled estimating the financial value of the owner's wood property. The value of one's property falls under the protection of privacy. These types of situations should be dealt with in advance, and that is why the authorities in charge should be included in the process of planning new ways to utilise data.

Data protection is part of information security, which generally refers to protecting the systems, data, services and data communications in parts, where peoples' basic rights are not threatened. The Data Protection Ombudsman in Finland has suggested that the current, scattered data protection obligations should be centred under one, general information security law (YLE 2007). The need for a general and technology-neutral information security law becomes in-

creasingly greater as cloud services – services provided via Internet – keep growing. When companies and private citizens start commonly storing their own files on the Internet and begin using cloud services, they preferably seek services provided by companies operating in countries people consider trustworthy. If the Data Protection Ombudsman's suggestion is supported, Finland could help implement such legislation on the EU level. However, information security is not examined any further in this guidebook.

3.2 Legislation and Regulations Regarding the Chargeability of Data

According to the Act on Criteria for Charges Payable to the State "a charge shall be made for public administration performances unless there is justifiable cause their being free of charge". The Act on Criteria for Charges Payable to the State gives general basis for chargeability of government performances, such as data transfers, services, or goods. Detailed regulations on charges and their amounts are determined by statutes from different ministries. According to the Act on Criteria for Charges Payable to the State, the size of the charge paid to the State, for a performance under public law, shall correspond to the total costs incurred by the State from producing the performance (cost price). However, the law allows for the charges to be waived completely, or the amounts to be determined lower or higher than the cost price.

The currently effective Act is facing pressure for change when it comes to transferring data (VM 2007). According to a disquisition from the Ministry of Finance in Finland (VM 2005), Finland might fulfil the legal requirements of the PSI Directive. Actions supporting the Directives goals (transparency of public data and making it more accessible through low charges, preferably no charges) have not been taken in Finland. The OECD recommendations emphasise using copyrights in a way that supports the reuse of government data. According to the recommendations, waving of copyrights is to be supported. In addition, pricing should make reuse and accessibility easier, and charges should not exceed the marginal costs.

3.3 Legislation Regarding Access Rights of Data

Copyrights do not protect ideas or facts, they merely protect the form they are given. This means that information content itself, consisting of facts, is not owned by anyone, unless it is stated so in a contract drawn up together with the producer of the data. In the EU, databases are protected as written documents by the copyright laws, or by the so-called Sui Generis protection.

If a database is protected as a written document, the protection covers the distinctive choice of information and the organisation of it in the database. The Sui Generis protection, on the other hand, is directed at the considerable workload related to the collection, manufacturing and presenting of a database. Copyrights on written documents are valid 70 years after the death of the author. The term of Sui Generis protection lasts 15 years after the database is completed, or after the last alteration.

If the data set is not protected by copyrights, the producer and the user of it can draw up a separate contract to control the reuse and redistribution of the

information. Conversely, even though the data is protected by copyrights, it does not mean that one has to control or license the use of it. The author can waive of all rights by simply stating it directly in connection with distribution. For example, Creative Commons Zero licence allows the data to be distributed to everyone through the Internet.

In both cases, the producer of the data set holds all the keys to encourage the reuse of the data. It is important to be familiar with the current state of government-produced data in relation to copyrights, because copyrights arise automatically and may limit the use of the data far more than the producer would have hoped. Different open licences, such as Creative Commons, Science Commons, and Open Database Licence, set the lowest restrictions on the use of the data. There are diverse contract practices between different actors in the public administration regarding the redistribution of data sets. Sentences like "distribution of data is allowed, but not in large quantities" are not uncommon. Contracts are often out-dated, and the Internet may not be considered as a distribution channel for digital information. Instead, the data has been conveyed, by virtue of contract, to be used, for example, for publication purposes. Many obscurities could be avoided, if the attitudes towards immaterial property laws were clearly defined in both municipal and governmental organisations.

3.4 Competition Legislation

The purpose of Act on Competition Restrictions (480/1992) is to protect healthy and safe financial competition from harmful competition restrictions. The law is applied to government agencies and institutions operating in the commodity market. However, a memo from the Ministry of Finances in Finland (11/2004) recommends that government organisations not operate on the commodity market. «*If any government produced information is available in the market, the government should consider the necessity of its data acquisition*». Participation of public sector organisations in the information market may distort the competition. Competition failures emerge when there is a question of actual market performance, i.e. the same information or service is available from other sources to be used for other purposes.

The chain of development of open data should follow the principles of competition legislation. Conforming to the competition legislation means that private companies operating in the market should be allowed to purchase the raw material and the semi-finished product at same cost as it takes to transfer it from one government organisation to another. This places enormous demands on the openness of price formation and the accounting in different organisations. This specifically calls for separate monitoring of economically priced services and advanced price control in accounting (VM 2003).

3.5 EU Directives Regarding Public Sector Data

Two EU Directives are directly linked to open government data: the PSI Directive, aimed at financially utilising open data, which came into effect in 2003 and the 2007 INSPIRE Directive on joint use of spatial data. The goals and methods for the implementation of these two Directives contain essentially the same information as this guidebook. The influence of the PSI Directive has not so far been great in Finland and the implementation of the INSPIRE Directive has only just begun. Time will tell what improvements it will bring.

3.5.1 Financial Utilisation of Public Sector Data and the PSI Directive

Compared to the INSPIRE spatial data Directive, the PSI Directive is fairly little known. According to the Commissions estimate, the Directive has significantly contributed to the formation of internal European market throughout the entire European Union. However, the implementation of the Directive should still be carefully monitored (EC Commission 2009). If the goals set in the 2003 PSI Directive were properly promoted in Finland, this guidebook would no longer be needed. Promoting the financial utilisation of data resources would also support non-financial use and further the development of inter-organisational exchange of information. If implemented, the Directive's articles would significantly further the clarification of licensing protocol and the availability of public data.

For example, article 9, Practical Arrangements, of the Directive states "*Member States shall ensure that practical arrangements are in place that facilitate the search for documents available for re-use, such as assets lists, accessible preferably online, of main documents, and portal sites that are linked to decentralised assets lists.*"

When it comes to pricing data transfers, the Directive sets a maximum fee and suggests the pricing be based on the technical costs of extricating and transferring data. Regarding digital data transmission or copying, billing for publication may result in higher expenses than the actual extrication. Already during the publication of the Directive, the instructions for pricing were a compromise. Since then, arguments for free distribution of public sector data have become stronger and stronger across Europe (see chapter 4).

According to a disquisition from the Ministry (VM 2007), possible instituting of the PSI Directive in Finland has resulted in changing section 34 of the Act on the Openness of Government Activities in a way that requires organisations to determine the fees for transferring data beforehand, and publish them.

Other measures supporting the goals of the Directive, such as the assets lists described in article 9, have not, so far, been taken in Finland. The authorities have created direct application-to-application connections to some of added value service producers. However, these connections are not being actively offered to other operators. The webpages of any authority do not often offer any information on the data resources available for reuse.

3.5.2 INSPIRE Directive and the Legislation on Spatial Data

Geodata adds up to a significant portion of all public sector data, and the opening of it has progressed all the way to the level of legislation. Due to the INSPIRE spatial data Directive, the environmental spatial data will be available, in standardised form, in all Member States. Legislation on spatial data refers to the Act on Spatial Data Infrastructure (421/2009), which came into effect June 17, 2009. The Act and its regulations implement the Directive in Finland. The law is consistent with the demands of the Directive and does not widen the scope of application.

A data set refers to a recognisable, electronic entity of data which can be published. A statute supporting the spatial data legislation defines the geodata affected by the law. In Finland, included are approximately 20 government agencies and institutions, some regional administration organisations and the municipalities. Over all, it includes over a hundred national data sets and several sets from the municipalities. In Finland, the law affects approximately 2,000 data sets.

According to the law, spatial data governing authorities should describe the information by using metadata and interlink them with a service provider. In addition, data sets which are meant for joint use should be placed visibly on the Internet, free for downloading. On the chargeability of the data, the law refers to the Act on Criteria for Charges Payable to the State, which means that possible changes to the Act affect spatial data, as well. However, using metadata is free for all and the possible charges need to be specifically explained. Electronic services and online payment methods need to be implemented before collecting fees. The terms of use and the contract model should be available on the web.

Spatial data legislation and the procedures of its implementation could function as a model for opening other public sector data.

Unfortunately, chargeability leads to incomplete usage of high-quality data resources and gathering the same information repeatedly in several locations.

4. Financial Views on Open Data

In this chapter, we take a look at discussions, arguments and viewpoints on the pricing of public data. Related to this issue are the Act on Criteria for Charges Payable to the State, the European Union's proposal for pricing based on the costs of distribution, the overall economic thinking supported by the government and the international views on open data as an enabler to a national innovation strategy. We will also examine possible solutions for funding open data. There is still rather little international research on the overall benefits and the cumulative effects of open data, so the finances are handled primarily through examples.

Two ideas have continuously come up in the context of public sector data: the presumably positive effects of open data and the concern over the financing of data production.

The financial benefits of open data are based on the idea of data as public goods. Previously, the distribution of information was restricted by the marginal costs of copying, printing, mailing and other logistics. The distribution of electronic information online, however, is virtually free; the small marginal costs consist mainly of moving bits from one data network to another. And the distributed data does not go missing or wear out, no matter how many times it is copied.

Information is always needed for ideas and innovations to emerge. Therefore, ideation and innovation are relevant topics in the discussion on open data.

Traditionally, the government has not been involved in the market in the US. The idea there has been that all tax-paid data should be freely available to everyone – even abroad, via the Internet. The European model of chargeability is being applied to the distribution of many geodata resources. The producer and the administrator of a data resource cover most of their expenses with the income from distributing the data. The pricing of information was widely discussed in Finland during the 1980s. The current Act on Criteria for Charges Payable to the State (1992) is a result of that discussion. The topic was revisited in the mid-1990s, when Finland started to create strategies for an information society (Kuronen 1998a).

Unfortunately, chargeability leads to the incomplete usage of high-quality data resources and gathering the same information repeatedly in several locations (Benson 2009). On the other hand, chargeability limits the development of data-based products. For example, in 2000, the market for risk management products, based on the reuse of weather data, was 50 times larger in the US than in Europe (Weiss 2002).

Marginal costs or non-chargeability, are often seen as alternatives for chargeability. Non-chargeability means that the production, maintenance and distribution of data resources are all managed completely with budget funds. Pricing based on marginal costs, on the other hand, means that budget funds are used for data collection and maintenance, but the additional costs of data distribution are charged from the users.

Free availability of data is an important way of adding to its openness. Free data does not exclude the possibility of business. Fine examples of this are the open source projects. Even though the software is free, chargeable installing, user support, maintenance and other tailored services may be offered to the users.

When considering open data, Finland would benefit from free distribution of data. The current model, based on the Act on Criteria for Charges Payable to the State, allows charges that mainly come from intra-governmental information

trades and small marginal costs. Collecting even the smallest fees need bureaucracy, which is not free to maintain. In addition to the financial aspect, free data equals democracy. People would have free and equal access to data to support their arguments, and they would be free to refine and use it as they see fit.

Case: Intra-governmental data transfers

«Making business with public records has moved money from one pocket to another, but there has been no increase in net income» (Kuronen 1998a)

Most of the intra-governmental data transfer is actually gratuitous transfer. This is due to the fact that the legislation regarding the receiving authority states, that it should have free access to another authority's data. Such decrees can be found, for example, in the Act on National Pension (568/2007), the Police Act (493/1995), the Statistics Act (280/2004) and the Customs Act (1466/1994). Regardless of this, almost half of the income from an authority's data transfers comes from another authority. On two separate disquisitions (2003 and 2004), the Ministry of Finance in Finland looked into the pricing of intra-governmental data transfers. The latter disquisition includes five suggestions for action, which have not been implemented. The first suggestion was to switch to pricing based solely on marginal costs.
In the disquisition, a questionnaire was given out to the 17 largest production and distribution organisations of digital data within the government. According to the survey, these organisations made over 28 million euro in data transfer fees in 2002. The costs of data transfers were reported at 13.5 million euro, which would leave 14.6 million euro as net income. The significance of data transfer income differed between organisations (table 4.1).

Table 4.1 The income from digital data transfers in some organisations in 2002 (VM 2004).

	State	Municipality	Companies and citizens	Total (thousand euro)	Share of income
Population Register Centre	2528	1223	4771	8522	98,0%
Finnish Meteorological Institute	4454	66	2127	6647	86,3%
Finnish Vehicle Administration	330	229	5017	5576	10,0%
National Land Survey of Finland	2728	236	1332	4296	9,0%
Tax Administration	757	151	87	995	27,0%

In 2007, earlier disquisitions were revised and the following was noted: "Since 2004, changes to the pricing of data transfers have been made in the administration of Ministry of Transport and Communication. Marginal cost pricing is applied to data transfers from Digiroad system and vehicle traffic system. Otherwise, the grounds for payments have stayed the same (VM 2007).

4.1 The Changed Operational Environment

Digital data, information and knowledge differ from other factors of production. They can be remoulded, copied and distributed while the original still remains with its owner and producer (Hintikka 1993). It is not free to collect and maintain high-quality data resources or to build a distribution infrastructure, even though the marginal costs from distribution and copying have come down because of digitalisation. The cost structure has shifted from distribution media to production and system investments.

The Internet becoming more common has helped people to see what all this actually means. Lines of business specialising in financial production and selling of information, such as encyclopaedias and mass media, are facing their worst crisis yet. At the same time, free information, products and services are a profitable business on the Internet.

Chris Anderson, the editor-in-chief of Wired magazine, calls this radical change freeconomics (2008). Broadband and the increase in storage capacity have led to free use of some Internet services. It is often more lucrative to offer free basic use of an Internet service to private users than to pay the costs of billing and transactions. Free basic use is financed through other means; according to Anderson, these means include adverts and subsidiaries. Subsidiaries are, for example, consultancy and other additional services that are charged monthly.

The changes in cost structure apply to digital services, as well. It is possible to create route suggestions for people without manpower, using public transport timetable databases. The labour costs are transferred to maintenance and development. Therefore, the capacity of the service can be scaled and it can be provided 24 hours a day.

Digital communication has become part of peoples' everyday lives rather quickly, and there are no signs of the speed slowing down. In the early 1990s, different government organisations in Finland were discussing if every organisation should have their own webpage. Today, the debate is about if the government officials should participate in social media and if machine-readable data should be freely distributed to everyone. Changes brought about by digitalisation are so great that they are going to have an impact on the quality of processes. It is no longer about doing the same things faster and more efficiently; the actors' roles have changed and people are doing completely different things because of automation.

The tax authorities in Finland are among the most progressive actors. They have taken full advantage of digitalisation and changed their operations model. Previously, citizens were struggling with their tax forms every January. Then, the tax authorities changed their system, and now most of the people receive a pre-completed tax return form and simply fill in the necessary information. This saves both the tax authorities' time and tax payers' nerves. The tax authorities did not have to go this far in implementing changes. They could have simply stuck with the old system only allowing tax forms to be filled in online. However, the total benefit would have been smaller.

4.2 The Pre-Internet Legislation on Payable Charges

The current legislation in Finland regarding the chargeability of government produced data is based on the Act on Criteria for Charges Payable to the State, which is from 1992. This was, of course, before the Internet and the changes to distribution structure it brought with it. When the first waves of information society came to Finland in the 1980s, there was a lot of discussion on government produced data. Then, references were made to the US model, where all data produced with tax income is considered a common good. On the other hand, people thought there should be a price tag on the produced data because the production was not free, either. In addition, government organisations wanted a minimum price for government produced data in order to avoid unnecessary requests from the public.

The Act on Criteria for Charges Payable to the State was passed in 1992, right in the middle of the recessionary period in Finland. The recession created pressure for change in the public administration.

The legislation on payable charges regulates the authorities' chargeable actions and gives overall grounds for the chargeability of services, goods and data transfers (chapter 4). When the legislation was passed, there was justifiable cause for the chargeability of data transfers. The data had to be dug up, printed out and mailed. Nowadays, the operational environment is completely different because of digitalisation. Any administration could easily and fully automatically upload their data resources on the Internet without any printing or mailing. Organisations transferring information need to invest on information service systems, but the criterion for charges is no longer applicable.

Some organisations transferring data estimate, that shifting to cost pricing in data transfers would diminish their capacity to develop and introduce new services. The current situation does not support this view; instead it is clear that the information services of organisations collecting high charges are not necessarily more diverse or modern than the services of other organisations.

4.3 EU's Proposal for the Upper Limit of Charges: Marginal Costs

Currently things change fast, and it is sometimes difficult for legislation to keep up. The PSI Directive (see 3.5.1), set in 2003, suggests forfeiting the criterion for charging and that charges should not exceed the marginal costs. At the time of preparation of the Directive, this was reasonable, since, for example, CD-ROMs were a common medium for data publication.

The Directive does not specify the marginal costs. Instead, terms such as costs of reproducing and disseminating are used. Terms like the ones mentioned show, just how old-fashioned the thinking was at the time of preparing the Directive: "The upper limit for charges set in this Directive is without prejudice to the right of Member States or public sector bodies to apply lower charges or no charges at all, and Member States should encourage public sector bodies to make documents available at charges that do not exceed the marginal costs for reproducing and disseminating the documents".

The greatest motive of the Directive is to add growth in the PSI market and, especially, to get companies to re-use the information produced in the public ad-

ministration. In euro, the size of the information market may grow as prices are reduced, allowing new user companies to enter the field. The PSI Directive is on the right path, but it is a product of its time. In the development of digital world, seven years equals eternity. In the time span from the first webpage to the present day, the Directive falls somewhere in the middle. The qualitative change was not taken into consideration when drawing up the Directive, instead it aimed to increase, fasten and facilitate the re-use of public sector data resources using the same process as in the past. What this in fact meant, was that companies were able to have access to data sets only by contract.

One can nowadays gain access to the Digiroad data set, a national road and street database in Finland, for marginal costs, which are approximately a few hundred euros for the entire database. There is a dramatic difference between this and the registers based on the criteria for charges. For example, 3.50 euro is charged for every digital inquiry from the Vehicle Registry in Finland, which means that the cost of the entire register would add up to millions of euro.

Currently, the Digiroad data is delivered on a DVD-ROM disc. Using the cloud services of the web, it would be possible to transfer this amount of data to the subscriber with costs less than one euro. In addition, no start-up investments or maintenance costs are required. The price does not, however, include the possible costs of automating the updates or the registration of users. One has to wonder, if it is not possible to finish what has been started, as we are so close to free data distribution. It is not hard to imagine that the last small fee and the drawing up of written contracts related to it could easily limit the new and creative use of data.

4.3.1 The Effects of Pricing on the Reuse of Data

Generally speaking, the time and money related to utilising data limit the potential for utilisation. However, the effects of pricing may differ from one actor to another. Government organisations and established companies are used to buying data and negotiating for the utilisation of it. It is a different situation with start-up companies, research institutes and the civil sector developing new, innovative operations. If one wants to try utilising data, even the smallest marginal costs and the bureaucracy related to those may diminish the desire to experiment, since the actor does not necessarily know if they are going to have any use for the data in the future.

Whether the data is available for free or for a small charge, does not really matter to companies which already have functioning services and business models based on public data. What is important to these companies is that the pricing and contract practices are precise and that the purchaser of the data has a contract-protected guarantee of the data interface's level of service. From the perspective of restricted competition, it might benefit these established companies to not allow too easy an access to free data.

Research and education are fine examples of situations where the mere lowering of charges would have no effect on the use of the data set. It would take getting rid of all payments to achieve an effect. Typically, the researchers in a research organisation want to experiment with data flexibly, but they have no actual authority to make any kind of purchases. That is why the charges have to be explained and taken through heavy bureaucracy. Experiments often go undone, if the data is not free.

For education and research purposes, different solutions have been created, where the data is available to a certain user group for free. For example, the spatial data service PalTuli, created by the IT Centre for Science (CSC), offers geodata

to Finnish universities to use for research and education. This, however, requires a definition of research use. Defining it can be frustrating, when the line between research and results and financial utilisation is blurry. In addition, the possibilities of starting companies to utilise data should be improved. For this purpose, the so-called start-off contracts were introduced. The contract would allow the distribution of data to new companies, and the payments would not fall due until the data sets would start showing profits (Hermans 2009).

In the long run, the previously mentioned models for charge discrimination are not recommended, no matter how good the intentions. They do not support comprehensively flexible utilisation of data, and maintaining several different contract practices simultaneously only burdens both the provider and the user of the data. The PSI Directive aims to dissolve exclusive and discriminating practices regarding pricing and contracts.

In many cases, it is difficult to say how much income actually comes from selling data or charges based on marginal costs, or how much of the expenses come from the production and distribution of data. The separation of expenses and income specific to each data set would require a highly developed and transparent cost accounting. In order to continue the discussion on the pricing models for the government, it would be beneficial to examine the relation between production, ownership, distribution and income in practice.

4.4 Data as Public Good

One way to approach public data is to think of it as a public good, or as part of a public infrastructure, available for everyone. For example, roads and traffic lights are used by everyone visiting Finland, irrespective of whether they pay taxes in Finland or not. Public good is a good that is non-excludable. Many users can utilise a public good simultaneously without it wearing down and the consumption by one individual does not reduce availability of the good for others. As an example of a public good, Kuronen (1998) mentions radio programmes. The programmes are meant for everyone to listen, and one individual listening does not prevent others from listening to the same programme.

The examination of financial possession of public data can, for the purpose of this guidebook, be divided into three sections: 1) production, 2) ownership and 3) provision and marginal cost. In market thinking, these three aspects are intertwined. The producer of the data becomes the owner and it can provide others with the data for a fee. Regarding public data, the finance of producing data sets should be kept separate from the question of sales.

Production

Collecting public administration's data resources is not exactly a market-based activity. The continuation of data maintenance does not depend on whether an organisation is capable of creating sellable information goods, or if it is done profitably (Kuronen 1998). Parts of the data sets are produced for internal use, and other parts are offered to others. On the other hand, there are many data sets in Finland, e.g. in personal data, where the collected data is overlapping and stored in many registers. In some cases, the municipalities are obligated by legislation to collect raw data to national registers without the collected data benefiting them in any way. In cases like this, the municipalities tend to compensate the costs of the legal obligation by selling the data sets.

By default, the public administration is, in any case, collecting data resources for their own use and for common good, but the legislation on criteria for charges has made this blurry. It is, in fact, a question of perspective: think of a situation

where a community is renting its own real estate to itself. Money is being moved from one unit to another, but in practice the transfer of money only creates expenses.

Ownership

Often, the ownership of information is perceived in the same way as the ownership of any other possessions. In the previous chapter, we examined copyrights and noticed that if copyrights emerge naturally, it is possible to give them up. Information in itself does not belong to copyrights, only the format it is presented in. If data produced in the public sector is seen as a public good, the ownership is hard to determine. Therefore, producing a data set does not automatically lead to ownership. That is why we talk about data held by public administration instead of data owned by public administration.

Provision and marginal costs

Once the data has been collected and organised as databases, through which data is published in downloadable format and through an interface to users, the marginal costs of one extra user are minimum or non-existent. This does not take away the fact, that it costs money to produce the data set and that there are initial investments to be made on a functioning distribution system. Public administration's data is, by no means, free but if so desired, it can be handed over to users for free.

If organisations collecting and maintaining the publication infrastructure are compensated, the result, raw data, can be distributed for free as a public good. As was mentioned previously, the collecting and the maintenance of data sets are usually done for other reasons and with budget funds. The remaining question is the maintenance of distribution infrastructure; whether or not it, too, should be paid with budget funds or covered by charges based on marginal costs.

Prior to the Internet, producing data was related to its distribution but that is no longer the case. Internet allows an effective and economic way to circulate data as a public good. Both the profit and the utility value of public data might be significantly larger if it was provided for everyone to use for free. In the name of equality and simplicity, same conditions should apply to handing out data to government organisations, average citizens and companies.

4.5 Time Consumption

An interesting viewpoint on the financial effects of open data is the time consumption – in other words, how much time do citizens and government officials spend on the acquisition of information and how many potentially useful actions are not taken because the acquisition of information is too laborious. With the help of time consumption, we can exemplify the ratio between opening data and the increase in productivity. From the perpective of innovativeness reducing the time consumed on routines and mechanic actions frees resources to creative thinking and actions.

Releasing time is a positive note from the perspective of democracy and civic activism. The less time is engaged in following the government processes, the better are the chances for citizens to take part in the decision-making in society.

Time consumption is a viable term when, for example, comparing a situation where some information, necessary for the user, is downloadable on the Internet to one where one would have to contact authorities and sign contracts in order to retrieve the same information.

The processes in the public administration can also be modelled based on the

time consumption. In the Netherlands, the government surveys the time citizens spend on their tax forms, reading government letters and other interaction with the government (Den Hurk 2008). The idea is, that time spent on the interaction with the government means less time to do something else. These surveys help create a measurable quantity, which can be used to assess parallel solutions and decrease the load inflicted by the government.

4.6 The Overall Profits of Non-chargeability

Opening government produced data is economically logical. There are still very few absolute statistics. Different preliminary calculations show the overall benefits for business life, citizens and state are greater than the current practice and the legislation based on the criteria for payable charges. When examining the benefits, the costs of the current model and the actual profits of it should be taken into consideration. In addition to economic benefits, the openness of public data enables other functional, societal and cultural benefits, measuring of which is more difficult.

For example, the Measuring European Public Sector Information Resources estimates the size of the European PSI market to be 27 billion euros (Dekkers & al. 2006). The estimates are indicative at best, and especially the estimates on the growth potential of the market depend on how much new business is expected to emerge and whether the calculations include only the direct effects. Whatever the case, we are talking about large economic effects.

Open government data should be seen as an investment, much like any other tools from hardware to software and their maintenance. The analogy is congruent with basic research or infra projects and the financing of those.

There is a strong force of the smallest common denominator in the web. Versatile, simple and easily adoptable solutions covering wide application areas may spread surprisingly widely.

5. Technical Preparations

This chapter gives an overview on the technical terminology related to opening data. We believe everyone working with open data projects should be able to understand the general, technological basics of opening data. We hope that the definitions clarifying the technology could help in practice, when technical experts, other professionals and decision makers are discussing the policies of opening data.

In previous chapters we offered guidance on the inventory of government data resources and examined data distribution from financial and legal perspectives. Hopefully, we have provided the necessary tools for decision-making. The next phase is to define the technical framework for publishing data, which is highly relevant regarding the utilisation and processing of data. If utilising data is too difficult, the possibilities of open data will not materialise.

In the same way we introduced the most important laws and regulations regarding open data in chapter 3, we now begin this chapter with a table listing all technologies and standards introduced in this chapter (Table 5.1). One should not be scared the number of standards, protocols and formats. Memorising or fully comprehending them is not necessary, since it is easy enough to view back on the lists. In this context, they are used mostly as means, through which to refer to the technical solutions of opening data. This is not a comprehensive list but we believe it contains all of the most important modern technologies.

Table 5.1 Technologies related to opening data

Standards for presenting data in a format that enables programmatic processing	**XML:** *Extensible Markup Language is a set of rules for encoding documents and it is expandable for different uses using new marking elements.* **CSV:** *Comma Separated Value* file format, which uses commas to separate values from one another. Files can be opened with spread sheet programmes. **JSON:** *JavaScript Object Notation* is a language-independent, lightweight text-based open standard. **RDF:** *Resource Description Framework* is a standard for linked data paradigm, where individual information resources are described through inter-linked vocabularies. (In this chapter, the abbreviation RDF refers to RDF files in XML format, most often named using the suffix .rdf).
XML-based specified markup languages developed for different uses	**RSS ja GeoRSS:** *Really Simple Syndication* is an XML-based information format for the transmission of feeds. If the feed is attached with geographical coordinates, it is a GeoRSS feed. **ATOM:** Atom refers to two standards, close to each other. Atom Syndication Format is an XML-based markup language for presenting feeds and Atom Publishing Protocol (AtomPub) is a simple HTTP based standard describing programming interface, meant for blog updates. **KML:** *Keyhole Markup Language* is an XML-based language for marking spatial data and using it in context with map services.

»

File formats for presenting documents	**HTML:** *Hypertext Markup Language* is the key file format for www, it enables both presenting the structure (but not the structure of the contents) of webpages and the linking of webpages to each other, forming a net of hypertexts. **PDF:** *Portable Document Format* is a common file format on the Internet; focus is on a good printability of the document.
Transfer protocol	**HTTP:** *Hypertext Transfer Protocol* is one of the key standards on the Internet. **PuSH:** *PubSubHubbub* is a transfer protocol used mainly for quick notifications of information updates.
Technologies for incorporating metadata on the webpage	**Microformats** are a bulk of small formats, comprised of HTML elements, used to incorporate machine-readable information into www pages. **RDFa:** a format used to incorporate machine-readable meanings into www pages.
Naming the resources	**URI:** *Universal Resource Identifier* is an identifier for Internet resources.
Architecture models for creating interfaces	**REST:** *Representational State Transfer* is an architecture model based on HTTP protocol, used for implementing programming interfaces.

5.1 Planning

The machine-readability of data is not a straightforward mechanical procedure. The following questions regarding different implementation possibilities are typical when a data set is published on the web:

- What is the preferred format for publication?
- Which standards to use?
- How should we describe the published information and what metadata should we offer?
- Should we create an interface, or put the data up for download in a file format?

Open source code, open data and open formats are strategic tools for data administration, designed to prevent dependence on a single provider. Open source code allows the government software development not to depend on a single supplier, and open data ensures that the government does not have an unintentional monopoly status on creating and developing new services. Transferring open, or closed, data calls for transfer formats independent of platforms and software.

Standards are important. Open standards are a prerequisite for open market. One might not think about standards daily, but without them the buyers would always be dependent on the manufacturers. In other words, there would be no possibilities to purchase add-ons or other compatible components for previously bought items. We all take for granted that the energy saving light bulb we bought is compatible with the light fitting at home. This type of thinking is relatively new to the field of information technology. For example, it was not until recently that, with help from the EU, the mobile phone manufacturers agreed on how to attach the charger to the phone. In future, the buyer of the phone is no longer dependent on the manufacturer in case the charger gets lost (EU 2009).

Open standardisation is not particularly high on the system supplier's list of priorities. A strong supplier may benefit from its customers being dependent on its products because of, for example, the supplier's use of non-standard file formats. Once the market changes, the supplier may make concessions. They might publish their standard for others to use, but still hold on to the development of it and charge for the use.

5.2 What is Machine-readability?

As a markup language, HTML (table 5.1) is well suited for describing the structure of documents (what is a heading and what is body text), but it does not present the information content in a machine-readable format. For example, if a municipality offers information on the opening hours and addresses of their libraries on their webpage, it is not difficult for a human being to find that information. However, it is relatively difficult for a machine to separate the opening hours and addresses from other information on the webpage.

Let's imagine a programmer creating a "*municipal services on your mobile*" application. The worst case scenario, the programmer would have to copy the opening hours from the webpage by hand and then attach the information to his application. The first step towards helping the programmer would be to offer the address and opening hour information to be downloaded in a comma or Tab separated text file. This would allow the programmer to automatically attach the information to his application.

In our simplified example, the next challenge the programmer faces is to combine the data collected from several sources, depicting the opening hours and addresses of libraries, day cares and public swimming pools. Machine-readability requires combining the metadata from several data sets so that it is mechanically possible to note that the "addresses" of libraries are comparable to "street addresses" of public swimming pools. Often, the harmonising of records is not as simple as our example demonstrates. The unemployment rates in Finland are a fine example. Statistics Finland and the Ministry of Employment and Economy have different ways of calculating unemployment. Their numbers are not easily combined.

5.2.1 Interfaces and Formats According to Data

Once the decision has been made to open data, the first question often is "in which format should we distribute data?" For the users, it is important to have access to data that is programmatically operable and offered in an open format. An open format means that processing the data does not require software from a specific provider. The openness of the format and machine-readability are minimum requirements for machine-processing of the data. Each change from one form to another takes effort, even though some changes are easier to execute than others. Therefore, the formats have a lot to do with the ease (or difficulty) of the use of the data.

Comma or Tab separated data is best suited for presenting name-value pairs in a table format. Almost all computers have a spread sheet application, which allows opening and processing such files. For presenting more complex data structures, XML, JSON and RDF are standards worth mentioning.

In inter-machine data transfer and automatic processing, the lingua franca of the web is XML. It is not a complete format for publishing data, more like a standard which can be used for defining application-specific markup languages in order to present the information. Any existing XML-based language can be

used for publishing more complex data structures, or one can define their own XML structure. To harmonise different markup languages, the linked data paradigm presents records in RDF format (see 5.6).

Compared to XML, JSON is a lighter way of presenting data in a format allowing easy transfer. With JSON syntax, one can present simple name-value pairs as well as more complex data structures in a way that makes them easy to process with common web programming languages.

In the initial stages, the data provider should publish the data in a way that he/she finds the easiest. Later, the usability of data can be improved by converting it to other file formats. As a ground rule, the conversions should be done by the provider. Otherwise every user has to make the same conversions themselves. The same information can be provided in XML, JSON, and RDF formats (see 5.5).

Once the data is available in a machine-readable format, it has to be documented. Documentation gives information on what exactly does each piece of information mean. For example, comma or tab separation allows presenting machine-readable table format data, but it does not give out information on which columns exist in the table. Many tables would be difficult to modify, if there were no column headlines. Therefore, in addition to machine-readability, it is important to offer documentation on what the structure contains. In our example, addresses and opening hours make it easy to figure out what the content is but most cases are far more complex and a clear manual on data is needed.

5.2.2 Machine-readable Licences

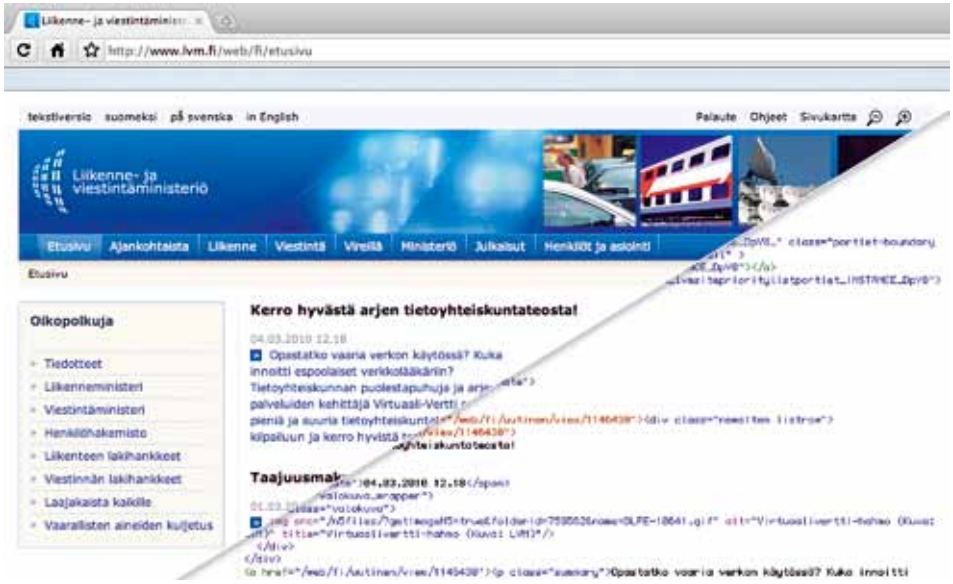
If the data is offered in a machine-readable format, the user licence should be, too. The machine-readability of licences makes it easier for the creators of mashups to monitor the following of rules, since the terms of use can be incorporated into the functionalities of the service.

Machine-readability also makes the information easier to find. For example, Flickr photo sharing application allows users to search for photos corresponding to a search word, whose licence terms allow reuse, for instance in the user's own power point presentation. Creative Commons has, for years, enabled machine-readable terms of use. Even if the data is considered to be a public good, it should be explicitly expressed. This only becomes more apparent as machine-readability becomes more common. For this purpose, Creative Commons Waiver (CC0) licence is a great tool. The PSI Directive, also, recommends using machine-readable licences.

5.3 Web Architecture

The terms www and Internet are often mistaken for synonyms. However, www is merely one of the many services of the Internet. Other services include email, chat groups and data transfer. The foundation for the Internet was laid in the US already in the 1960s, and currently it is a worldwide software infrastructure joining together information networks, through which computers can interact. Www is a scattered hypertext system based on links, and was developed by Tim Berners-Lee in the 1990s. When we in this book discuss the possibility of publishing data sets freely on the web, we refer to the www and incorporating the data to the Web architecture.

W3C is a standardisation organisation which furthers the development of www by creating joint agreements to further the compatibility of different parts



Picture 5.1: Hypertext is read with a browser, which looks for html documents (webpage) from web servers and shows them in a human-readable format.

of the Web. In 2004, W3C finished a recommendation that lists the central parts of Web architecture and design principles (W3C 2004). Unfortunately, these recommendations are not always followed when, for example, forming programmatically created URI identifiers. The recommendation offers a great frame for publishing data in the web.

Web architecture enables gradual development in a scattered web environment. It has been established as protean and has proven the force of one organically growing, linked information space. Later in this chapter we will discuss linked data. It gives us an idea on how to publish data on the Internet, linking it to the worldwide www-information space. REST-style service interfaces, introduced in the context with interfaces, are based on Web architecture and are therefore good platforms for publishing data in the open web.

There is a strong force of the smallest common denominator on the Internet. Versatile, simple and easily adoptable solutions covering wide application areas may spread surprisingly widely. One can save both time and effort, once the technology components, know-how, and practices can be applied to problem-solving. This is exactly what happened with the www, after it became a common tool for executing different services. Practically the entire Internet is based on a few, widely adapted standards (TCP/IP, HTTP, HTML, CSS, JavaScript etc.). These standards are well known by developers, administrators and software architects. In addition, there are many open source code software and advanced tools for implementing systems based on Web architecture.

One of the most important tools is the browser (picture 5.1). The features of the original browser, developed for browsing hypertext, have improved and they are used as user interfaces together with the www for other services on the Internet, such as email, data transfers and instant messaging in social media. Even the term www is beginning to fade from spoken language.

The central term in Web architecture is the global information space, which consists of clearly identifiable interlinked resources. A resource can be a document or an online computer programme which has an unambiguous identifier. The URI identifiers known to all users of the web include webpage addresses, such as http://www.suomi.fi. HTTP-URI is a recommended notation for things brought to the field of Web architecture.

A resource, for example the results of a vote in the parliament, can be presented in different formats, such as HTML, CSV, XML or RDF (see 5.4). In the very core of Web architecture lies the hypertext transfer protocol HTTP (Hypertext Transfer Protocol), which defines the possible functions (GET, PUT, POST and DELETE) regarding the interaction between browsers and www servers. This basic structure has been proven functional by both the users of the web and the software developers.

5.4 Content formats

Formats standardised by the IANA (Internet Assigned Numbers Authority) are used as content formats, or so-called MIME media types on Internet applications. The most common file formats and formats used only on the Internet are listed in these formats. There are many tools for programming languages and several software libraries related to processing open formats. Some manufacturer-specific formats are widely used on the Internet, such as the Microsoft Excel. The usage of such formats is more bound to the tools provided by the manufacturer.

5.4.1 Notifications on Updates through Feeds

Frequently updating information can be presented to the user as a stream of information containing real time updates. The best way to notify on updates is to publish a feed. The feed subscribers automatically receive information about the updates on their systems (e.g. feed readers, browsers or email applications). Information of the feeds is fairly easy to utilise programmatically and, in some cases, the feeds are a valid way of distributing data. However, feeds do not possess a power of expression on machine-readable meanings. In addition to update notifications, feeds are a fine way to mediate headlines, short texts and related links, as was done with rulings of The Courts of Appeal in Finland (http://www.oikeus.fi/rss/ho/hovioikeuksienratkaisut.rss).

There are two commonly used models for depicting feeds: RSS and Atom. Technically both RSS and Atom are XML markup languages. By distributing the data in a feed file format, the content producer enables the republication of the information, as well as regular follow-ups with a feed reader. The RSS model is currently very popular in distributing news, blogs and other topical issues, such as newsletters.

5.4.2 Real-time Web

New technologies enable worldwide and real-time production and following of information. Real-time distribution and follow-up of data (real-time web) has quickly increased its significance in Internet use. How is public administration data connected to real-time web? Waiting lists in health centres, trains arriving to stations on time, and weather reports are examples on situations where real-time data would make people's lives significantly easier and if delayed, the use of the same data would be more limited. The greatest advantages of data are achieved when it is published in real-time and stored for later use, in which case the data can later be used for developing operations, for example.

The problem with aiming for real-time operations is that the Internet was not designed for real-time actions. Until now, the web solutions have checked for

updates by sending "any updates" enquiries to the publisher, even though 99 % of the time the answer is "no". Protocols for creating real-time services have only recently become more common. These protocols operate according to a reversed logic; the publisher sends the updates to the inclined subscribers. Currently, the most popular protocol for real-time web is the PubSubHubbub (PuSH), which sends new contents directly to the subscribers.

5.4.3 Spatial Data in a Reusable Format

Much of any data has to do with a physical location, such as a place, region, or address. Spatial data covers a large portion of all government produced data since the location data, in the form of coordinates, regional domains and addresses, has to do with many government registers. Spatial data is somewhat unique, because it has been administered rather professionally by technically-oriented people, who have worked hard to achieve reusability and to improve distribution.

If location information is provided in a machine-readable format, the data can be illustrated on a map and combined with information from other sources based on the location. The existing structures, such as OpenStreetmap or Google Maps, have made it fairly simple for the programmers to create map services. The common map applications have made it easy to understand how the spatial data is being reused and the discussion around the topic has increased.

Table 5.2. Examples on data typically connected to location

Decision-making	Alteration proposals for zoning, motions and decisions on a specific region or address etc.
Services	Commercial and public services: libraries and hospitals and their opening hours, shops, restaurants, adventure services etc.
Traffic	Information on traffic jams, roadwork, public transportation stops and timetables, real-time locations of vehicles etc.
Weather reports	road weather, forecasts etc.
Media content	Photos, videos, stories linked to a specific location etc.

From a technical aspect, adding spatial data to RSS feeds is fairly simple. A feed containing location marking is called a GeoRSS feed. The interface for Google Maps contains features for depicting GeoRSS feeds on a map. KML Network Link allows an even more versatile publication of spatial data.

Google Maps and Google Earth can both utilise the KML feeds. Information from a specific area is provided to the user and the data is updated whenever the view changes. For this reason, feeds should contain information that does not update automatically but becomes relevant for the users when they are viewing a specific location. The same GeoRSS or KML feeds can be utilised in several map services (or any service utilising spatial data) and vice versa; GeoRSS and KML feeds from several sources can be attached to a single service.

5.4.4 Publishing Documents

Even though documents can be in electronic format, the connotation of the term often implies material artefacts, such as contracts or passports. Often, the minimum requirement for a digital document is that it can be printed out. A grand example of a completely unnecessary document is the flight ticket. Travel agencies send it to the client via email and the client then prints it out "just to be safe". In many cases, it would be reasonable to publish and mediate the information of the documents as such.

Typical examples of documents are the agendas for city council's meetings. In the city councils in Finland, different topics start up as initiatives, move up to meetings and statements, then climb up the ladder to committees and yet another meetings. Matters up for decision are often administered by separate case management systems, from where the agendas of a single meeting are published on the Internet before a meeting. A person interested in a specific matter has to go through a significant pile of agendas and minutes to find what they are looking for. If that person had the chance to search the internal database of the case management system, they could find the relevant information with just a few searches.

5.5 Interfaces, Applications and Services

What are open interfaces? A user interface of computer programmes is a junction between the human user and the programme, through which one can communicate and transfer information. Respectively, there are interfaces between programmes, through which the programmes or parts of the software communicate with each other. Often, the interfaces are meant primarily for internal use of the system and predefined integrations of information technology systems. However, it is fairly common these days to implement an open web API (Application Program Interface), a web programming interface, to a web service. Web APIs are useable through the Internet.

Typically, a web API is implemented into a system which provides other so-called application programmes with different services. It makes programming easier, since everything does not have to be done several times. The map interface of Google, popular among programmers, offers a convenient geocoding tool that converts text-form addresses (and much more) into corresponding geographical coordinates. Programming this yourself would be laborious, even if you had all the necessary data at your disposal.

In its simplest form, a service provided through an interface consists of sending a request and receiving a file in response. Services this simple are often not thought of as services, instead people talk about data interfaces.

Service interfaces are one way of providing machine-readable data for technical systems to use. Examples of services providing public sector-produced information through interfaces include Google Transit, API and World Government Data API, launched by the British publication The Guardian. Google Transit and API offer worldwide information on the routes of public transportation and, through the Guardian interface, access to public data catalogues in the UK, Australia, and New Zealand.

Services offered to programmes through interfaces are not to be confused with HTML-based web applications which are meant for people to read and use. For example, the journey planner of Helsinki Region Transport is one of the most popular web applications in Finland, but there is also an interface connected to the journey planner system, which allows searching information on timetables and stops. Therefore, the journey planner offers the data through the interface to other application programmes, such as the iPhone mobile application.

5.6 Linked Data

A Web architecture based on linked resources is a world-wide information architecture. Data publishers should include their data resources in the architecture to enable sufficient and wide use of the resources. When people spontaneously started linking HTML pages using links and URL addresses, the Internet as we know it was born: a web of interlinked documents. Correspondingly, it is nowadays possible, and fairly common, to link together not only HTML pages but also data. The development is leading to the emergence of a new Internet: a net of interlinked data. The term Linked Data often refers to the four principles (see table 5.3) created by the creator of www, Tim Berners-Lee.

Table 5.3: The four principles of Linked Data (Berners-Lee 2006)

In technical terms	The purpose
1. Use URIs (Universal Resource Identifier) as names for things.	1. To form a concept of the "thing", one that can be talked about and referred to.
2. Use HTTP URIs so that people can look up those names.	2. To offer information related to the concept from where it would naturally be looked from
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)	3. To make it easy to find additional information on named objects and resources.
4. Include links to other URIs, so that they can discover more things.	4. To form relationships between concepts – to create an ever-growing web instead of separate pieces of information.

Regardless of the format the information is being published in and the format it originally existed, RDF is a useful model to connect data resources together via the Internet. As a technology, RDF enables easy linking of things and concepts to each other, as well as the later linking of independent, separately designed systems to each other. Since the RDF enables depicting the same data using different vocabularies, harmonising the terminology in order to increase compatibility can be done where it is most cost efficient.

Regarding publishing open data, it is not necessary to search for joint standards or describe all data, for example information about a single school. The viewpoints of individuals and organisations differ greatly. For example, the school can be found in certain registers as a tenant, the ministry of Education could have a lot of information about the same school and the school itself is in possession of a lot of information, including opening hours etc. There is no reason to assume that all these actors would start using the same standardised terminology to describe the school. Even if a term list was created, it would be a result of a compromise, and therefore would no longer serve the needs of any of the actors.

Commonly accepted terms for description could significantly increase the integrated use of data reserves, and therefore creating such lists should be supported. The force of the smallest common denominator (see 5.3), operating on the Internet, has not in any way prevented the individuals and communities from creating more detailed practices on joint platforms. For example, the markup language for www sites, HTML, is a widely accepted standard. Yet more narrow practices such as the microformats, have been created on top of HTML to allow marking address information in the HTML in a machine-readable format. If these

narrow practices are applied, it is crucial to do it in a way that by no means restricts utilising the rest of the site. The site is visible even if the browser does not support the microformat, and using a microformat in one part of the site does not require the whole website to use all microformats.

In terminology, RDF offers balance between the easiness and the benefits of standardisation. The individuals and organisation publishing data have the right to choose how to describe their data. In addition, they have the chance to distribute their vocabularies and reuse parts of vocabularies created by others. Another point in favour of using RDF is the fact that vocabularies and term lists are easy to add to it later.

The Linked Data web is designed to organically grow and develop, same way as the web of linked documents has done. It grows when people and organisations, irrespective of each other, spontaneously add their own resources on the web and link them together. Self-organising and incoherence are parts of the development: links are broken, new links are made, and term lists and vocabularies are merged and separated.

Wouldn't it be great, if all public sector data could be easily found and available in one location?

6. Open Data Infrastructure

In this chapter, we introduce the idea of data catalogues and the cross-administrative actors contributing to the opening of data. In this context, those actors are referred to as the clearing house organisations.

The cross-administrative actors are consultants for opening data, who help unify the interests of parties. A data catalogue as a concrete website is a service aimed for the publishers and re-users of data, provided by clearing house organisations. A cross-administrative actor can, on a national level, be an independent office of data administration, like the OPSI (Office of Public Sector Information) in the UK or, operate as a supportive organisation regionally or within a city.

A data catalogue is a structured metadata register, which combines metadata from several public sector data sets. Data catalogues can be national (e.g. suomi.fi/datakatalogi or data.gov.uk) or regional (Washington D.C. or the recently published Helsinki Region Infoshare catalogoue data.hri.fi) and can be maintained by cities (San Francisco) or private parties (Sunlight Foundation – National data catalogue). In ideal cases, the administrators of data catalogues can offer support to the publishers regarding, for example, licensing or interfaces, and formats. The administrative party can also be seen as a contact to the users of the data by organising contests encouraging reuse or collecting experiences, feed-back and requests.

Case: Start off easy – The first data catalogue of Finland

The actions of suomi.fi in context with the 2009 Apps for Democracy contest are a fine example of quick response and taking action. Services relying on public, machine-readable information sources were developed in the contest. They needed a place where to gather all the links to the data sets. The editorial team at suomi.fi offered to keep a list of the information sources on a regular webpage found on their site for Laatua Verkkoon (Quality Online) project. The implementation was not ideal for the presentation and visibility of the information sources, but it was done in a very short period of time.

The early stage of the data catalogue (http://suomi.fi/datakatalogi) had a deeper impact than the editorial team might have anticipated. The contest and the catalogue proved to be highly important tools in opening data. Fast, light and well-timed reaction may have led to a better result than a wider but slower action would have.

The site was not merely a list of information sources, it also showed that the government was supportive of the contest and willing to open their data resources. The editorial team of suomi.fi acted as a mediator between other government agencies and the hosts of the contest. Without government support, the contest would have had a different tone. Publishing of the site goes to show the great multiplicative effects a small deed may have on the development of the data ecosystem in Finland.

6.1 Data Catalogue – All Public Data Available at One Place

Wouldn't it be great if all public sector data could be easily found and available at one place?

Through catalogues, the existence of data resources is introduced to the potential users. The need for the public data resources to be easily found was noted on the recommendation by the working group following PSI Directive (see 3.5.1). "Practical projects to create Data resource registers and other PSI infrastructure should be supported both nationally and across Europe". According to the recommendation, data catalogues containing metadata of published and unpublished government data resources in structural format are necessary for increasing the reuse of government data resources and, at the same time, they bring significant benefits to the producing organisations.

During 2009, the idea of all government data being available at one place broke through globally, possibly because of the example set by the Obama administration. Data.gov, operated by the government of the US, was released in May 2009. Since then, dozens of data catalogues listing the data assets of public administrations were published around the world on national and city levels (e.g. San Francisco, London, New York). In many places, such as Sweden, where the government has not yet released a data catalogue, people have been active. The year 2009 can, for a good reason, be called the year of opening government data.

Not including the few exceptions, data catalogues released prior to 2009 focused on more narrow domains, such as spatial data, required registration, and were available only to a limited group of people. In addition to publishing data catalogues, the governments of the US and Great Britain made a strategic commitment to producing open data (Digital Britain and Open Government Directive).

Case: data.gov.uk

So far, the most impressive government data catalogue is data.gov.uk, released on January 19, 2010 as a public beta version. This site, put together in six months under the surveillance of Tim Berners-Lee, beats the US data.gov by a country mile. At the time of publication, the UK catalogue contained approximately 3000 information sets, which is three times more than data.gov.

Not only larger, the UK data catalogue can be considered to be more interesting than the American one. Data.gov has faced serious criticism since its release, arguing that it only contains information that does not provoke political conversation. In the UK catalogue, one can find statistics on the deaths of soldiers on duty and other information interesting in the sense of transparency.

The UK catalogue was designed to support machine-readability and the characteristics of the semantic web. One can search through the database using SPARQL query language. The results – for example British schools - can be examined in the browser and nothing has to be downloaded to the readers' computer. The idea behind the solution used on the catalogue is that the technology it uses is easily adaptable.

6.1.1 Technical Compatibility of Data Catalogues

It is only natural that the government has its own official data catalogue, alongside which local catalogues operate. Regional and municipality-level authorities are significant actors in producing public data because most of the data, interesting enough for reuse, is found in their organisations. In different countries, high-profile national data catalogues have helped the other catalogues to gain publicity. However, the development often starts with a regional data catalogue. In Canada, for example, there are many city-level catalogues but the national catalogue is still in early stages.

In addition to official parties, private operators, including Sunlight Foundation in the US and opengov.se in Sweden, have set up their own public data catalogues. These catalogues were set up for two reasons; firstly because there were no official catalogues and secondly, to allow programmers and citizens to exchange ideas more freely on their own data community website, maintenance of which is independent of the authorities.

The national level is by no means the highest examination level of the utilisation of public administration's data resources. Interest towards cooperation between Member States exists at the EU level. Through cooperation, we might witness the emergence of multinational data catalogues and, later, the harmonisation of data sets to increase their usability.

The UK data catalogue was executed entirely using open source code solutions. Drupal was used as the content management system and the administration of the information's metadata was executed on CKAN platform, which has been in development by Open Knowledge Foundation since 2006. So far, the only commercial system provider to publish their own product for public administrations' data reserves is Microsoft. Their product is called Open Government Data Initiative and it first came into play in the city of Edmonton.

The compatibility of data catalogues has to do with the metadata people want to collect from the public sector's data reserves. Currently, there are no widely excepted standards for it. We should not just wait for them, since the standards develop from existing and emerging catalogues and protocols. At this point, it is important to take care of the compatibility between Finnish catalogues and the interoperability of those and the most important European catalogue, data.gov.uk.

6.2 Work Behind the Data Catalogue

As it seems, the data catalogues at their best are websites, through which one can gain access to public information. Looking deeper, the data catalogues represent the clearing house thinking. A clearing house organisation is an actor operating between the users and the producers of data. Its function is to collect, keep and spread information, metadata and data. Clearing house as a term derives from the world of finance.

For the users of data, such a clearing house organisation may become visible through a well-kept catalogue of public sector's data resources. All legally public information would be found in one place. Together with several other organisations, a clearing house organisation operating in the background of a data catalogue would solve technical, legal and organisational problems. In addition, it would harmonise and uniformly describe data from different sources.

It could also handle the maintenance and development of a public database, allowing the several organisations currently maintaining the so-called operative databases for their own use to outsource the functions to a clearing house organisation and focus on processing information.

In other words, a clearing house organisation would be responsible for the raw data being in a machine-readable format and available to every interested party both inside the government and outside. It would have no obligation to process data or produce any information services for ordinary citizens. Currently, the authorities are obligated to create information portals for citizens and the creation of machine-readable interfaces is often left with very little attention. Offering raw material and the interfaces could be outsourced to a clearing house, and the agencies would be free to focus on their current functions: providing basic services and processing information.

Naturally, processing creates new information that can no longer be called raw data but can be distributed through interfaces for free. In these cases, the agencies could outsource the distribution of produced information to a clearing house.

6.2.1 Unlocking Service of Public Data

Even though public data is available in principle, problems regarding the utilisation may occur when one wants to copy, process or republish the data, or connect it to other data sources. The challenges are often related to chargeability, terms of use, or the formats (see 2.2).

From the users' point of view, it would make sense to be able to report the problems to one instance, which would have the authority and means to improve the usability of the data together with the users. In the UK, the previously mentioned OPSI offers public sector unlocking services, where one can report e.g. the data missing form data.gov.uk. The unlocking service could be compared to the Consumer Disputes Board (in Finland), but at least in the UK, it carries a more positive connotation. Reports to the unlocking service are free for everyone to view and comment on. People can make suggestions for solutions and the reports are handled in order of importance.

References

Benson, Yrjö (2009). Valtion IT-toiminta. Valtion IT-johtaja Yrjö Benson alustus Kuntaliiton seminaarissa 4.11.2009. Helsinki: Kuntaliitto. qsb.webcast.fi/k/kuntaliitto/kuntaliitto_2009_1104_07_benson/. Viitattu 10.3.2010.

Berners-Lee, Tim (2006). Linked Data. www.w3.org/DesignIssues/LinkedData.html. Viitattu 10.3.2010

Cloud, John (2006). The Gurus of YouTube. 16.12.2006. Time magazine. www.time.com/time/printout/0,8816,1570721,00.html. Viewed 8.3.2010.

DCMS, & BIS (2009). Digital Britain. Final report. Building Britain's future. DCMS (Department of culture, media and sport) and BIS (Department of Business Innovations and Skills). www.culture.gov.uk/images/publications/digitalbritain-finalreport-jun09.pdf. Viewed 15.12.2009.

Dekkers, Makx, Polman, Femke, de Velde, Robbin, & **de Vries, Marc** (2006). MEPSIR Measuring European Public Sector Information Resources. Final Report of Study on Exploitation of public sector information – benchmarking of EU framework conditions. ec.europa.eu/information_society/policy/psi/docs/pdfs/mepsir/final_report.pdf. Viewed 10.3.2010.

EU (2007). Lausunto 4/2007 henkilötietojen käsitteestä. 20.6.2007. ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_fi.pdf. Viewed 8.3.2010.

EU (2007b). Communication from the Commission. Pre-commercial Procurement: Driving innovation to ensure sustainable high quality public services in Europe. SEC (2008) 1668. http://ec.europa.eu/invest-in-research/pdf/download_en/com_2007_799.pdf Viewed 17.3.2010

EU (2009). Cellphone charger harmonization. ec.europa.eu/enterprise/sectors/rtte/chargers/index_en.htm. Viewed 8.3.2010.

EU (2010). EUROPE 2020 A strategy for smart, sustainable and inclusive growth. http://ec.europa.eu/eu2020/pdf/COMPLET%20EN%20BARROSO%20%20%20007%20-%20Europe%202020%20-%20EN%20version.pdf Viewed 15.3.2010

Eups20 (2010). An Open Declaration on European Public Services. http://eups20.wordpress.com/the-open-declaration/ Viewed 18.3.2010

EY Komissio (2009). Julkisen sektorin hallussa olevien tietojen uudelleenkäyttö - direktiivin 2003/98/EY uudelleentarkastelu. Komission tiedonanto KOM(2009) 212 lopullinen. www.epsiplatform.eu/media/files/com09_212_fi. Viewed 5.2.2010.

EY (2003). Euroopan parlamentin ja neuvoston direktiivi 2003/98/EY, annettu 17. päivänä marraskuuta 2003, julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä. www.diges.info/pdf/PSI-direktiivi_suomi.pdf. Viewed 8.3.2010.

Forss, Marko (2009). Virtuaalinen lähipoliisiryhmä. Poliisin kotisivut. www.poliisi.fi/irc-galleria. Viewed 8.3.2010.

García-García, J., & **Alonso de Magdaleno, I.** (2010). Commons-based innovation, The Linux kernel case. In 2nd European Conference on Corporate R&D. iri.jrc.ec.europa.eu/concord-2010/posters/Garcia-Garcia.ppt. Viewed 9.3.2010.

HS (2010). Ministeriö pohtii viranomaistietojen vapauttamista verkkoon. Kansalaisista virkatiedon jalostajia. Lehtiartikkeli 7.2.2010. Helsingin Sanomat.

HS (2010). Suomi elää velaksi 2010. Lehtiartikkeli 9.1.2010. Helsingin Sanomat.

Hermans, Outi, & Hermans, Raine (2009). Paikkatietojen yhteiskäyttö ja jakeluperiaatteet. Hinnoitteluperiaatteiden analyysi ja kansantaloudellisten vaikutusten simulointi. Keskusteluaiheita – 1194. www.etla.fi/files/2372_Dp1194.pdf. Viewed 20.12.2009.

Hintikka, Kari A. (1993). Tieto - neljäs tuotannontekijä : tehtaasta televirtuaalisuuteen. Helsinki: Painatuskeskus.

Hintikka, Kari A. (2009). Toimintatapojen muutoksia eli kollektiiviset ja hajautetut toimintamallit. Alustus Suomi.fi:n seminaarissa 7.4.2009. www.slideshare.net/ubiq/suomifi-seminaari-742009-alustus-kari-a-hintikka. Viewed 7.3.2010.

Howe, Jeff (2006). The Rise of Crowdsourcing. Wired magazine. www.wired.com/wired/archive/14.06/crowds.html. Viewed 8.3.2010.

Joey Van Den Hurk, Peter Rem, M. J. (2008). Standard Cost Model for Citizens - User's guide for measuring administrative burdens for Citizens. Ministry of the Interior and kingdom Relations. whatarelief.eu/publications/standard-cost-model. Viewed 10.3.2010

Kansallinen ennakointiverkosto (2008). Internet ja vuorovaikutuksen uudet muodot. (Nieminen-Sundell, Riitta, toim.). www.foresight.fi/wp-content/uploads/2009/08/Internet ja vuorovaikutuksen uudet muodot.pdf. Viewed 19.1.2010.

Karvonen, Erkki (2000). Elämmekö tieto- vai informaatioyhteiskunnassa? Teoksessa Vuorensyrjä, Matti & Savolainen, Reijo (toim.) Tieto ja tietoyhteiskunta. Helsinki: Gaudeamus

Kuntaliitto (2010). Kuntien verkkoviestintäohje. Kuntaliiton verkkojulkaisu. Kuntaliitto. hosted.kuntaliitto.fi/intra/julkaisut/pdf/p20100127150612668.pdf. Viewed 10.3.2010.

Kuronen, Timo (1998). Tietovarantojen hyödyntäminen ja demokratia. Helsinki: Suomen itsenäisyyden juhlarahasto. www.sitra.fi/julkaisut/tietoyhteiskunta/sitra174.pdf?download=Lataa+pdf. Viewed 12.12.2009.

Kuronen, Timo (1998b). Tietovarantojen hyödyntäminen ja demokratia. Esimerkkejä tiedon prosesseista. SITRA 174. Helsinki: Suomen itsenäisyyden juhlarahasto. onko tähän Viewed? -Antti "Jogi" Poikola 10/03/2010 19:51

LVM, & VM (2010). Työryhmät edistämään julkisen tiedon saatavuutta. Tiedote 11.02.2010. Valtiovarainministerio ja Liikenne- ja viestintäministeriö. www.lvm.fi/web/fi/tiedote/view/1130523. Viewed 8.3.2010.

LVM (2009). Sähköisesti nouseva Suomi Viestinnän elinkeinopoliittisen työryhmän loppuraportti 22.10.2009. www.lvm.fi/c/document_library/get_file?folderId=339549&name=DLFE-9534.pdf. Viewed 3.3.2010.

Lasica, Joseph Daniel (2009). Identity in the Age of Cloud Computing. www.aspeninstitute.org/sites/default/files/content/docs/pubs/Identity_in_the_Age_of_Cloud_Computing.pdf. Viewed 5.3.2010.

Luoto, Karoliina (2009). Suomen tietoyhteiskunta 2020. Foresight.fi -blogimerkintä 17.12.2009. Sitra. www.foresight.fi/2009/12/17/suomen-tietoyhteiskunta-2020/. Viewed 15.2.2010.

Mannermaa, Mika (2006). Demokratia tulevaisuuden myllerryksessä. Yhteiskunnallinen vaikuttaminen uudessa viitekehyksessä. Helsinki, Eduskunta, tulevaisuusvaliokunta. www.eduskunta.fi. Viewed 10.3.2010. viitataanko tähän? -Antti "Jogi" Poikola 10/03/2010 20:07

MMM (2004). Kansallinen paikkatietostrategia 2005–2010, Maa- ja metsätalousministeriö MMM:n julkaisuja 10/2004. Retrieved from www.mmm.fi/fi/index/etusivu/maanmittaus_paikkatiedot/paikkatietojenyhteiskaytto/kansallinenpaikkatietostrategia.html. Viewed 10.3.2010.

Mokka, Roope & **Neuvonen, Aleksi** (2006). Yksilön ääni. Hyvinvointivaltio yhteisöjen ajalla. Helsinki: Sitra. www.sitra.fi/julkaisut/raportti69.pdf. Viewed 8.3.2010.

Neuvottelukunta (2009). Tietoyhteiskunnan vauhdittamiseen löydettävä ratkaisuja, Lehdistötiedote 24.11.2009. www.lvm.fi/web/fi/tiedote/view/994565. Viewed 10.3.2010.

Niiniluoto, Ilkka (1989). Informaatio, tieto ja yhteiskunta. Filosofinen käsiteanalyysi. Helsinki: Valtionhallinnon kehittämiskeskus.

Niiniluoto, Ilkka (2000). Uskomuksia ilman tietoa? www.uta.fi/~attove/niini.pdf. Viewed 7.3.2010.

O'Reilly, Tim (2010). Open Government: Collaboration, Transparency, and Participation in Practice. (D. Lathrop & L. Ruma) (p. 432). O'Reilly Media; 1 edition.

OECD (2006). OECD Principles and Guidelines for Access to Research Data from Public Funding. www.oecd.org/dataoecd/9/61/38500813.pdf. Viewed 20.1.2010.

OECD (2008). OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information. www.oecd.org/dataoecd/0/27/40826024.pdf. Viewed 10.3.2010.

Open Knowledge Foundation (2006). Open Knowledge Definition (OKD). Open Knowledge Foundation. www.opendefinition.org/okd/. Viewed 10.3.2010.

Otakantaa.fi (2010). Miten suomalaista tietoyhteiskuntaa pitäisi kehittää? Liikenne- ja viestintäministeriön kysely Otakantaa.fi -palvelussa. www.otakantaa.fi/forum.cfm?group=428&sort=Otsikko&id=&ref=Kaikki keskustelut&pg=4. Viewed 10.3.2010

Paukku, Paul (2009). Julkiset sähköiset palvelut / kehittämisen pullonkaulat. Selvitysmiehen raportti 12.6.2009. www.lvm.fi/web/fi/julkaisu/view/906227. Viewed 7.3.2010.

Pursiainen, Harri (2009). Kansallinen älyliikenteen strategia. Selvitysmiehen ehdotus 30.10.2009. www.lvm.fi/web/fi/tyoryhmat/tyoryhma/view/845060. Viewed 10.3.2010.

Shirky, Clay (2008). Here Comes Everybody: The Power of Organizing Without Organizations. Penguin Press HC.

TEM (2008). Kansallinen innovaatiostrategia. http://www.tem.fi/files/19704/Kansallinen_innovaatiostrategia_12062008.pdf Viewed 15.3.2010

TEM (2009). Kilpailulaki 2010 -työryhmän mietintö. www.tem.fi/files/21617/TEM4.pdf. Viewed 10.3.2010

TEM (2010). Kysyntä- ja käyttäjälähtöisen innovaatiopolitiikan toimenpideohjelma. www.tem.fi/files/26093/OSA_2_final.pdf. Viewed 20.2.2010.

Torkington, Nat (2010). Rethinking Open Data. Lessons learned from the Open Data front lines. O'Reilly Radar blog. radar.oreilly.com/2010/02/rethinking-open-data.html. Viewed 20.1.2010.

Turkki, Teppo (2009). Nykyaikaa etsimässä. Suomen digitaalinen tulevaisuus. www.eva.fi/files/2573_nykyaikaa_etsimassa.pdf. Viewed 8.3.2010.

UK (2009). Putting the Frontline First: smarter government. www.hmg.gov.uk/media/52788/smarter-government-final.pdf. Viewed 8.3.2010

VM (2003). Hallinnon sisäisten tietoluovutusten hinnoittelu, Hallinnon tietoluovutusten hinnoitteluhankkeen raportti 30.5.2003. Työryhmämuistioita 16/2003. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/04_hallinnon_kehittaminen/37893/40524_fi.pdf. Viewed 26.1.2010.

VM (2004). Hallinnon sisäisten tieto- luovutusten maksukäytäntöjä selvittäneen työryhmän muistio. Työryhmämuistioita 11/2004. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/04_hallinnon_kehittaminen/89111_fi.pdf. Viewed 5.3.2010.

VM (2005). Direktiivi 2003/98/EY julkisen sektorin hallussa olevien tietojen uudelleenkäytöstä. Kansallinen implementointi ja vastaavuus Suomen lainsäädäntöön. Muistio 26.5.2005. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/95083.pdf. Viewed 8.3.2010.

VM (2009). Valtiotason arkkitehtuurit -hanke. Julkishallinnon tietoarkkitehtuuri Määrittely 0.3 21.12.2009. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/20100105Valtio/04_Julkishallinnon_tietoarkkitehtuuri_v0.3.pdf. Viewed 15.1.2010.

VM (2009b). Sähköisen asioinnin ja demokratian vauhdittamisohjelman (SADe) toteuttamissuunnitelma 2009-2014. Valtiovarainministeriön muistio 16.6.2009. www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/suunnitelma_SADe_160609.pdf. Viewed 10.3.2010

VM (2009c). Valtiotason arkkitehtuurit -hanke. Valtionhallinnon arkkitehtuuriperiaatteet ja -linjaukset. Määrittely 0.62 Päiväys 31.12.2009. http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/03_muut_asiakirjat/20100105Valtio/01_Valtionhallinnon_arkkitehtuuriperiaatteet_ja_-linjaukset_v0.62.pdf Viewed 15.3.2010

VM (2010). Tilastotoimen tehostamista ja alueellistamista valmistelleen työryhmän loppuraportti. http://www.vm.fi/vm/fi/04_julkaisut_ja_asiakirjat/01_julkaisut/04_hallinnon_kehittaminen/20100302Tilast/Loppuraportti_2_3_2010_liitteet.pdfViewed 10.3.2010.

VNK (2006). Kansallinen tietoyhteiskuntastrategia 2007-2015. Uudistuva, ihmisläheinen ja kilpailukykyinen Suomi. www.arjentietoyhteiskunta.fi/files/34/Kansallinen_tietoyhteiskuntastrategia.pdf. Viewed 10.3.2010.

Voutilainen, Tomi (2009). ICT-oikeus sähköisessä hallinnossa. Väitöskirja Joensuun Yliopisto (p. 360). Edita Publishing Oy.

W3C (2004). Architecture of the World Wide Web, Volume One W3C Recommendation 15 December 2004. W3C. www.w3.org/TR/webarch/. Viewed 10.3.2010

YK (2010). Green Growth. UNESCAP. www.greengrowth.org/.

YLE (2008). Yleisradion hallituksen toimintakertomus v. 2008. yle.fi/yleista/kuvat/2008tilinpaatos.pdf.

YLE (2009). Tietosuojavaltuutettu vaatii yleistä tietoturvalakia. 20.2.2009. YLE Uutiset verkkosivu. yle.fi/uutiset/kotimaa/2009/02/tietosuojavaltuutettu_vaatii_yleista_tietoturvalakia_560232.html. Viewed 8.3.2010.

YLE (2009). Ylen mahdollistajastrategia. Yle. yle.fi/yleista/kuvat/YLE_mahdollistajastrategia.pdf. Viewed 8.3.2010.u_vaatii_yleista_tietoturvalakia_560232.html. Viewed 8.3.2010.

YLE (2009). Ylen mahdollistajastrategia. Yle. http://yle.fi/yleista/kuvat/YLE_mahdollistajastrategia.pdf. Viewed 8.3.2010.