
EPSI platform

European Public Sector Information Platform

Topic Report No. 2012 / 10

Local and regional level information architecture and open data initiatives

Author: Antti Poikola

Published: October 2012

Keywords

PSI, Public Sector Information, Cities, Municipalities, Local and Regional Data, Local and Regional Government, Smart Cities, Information Systems, Data Catalogues

Abstract

Cities around the world are developing local data catalogues and opening up PSI resources

for re-use. City organizations are also focusing on internal architecture work including information and application architectures. These two streams of development, “open data” and “good information management”, are not integrated in most cities.

When cities first start to open up their information resources, these efforts are often separate from their general ICT strategies. Is it possible to integrate these streams? If so, there might be a bright future for re-use of open municipal data.

In order to fully understand the extent of the value of open municipal data we need to know what kind of information resources are produced by city organizations (including those that are not open). This question is important not only from the information re-users perspective, but for the cities internally as well.

By reviewing only currently open data sets it is impossible to have the big picture on municipal data. For example, most data portals only list the data sources that are publicly available, not those that are for internal use only.

The best overall understanding of a city's vast information resources can be got from the people who work with the city’s information architecture and data management. However, even they often struggle with a fragmented landscape of hundreds of ICT systems and thousands of data sources.

This topic report helps in building a big picture on PSI resources of European cities by reviewing the content of currently active local and regional data portals and by covering a case study from Helsinki where over 500 city organization’s ICT systems were categorized and mapped based on their domain of use and information content. This report underlines the importance of integrating data management and architecture activities (especially information architecture) of a city’s organizations with their open data activities.

We suggest that by using the knowledge gathered by internal information architecture activities, opening up of data could be done with a clear strategy, instead of the usual “low-hanging-fruit” strategies employed by so many cities now.

Table of Contents

Abstract 5

1 Introduction 5

2 Municipal and regional data 7

2.1 Benefits of opening up local data 7

2.2 The priority challenge: what to open up first? 8

3 Mapping municipal data 10

3.1 Comparing the data portals 11

3.2 Information asset registers 16

3.3 Information architecture as part of the city's enterprise architecture 16

4 Case: Helsinki city organization 500 ICT systems mapped 17

4.1 What kind of ICT systems does the City of Helsinki have? 18

4.2 From information system listing to a visual map 19

5 Conclusions 21

Abstract

Cities around the world are developing local data catalogues and opening up PSI resources for re-use. City organizations are also focusing on internal architecture work including information and application architectures. These two streams of development, “open data” and “good information management”, are not integrated in most cities.

When cities first start to open up their information resources, these efforts are often separate from their general ICT strategies. Is it possible to integrate these streams? If so, there might be a bright future for re-use of open municipal data.

In order to fully understand the extent of the value of open municipal data we need to know what kind of information resources are produced by city organizations (including those that are not open). This question is important not only from the information re-users perspective, but for the cities internally as well.

By reviewing only currently open data sets it is impossible to have the big picture on municipal data. For example, most data portals only list the data sources that are publicly available, not those that are for internal use only.

The best overall understanding of a city's vast information resources can be got from the people who work with the city’s information architecture and data management. However, even they often struggle with a fragmented landscape of hundreds of ICT systems and thousands of data sources.

This topic report helps in building a big picture on PSI resources of European cities by reviewing the content of currently active local and regional data portals and by covering a case study from Helsinki where over 500 city organization’s ICT systems were categorized and mapped based on their domain of use and information content. This report underlines the importance of integrating data management and architecture activities (especially information architecture) of a city’s organizations with their open data activities.

We suggest that by using the knowledge gathered by internal information architecture activities, opening up of data could be done with a clear strategy, instead of the usual “low-hanging-fruit” strategies employed by so many cities now.

1 Introduction

Because of the number of municipalities and other public sector bodies, tracking and comparing the progress of local level PSI and open data policies is harder than comparing the progress at the level of member states. An earlier European PSI Platform Topic Report: Local and regional PSI re-use open data initiatives, written by Rob Davies, describes the situation as it was in late 2010 and early 2011.

The Open Data Scoreboard on the Openlylocal.com site[1] showed that in December 2010

1 UK Councils Open Data Scoreboard http://openlylocal.com/councils/open

only 37 out of 434 local authorities in the United Kingdom were publishing open data, of which 31 could be considered truly 'open'. A year and half later, in June 2012, the same scoreboard showed that 101 local authorities are publishing open data and 95 can be considered truly open.

In a reasonably short time frame (one and half years), the progress has been quite rapid in UK and elsewhere. This means that cities are now in a good position to learn from each other. Section 2 summarizes the benefits of opening up local and regional data and covers the strategies that can be used to prioritize what data resources should be opened up first.

More and more municipalities have started to publish at least some open data, but it can be argued that the progress has been more wide than deep. Many of the tough questions that were apparent at the time of the earlier topic report are still valid:

1. What business models and income streams for developers are achievable from the re-use of local data?
2. Is there likely to be an extensive (or measurable) expansion of the commercial market?
3. How can public (publishable) data be efficiently separated from data that cannot be opened up in order to avoid infringements of individual privacy?
4. How can the value of public investments in maintaining open data be demonstrated?
5. How can access be provided to those who wish to find data for re-use from more than one locality across local and national borders?

For re-users, the fragmented field of local and municipal data is challenging. At the moment it is extremely difficult to make any data-based application work in multiple cities: it is hard to find comparable data from multiple localities, and even if found, the varying licensing and pricing models make it hard to access. Finally, because of the lack of interoperability and standards, the work load related to re-formatting the data multiplies by the number of localities involved.

This topic report addresses the question of access and discovery of municipal data, by presenting ways to achieve better overall understanding on what information resources cities generally have. Furthermore, we suggest that integrating the open data initiatives with systematic development of a city's information architecture might bring additional internal value which would also justify the public investments in open data.

Improving the discovery of public sector information for external re-users requires that public sector bodies first have a comprehensive overall picture of the data and information systems internally. Section 3 introduces the concepts of open data portals (information about the published open data), information asset registers (information about data that is not yet published) and, finally, the city organization's information architecture activities (organizations' internal knowledge about their information resources) as tools for mapping

what data cities generally have.

The case study covered in Section 4 gives an overview of the information resources possessed by the City of Helsinki. The study also reveals that at present nobody knows exactly all the data and information systems that the city organization has in its dozens of different departments and public service corporations. It has been estimated that altogether the City of Helsinki has about a thousand information systems.

2 Municipal and regional data

A large share of all public sector information is produced at local and regional levels. When we talk about public sector information, the well-known basic registers usually come to mind, such as the population register, the property register and perhaps also statistical data and weather information. Maps and geospatial information are quite well known and some people also mention a variety of traffic information. Local and regional authorities, however, tend to have a slightly different set of data. Data related to cities particularly includes information on living, public transport, safety, social affairs and health, and other services.

The city data can be about schools, catchment areas and property prices; about bus timings and bus-stops, taxi ranks, car parks and traffic congestion; about energy use and CO_2 emissions; about decision making and public spending and so forth. This list can be lengthened and made more specific, but understanding the overall picture is nonetheless difficult.

2.1 Benefits of opening up local data

There a number of benefits to opening up local data. The most important categories are: (i) transparency and local democracy, (ii) economic benefits and (iii) efficiency of local government services supported by better ICT solutions. We explore each of these categories below.

Transparency and local democracy:
Publicly available and understandable information is a prerequisite for a properly functioning modern democracy. Access to relevant and future-oriented information contributes to better decision making and better lifestyles in terms of liveable cities, competitiveness and sustainability.

Economic benefits:
The economic importance of local government information as well as PSI in general comes through direct and indirect benefits. There is a both a potential and actual demand for some fields of local PSI such as land, property, urban planning and environmental data.

Opening up of local information resources may also lead to better services and cost savings for the authorities. Data-based applications developed by the community can cover the small 'niche' requirements of citizens, when municipal resources are insufficient or when it makes no sense to use them for that purpose. For example, in cities where public transport data was open, the first mobile journey planner applications appeared without public

sector investment soon after the rise in the penetration of the iPhone™ and Android devices.

Efficiency of local government:
However, it is not just about transparency and accountability or commercial re-use and mobile apps, important though they are. Internal benefits for the city organization are equally important. An open data policy is a mechanism for cities to "help themselves" to gain more efficiency and better use the existing information resources. Information systems and the operations of a city are closely connected. Knowing what data and what information systems the city has would help to understand the city. Having a clear picture of the city's information landscape is a prerequisite for developing ICT-based smart solutions.

Traditionally the last category, *efficiency of government services supported by better ICT solutions*, has not been a very strong motivation to open up local information resources, since the city organization may develop its own ICT infrastructure but without opening up the data. This may be the reason why PSI re-use and open data initiatives often start as separate actions that are not closely connected to the city organization's general ICT strategies. Over time the open data policy may evolve to be a more integral component of the municipal technology strategy. An example of this kind of integrated thinking is Chicago CTO John Tolva's intention to turn the city into an interactive platform where open data is used to increase the efficiency of government services[2].

2.2 The priority challenge: what to open up first?

The biggest advantages of open data are realised through well-described, standardized data, distributed through reliable interfaces. However, opening up data takes both time and resources, and everything cannot be built up at once. In order to have an impact, a city should look for actions that significantly increase the utility of data.

The question therefore is: Which municipal datasets are of "high value" to re-users? Weather data and geospatial data, for example, are often mentioned as cases that have a clear business potential for re-use of the data. But value can be measured in other ways as well. The number of people whose day-to-day lives can be changed by opening up the data (public utility), the value to scientific research or any of the other benefits mentioned above.

One approach is to determine which data resources have the highest demand. However, this often leads to a chicken-and-egg problem. A representative for a municipality 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".

There might be also hidden demand for certain information, only revealed once the data set has gotten publicity. For example, discussion on spatial data legislation has made spatial data sets better known and, therefore, has increased the demand.

2 Government Technology May 31, 2011. http://www.govtech.com/policy-management/Report-Chicago-CTO-John-Tolva-Open-Data.html

Instead of guessing the potential demand, it is better to ask re-users, lean on experiments, encourage data utilization 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.

One way to approach the question is to ask citizens, businesses or programmers, what data they want. Prior to opening the Berlin open data portal (daten.berlin.de), the City of Berlin polled its citizens[3] to find out which PSI themes it should initially release as open data. In the poll, people chose three themes out of a list of 20 options. The results show that the top-voted six themes (when the first, second and third options were counted together) were: (1) city planning, (2) administration, (3) environment, (4) controlling, (5) infrastructure and (6) population.

When polling potential re-users in this way we do get useful insights, but some do not show in the results. We should keep in mind that: (a) what will be popular in the future might not be popular today, (b) popular is not always the same as important — even if some data is not used by many it may lead to development that has a high societal value and (c) re-users may indicate as interesting only those data sets that they know of.

The Civic Commons wiki page lists four strategies for setting open data priorities: evidence, feedback, low-hanging fruit and high return on investment (ROI)[4].

(i) Evidence: Websites and information management processes often inherently track demand for requests of certain kinds of information. Freedom of information laws usually establish information requesting processes that can be analysed to determine where the demand is. Also website usage, search logs and search engine analysis (i.e. Google Insight) can be used to analyse the types of information that people are looking for on government websites. Naturally when a data portal is opened up, the download and usage statistics provide information on the demand for different data sets.

(ii) Feedback: An obvious strategy is simply to ask the potential and actual data re-users what they are interested in, as was done in the Berlin poll. It is helpful to provide also an on-going open feedback channel specifically centred on open information. These feedback channels are often associated with open data portals. A related strategy is to ask similar governmental bodies what demand they have seen in earlier open data initiatives. In the case of cities, for example, it is likely that citizens will have a similar demand for information in many cities.

(iii) Low-hanging fruit: Some data sets are easier and cheaper to publish than others. If the data is readily available internally and there are no technical or legal challenges in opening up the data, then it should be made available online even if the demand for it is not huge. Exposing the "low-hanging fruit" helps establish a process for other data to be released as well. Additionally, many obscure datasets that are easy to release can prove to have an unpredictably significant public interest. An example of this is the street tree data that was

3 http://epsiplatform.eu/content/first-results-berlin-survey-open-data

4 Civic Commons: Open Data Priorities http://wiki.civiccommons.org/Open_Data_Priorities

opened up in New York.

(iv) High return on investment: There are some datasets which can be determined to have a high return on investment by enabling public utility, safety, cost savings or economic activity. Examples of this are weather data and geospatial data where a small investment in making the data available can enable a relatively huge range of uses and value. Special consideration should be given to data with this potential. However, some of the most obscure data can have some of the highest value. For example, seemingly random environmental data might provide an important correlation to cancer cases.

The above-mentioned strategies can and should be used simultaneously to decide from where to start and how to prioritize. If only one strategy, let's say the "low -hanging fruit", is used, the results (how much the opened-up data is actually used) may not be as good. These strategies indicate which data is probably interesting for the re-users, but they do not help the "hidden demand" problem and, therefore, should not be used to exclude any data sets from the open data initiatives.

3 Mapping municipal data

At the level of this digital information, cities may be more similar to each other than at first glance; for example, when comparing the organisational structures of cities. Although there are differences in the municipal organization and the service production, the needs of citizens are, however, quite the same everywhere. Everyone needs health care, education, water, electricity, heating, the opportunity to move from one place to another, etc. Because of this, organisations responsible for municipal services — be they the city's internal bodies or other actors — record largely similar data in their information systems. Digital information offers a very interesting way to view the city as whole. Such an overall digital picture would be valuable for every city.

The PSI Directive requires member states to ensure that practical arrangements[5] such as asset lists and portal sites are in place to facilitate the search for documents available for re-use. For example, Portugal has implemented this requirement of the directive so that all public sector bodies have an obligation to publish (electronically when possible) lists of the archives of documents available for re-use, both independently as well as in a portal, as soon as possible[6]. The asset lists mentioned in the directive should include both published and unpublished information being held by the public sector bodies.

The local and regional data portals that have proliferated in different cities in recent times, such as Helsinki regional Infoshare (data.hri.fi) and the City of London's London Datastore

5 Directive 2003/98/EC on the re-use of public sector information, Article 9: Practical arrangements http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:345:0090:0096:EN:PDF

6 Tom Kroneburg, Topic Report 7/2011 State of Play: PSI Re-use in Portugal http://epsiplatform.eu/content/topic-report-7-psi-re-use-portugal

are promising initiatives that could be used for building general understanding and mapping data resources between the cities. At the moment the data catalogues mostly list datasets that are already published online and do not mention what data is not available. The realm of possibly publishable datasets (data that is not sensitive by nature, i.e. for privacy or security reasons) that local governments typically have is large but not unlimited and could be therefore presented in the data portals.

In order to avoid the chicken-and-egg problem, mentioned in the previous section, and to facilitate the discussion on the open data priorities, it would be logical to first publish a list of the data sets that the organization has, even if the data is not yet available. Unlike data portals, information asset registers (e.g. the "inforoute" service of the UK Office of Public Sector Information[7]) focus usually on information resources that have not yet been, or will not be, formally published.

Opening up of data for re-use is related to the development of an organization's information systems and the development of information architecture in particular. Organizations need to know what information resources they have in order to manage those resources well. In the case of public sector bodies the same knowledge of the information assets is valuable for potential re-users of the information. Opening up of data can be seen as a support measure for information architecture development, system design and cross-organizational standardization of data resources. The data catalogues and information asset registers should benefit the organization internally as well.

3.1 Comparing the data portals

Open government data catalogues are web portals that offer access to government data sets. The aim is to make it easy for the public to find relevant government information in a single register instead of having to search various sites or use multiple search engines. Data catalogues have existed in other sectors of society for many years; for example, in scientific communities, where there are many comprehensive data catalogues.

The first government data catalogue was the regional data portal of the District of Columbia (data.dc.gov) in the United States, which was launched in October 2008. Since then we have witnessed a wave of national, regional, city-wide and independent data catalogues being launched all over the world. The number of data portals has created interest in analysing the content, comparing the catalogues and building unifying search interfaces[8, 9] to them.

The most comprehensive list of public sector data portals can be found on the datacatalogs.org site, which is curated by a group of leading open data experts from around the world. At the time of writing this report, more than 200 data catalogues were

7 http://www.opsi.gov.uk/iar/index.htm

8 Guardian world government data API http://www.guardian.co.uk/world-government-data

9 CTIC - Public Dataset Catalogs Faceted Browser http://datos.fundacionctic.org/sandbox/catalog/faceted/

listed worldwide. We focused on the sub group called *Official EU Data Catalogues*[10], which is curated by the EPSIplatform team. There are currently over 60 local or regional data portals (and 15 national portals) in the EU member states, which is more than triple compared to the numbers at the end of 2011.

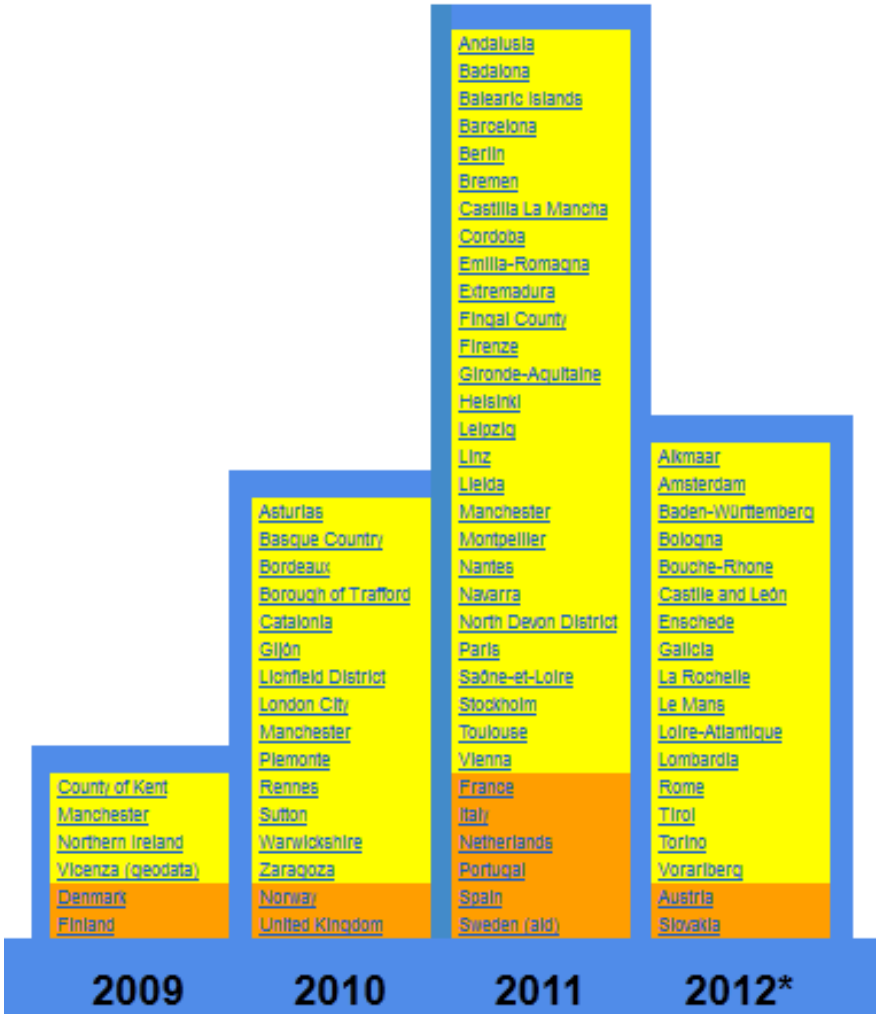

Figure 1: Official EU Data Catalogues from the datacatalogs.org organised by launching year. The orange background indicates a national data portal, and the yellow background, a local or regional data portal.

Koumenides et al (2010)[11] researched the keyword annotations in four national level data catalogues (data.gov, data.gov.uk, OPSI's Information Asset Registers[12] and australia.gov.au). They observed noteworthy variations in the four catalogues. In

10 Official EU Data Catalogues http://datacatalogs.org/dataset?groups=eu-official

11 Koumenides, Christos, Alani, Harith, Shadbolt, Nigel and Salvadores, Manuel (2010) Global Integration of Public Sector Information. In, Web Science Conference 2010, Raleigh, NC, USA, 26–27 Apr 2010. http://eprints.soton.ac.uk/271025/

12 OPSI's Information Asset Register http://www.opsi.gov.uk/about/

data.gov.uk, terms such as “health” and “social care” were prevailing annotators, the data.gov catalogue had more environment-related annotations, e.g. “toxic release”, “chemical release” and “facilities”. Australia’s catalogue gives emphasis to “education” and “environmental management”. And the OPSI concentrates on governmental affairs, e.g. “office services”, “supplier contracts” and “complaints”. More recently, Cyrille Vincey benchmarked[13] the French, UK and US national data catalogues (data.gouv.fr, data.gov.uk, data.gov) and did a quantitative analysis on the data domains, data producers and data quality.

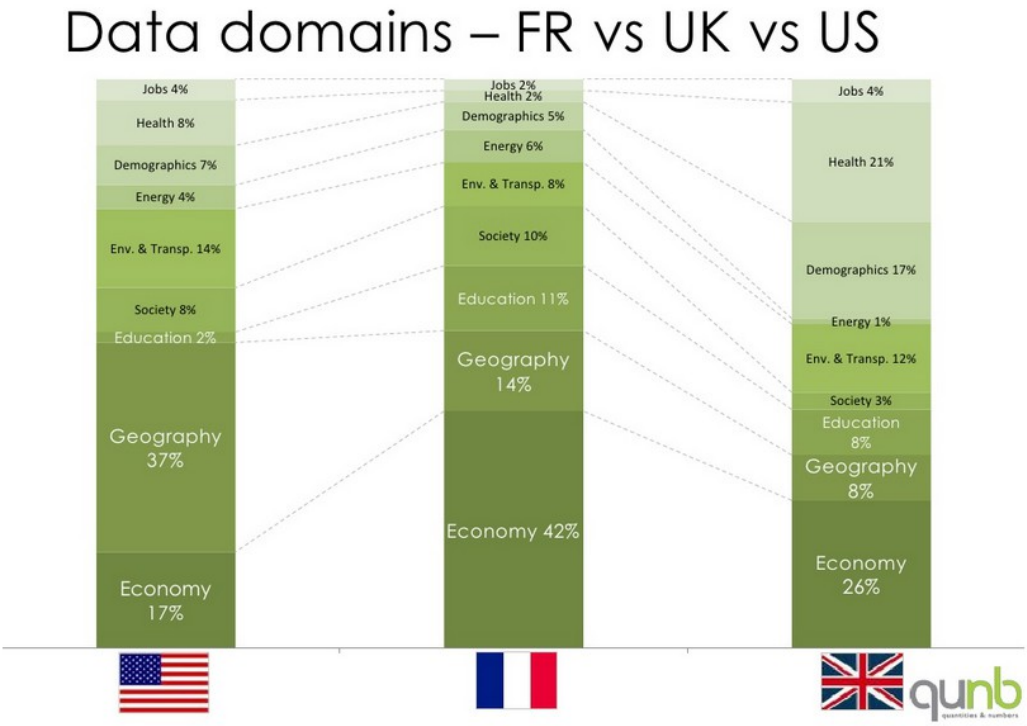

Figure 2: Benchmarking study by Cyrille Vincey compares the most typical categories of published data sets in the French, UK and US national data catalogues. In the French catalogue, the most typical category was the economy at 42%, in the US catalogue the economy had only a 17% share and in the UK catalogue it had a 26% share.

To our knowledge there is no similar content analysis that has been done for local and regional level data catalogues. As a sample we did a light-weight review of the content of six European local or regional data catalogues: Berlin, Helsinki, London, Paris, Torino and Wien. The aim was to shed light on the question: What general categories of information do the cities have and how are they similar? For comparison and benchmarking purposes more detailed analysis would need to be carried out.

Data portals are usually organised so that individual data sets are categorized (one category

13 Cyrille Vincey: Opendata benchmark — FR vs UK vs US, Qunb, July 10, 2012. http://www.slideshare.net/cvincey/opendata-benchmark-fr-vs-uk-vs-us

per data set), tagged (multiple tags per data set) or both. If the data portal uses categories, there are typically around five to twenty main categories. The number of tags may be in the hundreds. We focused on the general categories and found 21 categories that were used in more than one of the catalogues that were reviewed (see Table 1).

Table 1: List of 21 general categories that were used in at least two local data catalogues out of six. In the specific name column, the translated versions of the original category names are listed, which were interpreted to mean the same as the general category.

General Category	Specific names	Berlin	Helsinki	London	Paris	Torino	Wien
Culture	Art and Culture Culture	1	1	1	1	1	1
Education	Education Nurture	1	1	1		1	1
Environment	Environment, Environment and Climate, Environment, Pollution, Recycling and Waste	1	1	1	1		1
Health	Health	1	1	1		1	1
Land use and urban planning	City planning, Geography and Urban planning, Land use, Planning, Town planning, Urban development	1	1	1	1		1
Social and demographic information	Demographics, Family, Livelihood, Living conditions, Social indicators	1	1	1		1	1
Administration	Administration, Administrative units, Public administration, Budget and Tax, Social administration	1	1			1	1
Democracy, transparency and participation	Customers consulting, Democracy and Participation, Elections, Minutes and Resolutions, Transparency	1	1	1		1	
Economy	Economy	1	1		1		1
Services	Opening hours, Other services, Public facilities, Public services, Services		1		1	1	1
Transport	Public transport, Road works, Traffic, Transport		1	1		1	1
Business information	Business and Economy Commercial activities Companies		1	1		1	
Geographic information	Area, Base maps, Maps, Territory		1			1	1
Housing	Housing Housing and Real Estate Living conditions	1	1	1			
Labour	Employment and Skills Labour market	1	1	1			
Population	Citizen, People, Population, Population and Population changes		1		1		1
Security and crime	Control, Courts and Security, Crime, Crime and Community safety, Public safety		1	1		1	
Public finances and spending	Budget Municipal finances		1				1
Sports and physical activity	Sport Sports and Recreation	1		1			
Tourism and events	Events Tourism Travel	1			1		
Young people	Young people Youth work		1	1			

3.2 Information asset registers

Information asset registers (IARs) are registers specifically set up to capture and organise metadata about the vast quantities of information held by government departments and agencies. A comprehensive IAR would include databases, old sets of files, recent electronic files, collections of statistics, research, etc. An IAR does not provide direct access to the information holdings. IARs can include information that is held by public bodies but that has not yet been — and maybe will not be — proactively published. Hence they are the means of alerting the public to the existence of the unpublished information and letting them know whom to contact.

The practice of keeping IARs is done mainly at the national level in a few countries (at least, UK, US, Canada and Australia). In Finland, for example, such registers are not kept, but the legislation requires keeping records of public sector information systems (see Section 4) instead. Even if publicly available IARs are not kept, the organizations need to know, for internal purposes, what information they have and where.

In an EPSIplatform discussion paper[14] from 2008, John Sheridan shares one lesson learned from the UK practice of keeping IARs: *"for most officials a more compelling case can be made for IARs in terms of immediate benefit to public sector information holders themselves. An IAR gives an organization corporate knowledge of what information assets they hold. It is typically of secondary importance to the information holder, that by sharing that information with others, be they other public bodies, private companies or individuals, they can provide a way to let others know what information they may make available for re-use."*

3.3 Information architecture as part of the city's enterprise architecture

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 standardization of data is a key issue, both within the organizations and in co-operation with other organizations. In the 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 description of data sets and definition of information management methods are aimed to solve and prevent problems with compatibility, as well as create an overall view on the information resources in different organizations. The goal is to standardize information architecture and to find common ways of depicting information both within and between organizations.

14 John Sheridan, The role of Information Asset Registers, 10 September 2008, Office of Public Sector Information, UK. http://epsiplatform.eu/sites/default/files/ezpublish_media/Information%20Asset%20Registers%20(OPSI%20Discussion%20Paper)%2010%20Sep%2008.pdf

Compiling of information architecture requires time, hard work and co-operation between the parties. The wider the intended view, the harder the work. Despite the hard work, a good architecture is rewarding: systems are easier to update, working hours are saved, the data resources are of a better quality and, most importantly, data can be re-used even more widely.

4 Case: Helsinki city organization 500 ICT systems mapped

This case study was originally published in a report: *Information navigation in the city* (Viljanen, Poikola and Koponen 2012)[15]. As part of this topic report, the case study is re-published in a summarized form in order to give an initial answer to the question "What information resources might a European city have?"

The study was done by collecting and analysing basic descriptions of ICT systems of over 500 city organizations. The ICT systems were categorized and mapped based on their domain of use and information content.

The resulting information system map is a visual presentation of hundreds of information systems. It is helpful in understanding and discussing the City's digital information resources and gives insight on what kind of information systems other cities might have.

For practical reasons the study was a mapping of the ICT systems and not a mapping of the information resources directly. The City of Helsinki does not have as yet a centralized information asset registry or other database where the metadata of all published and non-published information resources would be stored (There is an open data portal hri.fi which contains the metadata of published data sets.), but the basic descriptions of the ICT systems were available since they are required by the law.

From the system level mapping, the domains of municipal data can be seen indirectly. Interesting questions like how often was data opened up or which data sets are shared between several ICT systems could only have been answered by mapping also the actual information resources, which was not done in this case.

Process:

> Finnish law requires an information system description for all public sector information systems. To receive the most up-to-date information on the systems, a freedom of information request was made at the municipal Registrar's Office.
>
> By collecting these descriptions from 33 municipal agencies in Helsinki, an overview was created of the information systems of the City of Helsinki.
>
> The information systems were then categorized based on their information content and visualized using a network graph.

The law may not require municipalities and cities to provide similar information in other

15 Kim Viljanen, Antti Poikola, Pekka Koponen: Information navigation in the city, City of Helsinki & Fireball project, April 20, 2012. http://www.hel.fi/hel2/ajankohtaista/kavo/information_navigation.pdf

countries. However, at least some information on every city's information systems will probably exist for administrative purposes and similar analysis could be done elsewhere.

4.1 What kind of ICT systems does the City of Helsinki have?

The number of information systems varied widely from agency to agency. On an average, the agencies had approximately 18 catalogued information systems. A few had only one system catalogued. The Port of Helsinki had the largest number of systems (54). To form an overview, the information in the various documents was compiled into one overall list. After faulty entries and duplicates were removed, it included a total of 595 information systems.

Many of the systems are connected to the running of the municipal administration. These systems are typical of all agencies: for instance, the systems related to staff, access control, monitoring of working hours, travel management and financial administration. However, the systems related to production of services are often agency specific. Among the more specialised systems are the traffic light control systems, the database on financial assistance for building of elevators and the registry of trees planted in the city streets. The following table includes more examples of information systems and their contents that do not come to mind very easily.

Table 2: Examples of the systems found by charting the ICT systems of the City of Helsinki.

Agency	Name of system	Intended use	Information content
Port of Helsinki	Vessel diary	Reporting of traffic by vessels	Information on vessels leaving and arriving at the Port of Helsinki
City Museum	MediaKsi	Management system for the City Museum's collections and content information	Information on cataloguing and acquisition of the museum's collections
Helsinki Philharmonic Orchestra	Opas	Tool for preparing sheet music and planning programmes	Information on composers, compositions and line-ups for sheet music; information on concerts; work lists
City Planning Department	Traffic light control systems	Direction of traffic	Use of traffic lights
Real Estate Department/ Housing Division	Database on financial assistance for the installation of lifts	Database software used to compile statistics on municipal 10% lift assistance granted by the City of Helsinki	Information on the recipients of assistance and their agents, amounts of assistance granted, expenditures
Youth Department	Hobby search	Facilitation of hobbies for youth	Information on hobbies, hobby arrangers, facilities; information on youth found in personal registry descriptions
Education Department	Camera surveillance system	Camera surveillance	The system records images of people moving in the cameras' area of operation; dates and times are also recorded, but speech is not recorded
Rescue Department	Merlot Medi Web Reporting	Reporting of paramedics' tasks	Information on clients and measures taken during paramedics' emergency calls
Public Works Surveillance Department	Facta registry of municipalities – Facta building surveillance	Granting of building permits, surveillance during building work, continuous surveillance, etc.	Building and building permits; information on surveillance during building work and continuous surveillance
Public Works Surveillance De-	Comet parking fee system	Loading of money into parking fee meters	

partment			

The general features, informational content and purpose of the information systems were described with free-form keywords, such as economy, billing and purchase bill. Proper names, names of agencies and words referring to geographic areas were avoided to make the keywords as comparable as possible with other cities. Housing, traffic, health, education, administration and many other similar concepts describe cities around the world, although organisations and areas of responsibility may vary.

4.2 From information system listing to a visual map

A list of information systems is interesting, but it is not yet in an easily understandable format. It is hard to see how systems or data sets are linked to each other, or to get an overview of everything that belongs to the digital field formed by the City's information systems

The accompanying image shows the map formed out of the lists of Helsinki's information systems. The visualisation helps one to see the areas in which the systems are grouped according to their content and purpose. On the map, the keywords used to describe the systems form clusters. These can easily be interpreted as larger functional wholes, such as "health" or "environment". The key functional domains have been named in the overall picture. Underneath is an enlargement of the group of keywords related to surveillance and security.

Technically, the map is based on a network graph; the keywords describing the information systems are its nodes. Two keywords (nodes) have been linked with a curved line where an information system has been described with both words.

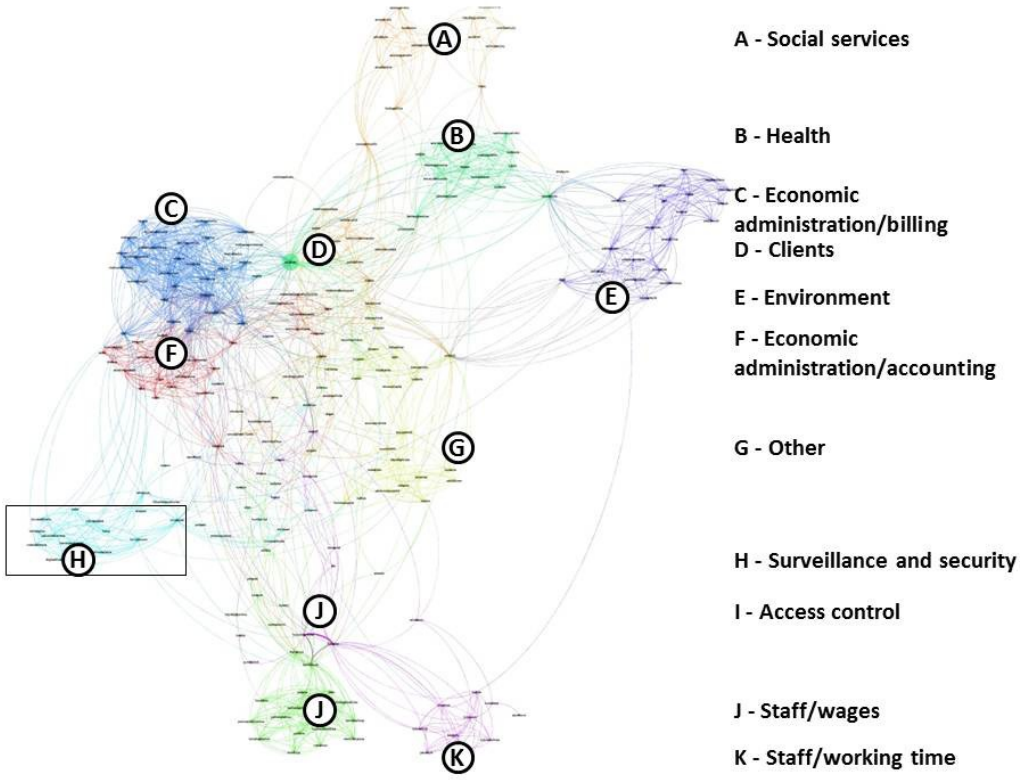

Figure 2: A map formed out of the keywords describing the information systems of the City of Helsinki. The key groups of keywords have been marked on the map. A: Social services, B: Health, C: Economic administration/billing, D: Clients, E: Environment

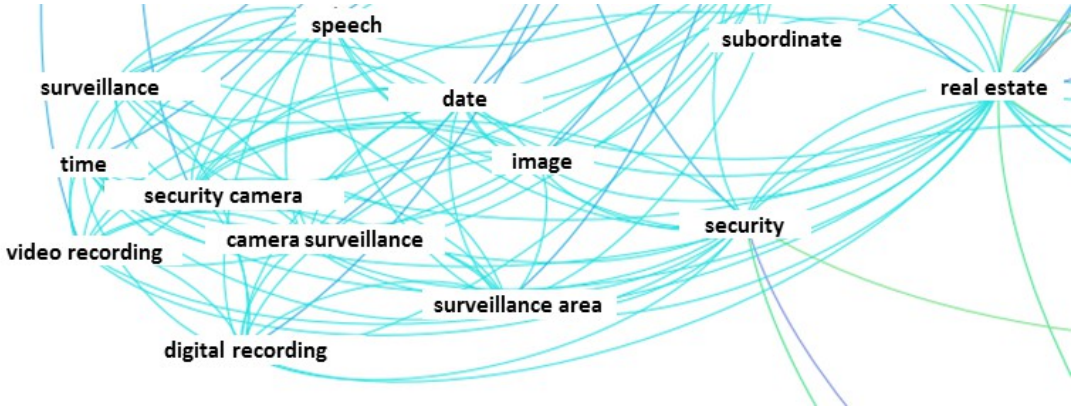

Figure 3: Translated enlargement of "Surveillance and security" (group H above). High-resolution version of the original image where the tags are in Finnish can be found in attachment 1.

The Helsinki information system map presented above is a first sketch of what a map of a city's virtual reality could be. The map gives an overview of the City's information systems that transcends organisational borders. Currently, the financial administrators are familiar with financial software and the geographic information people with geographic information systems; but nobody knows the whole. The information system map could help the persons responsible for the systems of the various agencies to see them as parts of a larger whole.

In developing the architecture of the information systems, the map can be compared to a city map used in city planning. It is a picture of the current state that makes it possible to plan the future.

5 Conclusions

When local level data is sustainably opened up, it is not only a service that a city is obligated to offer, but an integral part of the city's digital strategy. Opening up of data can be seen as a support measure for information architecture development, system design and cross-organizational standardization of data resources.

The public sector information holders as well as the re-users face the same challenge: they both need to know what information assets the organizations are holding. Organizations need to know what information resources they have in order to manage those resources well and the re-users need to discover what exists so that they may use it or request access to it if it is not yet openly available.

Currently the information landscape of cities is fragmented and obscure. It is hard (but not impossible) to build an overall picture on what information resources even exist for one city. The situation gets more complex, however, when cities are compared or when the same information resource, let's say data on public spending, is tried to access from more than one city. For the information re-users, improving discovery and access across localities and borders remains the key challenge.

There are currently over 60 local or regional data portals (and 15 national portals) in the EU member states, which is more than triple compared to the numbers at the end of 2011. The data portals greatly facilitate the discovery and use of the data, which is already published online, but usually they do not say what data is not available. However, IARs include information that is held by public bodies but that has not yet been — and maybe will not be — proactively published. Public IARs however are not common practice at the local and regional levels. The best overall knowledge on what information assets cities have may be gathered as part of the city's information architecture work, but usually kept for internal use only.

To conclude, we offer a future vision where the practices of maintaining IARs of non-published data sets and open data portals of the already-published data sets could be converged into one service. This would simultaneously benefit the data re-users in discovery and serve the organizations internally in their efforts to systematically develop information architecture and data management. Furthermore, this kind of holistic data registers would facilitate city-to-city learning and benchmarking and offer a starting point for harmonization work so that, in the future, it would be easier to access similar data resources in multiple cities.

About the Author

Antti Poikola is an independent consultant at HILA Open Ltd. and team member of the ePSIplatform. Before becoming a full-time entrepreneur he worked as a researcher in media technology at Aalto University. He is active and widely connected in eParticipation and Open Data movements in Europe. He specialises in the exploitation of social media, the development of collaborative solutions between citizens and public administration and the establishment of operating models based on openness. He was the main author of the guidebook: “Public data — an introduction to opening the information resources”[1], which was published by the Finnish Ministry of Transport and Communications 2010. Recently he focused on the cities and has co-authored the report on "Information navigation in the city"[2] for the City of Helsinki as part of the EU-funded Fireball project 2012.

[1] Poikola, A., Kola, P., & Hintikka, K. A. (2010). Julkinen data. Publication of the Ministry of Transport and Communications. http://www.julkinendata.fi/.
[2] Viljanen, K., Poikola, A., & Koponen, P. (2012). Information navigation in the city. City of Helsinki & Fireball project.
http://www.hel.fi/hel2/ajankohtaista/kavo/information_navigation.pdf.

Copyright information