



## Open data guidebook

Final report of the EUROCITIES working group on open data

---

### EUROCITIES

---

EUROCITIES is the political platform for major European cities towards the EU institutions. We network the local governments of over 130 of Europe's largest cities and 40 partner cities that between them govern some 130 million citizens across 35 countries.

[www.eurocities.eu](http://www.eurocities.eu)

# CONTENTS

Introduction .....	3
1. Open data - What? Why? How? .....	4
1.1 What is open data? .....	4
1.2 Why open data? .....	5
1.3 How to open data - organisational .....	7
1.4 City examples of how to start with open data .....	9
2. Technical and financial issues.....	11
2.1 Data collection and formatting .....	11
2.2 Making the data reuseable .....	12
2.3 Putting the data online .....	12
2.4 Financial issues.....	14
3. Open data in cities - EUROCITIES open data working group survey.....	16
4. Conclusions: Beyond open data, towards open government.....	18
Glossary.....	19
Annex 1 - Recommended open data formats.....	23
Annex 2 - Legislative overview .....	23

## INTRODUCTION

When the European Commission published its Directive on the reuse of public sector information (PSI) in 2003, a broader movement for transparency and participation emerged within the EU. Many member states, including France, the United Kingdom, Germany, Netherlands and Spain began to promote and implement open data policies. The directive provided an EU-wide framework for governments, at all levels, to begin opening data.

Since then, ten years of e-Government programmes across all levels of the EU public sector has led to the development of ICT infrastructure in public administrations. We now have the chance to open up policy documents, digital registers, statistics and maps for reuse. The reuse of public sector information (PSI) has a great economic potential. The European Commission estimates the economic value of the PSI market at approximately €40 billion per annum<sup>1</sup>.

*We are sending a strong signal to administrations today. Your data is worth more if you give it away. So start releasing it now: use this framework to join the other smart leaders who are already gaining from embracing open data. Taxpayers have already paid for this information, the least we can do is give it back to those who want to use it in new ways that help people and create jobs and growth.*

Commission Vice President, Neelie Kroes, 08.12.2011- press release

The 2013 revision of the European Commission Directive on the reuse of public sector information will further enable the opening of public sector data in a harmonised and more transparent way, and create the conditions for generating value, both economic and social, from this data.

Local authorities are playing a leading role in implementing open data policies and driving forward open data movements. City governments, together with different stakeholders, are increasingly tapping into this to deliver new services, improve liveability, stimulate business and engage and empower citizens. Open data stimulates the transformation of city authorities into modern open government that is capable of leading, responding to and reaping benefits from the expanding digital agenda.

As some cities push ahead with open data strategies and others just begin, this guidebook provides cities with an introduction and overview of open data. It shares lessons learned by our member cities, examines the reasons for opening data and gives an overview of some of the challenges and solutions in the process.

This open data guidebook has been developed by the EUROCITIES open data working group. EUROCITIES also worked in partnership with the EU Open Cities Project to develop sections of this guidebook. The guidebook is based on contributions from Barcelona, Berlin, Birmingham, Bologna, Copenhagen, Eindhoven, Ghent, Helsinki, Stockholm and Vienna. Particular thanks go to the chair of the working group, Wolfgang Both (Berlin) and the vice chair Bart Rosseau (Ghent).

---

<sup>1</sup> [eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=SEC:2011:1229:FIN:EN:PDF](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=SEC:2011:1229:FIN:EN:PDF)

# 1. OPEN DATA - WHAT? WHY? HOW?

## 1.1 What is open data?

Open data is based on the principle that certain data should be available for everyone to use and reuse, without restriction. There are a number of commonly accepted principles of open data (although some definitions have differences, e.g. on licence issue)

### **Data should be available and accessible**

Data should be available to all, preferably online, and in an accessible format

### **Data should be timely and quality ensured**

Datasets should always be checked and processed to ensure their quality and datasets should be kept as up to date as possible

### **Data formats should be machine readable and non-proprietary**

Data should be made available in open, non-proprietary and easily processable, machine readable formats that enable reuse and redistribution of the data. (See annex 1 for recommendations on data format)

### **Clear licences for reuse**

Establishing a clear licence between the licensor (in this case the local authority) and the licensees (those using and developing services from the data) will address concerns about liability.

In the context of this guidebook we refer to the opening of public sector data for reuse. It is important to point out the distinction between open data and freedom of information acts. The latter establish the citizen's right to ask the public administration for information in most European countries, whereas the former is any government data that is accessible and available for reuse.

## 1.2 Why open data?

The potential benefits of open data collaboration between public administrations, citizens and the private sector are varied and far reaching:

### Economic growth

Opening up public sector data and making use of available resources can unlock significant business opportunities. The 2011 European Commission communication on open data states that opening and re-using public sector information could potentially create economic gains of up to €40 billion annually in the EU.

### Transparency

Making public sector data available can increase transparency between government and citizen. It has the potential to redefine how governments work with their communities and create more open governance. For example, sharing budget data and development proposals online so that citizens can scrutinise, criticise and contribute ideas, is one way of rendering the decision making process more transparent and open. Transparency can improve the image of a local authority and increase the confidence of citizens in politics.

### Quick responses to rapidly evolving problems

Local authorities can use data to provide (real time) information to address issues from traffic congestion to peak load electricity management. Other services such as reporting tools can allow citizens to report local problems to the council just by locating them on maps.

#### City example: Helsinki: [www.hri.fi/en/](http://www.hri.fi/en/)

Developers working for the Helsinki region have tried to visualise the links between data on a map. The map makes large data publicly available and allows users to navigate between data clusters on issues such as health, education, traffic or libraries.

The map is based on a network graph; the keywords describing the information systems are its nodes. The map gives an overview of the city's information systems that transcends organisational borders.

Possible uses for the information system map:

- mapping of public administration and future city planning
- a tool for communication
- tracing unknown areas of open data

### Citizen engagement

Open data can stimulate behavioural change and citizen engagement as it relates to issues such as: finance, citizens' energy use; housing energy efficiency; real time public transport information; and easy access to online information (regarding everything from public bike rental to cultural

events). Hosting hackathons and open data days promotes co-creation and user-driven innovation.

**City example: Cologne:** [buergerhaushalt.stadt-koeln.de/2013/](http://buergerhaushalt.stadt-koeln.de/2013/)

Cologne was the first major city in Germany to implement a citizens' budget based on Neues Kommunales Finanzmanagement (NKF) ('new local finance management'). The pilot project for the 2008 citizens' budget saw the simultaneous launch by the city council of a new e-participation service, which supports citizen involvement. The e-participation approach used in Cologne goes beyond the usual methods of participation. The intention is to transfer the system of participation to be tested in the citizens' budget to all specialised procedures coming under public scrutiny.

### Improved eServices

Merging data digitally leads to improved collaboration between city departments and more efficient internal information sharing. This can also lead to improved e-government services being developed by public administrations.

### Addressing societal challenges

One of the key benefits of making data available on open platforms is its potential to address societal challenges, for example by:

- encouraging and developing digital inclusion
- enhancing the sustainability of healthcare systems
- improving sustainable mobility
- developing new services to measure and cut greenhouse gas emissions, change energy behaviour and increase energy efficiency
- increasing government transparency and building citizen trust
- activating and strengthening quadruple helix co-creation (research, government, industry, civil society)

**City example: Berlin:** [www.wheelmap.org](http://www.wheelmap.org)

A Berlin-based social agency for disabled people launched in 2010 with the aim of identifying problematic areas for disabled people on a city map (e.g. entrances, lifts and other points with poor access for wheelchairs). The project collected official information about public buildings and public transport stops, private buildings, shops, cinemas and other areas. The group 'SozialHelden' ('social heroes') developed and opened this collection on an internet platform (wheelmap.org). This map not only helps disabled people but also, for example, parents with prams, and can be more efficient than transport agencies in informing citizens about dysfunctional lifts at stations.

### Smart cities

Open data will be important for the further development of the European Commission's Smart Cities and Communities Initiative. It can be used to develop smart technologies integrating data from the ICT, energy and mobility sectors. For a city to be truly smart, it must have smart citizens. With open data, citizens can take the initiative, 'do it for themselves', innovate and to co-create.

## New services with public data

The creation of new services is the most visible and widely noted way of utilising public data. The term services is used here in a wider sense, including improved search, browse or query interfaces for end-users of individual datasets and interactive visualisations, but also new services in the original sense: service offerings for customers or citizens who are using open data, but go beyond the actual dataset itself.

### Examples of services built on public data

**Fix my street (UK and Ireland) - [www.fixmystreet.com](http://www.fixmystreet.com)**

This website allows citizens to report problems such as potholes, broken street lighting or similar issues. The site can then use open geographic data on administrative boundaries to direct these complaints to the right authority.

#### Data journalism

Data journalism is another type of service that in many instances is based on public data. By taking one or more public (and other) datasets and visualising them - often interactively - data journalists are able to tell the story behind the data and inform their readers in ways that weren't possible before. The Guardian's Datablog ([www.guardian.co.uk/news/datablog](http://www.guardian.co.uk/news/datablog)) was and still is one of the first prominent examples of data journalism. Infographic is a form of data journalism that is becoming increasingly popular.

## 1.3 How to open data - organisational

To establish open data as a new service there are a number of key steps to take that range from political and citizen engagement to the technical process of opening data<sup>2</sup>. In terms of organisational processes, the following are important:

### Clear open data policy

An open data policy with clear aims and goals is needed to ensure a clear direction and means to measure success.

### Deciding what data to open

When starting with open data it is a good idea for cities to focus on issues of high local relevance that have the greatest potential to raise interest. Gaining citizen feedback before opening data can be a useful means of determining what data is of the greatest interest.

### Political support

Agreement and support for open data policies is needed at the political level. This is vital for

<sup>2</sup> [de.slideshare.net/jkongga/open-data-current-state-next-gen](http://de.slideshare.net/jkongga/open-data-current-state-next-gen)

ensuring that the whole public administration is pulling in the same direction and that an open data policy can be pursued until its aims are achieved. Politicians and city management should be presented with a clear open data policy that is supported with case studies and success stories from other cities in order to convey the value (economic and social) of opening data.

### **City administration organisation**

Developing an internal open data strategy and a culture of data sharing can be challenging in terms of reorganising city administrations to alter existing management processes, legacy (existing IT departments, protective software) and purchase structures. All city departments should be involved in developing approaches to address disclosure and access to data. Developing common criteria for data disclosure and the creation of dedicated data management teams and contact officers can help. Additionally, it is vital to educate staff to understand what open data is, and on legal and technical issues, such as the use of software specific data modelling techniques.

### **Working with stakeholders**

A vibrant community of citizens and stakeholders is crucial for the development of open data initiatives. In most cases, the developer of services will not be the public administration itself, but companies, interest groups, NGOs, individual citizens and students. The administration can provide incentives, and in some instances work together with third parties to ensure the sustainability of new services. In order to promote the data and bring the right people together, the city administration should support stakeholder and citizen engagement through, for example, local hackathons and data demonstration days.

### **Encouraging data literacy amongst citizens**

Many citizens still lack the basic knowledge needed to understand and process information. Short courses, as well as blogs and podcasts can help to improve understanding, making it possible to significantly increase the number of data users.

### **Scepticism**

Open data is still a relatively new concept and whilst the value of the data is understood, there is still not the critical mass of useful applications developed to make it acceptable or even known about beyond the technical community. Commercial applications that are starting to reach the market place can overcome the scepticism.

### **Privacy**

All too often issues of data protection can be used as a barrier to publishing data. Many of these issues can be overcome by ensuring the data is aggregated to the correct level before publication to protect the privacy of the individual. It is also important to communicate to people that that open data refers to data that is not subject to valid privacy concerns.

## Quality of data

There can often be issues concerning the quality of data and a reluctance to open data that may have discrepancies. Rather than seeing this as an obstacle, public bodies can use it as an opportunity to receive feedback and improve the quality of the data.

## Legal support for open data initiatives

Legal concerns are one of the main obstacles to the development of innovative services and applications within administrations. Most of the regulations affecting data management in public administration were put in place before the spread of the internet. In several countries, extremely strict privacy regulations still make it impossible for public administrations to publish information such as the name, type, and geographical location of schools, and copyright issues make it impossible to disclose to the public domain pictures or multimedia contents.

The availability of legal support is crucial in the context of open data due to the inherent complexity of the matter and the need to undergo legal scrutiny before publishing every dataset. In addition, legal advice is required on the most appropriate licences to apply to disclosed data.

## 1.4 City examples of how to start with open data

Several cities have already opened their data files and launched a city data portal. London launched a portal in January 2010. This portal structure and the data register are a good example for other cities. Further data portals were launched soon after this, for example in Paris (launched in January 2011), Vienna (May 2011), and Berlin (September 2011). The experiences of these cities can help other cities in developing a city data portal. Below are some examples of how to set up and popularise open data platforms.

### Berlin open data day

The city of Berlin has been cooperating successfully with the internet community since 2010. This community-public-partnership designed and organised the first Berlin open data day (BODDy) in May 2011. More than 120 participants listened to the opening speeches by the deputy mayor and the city's chief information officer (CIO) as well as speakers from the net community. Several examples for running apps showcased a broad spectrum of applications based on open data.

The launch of the city data portal was prepared with a pre-launch test within the internet community. This provided a very helpful step in the portal's development. The 2012 and 2013 BODDy built on this success by attracting up to 200 participants.

### London data store - [www.data.london.gov.uk](http://www.data.london.gov.uk)

More than 200 datasets detailing life in London have been put online for free reuse by growing community of web developers and programmers. This includes information about planning decisions, crime rates, abandoned vehicles, house prices, road accidents and many other metrics. The developers with the most innovative ideas to harness the data are offered a grant to help them bring their idea to life.

### Vienna online survey

The city of Vienna conducted an online survey at an early stage of the implementation of their open data strategy in 2011. The survey looked at open government data and open government in general. A total of 376 persons participated in this survey. The results provided useful information for the further development of open data in the city

In particular, many participants responded to the open questions (e.g. 'What kind of data from the city of Vienna would you need?') and made specific requests. This feedback was analysed and helped to prioritise what data to publish. The raw data from this survey was published on Vienna's data portal.

#### **Important results of the survey:**

- more men than women participated in the survey
- women were more interested in health and men were more interested in traffic
- in the evaluation of using human-readable data, omen and men are practically at the same skill level - but the data is used three times more frequently by men
- traffic, public facilities and leisure were the most important local issues

## 2. TECHNICAL AND FINANCIAL ISSUES

Making sure data is available and reusable brings with it a number of technical challenges that must be overcome:

- data collection and formatting
- making the data reusable
- putting the data online

### 2.1 Data collection and formatting

Collected data can often be characterised by differences in terms of granularity (how datasets are sub-divided), and a lack of globally consistent identifiers, making it very difficult to correlate data from different sources. In addition, even though most data is managed by exploiting databases and data warehouses, city administrations may still organise and manage relevant data in the form of documents, e.g. Microsoft Excel spreadsheets. Available datasets are also often distributed across different branches of city administrations and, as a result of decentralisation in information management, available data can be in different formats.

#### Collecting and formatting data

Addressing the above issues is challenging. No single public sector organisation has all of the information about all of the services for the communities that it serves. Creating a centralised management and collection of data would be the ideal solution, but can be difficult in terms of capacity and finance.

There are a number of factors that can help ensure better coordination of data collection and formatting:

- all city departments use a common data collection procedure and format (as far as possible)
- each department publishes only the information that they are responsible for
- links within each data row are made with rows from other datasets and from other departments
- links are made to a small set of common definitions and identifiers so that it is clear when data from many sources is referring to the same 'thing'
- links are made from those definitions and identifiers to other data
- users can link their data and opinions to the public data

There are resources available online to help develop common identifiers and definitions for datasets. In particular, useful examples can be found for an ontology to describe open public services<sup>3</sup>; for common use of terms and vocabularies<sup>4</sup>; and for common identifiers<sup>5</sup>.

<sup>3</sup> Catalogue of public services in Germany (Leistungskatalog): [www.gk-leika.de/startseite/](http://www.gk-leika.de/startseite/)

<sup>4</sup> Local government business model: [www.standards.esd.org.uk/LGBMDiagram.aspx](http://www.standards.esd.org.uk/LGBMDiagram.aspx)

<sup>5</sup> URI - set of companies from Companies House: [www.companieshouse.gov.uk/about/mi](http://www.companieshouse.gov.uk/about/mi)

Making sure that data can be linked internally and that there are common formats and collection procedures will result in coherent and useable data that can be easily identified and utilised. Once this is in place then it is the task of the city administration to make the data available for reuse.

## 2.2 Making the data reuseable

- data assessment - privacy and quality
- data processing and cleaning
- formatting and transcoding data

As a first step, the data assessment process must identify institutional data owners, as well as determine if current regulations allow disclosure of information. It is then crucial to document exhaustively disclosed information, explaining its meaning and also the meta-data that describes the method (how and when data was collected) and how and why the data was pre-processed (e.g. how figures were calculated).

It is important for the credibility of the public authority that the quality of the data is verified to the greatest degree possible. Institutional audit mechanisms guarantee the quality of the data and ensure datasets are anonymised to protect personal information.

Once the dataset has been analysed and its schema<sup>6</sup> defined (ideally the schema should be defined with the data owner) the source datasets need to be transcoded to an open data format<sup>7</sup>. It may often be advisable to identify trade-offs between the level of detail of the data and simplicity when transcoding. This is because while in principle it is desirable to provide users with the all the data, in practice the risk is providing excessively detailed information that is difficult to understand. Decisions on the level of data granularity and schema should be specific for each different dataset.

In line with the principles of open data, the datasets should be transcoded and made available in open, non-proprietary and easily processable, machine-readable formats and published online.

## 2.3 Putting the data online

A common approach is to make data available for downloading in a zip file. This approach is pragmatic but does not allow for the development of mash-up<sup>8</sup> applications that utilise different data sources. In addition, this approach cannot be easily applied to large datasets, because of the amount of bandwidth and computational resources that are needed to distribute such large datasets.

Overcoming the above limitations requires the development of multiple distribution channels for data. One method of doing this is through linked data.

---

<sup>6</sup> A database 'schema' is a collection of meta-data that describes the relations in a database. A schema can be simply described as the 'layout' of a database or the blueprint that outlines the way data is organised into tables.

<sup>7</sup> This process is typically straightforward and, when possible, should occur automatically. The format and encoding of a dataset can be checked automatically, as can the validity of the data with respect to a certain format (e.g. perform a schema validation).

<sup>8</sup> A mash-up, in web development, is a web page, or web application, that uses content from more than one source to create a single new service.

## Linked data

Linked data is a method of publishing data so that it can be interlinked with other data, which often increases its value significantly. It allows the data consumer to create innovative services and applications by combining different datasets. There are many examples from different regions or countries, such as integrating mobility statistics with environmental data to see correlations or comparing data on the same issue.

### Sir Tim Berners-Lee's 5 star approach for Linked Open Data<sup>9</sup>

1. Data is available on the web (whatever format) but with an open licence, to be open data.
2. Data is available as machine-readable structured data (e.g. excel instead of image scan of a table). Thus, the reuser can process, export and publish the data easily, still depending however on proprietary software like Word or Excel.
3. Available in non-proprietary format so the user won't be confined to a particular software.
4. All the above plus use open standards from the World Wide Web Consortium - W3C (such as RDF<sup>10</sup> and SPARQL<sup>11</sup>) to identify things.
5. All the above, and link the data to other people's data to provide context.

However, data integration is often a difficult problem, requiring significant resources. Typical reasons for this are the use of different formats (CSV, Excel, XML, KML, etc.) or that data is structured in ways that do not easily match up, and importantly, that common global identifiers (e.g. geographical units, administrative departments, people, laws) are not used in different data sources. This may make it necessary to limit the use of linked data to selected datasets that are compatible with others.

## Creating a data portal

To make the data available and visible online, a good data portal is advisable. The portal should be as accessible as possible and meet the needs of: coders, who have a technical background and use the portal to retrieve the data; data owners, who want the portal to reflect the identity of the organisation; and the 'general' audience, who are looking for general information about the data and developed applications.

A good portal allows the city administration to manage and update the data when necessary. It should allow for feedback opportunities so the data quality can be upgraded and the usage of the datasets tracked. It should also have a good user interface, allowing users to easily identify and

<sup>9</sup> [www.w3.org/DesignIssues/LinkedData.html](http://www.w3.org/DesignIssues/LinkedData.html)

<sup>10</sup> Resource Description Framework - The Resource Description Framework (RDF) is a family of World Wide Web Consortium (W3C) specifications.

<sup>11</sup> SPARQL - (SPARQL Protocol and RDF Query Language) is an RDF query language, that is, a query language for databases, able to retrieve and manipulate data stored in RDF format.

navigate the website. A solid and unchanging URL is quite important to make datasets useable.<sup>12</sup>

The European Commission is currently developing a pan-European data portal, providing access to open, freely reusable datasets from local, regional and national public bodies across Europe: [www.publicdata.eu](http://www.publicdata.eu).

### Have a clear licence for reuse

One of the central and most crucial issues around open data is licensing. Failing to apply a clear and precise licence to a dataset in a visible and obvious manner can discourage potential users from using it. In fact, while closed and restrictive licences are problematic and should be avoided in favour of an open licence, any licence is always better than none: users should at least be clear about if and how they can use a dataset.

## 2.4 Financial issues

Working with open data requires careful planning, and city governments need to allocate a budget that guarantees the success of their endeavours.

### Additional costs

For some processes, such as building open data platforms that enable API<sup>13</sup> data release, linked data or server extensions, additional costs exist. Costs associated with technology, organisational changes, and promotion of the data (e.g. hosting hackathons or information days) have the greatest impact on the public budget. The cost of a project preparing an open data portal infrastructure can potentially be as much as €200,000. The annual cost of its operation can in some cases be €250,000.

### The availability of technology and know-how

City administrations need to acquire technology to store, manage and distribute open data. In the first stage this requires the availability of a straightforward web application. However, over time it becomes crucial to identify the best-suited technologies to allow the administration to manage growing amounts of data and to align its open data efforts with emerging standards in the field. This requires investment from the administration.

### Loss of public income

If data has previously been sold for reuse, opening it up for free represents a loss of income. In particular, geospatial data is sold by many cities. This loss can sometimes be balanced by tax income from new services using free public data.

---

<sup>12</sup> City data portal URLs: Barcelona: [opendata.bcn.cat/opendata/](http://opendata.bcn.cat/opendata/); Berlin: [daten.berlin.de/](http://daten.berlin.de/); Birmingham: [www.birmingham.gov.uk/open-data](http://www.birmingham.gov.uk/open-data); Bologna: [datacatalogs.org/catalog/bologna-dataportal](http://datacatalogs.org/catalog/bologna-dataportal); Eindhoven: [www.openeindhoven.nl/data-eindhoven/](http://www.openeindhoven.nl/data-eindhoven/); Ghent: [data.gent.be/](http://data.gent.be/); Helsinki: [www.hri.fi/en/data-search/](http://www.hri.fi/en/data-search/); Stockholm: [open.stockholm.se/oppna-data](http://open.stockholm.se/oppna-data); Vienna: [data.wien.gv.at/](http://data.wien.gv.at/)

<sup>13</sup> An application programming interface (API) specifies how some software components should interact with each other. A good API makes it easier to develop a program by providing all the building blocks. A programmer then puts the blocks together.

### Promotional activities

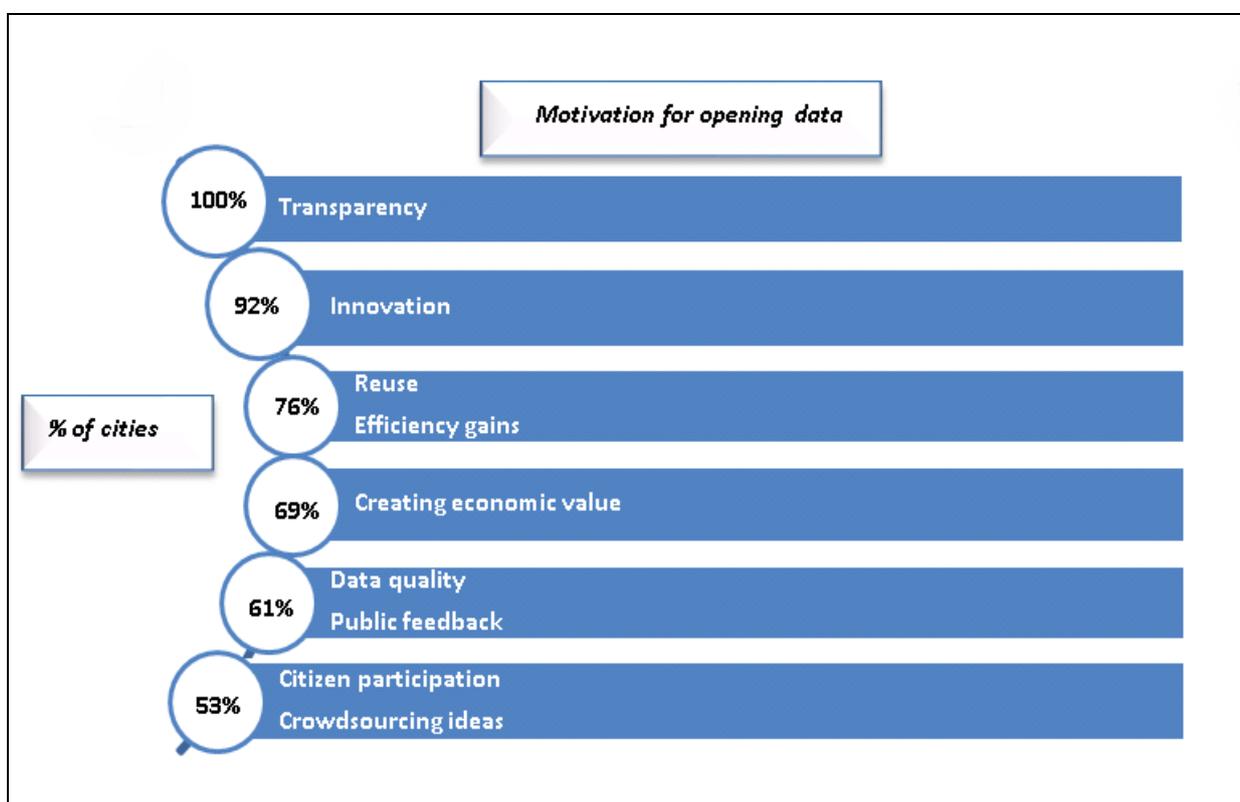
It is vital to guarantee the visibility of open data initiatives. Various experiences demonstrate the effectiveness of public events and the promotion of hacking initiatives devoted to the development of open data based services and applications. Resources are needed to gather and manage feedback these initiatives. Applying for EU or national funding can help local authorities fund some open data activities. Utilising new innovative funding schemes such as crowd funding may also provide a source of funding while also encouraging citizen participation and engagement.

### 3. OPEN DATA IN CITIES - EUROCITIES OPEN DATA WORKING GROUP SURVEY

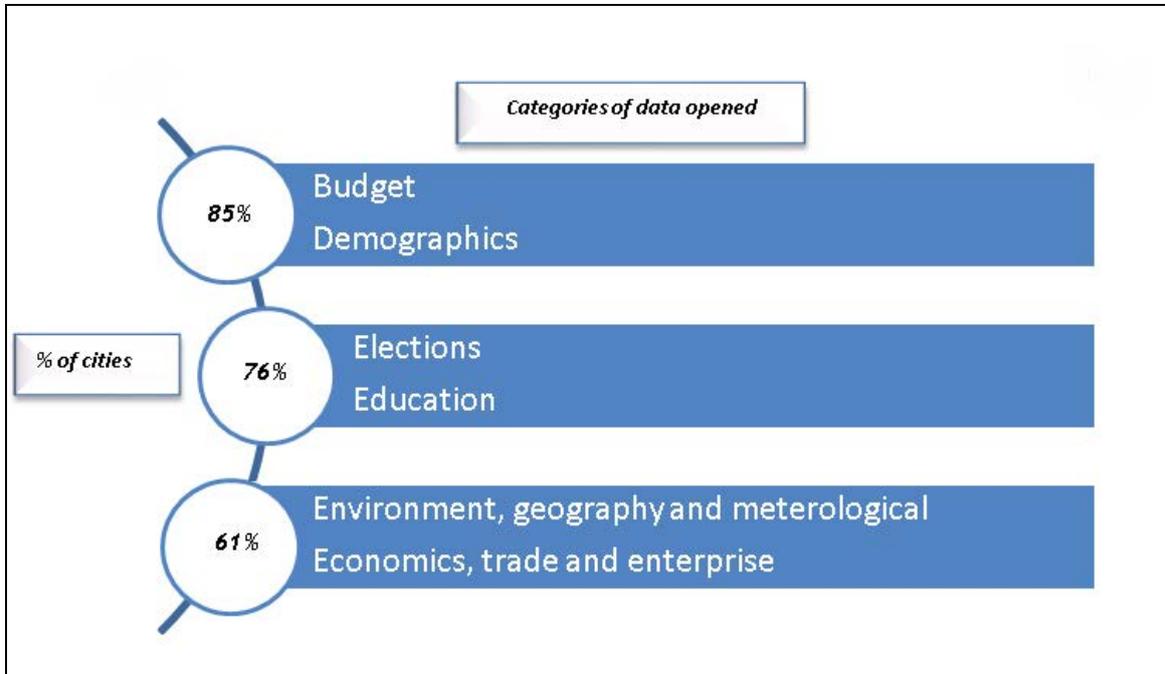
In 2011, the EUROCITIES working group on open data, in collaboration with the EU Open Cities project<sup>14</sup>, conducted a survey to examine the main reasons why cities decide to open data and what kind of data has been published. Thirteen cities responded to the survey.

The survey found that the greatest driver for open data is the aim of achieving transparency in governmental and administrative processes (100%). Likewise, of great importance for the cities is the topic of 'innovation' (92%). Reuse of datasets (77%), efficiency gains (77%) and the potential economic benefits of open data (70%) were also motives for opening public sector datasets. Only half of the cities surveyed currently view crowdsourcing or public feedback as a stimulus for open data activities (please see graphic below.)

The first datasets being opened digitally for reuse are municipal budgets, demographic information and election results. Of the cities surveyed, 85% view this as an entry point to open data. Data concerning education, the economy and labour market as well as geographic and meteorological data is also being supplied. Most cities are planning to host their datasets themselves (85%). The cities also expressed a need for support in aspects of standardisation, software tools for capturing, editing and visualising datasets, and maintaining data catalogues.



<sup>14</sup> <http://opencities.net/> - The Open Cities project is a platform of European cities, which are opening up their government data for citizens



## 4. CONCLUSIONS: BEYOND OPEN DATA, TOWARDS OPEN GOVERNMENT

The economic crisis has had a great effect on the political landscape within the European Union. The failure of governments either to foresee or to manage this economic crisis has profoundly shaken confidence in our political institutions. Citizens are increasingly cynical and disengaged from formal politics. Developing more open forms of government can help to address this loss of confidence by allowing citizens not just to engage, but also to actively contribute to the decision making process. Open government can form a new relationship between public administrations and citizens through increased transparency, accessibility of information and responsiveness to citizens needs<sup>15</sup>:

An open government implementation model was introduced by Lee and Kwak in 2011<sup>16</sup>. They describe four stages of open government:

1. increasing data transparency
2. improving open participation
3. enhancing open collaboration
4. realising ubiquitous engagement

Opening data is a first step towards open government. Making government information available increases transparency by breaking down information barriers and allowing citizens to participate in a meaningful way. This openness allows for closer collaboration with the community and increasingly involved citizens. Digital technologies act as a key enabler in this shift towards open government.

Digital technologies are allowing us to connect more widely, with increasing speed and ease. And we have seen how those technologies, particularly social media, are supporting new and different kinds of political as well as entrepreneurial activity. Government at all levels, from local to national, can embrace this change to develop more open, participatory forms of engagement with citizens.

---

<sup>15</sup> In 2005 the OECD identified three essential elements of open government - transparency, accessibility and responsiveness - [www.oecd.org/gov/pem/34455306.pdf](http://www.oecd.org/gov/pem/34455306.pdf)

<sup>16</sup> [www.businessofgovernment.org/sites/default/files/An%20Open%20Government%20Implementation%20Model.pdf](http://www.businessofgovernment.org/sites/default/files/An%20Open%20Government%20Implementation%20Model.pdf)

## GLOSSARY

The glossary is in part based material from the UK Open Data Whitepaper (OD-WHITEPAPER), licensed under the OGL (LICENCE-UK-GOV).

### **Anonymised data**

Data relating to a specific individual where the identifiers have been removed to prevent identification of that individual.

### **Application programming interface (API)**

An application programming interface (API) specifies how some software components should interact with each other. An API makes it easier to develop a program by providing all the building blocks so that programmers can write applications consistent with the operating environment.

### **App**

Short for 'application', i.e. software that a user would install or use for a specific purpose (as opposed to an operating system or a printer driver). In recent years, the term 'app' has increasingly come to mean 'mobile app' (software for mobile devices), but can also cover software on PCs or on web. In the context of open data, 'app' usually means a software that uses or visualises open data.

### **CKAN**

CKAN is a software framework developed by the Open Knowledge Foundation for data portals and catalogues on the web. It is open source and widely used by administrations.

### **Core-reference data**

Authoritative or definitive data necessary to use other information, produced by the public sector as a service in itself due to its high importance and value.

### **Copyright**

Copyright is an intellectual property right usually covering works of art and other works which involve creativity and originality in their creation.

### **Creative Commons Licences (CC)**

A family of licences widely used for works protected by copyright. CC licences exist in different forms conventionally identified by two-letter tags: attribution (BY), share-alike (SA), non-commercial (NC) and no-derivatives (ND). Combinations such as BY-SA are possible. Because CC licences up until version 3.0 are meant for copyrighted works, their use as data licences is often considered problematic. This is not the case for the CC0 rights waiver.

### **Data producer**

The department or office of the administration that creates or owns the raw data and makes it available as open data. When the data is covered by a licence, the data producer is also the licensor.

### **Data user**

The person or party that accesses and uses open data in any way (copies, re-distributes, modifies, etc.). If a licence applies, the data user is also the licensee.

### **Database rights**

Data and databases are not normally protected by copyright, since they ordinarily don't fulfil the necessary requirement of creativity and originality. In the EU, the 'Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases' ('database directive') therefore defines specifically coded ('sui generis') rules about the copying and re-distribution of databases. Database rights are independent of copyright (parts of a database might still be covered by copyright).

### **Dataset**

In open data, a dataset is a collection of data that together forms a larger whole. There is no strict definition, but the parts of a dataset would belong logically together (e.g. they are part of a time series, show the geographical distribution of a phenomenon, list companies, etc.). Ideally, a dataset will be modular and small enough to have a clearly apparent information content, which should be described in its metadata. A single dataset can have different physical representations ('resources'), such as a PDF document, an Excel table and an RDF graph all containing the same information.

### **Data portal, data catalogue**

A web portal to make public sector information accessible as open data. Typically, users can search and browse datasets, see metadata about these datasets (authors, publication dates, geographical coverage, licence, format, etc.) and follow links to the data itself.

### **Information**

Output of some process that summarises, interprets or otherwise represents data to convey meaning. The term 'information' usually subsumes data.

### **Intellectual Property**

A set of property rights that grant the right to protect the materials created by them. Intellectual property comprises copyright, designs, patents, certain confidential information and trademarks.

## Licence

A document whereby the owner of a work (the licensor) grants rights to a user (the licensee), such as the right to use, copy, modify or re-distribute the work. These rights would otherwise not be available to the licensee because of intellectual property rights restrictions, or database right restrictions.

## Linked data

Structured data where each entity in a dataset is provided with a unique identifier, to permit linking with other relevant data that might not otherwise be connected, improving discoverability and facilitating data integration. Linked data is usually expressed in the RDF data format. True linked data contains embedded links to other data.

## Metadata

Data describing other data. In open data, the metadata of a particular dataset usually covers a set of attributes such as data producer, creation date, geographical and/or temporal coverage, licence, links to resources, etc.

## Open access

Provision of free access to peer reviewed academic publications to the general public.

## Open data

Data that meets criteria for openness such as those defined by the open definition (<http://opendefinition.org/>). In addition, open data should be available via the World Wide Web, in a digital, machine readable format and interoperable with other data.

## Open data commons

A family of licences designed by the Open Knowledge Foundation specifically for data. ODC comes in three forms: attribution, share-alike and a rights public domain dedication

## Open government data

Public sector information that has been made available to the public as open data.

## Open Knowledge Foundation

The Open Knowledge Foundation (OKFN) is a non-governmental, not-for-profit organisation that is dedicated to promoting open data and open content. Among other things, the OKFN have published the open data Commons family of data licences, and the CKAN data catalogue software.

## Personal data

Data relating to a specific individual where the individual is identified or identifiable in the hands of a recipient of the data.

### **Public domain**

Works in the public domain are those whose intellectual property rights (such as copyright and database rights) have expired, have been forfeited (e.g. by use of a waiver) or are inapplicable.

### **Public sector information**

Data and information produced, collected or held by public authorities, as part of their public task.

### **Raw data**

Qualitative or quantitative statements or numbers that are assumed to be factual, and not the product of analysis or interpretation.

### **Rights waiver**

In contrast to a licence, a waiver surrenders rights that the creator or owner of a work may have. When a waiver surrenders copyright or database rights, those works are effectively put in the public domain. Examples are CC0 and ODC-PPDL.

### **Structured data**

Data that is explicitly and formally structured, so that it can be better processed, interpreted and 'understood' by software. Typical examples are spreadsheets and other tables, XML data, RDF data or JSON. In most cases, structured data is considered to be the more desirable form of open data, because it is easier to reuse and interpret.

### **Unstructured data**

Data that has no explicit, formal structure, and thus is difficult to process and interpret with software. Typical examples are raw text files, PDF documents, Word documents, image files, etc.

## ANNEX 1 - RECOMMENDED OPEN DATA FORMATS

Table and Text	Format	Geospatial Data	Format
Comma Separated Values	csv	Geography Markup Language	gml
Extensible Markup Language	xml	GPS Exchange Format	gpx
Hypertext Markup Language	HTML	Keyhole Markup Language	kml
Newsfeed/Webfeed Syndication	rss	Drawing Interchange File	dxf
Resource Description Framework	rdf	ESRI Shapefile Format	shp, shx, dbf
Java Script Object Notation	json		
Text			
Text file	txt		
Open Document	odt, ods, ..		

## ANNEX 2 - LEGISLATIVE OVERVIEW

The 2013 revision<sup>17</sup> of the European Commission Directive on the reuse of public sector information (2003/98/EC) establishes a general framework for harmonisation of national practices and regulations on the reuse of public sector documents. The revision attempts to address many of the changes that have occurred since the original directive was adopted in 2003, before open data movements really began to take hold. The directive seeks to ensure fair, proportionate and non-discriminatory conditions for the reuse of disclosed information across the EU. Its key principles are:

### Genuine right to reuse and extension of scope

The revised directive creates a right to reuse, in the sense that all legally public documents and data shall be re-usable for commercial or non-commercial purposes. The directive does not apply to documents and data that are commercially sensitive, important for national security, or that are restricted due to the access regimes of member states.

The scope of the revised directive has been extended to include museums, libraries and archives, however on a more limited basis.

<sup>17</sup> [eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0877:FIN:EN:PDF](http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0877:FIN:EN:PDF)

## Charging practices

Charges applied to data reuse will be limited to the marginal costs of reproduction, provision and dissemination. However there are exceptions for public sector bodies that need to generate revenue from the data, such as museums, libraries and archives. These bodies may charge above the marginal cost of production but should not exceed the cost of collection, production, reproduction and dissemination, (preservation and rights clearance charges for museums, libraries and archives) together with a reasonable return on investment<sup>18</sup>.

## Licensing

The revised directive does not obligate the adoption of any specific licence for the disclosure and reuse of data. However, licence conditions, if any, should be fair and transparent, and if member states attach conditions to reuse, a standard licence should be available that public sector bodies are encouraged to use.

## Formats and standards

The directive encourages the use of open standards and machine-readable formats where easily possible, but does not make it obligatory, stating that any pre-existing format and language is acceptable.

## Conditions and exclusive agreements

Public sector bodies may allow reuse without conditions or may impose conditions, where appropriate, through a licence. These conditions shall not unnecessarily restrict possibilities for reuse and shall not be used to restrict competition. Exclusive arrangements are prohibited as before, except for ensuring public interest services, or for digitisation projects by museums, libraries and archives. These types of exclusive arrangements will be subject to review.

## Redress mechanisms for citizens

Public administrations are required to inform interested parties that new datasets are available. The time limit for replying to requests for reuse of open data should be reasonable and in line with the equivalent time for requests to access the document under the relevant access regimes. If the competent authority decides to no longer make certain documents available for reuse, it should make these decisions publicly known.

The revised directive has updated redress mechanisms available to citizens. Any decision on reuse shall contain a reference to the means of redress. This includes the possibility of review by an impartial review body.

The 2003 directive has been implemented by member states. It is likely that in implementing the revised directive, member states will amend the current legislation they have in place.

---

<sup>18</sup> Any charges must be justified through transparent means, i.e. any applicable conditions and the actual amount of those charges, including the calculation basis for such charges, shall be pre-established and published, through electronic means where possible.