

Open Data, Open Society

*a research project about openness of
public data in EU local
administration*

*by Marco Fioretti
for the
Laboratory of Economics and Management
of
Scuola Superiore Sant'Anna, Pisa*

Project financed through the DIME network (Dynamics of Institutions and Markets in Europe, www.dime-eu.org) as part of DIME Work Package 6.8, coordinated by Professor Giulio Bottazzi

Table of Contents

1. Goals and structure of this report.....	3
2. Why data are important.....	4
3. Current status of Open PSI support in Europe.....	6
3.1. Austria.....	7
3.2. Finland.....	8
3.3. France.....	8
3.4. Germany.....	9
3.5. Iceland.....	10
3.6. Italy.....	10
3.7. Norway.....	11
3.8. Sweden.....	11
3.9. United Kingdom.....	11
3.10. An example of local Open Data policy from outside Europe.....	12
4. Why Open Data.....	13
4.1. Definition of Open/Linked raw data and their impact on government.....	13
4.2. The nature of data.....	15
4.3. Government transparency.....	17
4.4. Economic value of data openness.....	18
4.4.1. External value.....	18
4.4.2. Internal value.....	20
4.5. If openness is so good, why aren't all Public Data already open?.....	22
4.6. Open Data to restructure government.....	23
4.7. Why Start local.....	25
5. Real world examples of Open PSI at work.....	26
5.1. Geographical data (maps, land usage, cadaster and addresses).....	26
5.2. Local transportation.....	27
5.3. Demographics.....	29
5.4. Election support.....	31
5.5. Energy production and consumption.....	33
5.6. Budgets and taxes.....	34
5.7. Local economics activities.....	38
5.8. Real Estate.....	41
5.9. Environmental data and pollution measurements.....	43
5.10. Aggregate health, or health-related data.....	45
5.11. Security and legal.....	45
5.12. Education.....	46
5.13. Waste management.....	47
5.14. Water management.....	47
6. The social dangers of Open Data.....	49
6.1. Openness is not enough for real transparency.....	49
6.2. Optimistic assumptions about citizens data literacy.....	50
7. What to do.....	53
7.1. Political support for Open Data.....	53
7.2. Some short technical notes on data formats and software managing open PSI.....	57
7.3. Practical advice for Public Officers promoting Open Data.....	58
7.3.1. Don't worry (initially) about data quality.....	58
7.3.2. Beware of too much looking for a "business case for open data".....	58
7.3.3. Other practical advice.....	59
7.4. Citizens education.....	59
7.4.1. Education to Open Data, or Open Data for Education?.....	61
7.5. The role of the Public Sector in an age of Open Government: quality and reliability of openly generated PSI.....	61
8. Conclusions, next steps and contact information.....	64
9. Useful Open Data resources not specifically quoted in this report.....	65

Truly, nothing is so astonishing as figures, if they once get started - Mark Twain, 1897 (Following The Equator, Chapter XVII)

1. Goals and structure of this report

This report discusses the current and potential role, in a truly open society, of raw Public Sector Information (PSI) that is really open, that is fully accessible and reusable by everybody. The general characteristics of PSI and the conclusions are based on previous studies and on the analysis of current examples both from the European Union and the rest of the world.

Generation, management and usage of data constituting what is normally called PSI is a very large topic. This report only focuses on some parts of it. First of all, we only look here at really "public" PSI, that is information (from maps to *aggregate* health data) that is not tied to any single individual and whose publication, therefore, raises no privacy issues.

It is also important to distinguish between actual raw data (basic elements of information like numbers, names, dates, single geographical features like the shape of a lake, addresses...), their results (more or less complex *documents, policies, laws...*) and the procedures and chains of command followed to generate and use such results, that is to vote or, inside Public Administrations, to take or implement decisions.

So far, discussion and research on Open Data at national level has had relatively more coverage, even if much of the PSI that has the most direct impact on the life of most citizens is the one that is generated, managed and used by *local*, not central, administrations and end users (citizens, businesses or other organizations). Creation of wealth and jobs can be easier, faster and cheaper to stimulate, especially in times of economic crisis, at the local level. Finally, open access to public data is much more necessary for small businesses than for big corporations, since the latter can afford to pay for access to data anyway (and high prices of data may also *protect* them from competition from smaller companies).

For all these reasons, the main focus of this report will be on the raw data that constitute "public" PSI as defined above. This is the reason why in this report the terms "raw data" and "PSI" are practically interchangeable. We will also focus on the local dimension of Open PSI, that is raw data directly produced by, or directly relevant for, local communities (City and Regions), and on their direct impact on local government and local economy.

Chapters 2 and 3 summarize the importance of data in the modern society and some recent

developments¹ on the Open Data front in Europe. Chapter 4 explains why raw PSI should be open, while Chapter 5 shows the potential of such data with a few real world examples from several (mostly EU) countries. Chapter 6 looks at some dangers that should not be ignored when promoting Open Data and Chapter 7 proposes some general practices to follow for getting the most out of them. Some conclusions and the next phases of the project are in Chapter 8.

2. Why data are important

First of all, what are data? Borrowing from, and rearranging, a definition attempted by [Peter Murray-Rust](#) as summarized on the [Digital Curation Blog](#), by data we mean *single pieces of information of every nature (from pictures to numbers, textual definitions, maps, audio...)* that:

- *are* direct descriptions of facts (e.g. the path followed by a river, as drawable on a map, average temperatures in some city, tax brackets in some country...) or are closely related to facts, and as such are not copyrightable
- *are* reproducible without ambiguities when the method used to generate them is known in all its details. An aerial photograph is data because two identical cameras taking a shot from the same point in the same moment with the same settings would produce (to all practical purposes) the same picture, whereas oral description of the same scene by two individuals can be very different. Different people may, and very often will in the real world, produce different data to describe the same phenomena, since besides being different, they may use different methods and starting hypotheses to generate them.
- *are* parts, or can be immediately used as parts, of larger information or knowledge structures
- *have* (almost always) much more meaning and value when linked among them and completed by *metadata*. Metadata are simply data about other data, rather than about some facts. The day when a collection of digital pictures was taken would be a common metadata for all those digital pictures
- *can*, due to all the characteristics above, be expressed and stored in digital formats, even when they weren't originally generated in that form, and once digital can be processed by computers directly in those formats, to build other data, find metadata and take decisions.

An Economist [report on data](#) in February 2010 calls our age "the age of Big Data", because every year individuals, businesses and Public Administrations create (and rely on) amounts of digital data that are orders of magnitudes bigger than a few years ago. Data are digital when, whatever their

¹ as of October 2010, when the final version of this report has been delivered

nature is, they can be encoded as series of *digits*, that is bits representing ones and zeroes that can be stored in any kind of bit container, from computer hard disks to DVDs, floppy disks, SSD memory cards and so on, and can be directly transmitted in the same format, that is as sequences of bits, across all kinds of telecommunication networks.

Digital technologies have made terribly easy and cheap to generate, store and (when there's a will to do it) publish data. Quick and effective exploitation of digital data is every year more important for any organization at every level, from cost savings and transparent reporting to decision making. This is true partly because organizations must make decisions anyway, and today those decisions are based on data that are digital, and partly because digital data are so many that it's easier than ever to overlook, forget or misrepresent something. The same applies to single citizens whenever they must make important, well informed decisions, be it in the voting booth or in their work.

The same Economist Report sums the importance of data saying that they have become *"an economic raw input almost on par with capital and labour"*. The [Digital Britain Final Report](#) recognizes data as *"an innovation currency... the lifeblood of the knowledge economy"*. If all this is true, and it's hard to deny it is, giving data is like giving stimulus money, or at least sharing great lobbying power, but at a much smaller cost for taxpayers. Starting from these facts, this report looks at how much the value of data increases when they circulate and can be reused without restrictions.

How much are PSI data worth? It is hard if not impossible, for reasons that will be explained later, to give answers that are really complete, accurate and reliable. This said, here are a few numbers. According to a [MEPSIR study](#) conducted by the European Commission in 2006, the overall market size for PSI in the EU Member States and Norway was estimated at EURO 27 billions. A previous study (PIRA) had found in 2000 an 'investment value' (public sector investments in the acquisition of PSI) of EUR9.5 billions and an 'economic value' (part of national income attributable to industries and activities built on the exploitation of PSI) of EUR68 billions. Dr Rufus Pollock of Cambridge University, lead author of a UK report on the economic value of open data, has calculated that current plans to set UK government data free will create an estimated 6 billion GBP in additional value for the UK.

In Germany alone, the market for geo-information [increased](#) from EUR1 billion in 2000 (mainly from utility and engineering companies doing planning and maintenance systems) to EUR1.6 billion in 2006, with more than half the demand driven by a navigation market based on "free" private data. At about the same time, however, that is in 2007, the German government's revenue from PSI was only EUR164,000. In Denmark, open publication of the official [Danish addresses](#)

[database](#) had [direct financial benefits](#) around EUR 62 millions (~DKK 471 millions) in the period 2005-2009, with total costs until 2009 around EUR 2 millions. In 2010 it is estimated that social benefits from the agreement will be about EUR 14 millions (around 70% in the private sector), while costs will total about EUR 0.2 million.

Antoinette Graves, Office of Fair Trading OFT UK, noted in her 2009 presentation ["The Price of Everything but the Value of Nothing"](#) that:

- *"PSI is valuable and vitally important for businesses... a lot of products just could not be made, or could not be made in the form that they were, without access to and reuse of public sector information. When problems arose it was often due to public sector bodies that were doing something themselves that gave them an incentive to restrict access to the upstream level".*
- *"calculations indicated that the net value of public sector information in the United Kingdom is about 590 million GBP per year"*

3. Current status of Open PSI support in Europe

We're still at the very beginning in terms of large scale attention and usage for (public) Open Data. However, in the last year, this theme has got much more coverage than in the past and some interesting announcements have been made. Let's then try very quickly to sum up the status of open data across Europe, with a partial summary of what happened in 2010 in some European countries.

In 2003, the [EU Directive on the re-use of PSI](#) introduced a common legislative framework regulating how public sector bodies should make their information available for re-use. On the 7 May 2009 the Commission published a review of that Directive, encouraging Member States and Public sector Bodies to take proactive measures to promote reuse. In the context of the Digital Agenda for Europe, the review of the Directive has been signaled as the key action of the initiative and is foreseen for 2012.

As of July 2010, *all 27 EU Member States had [notified the Commission](#) that they had finished implementing these rules into national legal order.* In spite of this, a measure of the Economic Impact of the PSI Directive in the Context of the [2008 Review](#) showed that member states are not doing particularly well in implementing even the basic parts of the directive. The main reasons include lack of measurement tools and generally low understanding of and expertise with PSI. The

executive summary of the [MEPSIR analysis](#) *"clearly indicates that there still exist a considerable gap between the current situation and the one sought by the Directive"*.

In practice, many of the public organizations that do make the PSI that they generate available to others still do it by selling those data with more or less restrictive licenses. The reason for such a strategy is to, at least, directly recover in that way all the costs of the generation, maintenance and distribution of that PSI. This practice, however, doesn't appear so effective. Guarding the data is much more expensive than just publishing them on a server. It only makes sense if one is sure that there will always be enough users that **can** pay for those data to cover, in that way alone, both the initial costs faced to generate and maintain the data **plus** all the extra costs caused by enforcing access restrictions.

Besides, working in this way costs aren't even shared fairly among all the users of the data because, unlike what happens with fees of highways and similar services, once access to data is granted accurate metering of their usage is impossible. There are even cases where many potential users don't bother to pay simply because, thanks to the Internet, they can get the same or equivalent data for free... [from other countries](#)! For example SMHI, the Swedish Meteorological and Hydrological Institute, charges for access to weather data. As a result, (Swedish) people that needs to use Swedish weather data in their applications get them behind the corner, from the [Norwegian authorities](#).

The practical consequences are that *"recovering 1/5th of the development costs after years of sales is not uncommon and earnings for the public bodies that charge only the marginal costs [are very limited](#)".* In 2007, the [German government's revenue from PSI was only EUR 164,000](#). Graves says *"Marginal-cost pricing is not necessarily the answer. While public sector bodies may use differential pricing and recover more of their costs on certain products or users than on others, they may still restrict what is available. Moreover, when value is added, if a marginal price is charged, it is undercutting the competition."*

An even more serious fact is that, under such strategies, data are closed also to *any other government department* that may need them, [leading](#) to serious inefficiencies and duplications of efforts, even when all that would be needed is comparison of different data sets. That's why several states have started to pay more attention to the opportunities that can arise when data are opened. Here is a partial summary, in alphabetical order by country, of recent developments on this front.

3.1. Austria

In Austria (in April 2010) there have been [heated debates](#) around opening up databases from public

bodies (e.g. for farm subsidies): *"The European PSI directive from 2003 was implemented into national law as the IWG or Informationsweiterverwendungsgesetz, but a number of public bodies have violated the (actually very weak!) law by not responding to inquiries. A company providing high quality business data was even sued by the republic for collecting and using data from public databases (OGH decrees 4Ob11/07g, 4Ob35/09i, etc.). Many public bodies (don't even know) what's inside in their data silos, some of them collect equal data twice, and most of them are afraid of sharing anything."*

3.2. Finland

According to a [July 2010 report on Open Data in Finland](#) *"The general atmosphere for opening PSI is positive in Finland. Most activity in this area is connected to the implementation of the Infrastructure for Spatial Information (INSPIRE) Directive 2007/2/EC. The discussion has become more coherent and it is starting to reach (besides the civil society and the private sector) also the top level decision makers... However, progress in identifying PSI resources, opening new data sets and promoting re-use is still rather slow in Finland. A number of laws, directives and recommendations apply, from freedom of information (FOI) and the act on the criteria for charging for public sector goods and services to international recommendations and competition law. Unfortunately, while none of these laws explicitly prevents opening up and re-using of government data, current interpretation and practice doesn't support it either."*

According to the same report, the PSI directive 2003/98/EC had minimal effect in Finland because in 2005 a working group under the Ministry of Finance came to the conclusion that the existing national legislation in Finland already met the framework set out in the Directive.

3.3. France

The situation in France is under development and [changing quickly due to many influences](#). By the end of 2010, a data.gov style portal should be implemented by the French Government. There is increasing awareness by the public sector and community of the economic, political and social value of PSI.

The European PSI Directive 2003/98/EC was implemented into French law by texts very similar to the Directive: ordinance June 6th 2005 and decree December 30th 2005. On May 29th 2006, the Prime Minister's circular noted the obligations of this new law which specified the aims as economic development: the nomination of public representatives responsible for the re-use of public

information, the setting up of repositories ensuring the availability of key public sector information, the definition of standard licenses, and the analysis of licenses with exclusive rights.

APIE (the Agency for Public Intangibles of France) is working on the planning and implementation of a French PSI government data portal. Associations of citizens and non profit organizations such as Regards Citoyens and LiberTic are actively involved in open data discussions.

On the legal front, some open Licenses are available on the websites of the main French public government data producers:

- [Official Journal](#)
- [Legal information](#)
- [National Geographic Institute \(IGN\)](#)
- [National Institute for Industrial Property \(INPI\)](#)
- [Météo France](#)
- [Gas prices from the Ministry of Economy](#)

3.4. Germany

Germany implemented the PSI Directive in December 2006 with a Federal law (IWG) which has effect upon Federal authorities, Federal State authorities and municipal bodies alike. Daniel Dietrich, Chairman of the Open Data Network, [reported in April 2010](#) that the Network, a non-profit organization founded in September 2009 to promote open data, open government, transparency and citizen participation gained a lot of attention and positive feedback but the country seemed still far away from "data.gov.de" (that is having a national online portal and policy for Open Data), since local political situation, administrative structures and legislations are very different compared to the UK or the US. For example, he wrote, there is no central Office of Public Sector Information and an "Information Asset Register" simply does not exist.

A July 2010 [assessment](#) of the European and national regulatory framework impacting PSI re-use in Germany pointed out that one of the challenges for PSI re-use in Germany is to find out who has the legal competence to open up the data, since Germany is a Federation comprising 16 Federal States with great autonomy in generating, managing and publishing PSI. Data protection legislation can also close doors by being a ground upon which information requests are denied.

3.5. Iceland

According to [H. Gislason](#), the first examples of Open PSI in Iceland date to the late 1990s, when the government office [Statistics Iceland](#) concluded that their work was more valuable if openly accessible by anybody via the Internet than keeping selling access to their individual publications. This change was a success: *"today many Icelanders, from students to businessmen regularly use those data in their work"*. After the Icelandic financial system imploded in 2008 and following investigations revealed negligence by regulators and mistakes in governance, Open Data came to be seen as a high priority. More and more organizations and private sector companies have started their own efforts.

3.6. Italy

Italy has [adopted in July 2010](#) new legislation to comply with the EU rules on re-use of public data. Currently the most interesting Open Data initiative carried on by an Italian Public Administration, that is the single project with the largest scope and one coherent vision and road-map, is the [portal for open data](#) launched in 2010 by Region Piedmont, building on already existing common regional guidelines about PSI reuse. Piedmont is the only Italian region in 2010 that is explicitly moving to adopt an open license for all their currently available data ([CC0 license](#)), enabling unrestricted re-use and dissemination by anyone, even for commercial purposes.

A collection of Italian PSI data sets ([67 as of June 2010](#)) exists as an [Italian instance](#) of the Comprehensive Knowledge Archive Network (CKAN). That collection deliberately include data sets which aren't open, to help people get a "big picture" about what is available and how open it is. For example ISTAT, the national institute of statistics, put their data online for free use, but unfortunately commercial reuse is not allowed - which may inhibit the development of innovative applications and services.

On the research and advocacy front, an important initiative based in Piedmont is the [EVPSI](#) project (*Extracting Value from PSI*), whose goal is to study the status of PSI openness in order to maximize the benefits made possible by accessibility and reusability of PSI.

The [Nexa Center for Internet and Society](#), affiliated to the Politecnico di Torino in Piedmont, leads the European thematic network called [LAPSI project](#) (*Legal Aspects of PSI*). Unlike EPSI, LAPSI's goal is to find, study and overcome the current legal obstacles to PSI reuse. LAPSI will deal both with established PSI areas - such as geographic and land register data - as well as novel areas - such as cultural data from archives, libraries, and scientific information.

3.7. Norway

The EU's PSI directive was [implemented in Norwegian law](#) through changes in the Freedom of Information Act which came into force January 1, 2009. In the regulations, the Norwegian Mapping Authority has been permitted to continue its policy of charging for access to map data. Given the importance of map data for so many types of applications, the Mapping Authority's pricing regime has been heavily criticized for years.

A survey among state agencies found out that two thirds possesses data with potential for re-use that is not utilized today. In May 2010, however, the idea competition Nettskap 2.0, a Norwegian version of the Apps for Democracy contest, proved the local demand and interest for Open Data: [out of 135 applications received, 90 were based on reuse of data](#). In April a [Norwegian datastore](#) has been announced. Two urgent issues appear to be the need for country-wide standard licenses and licensing guidelines and how to face concerns that published data can be misinterpreted: in a survey of state agencies for the University of Bergen report, 43 percent of respondents agreed that *"private businesses and individuals can misunderstand data and disseminate misleading information"*.

3.8. Sweden

Swedish law 2010:566, [published in July 2010](#) implements in Sweden the European Union Directive 2003/98/EC. The law specifically purports to promote the development of an information market by facilitating re-use by individuals of documents supplied by the authorities on conditions that cannot be used to restrict competition. The website Opengov.se maintains a [registry of Swedish public datasets](#) with their formats and usage restrictions, showing what percentage of the data sets is [fully open](#), that is in open format and free for anyone to re-use and re-distribute without restriction.

3.9. United Kingdom

In this period, the United Kingdom is probably the European country where Open Data are getting the most attention from central government and major national parties. In June 2010 the data.gov.uk team [announced](#) the first meeting of a new Public Sector Transparency Board to:

- make transparency a core part of all government business
- ensuring that deadlines set for releasing the most essential and requested public data sets were respected
- setting common open data standards, according to the principle that: *"Public data policy and practice will be clearly driven by the public and businesses who want and use the data,*

including what data is released when and in what form."

Similar concepts were expressed in the [Speech on Smarter Government](#). In April 2010 Francis Maud, then Conservative shadow minister for the Cabinet Office, explained that for UK Tories [citizens are owners of their data and they will boost British jobs](#). The same concepts had been expressed in detail one month earlier in the UK [Conservative Technology Manifesto](#): *"We will create a powerful new Right to Government Data, enabling the public to request - and receive - government data sets This will ensure that the most important government data sets are released - providing a multi-billion pound boost to the UK economy. President Obama's administration has already implemented a 'Right to Data' policy. We will unleash an open data revolution..."*

3.10. An example of local Open Data policy from outside Europe

In May 2009 the City of Vancouver approved an [Open City Motion](#) stating that, *"since the total value of public data is maximized when provided for free or where necessary only a minimal cost of distribution ... and when data is shared freely, citizens are enabled to use and re-purpose it to help create a more economically vibrant and environmentally sustainable city"*, the City will:

- freely share with citizens, businesses and other jurisdictions the greatest amount of data possible while respecting privacy and security concerns;
- move as quickly as possible to adopt prevailing open standards for data, documents, maps, and other formats of media
- Identify immediate opportunities to distribute more of its data;
- Index, publish and syndicate its data to the internet using prevailing open standards, interfaces and formats;
- Develop a plan to digitize and freely distribute suitable archival data to the public;
- Ensure that data supplied to the City by third parties (developers, contractors, consultants) is delivered in a prevailing open-standard format and licensed under permissive terms that allow the data to be treated as in the public domain

4. Why Open Data

In this section we'll see in more detail the main reasons that call for making as much PSI information as possible open and linked in the sense described in the next paragraph. They are transparency, economic stimulus, savings in Public Administrations and effectiveness of non-profit organizations. The value of Open Data in education will be shortly explained later in the report.

4.1. Definition of Open/Linked raw data and their impact on government

Public data are really useful only when they are raw, really *open* and *linked*. We will now define, without going into technical details, what each of these three terms mean. Only the simultaneous presence of all these three characteristics allows to get the maximum benefits from PSI. The reason is that only when data are published online in that way every citizen or organization will be able to automatically analyze and present them in easy to understand forms like Google started doing in 2009 with its [public data search feature search](#).

Data are raw when each individual item is clearly labeled and can be immediately isolated from the others in order to be validated or reused, like the content of a single cell of a spreadsheet. Having the initial, raw data that are at the origin of some decision or action, instead of some aggregation of them, is extremely important when dealing with digital PSI. For example, publishing online in PDF format the spreadsheet containing the official budget of some city or ministry is certainly better than nothing, but it is still almost useless because those are not raw data.

A PDF file is, in fact, little more than a digital photography of the *printed* version of some document, that is of that part of its content, structure and meaning that immediately visible on screen or paper. Therefore, in the PDF version of any spreadsheet you can't see anymore the *formulas* and raw numbers and any macro or other hidden parameter that generate the final figures in the summary sheet, so you can't judge if those data and relations established among data *inside* the spreadsheet, are correct or not. In addition to that, in a PDF file you can't modify the content of some cell to verify if and how charts or totals change as a consequence of changes in the starting numbers. The consequence is that, when the native digital form of some PSI data under consideration is a spreadsheet, only the spreadsheet itself, or some equivalent version recorded in a database, could be considered "raw".

Similar consideration can be applied to any other form of PSI. Digital maps, for example, are made

of many numbers, text strings, images and more or less regular shapes (coastlines or road paths) displayed together in one coherent view. Taking a snapshot of an interactive digital map in JPEG format will yield a static picture of one of the countless messages it could have carried, and actually carries in its original form of dynamic aggregate of raw data. In the JPEG snapshot, instead, river paths, coastlines, roads, addresses, points of interests and elevations won't exist anymore as single elements that can be used and recognized independently by a computer.

That's why PSI must be made available in raw format. In all other cases the individual data cannot be reused anymore, not automatically at least!

Data are "open" when they are always published and updated online as soon and as often as possible, in a way that allows, at the lowest possible cost, to legally reuse them for free, for any purpose (including for-profit activities!) and to quick and easy *automatically* process them with any software. In practice, raw data are open when they have an open access license that allows what described in the previous sentence *and* are published in an open file format, or are directly accessible with open protocols not hindered by patents or similar restrictions, through the Internet.

Once we have open raw data, in order to make the most of them we still need (ideally in an automatic way, that is delegating all or part of the discovery and analysis work to some software program of our choice) the possibility to quickly *compare* them with other information from different sources. This need, and the related need to quickly *find* which other data may be relevant for comparison, is what leads to the concept of linked data and their [importance for Open Government](#). It is both impossible and not desirable, for economical, technical and political reasons, to have one single, huge database for all kinds of PSI. Consequently, it is necessary to facilitate as much as possible the automatic *linking, mixing and comparison* of the contents of different, independently owned and maintained online public databases. With linked data, [says the WWW inventor Director Sir Tim Berners-Lee](#) "*when you have some of it, you can find other, related, data*". This concept is also explained in the ["5 stars of open linked data" paper](#). The practical definitions that follow are, in a sense, a technical version of the "follow the money" mantra used in investigative journalism. Here is a synthesis of those definitions:

- each digital data object or resource should have a unique name and be accessible from the Internet with the same protocols used for normal web pages and services.
- data must be available in non-proprietary, structured formats that make it easy to discover them and to associate to them links to other related objects or resources

Please note that Open is not the same as Linked: all PSI that is "public" can and should be both

Linked and Open. In practice, though, it is equally possible to find Linked Data that aren't Open for licensing reasons and Open Data in formats that don't make automatic linking possible for purely technical reasons.

4.2. The nature of data

"Data is the new oil? No. Data is the new soil" David McCandless @ #TED

Understanding why it is crucial that PSI is published in raw, Open and Linked formats as defined in the previous paragraph is much easier if we look at the true nature of what we call data, that is to all the consequences of the definition attempted in Chapter 2.

In any domain, not just in the Public Sector, raw data are **only a starting point**. Just as it happens with soil, the intrinsic value of raw data, in and by themselves, is quite low, possibly lower than their cost. This happens because what really has value is what *grows* on top of those data and only thanks to their availability: the *decisions* taken by looking at data and, maybe even more, at the *connections* found among apparently unrelated data from totally independent sources. Data have value if and when they **affect decisions and change consequences**. A crucial corollary of this nature of data, that will be discussed later in this report, is the fact that, in politics, (open) data make a difference only if *enough* citizens use them as a basis to vote and participate in other public activities. The more data are used, the more they become valuable, because the amount of valuable *decisions, goods, products and services* based on them increases. The value of the data is embedded in the value of all those "products" and it is proportional to the improvements in that value versus the situation where the data were not available. In order for all this to happen, however, data must be both reliable and really open, that is freely accessible to everybody. Graves explicitly [notes that](#): *"When public sector bodies charge for PSI, those costs can actually inhibit others from adding value. The same is true with licensing restrictions"*.

In this particular moment, when many governments have already generated huge amounts of digital data but have barely started to ask themselves what openness means and whether they should bother about it, the increase in future value is much bigger for all the data that have been *already* created, maybe many years ago. Because in such cases all that remains (even if it is *not* a trivial task, of course) to create value is to open those data, that is (re)publish them in the right way. Many essential data already exist in digital format, even if not all their potential users already know it: re-generating them from scratch would be a huge waste of resources, but in some cases this is just what's happening. Maybe the best possible example of this problem is the [OpenStreetMap](#) project:

its volunteers must not only or simply *add to existing maps* data that weren't available anywhere else before. At least in some countries, they also have to spend huge amounts of time to re-create maps that already exist, that is to do *again* a job already made, probably with bigger precision and reliability, by their own governments with their own tax money. Another example, from a deeply different sector, of how much valuable PSI may just be lying in some closet, is in ["The Socioeconomic Effects of Public Sector Information on Digital Networks", 2009:](#)

*The (EU) Commission put our language resources online - gigabytes of pairs of languages from machine translations that allow translations into 23 languages. These resources, which are unique, are works of a team of, I would say, thousands of translators during many, many years. This is something for which it is **very difficult to substitute the work of private companies**... We put it on the Web, issued a press release, and had between 1,000 and 1,500 downloads of the whole data set in the first week.*

The other parts of this chapter explain these concepts in more detail by looking at several different spheres of activity, while the next chapter provides some concrete examples from several countries. Before that, however, it is important to answer a basic question: *why couldn't the public sector offer all these products and services based on PSI data by itself, regardless of the status of those data?* In other words, is opening PSI data the only way to accomplish what is described in the next paragraphs, or are there less radical alternatives?

As it will be evident from the next chapter, opening PSI data indeed is, if not the only one, by far the best solution. There are two main reasons for this statement, which are both related to the fact that raw data is like soil, therefore what is really valuable are not the raw data, but what is done *thanks* to their availability, that is their legal and technical openness and accessibility, *no matter who does it*.

First of all, no single Government apparatus (or any other single, more or less monolithic, organization) can know or figure out everything that is needed in societies as complex and interlinked as our ones. Data that may seem insignificant to the Public Administration that generated them can be valuable because they can be connected to something else, unknown to that PA, by somebody else. Besides, even if one single organization knew every way in which all PSI data can be used, it could never implement all of them by itself (even if it had the money, quite a rare condition these days). The [Open Declaration on European Public Services](#) says it clearly: *"The needs of today's society are **too complex to be met by government alone**".* This is why data have to be published with open formats and licenses, making it possible or just leaving to *all* possible end users (from public bodies to individuals and businesses) to [decide what to make with them](#).

4.3. Government transparency

The transparency in government that is achievable by opening PSI data can [reduce fraud and curb unnecessary spending](#): *"(in Canada) a \$3.2 billion tax evasion fraud was exposed when financial data was made publicly available"*. Other data can allow each voter to [measure in objective ways the distance between citizens and their representatives](#) by generating easy summaries of how they voted on a series of issues that he or she considers important. Any service of this kind, however, is possible and useful only as long as full access is granted not just to the *final* actions and decisions of those representatives (e.g. public budgets), but also to all the raw data and methods they used to come to those decisions. As the [Economist "Big Data" report](#) puts it, *"in a world of big data, correlations surface almost by themselves. Access to data creates a culture of accountability"* (maybe even more than laws punishing corruption). Transparency can also [save lives](#): *"If the inspection notes (of a mine in the United States) had been available, someone may have brought some much-needed attention to (failures and omissions in mining safety procedures) and maybe a disaster would have been averted."* In spite of all this, many citizens still [ask, or have available, far less information](#) about their representatives than they would of somebody they employ in their businesses or hire for any service.

Now, just as it happens with healthcare, prevention in government monitoring is much better and cheaper than therapy. Investigation and trials to discover and fix bribes or something else that might have happened many years before cost much more than putting every public process under thorough and *really* public scrutiny from its beginning.

Therefore, as far as making real transparency possible is concerned, the consequence is that data about public procedures, tenders and so on must become public as soon as they are generated, in formats suitable for immediate mash-ups in **one** table or diagram (cfr examples in the following chapters) that can summarize complex issues in the smallest possible space. In fact, in order to achieve concrete beneficial effects on public activities and services, transparency, or lack thereof, must be both very quick and easy to visualize.

In this context, a particularly interesting possibility and implementation of transparency only possible through Open, Linked raw data would be finding corruption (or any other anomaly for that matter, including **positive** ones like cases of excellence or innovative best practices that everybody could follow) almost in real time *automatically*, by anybody interested in doing so, with obvious beneficial effects for society. A UK citizen, for example, already proposed to [subject raw PSI data to Benfords Law](#) which states that *"in any list of numbers drawn from real life, the recurrence of digits*

from 0 to 9 follows a predictable pattern. Any deviation from this pattern would suggest that... some anomaly is occurring". Still from UK comes a similar proposal, that is [online publication of overdue tax payments from businesses](#), on the basis that "as a business owner I would like to know for free which businesses are late with those payments, by how much and how long... as this is the first indicator of their ability to pay me. I could choose whether to extend credit to them with much more knowledge than currently... to avoid me losing money and threatening the jobs of my staff when these businesses fail".

Another interesting trend or possibility in this space is crowdsourcing, that is delegation of basic tasks, from rough data analysis to entry and/or digitization data, to the crowds, that is to large numbers of casual volunteers without particular skills, but willing to contribute in any way they can to some specific cause or project. In December 2009, following the release of MP expenses documents in UK, [Simon Willison and others](#) built a web application for the [Guardian newspaper](#) that asked readers to help the newspaper dig through and categorize an enormous stack of documents - around 30,000 pages of claim forms, scanned receipts and hand-written letters, all scanned and published as PDFs, that is in absolutely non-raw and non-linked format, therefore very little useful.

The important thing in all the cases above, regardless of their feasibility, wording or the particular algorithms that should be used, is that they are *not* demands for the Public Administrations involved to do lots of extra work, that is to *add* other voices to already very tight budgets. The real request is to give all citizens (which could also analyze the data collaboratively on their own, with schemes similar to the [SETI@HOME project](#)) what they need to do the job by themselves, that is data.

4.4. Economic value of data openness

Economic value of open data falls in two very distinct categories that will be examined separately: the first is wealth generated *outside* public bodies, that is more opportunities for private businesses, job creation and so on. The other is savings *inside* the administrations themselves, because opening data makes it possible to cut some activities or handle them in more efficient manners. In this report, looking at the issue from the point of view of lawmakers and public officers evaluating when and how open the data they manage, we call the first type of value "External" and the second one "Internal".

4.4.1. External value

The first positive effects of opening PSI, specifically mentioned by the MEPSIR study, are "*more*

companies in the value chain, at various points and with more diversified products, all things that lead to increased tax revenues". Opening data as it will be shown in the next chapter makes new businesses easier to start and cheaper to operate: in such an environment they don't have to pay high fees just to access the data they need to work, or pay more than it's absolutely necessary for customization, normalization and conversion of the same data.

Costs from limited access to PSI data for the private sector also [include](#): for businesses, excessive time spent negotiating, managing and complying with licenses and indirect costs such as loss of opportunity and unfair competition; for citizens, lost job opportunities or higher taxes.

A few numbers that allow, even if in a non-systematic way, to have an idea of how much the value of open PSI may be comes from weather data in USA, which are [managed by NOAA and openly available](#): *"The underlying idea is that the information that NOAA generates has strong public good characteristics. First, it is difficult to exclude users. A 1977 study done to estimate the economic benefits of a major NOAA initiative to develop coastal and ocean observing systems estimated them at more than \$700 million annually, based on calculations of the value of information for a group of coastal and ocean-related industries—oil and gas, fishing, recreation, tourism, and two or three other large sectors."*

USA daily weather forecasts built on top of NOAA data and freely available on TVs, radio, newspapers, and online also have huge so-called non-market use benefits. Such benefits cannot be measured by multiplying prices times the quantities sold because the goods are not exchanged in a market. Instead, according to NOAA, by using state-of-the-art survey techniques and econometrics, it was estimated that there is a willingness to pay of about \$103.64 per household for the approximately 110 million households in the United States, which leads to an estimated total of \$11.4 billion in annual value (including \$3 billion in a typical hurricane season alone).

Looking at weather data in Europe, in ["Public Information wants to be free" \(2005\)](#) James Boyle estimates that Europe invests EUR9.5bn in weather data and gets approximately EUR68bn back in economic value - in everything from more efficient farming and construction decisions, to better holiday planning - a 7-fold multiplier. The United States, by contrast invests twice as much - EUR19bn - but gets back a return of EUR750bn, a 39-fold multiplier.

In her already quoted [report](#), Graves says that eliminating *"GBP20 millions from high pricing, GBP140 millions from restriction of downstream competition, and GBP360 millions from failure to exploit PSI... could lead to a doubling in the value of British PSI to around a billion pounds per year"*.

Economic benefits from data openness can happen outside governments at two other levels. Open PSI helps non-government organizations (NGOs), which constitute a significant part of GNP in EU, to understand where to work and how, both at country and city level. An already working example of the first case is the [Open Budget Index](#), which in 2008 found that 80% of the world's governments fail to provide adequate information for the public to hold them accountable for managing their money because: *"Nearly 50 percent of 85 countries provide such minimal information that they are able to hide unpopular, wasteful, and corrupt spending"*. Information of this kind, even [locally](#), can help charities to find who needs their services most and where or to boost their campaigns. Data openness also enables (foreign) investors to evaluate where to invest and how much to trust local administrations.

Finally, it is worth noting that Open Data can create business opportunities *even when not all potential customers or beneficiaries have Internet Access*: [Question Box](#), a mobile phone-based tool developed with support from the Grameen Foundation, allows Ugandans to call or message operators who have access to a database full of information on health, agriculture and education - a little like Google for people without Internet access. Such an approach could be viable even in Europe, even if the socioeconomic context and the average level of technical infrastructures are very different. In Europe, information and support services like the Question Box, which are possible only when there is unrestricted access to certain data, could be offered by NGOs and private businesses to any group of people who, due to any combination of low income, language difficulties, no familiarity with computers or lack of broadband connectivity, would not be able to use the same services by themselves: senior citizens and immigrants are just the two largest groups of potential users for this kind of services businesses that would made be possible by opening PSI.

4.4.2. Internal value

Quoting from a [2009 workshop on the socioeconomics effects of PSI](#) *"if efficiency improved in the public sector by only 1 percent as a result of free or improved access to the geospatial element of PSI, the sum saved would be the equivalent of eight times the cost to the state of collecting the data in the first place"*. As it happens with transparency, in many cases what's really great is not even the services that actually become available thanks to data openness: it's the fact that *others* did it at no cost for taxpayers. Giving away the data [saves the money](#) that otherwise should have been spent for building more or less complex websites or to provide the same services based on the same data, because **any** private businesses or group of volunteers can now offer them: ["the State brings its data and they do the rest"](#). Here are some cases of public officials explicitly mentioning that allowing

citizens to use PSI data for free saved public money:

*"[Something amazing](#) has happened in UK since the government spending recorded in the COINS database was made openly available to everyone. The impressive range of free, and in many cases open source, products to display the COINS data **beats the alternative of using public funds to pay for these tools** when the skills and enthusiasm are clearly out there in the community".*

Still in UK, the [London Datastore article](#) reports that "we simply put out an open call on Twitter for anyone interested in helping us free London's data and they have given us their time, energy and creativity in spades. The lesson? Do draw on the expertise and learning already there."

Skip Newberry, Economic Development Policy Advisor, City of Portland, OR, wrote in an email to the author that "in the case of New York and its [Big Apps](#) contest, the public investment was \$20k and the estimated return was \$4M in economic activity (\$100k per app; 40 apps created). This analysis is not terribly precise, but the point is that the citizens of NYC **received something valuable for a relatively modest investment of public dollars**".

[Click Fix in Bronx](#) is another cases where allowing citizens to enter data into an official, previously closed database lowered public expenses.

The first round of the Apps for Democracy competition in Washington DC saw [50 new software services and data analysis applications created in 30 days](#): "The city gained \$2.5m in development work outlaying just \$50,000 in prize money for the winner. The Californian government introduced a transparency website costing \$21k with \$40k annual operational costs. As a result of citizens reporting on unnecessary spending the state saved a whopping \$20m in a few short months."

The common thread in all these stories is that opening PSI often makes it possible (more on this later) to cut public expenses without cutting existing services or innovation. Savings may come from elimination of most *indirect* costs that an administration is *forced* to have when its data are closed. Djukstra, in [The business case for Open PSI](#) reports that: "the Dutch Ministry of Education finds that by providing standard information products as open PSI, the demand for specific information products declines, while the remaining specific questions are easier to answer... A lot of time is spent responding to requests for information from the public and journalists. (Opening data) reduces the time needed to deal with these requests and frees up resources where they are most needed." The same point is confirmed in a [2009 report on UK postal codes](#): "It was trivial for us to show that it costs more to restrict the use of the CodePoint database than it actually benefits the economy... the fees paid to lawyers are greater than the cost of the database license, and of the

benefits that would be paid to someone who can't find a job".

4.5. If openness is so good, why aren't all Public Data already open?

Much of the current civic activity around Open Data still happens in the conditions described in a [blog post from Mash the State](#): *the great independent civic websites using public data are mostly having to scrape and steal it. Very few councils will even acknowledge them, let alone co-operate with them.*"

Sometimes this happens because of [issues](#) which are much more general than PSI availability, from limits to freedom of speech to lack of affordable Internet connections and other physical infrastructures. Very often, however, at least in the EU, PSI data aren't available for a combination of much less serious reasons. The Danish addresses study, for example, also indicates that in the Central Business Register (CVR) and the utilities sector, usage of the official addresses is still limited due to technical, traditional and legislative barriers. Here's a summary of the most common reasons why PSI data aren't open yet:

- Pure and simple lack of real awareness about the importance and benefits of Open Data is still the norm in many government organizations (even if they should digitize all their procedures and documents anyway for their own good or to comply with some local e-Government directive, if they haven't done it yet). Side by side with ignorance, lack of explicit guidelines on data reuse from upper levels and fear to lose control are powerful motivators to do nothing, hence maintaining data locked.
- Legal barriers, or (even worst) serious confusion about the legal status of data. This happens when data come under restrictive or unclear terms of use, or simply without any terms of use at all, which is even worse. Under current legislation and international treaties, the default status of any creative work, including PSI data, is "All rights reserved" for many decades, so no re-use is possible without explicit authorization. But when data sets were produced assembling data by many different public and private bodies without a clear single policy (not an unusual case), even figuring out *who* is entitled to authorize reuse can become a costly legal procedure.
- Fear of embarrassment deriving from publishing low quality material: *"we can't publish this data, because there are errors in it"* (Zijlstra, Business case for PSI). Torkington [reports the same issue from New Zealand](#): *"serious problems exist in some data sets Sometimes corners*

were cut in gathering the data, or there's a poor chain of provenance for the data so it's impossible to figure out what's trustworthy and what's not."

- Last but not least, money. We explained that raw data are like soil: a generic foundation upon which wealth is created in many different ways which are basically impossible to predict. The "dark side" of this power is that the administrations that first see the extra money generated by Open Data almost never are the same who created and should have opened them in the first place. This makes quite difficult, for a public body without other sources of external funding and no policy imposed from the top, to see anything beyond its own real or perceived short term benefits coming from selling data, no matter if much more public money will be spent or not gained in the big picture. Even when data are already available at no charge to the public really opening them, that is deciding the proper license, getting approval for it and reformatting everything for online publication in the right formats is an extra expense that is very often perceived as not easily affordable or justifiable

4.6. Open Data to restructure government

Generation and management of PSI is related to efficient, cost-effective and transparent governance in deeper and more critical ways than those already considered. *Nations and cities are in desperate need of new ideas* ([J. A. Smith, Shareable Futures](#)). There is a need to rethink and review public services, to understand if and when there's still a need for them and if, when and how *Public Administrations and citizens can work together*. In such a scenario, opening PSI data can make the deep changes that will or should happen anyway in the next decades, happen in a less painful and possibly much more efficient manner.

Thanks to Open Data, and to computers in general, today it is possible, if not already necessary, to [move away](#) from the "vending machine model" in which all citizens get from government the same one set of *automated*, absolutely impersonal services, towards a model where citizens really participate because *they can finally do part of the job themselves as THEY need, with as little intermediaries as possible*. In the speech on [Building Britain's Digital Future](#), given on 22 March 2010, the UK Prime Minister said that *"Open Data transform not just the way services are delivered but, more importantly, allow citizens to control those services."* More than reducing government's role, Open Data can improve its public services and decision-making processes through real participation, competition and load sharing. There are many experts and "professional amateurs" who would never get into politics, but could contribute effectively to the transition away from the vending machine model.

Here is the reason why it makes sense to open the whole process of making use of PSI now that it is technically possible to do it. By definition, official public websites can only offer the ways in which the administration who owns them wants to interact with the public. In many cases that same administration will have to spend extra money to make its data and operations known to all citizens. But in the real world, citizens often don't know who's responsible for getting something done, *nor do they care*. Once systems like those presented in the next chapter become available to all citizens, people will have, much more than before, all the elements they need to form their opinions, plus public services available in ways matching their *real* needs, without wasting energies to understand the bureaucracy and unwritten customs of many independent offices and fight them.

This scenario has been described saying that "Open data allows software programs and services [to be designed by people for people](#)". Surely there is a lot of idealism in such a vision, but there is no doubt that it is also a very pragmatical, if not cynical one. Opening data quickly may be a very effective way, for any administration with a budget deficit, to cut on public expenses without greatly reducing the availability, for all citizens, of efficient and affordable public services, as well as the opportunity of more job or civic participation opportunities. Once many people and independent businesses can "play" with PSI to offer the same services, *under public, possibly real-time scrutiny from everybody else*, there are less expenses and reputational risk for the public sector, because it's much easier to have somebody else doing the jobs that are possible only having access to PSI data.

It is crucial to understand the difference between this kind of "restructuring" or transformation and the privatization/deregulation policies of the last decades. All too often, deregulation has turned up to just be the transfer of a service from an initial monopoly (by some State or local Government) to another monopoly or oligopoly ran by very large, private, for-profit companies. Opening data, instead, means making publicly available to everybody, for free and for any purpose, all the PSI needed to run that service at the smallest possible cost. This not only allows anybody to run that service. It also (and above all) makes it much easier for *everybody else*, from public officers and other single citizens to competitors, to verify in any moment if that service is offered in the best possible way. Unlike old-style deregulation, Open Data means engaging with, and trusting, all citizens to participate in the offering, management and control of public interest services, while spending the smallest possible amount of public money.

The conclusion is that the social and political costs of limiting access to PSI can only grow. Today, very often the main point is not how a Public Administration should build and run *by itself* the best possible online service and websites, but how to make it really possible for everybody with the right

skills to do the same or control the quality of the several services.

4.7. Why Start local

If we accept that data openness is good, the next question becomes where and how to start opening data. Intervention from above is necessary (see next chapters) in order to make the whole process happen in the fastest and most efficient way. This said, our observation is that opening PSI can bring very good results even if it only happens, at least initially, at the local level. Sometimes the reason is that this simply is the only way to go. In federal states like Germany, for example, many of the official registers so important for PSI are run on an entirely local basis. So for instance there are over 5,000 Population Registration Register. The Cadaster is also essentially a local responsibility.

In any case, opening and using PSI in cities or regions is the best way to stimulate local businesses as soon as possible and also to educate and engage citizens. Incentives for citizens to use Open Data, that is in analyzing PSI, reusing it and contributing to government, may be much greater and easier to achieve at the local level than at national or super-national levels. Starting local, but as soon as possible, can also be the best way to experiment cheaply, before expanding some initiatives at a national level. These assumptions are strengthened by the Communication ["A Digital Agenda for Europe"](#), Brussels, 19.05.2010, which states that the success of the Digital Agenda will require a sustained level of commitment *also at the regional level*, and by the ["Regional Dimension of Open PSI" article](#):

"Regional and sub-regional entities and Public Administrations are repositories of massive amounts of data, some of which produced or mostly useful and relevant locally. may be more detailed and more up to date, particularly if the institutional design gives responsibility to local institutions".

5. Real world examples of Open PSI at work

This section of the report describes the main characteristics and potential usage of some of the most useful categories of raw PSI. Whenever available, we also present one or more success stories of useful services and profitable local businesses made possible just by the openness of those data.

5.1. Geographical data (maps, land usage, cadaster and addresses)

Who draws and controls the maps controls how other people see the world. Mapping Hacks, O'Reilly.

Geographic data from elevations to addresses are "[*the first metaphor with which to represent reality*](#)." They are essential analysis and decision tools. Erroneous or incomplete geographic data lead to inefficiencies, errors and, in cases like an ambulance not finding injured people, loss of lives. Geographical data also describe and detect position and status of hydro-geological risks as well as protected areas, wildlife, fishery and forestry resources. Their importance has been already discussed in several papers and websites, from [Free Our \(spatial\) Data](#) to a [Canadian Government sponsored study](#). Geographical data are particularly important also because they add very important context and meaning, that is much value, to practically any other kind of PSI. Knowing that in some province the occurrence of some disease is higher than average is good. *Seeing* on a map that all cases happened very close to some particular type of soil or industrial facilities, that is *connecting* just through their location two otherwise unrelated groups of raw data, is much better. Especially when we think that, if all data are open, such linking can be made quickly and automatically via software that *everybody* with the right skills could write and use!

A first form of collaborative civic service based just on geographic data is [Open311](#), a standardized technology usually adopted to report, track and fix problems in public spaces and infrastructures like potholes, broken streetlights, garbage or vandalism. When somebody enters photos and description of some problem occurring at a given location, the report is automatically assigned to the public department that should fix it, while the status of the problem is continuously updated online, to monitor the effectiveness of that department.

Until 2002 in Denmark, the official address database was practically inaccessible: users had to make an agreement on prices and terms with each municipality. The bureaucracy was complex enough that some companies had developed alternative collections of addresses *even though the*

public sector already had the best possible data. Following an agreement in that year, everybody could order municipal address data via a [public server](#) by just paying distribution costs. A [study](#) performed in spring 2010 concluded that the direct benefits of free-of-charge data for the five years 2005-2009 can be estimated at about EUR 62 million and the total value of all distributed address data sets can be calculated to EUR 76 million. This figure includes the savings for private enterprises and municipalities made possible by not having anymore to negotiate, license and delivery data between them (municipalities' savings were calculated at about EUR 5 million from 2005-2009). The analysis doesn't include the supplementary financial benefits arising from more efficient emergency services and simplified managements of one data collection, with no more duplicates.

[Goolzoom](#) is a profitable local business built in Spain just on the free availability of (mainly geographical) PSI. Goolzoom started in December 2006 as a research tool to help people looking for a home, or land management, agriculture and real estate professionals to get many information about land parcel or specific buildings in one view. As of June 2010, Goolzoom not only displays cadastral or Google maps, but can integrate them in one view with about 200 other different kind of maps, published by local or central public administrations and all accessible online through the Web Map Service (WMS) standard. Building Goolzoom was made possible by the Inspire European directive, that recommends public administrations to make the data available using common standards. The Goolzoom business model is based on premium access (printable maps, export maps to image format, and brochures with different maps for a single place) and advertising for casual users. In the first months of 2010 Goolzoom had around 250.000 visits /month, 120.000 absolute unique users and 80.000 expected billing in 2010.

5.2. Local transportation

Productivity and other losses caused by car traffic amount to [40 billions Euros per year in Italy alone](#). Still in Italy, [time wasted for the same reason](#) amounts each year to 240 hours per person in Milan, 210 in Naples and 260 in Rome. Besides saving time and reducing stress, that is increasing productivity, public transit is also the single most effective way to cut one's contribution to carbon dioxide pollution.

Many people already know this, but don't use public transportation because it is, or is considered, much more unreliable when planning even a short trip in the city than using a private car. Having correct, real time information about how much time and money it will take to go somewhere with

public buses, taxis and trains or how much time one should spend waiting at some bus stop is a big, very important support and stimulus to *use* public transportation more. The practical consequences of having this information go even further than that. Knowing how long one will have to wait till the next bus comes can mean realizing that you may have time to take a coffee or buy something at the street corner, and it would be particularly useful for citizens with reduced mobility.

Knowing that their potential customers can have such information straight from the sources in real time on their smartphones is also good for shop owners, especially those in historical neighborhoods: if people can rely on such services they will be encouraged to come shopping with buses, therefore reducing merchants opposition to traffic and parking restrictions in the same areas. Finally, especially when linked with those about city budgets and/or pollution levels, these data can raise awareness of all the financial and energy-saving benefits of using public transportation.

Applications that provide real time information on local transportation are already available in several parts of the world, even if each of them is probably already used by only a small percentage of the people that could benefit from it. European examples include the [Helsinki Journey Planner](#), the [Kèolis portal in Rennes, France](#), and the Spanish ["Donde en Zaragoza"](#).

Kèolis, after only three weeks [from its opening](#) on March 1st 2010 had already given birth to 5 applications exploiting its original data, visible at [Levelostar](#). "Donde en Zaragoza" allows iPhone users to know where is the closest bus stop or wi-fi hotspots. As soon as more PSI data sets become freely available, the application will also signal libraries, ATM machines, pharmacies, parks and other public interest services. In the USA, Seattle has the [OneBusAway](#) Open Source Tool to find real-time transit and arrival information for trains and buses in the Puget Sound region.

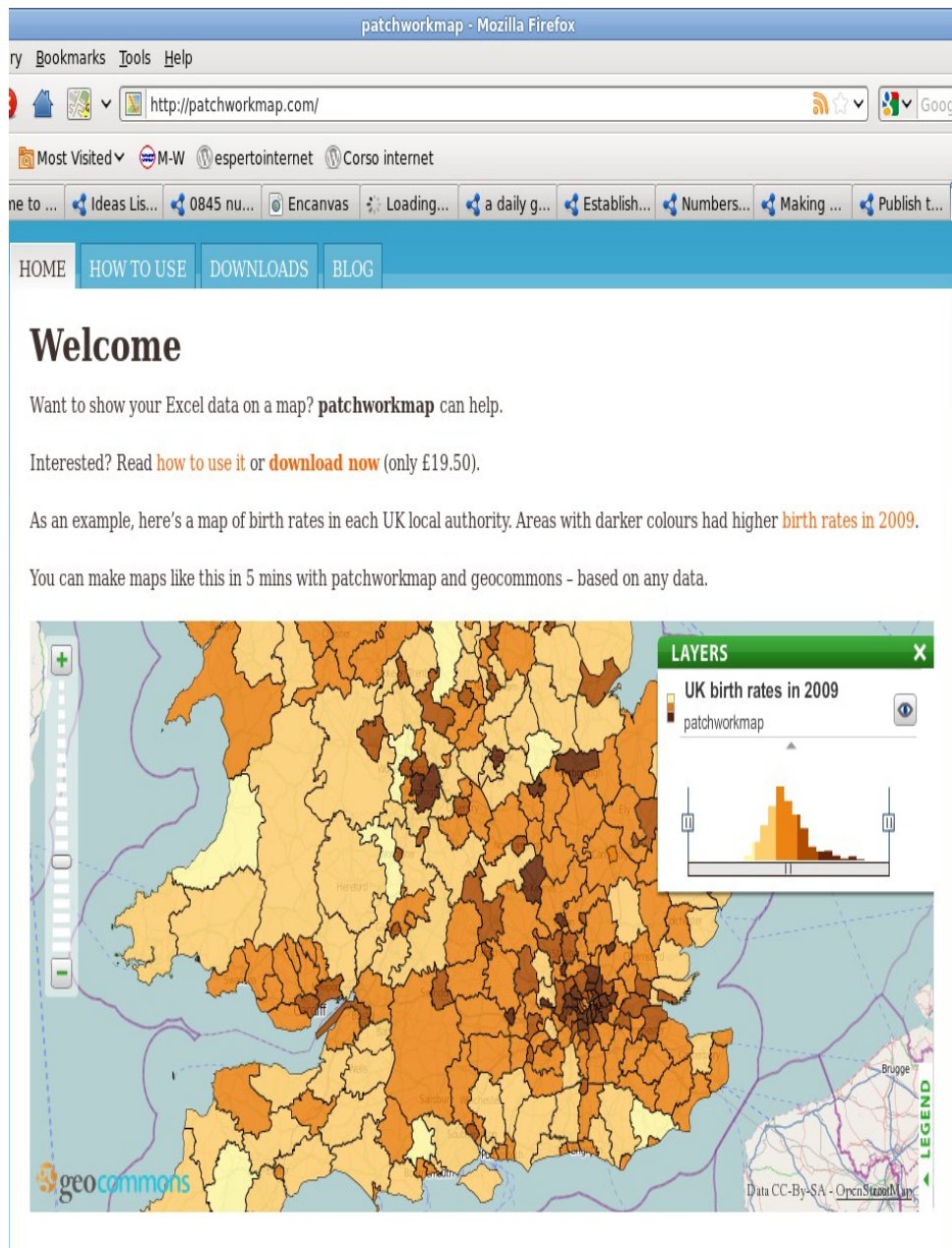
[Rodalía](#) provides real time information about schedules and status of local trains in the Barcelona area, collecting and displaying on its page, both in text format and on maps, official announces and accidents/status reports sent by train users via Twitter. Information of the second type is especially important at rush hours, while in other moments the most relevant contributions are sourced by official websites. The service became very popular from its very beginning, thanks to messages sent via Twitter and lots of media coverage. Initially the Government of Catalonia reacted to Rodalia.info negatively, because they had started their own similar service at the same time. However, when they saw the quality and speed of Rodalia (sometimes users add information to the website via Twitter before the administration itself adds it to the official web page) they changed attitude and started to support it. Initially, for example, the official website was not using open Web standards like RSS for announcing news. A few weeks after Rodalia administrators (in order to

simplify their own work) asked to switch to RSS and to classify incidents by lines, the administration did it. In other words, the existence of Rodalia (which is possible thanks to free use of PSI like train timetables) offers a public service at no cost for taxpayers and its competition has improved the quality of the official portal. According to Rodalia manager Roger Melcior: *"We make money with Google ads, but our main result is that we have changed the way the administration works: we set the agenda for us"*.

Open PSI related to transportation, roads or train networks isn't only useful for people that travel, or should travel, with public transportation. The data.gov.uk portal hosts an interesting [proposal](#) that shows how many different but always useful ways there could be to combine data when they are available: an addition to car GPS navigators like Tomtom and Garmin to inform drivers when they approach a road with a history of fatalities and casualties, so they could slow down and pay even more attention than usual. Tim Berners Lee [described in an interview](#) a case where something very similar has been actually done by several independent programmers in less than 48 hours: a map showing all the bike accidents within the last three years so bikers, so *"you can find your journey to work and maybe modify it to take another route, or put pressure on the government to deal with dangerous spots"*. Finally, in [Warwickshire](#) a web based application displays on Google maps the official height and weight restriction data for local bridges, helping freelance truck drivers and freight agencies to plan their trips.

5.3. Demographics

Data like population composition by age ranges, sex, birth or death rates, number of permanent or temporary residents and their schooling levels are useful to every public administration or private business that needs or want to offer some service to that population. Organizations aren't the only potential users of demographic PSI, however. Full access to demographic projections can, for example, help citizens to assess by themselves if and how much a City Council decision to (not) build in their neighborhood more hospitals, kinder-gardens, subway lines, parking lots or schools actually is in their interest or not. Even deciding *where* to start a certain business (think kinder-garden or shops selling clothes and gear for children or teenagers) benefits from wide availability of such data. An example of how they can be visualized for easier understanding of large scale trends is in the two following screenshots, that show how the website patchworkmap.com allows the user to select many kinds of data (birth rates in UK in this case) and display them on a map:



5.4. Election support

Participation to election is decreasing in several EU countries. Among the many causes for this trend there is the fact that many citizens find too difficult, or have lost interest for whatever reason, to know the candidates to each post, or don't trust the "canonical", top-down channels anymore, be they mainstream TV shows or even the official party websites of the candidates themselves. Online forums and social networks, or even "general-purpose" portals like Wikipedia haven't had much success so far in satisfying the need among voters for information on political candidates that is complete, relevant, reliable and easy to browse. A number of independent online initiatives has

started in the last years to fill this gap.

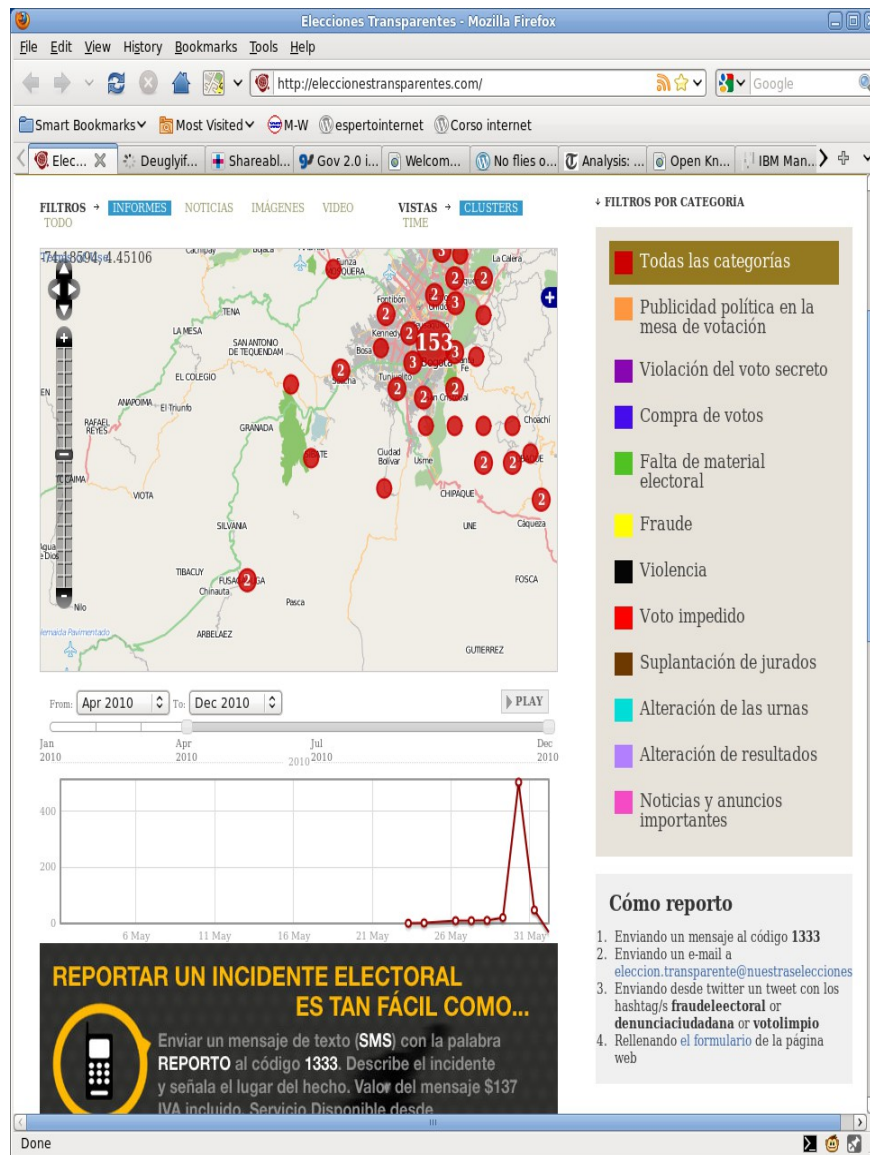
One of the most recent examples in this field is [Your Next MP](#) from UK, which is currently still focused on the candidates of the England 2010 general election. The [Straight Choice](#) group crowd sourced in UK 5173 election leaflets, from all parties and most constituencies. You can see a zoomable map of them, and a mosaic of the party leaders made of their leaflets, in this [blog post](#) where they report back on what they've found.

In Italy, the [OpenParlamento](#) group regularly scrapes data from the official government websites to build searchable databases that show how much each Parliament member is active, how he or she has voted on each issue (including the times where the vote was against the official party line) what is the status of each law proposal and other information of this kind.

The two examples above are particularly significative because they both show the usefulness of PSI data and how much effort must be duplicated without real needs when those data aren't open. What OpenParlamento and Straight Choice are doing is based on much manual work on their side that is not really necessary. All the services at OpenParlamento would be much easier to implement if it were possible to directly query through the Internet the official databases on which the Parliament websites are built. As a matter of fact, if those databases (that **must** be built and maintained anyway for official record keeping and internal operations of the Parliaments) were directly accessible through the Internet, a relatively standard procedure to set up, there wouldn't even remain much of a need to spend public money to maintain the official websites from which that same information has to be scraped afterwards!

Helping citizens to compare election leaflets could be even simpler for central governments. All it would take is one law mandating that all candidates in any post publish online all their leaflets with an open license, in a format that makes it possible to find, download and compare them automatically.

Besides the information that helps voters to know all they want to know to decide who to vote, public data about elections include those that help to see if voting happened regularly. This is the case of portals like [eleccionestransparentes.com](#), that collected and displayed on one map all kind of accidents, reported by the authorities or by citizens via SMS, Twitter or email, that could alter the result of Colombian elections in May 2010: votes not secrets, frauds, lack of voting material or booths, violence against voters, advertising in the voting booth and so on.



5.5. Energy production and consumption

Energy PSI is easy to define: these are the sets of data that would show, with at least daily updates, how much energy is used by a whole community and by all its public structures and offices and at which hours, together with information that explains where it was produced, by which sources (coal, oil, solar, wind...) and at which cost. Having this information constantly updated on public websites would allow everybody to build dynamic graphs and table that display supply and demand of electricity, highlight wastages or show which areas are more dependent from energy coming from other areas. Another reason to open up the data from electricity suppliers would be to justify and make more acceptable for their customers migrations to time-sensitive tariffs or other programs meant to reduce energy wastes.

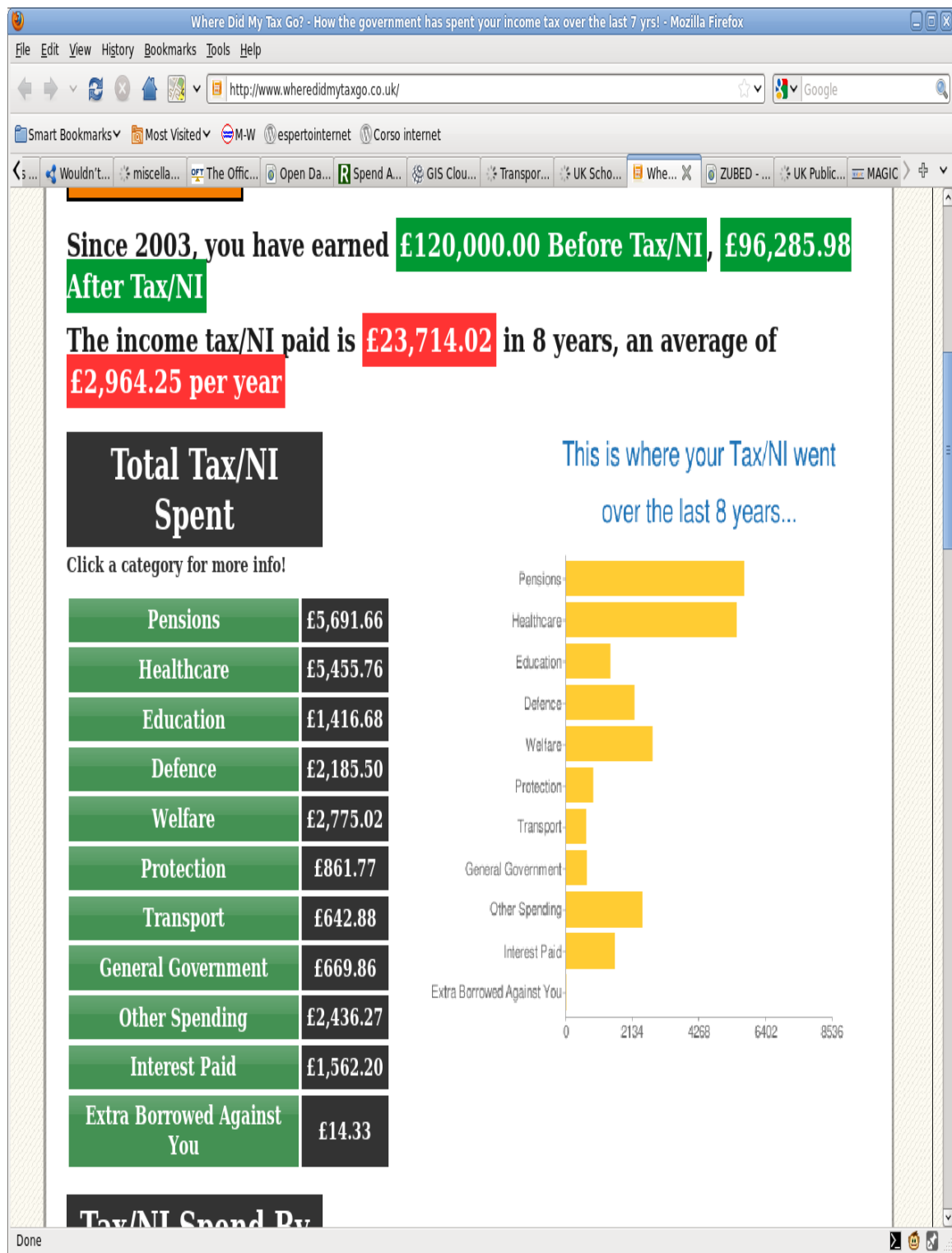
5.6. Budgets and taxes

Availability in raw, open formats of full budgets of both local governments and of all local public companies (including names and costs of contractors and consultants) as well as that of some tax-related information, can have two very important effects: one is obviously to prevent corruption or simply wastes of public money, spotting any symptom as soon as possible. Complete transparency of salaries for local administrators in the past years could have also avoided the "statewide outrage" caused by discovering only in 2010, and only through a Los Angeles Times investigation, that most council members of the city of [Bell, California](#), were paid nearly \$100,000 per year for part-time jobs. Open Data web services related to financial control can also be built ad-hoc to monitor specific topics, not just whole budgets. This is the spirit with which [StimulusWatch.org](#) was built in the USA: *"to help the administration keep its pledge to invest stimulus money smartly... by allowing citizens around the country with local knowledge about the stimulus projects in your city, to find, discuss and rate those projects"*. Similarly, the [NYCStat Stimulus Tracker](#) only tracks the City's use of federal stimulus/recovery funds provided through the American Recovery and Reinvestment Act of 2009. The important thing is that if the basic data are open, *what to do with them*, that is how to select and correlate financial raw data and how to correlate the result, is something that can be decided almost on the spot, without asking any time to the data maintainers to reformat them in any way.

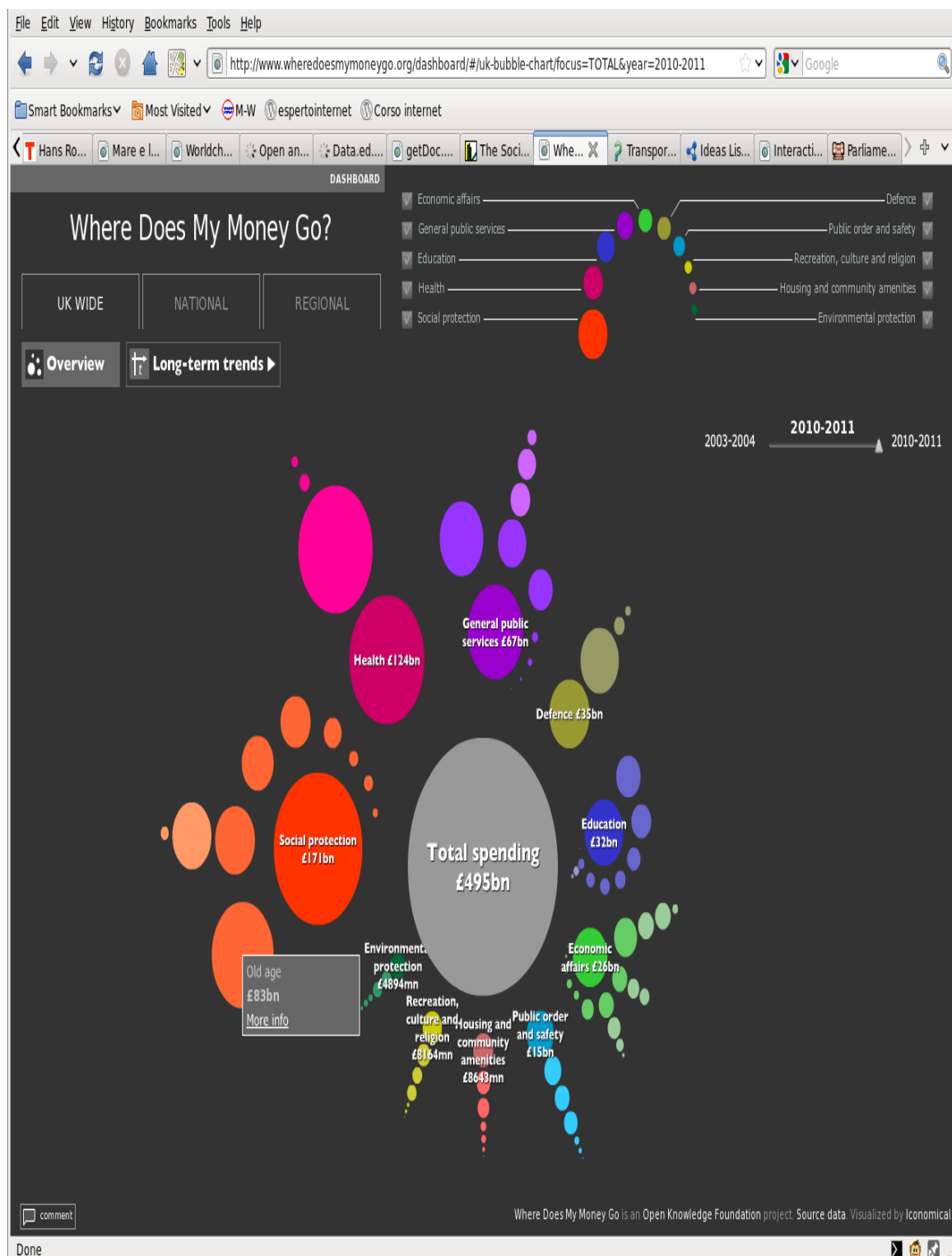
Another important benefit of completely opening financial PSI data is somewhat the opposite of spotting dishonest or incompetent administrators, or preventing them to make mistakes in good or bad faith. Raw, open budget data can help citizens to (re)gain more trust in their administrators, recognizing which ones are doing a good job and supporting them.

A simple application that, looking at official statistics published online, helps UK citizens to understand what their taxes are for is [Where did my tax go?](#), which is based on Public Expenditure Statistical Analyses (PESA) published by HM Treasury. The website consists of a web form in which the user must only enter Gross Income in each of the last seven tax years, current age and sex. The result is diagrams and tables showing total taxes due for each of those years and how much of the total was used for pensions, healthcare, education and other major budget voices.

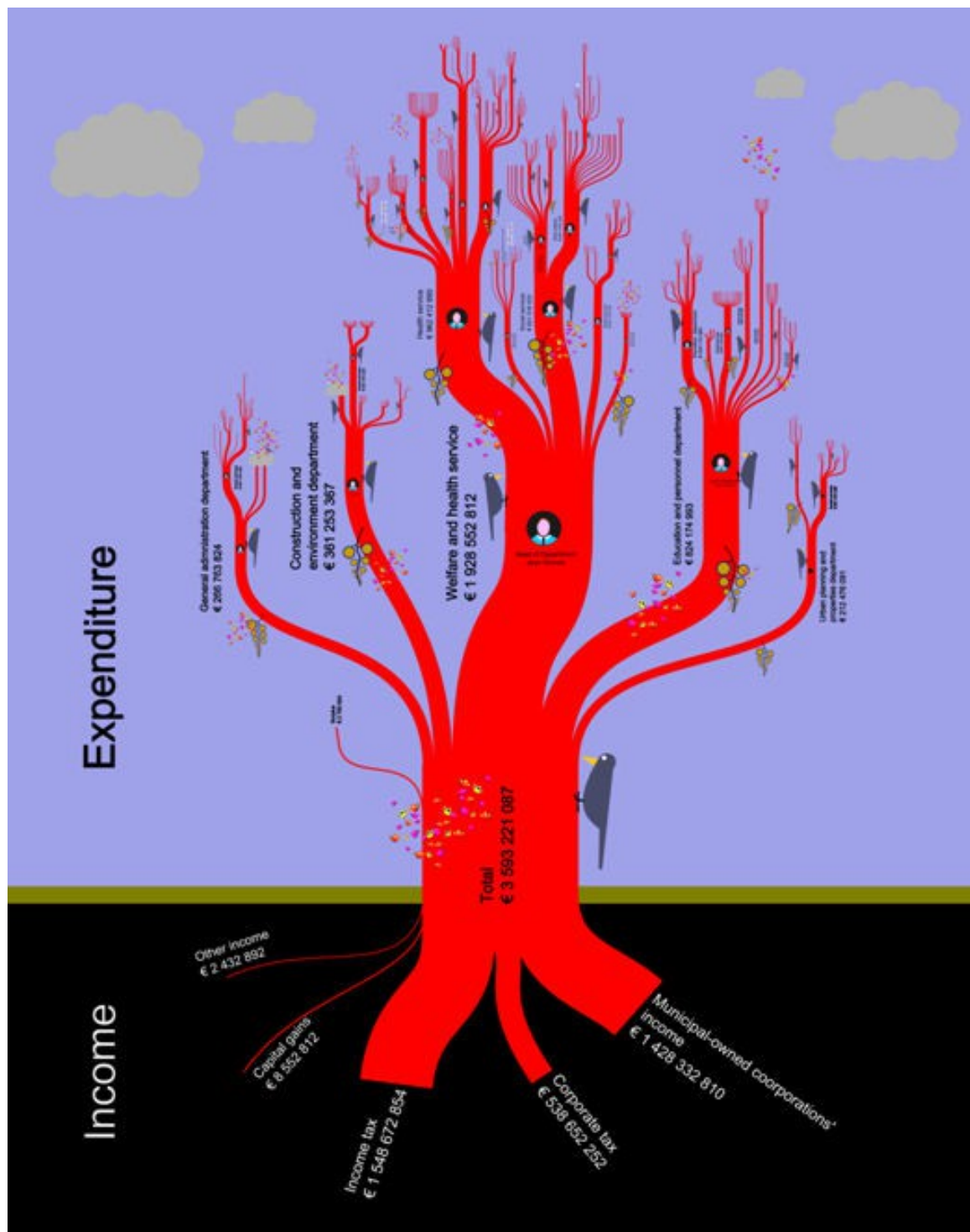
Here's an example of "Where did my tax go" calculation:



Another website offering the same services, but visualizing the same data in a different way, is Where Does My Money Go, shown in the following screenshot:

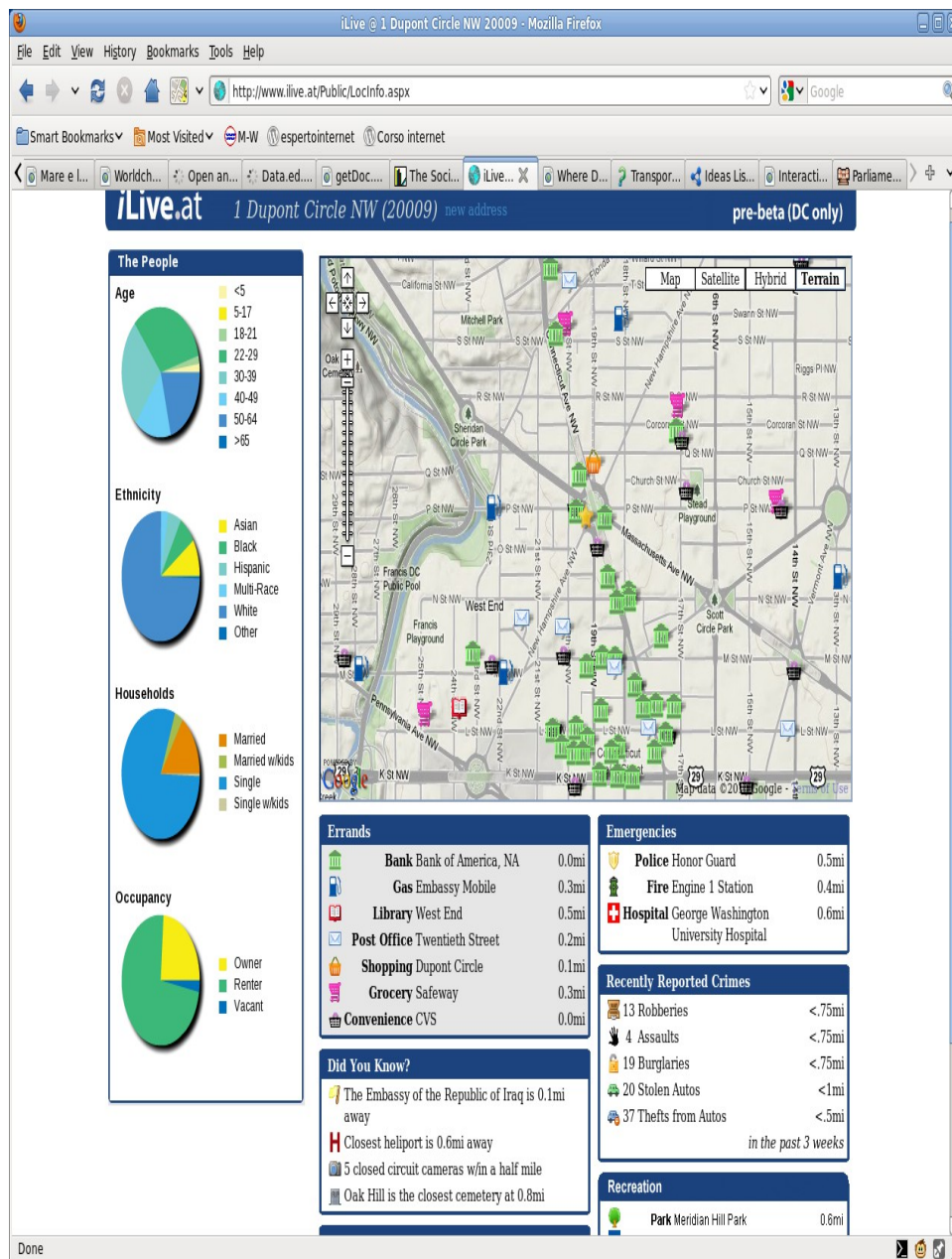


The Tax Tree shows the same kind of data, but visually combined in another different way, that may be easier to understand or simply more interesting to look at for many users:



5.7. Local economics activities

PSI regarding local economics activities refers to all data about local businesses: their number, location, activity sector, contact info, opening hours and possibly tax and other financial or labour-related information like numbers of employees. The simplest applications of these data, which are also the ones that are most immediately useful to the greatest number of citizens, are mashups that display all of them in one view, normally on a digital map, so that users can find location and information of the closest business and public services in each category. The following picture shows how the website www.ilive.at formats these data, merging them with local crime and other demographic statistics:



In many countries, services like these are already available on the websites of some telecom operators, since they may be considered as a digital, enhanced version of the traditional Yellow Pages. Making the raw data open allows everybody to mix and mash the data in all the ways that all end users may find useful. As we explained at the beginning of this report, what matters are the *connections* of the data, or even if and how they change over time. Therefore, only if all the underlying raw data are open it is possible to have them mixed time and again until the combination that the end users actually find more useful (at no cost for taxpayers!).

Viewing the PSI related to already existing local economic activities and services is useful to save time, stress and money, or to start and run other local businesses in the most effective way. The other usage of this category of PSI is to monitor *future* local developments, in order to spot inefficiencies or possibilities for corruption before it's too late, and to participate in the development of one's community. This is possible by releasing as Open Data all the PSI that makes possible services like [Planning Alerts](#) or data.seattle.gov. The first website regularly searches as many local authority planning websites as it can find, to email users the details of land development projects. The second one can, among other things, list all the building permits requested in a given part of the city, as shown in the following picture.

The screenshot shows the 'Active Building Permits' page on data.seattle.gov. The page includes a search bar, filters, and a table of permits. The table has columns for Title, Permit Type, Address, Description, Permit and Complaint Status URL, and Latitude. Below is a sample of the data from the table:

Title	Permit Type	Address	Description	Permit and Complaint Status URL	Latitude
Project 0001065 - Permit Issued as of 07/15/2012	Mechanical	12034 PINEHURST WAY	INSTALL FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.7167581
Project 0001070 - Permit Issued as of 07/15/2012	Mechanical	2507 MENOR AVE E	REPLACE 90,000 BTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6422447
Project 0001074 - Permit Issued as of 07/12/2010	Mechanical	4521 40TH AVE S	INSTALL NEW GAS FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5622833
Project 0001091 - Permit Issued as of 01/09/2007	Mechanical	8304 40TH AVE S	REPLACE 80 BTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5208065
Project 0001095 - Permit Issued as of 07/27/2009	Mechanical	3029 50TH AVE SW	REPLACE FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5760349
Project 0001104 - Permit Issued as of 07/16/2012	Mechanical	2338 41ST AVE E	INSTALL OIL FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6402065
Project 0001111 - Permit Issued as of 07/27/2009	Mechanical	2216 S LANDER ST	INSTALLING GAS FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5793891
Project 0001136 - Permit Issued as of 07/17/2006	Mechanical	11312 5TH AVE NE	FURNACE REPLACEMENT	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.7109716
Project 0001275 - Permit Issued as of 07/18/2012	Mechanical	4007 WOODLAND PARK	REMOVE OIL FURNACE AND INSTALL GAS FURNACE. BASI	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6556475
Project 0001320 - Permit Issued as of 07/27/2009	Mechanical	3433 ALAMO PL S	INSTALLING GAS FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5710636
Project 0001323 - Permit Issued as of 02/03/2010	Mechanical	359 WARD ST	INSTALLING GAS FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6278395
Project 0001340 - Permit Issued as of 02/03/2010	Mechanical	11709 SKIRRELHINE PARK	REPLACE ELECTRIC FURNACE WITH GAS	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.7139112
Project 0001351 - Permit Issued as of 07/19/2012	Mechanical	4228 1ST AVE NE	REPLACE OIL FURNACE WITH GAS	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6564658
Project 0001358 - Permit Issued as of 07/17/2006	Mechanical	2209 N 42ND ST	FURNACE & DUCTWORK	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6577927
Project 0001450 - Permit Issued as of 07/27/2009	Mechanical	5253 44TH AVE SW	INSTALL FIREPLACE INSERT (COSMETIC)	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5510351
Project 0001459 - Permit Issued as of 07/27/2009	Mechanical	4113 40TH AVE SW	REPLACE 60,000 BTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5661975
Project 0001463 - Permit Issued as of 02/03/2010	Mechanical	12740 6TH AVE NW	REPLACE 100 MBTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.7248222
Project 0001465 - Permit Issued as of 07/27/2009	Mechanical	3720 WALNUT AVE SW	REPLACE 80000 BTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5703342
Project 0001480 - Permit Issued as of 02/03/2010	Mechanical	2857 34TH AVE W	REPLACE 120000 BTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5666636
Project 0001564 - Permit Issued as of 01/01/2003	Mechanical	2600 FAIRVIEW AVE E	REPLACE 80 MBTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6437388
Project 0001565 - Permit Issued as of 01/09/2007	Mechanical	4906 S ORCAS ST	INSTALL 50000 BTU FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.5513333
Project 0001640 - Permit Issued as of 02/03/2010	Mechanical	7359 DOBBLE AVE NW	INSTALL GAS FURNACE IN NEW CONSTRUCTION	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6831251
Project 0001726 - Permit Issued as of 07/28/2012	Mechanical	812 NE 55TH ST	REPLACE GAS FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6688555
Project 0001772 - Permit Issued as of 07/03/2010	Mechanical	536 N 71ST ST	REPLACE GAS FURNACE	http://web1.seattle.gov/gis/PermitStatus/project.aspx?id=400	47.6855604

As usual, there's no intrinsic, technical limit to the amount of local economic PSI that can be mixed to give a very quick but effective representation of the status of a community or answer some specific question, in order to give all its residents to take informed decisions about voting, working or making the best of their own free time. The "This We Know" portal in the USA creates for all these purposes summaries of business, demographic, health and environmental statistics.

The screenshot shows a Mozilla Firefox browser window displaying the 'This We Know' website for Mountain View, CA. The browser's address bar shows the URL http://www.thisweknow.org/CA/Mountain_View. The website has a search bar with 'Mountain View, CA' entered and a 'SEARCH' button. Below the search bar, the page title is 'Mountain View, CA'. The main content area lists several statistics, each with a green circle icon and a 'tweet this' link:

- There are **2 Factories** (within 1 mi.) [tweet this](#)
- 781,441** pounds of **38 Pollutants** were released (within 1 mi.) [tweet this](#)
- 2 Officials** reported on **2 Factories** (within 1 mi.) [tweet this](#)
- 238 Violent Crimes** occurred or **3.4** per 1000 **people** (in this town) [tweet this](#)
- 56,733** people are **Unemployed**, while **474,094** have jobs (in this county) [tweet this](#)
- There are **127,889 Home Owners** and **51,247 Renters** (in this town) [tweet this](#)
- 46%** of people **Relocated** in the past 15 years (in this town) [tweet this](#)
- 4,933** people were diagnosed with **Cancer** (in this county) [tweet this](#)
- 1 Bill** have been introduced about this location by **1 Member of Congress** since 1993 [tweet this](#)

Below the statistics, a tip states: 'Tip: Click on any of the highlighted items above to explore the underlying data.' The sources listed are: '2005 Toxics Release Inventory * 2007 Crime in the United States * Local Area Unemployment Statistics * US Census 2000 * Cancer Incidence - SEER Registries * GovTrack'. To the right of the statistics is a map of Mountain View, CA, showing the city's location relative to surrounding areas like Martinez and Pacheco Blvd. Below the map, the following statistics are listed:

- Population: 2,468
- Households: 995
- Land Area: 0.3
- Water Area: 0.0

Below these statistics is a section titled 'NEARBY...' with a list of nearby cities and their distances:

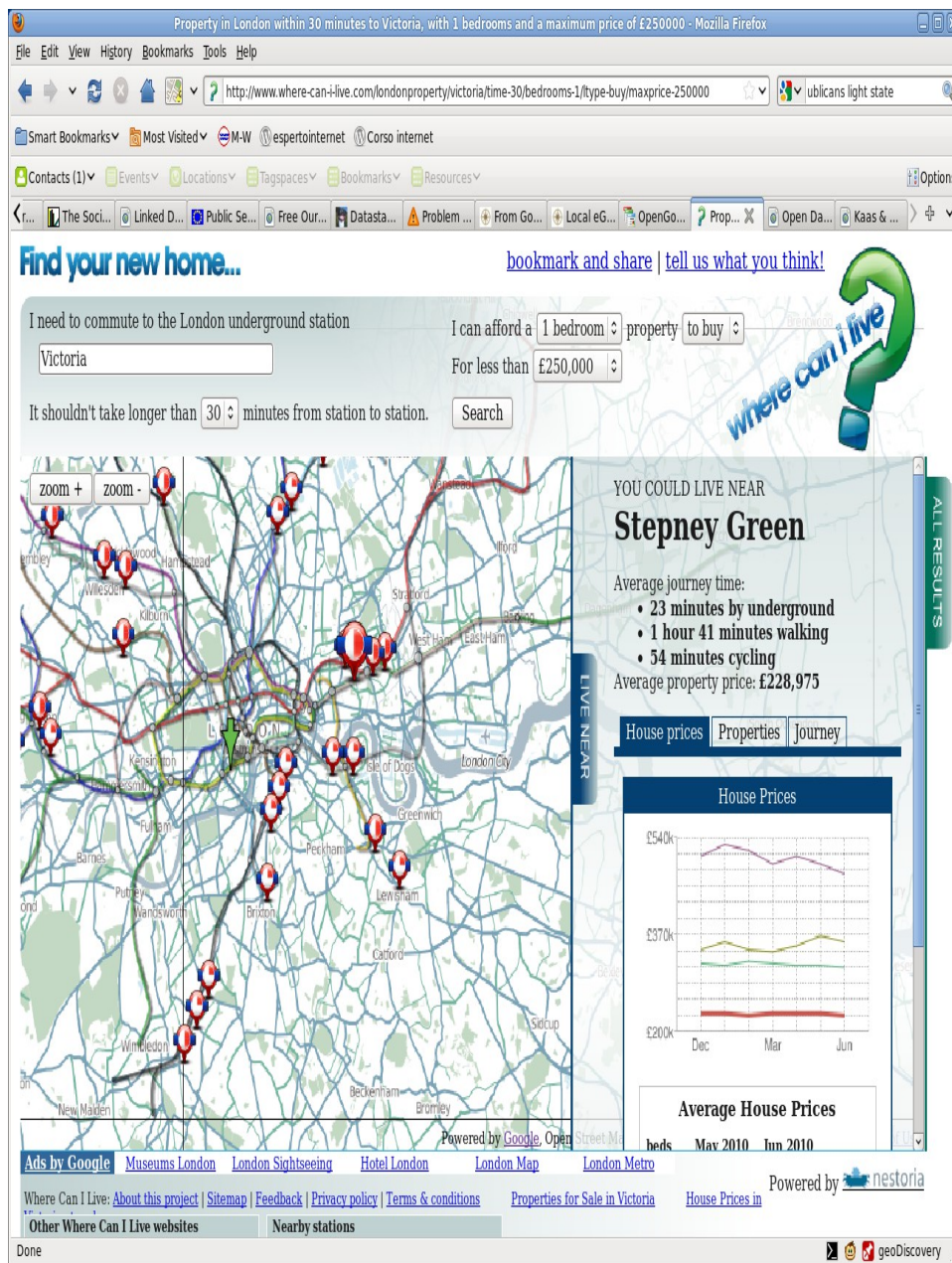
- San Jose, CA 11 mi
- Oakland, CA 30 mi
- San Francisco, CA 32 mi
- Sacramento, CA 88 mi
- Fresno, CA 134 mi

At the bottom of the page, there is a 'FEEDBACK' button and a 'Have some Feedback?' link. A footer message states: 'We Know More: This is just some of what is known about Mountain View, CA. We are actively loading more...'

5.8. Real Estate

Most current uses of Open PSI related to real estate fall in two categories. The first serves people looking to buy or rent a house.

The [London, where can I live?](http://www.where-can-i-live.com) looks at this problem from a commuter point of view: the users declare where they work in London and how much they can afford to pay for a home. The software, combining travel time between stations house-for-sale listings and average house prices, shows where they can live and how long it will take to go to work from each available place:



The other category of house-related Open PSI applications assists citizens who already are homeowners. A great example of this type is [Husets Web](#) in Denmark. Combining all sorts of PSI from local energy costs to weather statistics, maps and the technical characteristics of the materials used to build each house with extra information provided by the homeowner, Husets Web provides practical tips to reduce pollution and save on energy costs through an energy optimization calculator. What's particularly interesting is that the website also makes it very easy and quick to get a quote from local craftsmen (from plumbers to electricians) for remodeling a house in order to achieve those goals. In other words, Husets Web successfully uses Open PSI to stimulate creation and survival of local jobs that in and by themselves have nothing to do with the Internet, programming, or any other high-tech, "knowledge-economy" activity.

A proposal for the same type of service, that is helping owners of poorly insulated homes, [comes from UK](#): *"Houses of a similar construction and facing the same direction with respect to the sun would be expected to experience a similar rate of snow melt if they had similar insulation (and heated to a similar degree). Automatic comparison, using digital maps and aerial photos, of the proportion of dark vs light areas, which is roughly related to snow melt, that is to how heat each house loses, could be useful to find out automatically which owners could be more interested in insulating their homes better."*

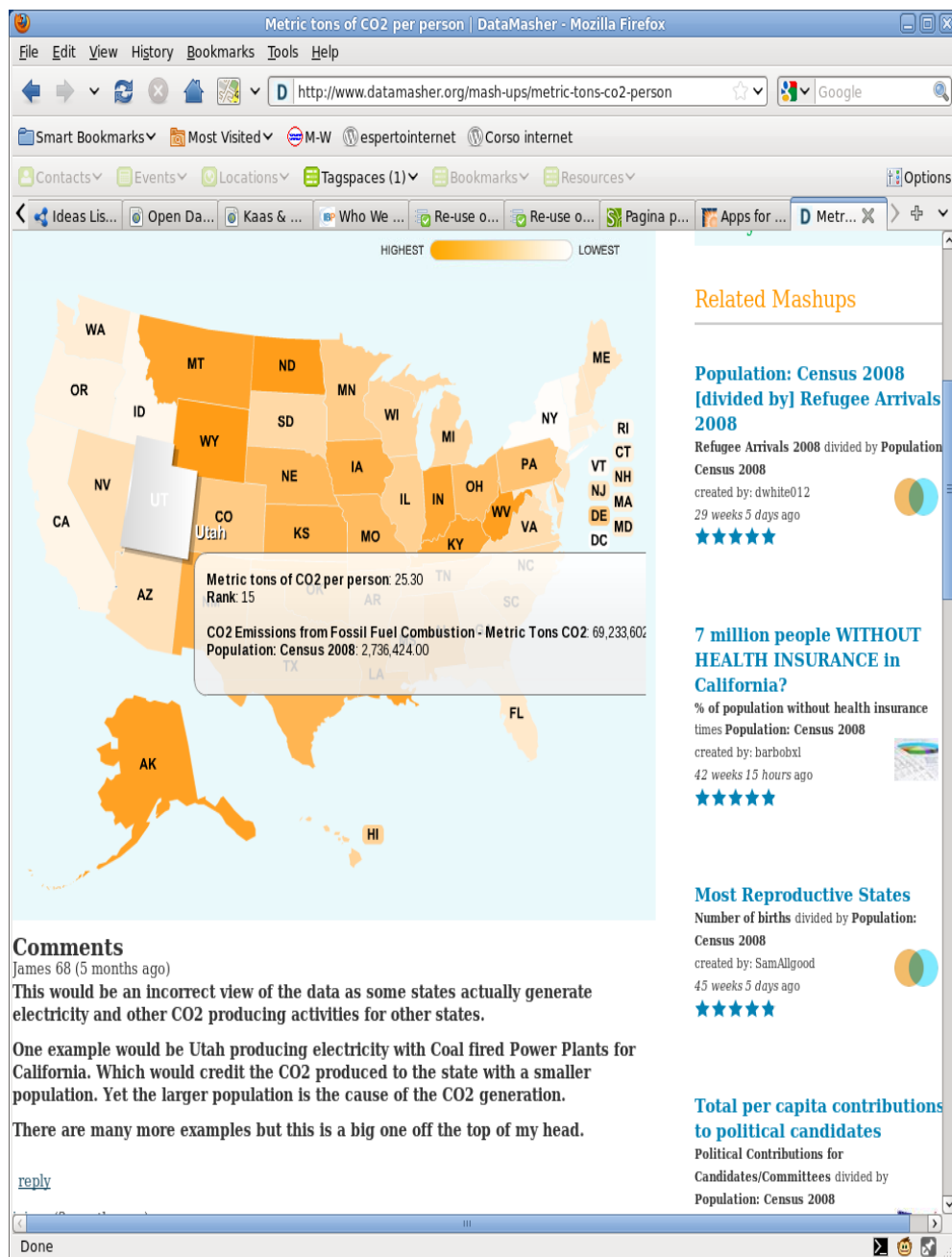
Some real-estate proposals and services based on Open public data go beyond the need of the single homeowner, to look at the status of whole neighborhoods or to actual urban planning.

[Fix My Street](#) in the UK allows citizens to report problems like graffiti, fly tipping, garbage, broken paving slabs or street lighting in a Web page and inform by email the council that would be in charge of fixing that problem. Of course, in order to work properly, Fix My Street or any other similar services like the Open311 websites in the USA, need to have unrestricted accesses to official digital maps as well as addresses and/or postcode databases. When the reports are actually and consistently used by the Public Administrations as input for their work, there are obvious savings coming both from more efficient usage of their personnel and from less delays and accidents for citizens or increased home values. Similar online databases have been proposed in the UK for [derelict buildings](#), in order to easily find their owners and pressure them to fix those buildings or just tear them down, to recover the space and therefore reduce pressures for building on green field sites.

5.9. Environmental data and pollution measurements

A website devoted to climate change issues asked in July 2010 ["What if the Public Had Perfect Climate Information 30 years ago?"](#): *"that would completely change the amount of information we have today. We would have seen that emissions reduction is inexpensive and straightforward, especially when you take a long time horizon. We would certainly be on a path to below 450 ppm.*

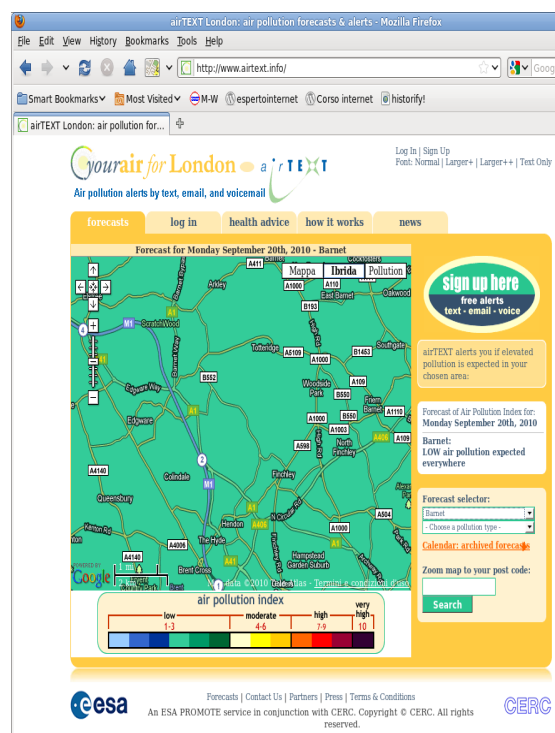
Lots of PSI of this kind already exists, and much more should be made available to make it as easy as possible to be informed about one's personal contributions to pollution, with statistics and graphs similar to those generated for the USA by DataMasher.org:



The [European Pollutant Release and Transfer Register \(E-PRTR\)](#) is an Europe-wide register that provides easily accessible key environmental data from 24,000 industrial facilities covering 65 economic activities across the European Union, Iceland, Liechtenstein and Norway, from amounts of pollutant released to air, water and land to off-site transfers of waste and of pollutants in waste water from a list of 91 key pollutants including heavy metals, pesticides, greenhouse gases and dioxins for the year 2007.

This register should allow citizens to know the emissions of industrial facilities across Europe, but in order to work as advertised it needs to be complete (meaning that industries should be required by law to provide and always keep up to date their data) and usable as a control instrument by whoever is interested in doing so. An [article](#) about this register said in November 2009 that *"it is fundamental for its success that EU states verify the quality of the data inside the register"*. Why shouldn't the all raw data that the states should use to performs such check also available online?

Those data could in such a case be compared or mixed with other independent sources, like the [UK air quality archive](#) that shows pollutants detected in many UK monitoring sites, or [airTEXT](#), a service for people who live or work in London and may be affected by higher than normal levels of air pollution because they suffer from asthma, emphysema, bronchitis, heart disease or angina.



airTEXT Subscribers receive free SMS, email or voice messages to know when they should be taking inhaler or angina spray with you or avoiding strenuous outdoor activity.

5.10. Aggregate health, or health-related data

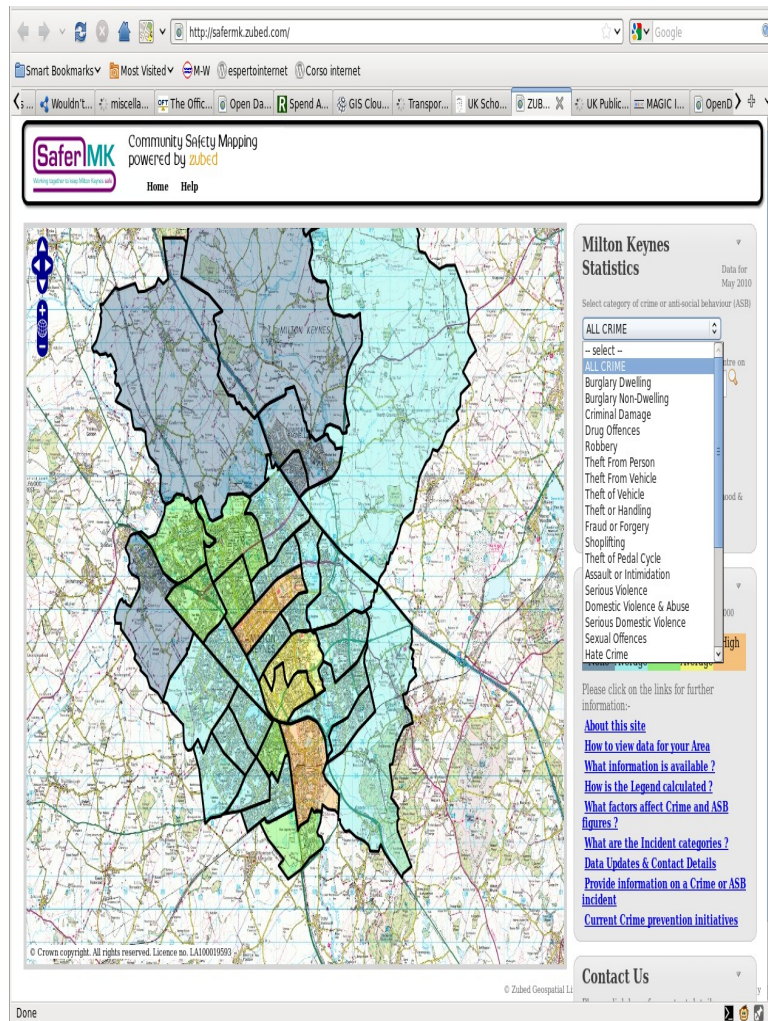
Health related PSI goes from hospital performances to hygiene ratings for food businesses to incidence of some diseases in each area of a country.

Citizens have already asked, for example, to publish online [the numbers of disability claimants](#), grouped by age ranges, disability types and area, or statistics about [Clinical Negligence cases sorted by hospital, death cause and costs](#), together with the [numbers of diagnostic tests performed by hospital laboratories](#). Meanwhile, at [Scores On The Doors](#) people can find the official local authority hygiene ratings for UK food businesses, that is how hygienic and well-managed the food preparation at any of the listed take-aways, clubs, pubs and restaurants are.

Of course, many of these data can be published only in aggregate form, to avoid privacy issues, but they can still be very helpful, on the transparency and prevention fronts, when presented in the right way. Health-related statistics may even help professionals, by making easier, cheaper and less risky for doctors to find what is the best area where to practice. Full disclosure and sharing as Open Data of all this information could also make much easier to pass new policies or expenses for national or local healthcare management. Still, all the comments about disclosure of security and legal data explained in the following paragraph apply.

5.11. Security and legal

PSI related to trials, other legal procedures, crime statistics or police operations is a delicate category of PSI, which is very powerful but requires special care, so to speak, in handling and presentation. The reason, which is also discussed in the chapter about the dangers of Open Data, is that this is PSI that requires much more context and user preparation than others to be used effectively without generating fear or confusion. One of the proposals for Open Data usage on the data.gov.uk portal is to [publish on online maps](#) *"where, when and by which power people are stopped and searched, together with their ethnicity and age"*. The Economist report on data mentions that in San Francisco citizens already come to public meetings armed with crime maps from the Crimespotting website to demand more police patrols. The [SaferMK Community Safety Mapping website](#) provides [comprehensive crime and anti-social behaviour data](#) for every estate, town or village in the Milton Keynes borough. It is easy to see how all these data can help citizens to understand and monitor the effectiveness of law enforcement and public security policies, if they are presented in the right way.



5.12. Education

Data about the education system include demographic summaries of students distribution, aggregated scores, school locations and costs, curricula, average age, salaries and specialization of teachers, grant programs. Access to this information can help families to spot deficiencies in the education system and ask that they be fixed, or students to choose which schools to attend. The latter application is the object of a [UK proposal](#): having an online database listing the number of predicted skill shortages in each area of employment, the number of university places and so on could give a general overview of the job market: *"This would be a great portal to make sure that we are not offering education at the tax payers expense or burdening people with debt with no real prospect of work - Or funding the education of those that we then lose to another country"*. Other citizens asked to [track the proportion of budget](#) that schools spend on teacher salary, resources, books etc. over time, to help understand if and how increased spending has actually increased quality of education. Websites displaying the location of UK schools according to the rating assigned to them by education watchdog Ofsted [already exist](#).

In USA, the Data.ed.gov launched in 2010 will increase access to education data. The site will ultimately serve as a one-stop shop where practitioners, researchers, and the public can access information about Department grant programs by providing tools that allow users to know which initiatives are funded in each community or see grant applications on a map that includes the option of overlays by congressional districts, filtering the results in several ways. For example, a user could search for all applicants in Texas that applied for grants to address a specific priority. Data.ed.gov also allows users to export data sets in a file format that can be loaded easily into common spreadsheet and data analysis tools.

5.13. Waste management

PSI related to waste management consists of information on how much garbage is produced in each part of the city, how much of it is recycled, what are the costs for its management, what is the status of local landfills, or even simple data like garbage collection schedules for each neighborhood. In Vancouver, [VanTrash](#) helps citizens to find out their garbage schedule, download it in their digital calendars or receive reminders by email, using garbage pickup times used scraped from the [official City website](#).

5.14. Water management

(this paragraph is a summary of an article published in December 2009 by the author, titled [Should water be public or private? Australian, of course!](#))

There are a lot of talks and public discussions in Italy these days about the "*privatization of water*" that should be soon approved by the national Parliament. Some people denounce a theft of all water that should be forbidden, period, while others declare that such concerns are just scaremongering, if not plain scams. Meanwhile, in Australia, they have dimply decided that lots of data related to water management must become automatically accessible online with open licenses, making possible for everybody to check through the Internet:

- how much water comes out of each spring and how much arrives in every neighborhood, that is how much water is lost and where
- when and where something breaks and how long it takes, on average, to fix it
- how much citizens pay for every single service or other activity performed by the management company and why
- etc...

therefore making much easier to discover which representatives should not be voted anymore because they delegated water management to organizations that (regardless of their nature) are obviously doing a bad job. Having those data would also be extremely useful in order to correlate them with other data: wouldn't it be great, for example, before buying or renting a house, to know how many times water distribution was interrupted in that street, or if it receives less water than other areas?

6. The social dangers of Open Data

As is the case with any other tool that is very powerful, Open PSI can also have negative effects, even if in the big picture, or in the medium/long term, their advantages still greatly outweigh the disadvantages. One first, potential disadvantage of opening PSI (more on this later) can be temporary disillusion and loss of interest for politics, if not disgust, in citizens. Another, more likely risk, is the fact that, at least initially, Open Data may only benefit people in the upper classes of society who have, on average, better Internet connectivity and much more familiarity with online services than the others, who could therefore may be damaged. A perfect, very recent example of this problem has been discussed in September 2010 by [M. Gurstein](#):

"A very interesting and well-documented example of this empowering of the empowered can be found in the [work of Solly Benjamin and his colleagues looking at the impact of the digitization of land records in Bangalore](#). Their findings were that newly available access to land ownership and title information in Bangalore was primarily being put to use by middle and upper income people and by corporations to gain ownership of land from the marginalized and the poor. The newly digitized and openly accessible data allowed the well to do to take the information provided and use that as the basis for instructions to land surveyors and lawyers and others to challenge titles, exploit gaps in title, take advantage of mistakes in documentation, identify opportunities and targets for bribery, among others".

For all the reasons mentioned in the previous paragraphs, Open Data (but this is true for any form of Open or E-Government in general) is also going to destroy jobs. An unavoidable consequence of large-scale adoption of services and initiatives like those we just described will be to make completely redundant several white collar jobs in the public sector, that is the sector which is the largest provider in many countries of long-term jobs, that is of social stability.

6.1. Openness is not enough for real transparency

Information is power. Availability and mass usage of data can make much harder for politicians and powerful lobbies to control public opinion and abuse of their power as it happened in the past. But the same abundance can also make much easier (and cheaper) to do the same things in other, more technologically advanced ways. When hiding information in order to maintain or gain power isn't an option anymore, it is still possible to achieve more or less the same result by providing *too much* of it, flooding relevant data under less important ones, or (not) linking and presenting them in

the correct way.

Open Data must be packaged in ways that most people care about and can *quickly* understand, in order to be effective. Above all, they must be *used* as much as possible, as soon as they are created. There is no guarantee that data will achieve a positive effect only because some generic Freedom of Information law has been approved and, consequently, data are put in plain sight. According to a [survey conducted in May 2010](#), nearly 80 per cent of local newspaper editors in UK believe that (in spite of the interest in UK for Open Data) public bodies such as the local council, police or health authority are becoming **more** secretive. 35% of editors had experienced having a reporter prevented from attending a public meeting or prevented from reporting details from it.

Similarly, there is no real guarantee of openness and transparency in the mere fact that some data *are or became*, somehow, available to the general public. A proof of this kind comes from [Estonia](#): after the first general elections in the country, the winning party donated many documents to the publicly accessible National Archives, where they sat ignored for 13 years. Only in 2006 a professional, Tallinn-based journalist Tarmo Vahter, found evidence that in 1993 party leaders had directly solicited and accepted payments from soon-to-be privatized companies. When Vahter published the story, it was too late from many practical points of view but one: the political parties terminated donations of their documents to the National Archives. We can't even be certain that those episodes would have been discovered earlier if in 1993 the Internet had already been as common as today and the data had been immediately published online. A badly indexed, non searchable website, full of PDF files with obscure names, could have hidden the facts almost as effectively as dropping paper documents in some basement of the National Archives.

6.2. Optimistic assumptions about citizens data literacy

Lies, damn lies and statistics. Leonard Henry, Baron Courtney of Penwith, 1895

Internet access greatly increases opportunities for access to information. However, it does not magically give all people the skills they need to interpret what they find. Even the so-called digital natives are simply citizens born into a world where digital technology was already commonplace. That's all the term really means and it has [nothing to do](#) with how digitally savvy they actually are. Assuming otherwise would be like assuming that all the people born after FM radio or analog TV became mass media are surely fully aware of all the ways those other technologies can influence their judgment.

Information is power and as such can be manipulated to actually *disempower* or manipulate people, especially when it's used as a tool of fear. This is particularly evident with data tied to public security, like sex offender registries or other [crime mapping tools](#), but is an absolutely general problem. Even simple lists of "risky" locations like the [Control of major accident hazards directory \(COMAH\) in UK](#) can generate panic, or at least confusion, if not released with [context](#).



An Anti-Social Behavior Order (ASBO) is a civil order made against a person who has been shown, on the balance of evidence, to have engaged in anti-social behavior in the United Kingdom and in the Republic of Ireland. In February 2010 the most popular free download in UK was the [ASBOrometer](#): a mobile application that measures levels of anti-social behavior at one's current location by looking at the number of ASBOs issued to residents of that area. The release of the ASBOrometer caused comments like ["a developer has seen the future, and it's anti-social networking"](#).

In and by itself, information doesn't necessarily lead people toward pro-active solutions. In worst cases, the extra information given first to the public may simply be the one that strengthens the position of the *one* power group already in charge.

A big part of the reason for this problem is that *transparency is not enough without real interest and literacy in the masses*. In this context, "literacy" means the combination of computer, digital media and traditional math skills necessary to correctly give context to sources, numbers and other information and to interpret everything as objectively as possible. For example, very often the *age*

or release date of some data is at least as important as their actual value or their source. The consequence is that the largest class of PSI end users, that is responsible citizens, should adapt to the idea of [data versions and version dependencies](#), just like they have already done, or should have, with versions of software programs. This kind of literacy is far from being widespread these days, is not evenly distributed across all segments of population and isn't something that people develop just because broadband comes to town or information is available online. It would be naive to assume otherwise.

If literacy is absent, data taken out of context or "assumed" without skills can have unintended consequences, like generating fear or loss of interest instead of engagement (here's another reason for linked data: they provide at least some context by themselves). As an example of these risks, Danah Boyd [quotes](#), in a talk on which part of this paragraph is based, *"the statistic from 2006 that 1 in 7 minors are sexually solicited online. Most people interpret this statistic as suggesting that 1 in 7 minors are sexually solicited by older sketchy adults seeking to meet minors offline for sex. But over 90% of sexual solicitations are from other minors or young adults, 69% of solicitations involve no attempt at offline contact and the term "solicitation" refers to any communication of a sexual nature, including sexual harassment and flirtation"*.

These are not theoretical concerns. The author [personally experienced](#) several bloggers republishing without problems, even *after* being told about that it wouldn't make sense, obviously absurd assertions like *"in 2003 Microsoft got from the Italian state more money than the state deficit in that year"*. In the USA, a [2010 study](#) concluded that *"about 70% of students in Grade 6 in the U.S. exhibit misconceptions" about the equal sign*". Tests performed in Italy in the same year on 125.389 primary and junior high school students showed a [decrease of math skills with age](#): correct answers to math tests were 61,3% among 10-year old students, but only 50,9% among 11-year old ones. Still in 2010, a report on ["Trust Online: Young Adults' Evaluation of Web Content"](#) concluded that students rely greatly on search engine *brands* to guide them to what they then perceive as credible material: over a quarter of respondents mentioned that they chose a Web site only because their preferred search engine had returned that site as the first result.

Obviously this report and many other sources still prove that it is necessary to open as much PSI as possible, if nothing else to give private entrepreneurs more opportunities to start new businesses. Our point here is simply to remind that opening PSI can be enough in that sphere, but is far from being enough when it comes to transparency in government, at any level. That can only happen if there is a *mass* interest, usage and understanding of Open Data.

7. What to do

All the examples and the analyses presented so far confirm that, even if there are some serious issues whose importance shouldn't be underestimated, opening all PSI that can be opened and use it as much as possible starting at the local level makes a lot of sense. How should this happen in practice? How can this process be sustained and stimulated? Several actions can be taken at the political, legal, technical and practical level. We will describe them, in this order, in the next sections, and immediately after we'll shortly discuss what is the role of the Public Sector in future scenarios where, just thanks to Open PSI, much of the work done today inside public organizations is done by (groups of) volunteer citizens or by private companies.

7.1. Political support for Open Data

In the next years, some smart politicians will surely realize by themselves that *opening data is a fresh and powerful promotional tool to gain votes and support*. Regardless of this, and in spite of our emphasis on the local usage and support of PSI, one thing is clear. Much of the activity now taking place in the UK is a direct or indirect consequence of the 2004 Freedom of Information Act [FOIA] and other steps [made years ago](#). In order to get in other countries the same benefits of the most successful initiatives of this kind in the UK and USA, said Becky Hogge in May 2010 in an [Open Data Study for the Transparency and Accountability Initiative](#), *"It has to start at the top, it has to start in the middle and it has to start at the bottom"*. We can only add to that the sooner the better, even if local administrations shouldn't certainly wait to open their data whenever they can, as soon as possible.

The support for Open Data that should come from the top, that is from central institutions, consists mainly of three things. The first is to clearly define by law what is public PSI and what isn't. It is not acceptable to have as a general, official or unofficial policy, to leave to the individual public departments creating some PSI to decide by themselves what data to open and what not. All data that, under existing laws, should already be given to citizens whenever they explicitly request them are good candidates for opening, but there are many others, especially because the definition of "data" isn't always so rigid and simple in the first place.

For example in Arizona, in 2009, it took an appeal to the Supreme Court of the State to rule that even *"the metadata attached to public records is itself public, and cannot be withheld in response to a public records request. Such a ruling on file metadata may not seem like a huge win for open government advocates, but it definitely is, given that metadata has unmasked more than one*

lobbyist's effort to influence Congress.". The [ruling](#) happened because a police officer demoted three years earlier had requested access to some computers and files to verify the creation time (that is, file metadata) of some negative performance reports that he suspected had been written as retaliation *after* he had denounced serious misconduct of some colleagues.

Another issue that should receive more attention is the very definition of "public", since it can create conflicts between the need for transparency, the one for privacy, the ways such needs are defined and protected by existing laws and the ways in which some Public Administrations more or less freely distinguish between the two. This issue is explained by, among others, a [blog post by Diego Ghisilieri](#), summarized here:

From a juridical point of view, the fact that a certain data is "public", meaning that it must be accessible to anybody who asks for it, doesn't mean that the same data can or should be distributed around without limits: there must be a balance between the need of citizens to know and the right to privacy (including the so-called "right to oblivion") of the specific people mentioned by, or directly identifiable by those data. In fact, if all such data were accessible online by anybody, including search engine indexers, without any restrictions, it would become possible to profile those people by mixing those and many other data (possibly outdated or relating to completely different contexts) for purposes that have no relationship whatsoever with the need for transparency in Public Administration.

In any case, government officials should be required to justify [why any public data should not be freely available to the taxpayers who paid for its creation](#) (taking into account what already exposed in this report, like the fact that charging for PSI to sustain the specific administration that creates them is almost always the least efficient strategy).

Different laws and regulations at this level are the main reason why some success stories from a certain EU country cannot be immediately replicated in others. As an example, the Webhusets initiative in Denmark relies on the fact that even all building designs and lots of technical information about them and the materials used is considered PSI that must be delivered to the City where a building is, and then be accessible by whoever requests it.

The second thing to do is to mandate law that all the PSI that has been defined as public, that is that can be opened without creating privacy, security or similar issues, be actually opened as soon as possible. In practice, it will be necessary to distinguish between PSI that already exists and PSI that will be created in the future. Data in the first category still are, sometimes, in non-digital formats and in a non clear legal status. Therefore, converting them to open formats and obtaining authorization to their publication are extra efforts that must be taken into account.

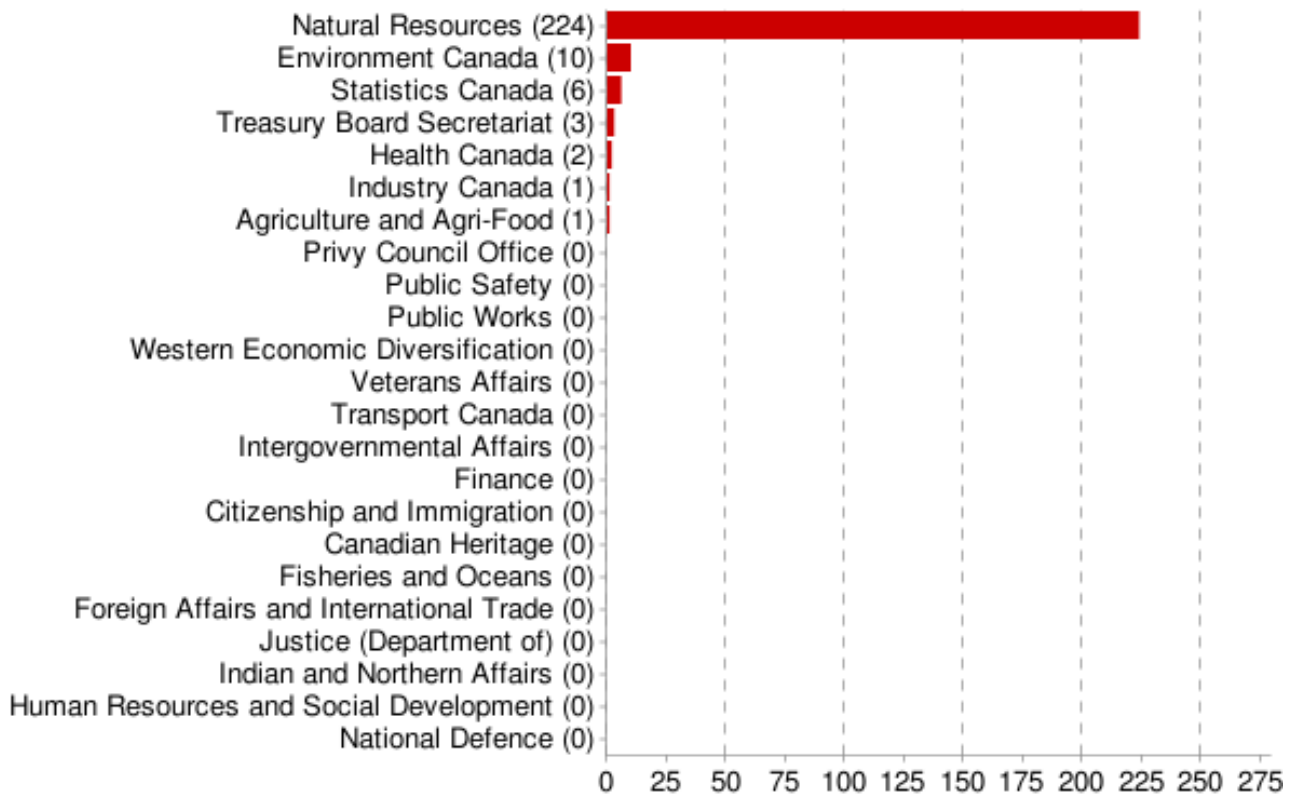
Data that must still be generated instead are easier to handle. In that case, laws must make it mandatory to publish all that public PSI online with an open license from the start, also for economical reasons: the final cost of 'adding openness' at the end because citizens and private businesses start asking for the data that an administration is forced by law to provide is higher than creating open PSI by default from the very beginning. It is equally necessary to establish the principle that all data of the same type created, with public money, by third parties on behalf of some Public Administration also are public PSI that must also be released with an open license. Only under such conditions private businesses and volunteer groups will be able to add value to the data at the lowest possible costs.

As we already mentioned, the advantages of mandating data openness are very easy to see both with strictly technical information like digital maps and in cases like the YourNextMP initiative in UK: all the basic work they are currently doing to collect and digitize candidate data is something that law should oblige each candidate to do by him or herself, without wasting public money or anybody's time. Every candidate should publish under an open license a full CV using technologies as RDF (Resource Description Framework) that allow to link the data, that is to declare their relationship with other data like unique codes in the companies databases, land or house ownership registries and so on. Once that is granted, services like YourNextMP could finally concentrate on adding value, maintaining simpler Web services that can immediately answer questions like: What companies has this candidate been director of? What charities does he or she support?

The choice of proper licenses for PSI is obviously essential. Common guidelines and recommendations on this topic at EU or at least country level would make much easier to exchange, correlate and reuse PSI, especially because the best choice also depends on the technical nature of the data. When it comes to databases, for example, several experts suggest to not use the popular Creative Commons licenses (with the exception of the one called CC0), but to adopt the [Open Database License](#). The reasons are explained in detail in the article [Why Need For Database License](#) and in the [Open Knowledge Definition](#), which comprises 11 clauses providing detail around the core premise that 'open' data should be freely available online for use and re-use. The [UK PSI Licence](#) is one useful example of the ways in which these generic principles may be practically implemented by governments. The already mentioned [LAPSI project](#) will look at all the legal issues related to PSI in much more detail than it would be possible in this report, so we invite readers to follow that project for in depth analysis and legal advice.

The last category of actions that should be promoted at all levels, but starting from the top, is to

create incentives and public demand to open public PSI without waiting for laws that mandate it and/or expose those public bodies that don't do it. An example of this kind is the website [Who's sharing in Canada?](#) which, in order to "encourage our government to share more structured data, publishes a graph showing which ministries share and which do not. It is a powerful metric of how transparent a given ministry is".



The [UK Councils Open Data Scoreboard](#) does a similar job: as of July 2010 it reports that 18 out of 434 local authorities publish open data (but only 9 are truly open). The same thing happens in [California](#), where CityGoRound reports in the same period that 691 transit agencies do not provide yet open data to software developers. Another good example from this point of view, still from the USA, is the [Public Transit Openness Index](#), which measures how much Public Transit companies across the USA are open to reuse of their data by publishing lots of parameters, from the file formats they use to whether or not they have sent cease-or-desist letters to third parties reusing their data.

7.2. Some short technical notes on data formats and software managing open PSI

Data file formats used to store and distribute PSI are also extremely important, as is the software used to process them. Here we only want to mention a couple of points about these subjects because we will elaborate more on them in the final report of this research.

Only if the formats are the simplest possible ones and are truly open, that is usable without asking permission or paying any royalty in any software program it is possible to speak of Open PSI. In the case of office documents, the best solution for new office texts, spreadsheets and presentations is the OpenDocument format ([ODF, standard ISO 26300](#)). ODF is an example of XML (eXtensible Markup Language), a generic technique to create data formats that are both open and linkable. As explained in an interview by [Simone Cortesi](#): *"XML is a format that allows data to be related to other data because it lets you refer to external content, that is to name an object and search about it via the Web, to retrieve further information on the same data... If we know that Rome is defined as city, we can go look in other databases on the web, always written in XML or accessible through XML, all information about objects of the city type that are called Rome and therefore interconnect the data. A further specification on XML, called RDF (Resource Description Framework), can connect to databases on the Internet. This enhances the value of a single database, because if we have a database it can be linked to other remote, independent databases in turn linked to other ones"*. RDF is already used in this way: Richard Cyganiak's [Linking Open Data Cloud diagram](#) now represents over 13 billion [RDF statements](#) connecting data from across a growing network of participating sites.

When it comes to software, practically all public services are based on it these days, so Free/Open Source Software (FOSS) that every developer can modify or reuse without license fees or similar restrictions is a necessary component of an open digital society. However, it is worthwhile to point out that, technically speaking, Open Data may happen even when proprietary, closed source software is used in their generation, analysis and distribution, as long as only really open standards are allowed for file formats and computer protocols.

7.3. Practical advice for Public Officers promoting Open Data

7.3.1. Don't worry (initially) about data quality

Just don't worry, initially of course, about data quality. Graves [explicitly recommends](#) that public sector bodies make PSI available at the earliest point that it is useful to businesses and citizens. In practice, this means as soon as possible and "quality" shouldn't be an issue. First of all, a corollary of the fact that data should be open because they are like soil and therefore not even their creators can possibly know all the ways to use them, is that the same creators can't even be always absolutely sure that *they* have all the elements to properly evaluate if quality of their data is good or poor. Even if quality were not sufficient for the Public Administration (which would then be a problem to solve regardless of openness!) it may be already good enough for third parties. In the second place, quoting [the Business case for PSI](#) *"quality of published data sets actually increases, both because of much more feedback from end users and because of more attention being given to generating the data in the first place"*. When the Greater London Authority asked developers community how should they release their data, [the response was clear](#): *"Go ugly early – don't worry about formats – just get the data out there and we will help you to clear it up... The sooner you can get data sets up and build sample applications to demonstrate the purpose and benefits of open data, the more likely you are to encourage other people to give you their data"*. Data quality is a case where the slogans of the Open Source Software movement, *"release early, release often"* and *"given enough eyes, all bugs are shallow"* really apply and can give positive results.

7.3.2. Beware of too much looking for a "business case for open data"

Speaking of how to justify opening PSI [Zijlstra](#) rightly says that: *"The business case has a long history of being abused to stop change: business cases are fine for investments that are one all or nothing decision about something of which all possible returns are known in advance and will happen in the same department that is considering the business case"* (something we already said it's impossible). When it comes to Open Data, instead, [according to many experts](#), *"it is very difficult, if not impossible in some cases, to quantify in advance the economic value of opening PSI"*. Even the already quoted MEPSIR report says: *"It turns out to be impossible to draw conclusions... at the level of the domains of PSI (e.g. legal information, social data, meteorological information, geographical information, and business information)... Generic business cases for open PSI cannot be made in a way that is relevant for a specific public body"*. Claims of huge savings at EU level mean very little for a local manager that has already finished her budget.

Another reason to not put too much faith in "business case" evaluations is that there will be costs that become *apparent* because of an open PSI project, but are not caused by it. Usage of, or conversion to, open file formats is already required by law in several EU countries, so it's a cost that sooner or later must be paid anyway regardless of openness. In practice, it may make much more sense to not look for a traditional business case but just start gradually, that is locally, as we already recommended, and/or by opening first, as soon as possible, easily available, non-controversial data sets. Another valid criteria to decide which data sets should be opened first is to start with those for which community-generated alternatives already exist (e.g. OpenStreetMap). The existence of such alternatives proves the public need and interest for those data, and therefore releasing the official ones will allow to those communities to concentrate on *adding values to the existing raw data and improving their quality, rather than re-generating everything from scratch*.

7.3.3. Other practical advice

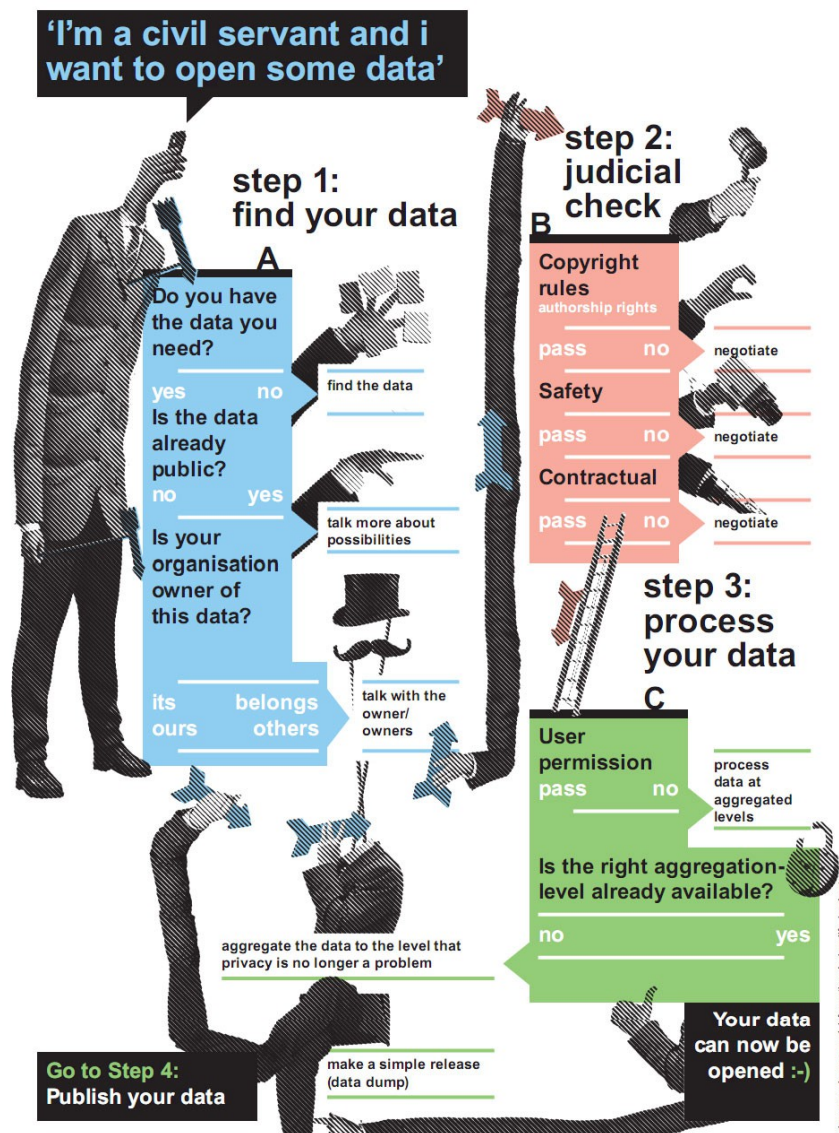
Other do's and don't's for Public Administrations, partly derived from the list first issued by the [London Datastore](#) are:

- Don't waste scarce resources on expensive consultancy firms unless you're already sure and have proved that no help will come at a lower cost from the community
- Avoid building new, possibly very expensive official public websites unless it's really, absolutely necessary. Just put the raw data online instead: *"Councils need stand-alone open data projects with their own resources and budgets. Lumping it in with general website work has demonstrably failed to give open data the priority it deserves"*. The only work that really needs to be done on many public websites, which by the way is only partially related to PSI openness, is to [make them search engine friendly](#)
- Publishing PSI online is not enough. It is necessary to establish and maintain connections with end users to get (often for free) their help to improve the data, measure reuse and publish those measures

7.4. Citizens education

Many citizens need education to understand raw public data, in order to put them in context and to make informed decisions when voting or otherwise interacting with their representatives. Such education should of course start in schools and Universities. Even outside those environments, specific actions to make sure that as many citizens as possible have all the skills they need to really benefit from Open Data are necessary.

Public employees also need to really understand the real nature and value of the PSI they generate or use every day, and how and why to make it open. In this context it will be helpful to promote initiatives like the [Open Government Data Poster](#), which tries to give an easy, visual explanation of the issues around open government data for civil servants.



Another class of citizens that should get special assistance are NGOs employees and volunteers. Many of the skills needed to create, access and use open data [are not yet widespread in the voluntary sector](#), and as open data becomes embedded in government, voluntary organizations which contract with government will have (see above) to be compelled to produce and share data as part of those contracts. At another level, Open Data Challenges, that is contests among developers

of software services that use or analyze open PSI like the one [successfully held in Spain in 2010](#) or those in New York and Finland, could and should become a regular occurrence all across Europe (including schools!).

One last but crucial thing is to make sure that education to Open Data does not ignore adults and senior citizens, *especially* when they haven't easy access to online services. The speech on [building Britain's digital future](#) explicitly mentions the need to *"put the 4 million people who are among the heaviest users of government services – but who have never used the internet – at the heart of our strategy rather than letting them literally slip through the digital net. Increasingly the digital net will be the social safety net – the only way to extend access to better services to all of our citizens"*.

7.4.1. Education to Open Data, or Open Data for Education?

In the previous chapters we have discussed at length the fact that raw, Open PSI can help a lot to achieve real transparency in government and really participated democracies. We have also explained why this will really happen only if enough citizens not only have physical access to online open data, that is sufficient Internet connectivity, but also the right skills to make sense of the data they find online. There is also another side of the Education/Open Data coin. In all of this report we have shown how Open Data can be a really powerful tool to give more opportunities, at many levels, to... *adult* citizens. If we look at Open Data from the point of view of an educator, instead, it's evident that they are also a wonderful opportunity to build more Open Educational Resources (OER), that is textbooks, exercise books and other courseware and to generally make teaching much more effective.

Mentioning just one of many possibilities, once all the financial data we've mentioned so far are open, raw and linked it is much easier, and less expensive, to write accounting manuals full of very up to date exercises and examples from the same area in which the students live. For the same reasons, it is also much easier to build interactive websites or software programs that explain the concepts introduced in class by mashing data or let the students practice with them at home. Of course, making this happen requires (again) specific education for adults, in this case the teachers, to help them use these resources.

7.5. The role of the Public Sector in an age of Open Government: quality and reliability of openly generated PSI

There is a key issue that, so far, has been deliberately ignored in this report to deal with Open Data in an ordered manner, but will become more and more important in the medium and long term. In

order to fulfill all their promises, raw PSI must not only be *open* and *linked* as previously defined, but must also be reliable and *valid in court*. This may be a problem wherever, as a consequence of opening PSI, citizen will be involved not just in the analysis and usage of public data, but also in their *generation*. Such a process is already at the base of Open311 systems. Other cases of citizens already generating PSI are described in the article ["Municipalities open their GIS systems to citizens"](#).

Citizen-generated digital maps have been already used for "official" purposes in places where there is not enough public money or market to create official, high-quality digital maps, for example in Gaza to help [ambulance drivers and other humanitarian relief personnel](#), or in [Albania](#), to support the population of Shkoder after a flood in January 2010. Other slums and lower income communities worldwide are [mapped only in this way](#), like the [San Javier La Loma neighborhood in Medellin, Colombia](#). In cases like Gaza or any other disaster site, open maps and other data sets quickly generated by volunteers are and will remain invaluable, as the only way to offer some services as quickly as possible. Such initiatives are wonderful and the only practicable solutions in scenarios like those above, but could they ever become the default way of working?

In other words, what are the value, or the limits, of community-generated maps or other data sets, and what's left to do in an Open Data world for the public bodies that once were the only creators of PSI and would normally keep it locked? Sticking to maps, let's consider with two extremely simplified but not-so hypothetical cases whether collaborative PSI-related efforts like OpenStreetMap may have any value in court:

Case #1: *"Your Honor, I didn't pay the new property tax for my house because it only applies to properties bigger than 10000 square meters, and everybody can see on this digital map that everybody can edit that my property is just 9899 square meters"*

Case #2: *"I am suing the City because I fell in a pothole which, as it is evident from this digital map that everybody can edit, lies two meters outside my back yard, that is in a location that the City, not me, must maintain safe"*

These two examples show the limits of user-generated PSI-data. In cases like these a Wikipedia model, in which vandalism or good faith errors in the data are sooner or later fixed by other volunteers patrolling the data, cannot work. The more PSI will be opened and heavily used, the more it will come natural to generate it collaboratively, and therefore it will become more and more necessary for all citizens to have real guarantees of PSI reliability and to know which public officer is responsible for it.

Today the problem of how much we can or should trust open PSI is still more theoretical than practical, simply because most data aren't accessible yet and there are very little possibilities for random third parties to alter them to their advantage. Eventually, however, PSI data sets must be reliable and continuously validated by somebody, in a way that holds in court, even if they were generated in an open manner. On one hand, this means that all procedures of online publication of open PSI will have to include as soon as possible both some digital signature mechanism and the name and contact info of the public officer responsible for the authenticity of those data. On the other hand, what we just said means that the role of Public Administrations and public officers directly responding to all the citizens they serve will remain essential and not replaceable: instead of being creators, exclusive users and guardians of secret data, they will have to become guardians of the *openness, usability, authenticity and quality* of the same data.

8. Conclusions, next steps and contact information

Practically all the real-world examples presented in these pages confirm a few general facts and principles of Open PSI. One is that what is really useful are the *relations* between different, apparently unrelated types of PSI generated by independent public bodies. Another fact is that, in practice, such relations are almost always found, analyzed and made available by third parties. Today it is very hard to make sure that PSI is regularly published, up to date and reliably. However, once these conditions are guaranteed, almost always somebody will use the data.

It also seems that geographical PSI is the most important PSI, at least for the general public, or as the first one that should be completely opened. The reason is that such PSI adds context and relevance to all other types of PSI for everybody, not just specialists, in what is probably the easiest and most effective way: showing *where* some PSI exists or has tangible impacts on everyday life.

Coordination in PSI production and management between citizens and Public Administrations will (have to) become more and more important. Working together through the Internet, citizens can do a lot to *create from scratch, digitize, validate and index* PSI. In many cases, they are already doing it for free, from OpenStreetMap to digitizing election leaflets or other non-electronic documents. Such efforts should be explicitly and officially encouraged and supported as much as possible by Public Administrations, for at least two reasons. The first is to have extra data sets of great public interests created for free or almost for free by volunteers, that is at the smallest possible cost for taxpayers. The second is to increase the legal and economic value to such data by *validating* them: the usefulness of many kinds of PSI (starting with geographical data) is maximized only when its quality and reliability are officially confirmed by a Public Administration. Such activities, however, are practically and legally possible in cheap and efficient manners only if all the PSI generated in this way is be raw, open and linked from the beginning.

Finally, the hardest problem may be to get enough citizens to *use* the open PSI made available online, on a regular basis, especially when taking decisions on political matters. Making raw PSI open can be enough if all one is looking for is more stimuli for economic activities, not when the rationale for open data is transparency in politics and active democracy. This is an issue that deserves both more study and specific educational initiatives targeting all citizens.

The "Open Data, Open Society" research project will continue with an online survey that will attempt to assess how many types of raw PSI are already released, in which formats and under which licenses, by the city and regional administration of the EU-15 countries.

A final report will summarize and comment t he results of the survey. For more information please contact:

Marco Fioretti (<http://mfioretti.com>) mfioretti@nexaima.net

Prof Giulio Bottazzi (<http://cafim.sssup.it/~giulio/>) bottazzi@sssup.it

9. Useful Open Data resources not specifically quoted in this report

- The [2004 UNESCO Policy Guidelines on the Development and Promotion of Governmental Public Sector Information](#)
- The [National Research Council/OECD report on The Socioeconomic Effects of Public Sector Information Online: Toward a Better Understanding of Different Access and Reuse Policies \(2009\)](#)
- [Toward Implementation of the GEOSS Data Sharing Principles, Journal of Space Law \(2009\)](#)
- the [2008 OECD Recommendation on Public Sector Information and related materials](#)
- [Measuring European Public Sector Information Resources \(2003\)](#)
- [Importance of Open Data stressed by the same inventor of the World Wide Web](#)
- the [Finnish Open Government Guidebook](#)
- [Models of PSI Provision via Trading Funds, Pollock, March 2008](#)