

Open Data: Emerging trends, issues and best practices

*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*

This report is part of the “Open Data, Open Society” 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. Introduction.....	3
2. Social and political landscape.....	3
2.1. Wikileaks and the Open Data movement.....	5
2.2. Data Openness in EU.....	6
2.3. Open Data in Latin America, Asia and Africa.....	8
3. Emerging trends and issues related to Open Data.....	11
3.1. Cost of not opening PSI is increasing.....	11
3.2. Creative, unforeseen uses of local Open Data increase.....	12
3.3. Legal issues remain crucial.....	13
3.4. The price of digitization.....	14
3.5. The nature of Open Government and the relationship between citizens and Government.....	15
3.6. Clearer vision of the real risks and limits of Open Data.....	16
3.6.1. Data alterations and financial sustainability.....	17
3.6.2. Real impact of data manipulation or misunderstanding.....	17
3.6.3. Unequal access.....	19
3.6.4. Lack of education to data.....	20
3.6.5. Lack of public interest.....	21
3.6.6. Unprepared Public Administrators.....	22
3.7. The privacy problem.....	22
3.8. Need to better define what is Public Data.....	23
4. Conclusion: seven Open Data strategy and best practices suggestions.....	27
4.1. Properly define and explain both Open Data and Public Data.....	27
4.2. Keep political issues separated by economics ones.....	27
4.3. Keep past and future separate.....	28
4.4. Impose proper licensing and streamline procurement.....	29
4.5. Educate citizens to understand and use data.....	30
4.6. Focus on local, specific issues to raise interest for Open Data.....	31
4.7. Involve NGOs, charities and business associations.....	32
5. Bibliography.....	33

1. Introduction

This report is the final deliverable of the Open Data, Open Society research project. It follows the publication of the [Open Data, Open Society report](#), finished in late October 2010 and published in early January 2011. That first report focused on explaining the critical importance of digital data in contemporary society and business activities; defining Open Data; giving examples on their potential, especially at the local level, on transparency and economics activities; finally, defining summarizing some general best practices.

This second report looks at what happened in the Open Data arena after October 2010. After some considerations on the general social and political background in late 2010/early 2011, it is divided in two main parts. The first describes some emerging trends and issues related to Open Data, that got minor or no coverage in the first report. The second part discusses some practices and actions to follow to deal with those trends and issues.

2. Social and political landscape

It is worthwhile to begin by mentioning several events, happened between the end of 2010 and the first months of 2011, that can help to understand what will be the place and role of Open Data in the future, as well as the challenges faced by its advocates.

The first two are the Spanish "Indignados" and the Arab Spring. The first movement has among its goals *"a change in society and an increase in social awareness"*. The Arab Spring, as L. Millar put it on the [New Zealand Computer Society website](#), *"demonstrated the potency of technology to reflect citizens' views of government systems that are not transparent."* As a consequence, noted the [Afrinnovator](#) blog, *"we have seen from the civil disobedience in the North of Africa and the Middle East, the appetite for more accountable and transparent government will only grow from here on"*. From this analysis it looks like, in a way, both the Indignados and the participants to the Arab Spring are (also) asking for Open Data, even if they aren't using the term and many participants to these grassroots movement may still ignore its definition, that was born inside hackers and Public Administration circles.

Two other important events that, in different ways and at different levels, prove the importance of Open Data are the Fukushima nuclear accident and the Cablegate, which we'll analyze in the next

paragraph. Whatever one may think about nuclear power, Fukushima remembered how important total transparency and accountability are in the management and maintenance of *all* power sources, and in the decision-making processes that create the corresponding public policies.

For the meantime, we'll note how all these events seem to hint that structural need and bottom-up demand for Open Data are mounting everywhere, even in cultural contexts very different than those in which Open Data was born, and even if sometimes they are not mentioned explicitly or consciously. Even in Western Countries, the high-level *motivations*, for the transparency and governance models that inspire Open Data, from positions different than those from which the movement started, are increasing. In 1931 Pope Pio XI wrote, in the Encyclical [Quadragesimo anno](#) that:

80. The supreme authority of the State ought, therefore, to let subordinate groups handle matters and concerns of lesser importance, which would otherwise dissipate its efforts greatly. Thereby the State will more freely, powerfully, and effectively do all those things that belong to it alone because it alone can do them: directing, watching, urging, restraining, as occasion requires and necessity demands. Therefore, those in power should be sure that the more perfectly a graduated order is kept among the various associations, in observance of the principle of "subsidiary function," the stronger social authority and effectiveness will be the happier and more prosperous the condition of the State.

This is the principle of subsidiarity, often summarized in a way that may sound familiar to many Open Data advocates: "*What men can do by themselves with their own resources can't be taken away from them and assigned as a task to society*". In March 2011, journalist Guido Gentili made [just this connection](#). After noting that the principle was also introduced in the Italian Constitution by the 2001 reform of article 118, he concluded that subsidiarity as a strategy for development *isn't an English invention and the "Big Society" vision (a proposal in which Open data is key) would do good to Italy too*".

At a more practical and economical level, digital information continues to increase. In spite of mounting cost pressures, large public and private organizations have to maintain massive amounts of structured and unstructured data, that keep growing, both for their own internal needs and to simply *comply with government regulations*. At the same time, signals that traditional public services and the whole welfare state won't remain sustainable for long with traditional means, continue to arrive, therefore strengthening the search for radical, innovative and cost-effective solutions.

Besides costs, another practical driver and justification for Open Data that is becoming more and

more concrete over time is damage control. In a world that produces digital data without interruption, uncontrolled and unpredictable data releases are facts of life that are very hard to predict, practically impossible to avoid and increasingly common. Opening public government data, that is providing plenty of officially verified information, becomes therefore also a damage control solution, to prevent or at least minimize damages from such uncontrolled releases. Without official Open Public Data, individual citizens, political parties or other organizations will start to process and compare (if they already aren't...) data from unofficial sources anyway, maybe from different countries. In such cases, it will be unavoidable not reach sometimes, even in good faith, wrong conclusions. This is not some theoretical possibility far in the future, as this real world example (from a comment to an [Open Data discussion in an italian blog](#)) proves:

"on the [non italian] Geonames website you can download geo-referenced data about... 47000 Italian municipalities. That worries me, because there are only 8094 of them. Besides, I grabbed a few random data about population, and I can guarantee you that not one was right. What should be done in such cases?"

From an Open Data perspective, all these recent stories have (at least) one thing in common: they suggest that, considering its current needs and problems, current societies want and need more Open Data than they already have.

2.1. Wikileaks and the Open Data movement

During the 2010/2011 winter the discussions around the Cablegate and other documents published by Wikileaks have, in some occasion, included hostility towards Open Data. This is a consequence of a more or less conscious mixing of the two themes, because in a very general sense, both Open Data and Wikileaks are about transparency, accountability and democracy.

As far as this study is concerned, two conclusions can be drawn from the Cablegate/Wikileaks scandal.

The first is that, in practice, it is necessary to find an equilibrium between secrecy and transparency whenever government activities are concerned. Citizens must be able to know what the state is *actually* doing but sometimes, be it for careful evaluation of all the alternatives or because of security, it must be possible to work behind closed doors, [at least temporarily](#). We'll come back to this point later in this report.

The second conclusion is that, while certainly both Open Data and Wikileaks are about openness and transparency in politics, not only there are deep differences between the two ideas but, in our

opinion, the Wikileaks experience proves the advantages of Open Data.

Was Wikileaks right to publish the cable? Were the specific facts and behaviors uncovered by Cablegate right or wrong? The answer to these questions are outside the scope of this document. Here we only wish to point out that Cablegate and Wikileaks, at least in the form we've known them so far, have been about:

- reacting to problems *after* they occurred
- without any intervention and involvement of the parties and organizations that may have behaved improperly

Open Data, instead, is about *prevention* of errors, abuses and inefficiencies, through conscious and continuous collaboration of citizens and governments officials *during* day to day operations, if not before their beginning.

Of course, citizens must always check that they aren't getting incomplete or biased data. But in any case, Open Data means that the involved government officials aren't just prepared to see that data published, they know and accept it from the start. In such a context, some risks associated to Wikileaks, like the fact that the leaker lacks the means to influence the downstream use of the information, and therefore may harm anybody connected to the linked information, are almost non-existent.

Above all, unlike the content of most Wikileaks documents, Open Data are almost always data that should surely be open, unlike wartime military reports, and that almost never contain any personal information. In summary, whatever the conclusions about Wikileaks are, they could not be conclusions against Open Data, because there are too many differences between the two movements.

2.2. Data Openness in EU

Both the interest and the need for data openness at the European Union level remain high. Here, without making any complete analysis, we'll only report and comment a few relevant episodes. While studies continue to point at the political and economical advantages of Open Data, great inefficiencies and delays still keep the time and cost savings that could be achieved a far goal for the European Union.

All the principles of the Open Declaration (collaboration, transparency, empowerment) have been declared key areas of action of the new EC eGov action plan. Particularly important, as explained

by David Osimo in [EU eGov action plan published: the good, the bad and the unknown](#), are the actions on Open Data (a EU portal and a revision of the EU PSI directive), and on citizens control over their data. However the Action Plan contains no reference to the need for a more open and collaborative governance.

In the case of European Structural Funds, as Luigi Reggi [reported in March 2011](#):

there is no single point of access to the data. Hundreds of Managing Authorities are following different paths and implementing different information strategies when opening up their data.

Many databases (often simple PDF lists) [. . . show . . .] huge variation not only in the way they can be accessed but also in content and quality of data provided.

... [. . . The results of . . .] an independent web-based survey on the overall quality of data published by each Managing Authority responsible for the 434 Operational Programmes approved in July 2009... can be summarized as follows:

The use of open, machine-processable and linked-data formats have unexpected advantages in terms of transparency and re-use of the data by the public and private sector. The application of these technical principles does not need extra budget or major changes in government organization and information management; nor does it require the update of existing software and infrastructures. What is needed today is the promotion among national and local authorities of the culture of transparency and the raising of awareness of the benefits that could derive from opening up existing data and information in a re-usable way.

The European Cohesion Policy is only halfway to accomplishing a paradigm shift to open data, with differences in performance both between and - in some cases - within European Countries.

Things don't go much better for the European Union in the energy field. Carlo Stagnaro wrote in [EU Energy Orwellianism: Ignorance Is Strength](#):

Energy is an active area of EU public policy. Yet authorities are not revealing information (data is surely has) that is crucial to determine whether its policies are distorting the market and come at too high a cost to society. This is a major fault in Europe's credibility in advancing its policy goals, as well as a serious limitation to the accountability of the policy making process

We realized that, while strongly supporting green investments the EU does not know, or does not make it public, how much is spent every year on green subsidies... With regard to green jobs, several estimates exist, but no official figure is provided.

More recently... I discovered that Eurostat does not tell how much coal capacity is installed - as opposed to natural gas- or oil-fueled generation plants. It is possible to know how much coal is used, but not the amount of fixed capital which is invested in

coal plants. If data are not available, every conclusion is questionable because it relies on assumptions or estimates.

2.3. Open Data in Latin America, Asia and Africa

Several countries in Latin America are studying and making experiments with Open Data both at the government and at the grassroots level. The same is happening, on a much smaller scale, in a few parts of Asia and Africa. On average, the volume of these Open Data experiments and the level of *local* interest and awareness around them is still lower than what is happening in Europe and North America. In spite of this we suggest that it is important, for public officials and civic activists in Western Countries, to follow these developments closely. The reason is that they may turn into very useful test beds for all the strengths and limits of Open Data, especially those not encountered yet where the movement was born.

In fact, the original discourse and arguments around Open Data are heavily Western centric. The problem they want to solve is how to make democracy work better *in countries where it already exists and which share a great amount of history and cultural/philosophical values*.

Other countries face very different challenges, from the philosophical level to the practical one. A common issue in developing countries, for example, is that there is very little to open simply because much PSI (Public Sector Information) doesn't exist in digital format yet. Therefore, the first thing to do is to *create* data, normally through outsourcing and crowd sourcing.

Other issues, that will be discussed in detail in other sections of the report because they are also present in Europe in different forms, are related to lack of equal opportunities for access to data and serious fears (sometimes, concrete, sometimes caused by confusion about what should be open and how) that data will be used *against* citizens. A commenter to Gurstein's [Open Data: Empowering the Empowered or Effective Data Use for Everyone?](#) said:

in Delhi and Mumbai, mobs and rioters managed to get information about particular identity groups through voter rolls: openness is, in certain situations, a precarious virtue. It is almost certain that Open Data would be used to rig election but here again openness is not the issue, they would find it anyway...

So far, the main interest about Open Data in Asian countries seems limited, so to speak, to its effects on transparency in politics. At a two-weeks programming contest held at the end of 2010 in Thailand, for example, one of the most appreciated entries was a software scraper of the Thailand's Member of House of Representative Website, that [made it possible for everybody](#) to create applications using those data.

Right now, one of the most active Asian countries in the Open Data arena is India, which also signed an [Open Government partnership with the USA](#) in November 2010. In January 2011 the Indian Congress Party announced plans for a new law to fight corruption among public servants and politicians. Anti-corruption websites (including ones in local dialects) like [Indiaagainstcorruption.org](#), already existed, including one, [Ipaidabribe.com](#), that collected more than 3,000 people reports of graft in its first four months.

As it happens in Asia, even Latin America is currently focused, at least outside Public Administration circles, on how to open public data to achieve actual transparency. This appears even from the way many projects are labeled, that is "[Civic Information](#)" instead of Open Data (which is an idea starting from data *reuse*) or Open Government.

The reason is that even where good Freedom of Information laws exist in Latin America, they still have too little practical effects. Mexico, for example, already has a digital system to manage Freedom of Information requests, but there are reports of complaints filed against municipal officials that either have no effect at all, or aren't possible in the first place, because relevant information has not been updated in years, or omits key data like (in the case of budget reports) "*descriptions of how the money was spent*".

Even with these difficulties, the Latin America Open Data/Civic Information landscape is active and definitely worthwhile following. The list of interesting Civic Information projects in Latin America include (from Sasaki's [Access to Information: Is Mexico a Model for the Rest of the World?](#)):

- Mexico
 - [Mexican Farm Subsidies](#) - an online tool to analyze how the federal government allocates those subsidies
 - [Compare Your School](#): compares aggregate test results from any school with the municipal, regional, and national averages
 - [Rebellion of the Sick](#) built for patients with chronic diseases whose expenses are not covered by the government subsidized health coverage.
- Argentina: [Public Spending in Bahía](#) analyzes how public funds are used.
- Colombia: [Visible Congress](#) monitors the actions of the Colombian congress
- Brazil
 - [Eleitor 2010](#): a website to submit reports of electoral fraud during the Brazil 2010

elections

- [Open Congress](#): a tool for political scientists to track the work and effectiveness of the Brazilian congress
- Paraguay: [Who Do We Choose?](#): lists profiles of all candidates for many public posts.

In Brazil, the principle that *"what is not confidential should be available on the Internet in the open data format"* is already discussed and, in principle, accepted, by some departments of the Brazilian federal government. However, the preferred practice for now is (if there are no other obstacles) to only publish data that have been explicitly requested by some citizens.

A report presented in May 2011 at the [First Global Conference on Transparency Research](#) mentioned a couple of Open Data issues in Latin America that are worth noting, because they're present even in Europe and North America, in spite of the different historical and social background:

- "Better coordination is needed between right to information campaigners and open data activists."
- "If activist manage to target particular topics to add "value" to the discussion, demand for open data could eventually increase in the region."

In Africa, mobile phones are much more available, and more essential than computer with Internet access, often bypassing the need for real desktop PCs with many applications. Therefore, from a purely technical point of view, transparency, accountability and efficiency in government are quickly becoming accessible to most African citizens through mobile networks rather than through the "traditional" Internet. However, there are still too few public departments and procedures that use digital documents and procedures on a scale large enough to generate meaningful volumes of digital data that could be then published online.

While we write, Kenya is laying the legal groundwork to support Open Data. Permanent Secretary for Information and Communications, Dr. Bitange Ndemo is reported as having been championing for quite some time. In practice, big challenges remain for Open Data usage in Kenya. The easiest one to solve is to technical, that is find skilled people that can package the data in ways that the public can consume (even on mobile phones...). The real problem, however, is the fact that (summarizing from [Thinking About Africa's Open Data](#)):

There is a lot of Kenya data but it isn't accessible. The entities that hold the most public and infrastructure data are always government institutions. Getting information from them can be very hard indeed. We don't know who to go to to get the data we need, and

there is no mandate to support one group to centralize it.

Kenya's own OpenData.go.ke website has only ever seen a small handful of data sets, none of which are now (early April 2011) available anymore. Groups like the Ministry of Education might publish some information on schools, but they won't give anyone the location data.

3. Emerging trends and issues related to Open Data

One of the most common activities for Open Data activists in this moment is the creation of country-wide catalogs of all data sources, to facilitate individuation and correlation of independent data sets. Normally, all initiatives of this type are announced on the Open Knowledge Foundation [blog](#) and/or its data hub [CKAN](#). Another relevant development is the publication of an [Open Data Manual](#) that *"can be used by anyone but is especially designed for those seeking to open up data, since it discusses why to go open, what open is, and the how to 'Open' Data."* Activists in several European countries have already published local versions of the manual, or equivalent documents. On this background, several interesting issues, some of which were anticipated in the Open Data, Open Society report, are coming in full light. They are presented, one at a time, in the following sections of this chapter.

3.1. Cost of not opening PSI is increasing

Much has been said on the *economic* benefits of opening public sector information, and much more remains to be said and studied. One part of this issue that is becoming more evident over time is that Open Data are the simplest, if not the only way, to save Public Administrations from the costs that they have *already* (and rightfully!) forced themselves to bear, through assorted laws and official regulations. This is explained well in the report from LinkedGov about the [economic impact of open data](#):

(p. 2) "As the costs of disseminating and accessing information have declined, the transactions costs associated with charging for access to information, and controlling subsequent redistribution have come to constitute a major barrier to access in themselves. As a result, the case for free (gratis) provision of Public Sector Information is stronger than has already been recognized."

Eaves provides a practical example from Canada in [Access to Information is Fatally Broken... You Just Don't Know it Yet](#): *the number of Access to Information Requests (ATIP) has almost tripled*

since 1996. Such growth might be manageable if the costs of handling each requests was dropping rapidly, but it has more than quadrupled.

Unfortunately, alternatives like charging for access to data or cutting the budget for providing them to citizens remain very common in spite of their negative effects. According to Eaves, the first practice has already caused *a reduction* in the number of freedom of information requests filed by citizens, while budget cuts invariably and greatly delay processing times.

3.2. Creative, unforeseen uses of local Open Data increase

Proofs that, as cited in the Open Data, Open Society report, "Data is like soil", that is valuable not in itself, but because of what *grows* on it, often in ways that the landowner couldn't imagine, continue to arrive. Here is an example from [Day Two: Follow the Data, Iterating and the \\$1200 problem](#):

*Ed Reiskin noticed a problem with street cleaning. Some trucks would go out, coming back with little or no trash depending on the day and route they took. After getting the tonnage logs, his team quickly realized that changing certain routes and reducing service on others would save money (less gas, parts, labor) and the environment (less pollution, gas consumption, water). A year later, the department realized a little over a million dollars in savings. The point? **Follow the data.***

The value embedded in data isn't only economical or political, but also social. Here are a few examples.

At the Amsterdam fire brigade, once a fire alarm starts, [all sorts of data is collected](#), to maximize the probabilities to save lives and property, about the location and the route to the emergency: constructions on the way, latest updates from OpenStreetMap, the type of house and if possible more data such as construction dates, materials, people living there and so on.

Using the geographical coordinates embedded in online photo databases like Flickr, digital cartographer Eric Fischer creates maps that highlight people behavior. For example, he documented how, in Berlin, most locals tend to stay in the same neighborhoods and don't go to West Berlin or to the outskirts of the city. This information has economic value, journalist Kayser-Bril noted: "*You can then sell this for instance to businessmen who want to open a shop in Berlin for tourists, and telling them where to go and where not to go.*"

Norwegian transport company Kolumbus has embedded 1,200 bus stops with barcodes in the square QR format, that can encode text or URLs. Scanning those codes with a free software application for smartphones loads a website that lists upcoming bus departure times. Later, Kolumbus [partnered](#)

[with a project called "Tales of Things"](#) to allow people to leave messages for each other (or just for the world) at the bus stops. Scanning the QR code now allows people to see not just the bus timetable, but also the notes other travelers have left on that stop, including *"what's nearby, who's waiting for whom, what number can you call for a good time. It's a cross between bus stop Facebook and digital graffiti"*, that happened thanks to the openness of the original bus stop data.

The [Social Life of Data Project](#) will study instead how particular datasets have been used, who used them, how those people are connected and what conversations happen around Open Data.

3.3. Legal issues remain crucial

Proper licensing of Public data is essential. The more Open Data activities continue, the clearer this rule becomes. What distinguishes Open Data from "mere" transparency is reuse. Paraphrasing Eaves, until a government get the licensing issue right, Open Data cannot bring all the possible benefits in that country. If there are no guarantees that public data can be used without restriction, very little happens in practice, and when it happens it may be something against the public interest.

Canadian Company Public Engines Inc, that is paid by local police departments to collect, process and analyze official crime data, also publishes online, with a proprietary license, anonymized summaries of those data. When in 2010 another company, Report See Inc, scraped those data from their website to reuse them, Public Engines sued.

Reporting this, D. Eaves [rightly points out](#) that *both* companies are right: one is trying to protect its investment, the other is simply trying to reuse what IS public data, by getting it from the ONLY place where it's available. This is what happens when public officials leave the ownership of *public* data to the third parties hired to collect them. Please note that, in practice, it makes very little difference whether those third parties are private, for-profit corporations or even other Public Administrations. Unless, of course, there are national laws already in place that define in advance what is the license of all present and future Public Data, *no matter how they were generated and by whom*, those data can be lost in any moment for society. In all other cases, the legal status of data will be either officially closed and locked, or uncertain enough to prevent most or all reuses. In February 2011, the [news came](#) that, even if they weren't the original copyright holders, Public Engines had been able to put together enough legal claims to convince Report See to give up.

Disputes like this should not happen and would not happen if all contracts regarding collection and management of PSI clearly specified that all the resulting data either go directly into the public domain (after being anonymized if necessary, of course) or remain exclusive property of the

government. Even ignoring data openness, this is essential for at least three other reasons. The first is to protect a public administration from having to pay *twice* for those data, if it needs it again in the future for some other internal activity, not explicitly mentioned in the initial contract. The second reason is to not spend more than what is absolutely necessary to respond to public records requests, that is to comply with Freedom of Information laws.

The final reason is to guarantee quality assurance and detection of abuses at the smallest cost, that is sharing it with all the citizens using the public services based on those data. A [real world example](#) of this point comes from the "Where's My Villo?" service in Brussels. Villo! is a city-wide bike-sharing scheme started in May 2009, through a partnerships with a private company: JCDecaux finances the infrastructure and operates it, in exchange for advertising space on the bikes themselves and on billboards at the bike sharing stations. The availability of bikes and parking spaces of each station is published online in real time on the official Villo's website.

When the quality of service decreased, some citizens started "Where's My Villo?", another website that reuses those data to measure where and how often there aren't enough available bikes and parking spaces, in a way that made it impossible for JCDecaux to deny the problems and stimulated it to fix them. Both this happy ending and the fact that it came at almost no cost to the city, because citizens could monitor the service by themselves, were possible just because the data from the official website were legally and automatically reusable.

3.4. The price of digitization

In practice, public data can be opened at affordable costs, in a useful and easily usable way, only if it is in digital format. As a consequence of this fact, demand for Open Data exposes a problem that already existed and must be fixed anyway, regardless (again) of openness. Any substantial increase of efficiency and reduction of the costs of Public Administrations can only happen when data and procedures are digitized. The problem is that such digitization (which, obviously, must happen anyway sooner or later) can be very expensive and we are only now starting to really realize how much. Actual, material costs are not the worst problem here. Activities like semi-automatic scanning of paper documents or typing again their content inside some database, are relatively low, one-time expenses that are also very easy to calculate and budget in advance with great precision.

The real costs are those at the social, cultural, historical and workflow reorganization level. What is really difficult, that is expensive in ways that are hard to predict, is to fit inside digital, more or less automatic procedures and file templates, formats, habits and customs developed, maybe over

several centuries, in the analog, pre-computer world. Developing countries are good case studies from this point of view, because they are often leapfrogging from oral tradition straight to computers in all fields, not just e-government.

Land ownership in India, discussed by Gurnstein in 2010, is a perfect example of the problems carried by digitization that requests for Open Data only expose, without creating them. Digitization can certainly increase efficiency, transparency and economic activities, but fully achieves these goals only by:

- standardizing as much as possible all concepts, formats and procedures.
- replacing *completely*, at least in standard day to day procedures, whatever other records and ways of working existed before

Gurnstein wrote:

"The problem of open access in the case of land records in India is... the manner in which the data tends to get encoded. Typically, digitization of land records would mean either scanning the record as it is, or inputting all the data on the record as it is, without changing any fields. But ways of maintaining land records are highly diverse... Private ownership is not the only means of holding a land parcel. When it comes to land ownership, for example, it may eliminate the history of land, how were subdivisions and usufruct rights negotiated and enforced."

Another risk of digitization and e-government (without openness, that is) is lack of contact between citizens and institutions:

"Prior to digitization, land records in India were available to people who made requests with village accountants for them. .. after digitization of several services, village accountants no longer personally visit the villages they are in charge of... What has happened with digitization is a reorganization of earlier forms of social and political relations. Accountability has moved from the immediate village level"

Of course, all these problems existed well before computers and return every time the political or social order changes. The demand for Open Data is only increasing, by orders of magnitude, the numbers of times in which we meet them.

3.5. The nature of Open Government and the relationship between citizens and Government

Open Data are an essential part of Open Government. Almost everybody agrees with this. Agreement on what exactly defines Open Government is, however, less universal. In January 2011 Lucas Cioffi, replying to Alex Howard, [wrote](#):

The biggest difference between Gov 2.0 and OpenGov seems to be how they approach transparency. Gov 2.0 is about transparency through open data and the "government as a platform" idea. "Open Government" is about Transparency for the sake of accountability, but not necessarily interaction, cooperation and reuse of data outside the government.

[who advocates] Open Data does so in order to make it accessible to citizens rather than to hold government accountable. This is not to say that one approach is better than another, but this is to say that there seem to be two very different motivations for advocating for transparency, and they do seem to correlate to whether people label themselves as part of Gov 2.0 or part of OpenGov.

In general, reflection and debate on this point is accelerating. At the moment, some characteristics of Open Government on which there is more or less agreement are that Open Government is about:

- deliberation, choice, influence on decisions and participation as a common citizen
- letting *all* citizens use technology to participate, monitor and define government activities. In other words, Government is really Open when it's based on interaction, not only on some set of infrastructures and methods imposed top-down
- diffused, seamless conversations, that are only possible with digital technologies, online social networks and so on, between public employees and citizens.

The obvious potential limit of these definitions is that they rely on a big, still largely unknown factor, that is actual citizen participation. When data are opened, the problem becomes to have everybody use them, in order to actually realize Open Government as defined above. This issue will be explored in detail in the next paragraphs, but we can already say that Open Data are highlighting the critical, weak points in the present and future relationship between citizens and governments.

While citizens participation is essential, especially in times of social and economic crisis, achieving it on a large scale won't be easy. Frustration and lack of trust in institutions in many countries are high, so it's no surprise when people express doubts that opening government data won't help much in fixing things.

3.6. Clearer vision of the real risks and limits of Open Data

Open Data, we already said, is about reuse. The point is, at least when the goal is Open Government and transparency in politics, reuse by whom? There is no *automatic* cause-effect relationship between Open Data and real transparency and democracy. On the contrary, several problems may occur, if administrators and citizens don't pay close attention.

3.6.1. Data alterations and financial sustainability

Some concerns about the limits of Open Data are about what may happen, or stop to happen, *before* they are published online. The most common concerns of this type are (from [Open Public Data: Then What? - Part 1](#)):

1. Opening up PSI causes those data to not be produced anymore, or to be only produced as private property by private corporations, because the public agencies whose job was to produce those data, can't sell them anymore.
2. total accessibility of data provides more incentives to tinker with them, at the risk of reducing trust in institutions and inhibiting decision-making even more than today.

Data manipulation is the topic of the next paragraph. Speaking of costs, a point to take into account is that, once data are open, routinely used and monitored by as many independent users as possible, even the cost of keeping them up to date may be sensibly reduced: in other words, in the medium/long term Open Data may reduce the need to periodically perform complete, that is very expensive, studies and surveys to update a whole corpus of data in one run.

Besides, and above all, even if opening data always destroyed any source of income for the public office that used to create and maintain them, this problem would only exist for the PSI datasets that are *already* sold today. Such data, even if of strategic importance as is the case with digital cartography, are only a minimal fraction of all the PSI that could and should be opened to increase transparency, reduce the costs of Government and stimulate the economy. In all these other cases:

- the money to generate the data already arrives by some other source than sales and licensing (but even with those data it may be possible to generate them by crowdsourcing, thereby reducing those costs!)
- the only extra expense caused by publishing those data online (assuming they're already available in some digital format, of course!), would be the hosting and bandwidth costs, that may be greatly reduced by mirroring and other technical solutions like torrents, already widely used to distribute Free/Open Source Software (FOSS) through the Internet.

3.6.2. Real impact of data manipulation or misunderstanding

The fix for the risk that data is manipulated is to not only open government data and procedures, but to simplify the latter (which eventually also greatly reduces cost) as much as possible. Abundance of occasions to secretly play with data and how they are managed is a symptom of excessive, or peak complexity: again, problems and risks with Open Data are a symptom of a [pre-

existing] problem that is somewhere else.

Regardless of the real probability of data alterations before they are published, the major problem happens after. We already mentioned in the first report the fact that, while correct interpretation of public data from *the majority of average citizens* is absolutely critical, the current situation, even in countries with (theoretical) high alphabetization and Internet access rates, is one in which most people still lack the skills needed for such analysis. Therefore, there surely is space for both intentional manipulation of PSI and for misunderstanding it. After the publication of the first report, we've encountered several examples of this danger, which are reported in the rest of this paragraph.

Before describing those cases, and in spite of them, it is necessary to point out one thing. While the impact on the general public (in terms of raising interest and enhancing participation) on the Open Data activity of 2010 is been, in many cases and as of today, still minimal, it is also true that there has been no big increase in demagogy, more or less manipulated scandals and conflictual discussion caused by Open Data. There has certainly been something of this in the Cablegate but that's not really relevant because, as we've already explained, what Wikileaks did is intrinsically different from Open Data. So far, negative or at least controversial reactions by manipulation and misunderstanding of Open Data haven't happened to such a scale to justify not opening PSI.

This said, let's look at some recent example of misunderstanding and/or manipulation based on (sometimes open) public digital data.

Nicolas Kayser-Bril [mentioned](#) a digital map of all the religious places in Russia, that shows [also] *"mosques that are no longer in use, so as to convey the idea that Muslims were invading Russia."*

In September 2010 the Italian National Institute of Geophysics and Vulcanology [officially declared](#) in September 2010 that they were evaluating whether to stop publishing online Italy's seismic data, as they had been doing for years. The reason was that, following the March 2009 earthquake in Italy, the data were being used to *"come to conclusions without any basis at all"*, both by the press, to sell more, and by local politicians trying to hide the lack of preventive measures, like enforcing anti seismic construction codes.

Still in Italy, Daniele Belleri runs a Milan crime mapping blog called ["Il giro della Nera"](#), making a big effort to explain to his readers the limits of the maps he publishes, and the potential for misunderstanding if they are used without preparation, or with wrong expectations. This is a synthesis of Belleri's explanation, also covered in [other websites](#), that is applicable to any map-

based PSI analysis and presentation, not just to crime mapping:

In general, a map is just a map, not reality. It doesn't always and necessarily provide scientific evidence. Crime maps, for example, are NOT safety maps, as most citizens would, more or less consciously, like them to be: a tool that tells them where to buy a house their according to the level of criminality in the district.

When used in that way, crime maps can give unprepared users two false impressions: the first, obvious one, is that certain areas are only criminal spaces, exclusively inhabited by criminals. The other is to encourage a purely egoistic vision of the city, where the need for safety becomes paranoia and intolerance and all that matters is to be inside some gated community. This doesn't lower crime levels at all: the only result is to increase urban segregation.

To make things worse, crime data not analyzed and explained properly don't just contribute to strengthen egoistic attitudes and lock the urban areas that are actually the most plagued by crime into their current difficult state indefinitely. Sometimes, they may even perpetuate beliefs that are, at least in part, simply false. Of course, when those beliefs not grounded in facts already existed, open crime data can help, by finding and proving the gaps between perception of criminality and reality. Belleri, for example, [notes](#) that residents of Milan consider the outskirts of their city more dangerous than downtown Milan, while Londoners think the opposite about London... but in both cities the truth emerging from data is exactly the opposite (at least for certain categories of crime) of what their residents believe.

3.6.3. Unequal access

Even ignoring crime mapping, in some worst case scenarios, data openness may be not only hindered by social divisions, but also create or enhance them. If citizens can't find and recognize real, relevant *meaning* and practical value in data, as well as way to use them to make change happen, there won't be any widespread, long lasting benefit from openness. How can we guarantee, instead, that such meaning and value will be evident and usable? What are the ingredients for success here?

Enhancing access to PSI it's harder than it may seem because it isn't just a matter of physical infrastructure. It is necessary that those who access Open Data are in a position to actually understand them and use them in their own interest.

This is far from granted also because, sometimes, the citizens who would benefit the most from certain data are just those, already poor, marginalized and/or without the right education, who have the least chances to actually discover and be able to use them. This is what G. Friedman was

speaking about when, in September 2010, he wrote about the [great divide caused by Open Health Data](#):

[in the USA] "statistically speaking, chronic disease is associated with being older, African American, less educated, and living in a lower-income household. By contrast, Internet use is statistically associated with being younger, white, college-educated, and living in a higher-income household. Thus, it is not surprising that the chronically ill report lower rates of Internet access.

Starting from this, and commenting a study of the performances, with respect to coronary artery bypass grafting, of several medical centers, Frydman expressed his concern that:

the empowered will have access to [this data] and will act upon it, while many of the people suffering from chronic diseases (the same population that would benefit most from access to this information) won't. Over time it is therefore probable that the current centers of excellence will treat an ever growing number of empowered while the centers that currently experience high mortality rates will get worse and worse result, simply because they will treat an ever growing number of digital outliers who haven't the possibility to obtain health data and apply filters.

Since one of the topics of this project is the *economic* value of Open Data, it is necessary to add a somewhat obvious observation to Frydman's concerns (regardless of their probability). Even if it is difficult now to make accurate estimates, such negative developments would surely impact also the costs of health services and insurances, not to mention healthcare-related jobs, both in the communities hosting centers of excellence and in those with the worst ones.

3.6.4. Lack of education to data

Boris Müller, professor for interface and interaction design at the University of Applied Sciences in Potsda, said in [an April 2011 interview](#): *"I think that really a citizen needs to know how visualizations work in order to really evaluate the quality of the data and the quality of the evaluation."* As data visualization and analysis becomes more popular easier to use (even as a tool for manipulating the public opinion), it's important for the public to:

- understand that, before becoming digital, information was coded, stored and used in many ways, through social norms and human interactions more complex than computer ones (cfr the digitization of India land ownership records), therefore making exact, one-to-one equivalence between analog and digital procedures hard or impossible in many cases
- think critically about where data comes from
- *remember* to always follow the *development* of data-based stories, or accusation.

Here's an example of why the two last things are important. In April 2011, during a prime time TV talk-show, Italian MP Enrico Letta asked Education Minister Gelmini to justify further cuts to Public Schools declared in the new State budget. Gelmini knew nothing about such cuts to the budget of her own Ministry, so all she could reply at the moment was that Letta's assertions were inconsistent.

Two days later, two bloggers "proved" that Gelmini was right and Letta's analysis wrong because he had cited gross figures instead of net ones and ignored that school budget cuts from 2012 onwards were not new at all, but had been already approved in 2008. Right after this debunking, a third blog [asserted](#) that *everybody* was wrong: Letta, Gelmini and also the first two bloggers who, for unknown reasons, had associated to the Education budget alone all the cuts to the whole public sector, and then based all their calculations on a different (and wrong) summary table, not the one used (still wrongly, but for other reasons) by Letta.

As far as we're concerned, the real issue here is not who was right and why, exactly, all the others made certain mistakes. The actual problem is: how many of the people who saw Gelmini unprepared on TV the day this case started *also* followed up the story in the next days and found out that things weren't exactly as they had looked in that talk show, even if Letta had "proved" his case with actual, exact "data"? How many citizens are educated to follow the analysis of some data *over time*?

3.6.5. Lack of public interest

After the October 2010 Government Open Source Conference in Portland, John Moore [reported](#) the surprise, among participants, *that people were not demanding more open data, that the push had not yet come from public*. If Open Data is about empowerment, transparency and saving public money, why aren't more common citizens already very excited about Open Data? Part of the answer is the already mentioned cynicism and lack of trust in institutions and in the possibility for individuals to participate effectively to politics and administration. Too many citizens still don't feel that it is their right to seek public information from their representatives and administrators, or that doing so will make any practical difference.

Another part of the problem is poor marketing from data activists and Public Administrations, that should start to act more like product developers, that is measure the outcome of their activity in terms of what has more appeal for the general public. One way to achieve this, especially at the local level, may be to highlight (only) the concrete cost savings and local jobs directly created by

the availability of Open Data. Of course, this isn't always possible.

3.6.6. Unprepared Public Administrators

It is undeniable that today, especially at the local level, most Public Administrators that should or may contribute to open the public data held by their organizations still ignore, and sometimes disdain, Open Data proposals, principles and practices. This happens for many reasons. We'll only mention two of them that are quite common. They are interesting because, while being somewhat related and sharing common origins, one is very hard to fix, the other, at least in comparison, very easy.

To begin with, most of these administrators are people that, albeit very competent and committed to their work, were not really trained to live with so much of what they perceive as "their" documents and daily activities as Open Data implies regularly exposed to the public. This is true even among administrators who are already well acquainted with mainstream "Web 2.0" practices. Many officers who already have a regular presence on Facebook, Twitter or other social networks and regularly use those platforms to discuss their work with their constituents feel diffident about Open Data in the same measure as their colleagues who don't even use computers yet. A cultural barrier like this requires both strong demand from citizens and detailed examples of how Open Data can be good for the local budget to be overcome in acceptable time frames.

Another factor that may keep administrators away from Open Data is the more or less unconscious assumption that, in order to use them, a City Major or Region Governor should be very skilled himself, if not with actual programming, with "Web 2.0" tools, modern online services and/or general software engineering principles. This is simply not true. Surely, Open Data is something that is made *possible* only by modern digital technologies and the Internet, but at the end of the day it's "simply" a way to increase transparency, efficiency and cost reductions inside Public Administration, and to create local jobs. If these hypotheses are as concrete as this and many other studies explain, there is no need for a Major to have programming skills, like social networks or have any other personal "2.0" skill or training to see the advantages of Open Data and delegate to his or her IT staff their implementation.

3.7. The privacy problem

Being perceived as a lethal attack to privacy remains one of the biggest misunderstandings that prevents adoption of Open Data. On one hand, there is no doubt that in an increasingly digital world it becomes harder and harder to protect privacy. But, exactly *because* the whole world is going

digital, attacks to privacy and to civil rights in general can and are coming by so many other sides that those from (properly done) Open Data are a really tiny percentage of the total.

This is a consequence of the fact that data about us end up online from the most different sources (including ourselves and our acquaintances), and that often it would be very hard to discover, never mind *prove*, that they've been used against our interest. There have been concerns, for example, that insurance companies may charge higher fees for life insurance to those among their customers who... put online a family tree from which it shows that they come from families with an average life expectancy lower than usual.

Assuming such concerns were real, would it always be possible to spot and prove such abuses of data, that weren't even published by any Public Administration? Of course, publishing online complete, official Census data of several generations, in a way that would make such automatic analysis possible would be a totally different matter.

Getting rid of all the unjustified concerns about privacy is very simple, at least in theory. All is needed to dismiss for good the idea that Open Data is a generalized attack to privacy is to always remember and explain that:

1. Most Open Data have nothing personal to begin with (examples: digital maps, budgets, air pollution measurements....)
2. The majority of data that are directly related to individuals (e.g. things like names and address of people with specific diseases, or who were victims of some crime) have no reason to be published, **nor there is any actual demand for them by Open Data advocates**
3. Exceptions that limit privacy for specific cases and categories of people (e.g. candidates to public offices, Government and Parliament members etc...) already exist in many countries
4. Very often, in practice, Open Data struggles only happen about *when and how* to make available in the most effective way for society information that was *already* recognized as public. *What* to declare public, hence open, is indeed a serious issue (more on this in the next paragraph) but is a separate one.

3.8. Need to better define what is Public Data

Together with citizens education, there is a huge challenge that Governments and the Open Data movement will have to face (hopefully together) in 2011 and beyond. This challenge is to update and expand the definition of Public Data and to have it accepted by lawmakers and public administrators.

What is, exactly, Public Data? A definition that is accepted almost implicitly is "*data that is of public interest, that belongs to the whole community, data that every citizen is surely entitled to know and use*". This definition is so generic that accepting it together with the assumption that all such data should be open as preached by the Open Data movement (online, as soon as possible, in machine readable format with an open license etc...) doesn't create any particular problem or conflict.

Real problems however start as it has happened all too often so far, whenever we assume more or less consciously that "Public Data" in the sense defined above and data directly produced by Governments and Public Administrations, that is what's normally called PSI (Public Sector Information) are the same thing.

There is no doubt that Governments and Public Administrations produce huge quantities of Public Data. But this is an age of privatization of many public services, from transportation to healthcare, energy and water management. This is an age in which many activities with potentially very serious impacts on whole communities, like processing of hazardous substances or toxic waste, happen *outside* Public Administrations. The paradox is that, as [Sasaki put it](#), this increased privatization is happening in the very same period in which "*we are observing a worldwide diffusion of access to information laws that empower citizens to hold government agencies accountable.*"

In such a context, "Public Data" is critical just because it is a much bigger set of data than what constitutes traditional, official PSI. "Public Data" includes all that information *plus* the much bigger amount of data describing and measuring all the activities of private companies, from bus timetables to packaged food ingredients, aqueducts performances and composition of fumes released in the atmosphere, that have a *direct impact* on the health and rights of all citizens of the communities affected by the activities of those companies.

Are such data "Public" today, in the sense defined at the beginning of this paragraph, that is something every citizen has the right to know without intermediaries or delegates, or not? Should they be public? If yes, shouldn't law mandate that all such data be Open (that is, published online as soon as possible, in machine readable format with an open license etc...) just like, for example, the budget of some Ministry? Answering these questions may be one of the biggest challenges for the Open Data community, and for society as a whole, in the next years.

Here are, in order to facilitate reflection on this issue, a few recent, real world examples of "Public Data" that are *not* PSI, and of the impacts of their lack of openness.

In April 2011, John Farrell [wrote](#):

solar and other renewable energy developers must find the best places to plug in to the grid, e.g. where demand is high or infrastructure is stressed. The cost to connect distributed generation may also be lower in these areas. Unfortunately, data about a utility's grid system is rarely public.

California utilities are changing the game. Southern California Edison (SCE) rolled out a map of its grid system, highlighting (in red) areas that "could potentially minimize your costs of interconnection to the SCE system." Since as much as a third of the cost of PV can be recaptured via its benefits to the electric grid when properly placed in the distribution system, having this information is crucial for solar developers. *Public data also levels the playing field between independent power producers and the utilities, since the latter can use federal tax credits and their proprietary knowledge of the electric grid to build their own distributed renewable energy at the most attractive locations.*

Having public data on distribution grid hot spots can make renewable energy development more cost effective and more democratic. Tell your utility to publish its map.

This, instead, is an excerpt of [This Data isn't dull. It improves lives](#) (March 2011, New York Times) that looks at public transportation and consumer safety:

The USA Department of Transportation is considering a new rule requiring airlines to make all of their prices public and immediately available online. The postings would include both ticket prices and the fees for "extras" like baggage, movies, food and beverages. The data would then be accessible to travel Web sites, and thus to all shoppers.

The airlines would retain the right to decide how and where to sell their products and services. But many of them are insisting that they should be able to decide where and how to display these extra fees. The issue is likely to grow in importance as airlines expand their lists of possible extras, from seats with more legroom to business-class meals served in coach.

Electronic disclosure of all fees can make it much easier for consumers to figure out what a trip really costs, and thus make markets more efficient, without requiring new rules and regulations.

Another initiative has been proposed by the Consumer Product Safety Commission. In 2008, Congress overwhelmingly passed and President George W. Bush signed legislation mandating an online database of reported safety issues in products, at saferproducts.gov. The Web site ran for a few months in a "soft launch" and went into full operation on Friday.

Thirteen years ago, two parents were told that their 18-month-old son had died in an accident in a model of crib in which other children had died, yet there was no easy way for any parent or child-care provider to know that.

What about food? Here is what [Christian Kreutz](#) said in January 2011:

Nutrition is another interesting sector to use open data, which I discovered lately. A last example for food is the whole potential behind bar code scanning - you take your mobile phone to the supermarket and scan products to get the information behind the fair trade certificate or behind the company. In the recent dioxin scandal in Germany, the company Barcoo took information from the ministry of agriculture in Germany, of which farms have intoxicated eggs and offer the info in their app. So, you can check in the supermarket the eggs that are fine and not with your mobile phone.

Food in supermarkets is only one of thousands cases of "Public Data" from a strategic sector of the economy that is huge, essential for creation of local jobs and in deep crisis in many countries in this period: traditional, brick and mortar retail and service businesses.

Consider this explanation by venture capital firm Greylock about [why they Invested in Groupon: The Power of Data](#)

Groupon is targeting a market that is huge and broken. Local advertising is a \$100 billion annual business in the U.S. and consumers spend something like 80% of their disposable income within a couple miles of their homes. Many local businesses still try to attract new customers through that heavy yellow book that gets dropped on your front doorstep until it rots or gets tossed in the recycling bin.

We think the technologies visible to consumers will be increasingly commoditized, while the data used to understand consumers better will become increasingly proprietary and valuable.

Offers to consumers can be intelligently served up based on a person's demographics, buying history and location. The merchant side of the equation is just as interesting. Local businesses need to be able to do more than just run a sale once or twice a year. The theater on Main Street or the children's museum across town should have the ability to revenue optimize, like United Airlines or Hilton, by appropriately pricing and marketing unsold capacity. We started really leaning forward in our chairs when the discussion turned to strategy, including the ways to use data to power Groupon's future consumer- and merchant-facing products.

We believe Groupon is the break-out leader in the massive local commerce space and its investment in data will be a critical ingredient in its long term march to build a meaningful and foundational company.

Groupon is the clear market leader in the local deals market in 2011. However, complaints from merchants about the money they can loss by offering deals via Groupon [already exist](#). Now, couldn't all the "local deals" raw information be considered as Public Data that merchants could (be trained to) directly publish themselves online, in ways that would allow everybody, not just Groupon, to present the deals to customers in ways more profitable for merchants? The point is, how many merchants, merchant associations and majors (whose budgets always and immediately

benefit when local businesses make more money) are aware of this opportunity?

4. Conclusion: seven Open Data strategy and best practices suggestions

Starting from the trends and conclusion described in the previous chapter, this section lists, in the most synthetic way possible, some strategic actions and best practices for 2011, that we consider important in making Open Data succeed and bring the greatest possible benefits to all citizens and businesses.

4.1. Properly define and explain both Open Data and Public Data

Just because Open Data is becoming more popular (and, we may say, more and more necessary every year), it is essential to intensify efforts to explain, both to the general public and to public administrators, that

1. **Privacy issues are almost always a non-issue.** Quoting from [What "open data" means - and what it doesn't](#)): *Privacy and/or security concerns with putting all the government's data out there are a separate issue that shouldn't be confused with Open Data. Whether data should be made publicly available is where privacy concerns come into play. Once it has been determined that government data should be made public, then it should be done openly.*
2. Defining as Public and consequently opening them in the right way, *much more data* than those born and stored *inside* Public Administration is an urgent task that is in the best interest of all citizens and businesses

4.2. Keep political issues separated by economics ones

Open Data can reduce the costs of Public Administrations and generate (or at least protect, as in the case of deals from local merchants) local jobs in all sectors of the economy, not just high-tech ones. There seems to be enough evidence for these two assertions to go for more Open Data *even if* they had no effect at all on participation to politics. This should always be kept in mind, also because some data that can directly stimulate business are not the same that would be useful for transparency.

4.3. Keep past and future separate

For the same reason why it is important to always distinguish between political and economical advantages (or disadvantages) of Open Data, it is necessary to keep decisions about *future* data (those that will arrive in the future, due to new contracts, public services and so on) separate from those about data that already exist. At the end of 2010, T. Steinberg [wrote](#) that the idea that Government should publish everything non-private it can **now** is "rather dangerous", and that it would be much better to release nothing until someone actually asked for it, and at that point doing it right, that is with an open license and so on. The first reasons for Steinberg's concern is that asking for everything as soon as possible would "*stress the system too much, by spreading thin the finite amount of good will, money and political capital*". The second is that many existing old data and data archival systems are, in practice, so uninteresting that it wouldn't make sense to spend resources in opening them.

Even if these concerns were always true, it is important to realize that they apply (especially the second) to already existing data, not to future ones. The two classes of data have, or can have, very different constraints. Existing data may still exist only in paper format and/or be locked by closed or unclear licenses, or not relevant anymore for future decisions.

Opening *future* data, instead, is almost always more important, useful, urgent, easier and cheaper than digitizing or even only reformatting material that in many cases is already too old to make immediate, concrete differences. While this argument is probably not always true when we look at Open data for transparency, it probably is when it comes to economic development.

Therefore, features and guidelines that should be present in all future data generation and management processes include:

- standardization: the less, obviously open, formats are used for data of the same type, the easier it is to merge and correlate them. The formats that have to be standardized are not only those at the pure software level. Even more important is, for example, to adopt by law standard identifiers for government suppliers, names and machine-readable identifiers of budget voices and so on
- preparation for future digitization: new digital systems should explicitly be designed from the beginning so that it will be possible, when non-digital records will be digitized, to add them to the databases without modifying losses.
- Open licenses

- better procurement

The first two features have obvious technical advantages regardless of data openness. The last two, being critical, are discussed separately in the next paragraph.

4.4. Impose proper licensing and streamline procurement

As with the first report prepared for this project, we will not delve into the details of *how* to license data because this topic continues to be followed and debated in all details by LAPSI and other projects or researchers. We will simply confirm the importance of establishing a proper license, at the national level, for all Public Data, that makes them Open in the right way and makes sure that what is opened stays open and that don't demand what isn't possible to enforce (e.g. attribution), because, quoting again [Eaves](#), "*no government should waste precious resources by paying someone to scour the Internet to find websites and apps that don't attribute*".

We want, however, to spend a few words about another legal/administrative side of the issue, that is procurement. Traditional procurement laws are very likely not flexible enough, in most countries, to handle the implementation of data-based public services. Here's why.

We know that if Public Data are Open, everybody, from volunteer activists to hired professionals, can very quickly write or maintain simple software applications that help to visualize and use them in all possible ways. Paradoxically, this is a problem when an Administration either wants to set up an Open Data programming contest (that besides being inexpensive, it's much simpler to organize and join than traditional tenders or grants) or *needs* to just pay somebody to write from scratch and maintain some new program of this type, or customize existing ones.

The reason is that, just because this type of software development is so quick, even hiring a professional to do it, or setting up a contest would be... too inexpensive to be handled with default procurement procedures. Quoting from [Day Two: Follow the Data, Iterating and the \\$1200 problem](#):

A big problem for cities is procuring products under \$10,000. How does a city pay for an awesome application like [SeeClickFix](#) when it doesn't fit the normal year-long planning and two-year implementation in the millions of dollars? In Tuscon, Andrew Greenhill tapped the Mayor's general budget for it, instead of trying to get the IT department to shell out. In San Francisco, Ed Reiskin uses discretionary spending. But every time, procurement gets messy. In reference to nepotism laws, Ed worries that he'll appear "like I'm giving my buddies dollars." Building great products for cities has to include finding great strategies to pay for them. In San Francisco, Jay Nath doesn't even have a budget...which, he says is 'liberating' because he doesn't need to go through

procurement.

The same issue is denounced as an obstacle to innovation and cost savings in [New recommendations for improving local open government and creating online hubs](#):

John Grant focused on a major pain point for government at all levels for tapping into the innovation economy: procurement issues, which civic entrepreneurs run into in cities, statehouses and Washington. "It is time to look at these procurement rules more closely," he said, and promote higher levels of innovation. "There are a lot of ideas are happening but a lot of rules restrict vendors from interacting in government," said Grant. Turner-Lee observed that traditional procurement laws may also not be flexible enough to bring more mobile apps into government.

Current procurement laws aren't partially incompatible with an Open Data world only at this level, that is when it's time to procure software that makes the data useful. Even bigger problems and inefficiencies can be introduced at the beginning of data life, that is when data collection and processing services are procured. We've already explained that forgetting to impose the right license is one of the problems, but it's not the only one. Even future *organization* of all the foreseeable data management activities should take advantage of the flexibility provided by data openness. Here is how Tim Davies [summarizes this point](#):

Right now [public] bodies often procure data collection, data publishing and data interfaces all in one block (as seems to be the case with Oxfordshires real-time bus information - leading to a roadblock on innovation) - and so without these layers being separated in procurement, some of the benefits here stand to be lost.

Changing procurement of information/data-rich public services would be, of course, only the first step of a general reform of procurement laws and regulations. After management of Open Data has been simplified, it becomes time to implement similar simplifications to procurement of everything else. In fact, in such a scenario, there would be much less possibilities for the loopholes, frauds and inefficiencies that forced local procurement procedures to become so slow and complicated: since the public budget and other relevant public data would already be fully open, errors and other problems would surface and be fixed much more quickly and reliably than today, even assuming that they would continue to appear with the same frequency.

4.5. Educate citizens to understand and use data

It is necessary to guarantee the widest possible availability of *all* the pre-requisites for effective use of Open Data. In other words, it is necessary to provide free and widely accessible training, oriented to average citizens, on how and why to visualize Public Data and use them to make informed

decisions. Ideally, this training should be provided at a local level with local programs, in a way that makes it possible to use it on local issues, for the reasons and in the ways discussed in the next paragraph. For example, visualization techniques like those used by ABC News to show the effects of the [March 2011 Japan Earthquake](#), in which all the user has to do to compare scenes from before and after the earthquake is to move a slider, should be routinely used to explain proposals about urban planning, zoning and related topics.

4.6. Focus on local, specific issues to raise interest for Open Data

Considering the continuous evidence and concerns about scarce interest and preparation of citizens to use Open Data in their political, economic and professional decisions, one of the final recommendations of the Open Data, Open Society report confirms its importance and needs to be repeated: it is very effective, if not simply necessary if the goal is to generate a critical mass of citizens that demand and use Open Data in the shortest possible time, to practice all the recommendations of this report *at the local level*,

Most people encounter their local governments much more often than their national ones. When working within a single city or region it is much easier to inform citizens, raise their interest and involve them, because they would be searching *local* solutions to improve *local* services and/or save *local* money. There may also be much more opportunities to do so, especially in this period of financial crisis that will see substantial decreases both in credit by financial institutions and in subsidies from central governments. Concreteness and, as they say in marketing, "customer focus" must be the keys for local activists and public employees working on local Open Data:

- work on specific issues and with precise objectives
- focus on immediate usefulness
- work on demand, on the *services* that people want. Required services define what data must be open, not the contrary

This is the most effective, if not the only strategy, to solve one of the biggest debates in open data: "*how do we get people to use the data that we publish?*". The right question, instead, is "*what data do people want?*". Even if citizens don't realize yet that what they actually want is more Open Data, or that what they need can be done more quickly and cheaply by releasing some information in that way.

A great example of what all this means is the [Great British Public Toilet Map](#): a public participation

website that tracks which councils have published public toilet open data, and which have not. A map like this solves one specific, concrete problem in the ordinary, daily life of many people: *"Many older people have continence concerns and only go to places where they know there is a toilet. "*

It is also possible and useful to pass the message that, when it comes to participation, activism and transparency in politics, Open Data are a concrete and pacific weapon that is both very effective and very easy to use for everybody. Dino Amenduni [explained the first point well](#) at the end of 2010 with words and arguments that, while tightly bound to the current situation in Italy, apply, in spirit, also to other countries:

in order to have your voice heard, it is necessary to threaten the private interests of politicians. The ways to achieve this goal are, in my opinion... Communication guerrilla: physical violence doesn't generate change anymore. Power is in the hands of those who have data. But those data must be communicated, made usable, fun to use, shareable, in order to give the feeling that knowledge brings a concrete (economic or intangible) personal advantage

Proofs that participation to generation and usage of Open Data is easy would include, instead, examples like [electionleaflets](#). All citizens who can use a computer scanner and have Internet access can upload on that website the leaflets distributed by the candidates during a campaign, making much easier (after other, more skilled volunteers have inserted the content of the leaflets in searchable databases) comparisons between the candidates positions or making public some disrespectful material (racist, insulting...).

4.7. Involve NGOs, charities and business associations

As a final note and recommendation of this report, we'll note that, in comparison with hackers and public officers, there are other parties that could and should play a role in Open Data adoption much bigger than what they have had so far.

NGOs and charities, as well as professionals or business associations, all have lots to gain from Open Data but don't seem, in many cases, to have realized this yet. Members of the first category should routinely ask for support directly to Open Data civic hackers to gather (either from government or citizens) more up to date information that is specifically relevant for their campaigns.

The other associations, instead, should be much more active both in publishing Open Data about their activities, to gain better access to customers and guarantee fair market competition, and in

officially lobbying Public Administrations to get the PSI they could use for the same purposes. As other suggestions made here, these are activities that should start at the city and regional level, first with custom-made education initiatives, then with specific data-based services. Engaging all these actors in the adoption of (local) Open Data will be one of the big challenges of the next years.

5. Bibliography

Besides those explicitly linked from the text, this report has drawn inspiration by many other resources. The most important ones are listed here, but the complete list should be much longer. We wish to thank first the authors of the works listed below and, immediately after, to all the activists, inside and outside governments worldwide, who are working on this topic.

1. [Are you prepared for the pitfalls of Gov 2.0?](#)
2. [Can we use Mobile Tribes to pay for the costs of Open Data?](#)
3. [Canada launches data.gc.ca - what works and what is broken](#)
4. [Creative Commons and data bases: huge in 2011, what you can do](#)
5. [Defining Gov 2.0 and Open Government](#)
6. [How Government Data Can Improve Lives](#)
7. [If you like solar, tell your utility to publish this map](#)
8. [Indian corruption backlash builds after "year of the treasure hunters"](#)
9. [Información Cívica / Just What is Civic Information?](#)
10. [Is open government just about information?](#)
11. [LSDI : In un click la mappa del crimine](#)
12. [La casta è online: dategli la caccia!](#)
13. [Linee guida UK sull'opendata](#)
14. [MSc dissertation on Open Government Data in the UK](#)
15. [Open Data \(2\): Effective Data Use.](#)
16. [Open Data: quali prospettive per la pianificazione?](#)
17. [Open Knowledge Foundation Blog " Blog Archive " Keeping Open Government Data Open?](#)
18. [Open data, democracy and public sector reform](#)
19. [Pubblicato Camere Aperte 2011 - blog - OpenParlamento](#)
20. [Reasons for not releasing data in government](#)
21. [The impact of open data: first evidence](#)

- 22.[Thinking About Africa's Open Data](#)
- 23.[Towards EU Benchmarking 2.0 - Transparency and Open Data on Structural Funds in Europe](#)
- 24.[UK Open Government Licence removes barriers to re-use of public sector information](#)
- 25.[Western Europe: A journey through tech for transparency projects](#)
- 26.[What open data means to marginalized communities](#)
- 27.[What's in a Name? Open Gov and Good Gov](#)
- 28.[WikiLeaks Relationship With the Media](#)
- 29.[WikiLeaks, Open Information and Effective Use: Exploring the Limits of Open Government](#)