POLITECNICO DI TORINO
Repository ISTITUZIONALE

Open Government Data: A Focus on Key Economic and Organizational Drivers

Original
Open Government Data: A Focus on Key Economic and Organizational Drivers / Iemma, R.. - ELETTRONICO. - (2013). [10.2139/ssrn.2262943]

Availability:
This version is available at: 11583/2520895 since:

Publisher:

Published
DOI:10.2139/ssrn.2262943

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

15 September 2026

Nexa Center
for Internet & Society

WORKING PAPER

OPEN GOVERNMENT DATA:
A FOCUS ON KEY ECONOMIC AND ORGANIZATIONAL DRIVERS

First version: November 2012

Abstract

Grounding the analysis on multidisciplinary literature on the topic, the existing EU legislation and relevant examples, this working paper aims at highlighting some key economic and organizational aspects of the 'Open Government Data' paradigm and its drivers and implications within and outside Public Administrations. The discussion intends to adopt an 'Internet Science' perspective, taking into account as enabling factors the digital environment itself, as well as specific models and tools. More 'traditional' and mature markets grounded on Public Sector Information are also considered, in order to indirectly detect the main differences with respect to the aforementioned paradigm.

JEL Codes: D45, H41, O31, O34.

Keywords: Public Sector Information, Information Economy, Internet, Open Data

Raimondo Iemma
raimondo.iemma@polito.it

Electronic copy available at: http://ssrn.com/abstract=2262943

Table of Contents

Introduction: knowledge-based economies and the Internet 3
The central role of knowledge in modern economies....3
Internet and the ICTs....3
1. (Open) Public Sector Information as a driver for innovation....4
1.1 What is Public Sector Information 4
1.2 Features of interest....5
1.3 Downstream applications....6
2. The European framework 7
3. Economic aspects....8
3.1 Market values....8
3.2 (marginal cost) Pricing....9
4. Open Government Data initiatives....13
4.1 The main costs for PSI holders....13
4.2 The potential benefits....15
4.3 Examples of key drivers and best practices....16
The role of transaction costs....16
Internal efficiency curves: small evidence....16
Fostering widespread positive effects through free-of -charge policies: two examples at national level....16
4.4 Emerging tools and models enabling PSI reuse....17
5. Examples of mature PSI markets in Italy....19
5.1 Business Register....19
5.2 Cadastral information....21
5.3 Legal information....23
Conclusions....24
References....26

Electronic copy available at: http://ssrn.com/abstract=2262943

Introduction: knowledge-based economies and the Internet

The central role of knowledge in modern economies

Advanced societies and economies are often said to be 'knowledge-based' or 'knowledge-driven' (Lundvall, 2003). A broad variety of conceptualizations of this paradigm, dating at least from the 1960s, can be acknowledged. Amongst several others, a simple definition has been provided by the OECD (2005), suggesting that knowledge-based economies are characterized by '*greater dependence on knowledge, information and high skill levels, and the increasing need for ready access to all of these by the business and public sectors*'. Such dependence seems not only to get stronger with the complexity of the interactions that take place within and between all categories of actors, but, even more importantly, it may assume several shapes.

Knowledge itself plays different roles along value chains (or, better, value networks), by representing at least: (i) an increasingly important asset or input to production, that can be privatized or commoditized (considering only the two extremes); (ii) a good that can be traded, exchanged or shared in market or non-market environments. According with the adopted assignment of property rights (or, using another perspective, with the 'openneness degree' of knowledge) both transformation and transaction costs can be affected, with a net balance that needs to be evaluated case by case. Moreover, attention by scholars has been (and is) devoted to the distinction between information and knowledge (Simon, 1999) and between different types of knowledge – e.g. codified, tacit but codifiable and inherently non codifiable (Witt et al., 2007) –.

Such trends are studied from different perspectives and have significant implications within various disciplines, ranging for instance from economics of innovation to industrial organization, from intellectual property studies to sociology, from public economics to institutional change. Businesses themselves are experiencing a transformation encompassing higher incentives towards the valorization of intangible assets (suffice it to remind the valuation of Internet-based companies that went public in the last years compared with the one of 'traditional' manufacturing companies), while in many countries the public sector increasingly produces efforts towards efficiency and transparency grounded on improved exchange of information and knowledge.

Internet and the ICTs

Information and communication technologies (ICTs), and the Internet in particular, offer unprecedented practical means to access, process, share, combine, organize and reuse vast amounts of information. On the one hand, such technologies allow to provide evidence of the existence of increasing returns (Arthur, 1996 and even before Romer, 1983) and the path dependence related with some specific features of ICT- and Internet-based businesses and processes, such as supply- and demand-side economies of scale, network externalities and the combinatorial nature of innovation in the digital environment (Varian et al., 2004). On the other hand, the development of the Internet and the ICTs is radically improving both the central role of knowledge in our societies and the ways knowledge is produced, managed, exchanged and reused (or, in case, 'locked' in proprietary contexts). Again, these kinds of processess concern individuals, firms, public bodies and all other kinds of organizations, including the interactions between them, with a relevant impact at social and economic level.

Not by chance, the emerging Internet Science specifically aims at understanding the impact of the inherent features of the Internet on societies and organizations, as for technological, economic and

Electronic copy available at: http://ssrn.com/abstract=2262943

social aspects, as well as the adoption patterns of the related innovations. The Internet Science lays on a multidisciplinary approach, highlighting the strong and mutual relationships between the development of the Internet and economic performance, the behavior of organizations and the way social challenges are addressed.

This workin papers attempts to describe some of the main aspects of the Open Government Data paradimg within the framework just described. The first section defines Public Sector Information and its potential for reuse. Section 2 briefly summarizes the content of the European prescriptions and approach on the subject. The third section discusses some key economic aspects, charging principles in particular. The fourth section aims at discussing costs and benefits related with Open Government Data, and presenting interesting examples of Open Data projects, models and tools. Section 5 analyses three mature markets grounded on (closed) Public Sector Information at national level in Italy.

1. (Open) Public Sector Information as a driver for innovation

1.1 What is Public Sector Information

Within their institutional mandate, organisms belonging to the public sector collect, create, maintain and update a huge flow of information and content, ranging from economic, demographic or meteorological data to geographic, toponymy or cadastral data. The nature, format, volume and frequency of update of this sets of information may vary according to the role, mandate and organisational features of the public body creating and/or holding them. Not unfrequently, information is generated by public sector bodies can be conceived as an incidental good, to the extent that collecting and maintaining information may not be the final goal itself of public bodies (whereas in others, it is indeed: one should think, for instance, of the case of a cadastral registry). For instance, the educational system is not in place to collect and release information about public schools, like number of teachers or students distribution; still, fulfilling the public task to run the educational system also leads public authorities to generate and maintain registries containing such kind of information. It may therefore seem that, whatever its origin (actual task, legal obligation, incidental creation, or a combination of the previous) the production of Public Sector Information (hereinafter PSI) is to be taken as exogenous. However, this holds true only in the short run, whereas, at least in principle, the incentive structure for generating PSI may vary across time and may not apply for all data streams indifferently, especially in cases where framework conditions such as available funding, regulation, market configurations and political commitment may vary.

The Directive 2003/98/EC indirectly defines PSI as “any content whatever its medium (written on paper or stored in electronic form or as a sound, visual or audio-visual recording)” and “any part of such content”[1] that is “held by the State, regional or local authorities, bodies governed by public law and associations formed by one or several such authorities or one or several such bodies governed by public law”. However, as noted for instance by the LAPSI Thematic Network (Position Paper No 3, 2011), PSI holders may be not be limited to sole public sector bodies in a strict sense, but also other organisms, like for instance public undertakings, especially where the activities of such bodies meet a general interest consistent with the public task[2].

[1] The Directive does not apply to specific parts of PSI whose disclosure and reuse (at least ‘as is’) might affect for instance privacy, third parties rights, statistical confidentiality and national security.

[2] Relevant policy documents on the subject (OECD, 2006; European Commission, ‘Vickery report’, 2011), it seems important to operate a conceptual distinction between PSI, which is continually generated and virtually readily usable in

1.2 Features of interest

First of all, the fact that PSI is or is not an incidental good has a direct implication on cost computation (related in turn with charging principles). Where PSI results incidentally from the public task, it seems reasonable to consider that its production has been already subsidized. Most probably, without the mandate to carry out the background set of activities that led to its incidental creation, a public sector body would not (have the incentive to) generate that information. In other cases, where the production, maintenance and dissemination of information is in itself a public task, the related costs can be more precisely allocated and, even more importantly, strong incentives emerge for the holder to consider PSI as a resource to be funded and, eventually, from which extracting direct returns.

Secondly, a vast amount of PSI is 'sole-source', to the extent that it could not be directly substituted by other sources. This particularly holds true where PSI is generated in light of a legal obligation or as an an outcome of the public task. Example of 'sole-source' information are legal information (e.g. verdicts) or public expenditure data. In general, PSI that describes and reports about functioning and performance of the public sector is 'sole-source'. Other sets of information - such as for instance business registries or cadastral databases - could be in principle generated by third parties, but the fact that such repositories are often registration-based (and Governments can make this registration mandatory, or link it to other activities that imply interactions with the public sector) and the significant sunk investment required to eventually duplicate the effort have made at least so far such domains of PSI as *de facto* 'sole-source'. In other sectors, PSI is not provided sole-source, but in practice substitutes (although not 'perfect substitutes') of PSI are produced: it is the case for example of geographical data (e.g. Google Maps, or the crowdsourced platform OpenStreetMap), and meteorological data.

Moreover, PSI is collected and stored in a variety of (digital) formats. This mostly relates with the nature itself of the information (for instance being it a document, a spreadsheet, or a database), and also with the intended scope and use within Public Administrations. The required effort to make PSI actually reusable may significantly vary from a case to another. Such activities may comprise, although not being limited to: extraction from a legacy database; adoption of an open format; application or update of metadata in order to make them meaningful for third parties; anonymization (where needed); update. Starting from a raw status, the path to get PSI reusable may be shorter or longer, entailing various types of fixed and variable costs.

It seems interesting to briefly discuss whether a relationship can be identified between the origin (incidental or not) and the nature ('sole-source' or not) of PSI. Table 1 suggests examples for each of the possible combinations. Information that is inherently 'sole-source', i.e. that cannot be directly substituted by other sources, can be generated either as the direct outcome of the public task or as a

third-parties applications, and Public Sector Content, which is often established and therefore static (e.g. a record), held but not necessarily produced by a public body, nor directly related with the functioning of the public body itself. Although legislations related with access and reuse of PSI are progressively starting to address also Public Sector Content, and in particular documents and works held by public libraries, museums and archives, we believe that the two domains - although they share common features and may partially overlap in some cases - hold radically different dynamics and therefore have to be kept separated in the analysis. This article focuses on Public Sector Information in a strict sense#, not including cultural goods and works. The OECD (2006) suggests the following PSI domains: geographic information; meteorological and environmental information; economic and business information; social information; traffic and transport information; tourist and leisure information; agricultural, farming, forestry and fisheries information; natural resource information; legal system information; scientific information and research data. In turn, Public Sector Content encompasses broad domains such as educational content, political content and cultural content.

‘by-product’. With respect to the latter, one could think of statistical information about public expenditure: raw data used to create it are exclusively in the hands of public bodies (which makes them 'sole-source'), and such data are produced somehow incidentally, to track public budget (where the actual task is its management and allocation). Similarly, information that is not inherently but, in most cases, *de facto* 'sole-source', is produced either incidentally or not. Finally, information that has (imperfect) substitutes is usually produced as a direct outcome of the public task, like in the case of cartographic data. It seems however particularly difficult to identify non 'sole-source' PSI that is generated incidentally.

	inherently 'sole-source'	**typically 'sole-source'**	**with substitutes**
non incidental good	Laws repository	Firm registry	Maps
incidental good	Public expenditure raw data	Data about the educational sytsem	

Table 1 - Example of PSI domains with respect to their origin (incidental or not) and nature ('sole-source' or not)

1.3 Downstream applications

PSI exists since well before the (rather recent) 'Open Data hype'. Better, we could say that the creation of PSI is as old as the public sector itself. This trivial observation leads to identify at least two separated kind of markets grounded on PSI reuse.

On the one hand we have datasets made available in digital formats (machine readable and, in the best cases, semantically linkable with each other) through web catalogues. In this case, information is somehow 'new', never made available as open before, and its reuse enables the creation of innovative services (such as apps and the like). This fosters 'low-end' markets of applications, some of which are able to generate a broad (and even cross-border) demand.

On the other hand, Public Sector bodies hold and manage huge amounts of information already feeding mature markets. Some key examples, just to mention a few, are land registries and firm registries. In those cases, where access and reuse of information is not open to anyone, downstream markets are typically populated by a limited number of medium-high sized players. This is what we could call the 'high-end' segment of PSI. The configuration of those markets (which are likely to be rather concentrated) matters for social welfare aspect, such for instance charges for end-users, quality of the service and innovation.

Therefore, under an economic viewpoint, whether PSI is incidentally generated or not, and whether it has substitutes or not are the key aspects to be evaluated. In fact, where PSI is in itself the expected outcome of a public task, empirics show that their public holder allocates a substantial part of its resources to the production, maintenance and update of information, seeking to recoup the related costs through charges. In the case of incidental PSI, what has to be eventually funded is the effort to actually make available to third parties this amount of information.

Jointly considered, such features imply a high potential for reuse, both for commercial and non-commercial purposes. From this point of view, one should consider that:

1. making PSI available in open formats is supposed to foster both simultaneous

innovation, i.e. 'forking' projects or services grounded on the same datasets (forking may relate with the way data are reused and/or with complementary assets held by single re-users) and cumulative innovation / 'inventing around' (which may be achieved by multiple re-users at the same time in a community-based framework and/or along the 'chain of reuse' of PSI);

2. an open approach for PSI panders the high degree of componentization of applications, products and services supplied downstream, which are likely to be grounded on multiple and heterogeneous sources, including but not limited to PSI;
3. supply- and demand-side economies of scale and scope may indeed represent a tangible driver both for PSIH and re-users. Once adopted a proper technical interface, which may partially overlap with the internal interchange infrastructure, the former are able to open new datasets with small incremental effort. Indirect network effects are besides important, to the extent that an increasing number of reusers may generate further incentives for PSIH to release new open data, with positive externalities on all reusers. Direct network effects do matter especially in the case of community-based projects based on PSI.

2. The European framework

European institutions have been issuing policies focused on PSI access and reuse since the the late 1980's. In 1989, the EC promoted a set guidelines aimed at improving the efficiency in the information market[3]. Nine years later, the EC issued the "Green Paper on Public Sector Information in the Information Society."[4], highlighting the relevance of PSI reuse in the EU, describing opportunities and obstacles within legislative frameworks of the Member States. In 2001, the EC adopted the "eEurope 2002" action plan[5], prescribing the online publication of public data as well as supporting an integrated approach to PSI at the European level. This Communication reported on the economic benefits related to PSI exploitation, also describing the remaining obstacles, especially for private reusers, stressing the importance of a cross-border perspective (including legal harmonization). In 2003 the European Parliament enacted a directive regarding the right of EU citizens to access environmental information[6]. In the same year, the European Parliament passed Directive 2003/98/EC on the re-use of Public Sector Information[7] (hereinafter 'PSI Directive'), currently the main reference on the subject at European level.

The PSI Directive aims "*(...) to facilitate the creation of Community-wide information products and services based on public sector documents, to enhance an effective cross-border use of public sector documents by private companies for added-value information products and services and to limit distortions of competition on the Community market (...)*" (Recital 25). More in particular, the PSI Directive has the purpose to facilitate (and, where possible, provide the tools for promoting) cross-border products and services based on the reuse of PSI. The Directive also intends to

---

3 http://ec.europa.eu/information_society/policy/psi/docs/pdfs/brochure/1989_public_sector_guidelines_en.pdf

4 ftp://ftp.cordis.europa.eu/pub/econtent/docs/gp_en.pdf

5 http://ec.europa.eu/information_society/eeurope/2002/documents/archiv_eEurope2002/actionplan_en.pdf

6 Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC

7 Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information.

neutralize competition distortions, enabling transparency and preventing discrimination (especially in terms of exclusive arrangements). Another important objective is to minimize the fragmentation of PSI reuse legal frameworks in Member States, harmonizing as much as possible the different approaches.

Amongst the other provisions, the PSI Directive mandates that:

- where documents[8] held by public sector bodies are made accessible, those documents must be reusable either for commercial or non- commercial purposes;
- documents must be made available in their pre- existing format or language through electronic means when possible;
- "where charges are made, the total income should not exceed the total costs of collecting, producing, reproducing and disseminating documents", also including a reasonable rate of return of the investment;
- non-discriminatory conditions for reuse for comparable categories of reuse (still, exchange of information between Public Administrations at no charge is allowed regardless of the charging regime applied downstream);
- PSI holders avoid exclusive agreements between themselves and private partners to the greatest extent possible (except where a service of public interest would not be provided otherwise).

As already mentioned, there are several categories of documents upon which the Directive's provisions do not apply: (1) documents produced outside the scope of the public task of the public sector body; (2) documents protected by the intellectual property rights of third parties; (3) documents that are excluded under a Member State's law, for reasons such as national security and commercial confidentiality; (4) documents held by broadcasters for public service remits, documents held by educational and research institutions, and documents held by cultural establishments

Two further recent trends can be briefly mentioned as part of the current and future PSI strategy at European level:

- the forthcoming launch of a European Open Data portal[9], with the twofold perspective to federate national and local portals and to make available for reuse datasets from the EC administrative level;
- the adoption of a European Open Data licensed (to be applied, for instance, to PSI).

3. Economic aspects

3.1 Market values

Several studies provide field evidence about the (downstream) overall market value of PSI. For instance, PIRA (2000) estimated investment value of PSI (i.e. government investment in the creation of PSI) and economic value of PSI (i.e. the national income generated by the downstream

[8] The EC defines a document as "any representation of acts, facts or information - and any compilation of such acts, facts or information - whatever its medium (written on paper, or stored in electronic form or as a sound, visual or audiovisual recording), held by public sector bodies".

[9] http://epsiplatform.eu/content/eu-data-portal-contract-awarded

exploitation of PSI) in the European Union, putting the former at around EUR 9.5 billion per annum in 1999 and the latter at around EUR 68 billion (equivalent to approximately 1.4% of EU GDP – a seven-fold return on investment).

Employing a large survey of PSI producers and users, MEPSIR (2006) sought to estimate the size of the PSI market in Europe. Based on the estimates of re-users they put the overall market for PSI in the EU plus Norway at around EUR 27 billion (approximately 0.25% of aggregated GDP). te Velde (2009) suggested that the value might drop further from EUR 27 to EUR 5 billion or even EUR 3 billion13 – only around 5% of the PIRA study estimate, and less than PIRA's estimate of investment value.

The Pricing Of PSI Study (POPSIS), issued in 2011, assessed different models of supply and charging for PSI and their effects through the analysis of 21 case studies. The cases cover a wide range of public sector bodies (PSBs) and different PSI sectors (meteorological data, geographical data, business registries and others) across Europe. The study has also produced a snapshot of the smartphone applications market based on PSI and a comparative analysis of several Open Data portals in Europe and beyond.

The case studies show a clear trend towards lowering charges and/or facilitating re-use (16 out of the 21 cases). Some PSBs only charge for commercial re-use and allow non-commercial re-use either against reduced fees (seven out of 21 cases) or for free (nine out of 21 cases). In almost all cases, PSBs allow free access to their PSI (i.e., viewing without copying). In some cases, free access has been the forerunner of a more flexible re-use regime. In all the case studies, the PSI re-use revenues of PSBs range from relatively small to extremely small when compared to the total budget of the PSB concerned. In half of the cases, these revenues constitute less than 1% of the PSBs' entire budget.

Based on their own raw data, the number of PSBs that exploit added-value products is limited (seven out of 21 cases) and appears to be decreasing over time. In those cases where PSBs moved to marginal and zero cost charging or cost-recovery that is limited to re-use facilitation costs only, the number of re-users increased by between 1,000% and 10,000%.

The market value of PSI reuse is substantial. According to a recent study by the European Commission ('Vickery report') the aggregate direct and indirect economic impacts from PSI applications and use across the whole EU27 economy are estimated to be around 140 billion Euros annually. Further analysis suggests that if PSI policies were open, with easy access for free or marginal cost of distribution, direct PSI use and re-use activities could increase by up to 40 billion Euros for the EU27.

3.2 (marginal cost) Pricing

Defining the optimal charging scheme for PSI is basically an answer to the question "Who should fund PSI collection, storage and distribution?". Two extremes can be identified: public funding (i.e. zero charge for reuse) and 'private users pay' (full cost recovery). In between, several other options are possible, among which average cost recovery (with charges allowing to recover long-term average costs, with no required funding) or marginal cost pricing (charges equalling marginal cost of reproduction, the latter being approximately zero in a digital environment).

At least three sets of arguments, based on the comparison between the aforementioned approaches under a 'social welfare maximization' perspective, support marginal cost pricing as a default rule. Those mainly relate with microeconomic analysis (supply- and demand-side) and public economics principles.

Newbery et al. (2008) and Pollock (2009) compare the effect of three basic charging policies: (i) profit maximization (resulting in a monopoly pricing); (ii) average cost recovery pricing (setting charges equalling to the long-run costs); (iii) marginal cost (zero cost) pricing.. The third option indeed allows to maximize consumers' welfare while implying government funding in order to carry out the supply of information. The overall social welfare is higher in case:

- the positive externalities given by 'zero' charges make the increase in consumer's welfare overcome government funding (the latter being then justified);
- the available financial resources actually allow government funding;
- lower ('zero') charges foster an increase in the demand volume.

Under such conditions, marginal cost pricing / government funding allows to maximize collective welfare.

Pénin (2010) tackles the same issue from a different perspective. The 'DIK' (Data, Information, Knowledge) paradigm is adopted as a model to branch PSI. Under Pénin's hypotheses, along DIK model, transformation costs and marginal cost of reproduction increase – the latter being higher than zero for Knowledge (i.e. contents that require organization and cognitive re-appropriation to be adequately disseminated) – while absorption costs for reusers decrease. Moreover, a link between PSI charges and reusers' willingness to pay is proposed. In short, the assumption is made that, the willingness to pay for raw data being zero (and the marginal cost as well), raw data should be released for free. For the same reason, the charge for PSI falling under the definition of Knowledge (treated contents) should be higher than zero. While agreeing on the first conclusion (marginal cost charging principle for raw data) we find the aforementioned assumptions not perfectly correct. First of all, grounding a charging strategy on the consumers' willingness to pay can be seen as a profit maximizing strategy based on second degree price discrimination (i.e. 'versioning') aimed at extracting surplus from consumers. Secondly, and even more important, the value of raw data seems radically underestimated. While raw data may indeed 'hide' valuable information (e.g. to be extracted through mash-up with other datasets), further elaboration before their release could make disappear such potential.

A further line of reasoning in favour of marginal cost pricing, suggested among others by the LAPSI Network (2011), is grounded on a simple but convincing public economics principle. Such as other activities carried out within the institutional tasks of Public Sector bodies, the production of PSI has been already funded by taxpayers. Therefore potential users cannot be asked to contribute a second time for the same piece of information. At the same time, should the dissemination of information entail extra-costs in response of a specific request by a reuser, such specific costs should be allocated to this specific reuser and not to the general taxpayer.

A *cavaeat* is moreover necessary. So far, we implicitly made the assumption that PSI only comprises *sole-source* information (i.e. datasets that cannot be reproduced by any other player than its original holder), therefore with no competition in the upstream segment of the market. Marginal cost charging then allows maximum distribution of data, exploiting the positive externalities of reuse and fostering downstream competition. However, as already suggested by a policy contribution issued by the LAPSI Network (Ricolfi et al. 2011), in some sectors private operators may also be able to produce the same set of information than the PSIH. In such cases of (at least potential) upstream competition, marginal cost pricing could be seen as predatory.

In general, a clear definition of 'marginal cost' is needed. As a general meaning, marginal cost is the cost of (re)producing an additional unit of a good. As previously reminded, the cost of reproducing a digital good tends to zero. In the case of PSI, the costs entailed by an additional download of a

single dataset by a new user seems less than significant, even if a strong increase in the demand may engender (non marginal) extra costs on the supply side (e.g. to expand band capacity). However, specific requests by single users may require extra-effort for retrieving data and make them available. The main terms of the discussion about pricing of PSI are summarized in Table 2.

Finally, the proposal of amendment of the PSI Directive released in December 2011[10] addresses charging. Marginal costs of reproduction and dissemination of PSI would be set as a default cap. However, this rule would not apply: a) in exceptional cases, "in particular where public sector bodies generate a substantial part of their operating costs relating to the performance of their public service tasks from the exploitation of their intellectual property rights"; b) to LMAs. The charging principles set in the PSI Directive (cost-recovery plus reasonable return on investment) would still represent a threshold not to be exceeded in any case, while PSI holders would carry the burden to prove compliance of their charging schemes with cost-oriented accounting principles.

[10] http://ec.europa.eu/information_society/policy/psi/index_en.htm

Charging principle	*Main supporting arguments*	*Critical points (or counterarguments)*	*Current fields of adoption*
Charges equalling marginal costs of reproduction and dissemination (or are set to zero)	• Maximization of the **dissemination level** of PSI, with related (upstream and downstream) **positive externalities.** • Increased **allocative efficiency** (benchmarK: perfectly competitive markets, where marginal revenues equal marginal costs). • Determining the market value of single PSI sets is frequently difficult, especially considering that the value is increased where several sources of information are mixed.	• Need of a **public subsidy** in order to sustain the process. • In case of non 'sole-source' repositories, marginal cost pricing by the public holder can be considered as **predatory** / anticompetitive.	• Open Government Data initiatives by local / national goverments.
Recouping long term average costs (+ *possible markup*)	• **Selection mechanisms** allowing to identify third parties with higher expected returns from PSI reuse (therefore with higher willingess to pay). • Increased **financial autonomy** for the PSI holder. • Incentives in defining appropriate and sustainable **terms of service**.	• Being part (or incidental to) the public task, creation and maintaining of PSI datasets are already funded through **general taxation**. • Applying charges above marginal costs may entail the exclusion of reusers holding **low willingness to pay**, even with socially valuable prospects of reuse (e.g. In the case of not-for-profit reuse aimed at increased transparency of the public sector) • Incentives for PSI holders to **over-invest** in PSI management and production. • **Double-marginalization** on end-users.	• National registries managed by public organisms with the task to maintain them by generating revenue flows (e.g. business registries).

Table 2 – Arguments in favour and against the two main charging frameworks for PSI.

4. Open Government Data initiatives

The global framework for PSI dissemination – from a legal, technical and economic perspective – is currently being improved and has so far achieved a good level of definition. As a preliminary observation, one should consider that not all PSIH may belong to the same category (being PAs, centralized bodies, departments, authorities, even public undertakings if we adopt the broadest definition of PSIH), with different degrees of financial and operative autonomy. Their incentive structure with respect to PSI supply may therefore differ, even dramatically. Some critical issues may then arise when adopting a marginal cost pricing scheme. Hereafter we remind some key aspects.

1. The actual financial commitment of the funding body, either being the PSIH itself or its managing organism. Budget allocation may be in fact subject to changes across time (e.g. for political or institutional reasons).
2. Incentives for cost-reduction and efficiency, which may fall in the case of PSIH that periodically negotiate with their managing bodies the amount of the subsidy.
3. Information asymmetry on the actual costs incurred for PSI-related activity.

Since, 2009, in several European countries, Public Aministrations are carrying out Open Data initiatives[11], i.e. the release of PSI as technically and legally open and reusable datasets. Such projects share some common points:

- they are carried out with internal (dedicated) resources, typically allocated both to back-end activities (such as coordination, legal clearance, data handling) and front-end activities (publication);
- the thematic coverage of the datasets is increased over time under an incremental approach;
- the datasets are made available under open licenses (CC BY is the most frequently used);
- some data holders are experimenting the publication of datasets as Linked (e.g. using the RDF formalism);
- the creation of value-added services grounded on Open Data (e.g. Apps) is usually left to private reusers.

4.1 The main costs for PSI holders

As previously discussed, the 'production' of information in a digital environment is characterized by high fixed costs (e.g. the collection of data, their processing and structuring, as well as the maintenance and updating of data bases) and costs of reproduction and distribution close to zero. Obviously, the cost structure has obvious implications in the choice of charging schemes and, more generally, in the patterns of release - open or restrictive – of information. Moreover, the computation of fixed costs is particularly difficult in the case of datasets / databases created as a 'by-

[11] In Italy, Open Data initiatives are being carried out by several Public Administrations in Italy at several levels: national (http://dati.gov.it), regional (e.g. http://dati.piemonte.it, http://dati.veneto.it/, https://dati.lombardia.it/, http://dati.emilia-romagna.it/) and local (cities), such as http://opendata.comune.fi.it/ (Firenze) and http://www.comune.torino.it/aperto/ (Torino). Other public bodies, like for instance the National Research Council (http://data.cnr.it/site/) or the National Social Security Institute (http://www.inps.it/portale/default.aspx?iMenu=1&iNodo=7719), are starting publishing datasets in Open modes.

product' of the public task.

The kickstart of Open Government Data initiatives typically consist, at least at local level, of three main subsequent steps, each one characterized by specific activities and cost types:

1. Preparatory activities, mainly including:

- establishment of a multidisciplinary working group;
- internat training about the main legal / technical features and implications of the Open Government Data model the characteristics and implications of the model open government data (legislation, guidelines, tools);
- internal coordination and discussion with rightholders;
- creation and testing of web platforms for Open Data publication;
- identification and adoption of the most approriate terms of use and licensing schemes for the datasets released;
- definition and formalization of internal processes and information flows.

2. Open release of the first sets of data, typically those already available without the need for further manipulation (e.g. necessary for compliance with the current legislation about privacy protection) or internal negotiation (for example in relation to intellectual property rights held by third parties);

3. Gradual increase in the volume of data available, typically beginning to address the process of opening sets of information that require non-straighforward technical and legal examination to be released. Those tasks may include:

- assessment of existing IPRs on the data;
- aggregation or other operations aimed at ensuring anonymization;
- adoption of fully machine readable (or even linked) data formats.

The main cost items previously discussed are summarized in Table 3.

Task	**Internal resources**	**Direct costs**
Coordination	*FTE Allocation (men-months)*	
Internal training		*(External assignements)*
Definition / adoption of terms of use and licensing schemes (legal clearance)		*(External assignements)*
Extraction of datasets from legacy databases and manipulation		*(External assignements)*
Web publishing and management		*(External assignements)* *(Software licences)* *(Cloud services)*
Foregone revenues		

Table 3 – Kick-starting Open Government Data initiatives: the main cost items.

4.2 The potential benefits

As noted by – amongst others - by Houghton (2011), there are three main levels in which tangible benefits from the release of PSI as open and free of charge can be experienced and assessed:

- at the PSI holder level, benefits can be entailed in terms of internal savings and increased efficiency, especially with respect to the following items:
  - reduction of transaction costs;
  - increased focus in core-tasks;
  - more efficient data exchange between different departments / offices
- for PSI reusers, in terms of savings (since fees and other kinds of charges are dropped) and in the increased availability of datasets, and in particular:
  - direct savings resulting from free access to the data;
  - indirect savings (e.g. less time spent on inquiries) resulting from the adoption of explicitly open terms of reuse (non-transactional licenses);
- for the ecosystem in its broader extent, in terms of new opportunities to create value-added services on top of PSI, also wth the purpose of increased transparency and democratic participation, directly linked with the following aspects:
  - increase in the availability of public information (and reduction of transaction costs);
  - creation of new products and value added services grounded on PSI;
  - development of complementary products to those directly based on PSI;
  - tangible and intangible returns from such kinds of services (network externalities).

4.3 Examples of key drivers and best practices

The role of transaction costs

Especially in the case of local administration (holding datasets with limited stand-alone overall value) transaction costs (i.e. the costs related with billing, accounting and management of the transaction) can actually overcome the expected income or the previous actual income from PSI release.

For instance, the regions of Piemonte and Emilia Romagna are examples where geographic data were available, until recently, under charges (typically a two-part tariff) but transaction costs turned out to be higher than the actual income, making PSBs decide to drop charging altogether.

In fact, the Geographic Information System of the Regione Piemonte encompasses a comprehensive set of cartographic indexes, updated over time. It is a set of geographic data, represented in the form of maps showing geometric information, topological and descriptive information of the region. Periodically updating such datasets entails significant costs. For instance, the making of the map of a large urban center (scale 1:2000, with an update necessary every two years) requires an investment of about 50 euros per hectare. Prior to the approval of the Guidelines for the reuse and interchange of information assets regional (November 2010), promoting Open Government Data, geographic data were released for consideration, according to a specific tariff (according with a price-list). The annual net revenues were estimated between 10 thousand and 20 thousand euros. The costs entailed by the non-automatized management of the transactions (with no relying on a distribution channel) – i.e. the allocated human resources – instead added up to about 50 thousand euro per year, thus making, in fact, such sale unprofitable. Besides allowing a wider dissemination of geographic data, dropping fees altogether has therefore generated savings for the Public Administration.

Internal efficiency curves: small evidence

In the case of the extraction of a single dataset from a centralized legacy database with the aim at publish it as Open, several types of costs can be computed: internal coordination meetings, querying, data sharing, web publishing, with a total cost (in terms of share-work of internal resources) that can reach 1500 euros. The overall resources (mostly internal) allocated by the Veneto Region between September 2011 and July 2011 to the regional Open Data project roughly amount for 35 thousand euros (as total costs related with management, communication, training, development of the web portal, web publishing, analysis of the datasets, rights clearance, licensing assessment) with an average cost per dataset published of around 300 euros. (*Source: interviews with the project manager of the Open Data initiative in Veneto).*

Fostering widespread positive effects through free-of -charge policies: two examples at national level

The Spanish Cadastre is a register containing periodically updated legal and economic information about rural and urban real estates, covering 32 million urban properties, 42 million rural parcels and 23,5 cadastral owners. While the Spanish government maintains such database mainly for fiscal reasons, cadastral data is a key ingredient of geographic / territorial studies and serves as input in a broad set of markets (from commercial information to real estate). Interestingly enough, the Spanish Cadastre switched from a paying / closed regime (requesting a fee to access data) to an 'open by default' regime, where cadastral data are available for free (via Internet or via 3000 Cadastral

Information Points distributed across the country), with the aim to foster both commercial and non-commercial reuse[12]. Amongst the other drivers of the change in their pricing policy (e.g. social welfare), the Spanish Cadastre experienced that transaction costs (pricing determination, management, invoicing, 'bureaucracy' in general) overcome the direct revenues from exploiting the data. Moreover, wider availability of the data is allowing to collect feedbacks on potential improvements by the users. Between 2003 and 2010, the number of downloads of cadastral data increased from 17.000 to more than 5 million.

The Danish Enterprise and Construction Authority (DECA), responsible for road names and addresses in Denmark, promoted a free of-charge agreement in 2002. Especially at national level, address data play a key role in several applications: fundamental activities such as emergency, fire and ambulance services, transport services and others strongly rely on the availability of reliable address data. In front of a total cost of around 2 million euros for the agreement, the direct overall financial benefits in the period 2005-2009 amount to around 62 million euros. Interestingly enough, the study does not include the supplementary economic benefits arising in later parts of the distribution chain, neither other indirect benefits such as the gathering of feedbacks form users and the rationalization of processes (e.g. no more duplications in data collection). Even if the value calculations in the assessment were based on the assumption that the economic value of the free-of-charge addresses corresponds to the 75% of the price to users (e.g. public-transport journey planners) actually paid for municipalities' address data before the free-ofcharge agreement, such figures are still quite considerable. The measured benefits also include the savings of public and private reusers from not having to enter into agreements anymore with the PSI holder.

4.4 Emerging tools and models enabling PSI reuse

During the last years, relevant models and tools have been designed by independent parties to evaluate the effectiveness of the publishing activity of (Open) datasets by public administrations and data holders at large.

With respect to data formats and formalisms, Tim Berners-Lee introduced a '5-star' model – each star representing a step towards 'maximum' reusability of the datasets:

- *(one star)* 'make your stuff available on the web (whatever format)';
- *(two stars)* 'make it available as structured data (e.g. excel instead of image scan of a table)';
- *(three stars)* 'non-proprietary format (e.g. csv instead of excel)';
- *(four stars)* 'use URLs to identify things, so that people can point at your stuff';
- *(five stars)* 'link your data to other people's data to provide context'.

The fourth and fifth level clearly refer to the Linked Data[13] formalisms. A set of costs and benefits are associated to each level[14].

In general terms, the ever-increasing role of data – constantly flowing from multiple sources (ranging for instance from sensors measuring air quality to the stream of social networks) – has been acknowledged by researchers and practitioners. Third parties consume data to create new

[12] See https://speakerdeck.com/u/epsiplatform/p/track-d-amalia-velasco-spain

[13] http://www.w3.org/standards/semanticweb/data

[14] http://5stardata.info/

businesses (e.g. via Application Programming Interfaces1, or APIs, made available to developers by big players like Google or Amazon and, increasingly, by smaller ones), accelerate scientific progress, improve social debate (Bizer and Heath, 2011). The so-called 'data deluge', i.e. the unprecedented availability of tremendous amounts of data, allows to rely on a constant flow of information, whose consumption is supposed to positively affect efficiency, transparency of markets and institutions, effectiveness of decision-making. However, this information needs to be properly represented and managed in order to become actually meaningful and reusable (Ericson, 2010), also preventing its misuse. Within this context, the underlying idea of the Linked data principles, proposed by Tim Berners-Lee in 2006, is to overcome such fragmentation by transforming the Web in a universally distributed database, within which data sets are identified, retrieved, represented and fused, with expressive query capabilities over aggregated data, similarly to how a local database is queried today. Several Public Administrations across the world are not only starting exposing their data sets as open[15], but also adopting Linked data principles within such initiatives , with tangible benefits related with internal efficiency and increased reusability of the datasets (Alani et al., 2006).

As another example of relevant guidelines on the subject, the W3C[16] has recently proposed a five-step framework to guide public administrations to publish the data they hold as open and engage external players in reusing them for all purposes. Again, a 5-star model is used. It encompasses:

- *(one star)* 'Be demand driven';
    - evaluating and considering communities' needs and specific demands;
- *(two stars)* 'Put data in context';
    - providing clear information about the data published, including links to projects already using it;
- *(three stars)* 'Support conversation about data';
    - enabling users comment the datasets published and the strategy itself;
- *(four stars)* 'Build capacity, skills and networks';
    - also through 'hands-on' activities, 'hackathons' and the like;
- *(five stars)* 'Collaborate on data as a common resource';
    - promoting all kinds of public-private partnerships.

The 'Open Data Census'[17] currently carried out by the Open Knowledge Foundation is aimed at exploring, for specific sets of data produced and managed at national level[18], its availability in a

[15] According to the definition provided by the Open Knowledge Foundation, "“A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike.”

[16] http://www.w3.org/2012/06/pmod/pmod2012_submission_5.pdf

[17] http://opengovernmentdata.org/census/results/

[18] Election Results (national); Company Register; National Map (Low resolution: 1:250,000 or better); Government Budget (high level – spending by sector); Government Budget (detailed – transactional level data); Legislation (laws and statutes); National Statistical Office Data (economic and demographic Information); National Postcode/ZIP database; Public Transport Timetables; Environmental Data on major sources of pollutants (e.g. location, emissions).

digital form; whether its format is machine readable; whether it is available free of charge; whether it is openly licensed.

5. Examples of mature PSI markets in Italy

As previously discussed, while new opportunities are being enabled by Open Data initiatives, some mature markets grounded on PSI already exist and have peculiar features. Their size (in terms of overall turnover) is anything but negligible. Moreover, being related with firms activities, land registries or real estate, such market have a broad set of end-users (potentially covering in at least one interaction a high share of the adult population). Applying Open PSI principles to such markets is therefore supposed to have a positive impact on competition dynamics and social welfare.

Hereafter are briefly described the cases of business register, cadastral information and legal information in Italy. A common structure is adopted. The nature of PSIH and the available datasets, the way datasets are supplied, the nature of reusers, the adopted charging principles, the estimated incomes for PSIH and the potential impact of an Open PSI approach are discussed. With ‘Open PSI approach’ we mean applying full open access and reuse to raw databases.

5.1 Business Register

PSIH and datasets

The Business Register ('Registro delle Imprese') is the widest repository of business and company intelligence information in Italy. It is supplied by InfoCamere, a company held by the system of the Chambers of Commerce (105 all across the Country), managed by their national Union ('Unioncamere'). The Business Register consists of the Register of Companies ('Registro delle Società') and the Register of Business Names ('Registro delle Ditte').

The data collection (upstream) is basically ensured by the Chambers of Commerce (for instance, each new firm has to register, under payment of a fee, to the local Chamber). Other information is provided by PAs and authorities, such as courts and social security institutions. InfoCamere is responsible for the digital and centralized database, including its real-time update.

The Business Register currently comprises information on 9 million individuals, 6 million firms and 900.000 financial statements (using XBRL standard). From a technical point of view, it is composed by three subsystems: the protocol; the business register and repository of economic and administrative facts ('REA'); the archive of official documents.

Data supply

Business information by the Italian system of the Chambers of Commerce is provided in several ways:

- specific data extractions can be requested (under payment of rights depending on the type of data, according with a price list) in each of the 105 premises;
- via the front-office service platform called 'Telemaco', which allows citizens and professionals to access information online and to upload information, under standard agreements (prerogatives and price list publicly available);
- via the online Business Registry, which allows to download single business profiles for free

and without registration;

- via reusers of raw data, such as the InfoCamere official distributors and Poste Italiane ('Certitel' service).

Reusers

The main reusers of data supplied by the Italian Business Register, i.e. the ones that develop value-added service based on the entire set of raw data, stipulated agreements with InfoCamere to gain direct access to the database. However, complying with specific prerequisites (regarding the size of the company, its technological equipment and the volume of information to be reused) is needed to become 'distributor', currently narrowing their number to 43. The reason of such policy mostly relates with the aim to preserve the quality of the information distributed. Interesting enough, the downstream market is rather concentrate, with the first 4 players (including Cerved, originally joint with InfoCamere, and the global player Crif) accounting for around 80% of the market share.

Business Register distributors supply value-added services – such as economic / financial assessments and market ratings – to a set of customers, especially banks (90% of which rely on such services in Italy). The overall annual value of such services reaches 250 million euros, with an estimation of business information market in Italy between 500 million and 1 billion euros[19].

Pricing model

As indicated by the Ministry of Economic Development (Law 580/93, art. 18) charging principles have to “take into account the average production costs and related services” (art. 18 Law n. 580/93). The charging strategy adopted by the Business Register is therefore 'cost recovery', regardless of the scope of the access and reuse.

As previously mentioned, fares depend on the type of information and its level of detail (i.e. the requested 'extraction' from the database), according to a publicly available pricing list, with no significant changes in the recent years.

Besides, raw data (i.e. direct access to the database) are available for 'distributors' via agreements that provide for the payment of an annual fixed subscription[20]), are supplied as direct access to the database. The price of access to the raw data is significant.

Incomes for the PSIH

Consistently with the multi-sided feature of a PSIH, the Chambers of Commerce have two main sources of income: (i) fees ('annual rights') annually paid by registered firms, including updates; (ii) fees paid by users and reusers, either fixed (annual subscriptions) or variable (single requests). For a single Chamber of Commerce, it is estimated as an average that 2/3 of the incomes relate with the former and the rest with the latter.

According with the 2009 official annual report, the total 2009 incomes for InfoCamere amounted for 99,7 million euros, divided as follows:

- 3,8 million euros from contributions by the Chambers of Commerce (to be negotiated each year in order to achieve break-even);
- 60 million euros from data supply (31 million of which from the 43 'distributors', the rest coming from 'Telemaco' agreements and single requests, the latter accounting for around 10

[19] Ancic (the association that gathers some 85% of the Italian business information players) estimates a (growing) value of around 750 million euros for this market in Italy.

[20] The average amount per distributor is around 700.000 euros.

million);

- 25,7 million euros from supply of services;
- 8,6 million from supply of products;
- 1,6 million from other sources.

Possible impact of an Open PSI policy

The example of the Italian Business Register shows how much the configuration of a mature market grounded on PSI matters for social welfare. On one hand InfoCamere seeks break-even in order to ensure continuity of data collection and supply: around 1/3 of its incomes derives from annual agreements with its 'distributors'. On the other hand, some current features of the downstream market structure seems to generate inefficiencies: (i) its high degree of concentration; (ii) its relevant entry barriers (as an effect prerequisites set by InfoCamere to officially become 'distributor'); (iii) its stagnating degree of innovativeness, especially within mid- and small-sized players. Moreover, such 'distributors' are *de facto* market intermediaries between citizens (even professionals) and the source data.

Granting a free and open access to InfoCamere raw data to all citizens would probably imply a positive selection in the downstream market, with potential new entrants, and boost competition (with a positive effect on quality and/or charges to end-users). The global value of the downstream market, as we know it so far, could probably decrease, but concentration as well. However, new branches could arise, mostly grounded on digital services. Besides, skilled end-users / citizens would have the opportunity to directly draw on source data, without having to rely on intermediaries for the most simple queries. This would however require to compensate InfoCamere from the 'loss' of missing sales, which can be estimated between 10-15 million euros (under the hypotheses that annual agreements remain in charge and are not affected by the new policy, with InfoCamere only losing incomes from single requests) and 60 million euros (under the unlikely hypotheses that the whole downstream market of distributors 'disappears' in light of the new policy).

5.2 Cadastral information

PSIH and datasets

The main national organism collecting and managing cadastral data (including land, cartography, data on real estate) is named 'Agenzia del Territorio' (henceforth labelled 'AdT'). Within its statutory mission, AdT carries out the following PSI-related tasks: registration of cadastral information, cartography and real estate rights and mortgages; setup, update and management of a nationwide comprehensive real properties register; decentralization of cadastral functions, integrating its activity with the ones under competence of local authorities.

Moreover, AdT manages a Real Estate Market Observatory ('Osservatorio del Mercato Immobiliare'), providing with updated information on the real estate market in Italy, used both as a resource for business services by private reusers and for tax avoidance assessment by public authorities.

The databases held by AdT result from a merge of the former 'Land Cadastre' and 'Building Cadastre'. They currently encompass 82 million rural parcels, 63 million urban audits and 340.000 maps in vector format; 180.000 real estate quotations and more than 40 million digital notations and deeds mainly concerning real estate rights transfer. A 'self-updating' system is provided by the

technical partner SOGEI, including online applications available for the update of digital records (e.g. by notaries) through the web-based platform 'Sister', which is used also for the supply of information.

Data supply

Online consultation of the cadastral maps and the land parcel where a particular building unit is located is free of charge via the AdT website. Online access is increasing and currently represents around 75% of the interactions with users and reusers. Around 50.000 user accounts are currently active, enabling online purchases. However, specific agreements to gain access to all databases and their update can be set with AdT (this is the case of the cadastral data reusers).

Substantial investments are being made, especially for the improvement of the existing information system (21 million euros for the years 2009 – 2011).

Reusers

Apart from the 50.000 individuals downloading single pieces of information (and the requests made at physical desks), cadastral and real estate data are mainly used by several types of actors (such as real estate agents, entrepreneurs, construction companies and public authorities, including the Italian fiscal agency) as an important source of information in support of their activity.

However, considering reusers only the players who develop value-added services grounded on public datasets, their number is currently close to 7000. The reuse market is basically composed of professionals and business information firms, the latter comprising around 100 firms, among which multinational brands (e.g. Crif) and Italian small / medium enterprises who provide Internet-based services to banks and other players.

The issues at stake are anything but negligible. Recently, the introduction by AdT of value-added services addressed to banks (typically the most important customers of cadastral data reusers) and the contextual increase by around 500% in the charges for raw data (the so-called 'elenco soggetti', a key resource for reusers since it allows a constant update of the data, which otherwise can be useless, since changes can occur 'overnight') represented an area of legal conflict between AdT and some of its reusers, who interpreted the strategy of the former as an attempt to foreclose the downstream market. Even if no final decision has been issued, the former pricing structure has been re-established.

Pricing model

As a general principle, AdT adopts a 'cost-recovery' model with additional fixed fees for (commercial) reuse. More in detail, cadastral data is free of charge for all PAs, while:

- database queries (via 'Sister') are subject to a registration fee (200 euros '*una tantum*') and a fee of 30 euros per account/year (contribution to the implementation of the information system);
- agreements for reusers have the same conditions as above, plus a fixed annual fee of 1000 euros for 're-use'.

Charges for the Real Estate Market Observatory data are zero for some basic services (such as quotations and sales volumes from 2009), while a 'cost-recovery' principle is applied to real estate quotations from 2002 (between 1.500 euros and 2.200 depending on the coverage). Specific agreements for data sharing with real estate intermediaries associations are as well put in place. Free of charge raw data are besides provided to research institutes and public bodies under special agreements.

Incomes for the PSIH

The incomes deriving from cadastral and real estate data supply by AdT are not disclosed. However, 'fees' by users and reusers are technically taxes and are therefore transferred from AdT to several Government branches. The estimated total budget is close to 600 million euros, with 100 out of 9800 full-time equivalents directly involved in PSI reuse facilitation. Some estimations by the POPSIS study assume a an annual cost of around 158 million euros for the cadastral service and 9,4 million euros for the real estate observatory.

Possible impact of an Open PSI policy

With respect to cadastral information, two undesirable market structures can be detected, both of which not complying with transparency, fair charging principles (for re-users and end-users) and non-discrimination principles (PSI Directive): i) a monopoly on the source data, with a concentrated downstream ecosystem of operators providing services to citizens, firms and banks (in a sort of 'double-marginalization' scenario detrimental for consumers' welfare); ii) an attempt of market foreclosure by AdT by raising charges for re-use of raw data (+ 500%) and, at the same time, by designing and releasing new value-added services, especially for banks. Besides, pricing principles set by AdT include a sort of 'ex-ante' fixed royalty for commercial re-use, which may be conceived as discriminatory.

In this case, the boundaries of the public task are they key issue: while it seems clear that AdT cannot enter the downstream playground without distorting competition, the re-user market itself is concentrated. The 'struggle' between PSIH and reusers to internalize the value of the cadastral data market is therefore detrimental for end-users in terms of charges and data availability. Free online access to single records could then improve efficiency, allowing citizens to rely on source data, while leaving the task to develop value-added services to the market.

5.3 Legal information

PSIH and datasets

In Italy, legal information usually subject to reuse mainly encompasses national legislation (stored in the Italian law archive 'Gazzetta Ufficiale'), regional legislation (held by regional administrations and commonly available through their official websites), administrative procedures (produced and held by national ministries and authorities) and official verdicts issued by the Supreme Court (through its Documentation Centre), the latter being around 54.000 per year.

Data supply

The aforementioned sets of documents are, by definition, of public domain (although '*sui generis*' rights may be applied the structured archives). Access to the maxims archive requires the payment of a fee (usually a flat subscription). Moreover, the unabriged versions of the verdicts are available only in papery format, therefore requiring digitization, partially carried out manually as a first step for reuse. Digitization costs reach around 200.000 euros per year for an average-sized reuser. The portal 'Normattiva' of the Italian Government supplies free of charge a structured archive of the laws currently in force in Italy.

Reusers

The widest part of the market grounded on legal information consists of publishing houses (3/4 big players, such as Zanichelli, Giuffrè and Wolters Kluwers, plus a few other minor operators) which provide users (mainly professionals) with editorial products including digitized and semantically linked archives, updated registers, as well as other publications customized by sub-disciplines of

interest. Professionals, lawyers and researchers represent their main customers. The size of the market, i.e. the overall revenues deriving from editorial products based on legal information, is roughly estimated to be around 200 million euros. Incumbent players have built a strong reputation across time, either because of the acknowledged high quality of their publications and/or in light of the improvement of their products, including digitization and semantic web-based services.

Pricing model

The Documentation Centre of the Italian Supreme Court applies a 'cost-recovery' model based on the required engagement of its servers, with an annual fixed fee depending on the category of the reuser (671,39 euros for public sector bodies, 1007,09 euros for lawyers and other professionals, 1342,79 euros for publishing houses) that grants access for 1000 minutes (only considering the actual time necessary to fulfil a request). Pricing on the exceeding time is between 0,50 euros and 1,25 euros per minute. Otherwise, a smaller annual fee can be paid (103,29 euros) regardless of the category of reuser, with a variable fee amounting to 1,25 euros per minute of connection.

Incomes for the PSIH

Not available.

Impact of an Open PSI policy?

The concentration degree of the legal information market holds some strong economic reasons. Substantial sunk costs, tacit knowledge, reputation and customer fidelization make entry barriers quite considerable. Moreover, the editorial work achieved by publishing houses (maximization, commenting, update) is most often necessary to correctly interpret the implications of laws and verdicts. However, with respect to the latter, open paradigms could be applied, at least on verdicts concerning specific disciplines of interest for professionals or researchers. This kind of approach would not affect the downstream market (for the reasons expressed above) while representing, at the same time, an opportunity of full disclosure of relevant contents for professional or research purposes. Moreover, counting on already digitized content would allow publishers to re-invest savings on product improvement.

Conclusions

According with the available examples, it is possibile to highlight several kinds of opportunities related with the Open Government Data paradigm. Where efficiently adopted, the latter may allow to:

- enable the creation of new value added services (at least partially) grounded on ‘sole-source’ datasets held and managed by public administrations;
- feed emerging markets massively grounded on the elaboration, aggregation (and / or intermediation) of data flows with new resources, also lowering entry barriers for newcomers and start-ups;
- explore new technological paradigms (e.g. Linked Data) in order to better exploit network externalities inherently characterizing data economies and the Internet;
- foster transparency and accountability of national and local public administrations;
- stimulate democratic participation, even at local level, around key issues and responding to emerging needs.
- increase the efficiency of public administration, mainly by renewing internal processes,

also benefiting from the work of community of reusers validating and integrating data.

At the same time, significant obstacles can be identified, and in particular:

- existing mature markets grounded on closed PSI still represent, by far, the highest share of PSI reuse; in those cases, the shift towards an Open model would require significant public investment;
- the volume and relevance of the datasets disseminated as open still seems insufficient; data sources are fragmented, real-time data are poorly available;
- (exclusive) arrangements with third parties hamper the opportunities of data reuse by potential new entrants;
- the potential benefits of the Open Data paradigm for public administrations and reusers are not perceived in their full extent;
- no structured process is currently in place in most of the Public Administrations in order to actually incorporate feedbacks, integrations and other kinds of data manipulation performed by third parties when reusing PSI;
- a need harmonization between Public Administration between the content, type and format (e.g. metadata formalism) of the datasets exposed as Open is globally perceived in order to increase the potential of data reuse.

References

Aichholzer, G., Burkert, H., *Public Sector Information in the Digital Age. Between Markets, Public Management and Citizen's Right*, EE, Celtenham, 2004

H. Alani, D. Dupplaw, J. Sheridan, K. O'Hara, J. Darlington, N. Shadbolt, and Carol Tullo , *Unlocking the Potential of Public Sector Information with Semantic Web Technology*, 2006

Arthur, W. B. (1996). *Increasing Returns and the New World of Business*. Harvard Business Review (July/August), 100– 109

Benkler, Y., *The Wealth of Networks*, 2006

T. Berners-Lee, W. Hall, J. A. Hendler, N. Shaboldt, D. J. Wieizner, *A framework for Web Science*, Foundations and Trends in Web Science, 1(1), 2006

T. Berners-Lee, *Linked data-design issues*, W3C, ACM – Press, 2006

T. Berners-Lee, M. Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*, HarperInformation 2000

Bizer, C., Heath, T., *Linked Data. Evolving the Web into a Global Data Space*, 2011

David, P.A., *Can 'Open Science' be Protected from the Evolving Regime of IPR Protections?*, Journal of Institutional and Theoretical Economics, 2004

J. Ericson, *Net expectations - what a web data service economy implies for business*, Information Management Magazine, 2010

Gallagher, William T., *What is a "Law and Society" Approach to Intellectual Property?*, Publications. Paper 30, 2009

Hess, C., Ostrom, E. 2003. *Ideas, artifacts and facilities: information as a common-pool resource*, Law and Contemporary Problems, 66 (1–2), 111–146

Hofmokl, J., *The Internet commons: towards an eclectic theoretical framework*, International Journal of the Commons, 4(1), 2010

Koshi, H., *Does Marginal Cost Pricing of Public Sector Information Spur Firm Growth?*, ETLA, 2011

LAPSI Position Paper n°4, *The "licensing" of public sector information*, 2011

Lundvall, B., *The New Knowledge Economy in Europe: A Strategy for International Competitiveness and Social Cohesion*, 2003

Mayer-Schönberger, V., Z. Zappia, Z., *Participation and Power: Intermediaries of Open Data*, 1st Berlin Symposium on Internet and Society October 26th – 28th

MEPSIR (Measuring European Public Sector Information Resources) (2006), *Final report of study on exploitation of public sector information – benchmarking of EU framework conditions*, Executive summary and Final report Part 1 and Part 2

MICUS Management Consulting (Fornefeld, M., G. Boele-Keimer, S. Recher and M. Fanning) (2009), *Assessment of the re-use of public sector information (PSI) in the geographical information, meteorological information and legal information sectors*, Dusseldorf, Germany

OECD, *The knowledge-based economy*, 1996

OECD, Working Party on the Information Economy, *Digital Broadband Content: Public Sector Information and Content*, 2006

Office of Fair Trading (2006), *The commercial use of public information (CUPI)*, OFT861, United Kingdom.

Newbery D., R. Pollock and L. Bently (2008), *Models of public sector information provision via trading funds*, Department for Business, Enterprise and Regulatory Reform, United Kingdom

Pollock, R. (2009), *The economics of Public Sector Information*, University of Cambridge Cambridge Working Papers in Economics 0920, Faculty of Economics, University of Cambridge

Pollock, R. (2011), *Welfare gains from opening up Public Sector Information in the UK*, University of Cambridge

Ricolfi, M., *Making Copyright Fit for the Digital Agenda*, 12th EIPIN Congress 2011 - Constructing European IP: Achievements and new Perspectives Strasbourg, February 24th – 25th, 2011 European Parliament

Ricolfi, M. et al., *The 'Principles Governing Charging' for Re-use of Public Sector Information*, Informatica e Diritto, 37(1/2), 2011

P. M. Romer, P. M., *Increasing Returns and Long-Run Growth*, Journal of Political Economy, 94(5), 1002–1037, 1986

J. Rowley, *The wisdom hierarchy: representations of the DIKW hierarchy*, Journal of Information Science, 33(2), 163 - 180, 2007

H. A. Simon, *The many shapes of knowledge*, Revue d'économie industrielle, 88, 1999

J. E. Stiglitz, P. R. Orszag, J. R. Orszag, *The Role of Government in the Digital Age*, 2000

van Eechoud, M., van der Wal, B., C*reative commons licensing for public sector information. Opportunities and pitfalls*, Institute for Information Law, University of Amsterdam, 2008

H. R. Varian, J. Farrell, C. Shapiro, *The Economics of Information Technology*, Cambridge University Press, 2004

Vickery, G., *Review of recent studies of PSI-reuse and related market developments*, European Commission 2011

Weiss, P., *Conflicting Public Sector Information Policies and their Economic Impacts*, U. S. Department of Commerce, 2002 Economic Analysis & Policy, Vol. 39 No. 1, March, pp. 127-142

U. Witt, T. Brökel, T. Brenner, *Knowledge and its Economic Characteristics. A Conceptual Clarification*, Jena Economic Research Papers, #13, 2007

Yiu, C., *The Big Data opportunity*, Policy Exchange, 2012

Legislative references and official documents

European Commission (2010), “A Digital Agenda for Europe”, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM(2010)245 final, 19 May 2010

DIRETTIVA 2003/98/CE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 17 novembre 2003 relativa al riutilizzo dell'informazione del settore pubblico

European Commission, Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL Amending Directive 2003/98/EC on re-use of public sector information, COM(2011) 877 final

D.Lgs. 36/2006 “Attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico”
DIRETTIVA 2007/2/CE DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 14 marzo 2007 che istituisce un'Infrastruttura per l'informazione territoriale nella Comunità europea (INSPIRE)

Legge 27 dicembre 2006, n. 296 “Disposizioni per la formazione del bilancio annuale e pluriennale dello Stato (legge finanziaria 2007)”, pubblicata nella Gazz. Uff. 27 dicembre 2006, n. 299, S.O. n. 244

Legge 29 dicembre 1993, n. 580, “Riordinamento delle camere di commercio, industria, artigianato e agricoltura”, pubblicata nella Gazz. Uff. 11 gennaio 1994, n. 7, S.O.

Legge 30 dicembre 2004, n. 311, “Disposizioni per la formazione del bilancio annuale e pluriennale dello Stato (legge finanziaria 2005)”, pubblicata nella Gazz. Uff. 31 dicembre 2004, n. 306, S.O. n. 192

Decreto-legge 31 marzo 2005, n. 45 “Disposizioni urgenti per la funzionalità dell'Amministrazione della pubblica sicurezza, delle Forze di polizia e del Corpo nazionale dei vigili del fuoco”, pubblicata nella Gazz. Uff. 1 aprile 2005, n.75

Regione Piemonte, Legge regionale n. 24 del 23 dicembre 2011 “Disposizioni in materia di pubblicazione tramite la rete internet e di riutilizzo dei documenti e dei dati pubblici dell'amministrazione regionale.” (B.U.29 Dicembre 2011, n. 52)

Infocamere, Bilancio di esercizio 2011, http://www.infocamere.it/doc/Bilancio_InfoCamere_2011.pdf