

Modelling the Economic impact of information reuse in Spain

Modelado del impacto económico de la reutilización de información en España

Tutora: Dra Carmen De Pablos Heredero

Alberto Abella García
Máster Oficial en Organización de Empresas
Departamento de Economía de la Empresa (Administración, Dirección y Organización)

Modelling the Economic impact of information reuse in Spain

Abstract

The increasing economic impact of Public Sector Information (PSI) reuse contrasts with a low level of understanding of how the economic value is created. A full theoretical model is drafted based on models developed for other disciplines. A subset of this model, the relation between reusability, scope and type of information and business value is tested with real data coming from the companies which reuse information (infomediaries) and data sources from the Spanish market. For testing with real data a new standardisation and an associated metric, MELODA, is defined by the author. Analysis about the relation between reusability dimensions, scope and type of information is tested against the information consumption by companies which provide economic value based on PSI reuse. Preliminary results show relation between access to information and information consumption.

Keywords

Public sector information, open data, MELODA, economic impact, open government data.

Index

1 Introduction.....	6
1.1 Background.....	6
1.2 Research question.....	8
1.3 Objectives of the study.....	9
1.4 Scope and previous considerations of the study.....	9
2 Theoretical concepts	11
2.1 Analysis of theoretical models	11
2.2 Theoretical Contribution 1: Global model of the economic impact of PSI reuse.....	16
2.3 PSI reuse value chain.....	19
2.4 Theoretical Contribution 2: MELODA metric of information reusability.....	22
3 Methodology.....	27
3.1 Methodology for data collection of infomediary companies.....	27
3.2 Methodology for sampling public data sources.....	27
3.3 Current situation of PSI reuse in Spain	30
3.4 Analysis and Time to Implement Reuse (TTIR).....	33
4 Results.....	34
4.1 Data consumption by infomediary companies	34
4.2 Reusability of public data sources.....	35
4.3 Impact of reusability conditions on data consumption.....	36
5 Conclusions and future goals.....	40
6 Annex I. Work Cited List.....	43
7 Annex II. Data sources.....	47

Index of Tables

Table 1, Legal dimension of MELODA.....	24
Table 2, Access dimension of MELODA.....	25
Table 3, Technical Standard dimension of MELODA.....	26

Table 4, Sampling of Spanish data sources.....	28
Table 5, Scope of PSI.....	29
Table 6, Type of PSI.....	29
Table 7, FOI and PSI reuse legislation in Spain and regional governments.....	33

Illustration Index

Illustration 1, Model based on DEA (Kauffman & Kriebel, 1998).....	12
Illustration 2, PSI value Germany (PIRA, 2000).....	12
Illustration 3, Model for PSI impact analysis (MEPSIR, 2006).....	13
Illustration 4: Demand / Cost functions (Pollock, 2008).....	14
Illustration 5, 4-stage model for Open Government Data (Kalampokis, 2011).....	15
Illustration 6, 5-stage maturity model (EU, 2010).....	15
Illustration 7: Global PSI reuse economic impact model. Own elaboration.....	17
Illustration 8, Proposed Model for infomediary economic impact. Own elaboration.....	17
Illustration 9, PSI reuse value chain (Abella, 2011) for Rooter.....	20
Illustration 10, Economic value of PSI types of information (PIRA, 2000).....	21
Illustration 11, Economic value of PSI reuse (MEPSIR, 2006).....	21
Illustration 12, Licensing conditions change in Spanish PSI sources Nov-11 to Sep-13.....	30
Illustration 13, New public opendata portals in Spain. Based on CTIC data.....	30
Illustration 14: Type of information used by infomediaries in Spain (Red.es, 2012).....	31
Illustration 15, Type of information used (Red.es, 2011) and (Red.es, 2012).....	34
Illustration 16, Increment of infomediary information consumption by type 12-11.....	34
Illustration 17, Global reusability of datasources in Spain. Own elaboration.....	35
Illustration 18, access dimension vs increment of information use 12-11.....	36
Illustration 19, access dimension vs increment of information use 12-11 by scope.....	37
Illustration 20, technical standard vs Increase of Information use 12-11.....	38
Illustration 21, technical standard vs Increment of Information use 12-11 by scope.....	38

1 Introduction

Science is the quest for truth. Consultancy is the professional field in which practical solutions are provided to customers. But consultancy is also a field in which customers buy products and services that are fashionable. I have worked for twenty years in the consultancy field and I can assure you that it is extremely difficult to sell new services which are not fashionable.

Information reuse is definitely fashionable today (2013). But the huge expectation about its economic and social impact contrasts with the actual data and lacks a sound theoretical model.

In this preliminary analysis we only expect to provide a closer look on the subject and to use a language that both, Academia, and Consultancy can understand and profit from.

1.1 Background

1.1.1 Political background

Information has been defined as the new oil for the knowledge industries. Information coming from the public sector (PSI) has been described by Vice-president of the European Union Neelie Kroes as a “gold mine of opportunities and the new oil”¹. Similar statements remarking this importance can be found in (Shapiro & Varian, 1999; Castells, 2000).

The PSI publishing has attracted the attention of policy makers since its beginning. Strangely, it received an extensive support before having a sound modelling of the discipline, but without showing real proofs of their economic impact (EU, 2003; UK, 2005; OECD, 2006; Spain, 2007). As Burton has stated “*The commercial re-use of public sector information (PSI) is clearly appealing for governments and their agencies*” (Burdon, 2009). Possibly due to it, some of these policies are being reviewed (EU, 2013a).

An improved access to and use of PSI has been identified as of major importance for all economies (OECD, 2006; Vickery, 2013).

¹<http://www.youtube.com/watch?v=9Jq4Qy1UeAE>. Speech of Neelie Kroes ofr EPSI platform event on the value of open public data: 'data is the new oil' (16-3-12)

The European Union provides a consistent support in its research priorities to PSI reuse², which is consistent because it started as soon as 2000 (PIRA, 2000) and remains now (2013)³. Future support is also drafted⁴ in the Horizon 2020 R+D programme focusing on information related with research, Open access, so public supported research results were available for other research bodies and initiatives (EU, 2013b).

1.1.2 Economical background

The overall impact of the public sector information reuse has been evaluated extensively, ranging from 0.25% of the total aggregated GDP for the European Union and Norway (MEPSIR, 2006), 1% of GDP in (PIRA, 2000), and up to 1.7% UE's GDP (Vickery, 2013), where the direct and indirect economical impact is taken into account.

These differences reflect the variability in the estimation of methodologies used, and, the scope of what is inside the PSI economic impact. Furthermore, estimations of the potential of the market contrast with the actual data coming from the surveys done. For example, in Spain, where this study will focus, 0.25% of GDP results in a potential market of 2.680 Million € for 2012, while surveyed data in (Red.es, 2012) only accounts for 400 Million (15% of the potential).

If previous analysis were right about the real market potential – a conservative mean could be 1% GDP – it is easily understood, if we compared it with EU economy growing rate in 2012 – -0.2% – , the drawing of interest on this field.

Summarizing, this background analysis shows that we are in a very attractive field of study, with a strong support from the EU policy-makers but showing a weak theoretical basis.

1.2 Research question

It is clear that PSI reuse is showing a growing economic importance. Very diverse policies supporting PSI publishing have been identified (Huijboom & Van der Broek, 2011). However most, if not all, of these policies show the expected importance of the economic impact of PSI

2 Open data is a subset of informations of the PSI.

3 <http://cordis.europa.eu/fp7/ict/content-knowledge/>

4 <http://bit.ly/GLOJh7>

(OECD, 2006).

In this situation the research question about what factors affect the economic impact of PSI reuse appears more relevant than ever. Even more, which are the drivers for making the real market to reach estimations? Does this impact depend on the type of information? Or would it depend on its legal framework, its basic pricing, or any other characteristics linked to the reusability of the information, or should we consider other factors? How long does it take since the information is released to create economic impact?

This work briefly analyses the theoretical framework in order to address these questions. It will focus in the search of any linkages between the reusability conditions and the real market. A first test with data will be carried out.

Impact of the research question

Therefore the research question proposed is: *“Is it possible to get more accurate estimations on the economic impact of PSI reuse?”* As far as we were able to provide such estimation, some important questions will be automatically answered too:

- Are our current PSI publishing policies creating real value for the overall society?
- Does the return on these policies justify further investments? To what extent?,
- Are the public administrations profiting from these publishing policies? Should we, therefore, include them into the economic impact of the PSI reuse policies?
- Should the freedom of information legislation include principles to promote economic reuse? Which ones?

The lack of a model is preventing us from coming up a quantitative analysis, a full estimation of their economic impact. As a result, public sector information publishing policies could be inadequate, since they are not based on the real impacts on the economy but on the satisfaction of some civil rights and accountability principles. If we were able to answer the research question we could design policies which would reduce the time to expand the real market to its potential. We, probably, could also be more effective on the definition of the standards to release information and of course we could focus our efforts on the information

which creates the biggest value for the economy.

1.3 Objectives of the study

The objective of this study is to draft a **theoretical framework to study PSI reuse economic impact**. The first milestones to reach it are :

- To specify a more detailed modelling of the best known part of the global model
- To test it against data in a real market.

As a consequence of the general goal four contributions were developed in this work.

- 1) A **theoretical framework for a global economic impact of PSI reuse**
- 2) A new **standardization of reusability** and the criteria to assessed it.
- 3) The evolution of a metric (and methodology attached) to provide **quantitative assessment of reusability of the Public sector information** released (Abella, 2011).
- 4) The definition of the **Time to Information Reuse (T_{TIR})** in order to assess additional effects to promote PSI reuse and its economic impact.

1.4 Scope and previous considerations of the study

Although many other aspects can be considered on PSI reuse, this study is just focused on the economic impact and not in other factors, as social impact, accountability of the public administration, transparency of implementations, etc.

These data only affects a part of the overall model (described in Illustration 7: Global PSI reuse economic impact model. Own elaboration.). Information analysed in this study includes only digital information, coming from those public entities described as Public Sector Information holders (PSIH) Pollock, 2008. These PSIH are the responsible entities for publishing the information. They, sometimes, are not the real aggregators neither the processors of the information.

In the sampling of data, only data coming from Spanish Market has been selected. These data include the companies which are creating value based on PSI based on the analysis depicted

for the years 2011 and 2012 by Red.es (Red.es, 2011; Red.es, 2012). Additional data has been collected for the PSI sources available. See it in 7 Annex II. Data sources.

Structure of the document

Next chapter focuses on the theoretical models, and and the value chain of PSI reuse. Also depicts the theoretical contributions of this essay. First: A global model to assess the economical impact, and with further detail for the case of *infomediary*⁵ companies. Second theoretical contribution is a metric and a methodology to assess data sources in order to relate them with the economic impact.

Third chapter explains the methodology for the collection of data. It includes both, companies' data and sources' data. Then a review of the Spanish market and a description of the analysis to be carried out are provided.

Results are analysed in next chapter from different points of view. Fifth chapter summarises the conclusions based on the analysed data while sixth chapter provides some guidance for further investigations.

The first Annex includes bibliography, and second further information about the sampling made, and provides examples on how the metric MELODA assesses a dataset.

5 In this essay infomediaries are companies profiting from the PSI reuse

2 Theoretical concepts

This chapter will set the background of the modelling of PSI reuse economy. After introducing the unique characteristics of the economy associated to PSI reuse, it will focus on the review of existing specific approaches, which provides partial analysis for our research question in the context of the organizational knowledge creation theory (Nonaka, 1994). Secondly, the proposed theoretical model is described and one of its parts, the one that will be tested with real data, is analysed in further detail. Then, for this part of the model, the value chain of PSI reuse is described in order to understand which agents are involved and what role they play. It ends with a description of the theoretical framework for assessing data information sources according to a new standardization of reusability.

2.1 Analysis of theoretical models

Importance and uniqueness of PSI reuse economy

Why is open access to public sector information important for the economy? Because the public sector is a major, even the dominant, producer of many kinds of information (Pollock, 2008; Janssen, 2011). Besides this, is somehow regulator on the flow of this information and it provides the legal frameworks which enable to create added value on top of the information.

PSI reuse presents qualitative and quantitative differences with other economic fields. The first differential point is that we are dealing with **digital information**. From the economic point of view, this information has a copy cost close to zero, but with a high production and maintenance cost (Shapiro & Varian, 1999; Pollock, 2008). The second difference to consider is that part of this information is published as a **result of transparency policies** and therefore, it is an assumed cost for the governments. Therefore economic modelling or return of investment calculus is out of the scope of the Public Administrations under this approach.

The third difference is that public sector information is a subject with a **virtuous economic circle**. The bigger ability to connect PSI sources and the higher availability, the more useful and more added value is possible to be created.

Analysis

Although a quite extensive search has been made, it has not been possible to find an economic model for the conditions of the public sector information reuse. Therefore similar or potentially related academic and consultancy documents will be summarised in this section.

It is evident, but a need to be remarked, that PSI reuse economy is a fully digital economy. Digital economy differs from 'regular' economy in that, once created, the product reuse presents a marginal replication cost, and for most of the analysis could be approached to zero. PSI reuse is an extreme case of this economy. Difficulty in its estimation is similar to the estimation of the economic impact of investments in Information technologies.

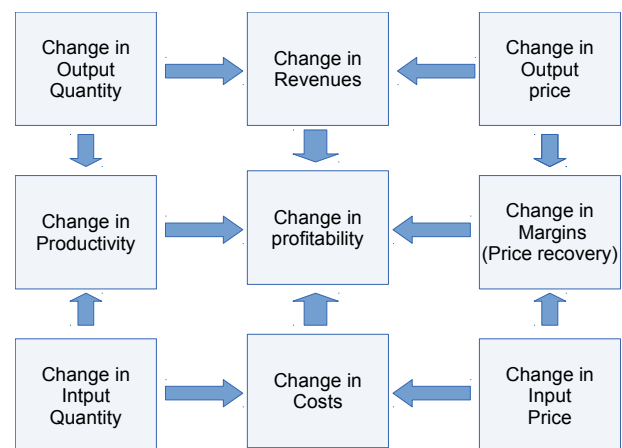


Illustration 1, Model based on DEA
(Kauffman & Kriebel, 1998)

A model to measure such impact was created by Kauffmann and Kriebel as early as in 1988, in which they apply the *Data Envelopment Analysis* (DEA) (Kauffmann & Kriebel, 1988). Some of these principles could be applied for a potential model to assess PSI reuse.

The PIRA analysis encompasses all firms that are in one way or another related to PSI, based on broad estimates using national accounts data (PIRA, 2000). PIRA takes the size of the information industry as an upper bound proxy for this market, particularly for estimating the US market. Based on this model, next figure shows the

expected impact for Germany.

The MEPSIR analysis is based solely on the surveyed added value by all first-order re-users, focusing on how much added value can be attributed to PSI re-users. The analysis excludes second order impacts as identified in Illustration 7.

It should be noted that this economic impact is based on expectations from the sampled sources but not the ones coming from real sampled data.

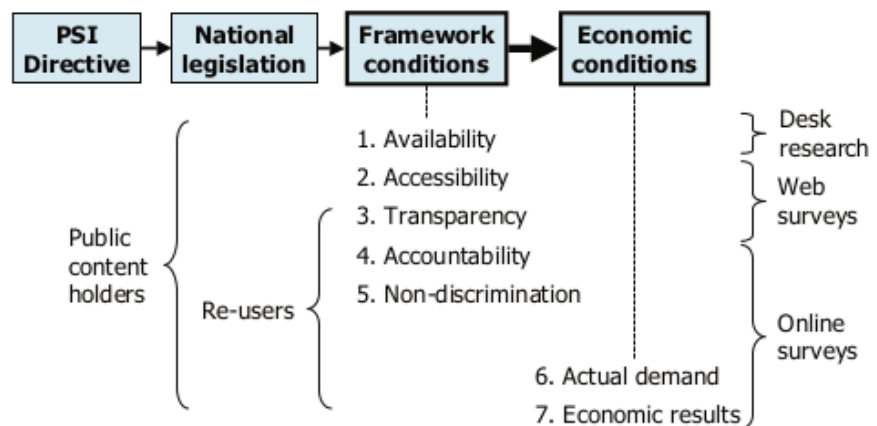


Illustration 3, Model for PSI impact analysis (MEPSIR, 2006)

The Pollock model, considers in its analysis only three main sources of PSI information: geospatial data, meteorological information and official statistics. His study is quite focused on how a single entity could be sustainable and how pricing and governance models interact. It does not take into account external effects of the information release and associated turnover for society. And in this sense this model attends only to the basics of the identified model in Illustration 7.

It is interesting to point out that incentives to subsidized models should compromise fixed costs and variable costs (those attending to the information output) in order to maximize incentives to improve performance and, at the same time, being innovative.

And most interesting of all is the statement “*any of the charging policies discussed could be used successfully if a independent, transparent and coherent governance structure were in place*” so it is not so dependent on the method used but in the coherence between charging policies and governance.

However the analysis only examines implicitly the impact on the overall society, impact that should be assessed in order to determine the level of subsidies to be assigned to the publisher (PSIH) entities.

This study is based on the regular price creation in this market, which is a hypothesis to be confirmed. In fact within research it is clearly identified that several industries could flourish if a marginal cost approach is used and therefore the linear hypothesis is a non-sense. Besides this, transaction costs (those associated to the payment, are also neglected as irrelevant and in favour of a more conservative hypothesis). Validation of such estimation is not provided and examples found by the author deny this approach.⁶

On top of that, regular price vs demand curve cannot be assumed as linear when price goes to zero, so most of the hypotheses presented in the research are not as solid as it should be required for a solid theory. Besides this valuable analysis of elasticity on the demand warns that the long-run could change as long as new applications and uses will be developed for the current information provided.

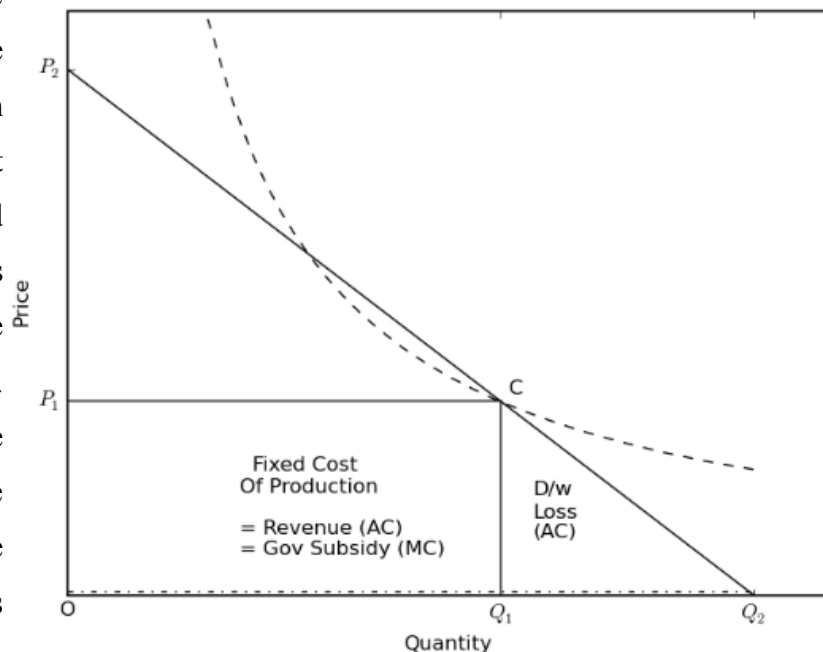


FIGURE 1. Illustrative demand and cost functions. Shown is a linear demand curve for a product with fixed costs and constant marginal costs (approximately equal to 0). Marginal cost (dot-dashed at very bottom of figure) and average cost curves (dashed) are shown.

Illustration 4: Demand / Cost functions (Pollock, 2008)

Total welfare = T(consumer surplus, producer surplus, government)

⁶ AEMET (Spanish Meteorology Agency) turn the free access to information into a pay-per-use charging policies. During the first year they spend more than 350.000 € only in the systems to support sales. This suppose a non-neglectible percentage of their revenue budget of 26 M€. <http://bit.ly/19JYyFh>

PSI holder entities, those public entities which are the owner and publishers of the information, have large fixed costs to collect, process and filter the information, whenever they have marginal costs⁷ on publishing and creating copies of it (Pollock, 2008).

Kalampokis refers a 5-stage model for e-government described in (EU, 2009) (see Illustration 5). In this model, open government data (which in this case is equivalent to PSI) is determined since stage 1 to be published. However this model does not pay dedicated attention to the fact that the public information released would be reusable as remarked by Alani (Alani et al, 2007). On the contrary, their analysis focuses on the complexity of the transactions that the citizens are able to interact with their administrations.

Kalampokis et al. model relates added value provided to the information and organizational and technological complexity as described in Illustration 5. It also addresses one of the biggest challenges for the information reuse which is the standardisation from different public sources. Especially when several of them publish the same or quite related information. However this model defines a stage two in which Public Administrations have a connection of their

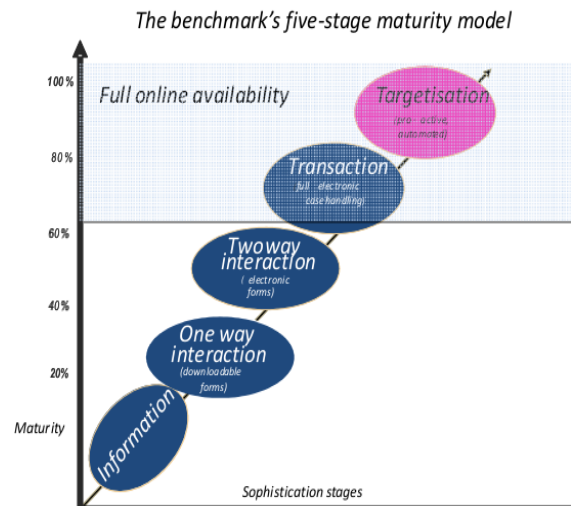


Illustration 6, 5-stage maturity model (EU, 2010)

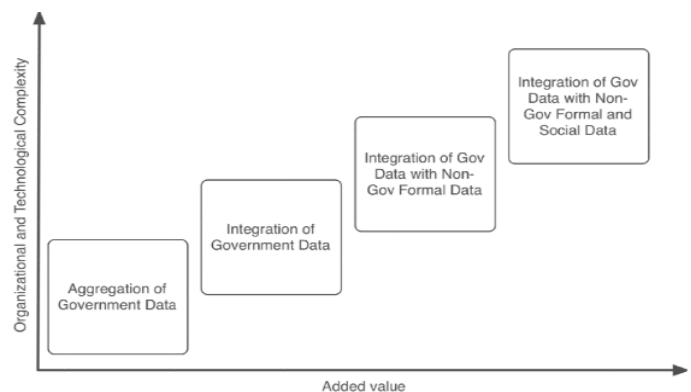


Illustration 5, 4-stage model for Open Government Data (Kalampokis, 2011)

⁷ Only digital information is considered in the scope of this study. Currently it does not impose an important restriction to the global effect.

data. It implies shared standards, compatible technologies and co-ordination procedures for the common publishing. This stage two is quite far from current and short term future situation. The emerging portals publishing information count on with weak implemented standards for the released information. So their applicability to current situations is limited.

2.2 Theoretical Contribution 1: Global model of the economic impact of PSI reuse

The model described in Illustration 7 provides a theoretical model for addressing most of the effects around economic impact of PSI reuse. The description of the proposed model comprehends these elements:

Line 1 reflects the most common studied line of value creation. This is the economic turnover of those companies which provide added value services based on PSI. Examples of this line include those companies providing financial solvency, geographical based services, etc. This will be the focus of the data analysis in this research.

Line 2 reflects the economic impact (savings, additional services) for the use that citizens do based on public information. Examples of this line include open science data information released under, etc.

Line 3 show the use that third sector organisations create based on PSI ranging from apps non-profit, to any other services which NGO provides to the citizens.

Line 4 reflects savings created in the public administrations as main users of the released data.

Line 5 reflects the impact on citizens because of the services provided by the NGO. Same way line 6 reflects the economic impact on products and services of the consumers of those products created by the professional providers (infomediaries).

PSI ECONOMIC IMPACT

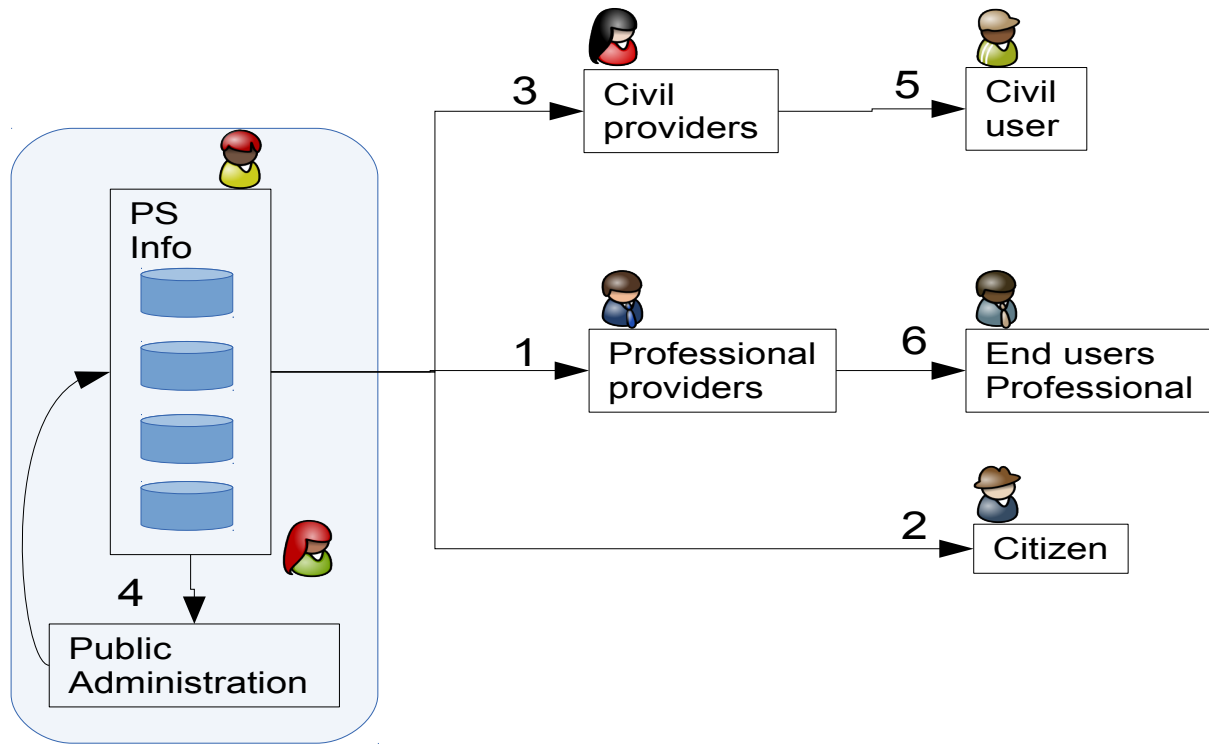


Illustration 7: Global PSI reuse economic impact model. Own elaboration.

2.2.1 Analysis of the Spanish case

The model analysed only studies the line 1 in the scheme of Illustration 7, meaning the relation between the revenue for the professional re-users and the reusability of information.

Next illustration

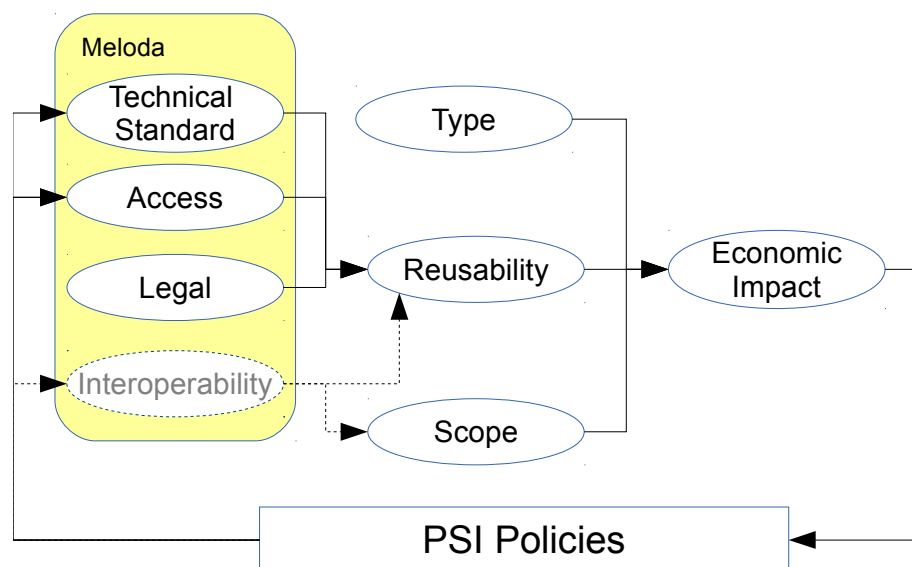


Illustration 8, Proposed Model for infomediary economic impact.

Own elaboration

describes the elements of the proposed model to be checked against the data of the Spanish Market.

PSI Policies

Public Sector Information policies represent those regulations and action plans from the public administrations stimulating the publication of information and how to do it. As described in point 1.1.2 there are strong differences between market potential estimations and actual data.

It is supposed, and this research tries to provide some modelling, that these policies will report benefits for the society, and that some of these benefits will be economical. To what extent should be determined in order to set the right amount of support to those policies.

Technical standard

Technical standard represents the technical mechanisms implemented by the public bodies in order to release the information. They could range from closed standards (so the re-user has to purchase some licensed tool in order to access the information) to open standards with additional meta data enrichment.

Access

Access represents the procedures needed for the re-users in order to retrieve the information. It could range from manual request (even not online) to an online request that could provide a single data into the dataset connected with other data in different sources.

Legal

Legal represents the license that the public body provides to the released data. Could range from copyright (and therefore, no rights to reuse) to only attribution. Other 'minor' differences are the request of not to distortion the information, and include updating dates in the reuse.

Interoperability

Interoperability will represent the extension and diffusion of how the structured information is released. It could range from a global standard with world-wide adoption to a, do it yourself, and no agreement on the structure of the information.

This factor has not been considered in the practical part of present study but it is noted here because some hints within the analysed data seems to assign it an important role.

MELODA

Three out of the four described factors compose MELODA metric. And in a close future interoperability will be included as well.

Reusability

It is the composed characteristic of released information which makes them simple, inexpensive, legal and connectible to other sources. It is composed of these 4 factors although in this analysis, only three of them have been included.

Scope

Scope represents the geographic range of affected information. It could range from local to international.

Type

Type represents the different types of information. In this work MEPSIR classification will be used in order to make possible the analysis with the provided data (MEPSIR, 2006). More detail is offered in point 7.1.1.

Economic impact

It is the real dimension that the society should take into account in order to determine the level of investment on PSI policies. Available analysis are quite partial as long as they only includes direct effects and possibly without a full coverage of the affected companies. A more general analysis should come from models like depicted in Illustration 7 in chapter 2.2.

2.3 PSI reuse value chain

Next illustration shows the main agents and tools being part of the value chain of PSI reuse.

Publishers

The first agents in the value chain are the sources of data (*Fuentes* in the illustration).

These can be public, or private, dividing the latter into for-profit organizations (companies) or without (foundations and NGOs).

The first challenge for these producers is how to get funding from their publishing activities.

Pollock identifies three main ones, but not

mutually exclusive, sources for the public bodies responsible for publishing PSI (Pollock, 2008). However, only two of them are really generalized. First one is the Government funding: fund from general government (or local ones) revenues and second one is user funding: charge those who use/ download/ access the dataset. A mix of both approaches is also generalized. A specific analysis of charging policies for trading funds can be found in Newbery et al article (Newbery et al, 2008).

As a result, different charging policies can be applied, from those, in which entities try to maximize their profits (in fact trying to be sustainable and not to depend on public budgets), to those who only charge as much as they have to reach a break even situation, and those who only charge a marginal cost (most cases zero for a digital access to the information). These last ones basically correspond with the entities that count on with a public budget for publishing the information and, at the same time, are the responsible entities which are forced to publish their information as a result of a legal obligation.

Different charging policies impact on the quality of data released and in the adoption of innovations in the publishing entities.

Legal and technical frameworks

The second agent in this value chain is legal and technical frameworks. They enable information reuse and include the regulations that promote publication, the public charging



Illustration 9, PSI reuse value chain (Abella, 2011) for Rooter

policies, the legal licensing of PSI, the mechanisms to access to information, privacy restrictions, etc.

There is also a whole infrastructure of technological regulations, data normalization, meta-data enrichment, etc. Public administration plays a key role

here, because it is not only the monopoly of publishing information but it is also the regulator of its potential uses.

Economic Value of Public Sector Information in the European Union, 1999 (€ Billion)

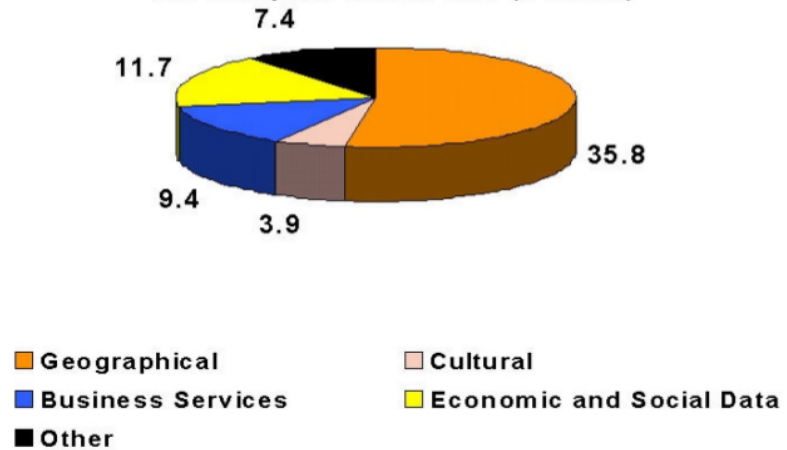


Illustration 10, Economic value of PSI types of information (PIRA, 2000)

	number	Average domestic market		EU25 + Norway	
		base	upper limit	base	upper limit
Average annual turnover from PSI per re-user		1.92	6.73	1.92	6.73
[A] Minimum number of re-users	238	456	1,602	11,881	41,645
Total domestic income from PSI (public content holders)	28	-62	-62	-62	-62
Net market size		395	1,540	10,279	40,044
[B] Base number of re-users	266	511	1,790	13,279	46,545
Total domestic income from PSI (public content holders)	28	-62	-62	-62	-62
Net market size		449	1,729	11,677	44,943

Illustration 11, Economic value of PSI reuse (MEPSIR, 2006)

It is worthwhile to note that the geo-referenced information (which possesses the most worldwide adopted standards) is also the type of information that provides the biggest economic impact according to PIRA (PIRA, 2000). Causality analysis for this relation is identified as a future research line. Next illustration shows the Economic Value of Public Sector Information from this reference and from (MEPSIR, 2006).

Infomediaries

The infomediaries are the creators of the products and services based on the published sources. Just as data sources, these entities can be public, private or third sector and, therefore, their objectives can range from mere profit interest, to the development of society itself, or the promotion of transparency or participation, among others.

Apps, services and derived information

Based on these data, infomediaries can generate either new data sets through treatment and combination, mobile applications for tablets or other devices, or purely services, either available online or as a consulting service.

Users

The users are part of the last link in this chain. These users can be citizens (adopting free or paid advertising business models) or from the professional area, who use the products as a part of their own business.

Standardisation bodies

Although it is not properly included in the value chain of the public information reuse (and therefore excluded from the graph), standardisation bodies can play a key role in the economic impact of this sector. Note that a quick normalisation could make easier for infomediary companies to develop global new products and services. Equally, the adoption of open standards could reduce interoperability costs.

Information reuse, both public and private sector has been practiced for decades, but it has been the rapid popularization of electronic access to information what is allowing the industry to move from a niche to a sector whose economic impact should be treated on its own.

The described value chain shows all agents and tools necessary to develop policies in terms of reusability of data. However, there are a number of additional factors, among which standardization standards, software tools, which could be considered as well as with relevance are included.

2.4 Theoretical Contribution 2: MELODA metric of information reusability

In order to relate reusability with economic impact, it is necessary to have a quantitative

metric for the concept of reusability and a practical methodology to carry on an assessment. Although some other metric with extensive support are available (5 stars of Tim Berners-Lee⁸), this metric lacks of a critical factor, which is the legal licensing of the released data. This factor is absolutely critical for professional reuse. Other metrics not specific of information reuse have also been analysed and some of their principles have been also assumed (Pipino et al, 2002) but with 16 dimensions (some of them subjective) it means these metrics are not really feasible to be implemented. Ren and Glissman (2012) proposed metric, although supposedly oriented to open data, considers dimensions which are not critical for professional re-user but for internal managers. (i.e. This metric consider Accessibility / Availability, understandability, completeness / correctness, timeliness, free of error and security dimensions) (Ren & Glissmann, 2012). The legal aspect is again deprecated. As a consequence of the inability to assess properly the concept of reusability, MELODA metric has been evolved in this work. It is based on previous works of the author⁹ (Abella, 2011). MELODA is a metric that implements 6 out of the 8 principles of open government data (Lessing et al, 2007) but in a way in which datasets can be assessed individually and a singular mark can be assigned to everyone.

MELODA comprehends, in current version 2.5, just three dimensions of information, Legal, Access and Technological standards, and they are assessed against a standard with 5 structured levels each. Every level is weighted and overall impact is aggregated. The next three points describe those dimensions and the different levels set of each one.

Legal reusability

Legal reusability is the characteristic of a dataset to allow reuse in different ways from private to fully commercial uses. In views of this study five levels have been standardised. Importance of legal licensing has been analysed (Jasserand & Bernt, 2012). They remark the importance of copyright licensing and provide recommendations on setting clear licensing models of the information released.

Next table shows the five levels included in this dimension. Other legal considerations (i.e.

8 <http://5stardata.info/>

9 <http://amedioentender.blogspot.com.es/2010/11/metodo-ltca-de-valoracion-del-open-data.html>

restriction of malicious use of the released information, attribution, misuse, requirements to include the last update's date, etc.) are not considered, because most of cases they are required as a group and does not limit potential fair uses of the information (Limits the illegal ones). An author cannot waive the attribution right accordingly to the Spanish legislation.

Additionally levels are weighted in order to provide a fair comparison.

Description	% weight
Level 1. Copyright. Data sources in this level reserve the copyright of data, thus restricting unauthorized use	0
Level 2. Private use Data sources in this level will allow the use of the data without processes of approval but only for private uses. Not public use is allowed.	10
Level 3. Non-commercial reuse Data sources in this level will allow data reuse but they do not allow commercial uses of the information re-use. (i.e. For a blog, for a non-profit organisation)	25
Level 4. Commercial reuse Data sources in this level will allow re-use of data, including the commercial reuse. (i.e. For any company in order to create new visualization of data)	90
Level 5. Only recognition Data sources in this level will only ask re-users the attribution of the data. i.e. Releasing data with CC BY license.	100

Table 1, Legal dimension of MELODA

Access reusability

Access reusability is the characteristic of a dataset to allow easy and automatic access to the contents of a data source. In this study five levels have been standardised. Every level is weighted in order to provide a fair comparison. Next table shows description of the different levels. In any case the free (no cost) access is a must for being considered in this dimension. Otherwise punctuation would be 0. This is supported by Cook (Cook, 2010).

Description	% weight
Level 1. No web access or manual request Access to information requires a non-automatic approval process to access to the datasets or to fill manually a form.	0
Level 2. Web Access URL with registration or with web interaction Access to information via the web, but it requires user interaction to select the data source.	10
Level 3. Web access or unique URL parameters Access to information via the web, and it is possible to access to the dataset individually, or through a unique URL, or by specific parameters in the query call.	25
Level 4. Web Access unique with parameters Access to information via the web, and it is possible to access individually to each data inside the datasets, or through a shortened URL, or by specific parameters in the query call and includes the date, the version, or the last update.	75
Level 5. API or query language Access to information provides access to specific data of the dataset, either by calling a documented API or through a query language of data sources (I.e. SPARQL).	100

Table 2, Access dimension of MELODA

Technical standards reusability

Technical standards reusability is the characteristic of a dataset to be stored in a standard which allows maximum reuse with minimum cost. This happens when the standard is open and there are free tools that implement it in order to retrieve the information. Qualified distinction is made based on the available documentation explaining how to determine the contents or the relevant ontology for semantic data. Every level is weighted in order to provide a fair comparison.

Description	% weight
<p>Level 1. Closed standard</p> <p>Data sources in this level are released on proprietary standards, or in open formats but not suitable for reutilisation (i.e. image formats for text). The definition of open standard can be found in national level legislation.</p> <p>I.e. xls, pdf image, doc, shp, etc.</p>	10
<p>Level 2. Open standard</p> <p>Data sources in this level are published on open standards but as individual files.</p> <p>i.e. csv, txt, odb, odt, ods, etc.</p>	25
<p>Level 3. Open standard with explanatory documentation</p> <p>Data sources in this level include those which release information as open standards in individual files but showing available explanatory information about the contents of the files.</p> <p>i.e. csv, odb, odt, ods, etc. with additional information about the structure of the fields inside the file, size, data type, range of registers of the file.</p>	50
<p>Level 4. Open standard, individual access/metadata</p> <p>Data sources in this level include those which includes metadata attached to any data. Additional information about the content of the data source is not available.</p> <p>i.e. rdf, json, xml (with meta-tags) without additional instructions</p>	75
<p>Level 5. Open standard with associated metadata and explanatory material</p> <p>Data sources in this level include those which release information attached with the metadata, and additional information about the content of the data source is available. i.e. ontology.</p> <p>I.e. rdf with additional documentation</p>	100

Table 3, Technical Standard dimension of MELODA

Assessment process

Assessment could be easily performed by filling a simple form (Error: Reference source not found. Error: Reference source not found). As a result MELODA provides a final figure for any dataset released composed according with this formula.

$$Meloda = \sqrt{100 \cdot legal \cdot access \cdot technical \ standard}$$

where each of the factors weights from 0 to 100%, so that MELODA ranges from 0 to 10. Further explanation about MELODA assessing process is included into annex 7.1.3.

3 Methodology

3.1 Methodology for data collection of infomediary companies

Methodology for data collection of companies, for this work corresponds to the raw data provided by the companies anonymously for two studies (Red.es, 2011) page 8 section 2.2, and (Red.es, 2012) page 6 section 1.3.1. It includes a sampling on 'every' infomediary company in Spain. These studies use classification of PSI information based on former studies (MEPSIR, 2006).

Answers comprehend information about 70 companies for year 2011 (Red.es, 2011) answering a 142 questions form and 72 companies in 2012 answering a 171 questions form (Red.es, 2012). This is the only source of this kind in Spain, and possibly the only one for a country providing information for two different years.

With these data it is possible to describe briefly the sector; however data do not provide enough insights on the relation between economic value creation and data consumption. Therefore the analysis will be carried out with the information consumption instead of the turnover of the companies.

3.2 Methodology for sampling public data sources

In order to carry on a contrast between the economic value created (based on information consumption) and reusability of data released, an extensive sampling of dataset sources has been performed. Unfortunately there is no sampled data about the creation of value classified by type or scope. Only information consumption is provided, and this indicator will be used as a hint of the economic impact.

First of all four categories in data sources accordingly to the scope of information published have been defined . Local, Regional, National and International. International has been rejected as long as there are not enough international data sources to contrast with. Additionally classifications of sources by type of information released share the criteria

former reports (MEPSIR, 2006; Red.es, 2011; Red.es, 2012).

Sampling of sources follows this grid of data.

Scope / type information	Business	Geograph	Legal	Meteo	Social	Transport	Other	TOTAL
National	10	7	10	5	6	6	6	50
Regional	14	14	12	3	8	13	14	78
Local	11	10	13	8	11	10	13	76
TOTAL	35	31	35	16	25	29	33	204

Table 4, Sampling of Spanish data sources

Finding the sources has been a challenging task because the traditionally most completed catalogue of public data sources (CTIC map¹⁰), seems to be outdated and without maintenance. Therefore some of registers have disappeared (i.e. open data Cordoba, Extremadura reutiliza) and some new ones were not available (i.e. Open Data Santander, Alcobendas INE, DGT).

Additionally, once the datasets have been determined for every source, in some cases dataset quality were so low that they were rejected as a valid sample and then changed to other equivalent sources. Detailed data about what entities have been sampled is included into the Annex in section 7. Annex II. Data sources.

All the data for these data sources can be freely downloaded for non-commercial uses¹¹.

3.2.1 Standardization

In order to be able to create a consistent analysis, the need of standardization has been detected. This standardization will include five dimensions: scope of the information, content type, legal license, ability of access and technical standards. The three last ones are described in 3.3. Current situation of PSI reuse in Spain .

Scope of the PSI information

Most of the PSI can be classified accordingly of the scope of the information. In this work

¹⁰ <http://datos.fundacionctic.org/sandbox/catalog/faceted/>

¹¹ http://www.meloda.org/form3/export_csv.php

will use this classification

Scope	Description
Local	Based of or published by a City or any of its public depending entities.
Regional	Based of or published by a regional government or any of its public depending entities. In the case of Spain this category will be exactly the same of Comunidad Autónoma
National	Base of or published by a regional government or any of its public depending entities.
International	Any information extending the scope of before mentioned of coming from other international sources (i.e. EU, world bank, IMF, etc.)

Table 5, Scope of PSI

The reason for this classification will be the need to find relations between data coming from the infomediaries and the sampled data from data sources (Red.es, 2011; Red.es, 2012).

Type of PSI

In order to make consistent comparisons with the available data is adopted this classification

Topic	Description
Business	It includes any Information on Business / Economy
Geographic	It includes any GIS / Cartographic
Legal	It includes any information about legislation or other
Meteo	Weather Information or other related with meteo phenomena
Social	It includes any information about Socio-demography or Statistics
Transport	It includes Information about Transport and traffic
Other	Any other sort of informations out of the previous categories

Table 6, Type of PSI

This classification matches (MEPSIR, 2006) where there is a bigger detail in every aspect, and it is the same than in Red.es studies (Red.es, 2011; Red.es, 2012).

3.3 Current situation of PSI reuse in Spain

Spain could be described as a strong and early supporter of PSI publishing policies. In fact in the period from 2010 to 2013 there has been a huge change in the licensing conditions of the released information as shown in next graph.

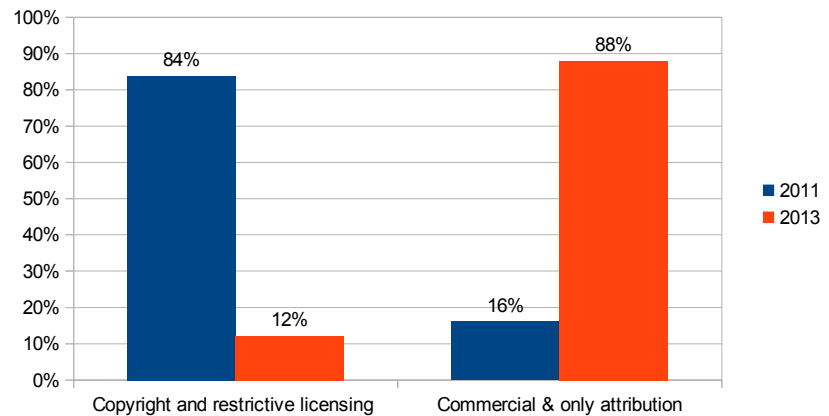


Illustration 12, Licensing conditions change in Spanish PSI sources Nov-11 to Sep-13

Besides this, a constant increase in the amount of available sources Illustration 13.

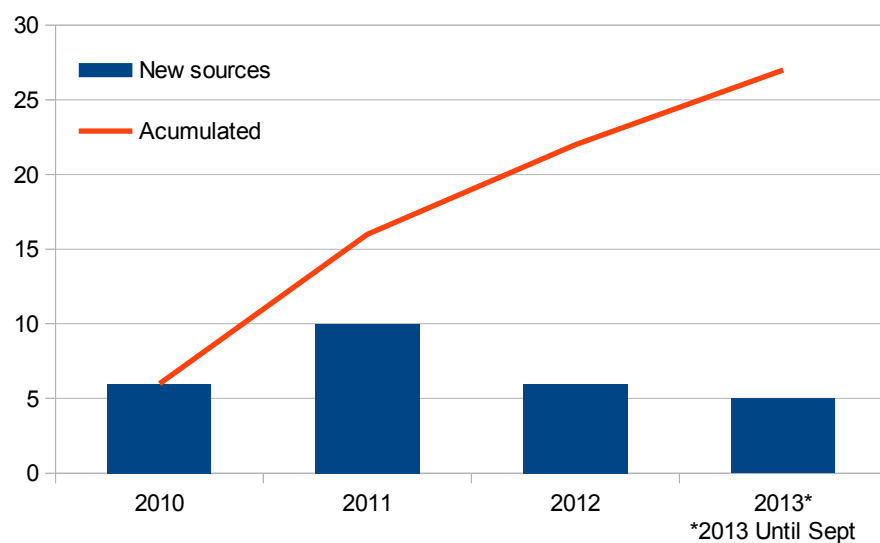


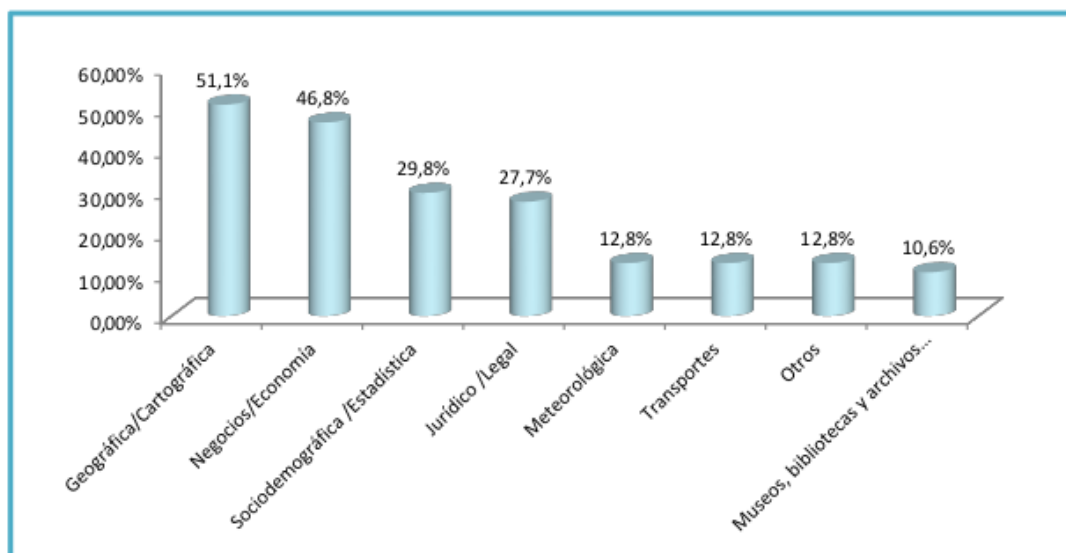
Illustration 13, New public opendata portals in Spain. Based on CTIC data.

and some of their results are translated here. For example (Red.es, 2011).

And (Red.es, 2012)

Estimation of activity depending on type of information. Spain 2011
Business 37,6%
Geographic/Cartography 30,5%
Legal 17,0%
Transport 5,2%
Socio-demographic/Statistics 1,9%
Meteo1,1%
Other: TIC 6,7%

Text 1, Infomediary business activity Spain by type of information (Red.es, 2011)



Fuente: Encuesta a empresas reutilizadoras ACAP-RED.ES 2012. Pregunta: ¿Sobre cuál o cuáles de los siguientes tipos de información desarrolla la actividad su empresa? Porcentajes válidos

Illustration 14: Type of information used by infomediaries in Spain (Red.es, 2012)

The use of PSI sources has become popular due to the impulse from civil society through app competition like the pioneer abredatos (2010)¹² and other social platforms. Currently PSI coming from Smart cities¹³ is attracting most of the attention¹⁴.

¹² Web site no longer available

¹³ <http://www.proacceso.org/>

¹⁴ <http://bit.ly/19uU6Kq>

The current quantitative assessments (Red.es, 2012) show a picture of an industry of about **400 million euros** for the Spanish economy (0.04% of GDP) ¹⁵ and restrict their study to the value created by infomediary companies. (Branch 1-6 in Illustration 8)

Transparency and freedom of information access

Transparency (or more precisely accountability of public institutions) and laws enforcing the freedom of access to information (FOI) are powerful tools which force public administrations to release public sector information. Here there is a table of FOI and PSI reuse passed laws in Spain.

Public administration	Situation	Passing date ¹⁶
Passed laws		
Spain	Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público.	17-11-2007
Spain	Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal.	8-11-2011
Navarra	Ley foral 11/2012, de 21 de junio, de la transparencia y del gobierno abierto ¹⁷	21-06-2012
Spain	Resolución de 19 de febrero de 2013, de la Secretaría de Estado de Administraciones Públicas, por la que se aprueba la Norma Técnica de Interoperabilidad de Reutilización de recursos de la información	4-3-2013
Extremadura	Ley 4/2013, de 21 de Mayo, de Gobierno Abierto	21-5-2013
In process		
Andalucía	Ley de Transparencia Pública de Andalucía ¹⁸ . Draft.	4-6-2013
Spain	Proyecto de Ley de Transparencia, acceso a la información pública y buen gobierno (To be approved in Senate) ¹⁹	18-9-2013

¹⁵ Red.es, 2011 and Red.es, 2012

¹⁶ Some of them are in the process to be passed. Date corresponds to the last step done in the passing process.

¹⁷ http://www.navarra.es/home_es/Actualidad/BON/Boletines/2012/125/Anuncio-0/

¹⁸ http://www.juntadeandalucia.es/export/drupaljda/ANTEPROYECTO_DE_LEY_DE_TRANSPARENCIA_PUBLICA_DE_ANDALUCIA.pdf

¹⁹ http://www.congreso.es/public_oficiales/L10/CONG/BOCG/A/BOCG-10-A-19-7.PDF

Public administration	Situation	Passing date
Castilla y León	Ley de Transparencia y Gobierno Abierto (Drafting after public consultation)	2-10-13

Table 7, FOI and PSI reuse legislation in Spain and regional governments

The various rights granted by these laws are reflected in the proactive publication of information, in many cases enabling reuse.

3.4 Analysis and Time to Implement Reuse (T_{TIR})

The performed analysis looks for relation between the reusability of data sources, based on their legal licensing, technical standard and access mechanisms with the data consumption by infomediaries and having as parameters the scope and the type of information. There is only aggregated information of the economic turnover of these companies classified by type of information used. It has to be noted that available data for companies were collected on June 2012, when data for data sources has been collected in September of 2013. (15 months delay).

It is clear that a time (*Time to Implement Reuse*) should be taken since the availability of information to the first dataset reuse.

$$R(t) = P(t - T_{TIR})$$

Where R is the information reuse in time t and P is the publishing of information.

$$T_{TIR} = T_{TIR}(\text{type}, \text{reusability}, \text{interoperability}, \text{promotion})$$

Where *type* refers to the type of information, *reusability* summaries the dimensions of MELODA, *interoperability* refers to the meaningful and global adoption of the technical standard adopted and *promotion* refers to the campaigns from different agents in order to promote reuse.

4 Results

4.1 Data consumption by infomediary companies

This graph shows consumption of information by infomediary companies classified by type for years 2011 and 2012 and next graph details the increase in the reuse. Several answers were possible for every surveyed company, results could exceed 100%.

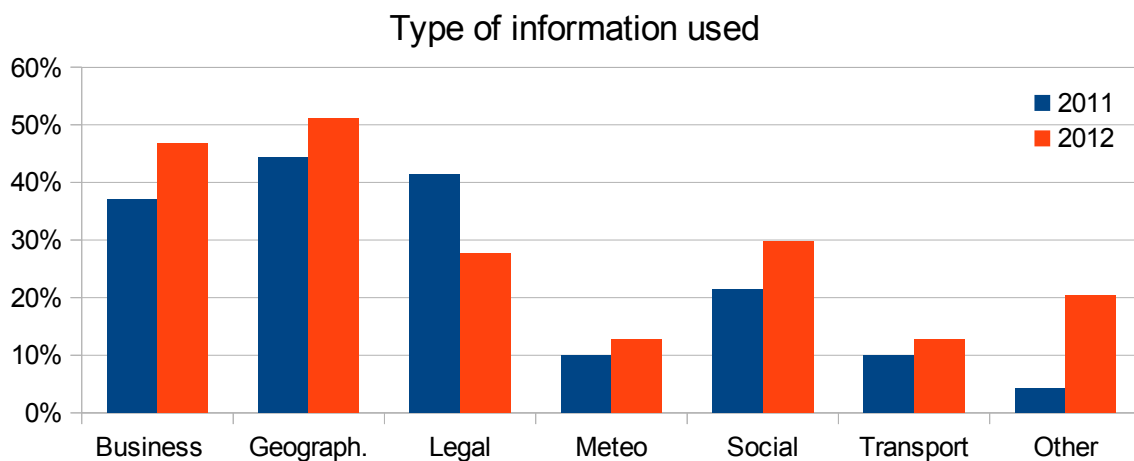


Illustration 15, Type of information used (Red.es, 2011) and (Red.es, 2012)

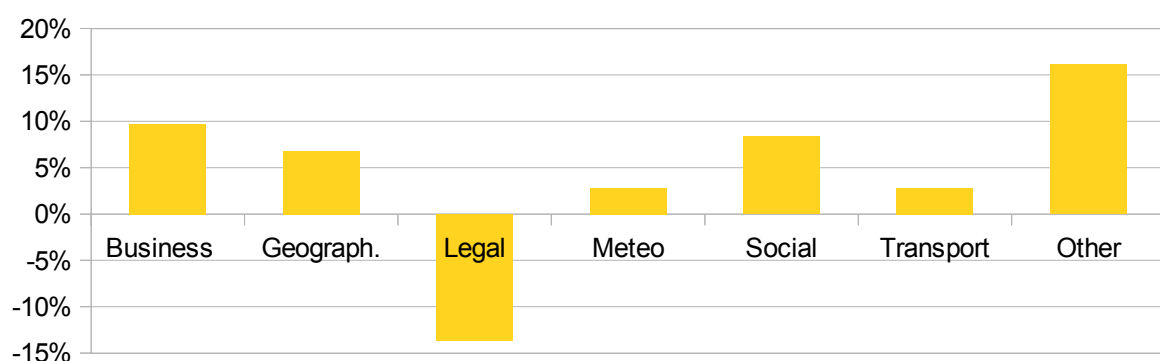


Illustration 16, Increment of infomediary information consumption by type 12-11

The information of this last graph will be compared with the analysis carried out in 4.3. Impact of reusability conditions on data consumption in order to determine if there is a

relation between the increase in the use of information and the reusability of data sources. Y-axis reflects the increase of consumption as a percentage, taking into account that in absolute levels it is likely a heavy decrease (turnover decline of these companies reach around 24% in 2012). Unfortunately these data reflect data consumption and not the turnover related with the type of information because such data are not available. Therefore in next studies it will be needed to determine which the relation between data consumption and turnover is in order to determine the right economic impact.

4.2 Reusability of public data sources

Next graph represents the reusability in terms of MELODA metric for every scope of information source and different types of information.

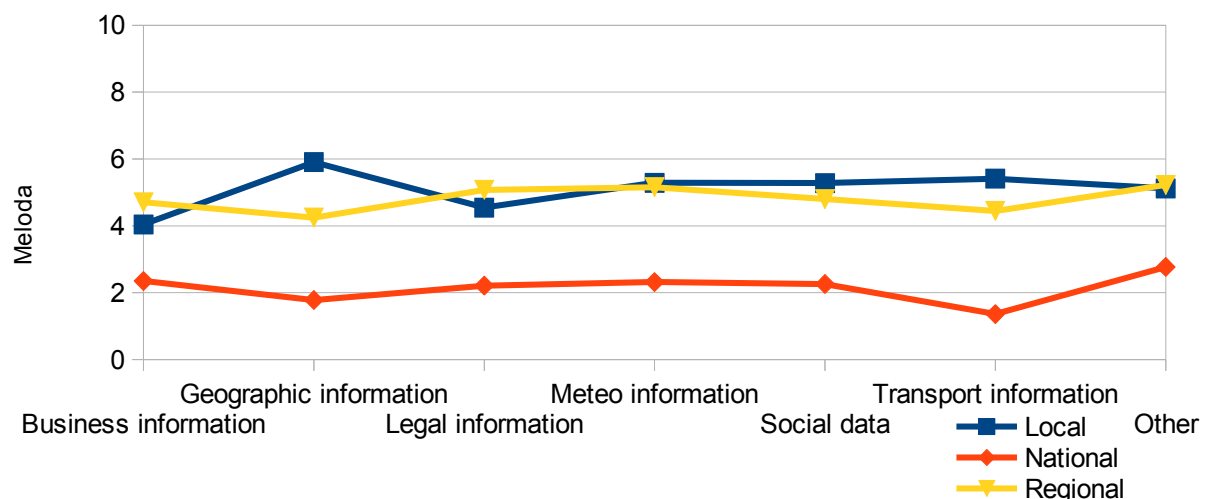


Illustration 17, Global reusability of datasources in Spain. Own elaboration.

It shows a remarkable difference between those sources coming from national scope and those coming from local and regional. It is related also, that most of the data sources coming from national sources are older (with less attention paid to reusability).

This graph only describes current situation of PSI data sources in Spain and it is the first in its kind currently (Nov 2013). It is created by classifying the data sources by scope and type of information and averaging the MELODA index. Y-axis reflects this average.

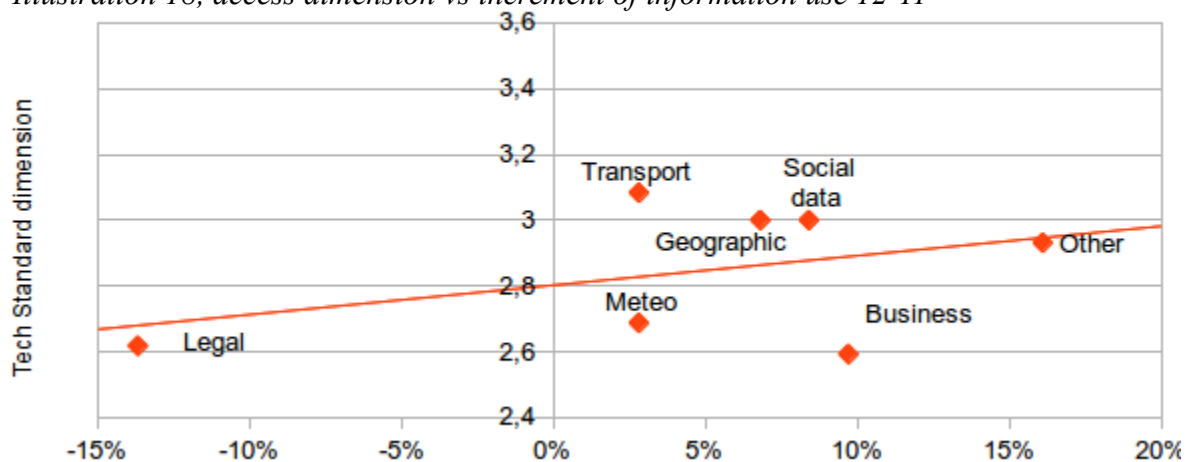
4.3 Impact of reusability conditions on data consumption

In order to make a fair comparison between the sampled data for infomediaries in years 2011 and 2012 (Red.es, 2011; Red.es, 2012) and the sampled data sources, it is necessary to filter those datasets which were not available for commercial reuse (Levels 1, 2 and 3 described in Legal reusability). This filtering is required due to the fact that products and services of infomediaries use the data which are allowed to be used commercially. . Fortunately, it is not a very restrictive condition because as we have presented in Illustration 12, 88% of sampled data sources meet this requirement. Therefore now a full comparison with MELODA reusability index makes no sense as one out of three of the dimensions (Legal) has been filtered. With this restriction in place results are shown in the next graphs for the other dimensions access and technical standards.

Access dimension

This graph represents the average of the MELODA index in Y-axis, vs the increase of use of information as described in Illustration 16 in the X-axis, trying to find a relation between the better reusability of information (access dimension) and the increase of its use.

Illustration 18, access dimension vs increment of information use 12-11



Next graph shows the same results of this, but with a split of the data sources according to their scope. Therefore 3 sets of points are determined for every scope. Additionally a trend line is determined and plotted in order to make easier to determine if there is a relation between reusability (Access dimension) and data consumption.

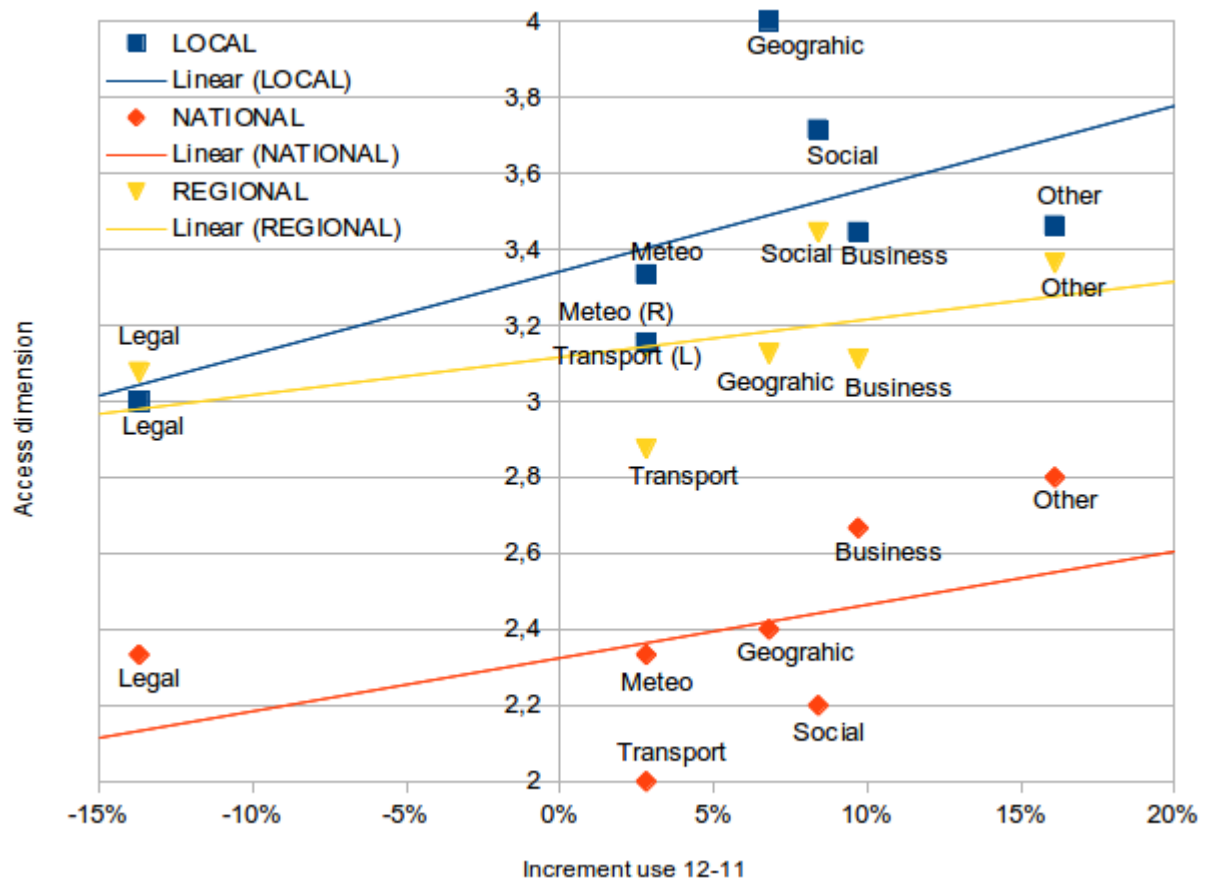


Illustration 19, access dimension vs increment of information use 12-11 by scope

Next illustration (Illustration 20) is similar to Illustration 18 but instead of showing access dimension reflects technical standard dimension in Y-axis. X-axis, is the increase of use of information by infomediaries as that illustration.

Similarly Illustration 21 is the replica of Illustration 19 changing access dimension by technical standard dimension.

Technical Standard

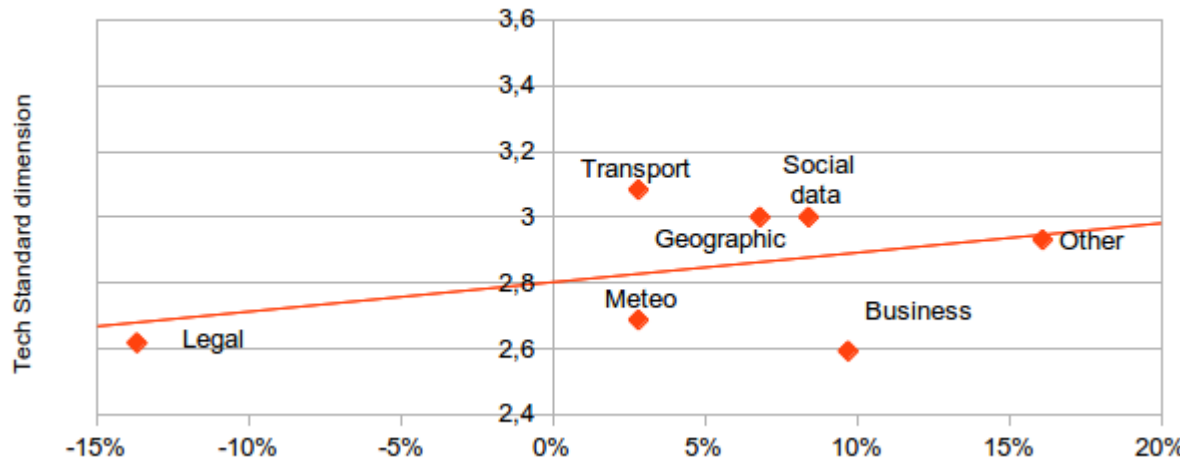


Illustration 20, technical standard vs Increase of Information use 12-11

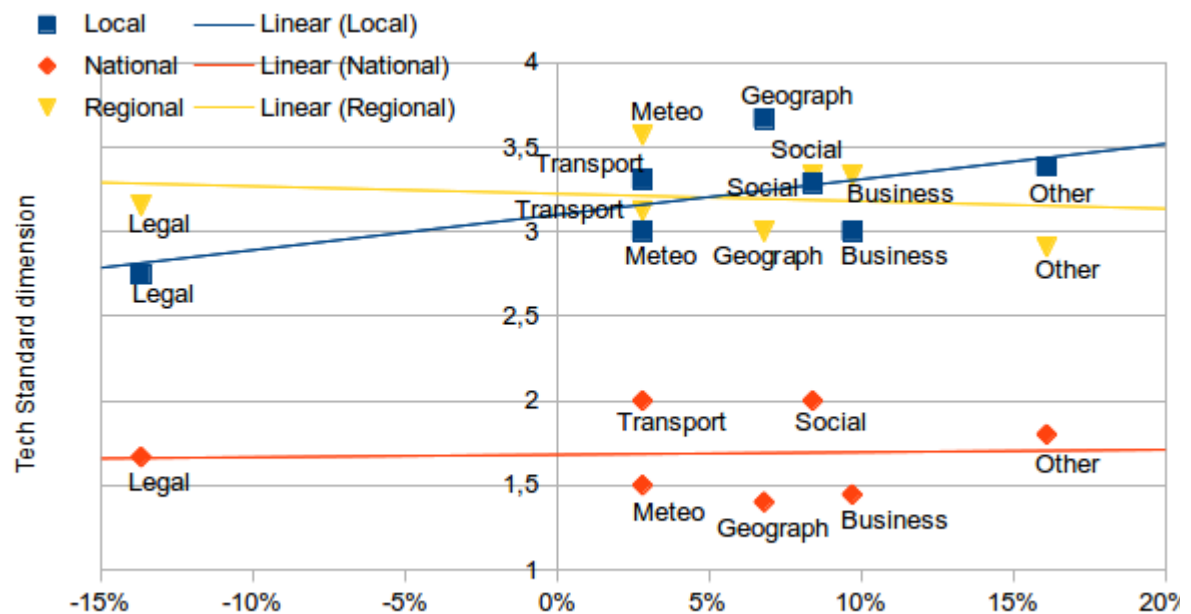


Illustration 21, technical standard vs Increment of Information use 12-11 by scope

Although both dimensions seem to have some kind of positive relation, a deeper analysis of data shows that while access dimension presents a robust positive relation, technical standard does not.

Not only because when we split into the different scopes, technical standard presents, in the case of regional data sources, a negative slope but also when we eliminate any of the points (i.e. the legal information point) relation remains in the case of access and not in the technical standard. This point will be further analysed in next chapter.

5 Conclusions and future goals

Access dimension shows a positive relation with data consumption

As shown in results for the restricted data Illustration 18 and Illustration 19, there is a positive relation between the data consumption and reusability dimension of access. Causality between both magnitudes should be proved based on an extended analysis. The next two factors could explain the low level of relation:

1. It takes a long time (T_{TIR}) to create a useful application, service or product based on new information available. See 3.4. Analysis and Time to Implement Reuse (TTIR). It could be counted in months.
2. As we can see in Illustration 13, there is constant increase of PSI (open data) portals, meaning that a big percentage of the sampled sources in 3.2. Methodology for sampling public data sources are newer than the sampling of companies data were performed. (If data sources available were growing at the same rate than portals—which is a conservative hypothesis-- 30% of data sources did not exist on June 2012, when data for infomediaries were sampled). Additionally newer sources tend to have a higher reusability MELODA index. So, their impact could exceed 50% of the overall result. This is commented in section 4.2, Illustration 17, on the difference between MELODA reusability indexes by scope and it would confirm this explanation for the weak relationship.

Technical standard does not show a positive relation with data consumption

Contrary to the expectations, our own experience and common sense, results do not show a positive relation between the reusability for the technical standard dimension. Although it seems that in Illustration 20 could show such relation, deeper analysis of data refutes that statement. We find three possible explanations,

1. The strong industrial crisis (which has reduced the sector in 24% from year 2011 to 2012) could hide results as this percentage exceeds potential relation effects.

2. As introduced in the previous section, and described in 3.4 Analysis and Time to Implement Reuse (TTIR), there is a time gap between sampling of companies results (May-June of 2012) and sampling of PSI sources (Sep-Oct 2013). It could reflect a time lapse between the availability of sources and its economic effects.
3. Technical standard integration is a one-shot investment. Once connected the data source for the infomediary company, contrary to the access dimension, there is no further costs for this dimension (unless that there were a change in the standard from the data source, which is uncommon). So this dimension, could be (like legal dimension), a pass non-pass dimension. If the forecasted cost is lower than potential profit, further increase in accessibility does not represent more reuse and therefore it would not show relation. Then this dimension could be only an economic threshold to start with its use.

Global conclusions

The economic modelling of PSI reuse is in its early ages. Not only is there a lack of theoretical models, but also there is a lack of data in order to oppose the theoretical model to the actual data. This research helps to create the theoretical framework to analyses systematically the problem of the economic impact of PSI reuse. Global description of the model, metric to assess the reusability of data sources and Time to Information Reuse would help on that. On the other hand the remarkable economic importance of the PSI reuse spurs further practical investigations on this field. Preliminary results in this work show that some factors affect the economic impact. Besides the legal licensing necessary for commercial reuse, access mechanisms appear as the first candidate dimension to drive this economic impact. It does not mean that it will be even the most relevant, and causality should be proved with further investigations. New data coming from infomediary sector would confirm to what extent such impact exists. On the other hand characterisation of the '*Time to Implement Reuse*' as described in 3.4. Analysis and Time to Implement Reuse (TTIR) could transform our practical vision of the economic impact evolution.

Short term goals

1. To review and complete MELODA metric in order to include interoperability factor.

Then review this analysis in order to assess its impact. Hopefully it should be carried out with real data, and not aggregated. It could help to determine the ratio of influence of every factor.

2. To study how time impacts on the start of information reuse and other specific support actions (i.e. promotion plans) influence the development of the economic impact.
3. If available, contrast the sampled data with the infomediary use of information for 2013, in order to check the hypothesis of the '*Time to Implement Reuse*'. Once validated it would help to weight the different factors of MELODA metric with a more accurate estimation.
4. To check the relation between reusability of data sources and real economic turnover of companies and not only with their data consumption. Again availability of data could be a challenging restriction.

Middle-term goals

1. If modelling could be successfully carried out and results proved to be useful, it would be possible to answer questions as : Who should afford the costs of PSI publishing and what regulatory structure should be adopted to support it?
2. To check economic reuse with the App economy (economy coming from the mobile use).
3. Extending the analysis by including data coming from the rest of branches of the model, specifically the performance increase of the public administrations accessing to updated and fully available data.
4. It is expected that in short future information coming from the private sources of regulated sectors (banking, energy, telecommunications, etc) could be considered public so it should be included as additional data sources.

6 Annex I. Work Cited List

Acknowledgements

Red.es and Borja Adsuarra and Pedro Martín by providing raw data about its studies for the intermediary sector in Spain for years 2011 and 2012.

Carmen de Pablos for guiding me on this job

Eva-Maria Mora and Marta Ortiz-de-Urbina for their invaluable support and guidance in the evaluation of data and statistical tools.

6.1.1 Work Cited List

Abella. A. (2011) “Reutilización de información pública y privada en España ”. *Rooter*. May.

http://rooter.es/documents/PAPER_REUTILIZACION_INFORMACION_PUBLICA_PRIVADA_OPENDATA.pdf

Alani, H.; Dupplaw D.; Sheridan D.; O'Hara K.; Shadbold, N.; Tullo, C. (2007). “Unlocking the potential of public sector information with Semantic Web technology”. *Semantic web proceedings, lectures on computer science* vol. 4825:708-721.

Burdon, M. (2009). ”Commercializing public sector information privacy and security concerns”. *Technology and Society Magazine, IEEE* . Vol. 28, Issue: 1: 34-40.

Castells, M. (2000). ”Internet y la sociedad red”. Capítulo “Internet y la Nueva Economía.” *Conferencia de Presentación del Programa de Doctorado sobre la Sociedad de la Información y el Conocimiento*. Universitat Oberta de Catalunya. 2000.

Cherry, M.; McMenemy, D. (2013). “Freedom of information and 'vexatious' requests - The case of Scottish local government”. *Government Information Quarterly* 30. issue. 3: 257-266

Cook, J. S. (2010). “Economic issues in funding and supplying public sector information”, *Spatial Information Systems Limited, as trustee for Participants under the Collaborative Research Centre – Spatial Information (CRC -SI)*.

Dekkers, M.; Polman, F. ;Te Velde, R.;De Vries. M. (2006). “Final Report of Study on

Exploitation of public sector information – benchmarking of EU framework conditions”. *European Commission*.

Directorate General for Information Society and Media. (2009). ”Smarter, Faster, Better eGovernment 8th Benchmark Measurement|November 2009”. *European Commission*.

EU. Directive 2003/98/EC on the re-use of public sector information. *Official Journal of the European Union*. 17 November 2003. 345/90-96. See more at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>

EU. Proposal for a Directive of the European Parliament and of the Council amending Directive 2003/98/EC on re-use of public sector information. *Official Journal of the European Union*. 27 Jun 2013. L175/1-8. See more at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>

Nicol, A.; Caruso, J.; Archambault, E. (2013) “Open Data Access Policies and Strategies in the European Research Area and Beyond”. *European Commission DG Research & Innovation*.

Hardin, G. (1968) “The tragedy of the commons”. *Science*. 1243.

Huijboom, N.; Van der Broek, T. (2011). ”Open data: an international comparison of strategies”. *European Journal of e-practice*. Nº 12, March / April. ISSN: 1988-625X: 1-13

Janssen, K. (2011). “The role of public sector information in the European market for online content: a never-ending story or a new beginning?”. *Emerald Group Publishing Limited*, ISSN 1463-6697. Nº 6: 20-29

Jasserand, C.; Hugenholtz, P.B. (2012). “Using Copyright to Promote Access to Public Sector Information: A Comparative Survey“. *WIPO*. June. (contract for the Institute for Information Law University of Amsterdam The Netherlands).

Kalampokis, E, Tambouris, E., Tarabanis, K. (2011). “Open Government Data: A Stage Model”. *10th IFIP WG 8.5 International Conference on Electronic Government, EGOV 2011*. LNCS 6846: 235–246.

Kauffman, R. J.;Kriebel, Ch. H. (1988) “Modelling and Measuring the Business Value of

Information Technology” *Information Systems Working Papers Series*. March. No. IS-88-24. Available at SSRN: <http://ssrn.com/abstract=1289710>

Lessing, L. and 29 authors. (2007) “8 Principles of Open Government Data”. *Open Government Working Group*. <http://www.opengovdata.org/home/8principles>

Marcos-Martín, C. Soriano-Maldonado. S.L. (2011) “Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta”. *El profesional de la información*. May-Jun. v. 20, n. 3 :291-297

Newbery, D, Bently, L. and Pollock, R. (2008). “Models of Public Sector Information Provision via Trading Funds”. Department for Business, Enterprise and Regulatory Reform (BERR) and HM Treasury. 2008. <http://www.berr.gov.uk/files/file45136.pdf>

Nonaka, I. “A dynamic theory of Organizational knowledge creation”. (1994). *Organization science* Vol 5 Nº 1. Feb.

OECD. (2006). “OECD recommendation of the council for enhanced access and more effective use of public sector information”. *OECD Committee for Information, Computer and Communication Policy (ICCP Committee)*, (Working Party on the Information Economy.) <http://www.oecd.org/sti/44384673.pdf>

Pipino, L.L.; Lee, Y.W.; Wang, R. Y (2002). “Data quality assessment” *Communications of the ACM*, April. vol. 45, No. 4ve.

Pira International. (2000). “Commercial exploitation of Europe’s public sector information”. *European Commission, Directorate General for the Information Society*, Oct.

Pollock, R. “The economics of public sector information”.(2008). *University of Cambridge*. Working paper.

Red.es. (2011). “Characterization Study of the Infomediary Sector 2011”. *Ministry of Territorial Policy and Public Administration, the State Secretariat of Telecommunications and Information Society and of the National Observatory of Telecommunications and of the Information Society (ONTSI), of the Ministry of Industry, Tourism and Trade*, with the advice and collaboration of Everis Spain S.L.

http://datos.gob.es/datos/sites/default/files/files/aporta_infomediary%20sector_2011.pdf

Red.es. (2012). “Characterization Study of the Infomediary Sector 2012”. (2012) *Ministry of Finance and Public Administration of the State Secretary of Public Administration, of the Ministry of Industry, Energy and Tourism of the State Secretary of Telecommunications and the Information Society and the National Observatory of Telecommunications and Information Society. July 2012.*

http://www.ontsi.red.es/ontsi/sites/default/files/121001_red_007_final_report_2012_edition__vf_en_1.pdf

Ren, G-J.; Glissmann, S. “Identifying Information Assets for Open Data. The Role of Business Architecture and Information Quality”. *2012 IEEE 14th International Conference on Commerce and Enterprise Computing*. Pp 94-100

Spain. (2007). “Ley 37/2007 sobre reutilización de la información del sector público.”. *BOE*. 17 November 2007. <http://www.boe.es/boe/dias/2007/11/17/pdfs/A47160-47165.pdf>.

Shapiro, C and Varian, H.R. “Information Rules: an strategic guide to the network economy”. (1999). *Harvard Business School Press*.

United Kingdom. (2005). “The Re-use of Public Sector Information Regulations 2005”. *The Minister for the Cabinet Office*. 1 July 2005. <http://www.legislation.gov.uk/uksi/2005/1515/made>

Vickery, G. “Review of recent studies on psi re-use and related Market developments”. *Information economics* . (2013). *Directorate for science, technology and industry Committee for information, computer and communications policies*. Paris.

W3C. (2009). “Publishing Open Government Data”. *W3C Working draft*. <http://www.w3.org/TR/gov-data/>

7 Annex II. Data sources

Here there is the list of data sources analysed and the type of information sampled in every source.

Identification	Scope	Business	Geograph	Legal	Meteo	Social	Transport	Other
Cabildo de Tenerife	Local		1					
Badalona Open Data	Local	1	2					1
City of Málaga Open Data	Local	1	2	1		1		1
Gijón Public Data Catalogue	Local	1	1	1	1		1	2
Lleida Open Data	Local	1	1			1	1	2
Observatori Ciutat Open Data Sant Boi	Local	1	1			1		1
Open Data Terrassa	Local	1	1	2		1	1	1
OpenData Barcelona	Local	1	1	1	1	1	2	1
OpenData Sabadell	Local	1	1				2	1
Pamplona Open Data	Local	1	1	2			2	2
Zaragoza Public Data Catalogue	Local	2	1	2	1	2	2	1
Open data Santander	Local	2	1			1	2	
Open data Alcobendas	Local			3				1
Open data Sevilla	Local	1						
** Expression is faulty **		**	**	**	**	**	**	**
SUB TOTAL Local		Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **
Catastro	National		1					
Ministerio de Agricultura	National	1	1					1
Biblioteca Nacional	National			1				
Ministerio de Justicia	National			3				
Ministerio de Hacienda y Administraciones Publicas	National	2		5				3
Instituto Geográfico Nacional	National		4					

Identification	Scope	Business	Geograph	Legal	Meteo	Social	Transport	Other
AEMET	National		1		5			
Agencia Tributaria	National	3						
Comisión Nacional Mercado Valores	National	1						
Ministerio de Industria Energía y Turismo	National	1						
INE	National	2				3	3	1
Fundación de los Ferrocarriles Españoles	National						1	
Ministerio del Interior	National							1
Ministerio de Fomento	National						1	
Dirección General Tráfico	National						1	
FECYT	National					1		
CSIC	National					1		
Servicio Público de Empleo Estatal	National					1		
Boletín Oficial del Estado	National			1				
		**	**	**	**	**	**	**
	** Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **	Expression is faulty **
SUB TOTAL National		faulty **	faulty **	faulty **	faulty **	faulty **	faulty **	faulty **
Abert@s (Galicia Open Data)	Regional	1		1	1	2	1	2
Andalusia Open Data Catalog	Regional	1		1	1	1	1	1
Open data Aragon	Regional	1	1	1		1		1
Asturias Public Data	Regional			1		1		1
Balearic Islands Open Data	Regional		1	1				1
Catalonia Open Data Catalog	Regional	2	1	1	3		2	1
Dato abierto Rioja	Regional	2	2			2		1
Instituto Canario de Estadística (ISTAC)	Regional	2	1	1			2	1
Open Data Euskadi	Regional		2	1	1	2	2	1
Open Data Junta de Castilla y León	Regional	1	1	1	1	1	1	2
Open Data Navarra	Regional			3	1			1
Portal de Datos Abiertos de JCCM	Regional			1				
Instituto Estadística Comunidad Madrid	Regional	1	1			1	1	

7.1.1 Extended definition of information categories

Business

This information category includes data related with the economy and does not include data about the investments, expenses, etc of public entities with are included into legal category. Global economy indicators are included into this type of information.

Geographical

This information category includes most of the data with georeference. It does not include other information in which some co-ordinates are included.

Legal

This information category includes not only data about the legal activities of entities but also data related with investments, expenses, bidding, etc of public entities.

Meteo

This information category includes data related with the weather and environmental information related with the weather.

Social and statistical

This information category includes data related with society which surrounds entity i.e. demography, social behaviour, etc.

Transport

This information category includes data related with traffic and transport by any means, flight, road, railway, etc

Other

This information category includes data not include in the other categories, for example cultural information, agenda, etc

7.1.2 Form to register information for the analysis

[HOME](#)
[ASSESS DATASET](#)
[ABOUT MELODA](#)
[DEFINITION OF M4J](#)
[FULL DESCRIPTION OF MELODA](#)
[BEST DATASETS](#)
[LOCAL, SPAIN](#)

Assess dataset

Meloda 2.0
Quality data sources

Fill this form with the data of the dataset. It will be assessed and listed in the [best datasets](#) page

Name * ⓘ

Link * ⓘ

Source * ⓘ

Country * ⓘ

Type of information *

Public/Private * ⓘ

☐ Private entity
 ☐ Public entity

Scope of information * ⓘ

Legal dimension * ⓘ

Access dimension * ⓘ

Technical Standard * ⓘ

Comments ⓘ

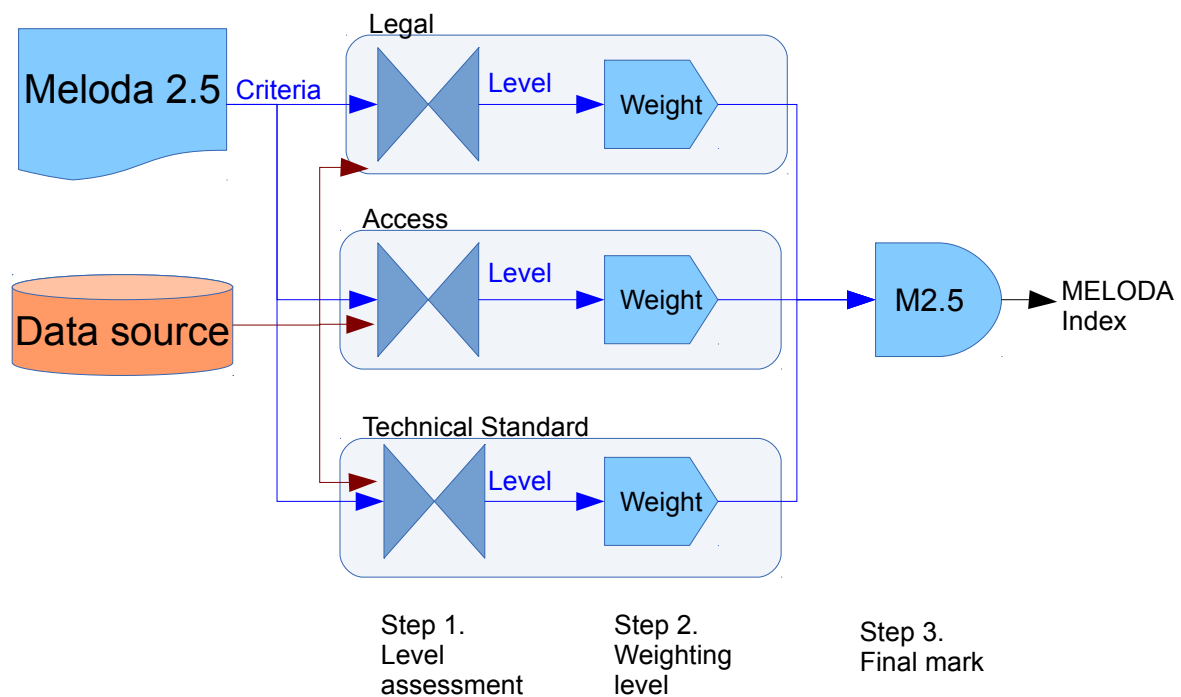
Assessor ⓘ

Generated by [phpFormGenerator](#)

7.1.3 Assessment process of MELODA metric

Here there are two examples of the assessment process with MELODA 2.5 metric.

MELODA 2.5 assessment process



Example:

Institution A releases a dataset D as linked data but without further explanations about the ontology used, licensed for commercial re-use, and being able to be queried as a SPARQL resource.

Step 1: Legal assessing: Level 4

Step 2: → 90% (Commercial use)

Step 1: Accessibility assessing: Level 5

Step 2: → 100% (Queryable through SPARQL)

Step 1: Technical assessing: Level 4

Step 2: → 75% (Metadata without details)

Step 3:

$$Meloda = \sqrt{100 * 90 * 75 * 100} = 8.22$$

We could say Dataset D is marked with a 8.22 / 10 in MELODA metric.