MELODA, A METRIC TO ASSESS OPEN DATA REUSE

Alberto Abella, Marta Ortiz-de-Urbina-Criado y Carmen De-Pablos-Heredero

Alberto Abella, a telecommunications engineer, is assistant professor at Universidad Rey Juan Carlos (URJC), in Madrid, and a PhD student at Escuela Internacional de Doctorado in the Social and legal sciences program at URJC. He is an expert in open data and open government within the public and private sector. He is president of the Spanish chapter of the Open Knowledge Foundation.
http://orcid.org/0000-0003-0243-4606
alberto.abella@meloda.org

Marta Ortiz-de-Urbina-Criado, PhD in Economics and business administration science, is an associate professor at the University Rey Juan Carlos, Madrid. She has published articles in impact journals including International business review, International journal of technology management, Journal of knowledge management, Revista española de financiación y contabilidad, etc.
http://orcid.org/0000-0001-7527-6798
marta.ortizdeurbina@urjc.es

Carmen De-Pablos-Heredero, PhD in Economics and business administration science, is an associate professor and head of the Master degree in business organization at the Rey Juan Carlos University. She has published articles in impact journals including CEDE, Universia, Interciencia, TIBE, Journal of entrepreneurship management, International journal of marketing research, Social sciences journal, World economy, Dyna, Pensée, Medical economics, REIS, etc.
http://orcid.org/0000-0003-0457-3730
carmen.depablos@urjc.es

Rey Juan Carlos University
Faculty of Social Sciences and Law
Paseo de los Artilleros, s/n. 28032 Madrid, Spain

Abstract
Current policies demanding an increase in transparency, open government, and the implementation of smart cities necessitate publishing open data. However, there are limited mechanisms for assessing the use and value of this information. The goal of this paper is to present a new metric, Meloda, that qualifies the information and helps to assess its reusability. Meloda’s four dimensions are described: technical standards, access, legal considerations, and data model. Finally, Meloda’s assessment process is explained.

Keywords
Open data; Information reuse; Metric; Open government; Meloda.

Titulo: Meloda, métrica para evaluar la reutilización de datos abiertos

Resumen
Las políticas actuales para el aumento de la transparencia, la implantación del gobierno abierto o de las ciudades inteligentes tienen en la publicación de información uno de sus pilares fundamentales. Sin embargo, la disponibilidad de mecanismos de evaluación del uso y valor de esta información es limitada. Se presenta una nueva métrica, Meloda, que permite calificar la información y evaluar su grado de reutilización. Se describen las cuatro dimensiones de la métrica: estándares técnicos, acceso, legal, y modelo de datos, y se explica su proceso de evaluación.

Palabras clave
Datos abiertos; Reutilización de la información; Métrica; Gobierno abierto; Meloda.
1. Introduction

The Open Government Partnership was born in 2011 and 60 countries have joined to date (Spain joined on April 2012). This proves the growing interest for open government worldwide (Ramírez-Alujas; Dassen, 2014). Open government is “a form of open, continuous and two-way communication between public authorities and citizens, businesses and civil society” (Hofmann; Ramírez-Alujas; Bojórquez-Pereznieto, 2012). Open government’s three pillars are participation, collaboration, and accountability. European institutions are supporting information reuse policies as catalysts of the digital economy and the basis of democratic transparency. The open data movement promotes the release of government data (Peset; Ferrer-Sapena; Subirats-Coll, 2011) to make it available to the public, commonly referred to as “open government data”. Ultimately, anyone or any organization could use the open government data to create new information and services (Marcos-Martín; Soriano-Maldonado, 2011; Weiskopf; Weng, 2013).

The release of public sector information allows for establishing mechanisms for transparency and encourages citizen collaboration. Some authors, however, suggest the policy of releasing data does not necessarily equal transparency, because it is necessary for data to be processed and presented in a way that can be reused to create studies, value added services, or to generate knowledge for society (Porter; Millar, 1985; Shapiro; Varian, 1998; Ferrer-Sapena; Peset; Aleixandre-Benavent, 2011). Reusing this information can lead to new digital products and services, which boosts the economy and business. New companies, even with limited available resources, could appear using new business models based on the reprocessing of such information and providing additional value to it (Marcos-Martín; Soriano-Maldonado, 2011). Spain is currently entering into an active and dynamic period of public data release through numerous new portals. The worldwide open data locator of CTIC Foundation provides a map of the global situation in which there are 267 websites publishing datasets, 32 out of them originate in Spain. However, our own study increases that figure to more than 50.

http://datos.fundacionctic.org/sandbox/catalog/faceted

Today, smart cities portals offer a lot of public information which is pointed through catalogs. These catalogs provide access to sources of information that can be used to create new services by the private sector and they are also valuable sources for big data business (LaValle et al., 2011; Kitchin, 2014). The update of Directive 2003/98/EC on the re-use of public sector information in the European Union, June 2013 (European Commission, 2013), confirms the importance of this issue at the highest European political level. The provision of public information in Spain is mainly regulated by Law 37/2007 of 16 November on the reuse of information (Spain, 2007), which aims to “harmonize commercial exploitation of public sector information and to publish all available documents that public administrations hold, considering that they are essential for the development of the right to access to knowledge, which is a basic principle of democracy.” This law is complemented by the National scheme for interoperability (RD 4/2010 of 8 January) (Spain, 2010), as well as the recent Law on transparency, access to information and good governance (Law 19/2013, of December 9) (Spain, 2013). Other national, regional, or local regulations set additional conditions to some areas of this subject. Open data is one of the main elements for transparency and open government policies.

“Current policies demanding an increase in transparency, open government, and the implementation of smart cities necessitate publishing open data.”

The release of data implies structural and organizational changes in public administration. After data is released, administrators need to establish policies of use and set privacy restrictions for the information; doing so it reduces resources and time needed to reuse the information when preparing reports (Ferrer-Sapena; Peset; Aleixandre-Benavent, 2011; LaValle et al., 2011). However, often the public sector has not established, or its not the core of its activity, organizational routines that allow users to extract the full value of open data. And this is where there is a gap filled by companies creating added-value services that make information more comprehensive and affordable (Ferrer-Sapena; Peset; Aleixandre-Benavent, 2011; Lee; Kwak, 2012). There are still very few governments or public entities that can ensure, or at least estimate quantitatively, the benefits of their policies on open government data. Most data reuse policies in public agencies are mainly based on principles such as civil liberties, compliance with existing laws, or public statements supporting transparency (Huijboom; Van-den-Broek, 2011).

In Spain, Red.es (Red.es, 2011; 2012), a public agency, analyzed the economic impact of the industrial sector that reuses open data information, but they did not establish a causal relationship with its data release policies. Almost all initiatives analysed show a lack of quantitative and qualitative metrics that can adequately estimate their impact. Fournier-Tombs (2011) explains that the analysis of the overall impact of data publishing sites is technically complicated. This author conducted a review of existing studies addressing the influence of the websites on the user community and notes that the methods based on measuring the use (surveys re-users and users) has a limited utility because
they cannot obtain statistically significant samples. Therefore, it comes as a strategic need to have metrics that assess reuse. These metrics can facilitate the efficient use of government resources and provide the most valuable information to businesses in order to create innovative business models. The aim of this paper is to explain the development and utility of a new metric to assess the reusability of this information.

2. Data reuse

Reuse is defined as, “To use something again.” The importance of reusing the information has already been studied in many other fields, including the financial sector (Chan; Greenbaum; Thakor, 1986), clinical medicine (Kim, 2005), and scientific information (Faniel; Jacobsen, 2010).

Spanish Law 37/2007 (Spain, 2007) defines the reuse of public sector information as, “The use of documents held by the administrations and public sector bodies, by natural or legal persons, with commercial or non-commercial purposes provided such use does not constitute a public administrative activity.” This regulation sets four general conditions for reuse:

- The content of the information should not be altered;
- The meaning of information should not be distorted;
- The source must be cited;
- The date of the last update should be mentioned.

In this paper we define reusable information as that which meets four minimum conditions that facilitate its use:

1) Absence of technical barriers to reuse;
2) Possibility of automated access to information (Eaves, 2010);
3) Existence of a legal framework that allows its use (Eaves, 2010);
4) Access to knowledge of the structure of the information published.

Some authors have proposed classifications and mechanisms for evaluating the quality of information, but they do not discuss reuse assessment.

For example, Berners-Lee five-star ranking is aimed at assessing whether the data is ready to be connected (linked data) with other datasets. The metric of Pipino, Lee, and Wang (2002) has 16 dimensions of analysis to evaluate quality of information, but not their reusability. Ren and Glissmann (2012) analyze the data architecture of the organization to release open data with the highest possible quality of information. In order to meet that requirement they present a 5-stage process to identify relevant information, check its condition, and design publication mechanisms. However, they do not include a proposal for evaluating the reuse of that information.

3. Meloda

Meloda was first designed in 2011 as a response to the lack of homogeneity in the released datasets of pioneering public sector open data portals. The goal of the metric was to analyze the reusability of the information.

For its preparation the three laws of open government data (Eaves, 2010) were considered:

- Be accessible online;
- Be readable by machines;
- That the legal framework allows uses other than the original, and the principles of open government data (table 1) (Lessig, 2007).

The Meloda metric allows qualifying the information and assessing its reusability.

In the first stable version of Meloda three dimensions were considered:

- Technical standards. Reflects the importance that the information is stored in a non-proprietary standard.
- Access. Access to all readable information automatically, with as much detail as possible and for the highest number of uses (complete, machine processable, primary, and accessible).
- Legal. Importance of the legal framework on the use of information and lack of legal barriers to their use, regardless of respect for privacy and security (license-free).

The first use of Meloda was performed on 200 data sources coming from open data portals (regional, local, and national) in Spain.

It was found necessary to add a fourth dimension to take into account the data model to publish, reflecting the importance of the data structure to process information (machine processable).


<table>
<thead>
<tr>
<th>Table 1. Principles of open government data</th>
<th><a href="http://opengovdata.org">http://opengovdata.org</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete</td>
<td>All public data is made available. Public data is data that is not subject to valid privacy, security, or privilege limitations</td>
</tr>
<tr>
<td>Primary</td>
<td>Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms</td>
</tr>
<tr>
<td>Timely</td>
<td>Data is made available as quickly as necessary to preserve the value of the data</td>
</tr>
<tr>
<td>Accessable</td>
<td>Data is available to the widest range of users for the widest range of purposes</td>
</tr>
<tr>
<td>Machine processable</td>
<td>Data is reasonably structured to allow automated processing</td>
</tr>
<tr>
<td>Non-discriminatory</td>
<td>Data is available to anyone, with no requirement of registration</td>
</tr>
<tr>
<td>Non-proprietary</td>
<td>Data is available in a format over which no entity has exclusive control</td>
</tr>
<tr>
<td>License-free</td>
<td>Data is not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security, and privilege restrictions may be allowed</td>
</tr>
</tbody>
</table>
3.1. Dimensions

The current version of Meloda has four dimensions.

1) Technical standards or technical structure in which the data are available. Rate the information on three levels (table 2).

Level 1: The information is offered on privative technical standards including intellectual property restrictions for use.
Level 2: Is reached when the data are stored in formats that do not have restrictions on their use.
Level 3: Is reached when Level 2 requirements are met and when the data has a metadata schema.

2) Access to information or mechanisms by which information is released. Rate information in five levels (table 2).

Level 1: Dataset that either does not have web access or requires a non-automated request.
Level 2: Access by manual user interaction or compulsory registration.
Level 3: URL allowing only access to the complete dataset (not single components).
Level 4: URL could include parameters so the user does not need to access the entire data set.
Level 5: It includes those datasets that provide the ability to perform queries on data and cross them with external sources.

3) Legal framework: License assigned to the dataset. Rate information in five levels (table 2).

Level 1: The use of the information is completely restricted without explicit permission of the author.
Level 2: The use of information is allowed only for private purposes.
Level 3: Only non-profit use of the information is allowed.
Level 4: Commercial use is allowed.
Level 5: A non-limited use and creation of derivatives is allowed as long as the authorship is acknowledged.

4) Data model: Model used to publish information and the number of times it is used in other data sources. Rate information in five levels (table 2).

Level 1: Do not have a published data model.
Level 2: Data fields of the data model are identified.
Level 3: Detailed and public specification of the fields of the data model is published (even being own or private).
Level 4: An external standardization body data model is used, although it is not widespread.
Level 5: A generalized data model is used coming from an external standardization body.

The latest version, Meloda 3.10, has four dimensions: technical standards, access to information, legal framework and data model.

Table 3 shows the relationship between Meloda’s levels and the definition of open data (green background). It can be seen that although Meloda is focused on open data sources it also includes data that do not match this definition; therefore, Meloda could be useful in other areas.

3.2. Meloda’s evaluation process

Figure 1 describes the evaluation process. Table 4 describes the weights assigned to each of the dimension’s levels. The final score for a dataset is calculated as the fourth root of the product of the scores on each dimension. This formula produces a more uniform distribution of results. It also allows analysis of the effect of a change in one or more dimensions over the final score.

Table 4 shows the relationship between Meloda’s levels and the definition of open data (green background). It can be seen that although Meloda is focused on open data sources it also includes data that do not match this definition; therefore, Meloda could be useful in other areas.

Meloda = 100 * \( \sqrt[4]{ \text{Technical standards} \times \text{Access} \times \text{Legal} \times \text{Data model} } \)
Table 5 presents the reuse ranges between 0 and 100, proposed after the first evaluation of the results with real data.

An example of an assessment for a dataset with Meloda is included.

Institution A publishes dataset as linked data but with their own data model, licensed for commercial use, and with a Sparql endpoint to query the dataset. Corresponding levels in each dimension are:

- Technical standards: Level 3 (RDF metadata) 100%
- Access: Level 5 (queryable via Sparql) 100%
- Legal: Level 4 (commercial) 90%
- Data model: Level 2 (ad hoc data mode) 30%

Meloda = 100$\times$1$\times$0,9$\times$0,3 = 72,08

4. Conclusions

In this work a new metric for the reuse of information has been described. Meloda characterizes reuse through four dimensions, and provides a quantitative approach for its assessment.

The most recent version (Meloda 3.10) is under review for improvement. In future versions time dimension and geolocation of information together with the refinement of the weights of the different dimension’s levels will be considered.

This metric should be of interest to any organization (public, private, for-profit, or not-for-profit) that releases data or provides products and services based on data sources.

An important area for future research is the analysis of information and data reuse in order to identify which types of information, or what specific features, have the greatest potential for reuse.

On one hand, background studies on open government tend to analyse the social value creation (transparency and citizen collaboration). On the other hand, literature on information reuse is focused on the use of information and on the potential economic value offered as a result of an efficient and innovative use of public resources (Jetzek; Avital; Bjørn-Andersen, 2013; 2014). However there is a shortage of literature providing empirical analyses of the economic impact of information reuse. Therefore, future research should analyse the economic value of information reuse.

Notes

1. Smart cities are a public-private ecosystem providing services to the citizens and their organizations with an extensive use of IT technologies.

2. Five stars are: Open license, structured data, open standards, unique resource identifier (URI) and connected sources. http://5stardata.info

3. Use of first version of Meloda provided indications that a sum function for dimensions’ scores resulted in incoherent global scores, and also that a uniform distribution of weights did not provide good insight on the reuse ranges.
Meloda, a metric to assess open data reuse

Acknowledgements

This work has been financed by the project ECO201236775 of the Ministry of Economy and Competitiveness (Spain) and it has been supported by the Apps4eu project funded by the Information and Communication Technologies Policy Support Programme (IC PSP).

Bibliography

http://dx.doi.org/10.1016/0378-4266(86)90008-7


http://dx.doi.org/10.3145/epi.2011.may.03

Hofmann, Andrés; Ramírez-Alujas, Álvaro; Bojórquez-Perez nieto, José-Antonio (coords.) (2012). La promesa del gobierno abierto, Itiap. http://www.lapromesadelgobiernoabierto.info


http://dx.doi.org/10.3145/epi.2011.may.07

http://dx.doi.org/10.3145/epi.2011.mar.06

Nueva colección EPI Scholar
Libros académicos y científicos de Información y Documentación

Primer título:
Cibermetría. Midiendo el espacio red de Enrique Orduña-Malea e Isidro F. Aguillo

A pesar del crecimiento de la disciplina de la Cibermetría desde mediados de la década de los noventa, son escasos los libros académicos o manuales dedicados en exclusiva a la misma desde un contexto de las ciencias de la información y documentación. Este libro pretende cubrir este claro hueco en la literatura tanto nacional como internacional.