Meloda 5: A metric to assess open data reusability

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

Alberto Abella ✉
http://orcid.org/0000-0003-0243-4606

Fiware Foundation
Infante Diego, 1
28050 Madrid, Spain
alberto.abella@fiware.org

Marta Ortiz-de-Urbina-Criado
http://orcid.org/0000-0001-7527-6798

Universidad Rey Juan Carlos
Facultad de Ciencias Jurídicas y Sociales
Paseo de los Artilleros, s/n.
28032 Madrid, Spain
marta.ortizdeurbina@urjc.es

Carmen De-Pablos-Heredero
http://orcid.org/0000-0003-0457-3730

Universidad Rey Juan Carlos
Facultad de Ciencias Jurídicas y Sociales
Paseo de los Artilleros, s/n.
28032 Madrid, Spain
carmen.depablos@urjc.es

Abstract

An updated metric developed to assess the degree of open data reusability, called *MEtric for the evaLuation of Open DAta: Meloda 5* is presented. Previous version of the metric, *Meloda 4*, had six dimensions: the legal licensing of data, the mechanisms to access the data, the technical standards of the datasets, the data model, the geographic content of the data and the updating frequency. With all these dimensions, the metric provides a quantitative evaluation about how reusable the datasets released are. During the last five years, this metric has been cited and used by some other authors but the extensive use of the metric has brought to light some of its limitations. In order to get deeper insights into the topic, a panel of international experts has been surveyed about two aspects of the metric. First aspect was what other factors should be considered in order to qualify the reusability of a released dataset. And the second aspect was the internal structure, the levels for every dimension of the metric; if they should be increased, merged, removed or divided. Considering the results of the survey, first, we identify the factors / dimensions that should be kept: legal licensing, access to information, technical standard, standardization, geolocation content and updating frequency of data. Second, we consider the inclusion of two new dimensions: dissemination and reputation. Then, we present the new internal structure, the levels for each dimension, and the measures to evaluate the degree of reuse of each dataset. Finally, a standardization of the metric for other steps of the data impact process, data reuse analytics and data-driven services generation are presented together with future research lines.

Keywords

Open data; Reusability; *Meloda*; Metric; Standardization; Geolocation; Dissemination; Reputation; Datafication; Value.

1. Introduction

The big data movement was born with the reuse and exploitation of available information, and it has increasingly grown as a consequence of the exponential increase in the amount of available data by several orders of magnitude (**Hashem** *et al.*, 2016).

Data reuse is the use of data for a purpose different from the one intended by the data producer (**Karasti**; **Baker**, 2008; **Zimmerman**, 2008; **Martínez-Méndez**; **López-Carreño**; **Pastor-Sánchez**, 2014; **Faniel**; **Kriesberg**; **Yakel**, 2016).

Yoon (2017, p. 946) explains how

> "benefits of reusing shared data have been asserted by a number of researchers, and data reuse has been a major concern in many disciplines".

Therefore, reusability analysis plays a key role in determining what factors and conditions can foster data reuse. The massive release of data helps to create added-value products and services and encourages scientific research through data reuse. The release of the data

> " Data reuse is the use of data for a purpose different from the one intended by the data producer "

implies an organizational change, a structural transition and a redefinition of the processes of private firms and public administrations. It requires establishing new data usage and the deployment of privacy policies (**Ferrer-Sapena**; **Peset**; **Aleixandre-Benavent**, 2011; **LaValle** *et al.*, 2011, **Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero**, 2018; 2019).

The literature on reuse of data focuses on data usage by the public sector, and on the analysis of their economic value. This economic value can be created through the efficient use of public resources –efficiency– and / or by creating new products and services –innovation– or increased competition –transparency– or allowing the birth of new social entities as identified by **Kassen** (2013). In this sense, **Chatfield** and **Reddick** (2017) state that previous research has found that the reuse of open government data can generate open innovation (**Zuiderwijk**; **Janssen**; **Davis**, 2014; **Susha**; **Grönlund**; **Jassen**, 2015) and economic and social value (**Jetzek**; **Avital**; **Björn-Andersen**, 2014). Indeed, some reports discuss extraordinary expectations for the social and economic impact of data reuse (*Pira International*, 2000; **Dekkers** *et al.*, 2006; **Vickery**, 2011; **Carrara** *et al.*, 2015; **Ceconni**; **Radu**, 2018). But, **Carrasco** and **Sobrepere** (2015, p. 663) explain that there is a

> "lack of a clear way to enable empirical analysis and quantitative measurement of open government data initiatives".

Datafication, that is to say, the progressive quantification of most of the aspects of life in terms of data (**Cukier**; **Mayer-Schoenberger**, 2013) has never been more possible than now. However, not all the organizations collect data for internal exploitation, conversely, some of them have the goal to share it with other organizations or individuals or even with the surrounding society. This is the case for most public administrations, many scientific initiatives and even some authors willing to share their data in order to achieve larger audiences (**Ceconni**; **Radu**, 2018). These collections of data, datasets, are released with legal licenses allowing data reuse. One of the most popular open data licenses, creative commons by, allows reusers to use the released data for any purpose, to process the data and to share them again with other users. Many benefits have been identified for open data (**Janssen**; **Charlabidis**; **Zuiderwijk**, 2012; **Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero**, 2015). Notably the release of data helps organizations to be more effective, to be more accountable and it stimulates competitiveness among several others.

To analyze open data, one of the most known metrics is the 5-star scale of Berners-Lee. **Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero** (2018) explain that the levels from 1 to 3 of that metric measure data availability, and levels 4 and 5 measure the capacity of being linked with other datasets. That metric just considers the data format; but it does not consider relevant aspects to measure the open data reuse potential such as geolocation and updating frequency of data (**Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero**, 2018). In that sense, **Abella**, **Ortiz-de-Urbina-Criado** and **De-Pablos-Heredero** (2014) analyzed what factors could affect the sharing of data and what features make data more prone to be reused. The metric *Meloda* condensed these findings and became a tool to assess the reusability of individual datasets. In 2014's version (namely version 4),

> " The metric *Meloda* condensed these findings and became a tool to assess the reusability of individual datasets "

six dimensions were taken into consideration. The legal licensing of data, the mechanisms to access the data, the technical standards of the datasets, the data model, that is to say, if the data meet some agreed model for their contents. Finally, the geographic content of the data and the updating frequency were also considered. For every dimension, several levels of accomplishment were defined and weights were assigned to each of them. With all these dimensions, the metric provides a quantitative evaluation about how reusable the dataset released are.

During the last five years, this metric has been cited and used by some other authors (i. e., **Solar** *et al.* 2014; **Gértrudix**; **Gertrudis-Casado**; **Álvarez-García**, 2016; **Vicente-Paños**; **Jordán-Alfonso**, 2017; **Zhao**; **Yan**; **Li**, 2018). *http://www.meloda.org/scientific-article-about-Meloda*

However, the extensive use of the metric has brought to light some of its limitations trying to explain why some datasets

differ in their reuse having similar values for the metric. Thus, this article presents an updated metric developed to assess the degree of open data reusability, called *MEtric for the evaLuation of Open Data* (*Meloda*). In order to get deeper insights into the topic, a panel of international experts has been surveyed about two aspects of the metric. First aspect was what other factors should be considered in order to qualify the reusability of a released dataset. And the second aspect was the internal structure, the levels for every dimension of the metric; if they should be increased, merged, removed or divided.

2. Open data reuse. A metric

The reuse of public sector information can be defined 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 that such use does not constitute a public administrative activity" (*Spanish Law 37/2007*).

Open data reuse is the reuse and redistribution of data including the intermixing with other datasets (*Open data handbook*, 2017). Following **Abella**, **Ortiz-de-Urbina-Criado** and **De-Pablos-Heredero** (2014), reusable open data meets four minimum conditions that facilitate

> This article presents an updated metric developed to assess the degree of open data reusability, called *Meloda 5*

its use: absence of technical barriers to reuse; possibility of automated access to information; existence of a legal framework that allows its use (**Eaves**, 2010) and access to knowledge of the structure of the information published.

2.1. Previous version of *Meloda*

To analyze the degree of reusability, the metric *Meloda* has been improved. *Meloda* is a metric that provides a quantitative assessment of the degree of reusability of data, and it is in a continuous process of improvement. The first version, version 1.0, was born in 2011. A further explanation of *Meloda 3.1* is described in **Abella**, **Ortiz-de-Urbina-Criado** and **De-Pablos-Heredero** (2014).

Meloda 1.0 to *2.5* identifies three dimensions to characterize datasets: legal license, access and technical standards. In 2012, 204 data sources were tested, more accurate results were provided, and one more dimension was added: data modelling of the datasets. The formula for calculating the metric value was also adjusted to include the new dimension (**Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero**, 2014). *Meloda 3* was tested in a sample of 30 European smart cities, and two new dimensions were included: real-time and geographic information, evolving to *Meloda 4* (**Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero**, 2017a). Version 4.13 was improved by redefining some of the levels (2 and 3) of the technical dimension.

In figure 1, a diagram describing the assessment process of a dataset with *Meloda 4* is presented (assessments for *4.13* remains equal to *4.01*). In step 1, the metadata of the dataset are validated using the criteria defined for every level in

$$MELODA = 100 * \sqrt[6]{Legal * Access * Technical\ standard * Data\ model * Geolocation * Real\ time}$$
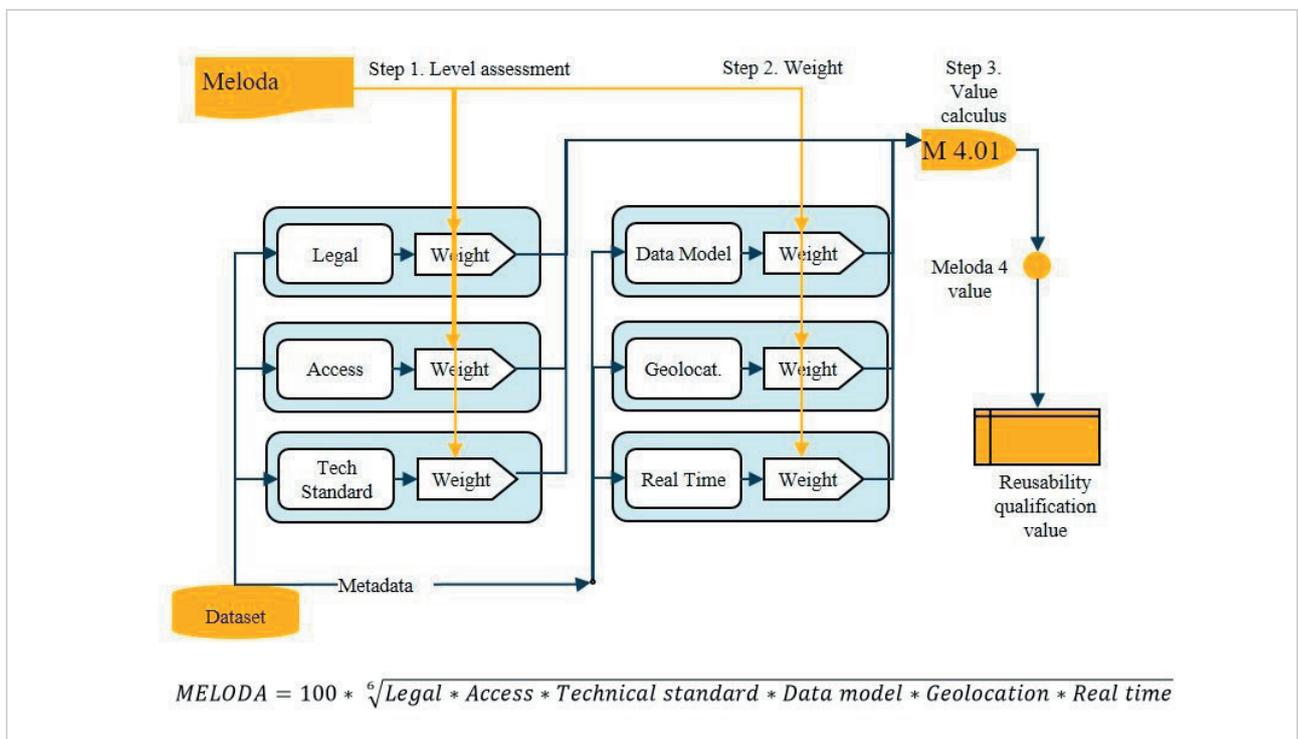
Figure 1. *Meloda 4.* Assessment process
Source: **Abella**; **Ortiz-de-Urbina-Criado** and **De-Pablos-Heredero** (2017a, p. 50)

Meloda and by determining the level reached in every dimension. In step 2, a weight for every dimension (Table 1) is assigned. In step 3, a formula is applied to the different weights assigned in step 2 (see the lower part of figure 1). Finally, in step 4, a level of reusability is assigned based on the results (Table 2).

For version *4.13*, these dimensions are the legal licensing of data, the technical standards used for the release of the data, the access mechanisms of the data, the data model used (i.e., characteristics of data structure), the geolocation (i.e., inclusion of geographic information in datasets) and timeliness or real-time (i.e., update period of the information and the time elapsed between versions).

Every dimension of the metric is divided into several levels (up to five). Every level presents a clear criterion of fulfilment. Each level of every dimension also has a weight (Table 1). To provide a final grade for reusability, those levels are weighted in a formula (Figure 1). Table 2 presents the global qualification based on applying the formula to quantify the degree of reusability.

Table 1. *Meloda 4.13*: Dimensions, levels and weights

Legal	Access to information	Technical standard	Data model sharing	Geolocated information	Real-time information
Level 5					
No restrictions or only attribution (100%)	API or query language (100%)		Global open data model (100%)	Full geographical information (100%)	Seconds (100%)
Level 4					
Commercial reuse (90%)	Web access unique with parameters to single data (90%)	Open standard, individual metadata (100%)	Local open data model (90%)	Coordinates (90%)	Minutes (90%)
Level 3					
Non-commercial reuse (25%)	Web access or unique URL parameters to dataset (50%)	Open standard reusable (60%)	Own ad hoc data model published (50%)	Complex text field (50%)	Hours (70%)
Level 2					
Private use (10%)	Web Access URL with registration or with web interaction (10%)	Closed standard reusable and open non reusable (35%)	Own ad hoc data model (35%)	Simple text field (30%)	Days (40%)
Level 1					
Copyright (0%)	No web access or manual request (0%)	Closed standard non reusable (10%)	Not known data model (15%)	No geographic information (15%)	Longer than week (15%)

Source: Updated from **Abella**; **Ortiz-de-Urbina-Criado** and **De-Pablos-Heredero** (2014)

Table 2. Qualification of reusability based on the value of *Meloda 4*

Meloda 4 range	0-25	25-50	50-75	75-100
Reuse qualification	Deficient	Basic	Good	Optimum
Symbol				

Source: **Abella**; **Ortiz-de-Urbina-Criado** and **De-Pablos-Heredero** (2014, p. 586)

2.2. *Meloda 5*

In order to improve the new version of the metric (version *5*), an online survey (see annex) was sent to a group of experts notorious for their publications on the open data field. The survey was sent to 28 experts with different profiles, some of them share profiles, concretely, academics and consultants, and 12 of them are from countries different to Spain. The profiles selected, apart

> It is presented a metric with a new internal structure, that is to say, the levels for each dimension and the measures to evaluate the degree of accomplishment for every dimension

from their remarkable experience, included open data managers from public administrations (6), academics (22), and consultants on open data and linked data (12) coming from these countries: Argentina, Brazil, Chile, Colombia, France, Greece, Netherlands, Spain, UK and USA. Survey was anonymous. The survey was completed by 11 of them (39.3%) during the period from 15/07/2018 to 14/09/2018. Results and the changes of the metric based on them are presented hereunder.

2.2.1. Dimensions' analysis

To analyze the dimensions, the experts were asked about the factors that are decisive for measuring data reusability (Table 3).

Table 3. Key factors for reusability of data

Factors	1	2	3	4	5	6	7	Mode	Mean
Legal licensing of datasets					1(9.1%)	5(45.5%)	5(45.5%)	6	6.36
Accessibility. How easy is to find and retrieve the data						1(9.1%)	10(90.9%)	7	6.91
Technical standard					2(18.2%)	2(18.2%)	7(63.6%)	7	6.45
Data model				2(18.2%)	1(9.1%)	4(36.4%)	4(36.4%)	6	5.91
Geographic content				2(18.2%)	1(9.1%)	4(36.4%)	4(36.4%)	6	5.91
Updating frequency				1(9.1%)	1(9.1%)	5(45.5%)	4(36.4%)	6	6.09
Topic of the datasets content				1(9.1%)	2(18.2%)	4(36.4%)	4(36.4%)	6	6.00
Reliability. If the dataset meets the announced updating frequency					1(9.1%)	4(36.4%)	6(54.5%)	7	6.45
Support to use. If there is a contact to make questions about the dataset			1(9.1%)	3(27.3%)	4(36.4%)	3(27.3%)		5	4.82
API availability If there is a documented API to describe access to the datasets				2(18.2%)	1(9.1%)	3(27.3%)	5(45.5%)	7	6.00
SEO friendly. If the dataset is located in a web site friendly to be indexed by search engines i.e. *Google*		2(18.2%)	3(27.3%)		2(18.2%)	4(36.4%)		7	5.27

The first six factors are the existing dimensions of *Meloda 4*, so the experts agreed that these aspects should be kept. In addition, they were asked about other possible factors and, in particular, the importance of considering the topic of the datasets, of analyzing their reliability and the API's availability. The experts also proposed other factors to consider such as popularity, knowledge of their existence, visualizations, updating notices, interoperability, standardization of data, harmonization of portals, data offers, and data literacy.

With this information, it has been proposed to modify, on the one hand, the name of one of the existing dimensions, formerly named "data model", now renamed as "standardization" to consider more clearly the issue of standardization and the harmonization of data. On the other hand, two new dimensions have been included into the metric, one to measure the dissemination of the open data activities of the repository and knowledge of the released data, and another, to analyze the reputation of the data source.

2.2.2. Analysis of the levels' structure for each dimension

To analyze the levels within each dimension, the experts were asked if they would be maintained, merged, included new ones or eliminated (Table 4).

Table 4. Dimensions and levels in *Meloda 4* and experts' recommendation on structure

Dimensions	Levels	Maintain	Remove	Do not know/ Do not answer
Legal licensing	1. Copyright	6(54.5%)	3(27.3%)	2(18.2%)
	2. Private use	8(72.7%)	1(9.1%)	2(18.2%)
	3. Non-commercial reuse	9(81.8%)		2(18.2%)
	4. Commercial reuse	9(81.8%)	1(9.1%) merge level 5	1(9.1%)
	5. No restrictions or only attribution	10(90.9%)		1(9.1%)
Access to datasets	1. No web access neither manual request	6(54.5%)	4(36.4%)	1(9.1%)
	2. No web access or manual request	5(45.5%) and 1(9.1%) joint level 1	4(36.4%)	1(9.1%)
	3. Web access or unique URL parameters to dataset	9(81.8%)		2(18.2%)
	4. Web access unique with parameters to single data	10(90.9%)		1(9.1%)
	5. API or query language	10(90.9%)		1(9.1%)
Technical standard	1. Closed standard non reusable	7(63.7%)	2(18.2%)	2(18.2%)
	2. Closed standard reusable and open non reusable	7(63.6%)	2(18.2%)	2(18.2%)
	3. Open standard reusable	8(72.7%)		3(27.3%)
	4. Open standard, individual metadata	8(72.7%)		3(27.3%)

Dimensions	Levels	Maintain	Remove	Do not know/ Do not answer
Data model	1. Not known data model	7(63.6%)	2(18.2%)	2(18.2%)
	2. Own ad hoc data model	7(63.6%)	2(18.2%)	2(18.2%)
	3. Own ad hoc data model published	9(81.8%)		2(18.2%)
	4. Local open data model	9(81.8%)		2(18.2%)
	5. Global open data model	9(81.8%)		2(18.2%)
Geolocation content	1. No geographic information	8(72.7%)	1(9.1%)	2(18.2%)
	2. Simple text field	6(54.5%)	3(27.3%)	2(18.2%)
	3. Complex text field	8(72.7%)	1(9.1%)	2(18.2%)
	4. Coordinates	9(81.8%)		2(18.2%)
	5. Full geographical information	7(63.6%)	2(18.2%)	2(18.2%)
Updating fre-quency of data	1. Longer than week	7(63.6%)	1(9.1%)	3(27.3%)
	2. Days	8(72.7%)		3(27.3%)
	3. Hours. Hours. Updating period ranges from 1 hour to 24 hours	8(72.7%)		3(27.3%)
	4. Minutes. Updating period ranges from 1 Hour to 1 minute	8(72.7%)		3(27.3%)
	5. Seconds. Updating period is lower than 1 minute	8(72.7%)		3(27.3%)

For the legal licensing dimension, several experts proposed to eliminate level 1 because copyright does not allow any reuse of data without explicit permission of the owner, and others proposed to merge levels 4 and 5. For the access to information dimension, some experts suggested that the first two levels had to be eliminated because they do not give access to data reuse. Accordingly, these two levels have been removed. For the third dimension, technical standard, some experts propose to eliminate or reformulate level 1, so it has been eliminated because it is a level that does not allow the reuse of data.

The dimension formerly called *data model* referred to aspects related to the normalization of the data. So, taking into account the experts' improvement proposals, its name has been modified and the levels have been adapted so that they better reflect the aspects related to the standardization and harmonization. For the geolocation dimension, following the opinions of some of the experts, several of the levels have been merged. For the updating frequency of data dimension, the five levels schema has been maintained but reformulating the time periods. Finally, two new dimensions with three levels each one have been proposed. The dimensions and levels of *Meloda* 5 are presented in next Table 5.

Table 5. Dimensions and levels of *Meloda 5*

Dimensions (maximum 61 points)	Levels
Legal licensing (max. 6 points)	1. Private use 2. Non-commercial reuse 3. Commercial reuse or no restrictions
Access to information (max. 6 points)	1. Web access or unique URL parameters to dataset 2. Web Access unique with parameters to single data 3. API or query language
Technical standard (max. 6 points)	1. Closed standard reusable and open non reusable 2. Open standard reusable 3. Open standard, individual metadata
Standardization (max. 10 points)	1. Own data model standardization 2. Own ad hoc data model standardization published (harmonization) 3. Local standardization 4. Global standardization
Geolocation content (max. 6 points)	1. No geographic information 2. Simple or complex text field 3. Coordinates or full geographical information
Updating frequency of data (max. 15 points)	1. Longer than 1 month 2. Monthly. Updating period ranges from 1 month to 1 day 3. Daily. Updating period ranges from 1 day to 1 hour 4. Hour. Updating period ranges from 1 hour to 1 minute 5. Seconds. Updating period is lower than 1 minute
Dissemination (max. 6 points)	1. Communication / dissemination not systematic 2. Available resources on updates (i.e., RSS feed) 3. Proactive dissemination / push dissemination (information automatic and timely)
Reputation (max. 6 points)	1. No information about the reputation of the data source 2. Statistics or reports published on user's opinions 3. Indicators or rankings on reputation of the data source

2.2.3. Qualification of reusability based on the value of *Meloda 5*

In version *4* of the metric, each level was assigned a weight and then a formula was applied to analyze the degree of data reuse. In version *5*, some of the ideas that experts have proposed have been considered, and the weighting for each level has been revised. In the new version, each level will be assigned the value it has (1, 2, 3, 4 or 5). To evaluate the degree of reuse of each dataset, two measures will be used:

1) the sum of the scores obtained in each dimension (being the maximum value that can be obtained 61 points) (see table 5); and

2) for each dimension, a descriptive analysis of the frequency of each level.

The first measure will allow obtaining a ranking of datasets according to their degree of reusability; while the second one allows to have a more detailed image for each dataset and to identify which dimensions need improvement.

3. Metric limitation for the analysis of open data impact

Assessing the reusability of the data published is only one step in the process of assessing the impact of data publication (**Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero**, 2017a; 2019). Two steps required to set an effective strategy for data impact are the analysis of the actual use of the data published and the assessment of the applications and services based on them.

3.1. Analysis of data reuse

Organizations need to have a system / metric to analyze the use and impact of their published data in order to justify the resources used for publishing data. Currently, there is not a generally accepted and used metric that would allow comparative analysis. Therefore, although managers of the data publishing facilities can assess their own reuse statistics, they cannot compare their results with other data publishers. Thus, their ability to improve in their publishing strategy is limited. **Abella** *et al.* (2019) explain that it would be possible to incorporate them into the main data publishing tools (e. g., *CKAN*, *Socrata*, *Open Data Soft*, *ESRI open data*, *DKAN*, etc.) as an added value service. The minimum fields to be included in this metric should include these:

- User type field (according to these categories: For profit organization, non-profit organization, publisher, other organization related with publisher, citizens, academics)
- What dataset/distribution is accessed
- If it is accessed as a direct download or via API
- Date and time
- Referrer (origin)
- Type of referrer (SEO, search engines, social networks, own or referred)
- Search keywords
- Service (if known, application or service reusing the data)
- Theme (*DCAT-AP*) of the data accessed
- Frequency of data update
- Technical format of the data accessed
- Type of query made to access the data/s (query)
- Accessed data has data model
- Data accessed has geolocation.

The functionality of *Meloda* helps when considering these criteria to analyze the use of data and proposes the minimum standard that they should have. Future work will develop this aspect and validate it.

3.2. Analysis of services based on published data

In addition, there is a lack of literature providing approaches (theoretical and empirical) for the impact of data reuse. Therefore, future research will analyze, considering *Meloda* results, their impact. In that sense, an analysis of the applications and services developed reusing data can be performed considering the proposal of **Abella** *et al.* (2019). The authors state that

> "it would be interesting to federate not only the data but also the applications developed to perform an analysis of the actual generation of services, and allow applications that are registered in several portals to be considered only once."

Authors propose a minimum list of the fields that should qualify an application based on open data: official id (unique code between portals), title, description, authors, creation date, last update, platforms on which is available, price, business model, type of authors, *NTI-RISP* (*Norma Técnica de Interoperabilidad de Reutilización de Recursos de Información / Technical Standard for Interoperability of Information Resource Reuse*) themes in which it serves, *DCAT-AP* (*Data Catalogue Vocabulary-Application Profile*) themes in which it serves, theme *NTI-RISP* from which it extracts data, *DCAT-AP* themes from which it extracts data, link to the website of the authors and a contact mail.

4. Conclusion

In this paper a new version of *Meloda*, a metric for the reuse of information, has been described. To develop the new version, an online survey was launched to a group of experts. Results of the survey exceeded author's expectations about the simplification of the metric and what factors would additionally have to be considered. The recommendations coming from the group of experts led to a simplification of most of the dimensions into three levels but for the standardization (4 levels) and the update frequency dimension (5 levels). In the most recent reports on open data reuse in Spain, open data-driven services with business models, in 2017, were 66% in real-time, while, in 2019, accounts 58% (**Abella**; **Ortiz-de-Urbina-Criado**; **De-Pablos-Heredero**, 2017b; **Abella** *et al.*, 2019). These figures show the relevance of short updating periods in open data reuse. But it is also true that there are some data with longer updating periods that could be attractive for reusers. Therefore, the dimension "updating frequency of data" presents 5 levels, two levels of which consider longer updating periods (levels 1 and 2). Additionally, the recommendations suggested that the reuse of data does not only depend on the dataset itself and the data contained, but also on the entity releasing them and its dissemination efforts. Thus, two additional dimensions have been included into the metric accordingly: dissemination and reputation.

The overall metric will be useful for open data portal managers, but also for data management systems' manufacturers, open data consultants and academics. The metric helps them to assess the open data to be published and to boost data reuse. Entrepreneurial data-driven initiatives will emerge from the potential of data.

The quest for the analysis of the open data impact led the authors to propose the standardization of other steps on the process for data impact. On one hand, the standardization of the statistics of data reuse and the services and applications created. Without them, the ability of data managers to implement successful strategies will be restricted to their own knowledge and experience and strong barriers to collaboration would limit data reuse and therefore its impact. And this is the goal for most of them.

5. References

Abella, Alberto; **Ortiz-de-Urbina-Criado, Marta**; **De-Pablos-Heredero, Carmen** (2014). "*Meloda*, a metric to assess open data reuse". *El profesional de la información*, v. 23, n. 6, pp. 582-588.
https://doi.org/10.3145/epi.2014.nov.04

Abella, Alberto; **Ortiz-de-Urbina-Criado, Marta**; **De-Pablos-Heredero, Carmen** (2015). "Information reuse in smart cities' ecosystems". *El profesional de la información*, v. 24, n. 6, pp. 838-844.
https://doi.org/10.3145/epi.2015.nov.16

Abella, Alberto; **Ortiz-de-Urbina-Criado, Marta**; **De-Pablos-Heredero, Carmen** (2017a). "A model for the analysis of data-driven innovation and value generation in smart cities' ecosystems". *Cities*, v. 64, pp. 47-53.
https://doi.org/10.1016/j.cities.2017.01.011

Abella, Alberto; **Ortiz-de-Urbina-Criado, Marta**; **De-Pablos-Heredero, Carmen** (2017b). *La reutilización de datos abiertos: Una oportunidad para España*. Informe elaborado como resultado del proyecto seleccionado dentro de la convocatoria 2016 del Programa de Innovación Abierta (PIA) de la Fundación Cotec para la Innovación.
http://informecotec.es/media/INFORME_REUTILIZACION-DE-DATOS.pdf

Abella, Alberto; **Ortiz-de-Urbina-Criado, Marta**; **De-Pablos-Heredero, Carmen** (2018). "Indicadores de calidad de los datos abiertos: el caso del portal de datos abiertos de Barcelona". *El profesional de la información*, v. 27, n. 2, pp. 375-382.
https://doi.org/10.3145/epi.2018.mar.16

Abella, Alberto; **Ortiz-de-Urbina-Criado, Marta**; **De-Pablos-Heredero, Carmen** (2019). "The process of open data publication and reuse". *Journal of the Association for Information Science and Technology*, v. 70, n. 3, pp. 296-300.
https://doi.org/10.1002/asi.24116

Abella, Alberto; **Ortiz-de-Urbina-Criado, Marta**; **De-Pablos-Heredero, Carmen**; **Vidal-Cabo, Christian**; **Ferrer-Sapena, Antonia** (2019). *La reutilización de datos abiertos en España II*. Report financed by desideDatum.
https://www.desidedatum.com/wp-content/uploads/2019/11/La_reutilizacio%CC%81n_datos_abiertos_espan%CC%83a_2019.pdf

Carrara, Wendy; **Chan, Wae-San**; **Fischer, Sander**; **Van-Steenbergen, Eva** (2015). *Creating value through open data: Study on the impact of re-use of public data resources*. European Commission.
https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

Carrasco, Carlos; **Sobrepere, Xavier** (2015). "Open government data: An assessment of the Spanish municipal situation". *Social science computer review*, v. 33, n. 5, pp. 631-644.
https://doi.org/10.1177/0894439314560678

Cecconi, Gianfranco; **Radu, Cosmina** (2018). *Open data maturity in Europe. Report 2018*. European Commission report.
https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf

Chatfield, Akemi T.; **Reddick, Christopher G.** (2017). "A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments". *Government information quarterly*, v. 34, n. 2, pp. 231-243.
https://doi.org/10.1016/j.giq.2017.02.004

Cukier, Kenneth; **Mayer-Schoenberger, Viktor** (2013). "The rise of big data: How it's changing the way we think about the world". *Foreign affairs*, v. 92, n. 3, pp. 28-40.
https://www.jstor.org/stable/23526834

Dekkers, Makx; **Polman, Femke**; **TeVelde, Robbin**; **De-Vries, Marc** (2006). *Mepsir (Measuring European public sector information resources). Final report of study on exploitation of public sector information – benchmarking of EU framework conditions, Executive summary and Final report Part 1 and Part 2.*
https://ec.europa.eu/digital-single-market/en/news/mepsir-measuring-european-public-sector-information-resources-final-report-study-exploitation-0

Eaves, David (2010). "The three laws of open government data". *Conference for Parliamentarians: Balancing openness and the public interest in protecting information*.
http://bit.ly/eaves2010

Faniel, Ixchel M.; **Kriesberg, Adam**; **Yakel, Elisabeth** (2016). "Social scientists' satisfaction with data reuse". *Journal of the Association of Information Science Technology*, v. 67, n. 6, pp. 1404-1416.
https://doi.org/10.1002/asi.23480

Ferrer-Sapena, Antonia; **Peset, Fernanda**; **Aleixandre-Benavent, Rafael** (2011). "Acceso a los datos públicos y su reutilización: open data y open government". *El profesional de la información*, v. 20, n. 3, pp. 260-269.
https://doi.org/10.3145/epi.2011.may.03

Gértrudix, Manuel; **Gertrudis-Casado, María-Carmen**; **Álvarez-García, Sergio** (2016). "Consumption of public institutions' open data by Spanish citizens". *El profesional de la información*, v. 25, n. 4, pp. 535-544.
https://doi.org/10.3145/epi.2016.jul.03

Hashem, Ibrahim-Abaker-Targio; **Chang, Victor**; **Anuar, Nor-Badrul**; **Adewole, Kayode**; **Yaqoob, Ibrar**; **Gani, Abdullah**; **Ahmed, Ejaz**; **Chiroma, Haruna** (2016). "The role of big data in smart city". *International journal of information management*, v. 36, n. 5, pp. 748-758.
https://doi.org/10.1016/j.ijinfomgt.2016.05.002

Janssen, Marijn; **Charalabidis, Yannis**; **Zuiderwijk, Anneke** (2012). "Beliefs, adoption barriers and myths of open data and open government". *Information system management*, v. 29, n. 4, pp. 258-268.
https://doi.org/10.1080/10580530.2012.716740

Jetzek, Thorhildur; **Avital, Michel**; **Björn-Andersen, Niels** (2014). "Data-driven innovation through open government data". *Journal of theoretical and applied electronic commerce research*, v. 9, n. 2, pp. 100-120.
https://doi.org/10.4067/S0718-18762014000200008

Karasti, Helena; **Baker, Karen S.** (2008). "Digital data practices and the long term ecological research program growing global". *International journal of digital curation*, v. 3, n. 2, pp. 42-58.
https://doi.org/10.2218/ijdc.v3i2.57

Kassen, Maxat (2013). "A promising phenomenon of open data: A case study of the Chicago open data project". *Government information quarterly*, v. 30, n. 4, pp. 508-513.
https://doi.org/10.1016/j.giq.2013.05.012

LaValle, Steve; **Lesser, Eric**; **Shockley, Rebecca**; **Hopkins, Michael S.**; **Kruschwitz, Nina** (2011). "Big data, analytics and the path from insights to value". *MIT Sloan management review*, v. 52, n. 2, pp. 21-32.
https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value

Martínez-Méndez, Francisco-Javier; **López-Carreño, Rosana**; **Pastor-Sánchez, Juan-Antonio** (2014). "Open data en las administraciones públicas españolas: categorías temáticas y apps". *El profesional de la información*, v. 23, n. 4, pp. 415-423.
https://doi.org/10.3145/epi.2014.jul.10

Open data handbook (2017).
http://opendatahandbook.org/guide/en/what-is-open-data

PIRA International (2000). *Commercial exploitation of Europe's public sector information*. European Commission, Directorate General for the Information Society, Oct. PIRA, 2000.
https://ec.europa.eu/digital-single-market/en/news/commercial-exploitation-europes-public-sector-information-pira-study-full-report

Solar, Mauricio; **Daniels, Fernando**; **López, Roberto**; **Meijueiro, Luis** (2014). "A model to guide the open government data implementation in public agencies". *Journal of universal computer science*, v. 20, n. 11, pp. 1564-1582.
http://bit.ly/2ejRBmG

Spain (2007). "Ley 37/2007, de 16 de noviembre sobre reutilización de la información del sector público". *Boletín oficial del estado*, n. 276, 17 de noviembre, pp. 47160-47165.
http://www.boe.es/buscar/doc.php?id=BOE-A-2007-19814

Susha, Iryna; **Grönlund, Åke**; **Janssen, Marijn** (2015). "Driving factors of service innovation using open government data: An exploratory study of entrepreneurs in two countries". *Information polity*, v. 20, n. 1, pp. 19-34.
https://doi.org/10.3233/IP-150353

Vicente-Paños, Adrián; **Jordán-Alfonso, Aurea** (2017). "Acceso a la información pública y su reutilización en las comunidades autónomas: evaluación de la reutilización de datos abiertos". *El profesional de la información*, v. 26, n. 3, pp. 381-391.
https://doi.org/10.3145/epi.2017.may.04

Vickery, Graham (2011). *Review of recent studies on PSI reuse and related market developments*. Information Economics, Paris.
https://ec.europa.eu/digital-single-market/en/news/review-recent-studies-psi-reuse-and-related-market-developments

Yoon, Ayoung (2017). "Data reusers' trust development". *Journal of the Association for Information Science and Technology*, v. 68, n. 4, pp. 946-956.
https://doi.org/10.1002/asi.23730

Zhao, Mengnan; **Yan, Erjia**; **Li, Kai** (2018). "Data set mentions and citations: A content analysis of full-text publications". *Journal of the Association for Information Science and Technology*, v. 69, n. 1, pp. 32-46.
https://doi.org/10.1002/asi.23919

Zimmerman, Ann S. (2008). "New knowledge from old data. The role of standards in the sharing and reuse of ecological data". *Science, technology, & human values*, v. 33, n. 5, pp. 631-652.
https://doi.org/10.1177/0162243907306704

Zuiderwijk, Anneke; **Janssen, Marijn**; **Davis, Chis** (2014). "Innovation with open data: Essential elements of open data ecosystems". *Information polity*, v. 19, n. 1-2, pp. 17-33.
https://doi.org/10.3233/IP-140329

Annex

1. To what extent do you think that these variables can be determining factors of the reusability of the data (scale 1: less important to 7 more important).

Legal licensing of datasets
Accessibility. How easy is to find and retrieve the data
Technical standard
Data model
Geographic content
Updating frequency
Topic of the datasets contents
Reliability. If the dataset meets the announced updating frequency
Support to use. If there is a contact to make questions about the dataset
API availability if there is a documented API to describe access to the datasets
SEO friendly. If the dataset is located in a web site friendly to be indexed by search engines, i.e. *Google*

2. What other factors can be included?

3. In *Meloda 4* version legal licensing is divided into 5 levels categorising the datasets in terms of is licensing. Do you agree with the existing levels? (keep, delete, do not know / do not answer). Justify your answer.

4. In *Meloda 4* version access to datasets' dimension is divided into 5 levels (categories). Do you agree with the existing levels? (keep, delete, do not know / do not answer). Justify your answer.

5. In *Meloda 4* version technical standards to store the data are classified into 4 categories depending on their openness. Do you agree with the existing levels? (keep, delete, do not know / do not answer). Justify your answer.

6. In *Meloda 4* version data model dimension is classified into 5 levels depending on their dissemination and popularity. Do you agree with the existing levels? (keep, delete, do not know / do not answer). Justify your answer.

7. In *Meloda 4* version geolocation content of the datasets is classified into 5 levels. Do you agree with the existing levels? (keep, delete, do not know / do not answer). Justify your answer.

8. In *Meloda 4* version updating frequency of data is classified into 5 levels depending on the period. Do you agree with the existing levels? (keep, delete, do not know / do not answer). Justify your answer.

9. Choose what classification of datasets by topic should be used: According to *NTI-RISP* Spanish regulation; according to *DCAT-AP*, EU regulation or others. Justify your answer.

10. Other questions. Answer yes or no.
- The number of tags condition the demand of datasets?
- Should the five stars of Berners-Lee be considered as a factor to assess the demand of an open data dataset?
- Should the number of distributions of a dataset be considered as a factor that conditions the demand?
- Can data downloads be an indicator of the degree of reuse of open data?