A User-Centered Model for Assessing and Improving Open Government Data Quality

Meryam Belhiah and Bouchaïb Bounabat

AL QualSADI Team-ENSIAS, Rabat IT Center, Mohammed V University in Rabat
meryam.belhiah@um5s.net.ma, bounabat@ensias.ma

Abstract: Growing open government data (OGD) initiatives are offering increased monetary and non-monetary benefits for various stakeholders, including governments, corporations, tech startups, civil society organizations and citizens. These benefits include: conforming to regulatory-driven compliance, increased transparency, increased commercial and social opportunities for innovation and growth.

A pluralist network of actors around the world is working to expand the availability of open government data by establishing the legal foundations and leveraging the technical capacity of public departments and agencies in different countries. Regardless of the numerous initiatives, the vast majority is focused on assessing readiness and implementation, in terms of legal and technological aspects, and only a few are providing assistance regarding data quality (DQ) aspects. However, inaccurate, incomplete and not up-to-date data are some of the most important challenges facing end users.

This paper presents an approach that combines data quality measurement (DQM) and recommender systems (RS) to provide suggestions of items (datasets) that may represent a potential interest for citizens for leveraging the value of open datasets, as well as planning data quality improvement actions that are cost-effective and have a highly positive impact.

Keywords: Open Government Data, Data Quality Assessment and Improvement, Cost/Benefit Analysis

INTRODUCTION

Open government data refers to data that is produced or commissioned by government or government controlled entities[1] and which is also open, according to the Open Definition[2]; that is, it should be accessible via the Web at no other cost than the cost of reuse and with no restrictions regarding the identity of the end user and its intention. Moreover, this data should be available in a digitized form.

The increasing attention paid to the open government data is motivated by the urge to conform to legislative constraints and to enhance the transparency and accountability in governmental actions, for instance, the public expenditure and revenue, public procurement contracts and election results, to mention but a few. Another leading factor is the need to extract business and social value from the open government data by their reuse by third parties (MEPSIR 2006; Vikery 2011; European commission 2015; Alexopoulos et al. 2014).

---

[1] The Working Group on Open Government Data. 2017. https://opengovernmentdata.org/

[2] Open Knowledge Foundation Network. The Open Definition version 2.1. http://opendefinition.org/od/2.1/en/

The problem of assessing readiness (The World Bank 2016) as well as the evaluation of implementation (Bogdanović-Dinić et al. 2014; Ceolin et al. 2013; Reiche et al. 2013; Harper 2012), has been given substantial attention in the litterature, both in research and industrial areas (Tim Burners-Lee 2010). While assessing readiness looks at whether the preconditions exist to set a successful open data project, evaluating implementation tries to answer the question of how good the current implementation is against a set of criterea including, accessibility, data format and open format, among others.

Even thougth the work cited above establishes the overall methodology for assessing the readiness and implementation of open government data initiatives; it only evaluates some aspects of data quality from the data provider's point of view. In fact, it lacks feedback from the general public in terms of quality, considering that the success of an OGD initiative could not uniquely be measured by the quantity of data that is released but by the use which is made of it.

As open government data is primarily intended to be accessed and reused by external actors to the entity that produces the data, the puspose of this paper is to put forward a new approach that combines recommender systems and feedback from end users to provide suggestions of items (datasets) that may represent a potential interest for them, and also to direct data quality improvement actions toward data that matters the most to end users and that lacks an acceptable level of quality. Accordingly, this will allow governments to monitor the DQ level of the most used datasets and program improvement actions that will have the greatest benefit-cost ratio. Therefore, the assessment is addressed to evaluate both the supply and the demand side of open data.

The organization of this paper is as follows: section 2 presents a definition of data quality dimensions, a benchmark of existing frameworks for open government data evaluation as well as open government data lifecycle. Section 3 describes the main steps of our approach. In section 4, the discussion and future work are summarized.

RELATED WORK

Data quality: definition and evaluation

Data quality may be defined as "the degree to which information consistently meets the requirements and expectations of all knowledge workers who require it to perform their processes" (IAIDQ 2015), which is epitomized in the expression "fitness for use" (Wang et al. 1996).

Building on the perspective of the people using the data (Belhiah et al. 2015), many researchers have tried to establish a classification for data quality dimensions. Below, Pipino et al. have identified 15 dimensions (Pipino et al. 2002):

- Intrinsic: accuracy, believability, reputation, and objectivity;
- Contextual: value-added, relevance, completeness, timeliness, and appropriate amount;
- Representational: understandability, interpretability, concise representation, accessibility, ease of operations, and security.

All case studies that aimed at assessing and improving data quality have chosen a subset of data quality dimensions, depending on the objectives of the study (Batini et al. 2012), (Narman et al. 2009), (Aladwani et al. 2002), (Catarci & Scannapieco 2002), and (Haug et al. 2011). Measurable metrics were then defined to score each dimension.

This is particularly true for open government data, as the poor quality of data may hinder these programs' objectives, by making public scrutiny to a great extent impossible.

Many frameworks have been designed in order to assess the readiness and the implementation of open government data initiatives. These assessments are intended to cover open government data projects from the earliest to the latest stages, including the design, launch and deployment phases. For each framework, a set of characteristics is chosen to be evaluated.

The next section summarizes the key aspects that are covered by the most prominent open government data evaluation framework, both in research and industrial areas.

Existing frameworks for open government data evaluation

The following tables describe the principles that define openness in relation to metadata and content, for the commonly used international standards.

We will focus on international standards (Table 1) and academic research (Table 2) that evaluate and assess the implementation of open government data projects.

Table 1: Overview of the Different International Standards of Open Government Data Evaluation

Standards	**Aspects of Open Data that are Covered**	**Quality of Data**	**Quality of Metadata**
opendefinition.org[3]	• Open license or status • Accessibility • Machine readability • Open format	x	x
5-star deployment scheme of open data (Tim Burners-Lee 2010)	• Availability		x
Sunlight principles for opening up government information[4]	• Completeness • Primacy • Timeliness • Ease of physical and electronic access • Machine readability • Non-discriminatory • Use of commonly owned standards • Licensing • Permanency • Free of charge	x	x
Data Catalog Vocabulary (DCAT)[5]	• Discoverability • Uniqueness		x
The ODI Open Data Certificate[6]	• Availability		x

[3] opendefinition.org
[4] https://sunlightfoundation.com/opendataguidelines/
[5] https://www.w3.org/TR/vocab-dcat/
[6] https://certificates.theodi.org/en

	• Open license or status • Timeliness		
Global Open Data Index (OKFN)[7]	• Open license or status • Machine readability • Free of charge • Primacy • Timeliness • Availability • Open format	x	x

Table 2: Overview of the Different Approaches in Research Area of Open Government Data Evaluation

Open data evaluation	**Aspects of Open Data that are Covered**	**Quality of Data**	**Quality of Metadata**
Reliability Analyses of Open Government Data (Ceolin et al. 2013)	• Reliability		x
How open are public government data? An assessment of seven open data portals (Bogdanović-Dinić et al. 2014)	• Completeness • Primacy • Timeliness • Accessibility • Machine readability • Non-discriminatory • Non-proprietary • License-free	x	x
Implementation of Metadata Quality Metrics and Application on Public Government Data (Reiche et al. 2013)	Metadata quality in terms of: • Completeness • Weighted completeness • Accuracy • Richness of information • Accessibility		x
Grading the Government's Data Publication Practices (Harper 2012)	• Authoritative sourcing (reputation) • Availability • Machine-discoverability • Machine-readability		x

[7] http://index.okfn.org/

We have undertaken an analysis of open data that is published by the Moroccan Government via its official portal[8]. According to the 2015 Global Open Data Index[9], Morocco ranks among the world's best countries in the category "Location datasets". The minimum data that should be available online is: zip codes, addresses and geographical coordinates at a national level.

However the published datasets are incomplete as they only contain zip codes and addresses; geographical coordinates are therefore missing. Also the published datasets cover only the 24 major cities of Morocco. The smaller cities and local communities were omitted. When it comes to timeliness, data should be updated once a year, according to the Global Open Data Index specifications. However, the 2015 survey points to 2011 data that is not up-to-date, owing to the fact that the Moroccan national administrative map has been changed since then. Lastly, the data is accurate as it exactly matches another authoritative data source. The figure below summarizes our findings:

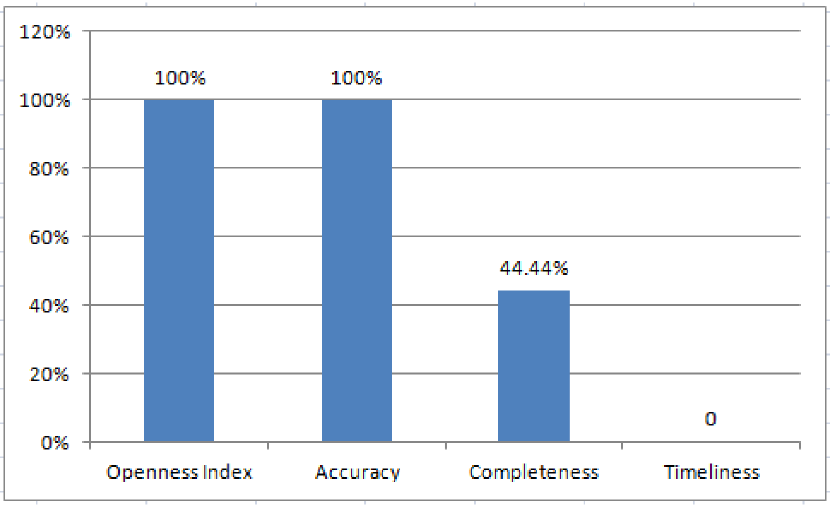

Figure 1: Openness Index and Data Quality Levels Comparison

This example shows how data openness indicators inform very little about the quality of published datasets. Even if a dataset is 100% open, it may still be inaccurate, incomplete and not up-to-date.

Open government data life cycle

Research in the area of data quality has shown that "poor data quality is a primary reason for 40% of all business initiatives failing to achieve their targeted benefits" and that "data quality effects overall labor productivity by as much as a 20%" (Gartner 2011). Poor data quality affects also downstream analysis and end users satisfaction.

As such, data producers have to evaluate different scenarios related to data quality projects to be considered. However, there is no general agreement on which set of criteria define the scenario with the highest level of contribution.

Prior to introducing our model in the next section, it is suitable to present the common phases of open government data life cycle:

[8] http://data.gov.ma/fr
[9] http://2015.index.okfn.org/place/morocco/

Figure 2: Open Government Data Life Cycle

Creating/Collecting Data – The open government data lifecycle commonly starts with this phase. Data that is intended to be published may be already available in the government entity Information System (IS), as part of its daily activities. Data may however be gathered from external sources for the purpose of publishing it.

Processing Data – This step consists of data selection and data harmonization. Indeed, data to be published should be selected and pass readiness assessments.

Analyzing Data – This step consists of adding value to data, using other data sources.

Publishing Data – This step requires giving access to data from external locations, such as open government data portals, public datasets or technical reports.

Using/Re-using Data – This step corresponds to the time-lapse where data is made available to the general public for use and integration. This is when economic and social value is extracted and created from data.

Improving Data – Data quality issues may arise, which will require designing and implementing improvement solutions on data and processes to meet requirements regarding the quality of data.

Validating and Monitoring Data Quality Levels – This step consists of defining thresholds for data quality acceptability. The appropriate actor will be notified in case of failure to comply with these levels, in order to remedy the situation.

Our field of intervention covers the last two phases.

MODEL FOR ASSESSING AND IMPROVING OPEN GOVERNMENT DATA QUALITY

Evaluating dimensions of data quality

After examining the existing standards and academic approaches for open data evaluation, we have developed a data quality assessment and improvement model that perceives the quality of open government data through the following indicators: accuracy, completeness, and timeliness.

While it is difficult to agree on the dimensions that will determine the data quality, it is however possible, when taking users' perspective into account, to be confined to a small number of dimensions, including, accuracy, completeness, and timeliness. These indicators were determined after performing a subjective assessment among users (citizen, developers, and integrators) of open datasets at the end of a hackaton, that was held in Rabat, Morocco (using a questionnaire). These indicators were the ones that present issues for end users. Feedback from an ODG producer corroborate these results.

The table below lists the most highlighted issues and links them to the correponding data quality dimensions:

Table 3: Data Quality Aspects Related Issues

Issues	**Data Quality Dimensions**
• Is there any duplicate data? • Is the data normalized, allowing syntactic validation?	Accuracy – *Does the data have an acceptable margin of error?*
• Is there any missing data? • Are cells complete? • Are rows complete?	Completeness – *Is the data complete?*
• Does the actual time of data delivery correspond to the documented time delivery? • Are there any delays in publication? • Is the data obsolete? • Is the data readily available (published online as soon as it is available)?	Timeliness – *Is the data current?*

Then, our approach allows users to rate the quality of datasets, in terms of: accuracy, completeness, and timeliness. The feedback from these users will enable us to:

- recommend the datasets that have the best ratings in terms of data quality;
- provide feedback about the most used datasets and an insight into their quality issues. This feedback will allow the OGD providers to plan data quality improvement with the most benefit/cost ratio.

We suppose nevertheless, that these datasets are accessible and available, according to the basic principles of open data. The schema below describes the iterations of our model:

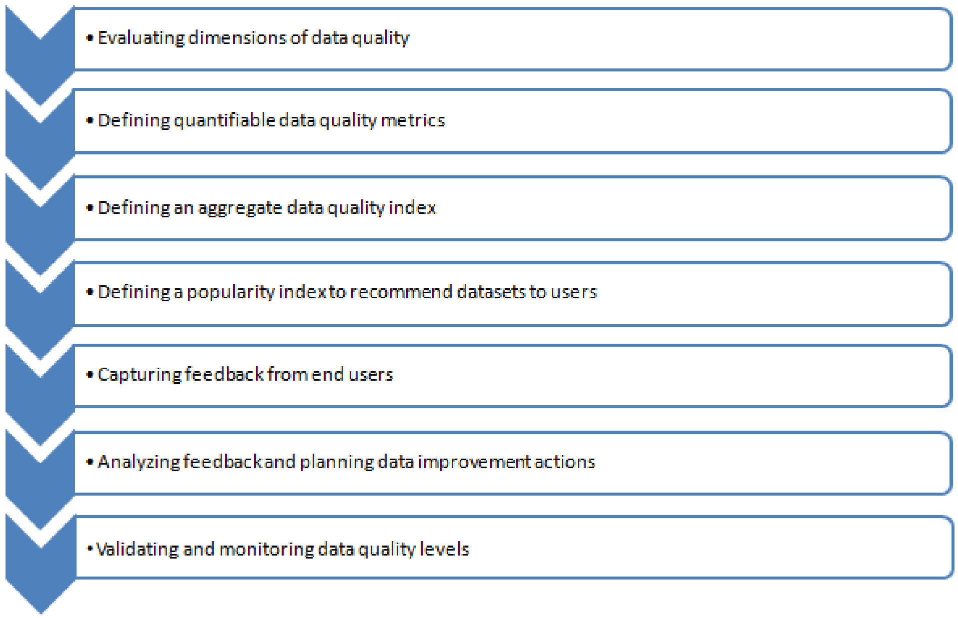

Figure 3: Model for Improving and Assessing Open Government Data Quality

Defining quantifiable data quality metrics

Accuracy is defined as “the closeness of results of observations to the true values or values accepted as being true” (Pipino et al. 2002). Wang et al. (1996) define accuracy as “the extent to which data are correct, reliable and certified”.

Completeness specifies how “data is not missing and is sufficient to the task at hand” (Batini & Scannapieco 2006). As completeness has often to deal with the meaning of null values, it may be expressed in terms of the“ ratio between the number of non-null values in a source and the size of the universal relation” (Naumann 2002).

Timeliness is a time-related dimension. It expresses “how current data are for the task at hand” (Batini and Scannapieco 2006). In the context of open data, it could also be defined as the comparison between the actual time of data delivery against the documented time range. As a matter of fact, even if a piece of data is accurate and complete, it may be useless if not up-to-date.

When performing an objective assessment, OGD publishers need to develop metrics that are specific to their needs. For each of the dimensions mentionned above, we use a simple ratio calculation method to measure these dimensions.

The indicators of the DQ model for OGD are presented in Table 4 below along with a brief description of their structure and grading as well as their functional form.

Table 4: Model Measurement Indicators

Dimension	Nature of the dimension	Functional form	Indicator score	Score
Accuracy	Intrinsic	$\frac{number\ of\ accurate\ values}{Total\ number\ of\ all\ values}$	Percentage	(0,1)
Completeness	• Dataset dependent • Domain dependent	$\frac{number\ of\ non-null\ values}{Total\ number\ of\ all\ values}$	Percentage	(0,1)
Timeliness	• Dataset dependent • Domain dependent	$\frac{number\ of\ values\ that\ are\ up-to-date}{Total\ number\ of\ all\ values}$	Percentage	(0,1)

Defining an aggregate data quality index

Due to specific aspects of each domain, and in order to provide a generic approach that can be implemented without any adjustment, the second step in our approach consists of defining a single index of data quality that aggregates these measures, with a weighing coefficient that does not cause a bias in the interpretation.

In a multivariate context, the provider should have a good understanding of the importance of each dimension depending on its domain of activity, to be able to define its contribution to the overall quality index.

The purpose behind using a weighing coefficient is to allow each organization to express the importance of each aspect of data quality, depending on its environment and strategy. A few examples where using different weighting coefficients is relevant are as follows:

- Producers of datasets that define master data (ex. geographical datasets) that are not updated frequently may give a greater importance to accuracy and completeness, rather than timeliness;
- Public organizations that produce financial datasets may give equal weighting coefficients to accuracy and timeliness due to the volatile nature of data;
- Other public organizations may give the same weight to all the factors above.

Using weighting coefficients that express the importance of each DQ aspect, one may define a unique index for data quality:

$$\text{DQ index} = \sum_{i=1}^{3}(Ri * wi)/100 \quad \text{subject to: } R_i > \text{threshold} \tag{1}$$

Where R_i is the rating for the DQ dimension “i” and w_i is the weighing coefficient that is associated with the dimension “i”, that was previously defined by the OGD provider. The obtained score ranges between 0 and 1, where “0” refers to “no quality” and “1” refers to “high quality”.

For “timeliness”, we, however, suggest incorporating a sensitivity parameter that takes into consideration how data becomes less timely faster or loses timeliness at a lower rate (Pipino et al. 2002).

Defining a popularity index to recommend datasets to users

When accessing an open government data portal, the user is confronted with a plethora of choices and a large number of datasets. For instance, UK open data portal has 1410 publishers to its credit. As our approach is user-centered, it allows users to filter information in order to spot the datasets that provide the best match to their needs, in terms of usefulness and quality.

As such, recommender systems (RS) are an Artificial Intelligence (AI) technology that is successfully applied in different e-commerce contexts to recommend the best options when it comes to books, movies and places, transforming how people shop (e.g., Ebay.com, Amazon.com, Netflix.com, Tripadvisor.com, etc.), which makes it possibly applicable in the context of open government data portals. RS provide suggestions of items that may represent a potential interest for users (Burke 2010).

In our context, two measures are associated with each dataset: popularity and quality index.

Popularity Index: for each dataset, the popularity index is defined, using one or a combination of these measures:

- number of views;
- number of downloads;
- number of reuses;

- data quality index that will be defined later in this section.

Many data portals have some of the first three metrics among datasets meta-data definition. Nevertheless, they do not use them in a sophisticated manner to recommend datasets to users.

Table 5: Popularity Metrics as Definied by 5 Open Government Data Portals

Country	**URL**	**Global Open Data Index 2016**	**Popularity Metrics**
Taiwan	data.gov.tw	1	• Number of views • Number of downloads
United Kingdom	data.gov.uk	2	• Number of views
Denmark	opendata.dk	3	None
Finland	data.gouv.fi	5	• Number of views
France	data.gouv.fr	10	• Number of reuse

As data portals have easy access to these metrics, incorporating such a popularity index requires minor adjustments. Incidentally, we could incorporate contextual data into the recommendation process by using the location of the requests and previous search topics.

It is important to note that these filtering techniques preserve personal data of users, as they do not require any personal information other that navigation history.

Quality Index: During the early ramp-up phase, the quality index is not known yet. It has to be calculated after feedback is received through ratings from end users (see step 5).

These two metrics will be attached to datasets and displayed as annotations. The popularity index, which is a function of data quality index, will control the listing and the display of datasets.

Capturing feedback from end users

Our model gives an insight for understanding the demand side of open government data, so data quality improvement actions will be aligned with those needs.

Users express their opinion of experiences when accessing data portals or datasets. Ratings are expressed on a scale of 1 to 5, 5 being the best. This can be performed using a simple 5-star rating system:

Table 6: Data Quality Levels

Rating	Quality level
★★★★★	very high quality
★★★★	high quality
★★★	moderate quality
★★	slight quality
★	no quality

Each aspect of data quality, namely accuracy, completeness and timeliness should be rated separately. In a further step, the aggregate data quality index, as defined in step 3 will be calculated and linked to the corresponding dataset.

In the ranking algorithm, older reviews may be given a lower weight, especially for datasets that are time-sensitive (transactional data v/s master data). The number of reviews is also a factor. Therefore, even with all else being equal, something that has a hundred associated reviews may have a different score than something with only one review.

Consequently, the overall quality index is a summary score based on quality, quantity and recency of ratings.

Analyzing feedback and planning data improvement actions based on the feedback

After collecting usage frequency and quality indicators about published datasets usage (popularity) and their perceived data quality by end users, datasets can be classified by priority and divided into categories, in order to plan actions to improve data quality. In fact, investment in terms of time, human and financial resources will be directed towards cleaning datasets that are used the most by users and that have data quality issues. This way, we spot data quality options with the greatest business value at the least-cost.

Because business processes access data objects in reading and/or writing modes, it is normal that the quality of the data has an impact on the result of business processes' execution and vice-versa. With this in mind, two business cases may be considered:

The first one is based on the improvement of data quality by determining and analyzing the sources of low quality, such as uncontrolled data acquisition, updating problems, etc. and then eliminating the source of identified data quality issues.

The second one is to improve the processes (reengineering, control, etc.), by enhancing their execution accuracy. This is a short-term option that is generally less expensive, but requires change management because it affects the work processes. In fact, while technology plays a key role in data quality improvement, changes in working methods are critical.

Validating and monitoring data quality levels

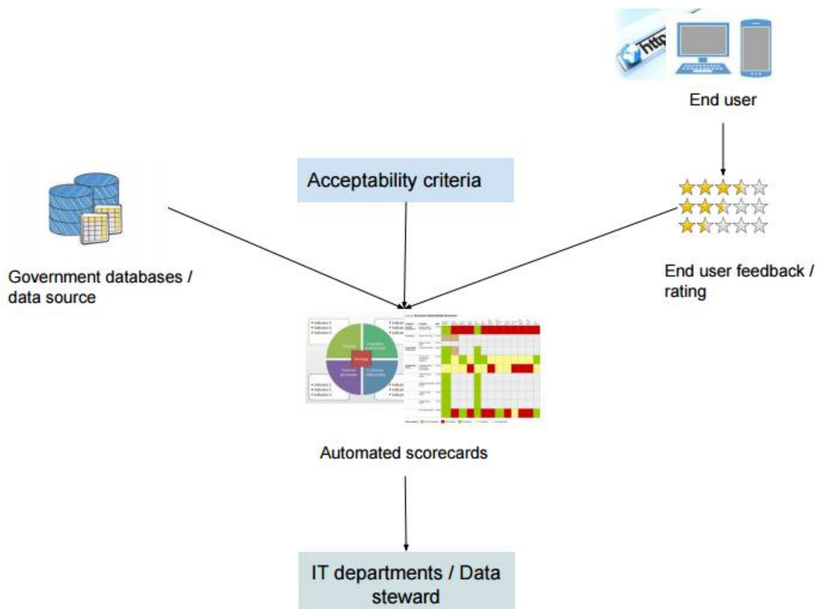

Figure 4: Automating the Collection and Reporting of Data Quality Levels

This step consists of defining thresholds for data quality acceptability: the degree of acceptability becomes the new conformance criterion against which data quality is measured. Hence, the quality levels for accuracy, completeness and timeliness of published datasets, should conform to end users expectations. The appropriate contact point, whether it is an IT department or a data steward, will be notified in case of failure to comply with these levels, in order to remedy the situation.

Automated scorecards allow notifying people when acceptability thresholds are not met. The figure above (Figure 4) summarizes the validation and monitoring data quality levels by automating the collection and reporting of data quality levels:

The following figure summarizes the architecture of our model:

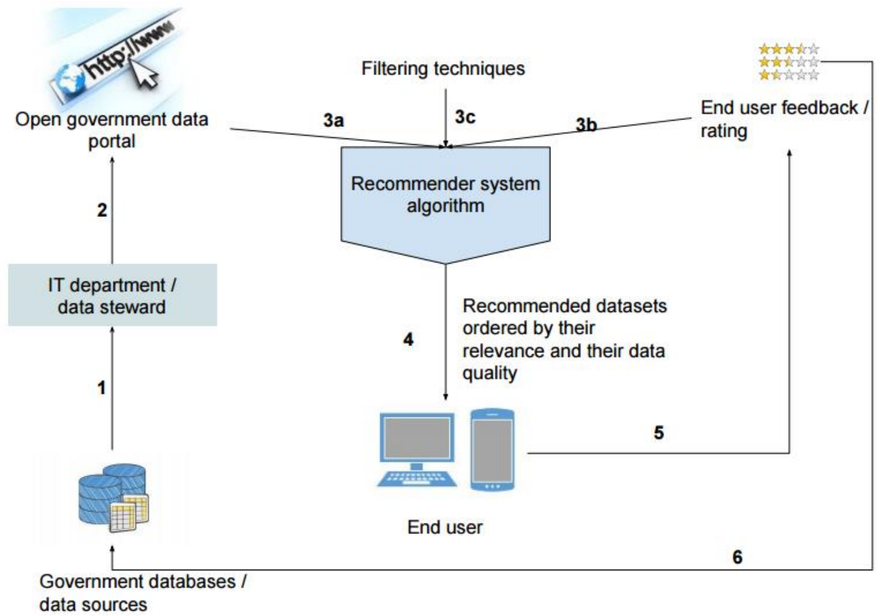

Figure 5: The Model Architecture

CONCLUSION

As public departments and agencies recognize information to be among their most valuable assets and that poor data quality has a significant impact on their digital economy plans, the demand for data quality assessment and improvement frameworks is maturing. Managing quality of open data adds yet another level of complexity to an already high demanding activity in the context of proprietary data. In actual fact, at the moment of publishing vast volumes of data, public organizations have no knowledge about how this information will be used, and for what purpose, what is the data that is the most valuable for end users, and what are the quality levels that are required to fully benefit from its potential?

Since governments are periodically collecting, cleaning, transforming and releasing gigabytes of raw data, actions must be directed toward carefully planning data quality initiatives that are cost-effective and that will have the most valuable contribution to end users. This guidance is particularly crucial in the context of open data as governments have no or little information about what data matter the most to end users. Our approach highlights the most cost-effective data improvement actions for open government data.

We have established a global indicator of data quality for datasets. This indicator aggregates data quality levels of accuracy, completeness and timeliness as expressed by and collected from end users. When used as a meta-data and integrated to open government data portals, this indicator enables end users, to identify trustworthy data. Also, emerging data issues are pushed upstream from end users to open government providers. When coupled with gathered information about datasets usage frequency, it allows data providers to plan improvement actions accordingly. The last step of our approach suggests putting practices in place in order to set acceptable thresholds for each aspect of data quality and implement automatic scorecards to monitor data quality levels. Thus, assessing and improving data quality is dictated by the users' needs and becomes a continuous practice in order to deliver high-quality data and provide confidence that end users can take advantage of its social and economical value.

The result of the work accomplished thus far shows how to measure in a quantitative manner, data quality levels as experienced by end users of open governments' datasets, by establishing an aggregate index of data quality.

It is now challenging to see how our model performs in a real environment. We are particularly interested in applying it to the context of Open Data in Morocco. Article 27 of Morocco's amended Constitution of 2011 enshrines the citizens' right to "access to information held by the public administration, the elected institutions and the organs invested with missions of public service" (Constituteproject.org 2012). It therefore follows that Morocco has its open data portal[10] since 2011. In the example mentioned in section 2, disparate data quality levels have been identified. Therefore, in an upcoming communication, we would like to highlight how putting our model in place could facilitate building trustworthy data.

REFERENCES

Aladwani, A. M., & Palvia, P. C. 2002. "Developing and validating an instrument for measuring user-perceived web quality," Information & management, 39(6), pp. 467-476.

Alexopoulos, C., Spiliotopoulou, L., & Charalabidis, Y. 2013. "Open data movement in Greece: a case study on open government data sources," in Proceedings of the 17th Panhellenic Conference on Informatics. ACM, pp. 279-286.

Batini, C., Comerio, M., & Viscusi, G. 2012. "Managing quality of large set of conceptual schemas in public administration: Methods and experiences," in Model and Data Engineering. Springer Berlin Heidelberg, pp. 31-42.

Belhiah, M., Bounabat, B., & Achchab, S. 2015. "The impact of data accuracy on user-perceived business service's quality," in Information Systems and Technologies (CISTI), 2015 10th Iberian Conference. IEEE. pp. 1-4.

Berners-Lee, T., Cailliau, R., Groff, J. F., & Pollermann, B. 2010. "World-wide web: The information universe," Internet Research, 20(4), pp. 461-471.

Bogdanović-Dinić, S., Veljković, N., & Stoimenov, L. 2014. "How open are public government data? An assessment of seven open data portals," in Measuring E-government Efficiency. Springer New York, pp. 25-44.

Burke, R. 2010. "Evaluating the dynamic properties of recommendation algorithms," in Proceedings of the fourth ACM conference on Recommender systems. ACM, pp. 225-228.

Catarci, T., &.Scannapieco, M. 2002. "Data quality under a computer science perspective," Archivi & Computer, 2, pp. 1-15.

Ceolin, D., et al. 2013. "Reliability analyses of open government data," in Proceedings of the 9th International Conference on Uncertainty Reasoning for the Semantic Web-Volume 1073 CEUR-WS.org, pp. 34-39.

[10] http://data.gov.ma/

Dekkers, M., Polman, F., te Velde, R., & de Vries, M. 2006. "Mepsir: Measuring European public sector information resources. Final report of study on exploitation of public sector information," Brussels, Helm Group of Companies.

European commission. 2015. "Towards faster implementation and uptake of open government," https://ec.europa.eu/digital-single-market/en/news/final-report-study-towards-faster-implementation-and-uptake-open-government-smart-20150041

Haug, A., Zachariassen, F., & Van Liempd, D. 2011. "The costs of poor data quality," Journal of Industrial Engineering and Management, 4(2), pp.168-193.

International Association for Information and Data Quality. 2015. IQ/DQ glossary. http://iaidq.org/main/glossary.shtml

J. Harper. 2012. "Grading the Government's Data Publication Practices," Cato Institute Policy Analysis, vol. 711.

Narman, P., Johnson, P., Ekstedt, M., Chenine, M., & Konig, J. 2009. "Enterprise architecture analysis for data accuracy assessments," in Enterprise Distributed Object Computing Conference, 2009. EDOC'09. IEEE International. IEEE, pp. 24-33.

Naumann, F. 2002. "Quality-driven query answering for integrated information systems," (Vol. 2261). Springer Science & Business Media.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. 2002. "Data quality assessment," Communications of the ACM, 45(4), pp.211-218.

Reiche, K. J., & Höfig, E. 2013. "Implementation of metadata quality metrics and application on public government data," in Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE 37th Annual, IEEE. pp. 236-241.

The Gartner Group. 2011. "Measuring the Business Value of Data Quality," https://www.data.com/export/sites/data/common/assets/pdf/DS_Gartner.pdf

The World Bank. Open Government Data Toolkit. 2016. http://opendatatoolkit.worldbank.org/en/

Vickery, G. 2011. "Review of recent studies on PSI re-use and related market developments," Information Economics, Paris.

Wang, R. Y., & Strong, D. M. 1996. "Beyond accuracy: What data quality means to data consumers," Journal of management information systems, pp.5-33.

Constitute Project. English Translation by William S. Hein & Co., Inc. 2012. https://www.constituteproject.org/constitution/Morocco_2011.pdf?lang=en