

Italian National Guidelines for the Valorization of the Public Sector Information

Gabriele Ciasullo, Giorgia Lodi, Antonio Maccioni, Francesco Tortorelli
{ciasullo, giorgia.lodi, antonio.maccioni, tortorelli}@agid.gov.it

Agency for Digital Italy

ABSTRACT

The Italian legislation for public sector digitalization has defined a cyclic process for the valorization of the Public Sector Information. This process involves three main elements, namely, the definition of a strategic agenda that identifies principles and objectives to be achieved by public administrations in valorizing the information they own and manage; a set of technical guidelines that contain recommendations that administrations should follow in order to meet the objectives indicated in the agenda; and a report including the principal results of an assessment of how well the objectives have been met by administrations.

The legislation assigns to the Agency for Digital Italy (AgID) the role of national body responsible for governing the life-cycle of the process mentioned above.

This paper discusses how AgID currently manages the process. In particular, the paper illustrates the principal recommendations included in the technical guidelines with a focus on the metadata management and the business models that can be enabled in Open Data initiatives. We advocate that the guidelines allow for supporting administrations in the creation of a uniform national single data market, as stimulated by the European Council.

1. INTRODUCTION

The Pillar I of the Europe 2020 Digital Agenda introduces the concept of digital single market to be the instrument through which generate growth, new jobs and prosperity. The digital market relies on the establishment of an European data ecosystem where information and knowledge are shared among different players (i.e. companies, entrepreneurs, data professionals, venture capitalists, universities, etc.). The basis of such a market is an efficient management of data owned and produced by public administrations. To meet this goal, some directives at over national level (EU PSI directive, G8 Open Data Charter, just to name a few) have been proposed.

To embrace local peculiarities, governments often introduce national strategies to govern their Public Sector Information (PSI). In Italy, following the European recommendations, the Law Decree 179/2012, that modified the National Digital Administration Code (Legislative decree n. 82/2005), defined a national framework for the valorization of PSI. The framework consists of three strategic documents to be produced; namely, an *agenda* [6], a set of technical *guidelines* [7] and a monitoring *report*. The three documents involve the definition of a uniform process at the national level for the production and publication of Open Data in the Italian public sector.

More in detail, the agenda identifies principles (e.g., interoperability, “Open Data by default”, usability, accessibility, etc.), objectives and actions to be carried out by PAs within a year in order to implement and sustain in the long term an interoperable Open Data process.

The technical guidelines include legal, organizational, economical and technical recommendations that PAs should comply with in order to fulfill effectively the objectives of the agenda. They comprise (i) an organizational model to follow in order to set up internal processes for the production and publication of interoperable Open Data sets: (ii) standards, data formats, ontologies and vocabularies to use as well as datasets already available on the Web of Data to link; (iii)

business models that can be enabled; (iv) legal aspects concerning the licenses associated with open data sets; and finally, (v) technical recommendations concerning infrastructural aspects of open data portals to implement (e.g., URI policies). These technical guidelines represent the natural update of previously published guidelines on “Semantic Interoperability through Linked Open Data (LOD)” [4], which identify and suggest Linked Data as the paradigm to use in order to solve, at the same time, interoperability, homogeneity and understandability issues affecting Open Data published by administrations.

Finally, the monitoring report includes the principal results of an assessment of the actual state of the Italian PSI and its improvements (or, in case, worsenings) with respect to the previous year.

The same Decree mentioned above appoints the Agency for Digital Italy (AgID) as the Italian PSI enabler. This means that it is AgID that annually proposes the agenda, defines the technical guidelines in collaboration with the Italian Public Administrations and, at the end of the year, evaluates how well the strategic policies are adopted by the administrations.

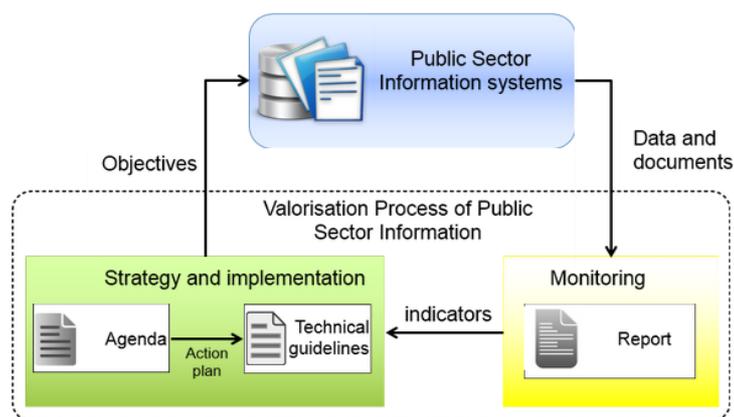


Figure 1: the Italian PSI enhancing process

The PSI valorization process is illustrated in Figure 1, where the PSI systems (the box on the top) are controlled by two logical components (the boxes on the bottom). The first component includes the “strategy and implementation”, which are represented by the “agenda” and the “technical guidelines”, respectively.

The second component is represented by a “monitoring” phase that involves the analysis of the current state of the PSI. This phase produces as output the annual report as discussed above.

The whole process can be viewed as a system (i.e., the PSI) regulated by a control-loop mechanism (i.e., the monitoring and the strategy and implementation components). This highlights the strategic role of the report that is used for the tuning of the forthcoming agenda and technical guidelines. These latter take a set of indicators, defined during the monitoring phase, as input and produce in output the objectives and the recommendations, which in turn are inputs for the PSI system managed by the public administrations.

This paper first illustrates in Section 2 the content of the technical guidelines with more focus on the model proposed for metadata management. Then, in Section 3 it discusses the recommendations on the business models that are included in the technical guidelines. Section 4 presents the conclusions identifying also possible future works.

2. THE TECHNICAL GUIDELINES

At the end of 2012, the Agency for Digital Italy (AgID) published national guidelines that paved the way to the use of LOD as the data paradigm for enabling semantic interoperability among PAs [4]. Since then, AgID continued to exercise its role of national PSI enabler by annually releasing national technical guidelines for the valorization of the PSI, as explained in the previous section. The technical guidelines include, among the others, several novel elements: (i) suggestions on the schema and ontologies for publishing data and in particular key datasets (e.g., official classifications

to foster an effective integration between even heterogeneous data [1]), (ii) a novel model to assess and increase the quality of metadata at national level, (iii) the definition of business models to convince public managers to release high-quality data and (iv) a schema for supporting the choice among a large plethora of existing licenses.

The organizational recommendations that are included in the technical guidelines are structured through an organizational model. This model aims at supporting the administrations that either want to introduce an Open Data initiative or want to enhance an on-going initiative. In this way, an administration can establish a process that is homogeneous, integrated with the processes already in place, sustainable and co-ordinated among the several business units.

The technical guidelines point out the importance of metadata management. Metadata play a crucial role when data have to be accessed by users: they facilitate understandability, search and discovery of open datasets. There exist different metadata models that significantly differ from the model included in the national guidelines since they do not provide a way to assess metadata quality. Differently with respect to existing evaluation models, we focus on the aspects related to the development of services over the data associated with them. In particular, the proposed metadata model considers the difficulty that developers may encounter in exploiting such metadata. Basically, it is independent from the physical representation of metadata and from the meaning of such metadata.

The model employs two factors: *data-metadata binding* and *metadata granularity*. Each factor has three possible values and their combination provides four levels. To differentiate from the five star model, we use the circles and consequently we also call it “4 circles” model. Figure 2 illustrates the graphical representation of the model.

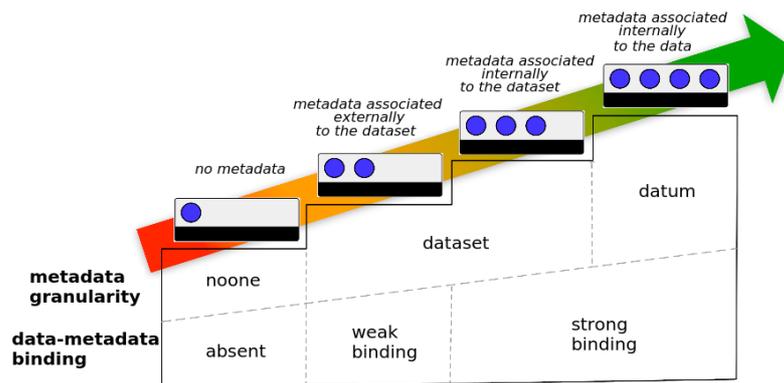


Figure 2: Metadata model

The *data-metadata binding* factor considers how the metadata are physically and logically bind to the data. It consists of 3 possibilities: *absent* (when no metadata is available – level 1), *weak binding* (when metadata are provided externally with respect to the data – level 2) and *strong binding* (when metadata are provided within the dataset – level 3 and level 4).

Of course, stronger is the binding and higher is the quality. In this way, we can use metadata on-the-fly since the proximity (i.e., in the graph of the Web of Data) between data and metadata is significantly reduced. This feature is very profitable when data have to be reused. In fact, if the binding is weak, a dataset can easily loose the related metadata after a transformation process.

The *granularity* factor represents the scope of the metadata. It consists of 3 possibilities: *no-one* (when metadata are not present – level 1), *dataset* (when metadata provide information about the dataset – level 2 and level 3) and *data* (when metadata provide information about a basic unit of information – level 4). Finer is the granularity and higher is the quality because, intuitively, the metadata are more specific and bring a larger quantity of information. Although the fourth level results to be the optimal one, there are cases (e.g., for privacy concerns, for statistical data published using the SDMX standard, etc.) where this cannot be pursued. In all these cases, we consider the third level as the best one.

Complementary to the metadata model, we introduced a common core of metadata attributes to be specified for each open dataset. This aims at enabling data interoperability at national level. The

metadata attributes have been selected among common ontologies and vocabularies, such as Dublin Core, Dublin Core Metadata Initiative, Data Catalog Vocabulary and Creative Commons. For the sake of space limit we do not report the list of the attributes, which can however be found directly in the Guidelines [7].

3. BUSINESS MODELS

The recent EU directive on the Public Sector Information foresees three cases where public bodies can release their datasets at a higher cost than the marginal cost of dissemination and reproduction: (i) the administration is a cultural institution such as a library, a university or a museum; (ii) the administration has to generate profit to cover a significant part of their internal costs; and (iii) the administration has to generate profit to cover a significant part of the costs for the gathering and the production of the data.

If the administration falls in one of these cases, it requires to choose a business model. AgID considered Open Data business models in the first released guidelines for semantic interoperability [4]. That document proposed a broad discussion about the value generation process with the partial inclusion of the study in [5].

In general, the suitable choice of the business model is a factor that affects the long-term sustainability and the success of Open Data initiatives. A business model is also useful to involve external stakeholders such as developers and SMEs in order to exploit the full-potential of the initiatives. For this reason the technical guidelines identify a small set of business models that public administrations can implement on top of their Open Data initiatives.

The set of business models does not expect to be exhaustive; rather it is an initial support for administrations that want to explore business modeling. The business models proposed in the technical guidelines are the following.

Fixed price. The administration sets a fare for the use of the data. The disadvantage of this model is the lack of flexibility, which also requires a careful analysis for the determination of the fare.

Dual licensing. This business model is based on the application of different licenses or different licensing terms to the same data. This allows administrations to employ flexible mechanisms for data pricing. For example, a dataset can be issued with a lower price if it is going to be used for research rather than business purposes.

Freemium. This business model is used extensively for the distribution of software. In the context of Open Data, it means that a portion of data is granted for free (e.g., in order to advertise the initiative) and the remaining part is charged (e.g., more refined data). It can be implemented with the dual licensing model introduced above. For example, different portions of a dataset can be subject to different prices. Let us imagine a data set representing all locations of a city. The portion of the dataset containing addresses can be granted for free since it is useful to all citizens, while the coordinates of each location can be charged.

Sponsoring. The administration provides free data and support to companies and developers in order to implement services and applications. The same data are released as Open Data, possibly with a fare. The administration can apply a constraint in the license that indicates that the implemented service/application has to make noticeable the provenance of the data. The administration takes advantage by marketing its data. Let us consider again the location data having a non-free portion of geographic coordinates. The administration can, in this case, grant this portion for free to those bodies that develop an application using the geographic coordinates also and advertise the Open Data initiative of the administration.

Equity on services. The administration provides data for free to whom create a service using those data. This is similar to the example mentioned above for the sponsoring through the application. Differently from the sponsoring, the administration shares the incomings of the service with the owner of the application.

In-House Development of applications. The administration can develop applications in-house with the use of Open Data and then exploits the business model of the application [8]. Referring to the

previous example, the administration can develop the application that uses geographic coordinates. Then, the administration can finance its Open Data initiative through the revenues of the application.

Note that this is different from developing an application without opening the data. In fact, when administrations are formed by many departments, it is often difficult to know the data owned by other departments. Moreover, in this case, opening the data has similar benefits of the open source code, where external people can validate the “open-sourced data”.

Donations. Data is free but the administration accepts donations from external users. This business model has been adopted, for instance, by Wikipedia.

Crowdfunding. The administration organizes a campaign of crowdfunding before launching an Open Data initiative. The administration has to declare the goals of the initiative and the amount of money needed to sustain the initiative. Many platforms exist to launch crowdfunding campaigns.

Support services. The administration grants Open Data for free and it generates business by providing support and consulting services. These services may include: tutorials for the use of data, data elaborations, data visualizations, data mashups with external sources, etc.

For a broader vision on the subject, we also suggested the administration to consult: the study on pricing models for the public sector published by the European Commission [3], the study of business models for Open Data published by the London Business School [9] and the survey on the business models in the context of Linked Open Government Data [10].

4. CONCLUSION

This paper illustrated the process currently in place for the valorization of the PSI in Italy. In more detail, we focused on the metadata management and the business models suggested to those public administrations that have to generate economical profit from Open Data or that want to involve external stakeholders.

AgID will continue to govern the process and, thus, support the administrations in the tricky task of finding a convenient business model for Open Data. Another aspect we want to introduce in our work are semi-automatic mechanisms for controlling the life cycle of our valorization process; for instance, an automatic assessment of the state of the Open Data.

BIBLIOGRAPHY

- [1] Giorgia Lodi, Antonio Maccioni, Monica Scannapieco, Mauro Scanu, Laura Tosco. Publishing Official Classifications in Linked Open Data. 2nd International Workshop on Semantic Statistics (SemStats 2014) in conjunction with ISWC 2014, Riva Del Garda, Italy, 2014.
- [2] European Commission. Pricing of public sector information study - models of supply and charging for public sector information (abc), 2011.
- [3] European Commission, “Pricing Of Public Sector Information Study - Models of Supply and Charging for Public Sector Information (ABC)”, 2011.
- [4] Commissione di Coordinamento SPC, “Linee guida per l’interoperabilità semantica attraverso i Linked Open Data”, Novembre 2012.
- [5] E. Ferro, M. Osella, “Modelli di Business nel Riutilizzo dell’Informazione Pubblica”, Osservatorio ICT della Regione Piemonte,
<http://www.osservatorioict.piemonte.it/images/phocadownload/modelli%20di%20business%20nel%20riutilizzo%20dellinformazione%20pubblica.pdf>, 2011.
- [6] Agency for Digital Italy, “Agenda Nazionale per la valorizzazione del patrimonio informativo pubblico (anno 2014)”,
http://www.agid.gov.it/sites/default/files/documenti_indirizzo/agendanazionalepatrimoniopubblico2014.pdf, 2014.
- [7] Agency for Digital Italy, “Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico (Anno 2014)”,
http://www.agid.gov.it/sites/default/files/linee_guida/patrimoniopubblicolg2014_v0.7finale.pdf, 2014.
- [8] M. Zarreh, “Apps for Europe - Business and Policy”, 2014.
- [9] Carla Bonina, “New business models and the value of open data: definitions, challenges and opportunities”, 2013.
- [10] P. Archer, M. Dekkers, S. Goedertier, N. Loutas, “Study on Business Models for Linked Open Government Data”, 2013.