

A federation tool for opendata portals

S. Castro, F. García, J.I Sánchez. *Red.es*¹. Madrid, Spain

M.A. Amutio, D. Hernández. *Ministerio de Hacienda y Administraciones Públicas*². Madrid, Spain

S. Soriano. *Ministerio de industria Energía y Turismo*³. Madrid, Spain

Abstract

The value of the public data being made available for reuse is becoming increasingly acknowledged not only by the private sector but also by the public entity that generates them. For this reason, more and more public entities are creating their own open data spaces in which to gradually expand the volume of information being made available to reusers.

These different spaces - which, generally speaking, comprise a series of unconnected islands - now require a single point where all the published information can be permanently and automatically referenced. In order to meet this need, the team responsible for the Aporta Project is developing a National Catalogue of Reusable Public Information at datos.gob.es and is publishing Technical Interoperability Standards on the Reuse of Information Resources. These two measures make it possible to develop a *catalogue federation tool* to enable automatic publication in the National Catalogue of the metadata corresponding to the data sets published on the websites of each public entity. A global index of reusable public information is thus created and can be accessed by companies or any member of the public to locate reusable data without the need to know and find the website of the public entity holding the data of interest to them.

The catalogue federation tool, which is being developed in *PHP* as an extension of the National Catalogue, ensures maximum coherence between the information being made available by the public entities in their own catalogues and the National Catalogue itself.

This solution enables the existence of a global reuse scenario that provides greater visibility for the public data made available by the three levels of government (central, regional and local), as well as a general overview of how public sector information is being reused in Spain.

Keywords: DCAT, federation, interoperability, legal framework, metadata, national catalogue, opendata portal, PSI, standards, tools, vocabulary.

1. Context

The opening up and reuse of public sector information (PSI) is a project being developed by the Spanish public administration services and has a direct impact on productive sectors and the public authorities themselves. On the one hand, opening up this information enables transparency mechanisms to be implemented and makes it possible for the general public to participate and collaborate, thereby promoting a more efficient democratic system. On the other, reuse of the information held by public administration services fosters the development of new digital products and services, thereby stimulating economic and business activity and ultimately providing value for society as a whole.

¹Sonia.Castro@red.es, francisco.garcia.vieira@red.es, JoseIgnacio.Sanchez@red.es

²miguel.amutio@seap.minhap.es, mdolores.hernandez@seap.minhap.es

³[sorsoriano@minetur.es](mailto:soriano@minetur.es)

Fully aware of the benefits involved, numerous initiatives on open data are currently emerging in Spain at central, regional and local government levels, as well as in universities, as shown in Fig. 1.

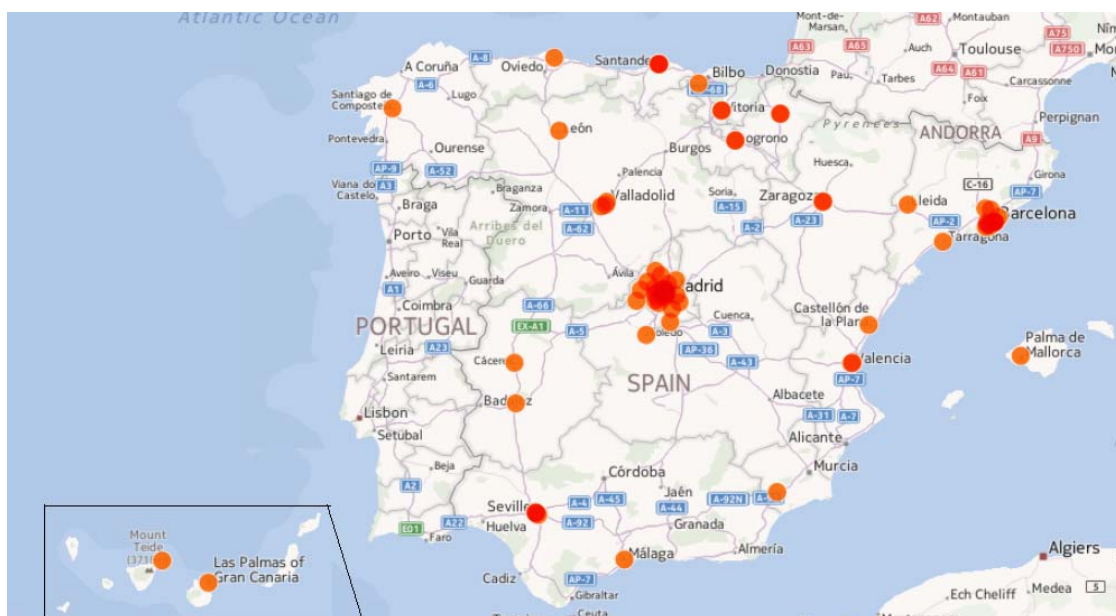


Fig. 1. Open Data Initiatives in Spain (April 2014)

All these open data projects were publishing their own data inventories in structured catalogues. Each one is an independent consultation point with descriptive and technical information on the data sets they contain. However, they lack a common, standard framework. This situation - i.e. numerous unconnected projects - has been highlighted repeatedly by the Open Data Community, which has been calling for standardisation between public services as a basic premise for continuing this work. This principle required a project capable of encompassing all initiatives and establishing guidelines to ensure standardisation with any other existing initiatives. Efforts are now being made within the framework provided by the Aporta Project to address this situation.

Developed by the State Secretariat of Telecommunications and the Information Society and the State Secretariat of Public Administration Services in collaboration with the public company red.es, a number of actions are being carried out within the Aporta Project. Among them, the following should be highlighted:

- The creation of a legal framework on data reuse: Law 37/2007, of 16 November, and Royal Decree 1495/2011, of 24 October, which develops the aforesaid Law; this legislation informs each entity of its duty to publish public data and make that data available for reuse, and provides guidelines for doing so.
- The development of a single information access point - the National Catalogue of Reusable Public Information (hereinafter, datos.gob.es), in which each public entity must reference the data made public. The latter must always be fully consistent with the data in its own information systems.
 - The drawing up of technical standards to establish common conditions for: the selection, identification, description, conditions of use and making available of data sets - the Interoperability Technical Standard on the Reuse of PSI (hereinafter, ITS-PSI), which define a DCAT profile for the public information catalogues at the various government and agency levels in Spain, **and closely linked to the DCAT Application profile for data portals in Europe.**

Based on these actions and in order to facilitate the task of public employees to publish reusable information in two different places, the team responsible for the Aporta Project is developing a federation tool that enables the metadata for their data sets to be aggregated and automatically published in their own catalogues and on datos.gob.es.

The main features of this project are: guidelines from agencies that specialise in this issue have been considered when creating the current legal framework established by ITS-PSI, as well as the technology solution implemented.

Proposals from experts belonging to the World Wide Web Consortium (W3C) - the body entrusted with supporting the development of open, free and interoperable standards that ensure the long-term growth of the Web - and the Spanish Open Data Community were considered when fine-tuning the requirements of the solutions developed.

Furthermore, the tool has been evolved while considering the common needs of the group directly involved in using the federation tool, i.e. PSI managers at public entities.

2. Main benefits

Functionality of the federation tool has been designed to meet a number of important requirements, including:

- The ability exists to interconnect open data initiatives at a single access point - datos.gob.es - that can be accessed by the reuser or any member of the public via a search tool to locate reusable information without needing to know and find the website of the public entity holding the data of interest to them, strongly contributes to the efficiency of the research processes. It facilitates enrichment of datos.gob.es through the large-scale upload of meta-information associated with the data sets made available by public entities for reuse.
- Maximum consistency between the information being made available by the public entities in their own catalogues and the information referenced at datos.gob.es.
- Reduced workload for public employees in their task of publishing the data sets subject to reuse by avoiding the need to upload information twice, both to the internal catalogue and datos.gob.es.
- It enables the existence of a global scenario that fosters the extraction of general conclusions and a general overview of the PSI situation in Spain, facilitating the use of this information to extract meaningful and actionable knowledge regarding the open data landscape.

3. The technological solution

Legislative and technological bases

As stated above, current legislation recommends that public entities establish mechanisms to ensure data standardisation, integrity and the constant updating of information published.

The federation solution, which is being developed to facilitate compliance with these recommendations, requires each entity to use a structured, predefined format for publishing the metadata that are defined as necessary for the data sets in question. As a result, technical standards are needed to ensure interoperability.

The aforementioned ITS-PSI was set up with that in mind and within the framework of the Interoperability National Scheme (INS). Specifically, the set of necessary metadata, a standard vocabulary and the range of values to be used have been defined with a view to ensuring precise uniformity and enabling the federation tool to function as required.

Description and how does it work

The federation tool - developed using the PHP open-source programming language - acts as an extension of the datos.gob.es portal, which was developed using Drupal 7.

Via a module integrated into datos.gob.es, federation is made possible by accessing the required metadata as specified in the ITS-PSI and that the public entity must make available in a DCAT/RDF or ATOM format feed. This feed, which can be generated automatically from the content of the entity's portal or manually by the editor, must be accessible at a known URL on the website of the entity in question according to a pre-established format. This

establishes a reuse method for the metadata associated with the data sets of the entity, which can be used by datos.gob.es or any other possible customers of the entity.

The solution developed is based on a “pull” system so that the datos.gob.es portal server is responsible for initiating the process by means of pre-established protocols, thereby accessing the information available from each entity for updating its records, the native format of which is already known.

The datos.gob.es federation tool accesses the metadata of each entity and updates the meta-information available at datos.gob.es on a pre-established regular basis.

The representative of the public entity in question, in possession of the permits required to upload data sets to datos.gob.es, can select the frequency with which the metadata from that entity must be updated to datos.gob.es (from a list of accepted options). Additionally, the entity’s representative is responsible for configuring whether to enable the federation and the URL at which the feed is available.

The diagram below shows the interactions between the various parts of the system (Fig. 2):

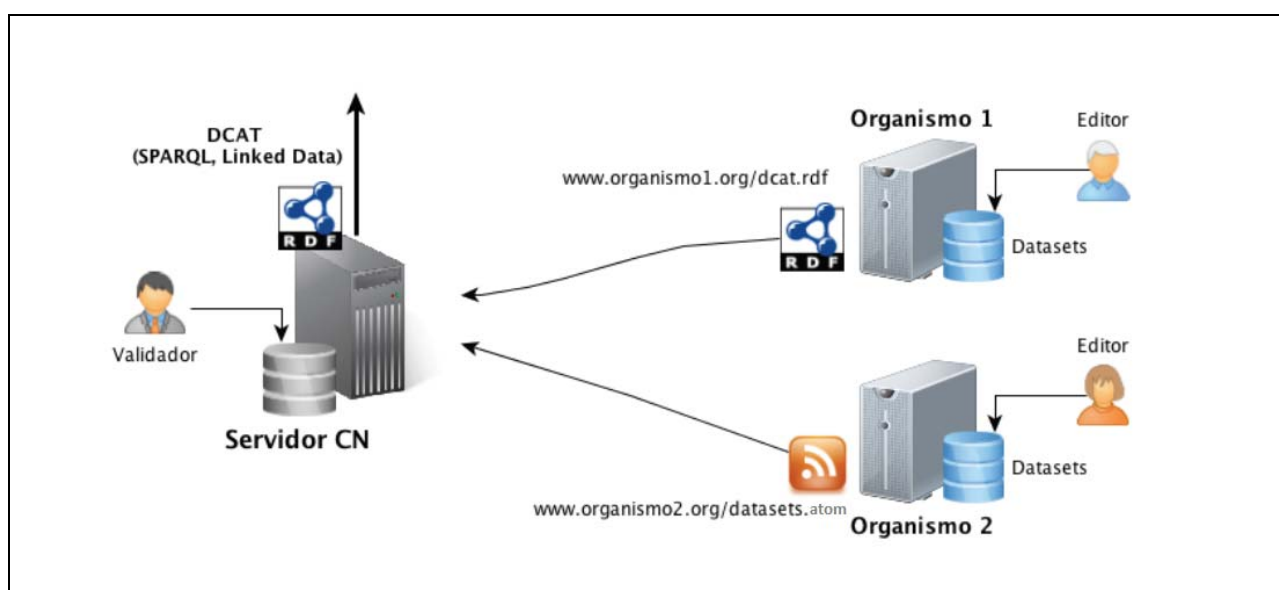


Fig. 2. Interaction between Components of the Federation Tool

The editors/administrators of an entity publish or modify a data set in their own open data catalogue.

1. The content manager of the entity automatically publishes the metadata for the data sets in an RDF or ATOM structured format, or the editor publishes the file manually.
2. At datos.gob.es, the entity manager configures the metadata publication URL and the update frequency.
3. datos.gob.es then imports the metadata from the previously-defined entities with the configured frequency.
4. The moderator at datos.gob.es may approve or invalidate any metadata import process.

Complementary tools (webservices and widgets)

For public entities that wish to make information available for reuse according to current regulations and ITS-PSI requirements but lack the material and human resources to do so, a series of complementary tools has been developed to make that possible.

The datos.gob.es portal enables each authorised user to manually register any data sets they wish to publish for reuse. Any information uploaded to datos.gob.es manually can be subsequently referenced on the website of the entity, either via a parameterizable widget or a web-based consultancy service that enables the meta-information

published in the catalogue to be obtained and processed according to various invocation parameters and various response formats (ATOM/XML, DCAT/RDF and JSON).

By using these complementary tools, it is possible to generate a global feed that enables convergence between datos.gob.es and the EU Open Data portal (<http://open-data.europa.eu/en/data>), which will facilitate the location and reuse of data from national, regional and local administration services throughout Europe.

4. Implementation results

Numerous public entities have been using the federation tool since it was launched. Specifically, the reusable data records from the Spanish National Statistics Institute are uploaded en masse to datos.gob.es via the data federation tool. The Regional Government of Castile and Leon, Lorca City Council, the Pablo de Olavide University and the Social Security Management Office publish the open data sets already present in the open data spaces of their websites on datos.gob.es, thus guaranteeing information integrity. Gijón City Council was the latest public body to federate the 423 data sets available in its catalogue with datos.gob.es, meaning that this information can be found in either data catalogue and is therefore more accessible to any open data consumer or user.

This federation system will soon be used to upload data sets to datos.gob.es from the catalogues of such public entities as the Regional Government of Aragon, Madrid City Council, Zaragoza City Council, the Regional Government of Galicia, the Centre for Sociological Research and Malaga City Council, among others. This will enable the number of data sets stored at datos.gob.es (currently 1,631) to be multiplied at least threefold.

The volume of federated data will continue to increase as public entities - with support from the technical team responsible for the Aporta Project - carry out the adjustments needed in the feeds they generate. This will lead to the large-scale upload and regular updating of the information published on datos.gob.es.

5. Conclusions

Project Aporta is a nationwide initiative commissioned with harmonising actions to encourage making public information more widely available, thereby enabling the implementation of transparency mechanisms and the development of the public sector information reuse market. This is being achieved by establishing interoperability criteria and making efficient use of the synergies that arise between the open data projects emerging from public administration services in Spain every day.

In order to promote the availability of public data, an official single point for hosting this data is being created at datos.gob.es. Data are displayed in a clear and structured fashion on a user-friendly interface for reuse. The catalogue federation tool enables data belonging to various local and regional authorities to be reused with the National Catalogue.

The federator - developed according to guidelines set down by experts in the field - ensures standardisation and data integrity, and enables automated publication and constant updating of published information, while also enhancing the visibility of the data sets made available by the various public entities.

The federation tool - integrated as an extra module on the datos.gob.es portal - accesses the metadata of each entity and updates the meta-information available at datos.gob.es according to a pre-established schedule.

This ensures effective federation with datos.gob.es of the open data catalogues of the public entities and, in a future step, with the Pan-European Open Data portal (<http://open-data.europa.eu/en/data>), which seeks to facilitate the location and reuse of data from national, regional and local administration services throughout Europe.