#

# Name of the Share-PSI workshop: SAMOS. Uses of open data within government for innovation and efficiency

# Title of the Best Practice: A federation tool for opendata portals

# Outline of the best practice

The Spanish National Catalogue *datos.gob.es* has developed ***A federation tool for open data portals*** that enables automatic publication 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 user to locate reusable data in datos.gob.es without the need to know and find the website of the public entity holding the data in which re-users are interested in.

# Management summary

## Challenge

More and more public entities are creating their own open data spaces in the web. These different spaces are unconnected.

To avoid that reuse agents have to look for the data all over the web, we needed a single point where all the Spanish public sector information can be permanently and automatically referenced at the National Catalogue of Reusable Public Information: datos.gob.es

## Solution.

The catalogue federation tool enables aggregation and automatic publication of the metadata corresponding to the data sets published in the own catalogues on the websites of each public entity and also at the National Catalogue datos.gob.es in a consistent way.

A global index of reusable public information is thus created and can be accessed. The tool developed in PHP as an extension of the National Catalogue data portal datos.gob.es 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, local and universities), as well as a general overview of how public sector information is being reused in Spain.



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

# Best Practice Identification

## Why is this a Best Practice? What's the impact of the Best Practice?

* The ability 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.
* 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.
* 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.
* It helps to accomplish PSI Directive to each entity of its duty to publish public data and make that data available for reuse, and provides guidelines for doing so.
* Fosters the development of new digital products and services, thereby stimulating economic and business activity and ultimately providing value for society as a whole.
* The use of the federation tool has leaded to the large-scale upload and regular updating of the information published on datos.gob.es with a considerable increase on datasets number in datos.gob.es.

## Link to the PSI Directive

*(Please use one or more of the categories listed on the last page of this document, as many as relevant)*

* **Policies and legislation (legal requirements**, licenses etc..)/ Licensing of information/data and metadata
* **Open Data platform(s) / Publication and deployment of information/data and metadata**
* **Charging issues** and proposals
* Techniques w.r.t. opening up of data **/ Technical requirements and tools**
* **Organisational structures and skills (coordination)**
* Dataset structures, formats, APIs / **Structuring of information/data, formats, APIs**
* **Encouraging (commercial) re-use (facilitates)**
* **Persistence and maintenance of information/data and metadata**
* Data quality issues and solutions / **Quality assurance**, feedback channels and evaluation **(standardization)**
* **Documentation of information/data, creation of metadata**

## Why is there a need for this Best Practice?

* To get the most out of scarce public resources that are available in our country
* To facilitate reuse by reuse sector in Spain. Datasets are displayed in a clear and structured fashion on a user-friendly interface for reuse.
* To get a set of guidelines to use standard metadata structure in data sets following W3C recomnedations that ensure the consistent growth of the Spanish Open Data Catalogue.
* To facilitate the task of public employees avoiding to publish reusable information in two different places.

# What do you need for this Best Practice?

* A legal and technical framework ensuring that each public entity will federate their datasets at the national data portal in a standard manner.
* A coordination structure between the different administrative levels (State, regional, Local) to agree the metadata considering the common needs of the group directly involved in using the federation tool to assure the further collaboration using the tool that means for the success on the initiative.
* A data portal as a single point where federate all published datasets of the different public entities. The federation tool is a module PHP open-source programming language that acts as an extension of the data portal, which was developed using Drupal 7.
* An agreed Metadata scheme following W3C recommendation should be available in a DCAT/RDF or ATOM format feed which must be accessible at a URI on the website of the entity origin of the data.
* Some complementary *webservices and widgets* that enable 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) to be referenced on the entity of the data website.

# Applicability by other member states?

* 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.
* 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.

# Contact info - record of the person to be contacted for additional information or advice.

soporte@datos.gob.es ; <http://administracionelectronica.gob.es/general/verContacto.htm>

# Categories for use in section 3.2

* Policies and legislation (legal requirements, licenses etc..)/ Licensing of information/data and metadata
* Open Data platform(s) / Publication and deployment of information/data and metadata
* Dataset criteria and priorities and value and scope w.r.t. datasets
* Charging issues and proposals
* Techniques w.r.t. opening up of data / Technical requirements and tools
* Organisational structures and skills
* Dataset structures, formats, APIs / Structuring of information/data, formats, APIs
* Encouraging (commercial) re-use
* Persistence and maintenance of information/data and metadata
* Data quality issues and solutions / Quality assurance, feedback channels and evaluation
* Documentation of information/data, creation of metadata
* Selection of information/data to be published according to various criteria
* Data discoverability