



EUROPEAN COMMISSION

Communications Networks, Content and Technology Directorate-General

Media & Data
Data Value Chain

Minutes of the 4th meeting of the public sector information sub-group on the pan-European Open Data portal – Luxembourg, 27 November 2013

1. Welcome and introduction by Chairman

The Chair summarised the key developments in relation to opening up public sector information that occurred since the last meeting, in particular

- The adoption of Directive 2013/37/EU revising Directive 2003/98/EC on reuse of public sector information in June 2013 that reinforces the European legal framework;
- The adoption of a "G8 Open Data Charter" approved by G8 leaders at a summit in June 2013, establishing among other things that government-held information should be open by default.
- The conclusions of the (Digital) European Council of 24 & 25 October specifying that "[o]pen data is an untapped resource with a huge potential for building stronger, more interconnected societies that better meet the needs of the citizens and allow innovation and prosperity to flourish. Interoperability and the re-use of public sector information shall be promoted actively." The Chair said that it would now also be up to the group convened to turn this mandate into concrete actions.
- A process of publishing guidelines by the Commission to Member States on selected elements of the new Directive, namely how to implement the rules on charging for public sector information, how to ensure interoperable licensing conditions and establishing an agreed list of datasets that should be released for reuse as a matter of priority.
- In the new framework programme of the Commission on research & innovation an emphasis will be put on the idea that research projects should try to use openly available data as input for their projects, reinforcing the development of the open data policy from the demand side.

The purpose of this working group is to assist the Commission in preparing the technical specifications for a pan-European digital service infrastructure for data. The Commission in the meantime has been able to work on finding the funding. This will come from the Connecting Europe Facility programme.

2. Supporting the publication of Open Data

Sander van der Waal, project leader at the Open Knowledge Foundation, responsible for the publicdata.eu project, explained that publicdata.eu is aggregating data from ever more portals, has

increased referencing from 19,000 to 30,000 datasets since last meeting in June 2013. User statistics show that people from all 28 Member States use the site. Recent re-launches are taking place on FR, SE, ES and publicdata.eu will work on integrating them. Harvesting from CKAN-based portals is relatively straightforward as publicdata.eu also uses CKAN. For each non-CKAN portals, a custom API needs to be developed requiring direct interaction with each such portal to design the specifications of the API.

Current research focusses on some automated transformation allowing converting CSV files into RDF compliant files. All project results are made available in open source.

Further development of publicdata.eu will include:

- Better support for multilingual data: Multilingual support at data set and API level;
- Further work on data catalogue interoperability and aligning with work of the Open Data Support project;
- Cooperation with Apps4EU competitions in order to showcase results on publicdata.eu.

Sander explained that all portals that are not yet registered under publicdata.eu should actively contact the team (info@publicdata.eu). The team would also be interested to learn about showcase examples and apps competitions.

In the question and answer round, Sander further clarified:

- Publicdata.eu harvests only from a national portal where this portal already aggregates content from other portals in that country, but can harvest directly from such other portals where no aggregation takes place;
- Beginning of 2014 a DCAT-AP compatible version of CKAN will be released;
- Implementing CKAN alone is not enough to guarantee full interoperability with publicdata.eu, other elements (compliance with the DCAT AP; API) are also important; this is usually addressed in bilateral communication between the publicdata.eu team and the respective portal;
- CKAN-based systems are scalable; more and more countries turn towards CKAN, also the US that currently references 88,000+ datasets ; publicdata.eu can harvest from portals not running on CKAN;

Nikos Loutas, project leader at PriceWaterhouseCoopers, responsible for the [Open Data Support project](#), explained that this project is about filling a future pan-European digital service infrastructure for data with high quality content and about building a capacity inside public administration in all EU Member States on opening up government data. Publicdata.eu is focussing on developing a pilot infrastructure interlinking national open data infrastructures, Open Data Support is focussing on the content for that infrastructure. Both projects are in regular contact in order to discuss the division of roles in order to avoid duplication.

The project has been running since January 2013.

The overall objective is to enhance the outreach of data published on national data infrastructures making them available to users throughout Europe and beyond. It provides a centrally-run metadata publishing infrastructure (ODIP) that harvests metadata from repositories in the Member States. No data are being centrally stored. Harvesting takes place only on the basis of an agreement by the national infrastructure in question. The project uses the CKAN API to harvest wherever possible, but also works with data infrastructures not using CKAN distribution. Metadata are transformed by the project into a [DCAT AP](#) compatible format on the basis of mappings. There is no obligation to change metadata specifications at the source. The project also tries to enrich the metadata harvested using thesauri and controlled vocabularies such as Eurovoc and the name authority lists provided by the Publications Office of the EU in order to fill the value fields in the metadata tables. The mapping process is semi-automatic. The mapping of the field descriptions is done manually, but can then be transformed into a SPARQL query. For the field values this is done in a similar way.

Currently, Open Data Support works with 11 Member States, the EU Open Data Portal and the INSPIRE geo portal operated by the Joint Research Centre of the EU. 32,000 harmonised metadata descriptions are thus already available.

The project team can work in a number of EU official languages; for some countries there may be the need for some interaction with the colleagues in order to overcome the language barrier in relation to metadata field descriptions and values.

The project also offers training services for public administrations, both in terms of making powerpoint slides available [online](#) for use and reuse and in terms of offering on-site trainings. On-site trainings should always have a speaker from the Member State concerned that will set the training activity into the national context.

In the question & answer part, Nikos specified that

- It is technically possible to make the content of the ODIP metadata repository searchable by any national infrastructure; ODIP has a SPARQL endpoint that can be accessed via a search plugin developed in some data portal; this would allow that any content from any portal from which metadata references are harvested by ODIP could be accessible from any other portal; however, the budget allocated to the Open Data Support project does not make it possible to offer this service beyond a small number of pilots to all open data portals in the Member States at this stage;
- The project is aware of the limitations of the thesauri and controlled vocabularies it uses, in particular their granularity and limits in adaption to specific national or sectorial contexts; however, these are the only resources available in all 24 EU official languages having a guarantee of continuity of service in the mid-term.
- The project is also aware that it may take a long time until the DCAT AP is a universally accepted standard used in all public administrations or that this may never be the case, but believes that it is still worthwhile to try to map as many metadata specifications as possible to the DCAT AP; achieving the objective of more harmonised metadata standards may need the involvement of many actors, also at national level;

- Mappings are currently done only once, but updates potential necessary if metadata specifications at the source change.

3. Towards the deployment of a pan-European Open Data platform

The objective of this session was to collect input for the design of the technical specifications of the future pan-European open data platform.

a) Portal end-user interfaces (data visualization, advanced search facilities, ...)

The session was kicked off by two presentations:

Input presentation by Ross Jones, data portal lead designer at data.gov.uk, UK:

The UK [open data portal](#) is designed to serve a quite diverse group of end-users: Public bodies (publishers), developers, data scientists, software companies, citizens. This represents a challenge.

End-user interfaces should be lean and easily understandable. Therefore the web interface of data.gov.uk was recently redesigned and the number of options visible on the homepage drastically reduced, pushing other content to the back without losing it. Interfaces need to be ready for mobile devices which make up a growing share in end-user devices.

Data publishers like to find an interface that provides metrics in relation to their data.

For the group of developers/ data scientists a documented and working API and documentation on the data are important as well as the quality of the data and the continuity of the service. Previews on the actual raw data are also interesting. In respect to metadata curation, quality can be challenging. Some publishers are really eager to have [Five-Star](#) data, others are being assisted by the data.gov.uk team. Data.gov.uk allows sorting by the number of stars in the five star scheme. It cannot be said that data having more 'stars' are used more often. Metadata should be human/ user-friendly.

For citizens the emphasis is on content curation in the sense of rendering the data in a format that is intelligible for the ordinary citizen is important (e.g. visualisation tools). Here the question is whether such tools should be provided by the public entity responsible for the data infrastructure or whether these tools should come from the market.

The site has a feedback function (forums & 'request data' workflows) and a 'data users group' is in place.

In the future the site should be enhanced in particular in relation to Linked Data and making content searchable in natural layperson language.

Input presentation by Salvador Soriana, Ministry of Industry, Aporta project, Spain

Salvador explained that the overall development of the [Spanish open data portal](#) datos.gob.es was done in an incremental ways, advancing in small steps and trying to focus on features and content that are useful, evaluating progress at any moment in time in order to define the course of action.

Salvador supported the statements of Ross on a lean and appealing site structure. The recent relaunch of the Spanish Open Data portal datos.gob.es is also inspired by this idea.

There is a legal obligation for the central government departments to make content available through the portal.

The user group that the Spanish portal is mainly focussing on is the group of developers.

The portal also contains training material and background information (e.g. legal information).

The search tool will lead to a list with a summary of the data. It allows for sorting and filtering according to a number of categories.

Applications are available through the portal. They are not developed by the Aporta project itself, but come from third parties.

There is no formal feedback channel on uses made of the data referenced in the portal. The Aporta project tries to use informal contacts with the developer community in order to have a sense of what is useful or could be useful in the future.

Federating content from all Spanish portals based on a Spanish National Technical Interoperability standard which is based on DCAT will happen progressively.

Specific advice when contemplating on a pan-European portal: Promotional activities on data re-use are very important, the sector still being quite immature.

In the question & answer round, Romania suggested to include some form to allow data to be displayed based on geospatial attributes (metadata values). Sander pointed to the example of France that seems to allow 'trusted' users to give feedback on data and improve data quality.

The EC pointed to the problem that metrics for data infrastructure have to rely on site visits and cannot measure reuse made of the data, because it would be undesirable to request any reuser to specify before or after download of the data what reuse is intended for the data. It may happen, however, that the number of site visits will drop in the mid or long term as end users will never access the data infrastructure in order to obtain the data, but go through intermediary providers. While this is a desired outcome, it may impact on the internal support given to data infrastructure projects based on metrics. It seemed that none of the participants knew a solution to this problem.

Sander van der Waal concluded that future work on a pan-European portal should put an emphasis on APIs so as to target developer-user.

The chair invited participants to share best practice examples together with potential feedback on the minutes.

b) Users interaction and user needs collection for an evolutive maintenance of the portal

**Input presentation by Jean-Charles Quertinmont, Administrative Simplification Agency,
Chancellery of the Prime Minister, Belgium**

Jean-Charles explained that Belgium currently has two portal sites at the level of the federal government, publicdata.belgium.be and data.gov.be. Apart from the fact that they are run by different services they also aim at different audiences, one for the citizen at large and one for developers. A merge in the future is planned.

Business demand for public sector information (open government data) is low and not well organised. Companies use traditional channels for obtaining data.

Civil society demand is stronger and better organised.

Main users of the portal sites are the public administrations, allowing them to see what data are being collected by other services.

User interaction is not straightforward. One of the questions to be tackled is whether the current differentiated access for citizens and developers remains useful.

In the question & answer part it was clarified that

- Neither portal actually hosts the data, but references to the original source and thus also always points to the licensing conditions chosen by the original publisher; a model licence exists, but is not widely used.
- Level of openness may vary from dataset to dataset depending on the choice of the original publisher;

With respect to this, the chair pointed to a paper recently published by the Legal Aspects of Public Sector Information 2.0 (LAPSI 2.0) network funded by the European Commission on interoperability of licensing conditions (available here: <http://www.lapsi-project.eu/outputs>).

The chair invited participants to share their experience on user engagement in order to obtain input on the development process of national portals:

- **Romania** reported that engaging with users was not straightforward in the Romanian case. From the outset of the development of a national open data infrastructure, the government engaged with NGOs, with mixed results. NGOs sometimes benefit from absence of openness if they can have a strategic advantage based on privileged access to government documents. Currently, interaction develops in a more positive way.
- In **Sweden** the open data portal is also used in order to encourage administrations to open up data.
- In **Spain**, virtual communities were created to support the development, but challenging to keep them engaged.
- In the **UK** engaging typically happens when there is a problem with the data; CKAN gives you 80% of the functionalities out of the box; changes to the interface are not necessarily driven by end-users, sometimes by developers, though, based on experimentation. The Open Data user group is a positive experience, but does not have a tremendous impact on the increasing the number of data made available. The impact of the "request data" feature in

data.gov.uk in terms of human resources is not easily measured as the main work is done by the service holding the request data.

c) Towards a portal governance structure and its long-term sustainability

In its introductory statement, the EC presented a number of questions to the participants in relation to the governance of a pan-European digital service infrastructure and its long-term (financial) sustainability:

What body should oversee the setting up of the infrastructure (including designing the technical specifications) and the operations? What is the division of roles between the EC and the Member States? What decisions would need to be taken at what (hierarchical) level)?

In relation to the budget, the EC specified that funding for an initial phase would come from the Connecting Europe Facility programme. In the mid- and long term, other options need to be envisaged. Should it continue to be run exclusively by public money? Could revenue resulting from services private operators develop on the basis of the data be used?

The EC specified that many options are possible and the EC has no preconceived ideas.

Malta pointed to the fact that governance also involves having a solid data management policy and structure in place within the entire public administration.

Germany suggested to envisage an evolutive approach. Instead of designing a potentially heavy and complex governance structure from the start, it may be advisable to start light in order to facilitate take-off and adapt governance and funding later as the infrastructure grows.

Austria explained that its federal portal federates data also from regional ('Länder') level, but funding came and continues to come always from the (federal) Prime Minister's Office. This could only be an interim solution, however.

The EC noted that

- All the portals so far are funded with public money.
- However, experience with other federating initiatives on the EU level (e.g. Europeana) suggests that the governance and financing issues will inevitably come back and so, a reflection process should rather start faster.

In its conclusions the chair invited the participants to reflect on the questions raised in this last part of the meeting and submit comments and suggestions, including best practices.

He reminded participants that national portal infrastructures need to be actively registered with publicdata.eu in order for harvesting to be possible (by sending an email to info@publicdata.eu) and that the Open Data Support project's services can be solicited.