



Luxembourg, 24 June 2013

CONNECT.G3/DR V2

**Subject: Minutes of the 3rd Meeting of the Public Sector Information sub-Group on the pan-European Open Data portal - Luxembourg, 18 June 2013**

**1. WELCOME AND INTRODUCTION**

The meeting chairperson, Yvo Volman, deputy Head of the Data Value Chain Unit of DG CONNECT, welcomed the participants.

As an introduction the chair gave a short overview of the on-going initiatives at the European level related to open data and the open data portals.

Open data are high on the agenda of the G8 summit at Lough Erne (Northern Ireland) on 17-18 June. An open data charter<sup>1</sup>, including basic principles to be adopted by G8 members as “government information as open data by default” and commitments for a short term release of core datasets as open data, was drafted and will be endorsed at the G8 summit.

The revision of the PSI directive has been finalised and is in the process of being adopted at first reading. The European Parliament already positively voted the agreed text between Commission, Parliament and Council, and the final adoption is expected in the coming weeks. Main points included in the revision are free of charge or minimum costs by default, cultural sector now in the scope of the Directive, all information accessible also reusable for commercial and non-commercial purposes. Open and machine readable formats are recommended for dissemination to facilitate reusability.

The Commission will issue recommendations on charging, core datasets and licensing.

The Commission developed an EU Open Data Portal<sup>2</sup> to progressively give access to all datasets held by the Commission and other EU Institutions and bodies. The portal, operational since December 2012, is the result of the cooperation between DG CONNECT and the Publications Office, which is also hosting the website. At the moment around 6000 datasets are available, from several Commission DGs and the European Environment Agency.

The Connecting Europe Facility (CEF) programme, proposed for the period 2014-2020, should be a fundamental instrument for implementing a consistent pan-European infrastructure to support open data access. Severe cuts were applied to the budget initially planned for the whole programme. Even if the cuts will in the first place affect the

---

<sup>1</sup> <https://www.gov.uk/government/publications/open-data-charter>

<sup>2</sup> <http://open-data.europa.eu/>

deployment of broadband infrastructure, the budget foreseen for the development of Digital Service Infrastructures (DSI) is also considerably reduced. This may affect the funding available for developing a pan-European platform for accessing open data public sector information. The proposal for a Regulation on CEF is currently under discussion with European Parliament and Council. It is of the utmost importance that support is provided by MSs representatives in order to maintain the Open Data DSI in the programme together with an appropriate budget. At the same time it will be important that MSs already make a reflection on possible ways of guaranteeing the long term sustainability to the open data platform. Support and feedback from participants is necessary and very welcome.

The chair finally underlined that open data portals are now at the forefront of public administrations priorities, and congratulated the winner of the European Prize for Innovation in Public Administration 2013, a national open data portal.

## **2. REINFORCING PUBLICDATA.EU AS AN OPERATIONAL PROTOTYPE OF A PAN-EUROPEAN OPEN DATA PORTAL**

A member of the LOD2 project presented the current status of the publicdata.eu web site, the one-stop shop for Open Government data proposed to be used as the prototype of the Commission pan-European Open Data Portal<sup>3</sup>.

The portal currently references around 20000 datasets from 12 countries and 26 different sources, mostly harvested from existing CKAN national portals. Links with other types of portals have been initiated, e.g. through a pilot with the Spanish open data portal. Features developed for publicdata.eu are being included in the next version of CKAN, 2.0, thus making solutions developed in the context of the project available to the whole community.

Plans envisaged for the remainder of the LOD2 project (ending in August 2014) are to integrate additional open government data portals, in particular the national ones, increase the uptake of data, e.g. through the Apps for Europe (Apps4EU) project<sup>4</sup>, automate the harvesting process (currently launched manually), connect non-CKAN portals, provide DCAT metadata compatible with the newly developed application profile (see point 3 on the agenda) and develop an API for direct access to datasets or a dump of all metadata in RDF/XML or JSON. These plans also constitute the main project challenges, and envisaged solutions must include a strong involvement of the community as well as the coordination with other on-going initiatives.

### Discussion

Some MSs explained that they are in the process of creating a national open data portal, expressing interest in getting guidance and access to technical know-how. Some support can be provided by the Open Data Support project (see point 4 on the agenda).

The problem of data duplication (the same or different versions of the same dataset harvested and referenced more than once) was discussed. A possible solution would be the definition of a “reference” dataset at the level of national authorities, possibly with the aggregation of datasets at the national level.

---

<sup>3</sup> <http://www.publicdata.eu/>

<sup>4</sup> <http://www.appsforeurope.eu/>

Participants were reminded of the existence of a directive, INSPIRE, establishing a EU-wide cross-border data and service infrastructure in support to environment policies. The INSPIRE geo-portal<sup>5</sup> and its content may be connected to publicdata.eu, e.g. through its catalogue service.

The need for agreed and stable URIs, and for the adoption of multilingual common vocabularies was also debated.

### Conclusions

**The participants agreed to build on publicdata.eu to work towards a pan-European open data portal, and to cooperate with the project and the Commission in providing or extending the access to their open datasets.**

**It was also agreed that all members of the group not yet in touch with publicdata.eu would contact the portal representative<sup>6</sup> in the week following the meeting.**

### **3. AREAS OF INTEROPERABILITY**

The chair opened the floor asking participants to share their experiences regarding metadata interoperability.

A problem raised by some delegation is the different degree in the implementation of a coherent open data policy at different levels of government and in different domains, as well as the development or the adoption of different domain related metadata models. Metadata catalogues already exist but they are not yet interoperable. Some domains (e.g. spatial information) are already bound to implement specific standards defined through European legislation (INSPIRE).

A common requirement for overcoming this problem, at the national as well as European level, is the definition of a simple common model, extensible in order to satisfy the specificities of different domains. The necessity of developing solutions to make different technical implementations (CKAN, DCAT) interoperable, was underlined.

The importance of sharing experiences and exchanging best practices, at the national and European level was also highlighted.

Regarding non-technical issues, common/interoperable licensing schemes were considered essential to implement a common framework for open data. The Commission reminded the participants that guidelines on licenses will be issued following the adoption of the revised PSI directive, and that the first step will be the launch of a public consultation. Already existing national experiences regarding a common license framework are very interesting and will be duly taken into account.

#### Interoperability at the level of the metadata schema

In the framework of the European Commission ISA programme (Interoperability Solutions for European Public Administrations), the SEMIC project has set up an international group of experts for the definition of a DCAT application profile (DCAT-AP) to be used as a common metadata schema for describing public sectors datasets in

---

<sup>5</sup> <http://inspire-geoportal.ec.europa.eu/>

<sup>6</sup> LOD2/publicdata.eu : [info@publicdata.eu](mailto:info@publicdata.eu) ; Sander van der Waal : [sander.vanderwaal@okfn.org](mailto:sander.vanderwaal@okfn.org)

Europe in a standardized way. The specifications have already been drafted and have been submitted to public consultation<sup>7</sup>, open until 15 July 2013.

The adoption of a common metadata schema to describe datasets would benefit both data publishers and data users/reusers, the former increasing discoverability of their data, the latter being able to uniformly search across platforms. The developed specifications include a simple model based on mandatory, recommended and optional classes, prioritization being based on usefulness and the desired degree of interoperability. The proposal also recommends the use of multilingual controlled vocabularies (such as EuroVoc) to achieve semantic interoperability.

The goal of the proposal is not to enforce the use of a common metadata model at the source, but to adopt an agreed common metadata model at the European level. National/local/sectoral models should be easily and automatically mapped on this model. Some examples were shown to demonstrate the feasibility and easiness of mapping different models adopted by national portals onto the proposed DCAT-AP.

## Discussion

The compatibility with existing standards was confirmed; the proposed DCAT-AP will take into account the specifications of INSPIRE and SDMX metadata models. It was also confirmed that schema.org schemas can be extended to the DCAT-AP. Representatives of the W3C consortium make part of the DCAT-AP working group and support the initiative. A field of the proposed profile allows the inclusion of information related to license conditions.

## Conclusions

**Participants were invited to take note of the consultation, take care that their stakeholders are aware of the process and provide feedback within the deadline of 15 July.**

## Interoperability and multilingualism at the level of the metadata: use of thesauri and controlled vocabularies

The Publications Office presented its activities supporting the interoperability of metadata at the level of the European Institutions and the European Union.

The Interinstitutional Metadata Maintenance Committee (IMMC) works towards the harmonisation and standardisation of metadata used in the Publications Office itself and in the context of EU interinstitutional data exchanges. The results of the work of the committee are the creation and adoption of more than 10 “authority tables” (controlled lists of metadata values), such as resource type, file type, language code, ... in all the 23 official languages of the EU. These tables, the maintenance of which is guaranteed by the Publications Office, are freely available, which will contribute to a wider harmonisation at the EU level.

The Publications Office is also responsible for the EuroVoc multilingual and multidisciplinary thesaurus (controlled vocabulary). While developed initially to cover the fields of interest of the European Institutions, it is now developing beyond this initial scope and is progressively offering a tool for indexing and retrieving an even broader range of information. Its last version, released in December 2012 (EuroVoc 4.4) contains

---

<sup>7</sup> [http://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](http://joinup.ec.europa.eu/asset/dcat_application_profile/description)

6883 concepts, 200 new concepts having been added to the previous release. EuroVoc can be downloaded free of charge from the EuroVoc web site<sup>8</sup>, and is also referenced in the EU Open Data Portal.

EuroVoc uses semantic web technologies and has already been converted to SKOS (Simple Knowledge Organisation System) as RDF resource. Its concepts are all referenced with a URI.

EuroVoc can be enriched with specific concepts, in order to broaden its generic scope and align it with domain specific thesauri. Integration of new concepts can be made directly (by integrating new concepts) or, preferably, indirectly, through the connection with existing specific vocabularies. An on-going project is working on the creation of semantic correspondences between EuroVoc and specialized vocabularies as GEMET (the thesaurus of the European Environment Agency, one of the thesauri used in INSPIRE), AgroVoc (FAO), ETT (Cedefop), TESE (EACEA), etc.

The results of these alignments will be made available on the EuroVoc and EU Open Data Portal-websites; permanent URIs will be attributed to aligned concepts; a searchable SPARQL endpoint will be made available. The Publications Office will guarantee maintenance and permanent availability.

### Discussion

The Publications Office reminded the participants of the importance of being made aware of the needs of the users, in order to incorporate as much as possible alignments with national/sectoral vocabularies into their offer. It was confirmed that EuroVoc and possible adaptations to national needs is flexible enough to accommodate the needs of individual Member States.

Concerning the current necessity of establishing bilateral licence agreements before using EuroVoc, the Publications Office indicated that a click-license download mechanism will be soon implemented. Users already licensing previous versions of EuroVoc are entitled to updates without any further procedure.

At the moment authority tables are not available outside the European institutions through a SPARQL endpoint. The Publications Office will investigate the possibility to change this.

The alignment of EuroVoc with the INSPIRE Spatial Data Themes, the mandatory vocabulary for INSPIRE, was confirmed. The tool for automatic alignment of EuroVoc with other vocabularies is the result of a specific development and cannot be made available to third parties at the moment.

### Interoperability at the level of data catalogues: the Open Data Interoperability Platform and Protocol

The Open Data Interoperability Platform and Protocol (ODIPP) is a protocol (the DCAT-AP presented earlier) and a platform providing a central metadata infrastructure storing harmonised metadata of datasets published in open data portals across Europe.

ODIPP is being developed through the Open Data Support project (ODS), financed by the European Commission, and is supposed to host by the end of 2015 at least 15.000 metadata referencing corresponding open datasets from European national and local portals. The platform will be accessible through the pan-European Open Data portal, publicdata.eu.

The platform will make it possible to reference a huge number of datasets in a harmonised way; the creation of metadata will be facilitated through the services offered by the ODS

---

<sup>8</sup> <http://eurovoc.europa.eu>

project (see point 4 on the agenda). The complete process includes the harvesting of metadata descriptions in existing open data portals, the transformation of metadata to RDF format, their validation, their harmonisation with the newly developed DCAT-AP, the publication of RDF metadata as linked open data. Search will be possible through a SPARQL endpoint. No strict barriers will be applied for the acceptance of metadata; quality controls will be applied, generating warnings on non-compliant metadata, thus implementing a soft enforcement of metadata quality. For the selection of the 15.000 metadata to be generated and made available on ODIP, priority criteria have been defined: referenced datasets should, in order of preference: cover all EU MSs; belong to a high value domain; be available under an open license; be already described with high quality metadata.

The ODIP platform, which is not yet public but already available for internal testing, was demonstrated. The objective is to make available a “self-service” platform to allow the transformation between specific metadata profiles and the DCAT-AP by users.

### Discussion

Regarding the relationships between ODIP and publicdata.eu, the Commission explained that ODIP will be a common metadata catalogue used by the pan-European open data portal, but also freely accessible and useable by other portals.

## **4. SUPPORTING THE PUBLICATION OF OPEN DATA: THE OPEN DATA SUPPORT PROJECT**

The Open Data Support (ODS) project, financed by the European Commission and operating from 1/1/2013 to 31/12/2015, has the general objective of improving the visibility and facilitating the access to datasets published on local and national data portals in order to increase their re-use within and across borders.

This objective is implemented through the provision of three services: publishing services, through the definition of a common metadata profile and a platform (DCAT-AP and ODIPP, see point 3 on the agenda) and the support to MSs for the creation of metadata; training and consulting services offered to MSs.

Regarding publishing services, the project will provide practical support for selecting high-value datasets, cleansing and modelling selected data and metadata, mapping metadata onto the DCAT-AP and uploading them on ODIP.

In order to establish an efficient relationship for the provision of services, Member States are invited to nominate a contact point, give access to datasets and metadata and their relevant documentation, to cooperate to define the mappings between specific metadata models and DCAT-AP. It will also be useful to establish where possible a sustainable cooperation by ingesting harmonised metadata and referencing the ODIP catalogue in MSs portals.

Training services are offered in the form of a portfolio of on-line modules, which will be made publicly available, and dedicated on-site trainings, to be agreed with the project. 36 on-line trainings in total have been planned. Both kinds of training will cover a wide range of subjects, from the more general (The PSI Directive and Open Government Data, Introduction to Linked Data, to RDF and SPARQL, ...) to more specific (Open Data quality and cleansing of data and metadata, Design and manage persistent URIs, open data and ODIP, ...) targeting different audiences (policy makers, IT strategists and s/w engineers).

In order to optimise the use of the available resources for on-site training, Member States are invited to indicate a suitable training date and location, and to specify the intended audience of the training. MSs should also take care of the training logistics, making available appropriate facilities, sending invitations, etc.

Consulting services are small-scale assignments (in the order of around 5 person days), to be agreed on request. The Commission will examine and whenever appropriate validate the request.

## 5. WRAP-UP AND CONCLUSIONS

The chair thanked the presenters and all the participants for their contributions to a very informative and rich meeting.

### Wrap-up

- The movement towards open data is progressing rapidly, and there is a need to further build and consolidate
- The Commission is committed to continue to work on the development of a pan-European open data portal and its corresponding platform
- A number of projects are supporting these developments (LOD2, ODS)
- The development of a common standard metadata model to be used at the European level (DCAT-AP) is an important building block; the EuroVoc multilingual vocabulary, including its potential extensions/linkages to specialised/national vocabularies, is a resource to be exploited
- The ODS project is a resource to be used for supporting mutual developments at the EU and MSs level

### Agreed actions

- Member States are invited to support Open Data as a Digital Service Infrastructure in the Connect Europe Facilities (CEF) programme, currently under discussion with the European Parliament and Council, together with an appropriate budget. Member States should also reflect on possible ways of guaranteeing the long term sustainability of the open data platform.  
The Commission will send the Commission proposal on CEF to the participants.
- Member States agreed to work towards a pan-European open data portal, and to cooperate with the project and the Commission in providing or extending the access to their open datasets.
- Representatives of Member States not yet in touch with publicdata.eu will get in touch with the portal contact point<sup>9</sup> in the week after the meeting.
- Member States will take note of the consultation on DCAT-AP<sup>10</sup>, and take care that their stakeholders are made aware of the process in order to provide feedback within the deadline of 15 July

---

<sup>9</sup> LOD2/publicdata.eu : [info@publicdata.eu](mailto:info@publicdata.eu) ; Sander van der Waal : [sander.vanderwaal@okfn.org](mailto:sander.vanderwaal@okfn.org)

- Member States will provide the ODS project with a contact point and all the information necessary for the support to the creation of metadata as well as the organisation of on-site training; MS will also reflect on how to best use the available resources in order to support open data activities at the national and local level
- Member states will send to the Commission existing documentation and any material regarding open data useful to be shared with the members of the group. The Commission will take care of making this material available to all members.

#### Next meeting

It was suggested to organise the next meeting back to back with the next meeting of the PSI group (September/October 2013, to be confirmed).

Copies:        Participants

---

<sup>10</sup> [http://joinup.ec.europa.eu/asset/dcat\\_application\\_profile/description](http://joinup.ec.europa.eu/asset/dcat_application_profile/description)