Report on the Budapest ePSI Platform Workshop

Submitted on 22 Oct 2013 by Martin Alvarez-Espinar

A few weeks ago, the ePSI Platform organized the first of a series of public workshops aimed at strengthening the PSI re-use community in Europe. This ePSI Platform workshop was held at the Hungarian W3C Office's headquarters, the MTA SZTAKI Research Institute, in the city of Budapest, Hungary.

Although this workshop was addressed mainly for the Hungarian community –not very active in Open Data–, it was publicly broadcasted on video, enabling the remote attendance from various countries.

A dozen of speakers from all around Europe were invited to take part in various sessions, addressing different topics from general introductory talks to in-depth discussions on the status of PSI re-use in different sectors and regions.
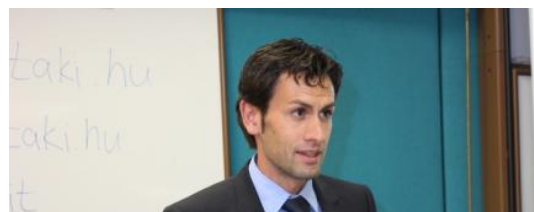
The EC's Vision

Márta Nagy-Rothengass, *Head of the Data Value Chain Unit (DG CONNECT) at the European Commission*, opened the event welcoming the attendees with a salutation both in Hungarian and English. She gave a talk titled Open Data in Europe, where expressed the commitment of the Commission to boost the PSI re-use in all the Member States.

As part of the data value chain, Márta said: "*data increase business intelligence and the efficiency of private and public sectors.*" And introduced the Europe 2020 programme, as a strategy for smart, sustainable and inclusive growth of business and public services for citizens, always bearing in mind issues such as privacy and seizing opportunities like the ones Big Data processing may offer – **Volume, Velocity, and Variety** are the aspects to take into account in Big Data. She showed some examples of the real benefit of these new economic opportunities.

After explaining her vision of the *data value chain*, Márta highlighted the current transitional state of the recently modified PSI Directive. She asked for public review of the new document, and regarding that the **consultation is open by November 22**.

One of the Data Value Chain unit objectives is to create a **friendly policy environment** in order to enable a smooth adoption of the Commission's recommendations by national governments. Thus, during the period of 2-3 years for the implementation of the PSI Directive in the Member States **the Commission will provide guidelines and tools** for them.

Márta unveiled some aspects of the future of PSI re-use in Europe, valuing the work of many local, regional and national Open Data portals, and commenting the challenge of multilingualism in the **future Open Data Pan-European to be launched**

in November 2013. Apart from the next ePSI Platform workshops, she made a call for contributions for workshops, and speeches at the 2014 European Data Forum to be held in Athens, and focused on Public Data.

I gave a brief overview of ePSI Platform and the opportunities this tool offers to strengthen the community. I showed some figures about the contents published on the platform and I asked for feedback on the PSI Scoreboard and the platform in general.

Open Access to Research Data

András Micsik, Researcher at MTA SZTAKI, presented the work this prestigious institute is doing in Hungary and Europe, always aligned with standards and the principles of openness. His talk treated a different approach of public data, **data used in sciences and research**. András explained the outlines of the **Open Access Mandate for Research in Hungary** that regulates researchers' publications. In a few words, researching works funded by the government have to be open. This is the next step towards open and reproducible research results. Since January 2013, the MTMT (Hungarian National Scientific Bibliography) Service collects the web addresses (URLs) of all publications. Results of public funded research projects must be public as well.

András also presented the **NextGeneration** project as the next step towards open and reproducible research results, through the publication of data and algorithms underlying the research.

Following the metaphor of the Tim Berners-Lee's 5-star classification of Linked Data, András explained the **5-star classification of online journal articles**: (1) Open access; (2) Enriched content; (3) Available datasets; (4) Machine-readable metadata for datasets; (5) Peer reviewed.

András mentioned a couple of projects SZTAKI has developed: lod.sztaki.hu is a metadata store which includes **11 million of RDF triples about cultural assets in Hungary** (movies, books, publications, etc.), and over **130,000 RDF triples about SZTAKI's publications**; and an interesting and intuitive Linked Data front-end called LODmilla.

PSI in Europe

Four panellists discussed about the status of PSI re-use in Europe.

Phil Archer, *W3C's Data expert*, was the first speaker to talk. He pointed that defining the status of the PSI in Europe is impossible, because the answer depends on the sector and region. Referring to the PSI Scoreboard, he commented that we cannot measure the real status with such as tools: "*Region of Flanders is not well represented in the scoreboard, they should have a higher mark.*"

Phil talked about technology, and his experience leading

successful works within W3C. He mentioned formats and technologies of Open Data. **Although its economic benefits are not reflected, Linked Data has a huge potential when governmental departments need to share and manage internal information** –e.g., the Italian Government and the BBC represent and manage Linked Data for internal purposes but they do not publish information, it is just for themselves for internal usage.

Emilio García, *Technical Advisor for the Spanish Government*, presented the government's measures to boost the PSI re-use in Spain. From the beginning, he showed his commitment with Open Data: "*If governments publish data, citizens will be able to understand the limits of the state.*" **He sees PSI-reuse as an opportunity, not only for citizens or industry, but for the public administration itself,** so with data and knowledge, citizens are able to offer better services the government should provide.

Emilio identified the gap between the demands of the potential re-users –open standards, permissive licenses, easy access, granularity, etc.– and the ways the government supply data, which are not always the best. In order to bridge this gap, the Spanish Government has performed different actions: development of laws and regulation; establishment of partnerships between the private and public sectors; embracing standard technologies –RDF, DCAT application profile to describe datasets. The results of it are under in the Spanish National Open Data portal (datos.gob.es):

- Over **6,000 datasets federated** from various public bodies at different governmental levels.
- In 2012, **3600-4400 jobs** were created in the PSI re-use industry, generating estimated **revenues of €330M-€550M**, figures higher than in the video games industry.
- It has been proposed a clear and **single common license framework**, to engage public bodies to embrace it.

Thanks to these measures for standardization, **datos.gob.es is ready to be connected with other initiatives such as the future Pan-European portal**.

The third speaker in this session was **Michal Kuban**, *Open Data Coordinator at Open Society Fund Prague*, and ePSI Platform Correspondent in Czech Republic and Slovakia. Michal presented the state of the art on PSI re-use in Czech Republic, explaining the **commitments the government acquired** after signing and submitting the plan for the Open Government Partnership.

The main commitments are: the adoption of an Act on Civil Servants, ensuring the **depoliticization and stabilization of the public administration sector**; enhance the access to data and information through the **publication of open data** –focused on specific datasets– under open licenses; and **developing an infrastructure** for that. Open Data has been adopted for the Czech Government in different

strategies such as the Digital Czech 2.0 programme for digital development, and other actions against corruption.

The Government of the Czech Republic have already created a methodology and defined the requirements for cataloguing public datasets. Also, it is expected the government to release an Open Data catalogue in December this year.

Up to date, there are three institutions planning different actions to promote Open Data and PSI re-use in the Czech Republic. Michal highlighted their most relevant projects: nasstat.cz (our state) is a watchdog project; vasmajeterk.cz (your property) with information on public auctions and sales –offering free and premium services; budovanistatu.cz to monitor how the government spends money.

Michal stated that expressions of interest are not enough: *"Without punishments there is no results."* Another problem he identified **is the lack of user-friendly Open Data guidelines for public bodies**.

Szymon Lewandowsky, *Policy Officer at the DG CONNECT*, complemented the panel with the vision of the European Commission. He mainly talked about the changes in the revision of the Directive on PSI re-use (published as 2013/37/EU). **This set of rules for Member States has to be transposed by July 2015**.

Szymon explained the main changes of the directive, including the **establishment of the genuine right to reuse public information** –thus all accessible information will become re-usable; now, the **charging rule by default is based on marginal costs** –i.e., in case a public body wants to charge for the information, the upper threshold is the cost of the operating cost of publication. Another change is the **broadening of scope**, now including cultural institutions -–i.e., universities, museums, archives, etc.

Complementing Márta's speech, Szymon explained the next steps of the Commission related to the creation and adoption of common guidelines for licensing, charging, publishing, etc. Also, he asked for provide feedback on the Directive consultation, open until November 22. Directive can be enriched, not only by governments, but grassroots and NGOs.

PSI in Hungary

Sandor Lederer, *CEO at K-Monitor,* an anti-corruption NGO with three pillars of operation: **anti-corruption advocacy, research, and Open Data Projects**. k-monitor.hu is a database for fighting against corruption. Information is collected through regular media monitoring –i.e., articles from the online press.

Sandor is involved in another project called *ahalo.hu,* to represent people's connections where the information is scraped from the web, so its extraction is very hard and

data has to be "manually" checked.

Sandor, as Hungarian expert, assessed the status of the Open Data in Hungary. He mentioned the EC Directive on PSI was transposed into the Hungarian FOIA (in 2003) –NGOs' suggestions were mostly not accepted. **This act does not include standards in publication**, neither sanctions for data owners who does not publish the demanded information.

The Hungarian OGP Action Plan was also cited in his talk. Proposed actions are aligned with the anti-corruption strategy of the government, thanks to Civil Society Organizations had to participate in elaborating the Action Plan.

Later this month, in the annual OGP Summit, the Hungarian Government will present an updated action plan, but there have not been reflected actions associated to NGO's suggestions on Open Data and FOI legislation.

Gabriella Ivacs, *Chief Archivist at Open Society Archives, and Central European University*, is an **activist advocated to open standards and open source** –she always have believed open source offers cheaper solutions, so this kind of solutions should be adopted by governments. Her association does not receive public funds. She commented they maintain a black-list with of Hungarian public bodies that uses and publish information in proprietary formats.

Her role of archivist in osaarchivum.org –a private archive– offered another point of view about PSI re-use. She said: "*There is a gap with the government data and other data such as the data of archives. The EC Directive doesn't cover this problem."* Archives are crucial to preserve cultural resources, as well as for research purposes. **Digital repositories apply specific standards that should be considered as the rest of Open Data**.

How to Open your Data

After the lunch break, the third panel gathered together three experts in PSI reuse and Open Data with different points of view.

Zdenka Zubcic attended in representation of *Tenderi Ltd*, a Croatian company that collects public procurement information and distributes it to its users. Zdenka –the only Croatian in the room– explained their business model based on providing email services after processing PSI from the Croatian government.

Tenderi, as well as other similar companies, **are required to pay up to €4,000 per year** (and probably more in the future) **to get the information about public contracts**. This high fee seems to be arbitrary since the Croatian administration does not provide a clear explanation about this charge. As Zdenka said: "*Nobody in Croatia sees this [payment for the information] as a problem.*" Read the full case.

Szymon and Marta were interested for this case where Croatia –our newest Member State– is

involved. The Commission will follow up the case, so I hope in the next event Zdenka can be invited again and presented some advance in her business.

Zoltan Varju, *a Hungarian computational linguist interested in translations from Precognox*, presented his company's work with data and explained that openness is already a great achievement for their work. He remarked they are **just** interested in data: "*It doesn't matter in which format data is published, just publish it, please.*"

As Semantic expert, Zoltan prefers processing information in Linked Data, but the lack of information puts this requirement in a second term. He mentioned several times: "***data in one-star format is enough.***"

Phil Archer explained how to open data according to W3C's recommendations. Showing a slide of the Tim Berners-Lee's five-star Open Data classification, Phil stated that **the most difficult status to achieve is the first one (one start)**. Convincing people to expose data is the hardest step.

W3C is merging two activities, eGovernment and Semantic Web, into one "Data Initiative". During these years involved in Open Data, he recognized that they are trying hard to achieve the five-star status in the Open Data initiatives, but "**CSV is King.**"
"Spreadsheets are everywhere and many people knows how to manage them, so we should work more on this."

Phil commented several times the need of using **URLs as unique identifiers for resources**. Many other participants agreed with this statement.

Open Data Support

After this theoretical panel, Nikolaos Loutas, *Manager at PwC Technology Consulting and member of Open Data Support*, presented a complete tutorial on how to develop open data initiatives. He talked about **standards and different approaches in Open Data strategies**, and why metadata and Linked Data is important for the interoperability among public bodies, as well as for the management of information.

Nikolaos reviewed some successful cases in Europe, and explained step-by-step how to define and develop Open Data strategies. Nikolaos encouraged the public to follow the Open Data Support programme and take part in their conversations.

See his complete tutorial.

We would like to thank all the speakers, the public and organizers to make this event possible. We hope it helps, at least, in the development of PSI reuse in Hungary. See you in the next event!!!

File:

 Márta Nagy-Rothengass

- [Martin Alvarez-Espinar](#)
- [András Micsik](#)
- [First panel: Emilio García, Szymon Lewandowsky, Michal Kuban](#)
- [Emilio García](#)
- [Michal Kuban](#)
- [Phil Archer](#)
- [Szymon Lewandowsky](#)
- [Sandor Lederer and Gabriella Ivacs](#)
- [Sandor Lederer](#)
- [Gabriella Ivacs](#)
- [Panel: How to open and reuse Data](#)
- [Zdenka Zubcic](#)
- [Zoltán Varjú](#)
- [Phil and Zoltán during the panel](#)
- [Nikolaos Loutas](#)
- [Public after the workshop](#)