

Report

Technical workshop on the goals and requirements for a pan-European data portal

Wednesday 3rd November 2010, 09:45 – 16:00
Euroforum Building, room 2292
10 rue Robert Stumper – L2557 Luxembourg

Table of Contents

- 1. EXECUTIVE SUMMARY.....2
- 2. INTRODUCTION.....3
- 3. STATUS OF EXISTING EU 27 NATIONAL GOVERNMENTAL INITIATIVES3
- 4. REASONS FOR AND BENEFIT OF A PAN-EUROPEAN DATA PORTAL.....3
- 5. CHALLENGES CONCERNING A PAN-EUROPEAN DATA PORTAL.....5
- 6. IMPORTANT ISSUES FOR A PAN-EUROPEAN DATA PORTAL5
- 7. SUMMARY AND OUTREACH7

1. Executive Summary

On the 3rd of November 2010 the European Commission organised in Luxembourg a technical workshop on the goals and requirements for a possible pan-European data portal. Experts with practical experience in their respective countries were invited to share their experiences and ideas.

The experts consider that such a portal would add value to existing regional and national initiatives by improving transparency on issues of EU-wide interest, providing evidence for better policy making, improving the efficiency of data-dependent administrative and business processes and stimulating economic development through EU-wide reuse of data.

Several issues of legal, technical and socio-political nature must be addressed for such a portal to function effectively, among them the need for high level political support, the systematic adoption of reuse-friendly data licences, the promotion of established data standards for maximal interoperability and the organic involvement of European software developers and data-literate citizens.

A pan-European portal should be able to expand rapidly in breadth (thus fostering the interest of the public with large numbers of relevant datasets) while at the same time also showing the value of deeper data integration, starting from a core set of statistical, financial, geospatial data of high quality. Agile prototyping and development models are recommended, given the extremely fast pace at which data initiatives are developing in Europe.

A small working group should be created to drive the issue forward and meet regularly to identify more precisely technical requirements. The group should connect with other open data stakeholder groups established at the national or European level and contribute to the definition of European datasets, government open data conferences and software development competitions, with first results visible and publicised by mid-2011.

2. Introduction

On 3rd of November 2010 a technical workshop on the goals and requirements for a possible pan-European data portal took place in Luxembourg organised by the European Commission.

The goal of the meeting was to start a discussion with recognised experts in order to determine:

- the expected benefits and value added of building a pan-European data portal;
- its technical and operational requirements;
- the next steps and activities to consider to evaluate its feasibility

Eight experts were invited. They were selected based on their recognised status as subject matter experts and their practical experience with existing or soon to be launched regional or national government open data projects:

1. Jose Manuel Alonso, CTIC and W3C, Spain
2. François Bancilhon, Data Publica, France
3. Bastiaan Deblieck, TenForce, Belgium
4. Jarmo Eskelinen, Forum Virium, Finland
5. Jonathan Gray, Open Knowledge Foundation, United Kingdom
6. Michael Hausenblas, DERI, Ireland
7. Martin Kaltenböck, Semantic Web Company, Austria
8. Nigel Shadbolt, University of Southampton, United Kingdom

3. Status of existing EU 27 national governmental initiatives

In order to ground the discussion on factual and current information, an overview of existing national initiatives was given: in the European Union there already exist five nation-wide governmental data initiatives that provide data portals to the public. Thirteen additional European countries are already working on national data portals or have concrete plans to do so.

In addition, there exist across Europe several regional government data initiatives as well as private initiatives that provide data portals as well as software application competitions on open government data to raise awareness and stimulate reuse.

4. Reasons for and benefit of a pan-European data portal

The workshop participants identified four main value added to be expected from a pan-European data portal.

Based on the experiences of Member States, two of them are related to good governance: **transparency** and **evidence based policy**.

Transparency means that all EU citizens would be able to access detailed information on EU-wide issues such as taxation, funding, crime, environment, energy, transportation, etc... and become better informed citizens in the process.

Evidence based policy means that EU institutions and Member States would be able to ground their policy decisions on empirical data open to public scrutiny and draw legitimacy in the process.

The other two reasons are **efficiency** and **economic development**.

As an example of **efficiency**, it was pointed out that the government of a European region has already represented as open data all its physical resources (buildings, facilities, etc...). This data set is now being jointly used by several administrations all of whom previously had to come up with and maintain their own database of resources

As an example of **economic development** the experts mentioned geo-location and public transportation applications that have been already built on the reuse of open data.

Further concrete reasons have been mentioned in support of the notion that a pan-European data portal would add value to the existing or upcoming national portals:

A) For European citizens

- Single point of access on European information
- Enabling services for citizens that live at country borders and/or work abroad
- knowledge of successful open government data initiatives in some Member States can drive further initiatives in other Member States

B) For administrations

- Improvement of interoperability across processes thanks to greater availability of data
- Improved comparability of EU 27 information and data
- Reduction in administrative costs
- Avoiding / cutting existing costs of re-publication of official information
- More efficiency in servicing Freedom of Information requests
- Involvement of European citizens (crowd sourcing approach) can have positive effects on transparency and quality of data.

C) For economic development

- Planning and monitoring resource for companies operating across EU borders
- Driving the European innovation process
- Driving force for European economy (information technology, new location based services, analyzing services et al)
- Harmonisation of standards and guidelines for open government data across Europe

5. Challenges concerning a pan-European data portal

Several challenges have been identified and should be considered in future work on this topic. They fall under three main categories: legal/economic, technical and socio-political.

1. Legal/Economic

- a) differences in legislation across Member States (for example, data protection laws or laws preventing certain datasets to be stored outside a certain region) need to be addressed;
- b) high value data sets are today distributed for a fee and represent a source of income to some administrations

2. Technical

- a) mechanisms for versioning data sets that are often updated
- b) mechanisms for measuring data reuse, methodologically sound indicators of success and return on investment
- c) ease of data exporting by harmonisation of data management tools with the formal requirements of national authorities
- d) support for multilingual data search and interpretation
- e) identification of an appropriate portal architecture to support a seamless transition between federation of remote resources and integration of core data sets
- f) appropriate resourcing to ensure the stability of the portal, its adequate performance under the expected access patterns and its scalability

3. Socio-political

- a) need to secure high level political support and funding
- b) need to involve citizens and support community building among users of open data
- c) need to involve application providers and software developers as beneficiaries and value multipliers of an open data infrastructure

6. Important issues for a pan-European data portal

The experts identified two orthogonal strategies that would maximise the value added of a pan-European portal (particularly in the short term):

- A) very quickly publish a catalogue of a very large number of diverse datasets from member countries in order to achieve breadth of coverage and thus sustain citizen interest;
- B) deeply integrate a small set of very high quality datasets demonstrating immediate value and, in time, capable of acting as a scaffold for the integration of many other datasets. Candidates in this second role are geospatial, transportation, statistical and financial datasets.

The participants of the workshop furthermore identified **appropriate data licensing** at the source as the conceptual precondition for any value to be extracted by data reuse (developers will not reuse data if it is not clear that they have the right to do so). This appears to be mostly an issue of educating data publishers on the selection of an appropriate license. There may be however contexts in which this might turn out to be a legislative issue, to be considered in the context of the review of the Public Sector Information Directive.

There was also consensus on the fact that a clear licensing policy should be created and enforced on a pan-European data portal so as to maximise the opportunity for data reuse.

The experts suggest that very **high level political support** would be required to establish a culture in which government data are always published as open data unless there are good reasons to the contrary.

The experts report on their experience with '**applications challenges**' as an effective mechanism for involving software developers in open data initiatives. Similarly, opinion campaigns as have been run by major newspapers have proved effective in stirring the interest of the general public.

No clear consensus was reached on the desirable architecture of a potential pan-European portal at this time. This reflects the dichotomy of the recommendations above.

On one hand, deep integration of high quality datasets would require for them to be copied wholesale from the member state publishers and then actively curated and interlinked on an EU-wide data portal (the Eurostat model).

On the other hand, indiscriminately copying and cleansing every dataset from every EU country data publisher would impose highly non-trivial operational costs on a EU portal. For this reason it would be best if, for most datasets, the EU portal acted only as a data register, with links to the respective national initiatives where the actual data would reside and be maintained (the Europeana model).

Deciding which datasets should be imported and integrated and which just registered and linked to could be solved by creating several 'mechanisms of trust' for the datasets. Such mechanisms could be based on

- quality assurance certificates (awarded by the EC or by officials of member countries) as well as
- crowd-sourcing mechanisms where end users rank and comment on data sets.

Clear consensus was reached on the fact that an EU data portal should be ecumenical in accepting all structured data formats and not dictate a specific format as a requirement. That notwithstanding, existing data formats (and open data formats in particular) should be promoted and used whenever

possible. There was also a consensus on the facts that there is a clear progression in the usefulness and ease of integration of various formats, with Linked Open Data (open data published on the web by means of shared URI) being found at the end of that progression.

The 'five star' methodology outlined by Sir Tim Berners Lee was recommended as a pragmatic and useful guidance in that progression:

- 1 star - Put your data on the Web (any format)
- 2 stars - Make it available as structured data (e.g. Excel, CSV, instead of PDF)
- 3 stars - Use open, standard formats (e.g. XML, RDF)
- 4 stars - Use URLs to identify things (so people and machines can point at your data)
- 5 stars - Link your data to other people's data (Linked Data approach)

Two additional specific technical roles have been identified as a source of added value for a pan-European portal:

1. Becoming an authoritative registry of conceptual models (e.g. the definition of a 'vehicle' or an 'employment skill') to be (voluntarily) used by national initiatives for structuring their own data. This would make it easier to explore issues of interest across data sets contributed by several independent data providers.
2. Becoming an authoritative provider of codes or identifiers (e.g. the identifier for 'the Grand Duchy of Luxembourg') to be (voluntarily) reused by national initiatives for structuring their own data (e.g. for a Portuguese administration to record the number of Portuguese nationals residing in 'the Grand Duchy of Luxembourg'). This would make it possible for information about the same entity to be retrieved across independently developed datasets.

7. Summary and outreach

In conclusion, the participants agreed that a pan-European data portal with the characteristics described above would add value to open data initiatives from the Member States. Such an initiative should be pursued without delay in order to exploit the current momentum of open government data initiatives across Europe.

The participating experts agreed that data reuse from regional/national portals has not yet been massive but observe that we don't yet have a well established methodology for appropriately measuring reuse and return on investment.

As for next steps, the experts expressed reservations with respect to treating the establishing of a pan-EU data portal as a top-heavy IT project (feasibility study followed by request for tender, followed by project award, followed by implementation). In their opinion, such a process, while respectable in other contexts, would be inappropriate in this one because open data EU developments are currently taking place very rapidly and precious momentum would be lost if such a formal approach were followed.

What they recommend instead is for a first core of datasets to be collected in an EU environment (possibly in the context of existing projects funded by the Commission) and that a broader

community of civil servants and developers should be mobilised to experiment with this core. This core could then informally become a proof of concept of what a bona fide pan European portal could become.

The experts suggested that a first such experiment could be attempted as early as Q1 of 2011 with a second, higher visibility, event organised later in the year.

Furthermore it was pointed out that a small working group should be created to drive the topic of a pan-European data portal as efficiently and flexibly as needed. Broader awareness building across Europe for all open government data stakeholders is also needed and could be supported by the European Commission in the form of open data competitions and/or events as for example a European open government data conference and/or the participation of EC representatives at national open government data events and conferences to bring together European stakeholders.