TOWARDS OPEN GOVERNMENT METADATA

September 2011

1. BACKGROUND

Governmental agencies are considered to be the most significant data owners and providers in modern societies. The sheer volume and wealth of this data makes apparent the potential benefits of reusing, combining, and processing governmental data. However, administrations typically express reluctance to make their data and metadata (information about the data) available, for various cultural, political, legal, institutional and technical reasons. They keep data and metadata within their legacy systems, fenced and isolated.

The Open Government Data movement which recently becomes visible in various countries[1] promotes openness for public sector information. This is often materialised with public sector catalogues becoming available through governmental portals where public agencies make their datasets available to the general public. There are two main prerequisites for establishing an Open Data policy: a) high availability of the data, preferably in formats that are both human and machine friendly and b) openness, meaning allowing reuse of the data for all purposes and with no restrictions.

Interestingly, and despite the clear move towards Open Data, there is currently very little discussion, if any, on Open Government Metadata. Is such discussion useful and relevant in the more general topic of eGovernment interoperability at a European level?

2. SEMANTIC INTEROPERABILITY AND METADATA MANAGEMENT

According to the EU Digital Agenda[2], the lack of interoperability has been identified as one of the seven major obstacles to the digital economy, while semantic interoperability barriers are identified as a major element of this interoperability problem. In the European Interoperability Framework[3] semantic interoperability is described as relevant to *"… the meaning of data elements and the relationship between them. It includes developing vocabulary to describe data exchanges, and ensures that data elements are understood in the same way by communicating parties.*" In the EU eGovernment Action Plan[4] semantic interoperability is mentioned *"…as an essential precondition for open, flexible delivery of eGovernment services*".

Commonly agreed and (re-)used **Semantic Interoperability Assets** can be perceived as an important infrastructure for facilitating semantic interoperability.

We define Semantic Interoperability Assets as highly reusable metadata (e.g. xml schemata, generic data models) and reference data (e.g. codelists, taxonomies, dictionaries, vocabularies) which are used for eGovernment system development[5].

---

[1] http://lod2.okfn.org/eu-data-catalogues/

[2] http://ec.europa.eu/information_society/digital-agenda/index_en.htm

[3] http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

[4] http://ec.europa.eu/information_society/activities/egovernment/action_plan_2011_2015/

[5] For simplicity reasons we refer to the **Semantic Interoperability Assets** here as "*Metadata*", although not all metadata qualify as Semantic Interoperability Assets (only those which are highly reusable) and there are semantic interoperability assets that are not of metadata type (reference data).

This particular type of *Metadata* as defined above is an important asset for all MSs as they provide the basic informational blocks for developing eGovernment applications and systems. This *Metadata* can be defined once and can then be shared and reused. This reuse results in various benefits including:

a) Increase interoperability between systems that share common or similar *Metadata*

b) Reduce development costs for eGovernment systems

c) Reduce integration, and sharing of information costs especially across agencies and ministries

d) Avoid reinventing (and paying for) the wheel dozens of times in separate projects

Furthermore, if this *Metadata* is discussed, aligned, agreed and reused at a European level, then Member States will be prepared for providing European cross-boarder public services.

To harvest these benefits, *Metadata* should be carefully documented and managed.

For managing this valuable resource, Member States should start identifying and documenting in a systematic way the *Metadata* they use. However this is only the first step towards establishing concrete *Metadata* management policies and treating *Metadata* as a valuable resource. Some countries (e.g. Denmark, Finland) already develop advanced and promising Metadata Management Systems, as specialized web portals from where users can find the *Metadata* they need for their projects. These portals are often part of broader national Information and Metadata Architectures.

To provide insights for the current level of advancement, as well as to help Member States to identify what needs to be done for improving their own *Metadata* management policies, in the following part, a maturity schema for *Metadata* management is proposed.

3. MATURITY OF METADATA MANAGEMENT

Five levels of maturity for metadata management are listed below[6]:

Level 1: **Metadata Ignorance** – *Metadata* is not documented, mainly because administrators are not aware of its importance. This situation results in serious semantic interoperability (IOP) problems within each country as developers use ad hoc data models, metadata, codelists, taxonomies, etc for developing eGovernment systems which are very hard and expensive to interoperate later due to competing data specifications.

Public administrations should become aware of the importance of Metadata in eGovernment and the need for coherent relevant management policies.

Level 2: **Scattered or Closed *Metadata*** – *Metadata* may be partially documented but a) not in a centralised and structured way or b) it is not available and accessible under an open license framework, in other words as "Open Metadata" for developers to share and reuse.

Public administrations should organize the scattered Metadata in structured repositories, catalogues or libraries and provide open access to the collected resources.

Level 3: **Open *Metadata* for Humans** – *Metadata* is documented and becomes available as "Open Metadata" for reuse, but are not systematically published in a reusable format, e.g. may only be available in .pdf or .doc documents.

---

[6] Following a similar – though in a different context – approach to Tim Berners-Lee's five star Open Data schema, http://inkdroid.org/journal/2010/06/04/the-5-stars-of-open-linked-data/

Public administrations should provide services to query, browse and export their Metadata in a machine-readable and preferably non-proprietary format (e.g. CSV, XML).

Level 4: **Open Reusable *Metadata*** – *Metadata* is centrally managed, and published as "Open Metadata", in a machine readable format and/or an API is provided for computers to access, query and reuse the available metadata repositories, catalogues, libraries, etc. Electronic Metadata Management Systems are introduced - e.g. platforms like SEMIC.EU[7], Digitalisér.dk[8], Yhteentoimivuus.fi[9], NIEM.gov[10] in the USA – to support metadata architectures and policies. Through these systems users can find, browse, compare, download and use *Metadata* that better fits their needs and projects.

Public administrations should consider applying linked metadata policies, including use of RDF to document their Metadata, persistent design, use and maintenance of URIs[11], linking to external vocabularies/schemata, harmonize their resources to third parties' resources etc.

Level 5: **Linked Open Metadata** – Semantic Assets are documented using linked data principles[12] and are managed by advanced Metadata Management Systems. At this level a graph of interlinked concepts emerge, as *Metadata* definitions systematically use definitions from other vocabularies. Each concept represented in this graph corresponds to a unique URI.

Currently, very few EU Member States have reached level 4, whilst most of them could be rather placed in levels 1 and 2.

4. RECOMMENDATIONS AND CONCLUSION

The European Commission through the ISA Programme[13] and more specifically through the Semantic Methodologies Action[14] promotes the awareness of proper *Metadata* management as an instrument to facilitate semantic interoperability amongst the MSs. This awareness is of critical importance towards the provision of seamless cross-border services and for enabling transparent and user-friendly European Public Services.

Best practices in this area are already available from Denmark and Finland at the national level where metadata management becomes a major element at national interoperability and information architectures, and from SEMIC.EU and NIEM at the European and international level respectively. Sharing experiences and reusing available solutions remains at the heart of the ISA Programme and will be systematically promoted.

The ISA Programme also supports work towards the definition of the Asset Description Metadata Schema (ADMS)[15] which can play the role of a common language for describing and publishing "Open Government *Metadata*". Moreover, a federation of metadata repositories will be based on ADMS[16]. Member States are invited to participate in this work to collectively draft the ADMS specification.

---

[7] http://www.semic.eu/semic/

[8] http://digitaliser.dk/

[9] https://www.yhteentoimivuus.fi/

[10] http://www.niem.gov/

[11] http://www.cabinetoffice.gov.uk/resource-library/designing-uri-sets-uk-public-sector

[12] http://www.semic.eu/semic/view/snav/Conformance/LOD-metadata.xhtml

[13] http://ec.europa.eu/isa/

[14] http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm

[15] http://www.semic.eu/semic/view/documents/federation-of-repositories.pdf

[16] http://share-psi.eu/papers/DERI-Peristeras-init-federation.pdf

Another relevant line of work related to Core Vocabularies[17] has already started. The goal is to provide a first small set of generic and highly reusable vocabularies that could be used as core *Metadata* across the MSs. These vocabularies are also developed in a collaborative fashion with the Member States.

In a similar way the Open Data movement and projects all over the world pursue openness and availability for public sector information, governments should also pursue openness on government *Metadata*.

Public administrations should be aware that *Metadata* is an important asset for eGovernment systems development and as such should be carefully and professionally managed. This Metadata Management includes the following actions:

- **Public administrations should identify important *Metadata* with reuse potential for developing eGovernment applications and systems.**

- **Public administrations should systematically document this *Metadata*.**

- **Public administrations should check their *Metadata* to identify inconsistencies, overlaps, conflicts and opportunities of harmonization.**

- **Public administrations should make their *Metadata* open for reuse.**

- **Public administrations should provide their *Metadata* both in human and machine readable formats.**

- **Whenever possible, *Metadata* definitions should reuse existing specifications to avoid duplication of work, and overlaps. Linked data approaches and techniques applied for *Metadata* modelling can substantially improve the quality, cross-reference, integrity and reusability potential.**

Based on the above recommendations, the European Commission, through the ISA Programme, will identify opportunities for *Metadata* alignment at the European level, animate and facilitate relevant discussions. This work will follow the directions and recommendations provided by the European Interoperability Framework with the goal to facilitate and enable the seamless provision of European Public Services.

ISA Contact:
Vassilios Peristeras, DG DIGIT, ISA Unit, vassilios.peristeras@ec.europa.eu
Montoyer (MO) 34, Floor: 3, Office: 103; B-1049 Brussels, Belgium
Tel: +32 (0) 2 29 81014

---

[17] http://www.semic.eu/semic/view/documents/egov-core-vocabularies.pdf