Parallel Sessions D: Core Vocabularies and Grammar for Public Sector Information Management & Interoperability

Facilitators

Chris Harding (TOG) [presentation]

Yannis Charalabidis (UAEGEAN) [presentation]

Scribe: Harris Alexopoulos (UAEGEAN)

The first presentation of this session was given by Yannis Charabidis who presented the motivation on the core vocabularies based on:

- Governmental resources metadata: an important ingredient to achieve standardised service description, retrieval, composition and cross-organisational collaboration

- Existing schemas present certain shortcomings:

    - Pan-European standardisation of service metadata is essential. The ISA Core Standardisations are in the right way, but still cover very basic data.

    - Metadata sets form parts in larger Government Ontologies, with applications in public knowledge management and semantic interoperability (Opendata, etc)

Yannis continued on describing the sources of eGovernment knowledge in terms of data, processes, workflows, web services, knowledge bases, documentation, legal elements and other highlighting the importance of citizens' participation and the ways this knowledge could be managed ontologically based on the UKs standard eGSM+ and the Greek Interoperability Framework (eGIF).

Then he emphasised on the need for describing resources in a systematic way with great acceptance from the audience and he continued on the metadata definitions within a service registry and an analysis of the core metadata fields of services, documents, XML schemata, legal frameworks, Information Systems and public bodies.

Finally, he concluded on the following:

- Standardization of an ontology-based extended metadata set embracing the Government knowledge, from services and documents to code lists and information systems, which:

    - Effectively supports the Greek e-Government Interoperability Framework and the Interoperability Registry Prototype implementation.

    - Formalizes the exchange of information between portals and registries.

    - Includes metadata around service delivery scenarios that can guide any business process re-engineering effort in the public sector.

- Future steps include exploration of how such a metadata set can:

    - (a) embrace policy modelling and intelligent governmental service front-ends

    - (b) be further elicited in order to take into account citizens' feedback when designing public services.

◦ (c) automated metadata filling

The final conclusion of automated metadata filling triggered a very interesting discussion on if this could ever be happened among the audience. An example of how this cannot be happened has been given by transferring his experience of such an initiative. But still alternative ways in how automated metadata filling could be achieved were also presented.

After that, a PwC employee argue that Core vocabularies are domain specific and that point to point mappings are very time consuming in terms of velocity of fields (tens of hundreds). So, adaptations is not the solution – Standards and conformance to them is.

Second and final presentation of this session by Chris Harding underlined once more the need for a data classification system. Chris put emphasis on the need of not only vocabularies but the existence of a grammar too. A lot of classifications exist (vocabularies) in different formats (rdf, relational databases). The thing is how we can establish a common grammar in the world of structure data.

"Data is the new oil" stated by Chris: explaining his argument "Extracting and refining minerals requires work". So does providing and consuming data.

He continued on presenting the information reuse issues that have been identified in previous SHARE-PSI workshops:

- Even though valuable data is released, additional effort is required to energize external stakeholders to create something with it (Samos).

- Open data, even when freely available, is not free to use since so much time has to be spent cleaning it up, converting it, integrating and maintaining it (Lisbon).

- There is a need to describe the quality of data in a consistent manner if potential consumers are to make informed choices (Timişoara).

- Raw data is of almost no value except to a small number of people with the skills and motivation to work with it. Commercial re-users add value to the raw data by analysis, transformation, and enhancement (Krems).

He carried on by analysing the ways oil produces value giving examples on the open data use case. "Understanding the value of minerals in their raw state can be hard, So can understanding the value of raw data" he said. Classification enables analysis and aids understanding and use. **We Need a Data Classification System** (a) To enable analysis, (b) To aid understanding and use and (c) To make interoperability possible.

Chris presented the cycle of knowledge (Analysis → Understanding → Integration) based on data units. A unit of data that is considered in context to be indivisible. [ISO 2382]

He gave examples on different data units in different contexts (poem, structured data, data in the Internet of Things) and he concluded that

- Some applications are heavily used: They justify major development effort.
- Others are hardly used at all: They are only viable if they need minimal effort.
- Software productivity and rapid development enable a "long tail" of lightly-used applications
- The customer experience application is near the end of the tail

After that Chris continued his presentation on where we stand nowadays in terms of vocabularies and grammars.

Many vocabularies exist

- Flat: Dublin Core, ISA core vocabularies
- Domain specific: INSPIRE for geo-information, UNSPSC (United Nations Standard Products and Services Code), SNOMED-CT for health and
- Public Service Vocabularies: Data Catalog Vocabulary (DCAT) and the DCAT Application Profile, Core Public Service Vocabulary: The Italian Application Profile, Controlled Vocabularies and Metadata Sets for Public Sector Information Management

The Basic Grammar is needed for consistent interpretation by people and especially for interpretation by machines. Basic Grammar concepts are:

- Relational Database and Data Modeling
- Resource Description Framework

Chris finally concluded on the following:

- Data is the new oil, but full exploitation of its potential requires a data classification system
- The Data Classification System Should
    - Enable use of the existing vocabularies that have been developed by industry bodies and standards organizations
    - Provide a basic grammar for data descriptions
    - Be consistent with relational database usage
    - Be able to accommodate other data representation approaches
    - Use RDF to facilitate semantic processing
- The Open Data Element Framework which is currently in technical review
    - Index and method for using it to classify data elements
    - Will meet the requirements
    - Based on the UDEF but Has roles as well as object classes and properties and Has plugins

A discussion started about the oil phrase" " There are bad connotations about oil. Yes, but the point is to understand the value of structuring – analysing and using data. Could be the water circulation or something else in order to show the potential growth the will come out from data.

What do you mean by grammar ? More appropriate grammatical properties…