CEUR Workshop Proceedings

http://ceur-ws.org

ISSN 1613-0073

ORES-2010
Ontology Repositories and Editors for the Semantic Web

Proceedings of the 1st Workshop on Ontology Repositories and Editors for the Semantic Web

Hersonissos, Crete, Greece, May 31st, 2010.

Edited by

Mathieu d'Aquin, The Open University, UK
Alexander García Castro, Universität Bremen, Germany
Christoph Lange, Jacobs University Bremen, Germany
Kim Viljanen, Aalto University, Helsinki, Finland

A pan-European repository:
SEMIC.EU as the point of reference for eGovernment ontologies

Klaus Reichling, Martin Luts, Renke Fahl-Spiewack

Semantic Interoperability Centre Europe,
]init[AG für digitale Kommunikation, Berlin, Germany
ELIKO Competence Centre in Electronics-, Info- and Communication Technologies, Tallinn,
Estonia; Department of Informatics, Tallinn University of Technology, Estonia

Abstract. The Semantic Interoperability Centre Europe (SEMIC.EU) is a Service of the European Commission to foster seamless data exchange. One of the principal objectives of the initial project setup was the creation of a repository of semantic interoperability assets for pan-European eGovernment projects. eGovernment arrived at the opportunities of semantic technologies. A recent poll by SEMIC.EU showed: Over 80 per cent of respondents indicated they have used or are going to use ontologies in the next 12 months. European public administration has realized the potential of ontological solutions for its purposes.

Keywords: semantic interoperability, European Commission, eGovernment, repository, ontologies, ISA, IDABC

1   Introduction

The Semantic Interoperability Centre Europe, SEMIC.EU, is a service offered by the European Commission: It facilitates the reuse of syntactic and semantic assets needed for semantic interoperability. It was initiated in the framework of the programme 'Interoperable Delivery of European eGovernment Services to public Administrations, Businesses and Citizens' (IDABC) and is carried on under the successor programme "Interoperability Solutions for European Public Administrations", ISA[1].

A standardised clearing process, supported by platform functionalities, governs the evolution of the data models, XML schemas, code lists, taxonomies and ontologies which are shared through the open repository.

---

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

The service addresses seamless data exchange in Europe, particularly in public administration. It is understood that countries have different administrative, technical and linguistic backgrounds. Solutions for the technical, semantic, and organisational dimensions of interoperability are needed. SEMIC.EU's focus is on the vital semantic dimension of interoperability.

As a collaborative service and with a repository of assets for semantic interoperability, SEMIC.EU is a catalyst for the preservation of meaning in the data exchanged at all levels of administration.

2   The asset principle

The SEMIC.EU web portal comprises an open repository of "interoperability assets". The term "interoperability asset" describes a resource that supports the exchange of data in distributed information systems, roughly to be understood as data models that help overcome differences in the systems involved in the exchange of certain data.

SEMIC.EU's most common semantic specifications used in interoperability assets are code lists, taxonomies and also ontologies including their respective mappings.
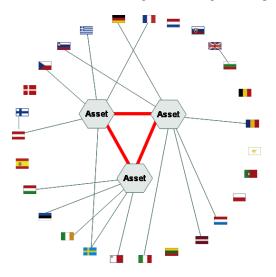

Fig. 1 Multilateral assets and mappings

They are collected, quality checked and made available for download and reuse in the web repository. The assets are taken through an open clearing process [SEM01].

The common features of all these interoperability assets are their design and suitability for concrete data exchange and data modeling purposes in a wider eGovernment context.

2.1   Life cycle of an interoperability asset

SEMIC.EU has defined a collaborative process of incremental quality enhancement and adaptation to a comprehensive set of criteria. Any asset submitted to the repository enters this *clearing process*.

An asset is a package of related documents that form a specification. The contained code list, taxonomy, XML schema or ontology is just a component of the asset as a package.

Fig. 2. Screenshot: Asset Repository

From a rough draft to a certified recommended means of facilitating data interchange, the process requires actions from the owner of the asset, SEMIC.EU as the broker and quality assurer, the general public for consultation and enhancement and expert communities for validation and approval. Technical, semantic and pragmatic aspects are assessed during all steps. Assets of interoperability and their respective releases are scrutinised by a staff member - the clearing process manager, as well as through public consultation and by experts in the domain of the asset.

During the process, assets are labelled with their current state of quality assurance. The successive stages of the process depend on actions by the asset agent (the person in charge of the asset) and the clearing process manager, supported by the platfrom (see chapter 3).

Although every asset has its own, individual life cycle, the paths and development stages are defined by the clearing process of the SEMIC.EU platform [SEM02].

Assets pass a number of successive stages consisting of quality assessment and feedback. A certain label indicates the maturity level of an asset (or one of its releases) – known as the "asset/release state":

Unpublished

For an unpublished asset to be submitted to SEMIC.EU, there are no technical or formal requirements other than the registration of the asset agent.

The asset agent must submit a set of metadata and select a topical category (domain) for an asset, but has not requested its publication yet.

The asset is not displayed in any search results or lists outside the asset agent's personal area on the website. It can only be viewed by the agent and the Clearing Process Manager.
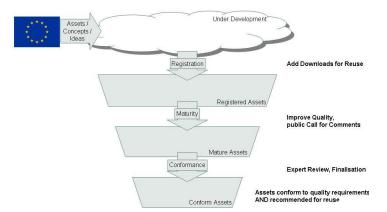

Fig. 3. Basic three-step scheme of the clearing process

Development

There are still no technical requirements, but the asset is now published and visible to all users.

To put an asset "under development", the asset agent needs to request its publication which will then be approved or rejected by the clearing process manager. This is done via repository functionalities.

This state is chosen by the agent if he or she wants to get feedback and support by other users. With an asset "under development", asset agents can make use of community features and seek advice from others in the development of their data model. For each published asset, a dedicated forum thread is created at www.semic.eu.

Registered

This is the first state of the actual clearing process since for this state to be reached, an asset needs substance (i.e. a first release):

The asset agent must upload the actual artefacts (data model files, a licence specification and other documentation) or provide a link where this information can be obtained.

If the clearing process manager approves the publication of the release, both the release and the asset are marked as "registered".

Mature

The requirements of negotiation and collaboration take centre stage here: An asset marked as "mature" has taken the important first step of the quality assurance and enhancement process.

At least one release of an asset has been approved as mature by the clearing process manager and after a community review. The artefacts it consists of meet quality criteria concerning their completeness, syntactic requirements and usability for practical applications.

The focus in granting "mature" state to an asset is on its potential for reuse.

This is a cyclic process in which the asset enhances its quality incrementally (in its respective releases).

Conform

An asset accepted as conform by the clearing process manager and the conformance committee – a group of domain experts - has reached the highest possible quality state, i.e. one of its releases was accepted as conform.

From a clearing process perspective, the development of an asset ends here. It is considered conform to all the criteria defined in the Clearing Process Framework.

Modifications (especially for reuse in other contexts) will make it a new release or, in most cases, a new asset which will undergo a separate clearing process.

These stages, the community review and the role of the conformance committee reflect the interactive nature of the entire process as well as the room for negotiation about appropriate semantic properties, granularity, usability and other quality criteria foreseen in the process. The basic process is illustrated in Fig. 3.

The controlled vocabularies [SEM03] and data models provided enter the clearing process, i.e. they are subject to a quality assurance procedure and actual peer review, possibly resulting in the publication of revised and enhanced versions. Through this instrument, SEMIC.EU introduces a new layer of negotiation in the creation of semantic interoperability.

3   Roles and platform features in the clearing process

The clearing process involves people in various roles that are given to registered users. They are supported by the SEMIC.EU platform in executing their tasks. With respect to their duties and rights in the context of an interoperability asset, roles are assigned to these users. The roles are attached to permissions which enable the users to execute tasks on the SEMIC.EU platform. Slightly simplified, the roles are divided into: audience, asset provider (agent) and SEMIC.EU as an organisation (including the clearing process manager). These groups carry out the clearing process in its successive stages.

Assistance is provided by SEMIC.EU in questions of documentation, licensing and publication of assets. After the publication of an asset, the owner can request additional quality assessment.

Asset assessment

Asset agents receive feedback on their assets from the SEMIC.EU clearing service, practical advice on development and avoidance of potential semantic conflicts compiled in the form of a comprehensive SEMIC.EU asset assessment report.

Asset management / Repository

The platform manages the versioning and assists in the collaborative development of the asset. When the asset agent uploads an asset to the repository, the asset receives an initial version number, incremented each time a new version of the asset is uploaded or the asset progresses to another state in the maturity process. Using the forum, interested users can discuss the design and development decisions.

Asset refinement

During the asset refinement the clearing manager gives feedback on the asset concerning its design, documentation, modularisation and semantic accuracy. A focus in the refinement of assets is their respective potential for reuse.

Asset harmonisation

During the asset harmonisation assets with similar focus are studied and if possible consolidated. The SEMIC.EU clearing manager works together with the asset-agents and requests feedback from partners, experts and interested stakeholders.

Asset development assistant

Additional support in the development process can be requested via the asset development assistant, an interactive feature linked to the registration and development processes. The assistant suggests issues to consider, gives examples and guides the user through all relevant phases of asset development, asset clearing, licensing and quality improvement. Common user requests and issues of general interest are successively included in the assistant functionality.

Contents on the platform are interlinked and related (e.g. actors and assets), so that users can recognise the relationships of artefacts (derivates, similar content or use case). One of the most important cases is the reference from a mapping asset to its respective source and target artefacts (taxonomy or ontology).

4   Exploiting the potential of ontologies

The potential of ontologies to generate semantic interoperability in eGovernment services and general data exchange is commonplace. However, empirical data on the actual use of ontologies and their acceptance by practitioners has yet to be gathered.

The activities coordinated by SEMIC.EU show a more widespread use of single data models, code lists and taxonomies in every-day eGovernment contexts. They are shared through the open repository of SEMIC.EU, undergo an open, peer-review

based quality assurance process and are the object of community work and coordination.

However, a recent poll by SEMIC.EU among its user base has shown that well over 80 per cent of the respondents claimed to have used or are planning to use ontologies for data exchange within the next 12 months[2].

Based on this we expect that more than 30 eGovernment projects will request for SEMIC.EU coaching services to develop, implement and reuse of ontologies in practical eGovernment projects within the next year.

In a Union of 27 Member states with a number of other associations, candidate countries, the European Economic Area, etc., federating existing sources and making them accessible through a central resource, is a pivotal task.

As a European centre for eGovernment services, it is SEMIC.EU's duty and aim to establish the link between ontologies and eGovernment practice.

From the point of view of government users the benefit is often not apprehensible, in contrast to other semantic specifications with simpler structures like code lists and glossaries.

Public administration bodies themselves have not shown readiness for implementation or usage of ontologies (or related methods).

5 eGovernment requires reliability, soundness, usability

All SEMIC.EU's activities and initiatives are performed in the public interest. Therefore they are always driven by the demand for practical implementation. This has a number of consequences also for strategic and organizational considerations:

SEMIC.EU's mission is not to be fast or avant-garde but rather solid, stable and reliable.

The demands of public services require that solutions are concrete, reliable, in line with common standards and legislation, and – to a certain degree – transferable. The latter is a specific European requirement due to the heterogeneous linguistic and legislative landscape in the European Union despite the existence of European directives and regulations[3].

What about SEMIC.EU's relationship to the semantic web community and existing semantic technologies? Knowledge is indeed extracted from silos, made available and accessible but only as a function of its usability and purposefulness in matters of public administration and data exchange for eGovernment services. Peer-reviews and community work warrant accuracy and practicability of the proposed solutions.

---

[2] The survey was conducted via e-mail among 900 users (experts in eGovernment, IT) of SEMIC.EU representing predominantly EU Member States. The response rate was 20 percent. Respondents were asked to choose one out of three statements ("Yes, we have already used ontologies", "Yes, we plan to implement ontology based applications within the next 12 months", "No, ontologies have never been relevant in the context of our work").

[3] Ferrario and Guarino givan an assessment of the potential of ontologies in services contexts [6]

5.1   Is government ready to work with ontologies?

Against this background, we can state that eGovernment has caught up – it has, in many cases and ways, embraced the opportunities of semantic technologies. The poll mentioned above illustrates the demand and willingness of an eGovernment audience to follow suit, or, respectively, lead the way.

European public administration has realised the potential of ontological solutions for its purposes and for the extension of its services and efficiency measures.

Austria has once more proven to be a frontrunner in the adoption of technologies for government services. The Austrian chancellery and municipalities already offer ontology-based citizen services [7].

The European Commission's DG Environment is developing a novel information platform and service, making use of dispersed information from various domains and resources in order to generate added value and integrated information in climate change matters.

5.2   Are ontologies ready to deliver to government?

Researchers and developers are now called on, it is their task to produce usable ontologies that meet the needs and requirements of practitioners, i.e. secure, reliable, proven ontologies and methodologies that have the power to be used in real-life scenarios. Theory is a precondition, practicability is what counts at the end of the day.

Moreover, it will be advantageous to create and find links between ontologies and other types of semantic interoperability assets already available on SEMIC.EU.

For an ontology to be of use to administrative services, quality requirements must be clear, precise and comprehensive.

6   Making implementation possible: Federation and Reuse

The aspects of reuse and exchange are imperative for SEMIC.EU's approach. The Semantic Interoperability Centre Europe is therefore not only organized around the online platform, but also with its communities and as a coaching and clearing service.

How can eGovernment projects and ontology developers benefit from each other? SEMIC.EU provides the organizational structures for coordination and federation of existing resources [SEM04] & [EPR01]. The approach of collaborative issue solving and network building creates an environment that can work as an interface between theory and practical requirements.

At the same time it is an information space where latest developments, unresolved issues and new ideas can be shared.

Much as it is SEMIC.EU's task to coordinate cross-country and pan-European cooperation, it is also vital to monitor and display trends and developments. This is, however, always with a view to quality and practicability criteria.

What does SEMIC.EU have to offer to the scientific community and to practitioners at the same time? It is the European source for cooperation partners, experiences, real life application of technologies (proof of concept), bringing together dispersed initiatives and knowledge. Making solutions accessible as a clearinghouse is one part of this function.

In order to turn this general mission into actual benefit, ontologies (including their drafts and beta versions) must be documented and available at SEMIC.EU in order to become ready for use (see Fig. 4 for the total number of assets available through SEMIC.EU).

Only if this is accepted and put into practice, can the SEMIC.EU community yield results from its opportunities for joint and coordinated development of ontologies, to double-check against practice requirements and feasibility.
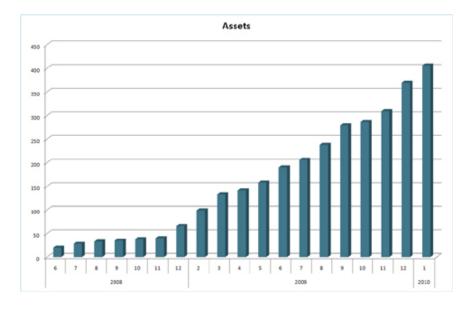
Fig. 4 Total assets registered at SEMIC.EU

6.1   What does SEMIC.EU not do?

SEMIC.EU does not develop ontologies itself. Coordination, coaching, investigation, demand and applicability are the main driving forces for ontology coordination through SEMIC.EU. Therefore, it offers academics, government officials, IT professionals and project managers assistance in

- using an existing and established infrastructure with access to decision-makers, practitioners, voices in both technical and administrative/political issues.

- Building on a network and getting the backup of the central European administrative body, the European Commission.

7   SEMIC.EU – The role as a point of collaboration for pan-European eGovernment initiatives

SEMIC.EU aims to facilitate seamless data exchange in Europe, particularly in public administration. Therefore questions of semantic interoperability have to be solved. Ontologies are one of the most relevant concepts to enable semantic interoperability.

Currently SEMIC.EU initiates collaboration with other Member States initiatives and domain specific repositories to facilitate the integration and federation of different repositories. Cooperations have been established with ontology repositories such as the United Nation's Food and Agricultural Organization's AIMS (Agricultural Information Management Standards) and the United States National Center for Biomedical Ontology (NCBO).

Looking from the practical point there are quite a lot of open questions to be solved.

- What is the most promising and useful strategy to link them?
- Which metadata are relevant and need to be shared?
- How can we establish interoperability between ontology repositories?

Our next step is to create a common metadata schema that allows for data exchange between different repositories. Furthermore it will be necessary to develop different scenarios for repository integration, and define use-cases and collect requirements for these scenarios. In the context of ontology development and mapping, this will also mean that visualisation, relationship definition and navigation support can be realised.

SEMIC.EU invites all other initiatives of open repository development to collaborate.

References

1. [SEM01] Fraunhofer ISST and]init[.Quality Framework for Interoperability Assets. European Commission – IDABC. 2008.
2. [SEM02] Fraunhofer ISST and]init[. Vision of the Clearing Process. European Commission – IDABC. 2008.
3. [SEM03] Fraunhofer ISST and]init[. Study on Methodology. European Commission – IDABC. 2009.
4. [SEM04] Community on SEMIC.EU Technology Reuse http://www.semic.eu/semic/view/snav/network/Communities/Technology-reuse.xhtml?cid=378223
5. [EPR01] SEMIC.EU Technology re-use Community http://www.epractice.eu/community/semictechreuse
6. Ferrario, R.; Guarino, N.: Towards an Ontological Foundation for Services Science. In: J. Domingue, D. Fensel & P. Traverso (eds.) Future Internet - FIS 2008, First Future Internet Symposium, Vienna, Austria, September 29-30, 2008, Revised Selected Papers (Lecture Notes in Computer Science 5468), pp. 152-169 (2008)
7. Salhofer, P., Stadlhofer, B.; Tretter, G.: Ontology Driven e-Government. In: P. Hahamis (ed.), Proceedings of the 9th European Conference on e-Government, London, United Kingdom, June 29-30, 2009, Reading: API, pp. 540-549 (2009)