A Market User Perspective is Important
Arild Haraldsen, NO

Arild Haraldsen, NO

Semantic interoperability

The European Interoperability Framework (EIF) of the EU describes three layers of interoperability: technical, semantic, and organisational. The Internet and World Wide Web, along with technologies like SOA and Web Services, are providing protocols and standards necessary in order to link different systems at a technical level. This helps open up the information silos that have traditionally existed between computer systems in different departments and enterprises. However, technical interoperability also lays bare the challenges that exist at the semantic level, since it does not address the issue responsible for most of the costs involved in information integration projects – that is, a common understanding of the precise meaning of the information.

Introduction to the selected technologies

This report focuses on technologies that purport to offer Semantic Interoperability. Many of them are either directly based on XML, or else utilise XML in one way or another. Because it is actually a meta-language, XML has given rise to a whole flora of markup languages designed to allow the representation and interchange of information and knowledge in various domains. As noted above, all of these involve semantics to some degree or another, but only a few of them exhibit sufficient generality to be termed semantic technologies.

RDF/OWL

RDF (Resource Description Framework) and OWL (Web Ontology Language) are two of the technologies that form the foundation of the Semantic Web.

The Semantic Web is an extension of the existing Web in which information is given well-defined meaning, enabling computers and people to work in better cooperation. The term “Semantic Web” was first coined by Tim Berners-Lee, the creator of the Web, in his article in Scientific American in May 2001 where he presented his vision for the next generation Web. His vision was to annotate the data on the web so that they can be clearer understood by machines, not only by humans as the situation is today. In that way it is possible to have Web requests answer your questions in a much more intelligent way than just refer to documents that contain your search strings.

The vision of the Semantic Web is to extend principles of the Web from documents to data. This extension will make it possible to fulfil more of the Web’s potential, in that it will allow data to be shared effectively by wider communities, and to be processed automatically by tools as well as manually.

The Semantic Web allows two things.

1. It allows data to be surfaced in the form of real data, so that a program doesn’t have to strip the formatting and pictures and ads off a Web page and guess where the data on it is.

2. It allows people to write (or generate) files which explain – to a machine – the relationship between different sets of data. For example, one is able to make a “semantic link” between a database with a “zip-code” column and a form with a “zip” field that they actually mean the same – they are the same abstract concept (described in an ontology). This allows machines to follow links and hence automatically integrate data from many different sources.

Semantic Web technologies can be used in a variety of application areas; for example: in data integration, whereby data in various locations and various formats can be integrated in one, seamless application; in resource discovery and classification to provide better, domain specific search engine capabilities; in cataloguing for describing the content and content relationships available at a particular Web site, page, or digital library; by intelligent software agents to facilitate knowledge sharing and exchange; in content rating; in describing collections of pages that represent a single logical “document”; for describing intellectual property rights of Web pages (see, for example, the Creative Commons), and in many others.

Topic Maps (ISO/IEC 13250)

Topic Maps is a standard technology for describing knowledge structures and using them to improve the traceability of information. It is based on a formal model that subsumes those of traditional finding aids, such as indexes, glossaries, and thesauri, and extends them to cater for the additional complexities of digital information. The model is defined in an ISO standard (ISO 13250), along with interchange syntaxes, a formal semantics, and a graphical notation. Ancillary standards define a query language (TQML), a constraint language (TMCL), and mappings to other knowledge organisation specifications, such as Dublin Core.

The value proposition of Topic Maps is that it provides the ability to control infoglut and share knowledge by connecting any kind of information from any kind of source based on what it means, i.e. its semantics. It is increasingly used in Enterprise Information Integration, Knowledge Management, and e-Learning, and as the foundation for web-based information delivery solutions.

UN/CEFACT Core Components

The Core Component Technical Specification (CCTS) is a cornerstone of the UN/CEFACT standardisation e-business activities envisioning simple, transparent and effective processes for global commerce. Its focus is on machine-to-machine exchange of business documents like order, billing, transport documents between business-critical applications hosted by collaborating business partners. Core Components are the syntax-neutral and technology-independent building blocks that can be used for data modelling. Major benefits of CCTS include improved reuse of existing data artefacts and improved enterprise interoperability.

ISO 15926

The purpose of this standard is to facilitate integration of data to support the life-cycle activities and processes of process plants. To do this ISO 15926 specifies a data model that defines the meaning of the life-cycle information in a single context supporting all the views that process engineers, equipment engineers, operators, maintenance engineers and other specialists may have of the plant. The data model is generic and can be used in any industry.

Industry adaption can be done by defining the appropriate semantics by extending the Reference Data Library. See below.

Traditionally, data associated with a process plant have been concentrated on some individual view of the plant at a point in time. Such data are usually defined and maintained independently of other groups of users, resulting in duplicated and conflicting data that cannot be shared either within an enterprise or with business partners of an enterprise. ISO 15926 enables integration of such data, and as it is based on 4D (handling the space and time dimensions) paradigm it can handle change over time.

UML

In the field of software engineering, the Unified Modeling Language (UML) is a standardised specification language for object modeling. UML is a general-purpose modeling language that includes a graphical notation used to create an abstract model of a system, referred to as a UML model. UML helps you specify, visualise, and document models of software systems, including their structure and design, in a way that meets all of these requirements.

UML is officially defined at the Object Management Group (OMG) by the UML metamodel, a Meta-Object Facility metamodel (MOF). Like other MOF-based specifications, the UML metamodel and UML models may be serialised in XMI. UML was designed to specify, visualise, construct, and document software-intensive systems.

UML is not restricted to modeling software. UML is also used for business process modeling, systems engineering modeling, and representing organisational structures. UML has been a catalyst for the evolution of model-driven technologies, which include Model Driven Development (MDD), Model Driven Engineering (MDE), and Model Driven Architecture (MDA). By establishing an industry consensus on a graphic notation to represent common concepts like classes, components, generalisation, aggregation, and behaviours, UML has allowed software developers to concentrate more on design and architecture.

UML models may be automatically transformed to other representations (e.g. Java) by means of transformation languages.

With regard to semantics and Semantic Interoperability these issues have not been the core issues that UML was designed to handle. The focus of UML has been primarily on the development of software systems, not to secure interoperability between these systems on a semantic level. However, UML delivers modeling notation that is very useful for information modeling and with the use of profiles UML is often used as a vehicle to secure Semantic Interoperability.

Conclusion

This work has been the participants' first step in the direction of improved and common understanding across technologies. And it has been useful to identify the strengths and weaknesses of the technologies and become conscious of how one in some cases can opt to use different technologies together to give the maximum effect when using semantic technology in some projects.

In an emerging field like semantic technology the process of writing this report has brought together companies and researchers with the goal to share and increase knowledge. We hope to receive comments on this work and we know it is not complete as such.

Contributions from:

- Arne-Jørgen Berre, SINTEF
- Stian Danenbarger, Bouvet
- Lars Marius Garshol, Ontopia
- Johan Wilhelm Klûwer, Det Norske Veritas
- Magne Valen-Sendstad, Det Norske Veritas
- Per Myrseth, Det Norske Veritas
- David Norheim, Computas
- Steve Pepper, Ontopedia
- Mariann Sundvor, NorStella
- Øyvind Aassve, NorStella

__Arild Haraldsen, NO,__ is general manager of NorStella, a private non-profit foundation for e-business and trade facilitation. Arild Haraldsen is Head of Delegation to UN/CEFACT Management Forum. His organisation is actively participating in UN/CEFACT and CEN/ISSS working groups. Haraldsen is head of NordiPro – a co-operation between the "pro"-organisations in the Nordic countries. Haraldsen has also written several textbooks on strategy and e-business strategy.