**XML related Data Exchange from the Test Machine into the Web-enabled Alloys-DB**

Nagy M, Over HH, Smith A

Joint Research Centre of the European Commission – Institute for Energy, Petten, The Netherlands
{miklos.nagy@jrc.nl hans-helmut.over@jrc.nl, smitha@tiscali.nl}

Abstract

The engineering sciences place considerable reliance on electronic systems to produce, store and process experimentally measured material data. One of the JRC Petten tasks is to support European Research and Development (R&D) projects in the material and energy related areas with the management and dissemination of research results. The 'Alloys Database' (**Alloys DB**), developed at JRC Petten, manages mechanical and physical properties test data and covers materials behavior for base materials and joints. By using XML technology such test data can directly be entered from the test machines into the Web-enabled **Alloys-DB**.
Test data, which are kept in XML format and sent by R&D project partners via the World Wide Web to the Petten Server are stored within the **Alloys-DB** XML database module. There they can be checked and updated on-line before they are uploaded into the database. After validation by the source administrator they can immediately be retrieved and evaluated by all project partners. A pilot test with the new XML related data exchange module from test machine into **Alloys-DB** has currently been started within the European 'TMF Standard' R&D project.
The use of XML standards is already a very promising step into the next generation Web, often referred to as the Semantic Web. It aims at machine-processable information that will enable true interoperability between sort specific databases, different in structure and language. Data on the Semantic Web is characterized by semantic metadata called ontologies that facilitates sharing and interoperability between systems and languages. It is our aim for the future to realize this step and thus to enable JRC Petten to built up a European network for collecting and disseminating publicly available experimentally measured materials data within national and European R&D projects using the Web-enabled **Alloys-DB**. Further on it would allow exchanging materials data on an international level.

## 1. Introduction

One of the JRC Petten data management and dissemination tasks is the establishment of a European network for collecting and disseminating publicly available experimentally measured materials data within national and European R&D projects in the material and energy related areas. The most promising and efficient data storage method is the implementation of a direct link between the test machine, respectively the post processing software of a test machine and a database.
Test data, which are logged in XML format and sent by project partners via the World Wide Web to the Petten Server can be stored within the Alloys-DB XML database module. There they can be checked and updated on-line before they are uploaded into the database. After validation by the source administrator they can immediately be retrieved and evaluated by all project partners. A pilot test with the new XML related data exchange module from test machine into

**Alloys-DB** has currently been started within the European 'TMF Standard[1]' R&D project. This project creates huge amounts of data; each of the 20 partners performs 20 thermo-mechanical fatigue tests. Each thermo-mechanical fatigue test contains information about source, material, specimen and test condition, which results already in major set of metadata but the test result itself contains test specific information, data for up to 100 cycles and up to 300 curve data for each cycle.

## 2. XML and semantic interoperability

XML - eXtended Markup Language is accepted as the emerging standard for data exchange on the Web, aimed at solving the problem of interpretation and syntactic interoperability but still not meaningful enough to describe information sources at the semantic level. One reason why semantic interoperability became an important issue is that there are currently over 300 XML eBusiness formats, essentially amounting to domain specific taxonomies. As envisioned the Semantic Web is the next generation of the World Wide Web which will provide machine processable information to intelligent agents or software applications in order to help humans find their way in the huge amount of data available on this media. Knowledge representation researchers have already developed different ontology languages to represent machine-readable and understandable web languages based on current exchange standards such as XML. One of the latest proposals of the W3C is OWL[2] , which became a W3C recommendation in 2003 as web ontology language. This approach assumes that information available on the web is expressed or annotated with these languages, which means that concepts are expressed by the means of classes, properties and relations among them. Scientific databases containing vast numbers of experimental data became Web enabled in order to facilitate better knowledge sharing and reuse between the scientific communities. Although these databases are accessible, the seamless data exchange between differently standardised databases is still an unsolved problem in spite of the fact that different XML based languages were defined by the different scientific communities e.g. MatML[3] on the field of material science to facilitate a standardized XML based data exchange. This solution solved a number of interoperability issues but makes the assumption that both parties agreed the syntax of the data exchange. This assumption fails when one search for existing experimental data on the web since neither the syntax nor the semantics of the requested data is known before the submission of the query. The problem is that different research institutions and companies use different standards and naming conventions in their logical data model for the same data, additionally these data models are not always even accessible on the web. Hence a vast number of experimental data are remaining inaccessible or unanalyzed.
As different communities defined their own XML schema (MatML , femML[4] , CML[5] , MathML[6] ) to resolve interpretation and interoperability issues during data exchange via World Wide Web a number of problems from the data exchange perspective have been resolved. From the computation perspective the tag like <Material> carries the same semantic information as tag

[1] Thermo-mechanical fatigue - the route to standardization, Project Reference: G6RD-CT-2001-00526
[2] Web Ontology Language
[3] Markup language for the Exchange of Materials Property Data
[4] Finite Element Modeling Markup Language
[5] Chemical Markup Language
[6] Mathematical Markup Language

<H1>. The computer just simply does not know what the material means and how the material as a concept relates to another concepts like data source. From this perspective information between tags can help for humans to predict what information can be found between the tags but for an XML processor this information is completely meaningless. To address the semantic interoperability problem Tim Berners Lee[1,2] the founder of the World Wide Web proposed different levels of information that can be expressed by the language on the web (Figure 1).

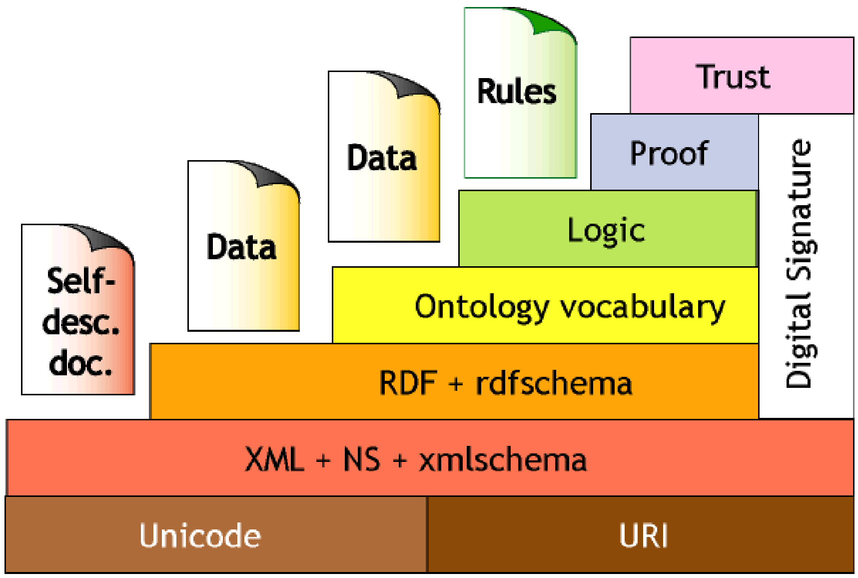



**Figure 1: Different levels of information on the web expressed by the language**

The functions of the different layers are as follows:

- XML, Namespaces and Schema provide a basic format for structured documents, with no particular semantics.
- RDF[7] and RDF Schema is a basic assertion model that allows an entity-relationship-like model to be made for the data and allow the document structure to be constrained for predictable computable processing.
- The Ontology layer provides more powerful schema concepts, such as inverse, transitivity, uniqueness of properties that allows a system to recognise different identifiers and their relations, which in fact are talking about the same thing.
- Logical layer turns a limited declarative language into a language with inference and functions.
- Proof layer enables different parties to exchange assertions between each other, together with the inference path to that assertion.
- Web of trust: The proof layer together with the digital signature turns the web into the Web of trust.

To date industrial and commercial implementations of data exchange on the World Wide Web can be placed into the second layer where the XML Schema defines XML data. Research

[7] Resource Description Framework

projects from different fields have resulted in a number of prototype systems that make use of the higher layers namely ontologies to resolve semantic heterogeneity between different data sources.

For true semantic interoperability, which means sharing knowledge and information between different applications a shared set of terms describing the application domain with a common understanding is necessary. More flexibility could be achieved if not just a set of terms is defined but the relations between them. On this way application are at least partially aware of the domain, which increases flexibility considerably.

To address the above-mentioned challenges, build on XML and XML Schema more expressive ontology languages has been created that represent the conceptualization of the domain with explicitly representing the meaning of terms in vocabularies and relationships between those terms. It also describes a vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties, characteristics of properties (e.g. symmetry), and enumerated classes. These extensions of the markup language makes it possible not only to read the data but interpret and infer over the data, which can enable the development of applications that can autonomously retrieve and manipulate information available on the sematic Web. While the complete realisation of Tim Berner's Lee view seems far away, achievements that has been reached so far are worth to utilizing to enrich data exchange with semantic information.

## 3. Architectural aspects

Web services based on XML messaging technologies, such as SOAP[8] and WSDL[9] are the backbone for our pilot application, which includes SOA (Service-Oriented Architecture) and Oracle XML DB that provides a high-performance, native XML storage and retrieval technology. Our current implementation runs under oracle 9i and Oracle 9iAS. Due to performance issues related to the Oracle XDB, the JRC Petten Data Management and Dissemination (DMD) sector migrating to Oracle 10g, which can provide improved performance concerning the XML storage and manipulation. From the knowledge engineering perspective the OWL ontologies are constructed by the Protégé ontology & knowledge base editor that provides a user friendly and easy way to create OWL files. Since the Semantic Web is still in its infancy, which implies that well-developed ontology engine is still non existent the DMD is working on a ontology engine that based on pre defined domain specific rules that can support the XML based data entry process.

## 4. Alloys-DB Content

The pilot study has been carried out for data exchange between test machines and the Alloys database (Alloys-DB) that stores experimentally measured mechanical and physical properties data of engineering alloys produced by the European R&D projects. It covers the materials behavior at low, elevated and high temperatures for base materials and joints and also includes irradiation materials testing in the field of fusion and fission and tests on thermal barrier coating for gas turbines. Its emphasis is on data from standardised tests and on evaluation methods,

[8] Simple Object Access Protocol
[9] Web Service Description Language

which are well established and widely accepted. Alloys-DB has more than 130 entities with more than 1850 attributes. Detailed description of Web enabled retrieval and evaluation can be found in [3,4].

## 5. Alloys-DB-Ontology

Ontology as defined in the knowledge-sharing context is a specification of a conceptualization of a specific domain.
The basic structure of the Alloys-DB ontology is depicted in figure 2. Development of the Alloys-DB ontology is under progress and will be extended for all entities that Alloys-DB cover. The ontology corresponds to the Alloys-DB XML Schema that extends the MatML Schema with concepts like specimen, test condition that is not described by MatML. The main purpose of creating such a metadata repository is to represent the semantic concept mappings between the different research partners and/or external clients without imposing a standardized syntax structure on them. The main advantage of this solution is to let different institutions use their own well-established naming structures but facilitate sharing the data by translating between the naming without explicitly hard coding it into software applications.

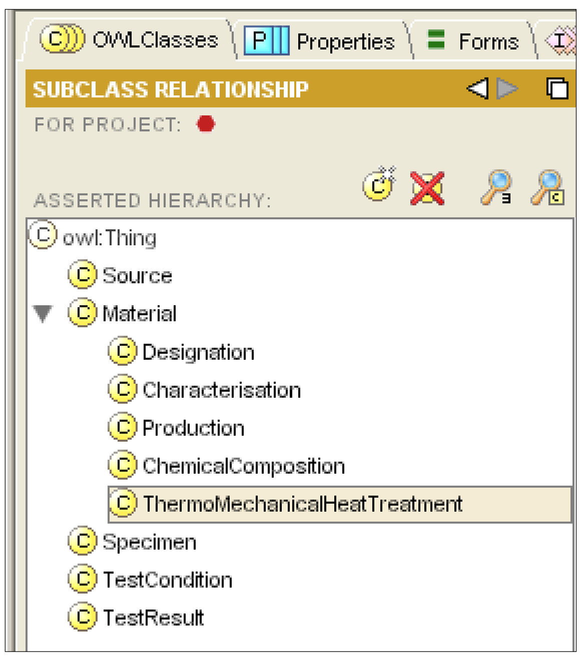



**Figure 2: Alloys-DB ontology class hierarchy**

## 6. XML data in Alloys-DB

The high-level system architecture of the pilot XML upload study that has been carried out for uploading complex thermo-mechanical fatigue tests into the Alloys-DB is shown on figure 3. The scope of the test types will be extended for all tests that Alloys-DB cover. Research partners post process the raw test results supplied by the test machine and create XML file from it that conforms to their naming structure that has been registered in the OWL ontology repository.

Based on the mappings in the ontology repository the data file can be transferred into the Alloy DB XML module through Web-services where the data are temporarily stored before they are uploaded into the relational database. Keeping the XML files in a temporary area enables research partners to check, update and validate the data set on-line and share the data in XML format between each other where based on the ontology repository one can resolve the structural and naming differences between formats in case they have not agreed on a standardised format.

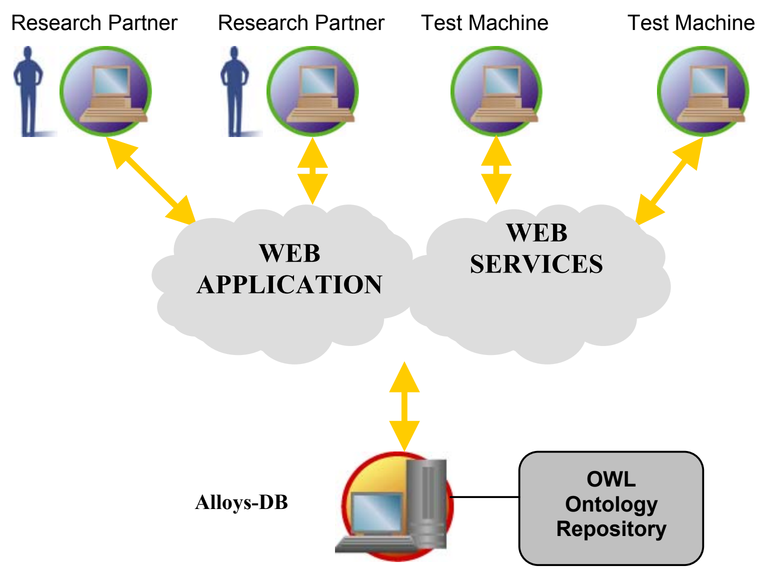
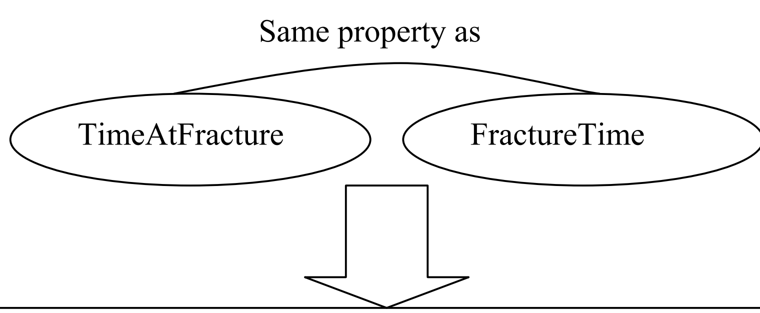



**Figure 3: System architecture of the AlloysDB XML module**

Since OWL is based on description logic it is possible to use a reasoner to determine inconsistencies in the ontology. However this proved difficult in practice because of the lack of mature ontology to reasoner language converters. In the Alloys-DB ontology besides the concept and property descriptions different instances represent the different research partners. In our pilot study we are currently able to resolve the following differences:

- Concept and attribute naming




```
<owl:DatatypeProperty rdf:ID="TimeAtFracture">
   <owl:equivalentProperty rdf:resource="#FractureTime"/>
   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
</owl:DatatypeProperty>
```

**Figure 0: Concept and attribute similarity in OWL**

Figure 4 depicts that different vocabulary that describes attributes can be mapped directly into OWL <owl:equivalentProperty> construct while different concepts can be mapped into <owl:equivalentClass>.

- XML Structural and linguistic

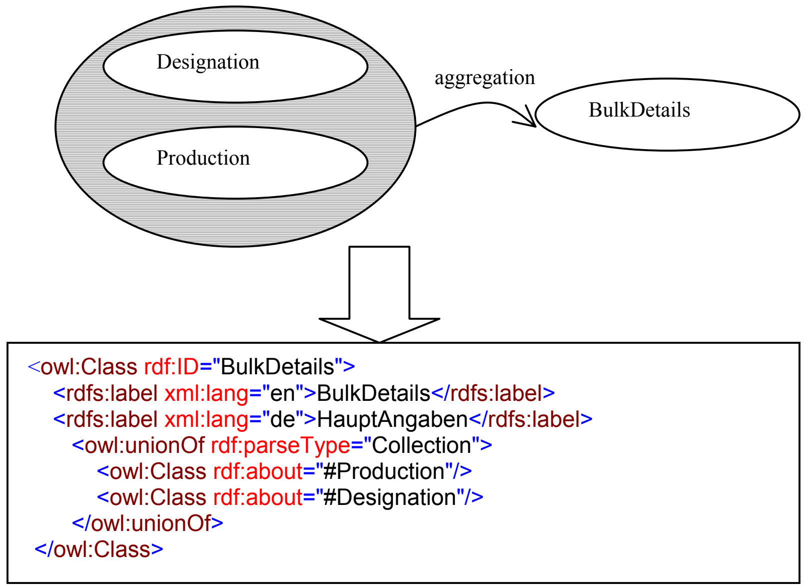



**Figure 5: Structural similarities in OWL**

As depicted on figure 5 the structural differences can be mapped into OWL complex class construct that support basic set operations such as </owl:unionOf>. The <rdfs:label> entry provides solution for the linguistic differences but does not affect the logical interpretation of the ontology.

## 7. Data upload from the test machine to the Alloys-DB

Scientific data servers, and in particular the ones publicly available on the Web, usually provide information retrieval techniques to access data. Uploading experimental data from the other hand is a very important factor in data management. Raw data produced by the test machine can easily be converted into XML format by a post processor and uploaded directly to the database. However in practice this process is not so straightforward as described above. Metadata needs to be created, which can be used to mediate between different XML formats and standards used by the different institutions.

The overview of the uploading process from the test machine to the Alloys-DB is as follows and depicted on figure 6.

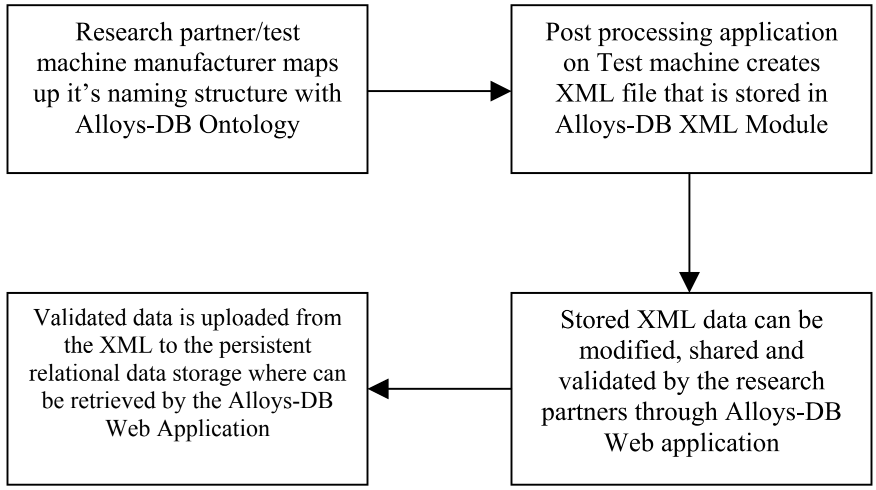



**Figure 6: Flowchart of the XML Data upload**

Research partners and the Data Management and Dissemination group creates mappings between the different naming conventions, which can cover structurally and semantically different expressions and they include them into the Alloys-DB OWL ontology, e.g.:

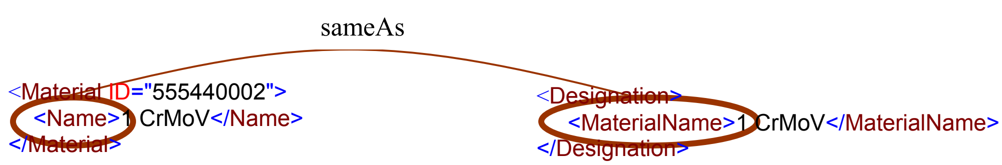



**Figure 6: Attribute similarity**

The post processing application instead of only hard coded decision-making using the Alloys-DB OWL ontology to resolve semantic differences between the raw data files or source XML files and the Alloys-DB XML Schema. As a result of these process the XML file can be stored in the Alloys-DB XML module.Alloys-DB Web application provides a user-friendly interface to add, modify, validate and share the data set stored in XML format. In order to avoid the possibility of typical human errors like mistyping important field values (e.g. test standard, fracture mode) Web application maintains a thesauri list for the specific entities in the Alloys-DB that helps to guarantee that the field values will not be entered incorrectly. Once the mandatory fields has been filled in and the data set has been validated by the authorized person the content of the XML file can be uploaded into the relation database where it is immediately available for Web enabled retrieval and evaluation.

## 8. Conclusion

In this paper we investigated the use of ontologies for materials data exchange between research partners in the European Research Projects. The results are relatively modest comparing the full potential of using ontologies to improve data discovery and exchange in a distributed and heterogeneous environment. Using a standardised ontology between material databases [7] and from machine to database for facilitating data exchange is just the first step in the direction of achieving semantic interoperability between heterogeneous data sources and fully in line with our future vision of using semantically rich data sets in the scientific domain. We argue that using only standardized schema based metadata repositories will give solution for today's interoperability problems in the scientific community. Moving towards using distributed ontologies and mappings for each data provider is a challenging task and an active research area that will contribute to our future research interests. According to our view the main future challenge of data interoperability and exchange for the material society is to address the semantic heterogeneity problem with creating an industry wide ontology repository. Besides these challenges it has to be acknowledged that moving forward in this direction should require cautiousness since the majority of practically usable semantic web related technologies are in their infancies and the area itself is an active research topic among computer science and knowledge representation researchers. The biggest shortcoming that we identified is the OWL reasoner connection and the lack of reliable ontology engine that make use of the information coded in the metadata.

## References

[1] Berners-Lee T., Hendler J., Lassila O., The Semantic Web, Scientific American, May 2001

[2] Berners-Lee T, M. Fischetti, Weaving the web: The original design and ultimate destiny of the World Wide Web by it's inventor. Harper, San Francisco.

[3] W3C: OWL Web Ontology Language Semantics and Abstract Syntax: URL:http://www.w3.org/TR/2004/ REC-owlsemantics- 20040210/, 2004

[4] W. Dietz, H. H. Over, E. Wolfart, "Web-enabled Materials Databases of the European Commission, a Tool to Manage Experimentally Measured Data", Materialsweek 2004, Munich, Germany, 21 - 23 September 2004

[5] H.H. Over, E. Wolfart, E. Veragten, "Data Retrieval and Evaluation within the Web-enabled Alloys-DB", 19th International CODATA Conference, "The Information Society: New Horizons for Science", Berlin, Germany, 7-10 November 2004

[6] Patrick Lehti, Peter Frankhauser, XML Data Integration with OWL: Experiences & Challenges

[7] VAMAS - Versailles Project on Advanced MAterials and Standards - TWA workshop: 'Generic Platform of Materials Database', Ryukoku University, Japan, 5. February 2001