Implementation of Materials Data Integration using Ontology

Toshihiro Ashino[1], Nobutaka Nishikawa[2], Takuya Kadohira[3] and Masahiko Demura[3]

[1] Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, JAPAN
[2] Mizuho Res. & Technol., Ltd. 2-1 Kanda-Nishikicho, Chiyoda-ku, Tokyo, 101-8443 JAPAN
[3] National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, JAPAN.
ashino@acm.org

Abstract. For materials design, it is necessary to refer a variety of heterogeneous information resources, such as databases, formulae, and computational simulations. Ontology-based data integration is one of the approaches to combine data from multiple heterogeneous data resources. In this paper, an implementation of ontology-based data integration with Semantic Web standards and its application for materials data integration are presented.

Keywords: Ontology-based data integration, Materials integration, Semantic Web.

1 Introduction

In the development of research into data driven materials design, the importance of data integration has been recognized. The materials design process requires utilization of a wide variety of information, including traditional data sheets, data from new automated experimental devices, knowledge of materials science and engineering, and simulations using physical models. In order to support researchers, it is necessary to develop a platform to access these data, process workflows, and visualize them. Such platform requires to handle wide variety of information resources [1].

There are several approaches to data integration [2] and ontology-based integration, utilizing ontology to overcome semantic heterogeneity [3], is one of them. One of the implementations of ontology-based data integration is to create a single ontology, called global ontology, and map different data to it [4][5]. In our implementation, the extended version of Materials Ontology [6] is used as a global ontology to map material information such as datasheets and formulas for creep, fatigue, and other properties. Fig.1 illustrates the concept of data integration for creep properties, where heterogeneous data resources such as experimental data and computational data are linked into a single ontology. We developed a web application using this as a back end and realized data navigation using an ontology which is based on the knowledge of experts.

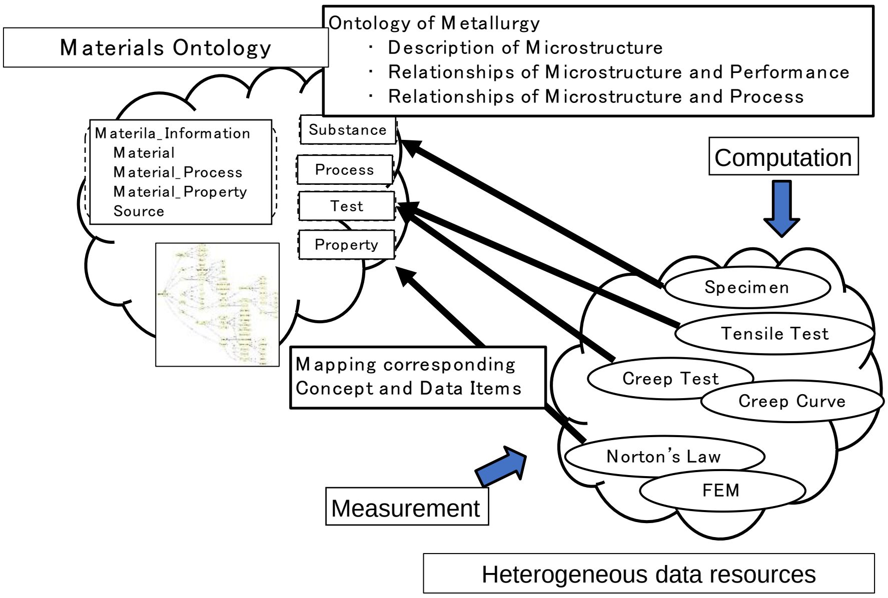

Fig. 1 The concept of heterogeneous data integration for creep properties.

2 Implementation

This implementation is based on Semantic Web technology [7] and consists of following three basic components. (see Fig. 2)

- Materials Ontology - the global ontology written in OWL (Web Ontology Language).
- RDF (Resource Description Framework) files - describe metadata of the XML data files. stored in the SPARQL endpoint.
- Data resources: XML formatted data files, RDBMSs, or Computational resources.

The RDF files play key role to map the global ontology and the data resources. Th refer classes defined in ontology and describe the corresponding information resource, such as file location, name of XML element, names of data fields and access methods of data entity.

Fig. 2 shows a typical data navigation flow. When navigating a database, the user first traces the conceptual structure of the ontology and find the target concept. The system generates a SPARQL query to search the selected OWL class and issues it to the SPARQL endpoint which stores the RDF metadata files. Each RDF file describes metadata about a single information resource.

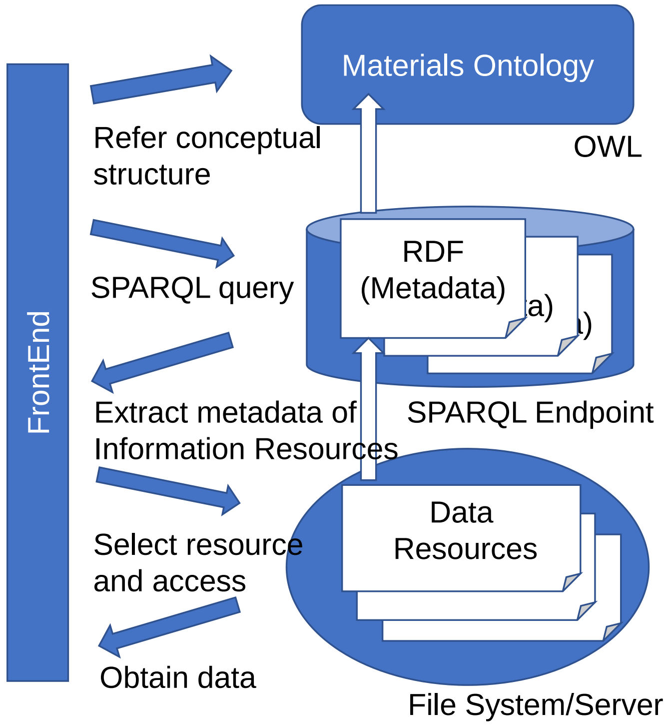

Fig. 2 Three layers, global ontology (Materials Ontology), RDF files describing metadata of data resources and heterogeneous data resources.

Fig. 3 shows an example which using XML formatted data file as a data resource. The RDF metadata has entries corresponding to the classes defined in Materials Ontology. The metadata entry contains information, which is required to access data file, the location of file XML tag name of data item and access method to the data value, in this case, XPATH expression. The tags in RDF files are conforming to standards such as Dublin Core, no original tags are defined.

In this example, XML data is conforming XML schema definition (XSD), but it is not mandatory. However, if it conforms XSD, the metadata can be extracted from XSD and it makes easier to create RDF files. Although this three layer structure takes time to create RDF files, makes it possible to map multiple data schemas or other information resources, such as numerical calculations, to a single ontology through the RDF intermediate layer.

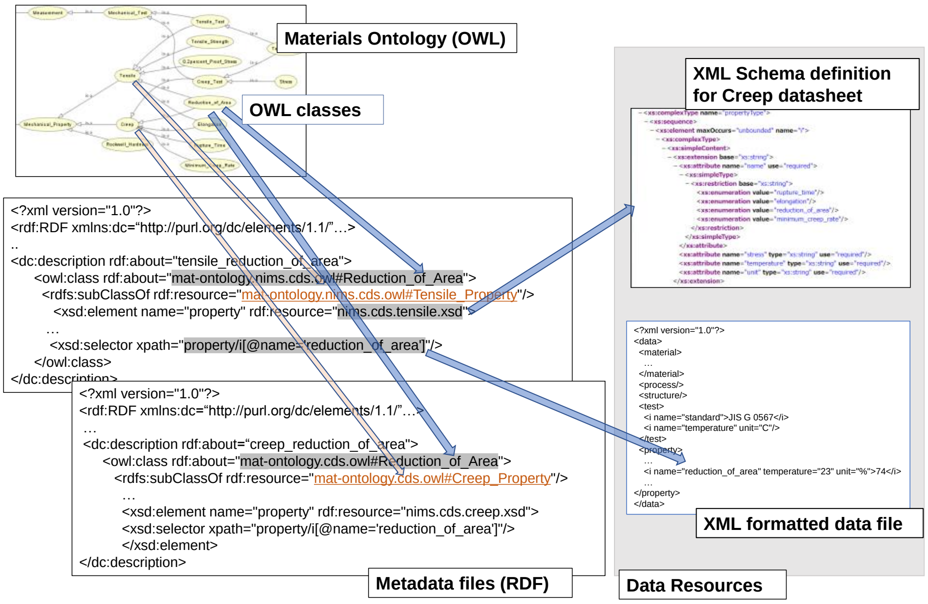

Fig. 3 Mapping global ontology to heterogeneous data resources using RDF

3 Discussion

Since in current implementation, the target data resources are digitized version of NIMS (National Institute for Materials Science) datasheets of creep, fatigue and others, they are conforming XML Schema definition, we are considering RDF format only in such data resources. Appropriate metadata tags used in RDF is under consideration for other forms. It may be possible to support data stored in RDBMSs, SQL queries may be used; for computational results, we can use the name of the software for the specified computation, the parameters to be set, etc.

Also in this system, it is inefficient to retrieve data by its value and some common data items are distributed several data files. To find data by their values, load XML data files into frontend and scan all values in these files. In order to improve this flaw, we are considering creating RDF files which consist frequently used data item, such as material type, and list of its values, pointers to data file. These RDF files are stored in the same SPARQL endpoint, automatically generated, and updated by scanning all XML data files, and retrievable by SPARQL engine.

4 Conclusion

Data integration plays an important role in material data platforms that handle many types of data and heterogeneous data resources. In this study, we implemented ontology-based data integration in a three layer structure of Ontology, RDF for metadata and

Data Resources, using “Materials Ontology” as a global view. Currently, it demonstrated by XML schema conforming data files as the data resource, which have been developed for NIMS datasheets such as creep and fatigue.

Acknowledgements. This work was supported by Council for Science, Technology and Innovation(CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Materials Integration” for Revolutionary Design System of Structural Materials (Funding Agency: JST).

References

1. Ashino, Toshihiro, Nobutaka Nishikawa, and Takuya Kadohira. "Data Analysis Environment for Materials Science and Engineering Integrating Heterogeneous Data Resources." In Data Analytics and Management in Data Intensive Domains: XXI International Conference DAMDID/RCDL'2019 (October 15–18, 2019, Kazan, Russia): Conference Proceedings. Edited by Alexander Elizarov, Boris Novikov, Sergey Stupnikov. Kazan Federal University, 2019, p. 420-425. 2019.
2. Lenzerini, Maurizio. "Data integration: A theoretical perspective." In Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, pp. 233-246. 2002.
3. Uschold, Michael, and Michael Gruninger. "Ontologies: Principles, methods and applications." TECHNICAL REPORT-UNIVERSITY OF EDINBURGH ARTIFICIAL INTELLIGENCE APPLICATIONS INSTITUTE AIAI TR (1996).
4. Wache, Holger, Thomas Voegele, Ubbo Visser, Heiner Stuckenschmidt, Gerhard Schuster, Holger Neumann, and Sebastian Hübner. "Ontology-based integration of information-a survey of existing approaches." In Ois@ ijcai. 2001.
5. Gusenkov, A., N. Bukharaev, and E. Birialtsev. "On ontology based data integration: problems and solutions." In Journal of Physics: Conference Series, vol. 1203, no. 1, p. 012059. IOP Publishing, 2019.
6. Ashino, Toshihiro. "Materials ontology: An infrastructure for exchanging materials information and knowledge." Data Science Journal 9 (2010): 54-61.
7. W3C Semantic Web page, http://w3.org//standards/semanticweb/, last accessed 2021/5/17.