NISTIR 7729

Requirements to Support a Modern Architecture for Industrial Data and Industrial Data Standards on the Semantic Web

Evan Wallace
David Leal
Allison Barnard Feeney

National Institute of
Standards and Technology
U.S. Department of Commerce

NISTIR 7729

Requirements to Support a Modern Architecture for Industrial Data and Industrial Data Standards on the Semantic Web

Evan Wallace
Allison Barnard Feeney
Manufacturing Systems Integration Division
Engineering Laboratory

David Leal
CAESAR Systems Limited

September 2010

U.S. Department of Commerce
Gary Locke, Secretary

National Institute of Standards and Technology
Patrick D. Gallagher, Director

Requirements to support a modern architecture for industrial data and industrial data standards on the semantic web

Evan Wallace, National Institute of Standards and Technology, ewallace@nist.gov
David Leal, CAESAR Systems Limited, david.leal@caesarsystems.co.uk
Allison Barnard Feeney, National Institute of Standards and Technology, abf@nist.gov

Introduction: The subcommittee of the International Organization for Standardization (ISO) tasked with developing standards for industrial data, TC 184/SC 4, has recently undertaken an effort to define a new, more flexible architecture based on widely-used information technologies and standards. The goal of this effort is to reduce barriers to the use of SC 4 standards by industry and increase the utility of SC 4 standards beyond data exchange. This paper provides an overview of the proposed architecture, the types of data it must support and the languages/technologies chosen to support it, and the challenges that it may impose on the use of the Resource Description Framework (RDF) for implementing this architecture. These challenges are similar to those experienced in other uses of RDF, and with this report we hope to encourage further enhancements in standards and tools for RDF to meet these shared needs.

Background:
While the nominal area of interest of ISO TC184/SC4 is Industrial Data, its official scope of work is described as follows:
Standardization of the definition, integration and quality management of information which is shared, exchanged and archived in the context of:

- *the definition, description and classification of products and facilities throughout their lifecycle*
- *industrial management and operations*
- *other information requirements supporting automation systems*

The focus early in SC 4 was on the exchange of product data, and the tools and architecture developed at the time reflect that purpose. Principal among these was a data modeling language called EXPRESS[1],[2] developed specifically to support SC 4 information standards. EXPRESS was, and is, a quite expressive data modeling language with a powerful constraint capability. The SC 4 architecture decoupled the model from the implementation technology, enabling reuse of EXPRESS models with different encoding methods or data sharing techniques. SC 4 is now a rich source of consensus-based, domain-specific information models; however, there are barriers to SC 4 standards use by and utility to industry.

- EXPRESS has not gained wide acceptance outside of the community associated with SC 4 standards
- SC 4's focus on data exchange means SC 4 models aren't part of a broader information technology architecture that supports all of the information needs of an enterprise

A new architecture:
Recognizing the need to adapt to the changing information technology landscape, SC 4 formed a project to propose a future SC 4 architecture. The stated objectives of the

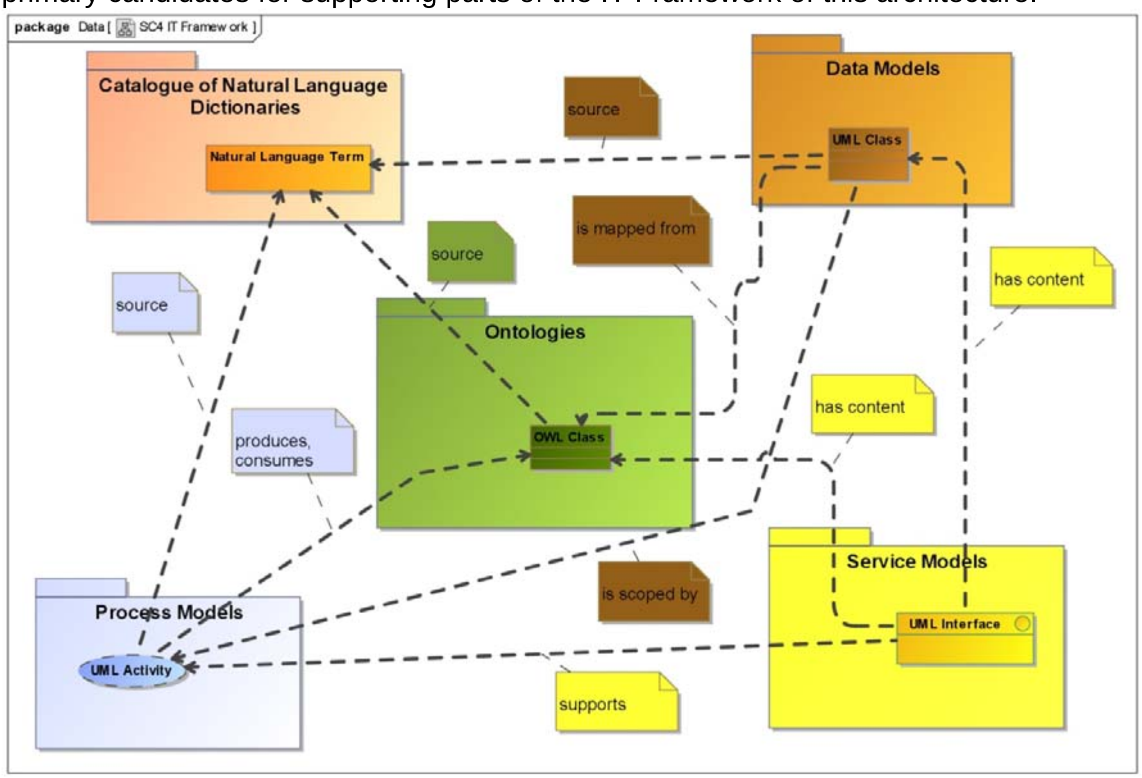
architecture are to 1) reduce greatly the barriers to the use of SC 4 standards by industry, 2) reduce barriers to the use of SC 4 standards in conjunction with standards from other bodies such as the Object Management Group (OMG) and the Organization for the Advancement of Structured Information Standards (OASIS), and 3) increase the utility of SC 4 standards to industry [3].

The design drivers for this project can be summarized by the following:

> *Provide SC 4 with a framework/architecture enabling its projects to create standards using the same technologies, methodologies and approaches that are the best practice in industry and in other standards-making bodies.*

The evolving architecture, referred to as Industrial Data Integrated Ontologies and Models (IDIOM), consists of three parts: 1) an information technology (IT) framework of software technologies and tools; 2) methodologies, policies, and guidelines for the use of the tools; and, 3) a framework of core concepts. Most effort to date has focused on the IT Framework as it is the basis for the other parts.

Ontology Summit 2009, on the topic “Toward Ontology-based Standards” recommended the use of ontologies as the best practice for specifying information standards [4]. Accordingly, IDIOM has placed ontologies specified using logic-based languages at the center of its approach. This makes the Web Ontology Language (OWL) and RDF primary candidates for supporting parts of the IT Framework of this architecture.



Components of the IT framework include

- natural language dictionaries
- ontologies
- process models
- data models

- service models
- mappings and traces

The components of the IT Framework include models at varying levels of abstraction. Some technologies are used to express more abstract models and others declare elements that are directly implementable. The Future SC 4 Architecture project surveyed ISO, OMG, OASIS and other bodies that are standardizing process models, services models and data models. The IDIOM approach was informed by the following findings

- When industry organizations and standards bodies develop ontologies, the standard language used more than any other ontology language is OWL.
- When industry organizations and standards bodies develop process models, the standard languages used more than any other process modeling languages are the Unified Modeling Language (UML) and Business Process Modeling Notations (BPMN).
- When industry organizations and standards bodies develop service models, the standard languages used more than any other service modeling language are UML, a profile of UML corresponding to the Web Service Definition Language (WSDL), and/or WSDL itself.
- When industry organizations and standards bodies develop data models, the standard languages used more than any other data modeling language are UML, EXPRESS (primarily in SC 4), and XML Schema.

The goal of IDIOM is to document industry best practices for industrial data. While the draft specification of IDIOM names a set of technologies for each component, best practices change. IDIOM is intended to be flexible and to be extended as new technologies become available.

The brief description above of the proposed new architecture may suggest to the reader that it must support quite a variety of types and formats for information. This is quite true. But the demands on the IDIOM framework are even greater since it is intended to support the model and data sharing needs not only of enterprises using SC 4 standards but also of those developing and disseminating the standards.

Demands on RDF to support IDIOM:
The web-based nature of RDF, its flexibility, its tool support, and its formality make it a desirable foundation for IDIOM. The primary challenges that we see to using Semantic Web languages for the IDIOM IT framework are

1. the expressiveness needed for its diverse content
2. the different semantics that must apply to different types of content

The latter point covers two different categories of concern. The semantics of the language needs to support notions for instance of time or change (for process or service models), as well as to close worlds such that completeness or other integrity constraints could be specified for classes of data or individual data sets (this is a typical interpretation of data models). The language also needs to provide the necessary context for interpretation to qualify a truth statement. Two common kinds of product data where this second concern would apply are measurements and simulation results. The meaning of such data can't be understood without either more information about the measurement or simulation (such as when, where, how, and under what conditions it was generated and by whom), or some claim made by some agent about what can be inferred from this measurement and its context.

We believe that making this claim should be possible with Named RDF Graphs while the Proof Markup Language (PML)[5] may provide a means of documenting a justification for the claim.

The use of the IT framework for standards development and for product design/lifecycle support will also require similar annotations to make claims regarding the status of standards or product designs with respect to standard workflows. Here again Named RDF Graphs may be useful for identifying the content having the status, and PML may be usable for describing how it came to that state (much as described in [6]).

Examples:
The following are some plain text examples illustrating some of the kinds of information that would be captured using IDIOM that may push the boundaries of standards or recommended practices for RDF.

Measurement: Temperature sensor TS_101 records that stream F_567-at-2010-04-05T11:36 has a temperature of 410 deg C.

Simulation results: On the basis of finite element analysis A (to find the peak stresses), scenario definition D (to specify the number of cycles), and material property specification M (defining the high cycle fatigue S-N curve) it is predicted that part P_101 in role R_101 will fail after 100,000 hours.

Engineering design approval/status: Process design information PD_E_101 about unit U_101 is approved for release to Fred Bloggs and Co. as the basis for mechanical engineering design.

Reference data status: The thing assigned the URI <http://standards.iso.org/iso/12345/-6/tech/widget> is defined by the English language text "A thing that is a blabla ..."; has the English language label "Widget"; has the French language label "Blabla"; and has ISO stage code 60.60 (i.e. is published as IS).

Summary:
It is important for standards and data related to the design and production of manufactured products (i.e. industrial data) to be clear, correct, and unambiguously interpreted. Misinterpretation of this information could lead to production problems, user dissatisfaction, unnecessary expenses, or even injury or loss of life. The formality of logic languages offer capabilities to address this need previously unavailable in standards-based information technologies. The IDIOM framework recommends the use of RDF and OWL to capture, relate, and describe this information because of their formal underpinning as well as their pragmatic features for web use. But the expressivity limitations and interpretation as truth statements that come with this formality create new challenges when applied to the richness and diversity of industrial data.

We ask that the World Wide Web Consortium (W3C) consider these challenges while it evaluates if and how RDF should evolve. This may be key in determining the level of use of RDF in the development of standards, models, and data for the industrial domain, as well as for domains with similar characteristics such as electric grid management and building construction.

Specific things that we would like to see from the W3C and the RDF community are

- guidance on how to use RDF in a coherent manner to describe and manage information in a heterogeneous environment that mixes content captured as RDF with content in other modeling languages such as EXPRESS, UML, and BPMN
- maturation of RDF named graphs to Recommendation
- maturation of a standard way to express workflows that describe the meaning of status metadata associated with an RDF graph
- guidance on the use of RDF and RDF Named Graphs to represent data with associated provenance metadata
- guidance on the use of RDF and RDF Named Graphs to represent data with associated metadata for expressing context such as for expressing measurement information
- guidance and potentially standardized vocabularies to express beliefs, and to relate those beliefs to the graphs that motivate them

References:

1. ISO 10303-11:2004, Industrial automation systems and integration – Product data representation and exchange – Part 11: Description methods: The EXPRESS language reference manual.
2. Wikipedia entry for EXPRESS http://en.wikipedia.org/wiki/EXPRESS_(data_modeling_language)
3. SC 4 Future Architecture Wiki, http://futurearch.wikispaces.com
4. Ontology Summit 2009 Communiqué: Toward Ontology-based Standards, version 1.1.1, http://ontolog.cim3.net/cgi-bin/wiki.pl?OntologySummit2009_Communique#nid1WLI
5. Proof Markup Language (PML) Primer, http://iw.stanford.edu/2005/wd-pml-primer/pml-primer-20051128.htm
6. James R. Michaelis, Li Ding, Zhenning Shangguan, Stephan Zednik, Rui Huang, Paulo Pinheiro da Silva, Nicholas Del Rio, Deborah L. McGuinness: Towards Usable and Interoperable Workflow Provenance: Empirical Case Studies Using PML, In Proceedings of the First International Workshop on the Role of Semantic Web in Provenance Management