Semantic Information Modeling for Federation

Eclipse Open Source Software and OMG Open Specifications

March 25th 2012

Cory Casanave  (cory-c at modeldriven dot com)
Introduction

Cory Casanave

CEO, Model Driven Solutions

Actionable Architectures & Agile Solutions

Information Federation, SOA, EA and MDA Development

Open Source Supporting a Model Driven Approach (ModelDriven.org)

Object Management Group

Board of Directors

Standards work: UML, SoaML, BPMN, EDOC, AESIG, SIMF, Etc.

Chair – GovDTF - Open Government Workgroup, Architecture Ecosystem SIG

W3C

Government Linked Data (GLD) Workgroup

Government

NIEM: Co-chair NIEM-UML PIM Submission Team

CIO Council/DAS: Open Government Vocabularies Workgroup
What is SIMF?

- Semantic Information Modeling for Federation is an in-progress standards process within OMG
- The goal of SIMF is to provide the modeling capabilities to support information federation by leveraging conceptual and logical information modeling with model bridging relations
- Initial submissions for SIMF are due August 13th, 2012
Problem statement

- **Federation** (information sharing, interoperability, shared services, etc.) is the **problem of this decade** – it is costing productivity, lives and billions trillions of dollars annually. It is the pre-requisite to solving many problems in the large. It is a problem faced by most CIOs in government and industry.

- We are calling this the “data problem”

A problem not solved...

- None of the standards we have **directly target this problem**. Not: UML, OWL, LoD, E/R, SOA, DoDAF, XML Schema, Common Logic or SBVR, etc.

- With all these solutions – we still have a pervasive problem!

- **While not ideal, the standards above can and are used for federation**, but, they are all built for other purposes and **repurposed to solve the data problem**. Experts can pull these technologies together to solve a specific problem, we want to make it easy to do so with an integrated and standardized approach supporting mainstream solutions and **internet-scale federation**.

We can make a substantial dent in the data problem with new standards derived from current technologies and practices. This is the “SIMF” Initiative.
What is Information Federation?

Combining multiple *independently conceived* data sources and using them together for analytics and other purposes.

**Example:** A sales department may want to combine public, internal and external information about prospect companies as part of their CRM system.

**Key term:** *Independently conceived*

- Different data sources may use different structures, technologies, vocabularies, identifiers or theories when expressing information about the *same things*.

Sharing information between potentially independent organizations (and their independently conceived systems).

- **Example:** U.S. Government Information Sharing Environment (ise.gov) initiative to combat terrorism and other threats to the U.S.
What is Information Federation?

Enabling collaborative processes that may cross organizational boundaries.

- **Example**: An agency wants to outsource human resources but needs to understand how the processes, services and information of their internal department can be satisfied by an external provider. Information federation is essential.

Service Oriented Architecture Mediation and Brokering

- **Example**: U.S. States provide services to access healthcare information but each State’s service is different. The federal government as well as other states need to interact. Some level of mediation is required across these independently conceived services. Information exchange and federation is the essence of SOA.
Importance of Information Federation

Life and death
- Combatting terrorism
- Joint forces acting safely and effectively on the same battlefield
- Situational awareness
- Intelligence

Efficiency:
- Better decision making through analytics
- Shared Resources
- Shared Processes
- Shared Services
- Shared Information

Agility
- Componentized development
- Agile iteration
- Repurpose data & other assets
- Reuse
- Stakeholder centric design

Cost Reduction
- Reduced time/cost to integrate systems
- Reduced time/cost to build and maintain systems
- Reduced time/cost to find and use data

This is a multi-billion (trillion?) dollar problem that has not received the dedicated attention it deserves!
Federation - State of the art

Point-point structural transformation of data {**Let's hack a solution point**}

- Representations: XSD, XSLT, Copybooks, Code {Which are not accessible to most stakeholders}

Standardized or centralized data structures or APIs {**One size fits all**}

- Representations: XSD, SQL-DD, UML, “Master Data Management”
- Service/API Definitions: WSDL, Corba, SoaML

Canonical data model with proprietary/structural mapping {**Convert to MY WAY**}

- Representations: E/R, UML Classes, RDFS, Code/Proprietary, Data Warehouse

Web data with point-point links {**Publish now, federate later**}

- Representations: RDFS+ (SEMWEB/LOD)

Conceptual or logical models (sometimes) with logical links {**Abstraction**}

- Representations: Ontologies, Rules, UML (With Extensions), SBVR
The conceptual pivoting approach

- A common and growing approach to the data problem leverages abstraction: Defining a domain focused vocabulary with integrity rules and assertions as part of a conceptual model that captures domain semantics. Federation and integration is achieved by relating various logical and physical information structures to the conceptual model.

- Information federation and integration is achieved via a “pivot” through this conceptual semantic layer.

- This approach is used, in part, in existing standards such as CCTS (Core Components), ISO 20022 and is currently being utilized in OMG for finance.

- In the majority of cases the “tool” used to represent these common semantics and links is a spreadsheet, but UML and OWL are also used.
Example of “Pivoting” through a conceptual model

There is an actual “Person”, Cory Casanave

- There is a concept of this person shared in this room, right now
- Here is one representation of him
- “Person” is a shared concept, independent of data structures
- There may also be shared agreement that Cory is a person and some other “facts”
  - “Cory Casanave” is a name for this person
  - He weighs 240 LBS
- There are multiple data representations about Cory Casanave which may or may not agree
- Those representations can be grounded in concepts (semantics), assisting federation
Conceptual modeling with relations to structural models is not new

- It is done with a variety of representations
  - UML, OWL, RDFS, E/R, Spreadsheets, FOL Ontologies, SBVR
- With a variety of linking and transformation mechanisms
  - Code, XSLT, FOL, OWL, Rules, QVT, Proprietary
- **What seems to work now – working with what we have**
  - Conceptual UML models with extensions for linking, transformed to RDF-LOD
  - RDFS models with rules and a bit of OWL
  - Structured English (i.e. SBVR) representations of conceptual models
  - A bit of structural mapping, some proprietary solutions
- None of these approaches seem ideal for the task and all require substantial expertise, more than is practical for mainstream adoption. But, they can inform SIMF “built for purpose” standards and tools.
Semantic Information Modeling for Federation

OMG RFP focused on Federation through Conceptual Modeling
SIMF Architecture

Subject focused conceptual models define the concepts, predicates, integrity rules and terms of a domain that can be related to each other.

Solution focused logical information elements represent information structures and integrity rules that can use and extend other information.

Technology focused physical data schema are grounded in logical data models which define their context and semantics.
SIMF Language Definition

OMG Diagram Definition
- Defined Using
- Represented In
- For
- Optional
- Defined Using
- Defines

SIMF Notation Graphical + Textual
- Defined Using
- For
- Optional
- Defined Using
- Defines

SIMF Metamodel
- Defined Using
- Optional
- Defined Using
- Defines

OMG MOF/SMOF
- RDF
- XML

SIMF Conceptual Model
- Conceptual Domain Concepts
- Model Bridging Concepts
- Logical Information Concepts

Kernel Ontology
- Defined Using
- Grounds
- Defined Using
- Defined Using
- Defined Using

Formal Logic
- Defined Using
- Defined Using
- Defined Using

Semantic Bridge
- Defines
- RDF
- XML
SIMF Language Federation

OWL Metamodel

XSD Metamodel

SQL DDL Metamodel

UML Metamodel

E/R Metamodel

SIMF Metamodel

Semantic Bridge

Semantic Bridge

Semantic Bridge

Semantic Bridge

SIMF Conceptual Model

Conceptual Domain Concepts

Model Bridging Concepts

Logical Information Concepts

Abstraction of information modeling concepts found in more than one reference language
Summary of SIMF Requirements

Conceptual Model of SIMF expressed in SIMF notation
   Conceptual Domain Model
   Logical Information Model
   Model Bridging Relations

Grounding in formal logic (Common Logic Default)

Textual and graphical notations

Bridging to common information modeling languages
   ER, SQL DDL, XSD, UML, SBVR, OWL, RDFS

Metamodel and exchange format for OMG-MOF and (Optionally) RDF
How does SIMF relate to...

As a **federation capability**, “**overlap**” with other views of information, semantics and conceptual models is **required and intended**.

So if your saying : We can do that with {OWL, Rules, UML, EMF, XSLT, CL, My Product...} we want to listen.

It is expected that other standards will be proposed by submitters to fulfill requirements as part of the SIMF specification.

Since there are multiple choices for what to reuse and how these existing standards should be integrated into the SIMF solution, the choice of standards to leverage for the SIMF domain specific language is the **purview of the submitter and not prescribed by the RFP**. It is intended that SIMF build on existing languages!
SIMFTeam Introduction – who we are

SIMFTeam is one SIMF submitter group

• Model Driven Solutions - Cory Casanave, Ed Seidewitz, Tom Digre
• PNA Group - Sjir Nijssen, Mathieu Klinger, Koen van Leeuwen, Jean Paul Koster, Inge Lemmens
• TIBCO – Paul Brown
• TMForum - Alex Zhdankin (Cisco), Nigel Davis (Ciena)
• European Space Agency - Serge Valera
• Laboratory for Applied Ontology (Brazil) – Giancarlo Guizzardi
• Deere – Roger Burhart
• Agile Birds SPRL - Sylvain Guérin
• ABN AMRO Bank - Andre Le Cat
• Turien Insurance - Jos Rozendaal
• ING Bank - Lex Bruil
• Pension Fund - Jos Vos
• Individuals - Miriam Wesseling
SIMF and Eclipse

• SIMF Requires a MOF metamodel representation
  • However – the foundation is not necessarily MOF as the foundation must have a formal logic grounding
  • It is not decided if it is “primarily” a MOF specification or based on something else like OWL or Common Logic
  • SMOF (Multiple classification) would probably be required for a reasonable representation
• Eclipse could make an excellent foundation for a SIMF modeling capability
• Eclipse could be used during the submission process, to build a proof of concept
• Eclipse could also be used for development based on the adopted standards
• We are interested in engaging with the Eclipse community to provide solutions for information federation
SIMF and Linked Data

- Linked data provides a platform for data to be ubiquitously available and linked.
- The linking and semantics of the links is not well defined.
- OWL has largely not proved successful for wide-scale federation of independently conceived data – it is fragile, lacks expressability and is not stakeholder friendly.
- Linked Data (RDF) could become the delivery platform for data described using SIMF.
- OWL and RuleML Semantics can contribute to SIMF semantics, perhaps also be used as part of an implementation.
- XSD Data structures will be federated with the SIMF conceptual model.
- RDF representation of SIMF models provide for a SEMWEB definition.
- Perhaps this could become the design language for Semantically Linked Data (SDL)?
SIMF RFP Status

• Issued by OMG
• Submissions teams are forming

Find more about SIMF here:
• http://tinyurl.com/SIMFRfp

Email list and wiki are open (OMG membership not required) – get involved!

Contact

(cory-c at modeldriven dot com)