The Model Driven Semantic Web

David Frankel
David Frankel Consulting
741 Santiago Court
Chico, CA 95973
df@davidfrankelconsulting.com

Pat Hayes
Institute for Human & Machine Cognition
University of West Florida
40 S. Alcaniz Street, Pensacola, FL 32502
phayes@ihmc.us

Elisa Kendall*
Sandpiper Software
2053 Grant Road #162
Los Altos, CA 94024
ekendall@sandsoft.com

Deborah McGuinness
Knowledge Systems Laboratory
Stanford University
Gates Bldg. 2A, Stanford CA
dlm@ksl.stanford.edu

ABSTRACT

Researchers in knowledge representation and practitioners of the OMG's Model Driven Architecture® (MDA) are beginning to recognize that there is overlap in these technologies and that there may be synergies that could be exploited to enable new, breakthrough capabilities in software engineering. The purpose of this paper is to highlight some of the commonalities and unique characteristics of each that may benefit the other and discuss areas of potential synergy. We also summarize features of the current draft revised ODM submission, and conclude with thoughts on areas where further collaboration might benefit both communities.

INTRODUCTION

MDA® consists of a number of the Object Management Group (OMG)'s standards and best practices that collectively support a wide range of software engineering disciplines, with the goal of insulating business applications from the constantly evolving technologies that implement them. The family of specifications under the MDA umbrella includes:

- The Unified Modeling Language (UML®) [1].
- The Meta-Object Facility (MOF™) [2].
- The Common Warehouse Meta-model (CWM™) [3].

Of these, MOF is most central to the MDA approach.

The MOF defines the metadata architecture that lies at the heart of MDA. MOF technology helps to automate metadata management. What the MDA world calls metadata includes database schema, UML models, workflow models, business process models, business rules, API definitions, configuration and deployment descriptors, and so on.

MDA seeks to deal with the current situation in modern organizations in which each kind of metadata is an island that is disconnected from other metadata. MOF defines standards for automating the physical management and integration of different kinds of metadata. To add a kind of metadata to the kinds of metadata that MDA tools can manage, it is necessary to define a MOF model of that kind of metadata. Such a model is usually called a "metamodel."

MOF-based tools use metamodels to generate code that manages metadata embodied as XML documents, Java™ objects, and CORBA® objects. The generated code presents and implements Java or CORBA APIs that make it possible to read and manipulate the metadata, and also serializes metadata from Java or CORBA objects to XML documents and vice versa. By automating the production of such code according to standardized patterns, MOF tools relieve programmers of tedious, repetitive chores and enforce a consistent approach to managing disparate kinds of metadata.

MOF is actually a core standard that is augmented by the following adjunct standards[1]:

- XML Metadata Interchange (XMI®), which defines how to manage MOF metadata as XML documents [5].
- CORBA Metadata Interface (CMI), which defines how to manage MOF metadata as CORBA objects [6].
- Java Metadata Interface (JMI), which defines how to manage MOF metadata as Java objects [7].

XMI and CMI are OMG standards, while JMI is a Sun Java Community Process standard. Some MOF-based tools also support Web service interfaces to MOF metadata.

Kinds of metadata for which the OMG has defined or is defining metamodels include relational database modeling, hierarchical database modeling, online analytical processing (OLAP), business process definition, business rules specification, XML, UML®, and CORBA IDL. As the OMG continues to define new metamodels, an infrastructure for connecting organizations' metadata islands is forming.

Notably, the Eclipse integrated development environment uses MOF technology to create an ecosystem in which multiple development tools can coexist and share metadata. Eclipse was originally an IBM open source product that now is under the control of an independent organization and has gained enormous traction in the industry over the past few years. It has extensive support for plugging third party tools into its environment, and uses technology based on MOF and XMI to integrate the metadata that drives multiple cooperating tools in its ecosystem.

The Semantic Web activity of the W3C also builds on an extensive body of research in knowledge representation conducted over more than two decades. Examples of results from the KR community include a number of first-order and higher order logic languages such as Knowledge Interchange Format (KIF) [8], and its recent successor, Simple Common Logic (SCL) [9], the family of languages under the description logics umbrella [10], including the Web Ontology Language (OWL) [11,12,13], and many others of varying levels of expressivity. Prior to development of the Resource

[1] See [4] for more detail on the MOF, XMI, JMI, and CMI and how they are used together to facilitate MDA

* Contact Author

Description Framework (RDF) [14], which OWL extends and refines, researchers in KR had recognized the need to address issues such as component-based, collaborative ontology development, notably in Ontolingua [15], ontology analysis and alignment [16,17,18], and ease of use [19,20,21]. The emergence of a family of languages and approach to knowledge representation compatible with other emerging standards for representation of information in XML was inevitable.

Despite such progress, KR researchers are only beginning to address the fact that emerging applications in collaboration, application integration, web services, and content management require large, complex ontologies that must be built and maintained by distributed teams. Subject matter experts with little or no background in knowledge representation methods need tools that will enable them to develop knowledge bases and related intelligent systems, tools providing capabilities for directly importing knowledge not only from formal knowledge bases but also from reference vocabularies, other repositories, and relevant applications. The portions of knowledge bases that are imported from disparate resources then need to be merged or aligned to one another in order to link the corresponding terms, to remove redundancies, and to resolve conflicts. Scalable, graphical, multi-user, component-based construction, diagnosis, and analysis tools will be essential if large, distributed teams are to effectively develop and compose reusable ontology fragments for use in new applications as envisioned for the Semantic Web [22,23,24].

POTENTIAL AREAS OF SYNERGY

Many current ontology development systems, such as Ontolingua and Chimaera [16,17,18], Protégé [25,26], OilEd [27], WebOnto [28], SymOntos [29], and LOOM [30,31,32] were developed by the research community and provide limited standards-based interoperability support or inherent analysis, merging, alignment, or composition of ontology components [33]. Their use requires deep understanding of knowledge representation languages and methodologies, and none scale to the degree required for the construction of complex and sizable ontologies needed for many commercial or government applications. Most commercial tools have similar scalability limitations and require significant expertise to use. Examples include Network Inference's Construct tool, EPM Technology's VisualExpress, Ontoprise's OntoEdit tool, SemTalk's ontology editor, IODE from OntologyWorks, and Coherence from Unicorn Software. Many of these tools are single user; few support component-based development, integrated consistency checking or validation, and most provide only limited configuration management, versioning, or other software engineering support.

Additionally, most knowledge representation based development environments are oriented towards ontology and taxonomy formation rather than terminology reconciliation and knowledge base development [34]. Vulcan estimates the cost of encoding 50 pages of high-school level chemistry knowledge to be approximately $10,000 per page of text. The costs associated with knowledge base development and lifecycle maintenance must be dramatically reduced for enterprise scale projects to succeed over time.

There are a number of areas of overlap between Semantic Web languages and UML and MOF specifications, including subsumption relations, relationships among classes and individuals, and so forth. That is not to imply that there is a one-to-one mapping among these constructs, but that there are similarities that can be used to facilitate the use of MDA technologies as a basis for ontology modeling and management in MOF repositories. These similarities, and the fact that UML and MOF are graphical in nature (and thus more accessible to subject matter experts than the typical knowledge representation language), form the basis for emerging work in development of an Ontology Definition Metamodel (ODM) by members of the OMG's Ontology PSIG [35,36]. Most UML tools are multi-user and many are integrated with commercial-quality configuration management capabilities. Some also provide facilities for round-trip engineering of resources that may be of interest as source material for ontology development. Many of these features can be extended to support ontology authoring given that the abstract syntax of the languages can be successfully mapped [37].

One of the primary distinctions between Semantic Web and MDA technologies is their focus – MOF, in particular, is concerned with automating the *physical management and interchange* of metadata, while knowledge representation is focused on the *semantics embodied in the content* of the metadata as well as on automated reasoning over that content. This is a critical distinction that also points to the potential areas where each technology can benefit from the other. MDA technologies are used by industrial and government information technologists developing large-scale enterprise, real-time, and embedded systems, without grounding in the semantics of the domains they support. Where such grounding does exist, it is often tightly coupled with the applications themselves, such as in current network management and security applications. Emerging Semantic Web technologies have been used primarily in applications involving terminology resolution, or to ensure unambiguous, machine interpretation of content, but few such projects have been implemented on a broad scale (with notable exceptions in the bioinformatics domain).

Emerging applications in finance, healthcare, security, communications, business intelligence, and many other vertical markets are content and context sensitive, and also require enterprise scalability and performance. MDA and Semantic Web technologies together can potentially provide breakthrough solutions to these kinds of problems, for example:

- Providing interoperability among multiple controlled vocabularies with interdependent and sometimes conflicting semantics (*e.g.*, in healthcare applications with numerous insurance industry encoding standards, drug formularies, disease vocabularies, and the like).
- Where policy-based applications with declarative, rather than procedural, rules representation can

increase scalability, such as for network security and management.

- Where reasoning capabilities can provide insight into, or limit potential, policy violation, predict optimal communications paths, or answer previously unanswerable questions, such as for business intelligence applications.

Many businesses and government agencies have accumulated a wealth of knowledge represented in relational databases and related UML and entity-relationship models, data warehouses, XML schemas and documents, and other metadata repositories that can provide the basis for ontology development. By leveraging MOF metamodels of such metadata silos and mappings from such models to and from the ODM, MOF and ODM-based knowledge engineering tools can automate the import of relevant portions of the metadata to jumpstart the ontology and knowledge base development process. This capability alone can dramatically reduce the cost of development, which, given Vulcan's estimates, is a tremendous impediment to widespread adoption of Semantic Web technologies today.

Knowledge representation and Semantic Web technologies can also provide significant benefits to more traditional enterprise software engineering and current MDA best practices. Declarative, policy-based approaches to complex systems development and management can dramatically improve scalability and maintainability, particularly where:

- Numerous, interrelated or dependent and complex business processes are required, such as in network security and management and telecommunications operations.
- There are large numbers of complex production rules, as in many manufacturing environments.
- MDA's design by contract (DBC) as been rigorously applied (resulting in a large number of invariant rules and constraints).

In such cases, the rules can easily be migrated to a knowledge representation formalism, uncoupling the rules from the source code and dramatically increasing scalability as a result. This migration would also enable reasoning systems to provide analysis capabilities, determine whether or not there are inherent inconsistencies among the rules or policies, and uncover any conflicts among them, eliminating inherent processing inconsistencies that are currently difficult to detect.

TOWARD A MODEL DRIVEN SEMANTIC WEB

In March 2003, the OMG initiated the formal process of soliciting proposals for an Ontology Definition Metamodel (ODM), with goals including:

- Enabling development of ontologies in UML tools
- Implementation of such ontologies in OWL without loss of fidelity
- Forward and reverse engineering of ontologies

That initial RFP led to several formal initial submissions in August 2003, and an ongoing, joint revised submission effort by companies and organizations including:

- Distributed Systems Technology Centre (Dr. Robert Colomb, University of Queensland).
- Gentleware AG (Marko Boger), with assistance from AT&T Government Solutions (Patrick Emery and Lewis Hart)
- IBM (Dr. Daniel Chang and Dr. Yiming Ye)
- Sandpiper Software (Elisa Kendall), with assistance from David Frankel Consulting (David Frankel), Stanford University Knowledge Systems Laboratory (Dr. Deborah McGuinness), and the Institute for Human and Machine Cognition, University of West Florida (Dr. Patrick Hayes).

The results of this joint effort will enable interoperability among a number of key knowledge representation and metadata management technologies in addition to meeting the goals set out by the OMG in the initial RFP. Interoperability will be accomplished through a set of distinct, but related metamodels that comprise the current ODM specification, including:

- A high-level metamodel representing core description logics (DL) concepts to be used primarily for mapping strategies among the other metamodels
- Metamodels representing the abstract syntax of RDFS and OWL
- A metamodel representing the abstract syntax of Simple Common Logic (SCL), which may be used either as a constraint language for the ODM or as a native knowledge representation language
- A metamodel representing the abstract syntax for Topic Maps
- A metamodel representing key concepts central to entity-relationship modeling, which is intended to support the use of ER diagrams as input to the ontology development process.

An incomplete draft revised submission document containing these metamodels is available for comment from the OMG Ontology PSIG web site [38]. In addition to this set of metamodels, the submitters will provide a mapping strategy that limits loss due to differences in expressivity among the various metamodels and the languages they represent, including bidirectional mappings from the DL core metamodel to the other metamodels that comprise the ODM, mappings from the DL core metamodel, RDFS, and OWL to UML 2.0 using MOF 2.0 Query / View / Transformation (QVT) [39], and one or more UML profiles for use by tool vendors in implementing the ODM, as appropriate. As of this writing, the mappings and related strategy are in process, and thus we anticipate that there may be changes required due to issues uncovered or as a result of feedback received on the draft document.

Anticipated capabilities that the resultant ODM specification will support include:

- Interoperability between native OWL ontologies and XMI-based OWL ontologies.
- Use of UML, ER, and CWM-based models as a basis for ontology development.
- Use of UML based development tools for ontology modeling, supporting not only generation of OWL, but of SCL and Topic Maps.
- Interoperability with MOF-based tools that can provide repositories for model management as well as interchange with other metadata and applications for which MOF metamodels have been defined.

Participants in the submission have also developed early implementations of a number of the metamodels and features of the ODM, and anticipate further work in this area prior to final adoption of the specification.

Areas for Further Research

Beyond the ODM, significant work remains in determining strategies for applying other capabilities that UML tools and MDA development practices can bring to the knowledge representation community, such as configuration management and version control. Configuration management is much less straightforward for ontologies than for software components, particularly in cases where ontology evolution is not necessarily under the control of the organization depending on a particular ontology. How ontology evolution affects reasoning and the applications that embed inference technologies is an area of high interest. Ontology and knowledge base analysis, alignment, merging, and composition capabilities can also benefit from MDA technologies, but more work needs to be done to understand the issues and opportunities once ODM-based tools are available. How MDA based applications and practitioners will ultimately leverage ontologies and knowledge-based capabilities is also an open question, which availability of ODM tools will help answer over time.

Acknowledgment

The authors wish to thank the other ODM submission team members, including Dan Chang, Yiming Ye, Patrick Emery, Lewis Hart, Bob Colomb, and Mark Dutra for their contributions to this effort. The opinions expressed herein represent the opinions of the authors, and do not necessarily reflect the views of the entire ODM submission team.

References

[1] *Unified Modeling Language™ (UML®) Infrastructure Specification, Version 2.0*, Object Management Group, Inc., Needham, MA, December 2003.

[2] *Meta-Object Facility (MOF™) Specification, Version 2.0*, Object Management Group, Inc., Needham, MA, August 2003.

[3] *Common Warehouse Metamodel (CWM™) Specification, Version 1.1*, Object Management Group, Inc., Needham, MA, March 2003.

[4] David S. Frankel, *Model Driven Architecture: Applying MDA to Enterprise Computing*, John Wiley & Sons, 2003, ISBN 0471319201.

[5] *XML Metadata Interchange, Version 2.0*, OMG document formal/03-05-02.

[6] *Java Metadata Interface Specification*, Java Specification Request 40, Java Community Process, http://jcp.org/jsr/detail/40.jsp.

[7] MOF 2.0 IDL, OMG document ad/04-01-16.

[8] Michael Genesereth & Richard Fikes, *Knowledge Interchange Format, Version 3.0 Reference Manual*, KSL Report KSL-92-86, Knowledge Systems Laboratory, Stanford University, June 1992.

[9] Harry Delugach, *ISO/IEC WD 24707 Information technology – Common Logic (CL) – A Framework for a Family of Logic-Based Languages*, Pacific Northwest National Laboratory, Chantilly, VA, 7 June 2004.

[10] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, P.F. Patel-Schneider (editors), *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press, January 2003.

[11] Deborah L. McGuinness and Frank van Harmelen, editors, OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004, http://www.w3.org/TR/2004/REC-owl-features-20040210/.

[12] Michael K. Smith, Chris Welty, Deborah L. McGuinness, OWL Web Ontology Language Guide, W3C Recommendation 10 February 2004, http://www.w3.org/TR/2004/REC-owl-guide-20040210/

[13] Mike Dean and Guus Schreiber, eds., Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. OWL Web Ontology Language 1.0 Reference, W3C Recommendation, World Wide Web Consortium, Amsterdam, The Netherlands, 10 February 2004. Latest version is available at http://www.w3.org/TR/owl-ref/ .

[14] Dan Brickley and R. V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, World Wide Web Consortium, Amsterdam, The Netherlands, 10 February 2004. Latest version is available at http://www.w3.org/TR/rdf-schema/ .

[15] Farquhar, Adam, Richard Fikes, and James Rice, *The Ontolingua Server: A Tool for Collaborative Ontology Construction*. KSL Report KSL-96-26, Knowledge Systems Laboratory, Stanford University, September 1996.

[16] McGuinness, Deborah L. "Conceptual Modeling for Distributed Ontology Environments." In *Proceedings of the Eighth International Conference on Conceptual Structures Logical, Linguistic, and Computational Issues (ICCS 2000).* Darmstadt, Germany. August 14-18, 2000.

[17] McGuinness, Deborah L., Richard Fikes, James Rice, and Steve Wilder. “The Chimaera Ontology Environment”, in *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI 2000)*. Austin, Texas. July 30 - August 3, 2000.

[18] McGuinness, Deborah L., Richard Fikes, James Rice, and Steve Wilder. "An Environment for Merging and Testing Large Ontologies", in *Proceedings of the Seventh International Conference on Principles of Knowledge Representation and Reasoning (KR2000)*, Breckenridge, Colorado, April 12-15, 2000.

[19] Stephen Cranefield and Martin Purvis, "UML as an Ontology Modeling Language", in *Proceedings of the Workshop on Intelligent Information Integration, 16th International Joint Conference on Artificial Intelligence (IJCAI 99)*, 1999.

[20] Kogut, Paul, Stephen Cranefield, Lewis Hart, Mark Dutra, Kenneth Baclawski, Mieczyslaw Kokar, Jeffrey Smith. "UML for Ontology Development", in *Knowledge Engineering Review, Special Issue on Ontologies in Agent Systems*, December 2001.

[21] Kendall, Elisa F., Mark E. Dutra, and Deborah L. McGuinness. "Towards A Commercial Ontology Development Environment", in *Proceedings of the International Semantic Web Conference, Late-Breaking Topics*. Sardinia, Italy, June 9-12, 2002.

[22] Tim Berners-Lee and Mark Fischetti, *Weaving the Web: the Past, Present and Future of the World Wide Web by Its Inventor,* Orion Business, London, 1999.

[23] Tim Berners-Lee, James Hendler, Ora Lassila, "The Semantic Web", *Scientific American*, May 2001.

[24] Michael C. Daconta, Leo J. Orbst, and Kevin T. Smith, *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*, Wiley Publishing Inc., 2003, ISBN 0471432571.

[25] Gennari, J. H., H. Cheng, R. B. Altman and M. Musen, *Reuse, CORBA, and Knowledge-Based Systems*. SMI Report SMI-97-0687, Stanford Medical Informatics, Stanford University, 1997.

[26] Q. Li, P. Shilane, N. F. Noy, M. Musen, *Ontology Acquisition from On-line Knowledge Sources*. SMI Report SMI-2000-0850, Stanford Medical Informatics, Stanford University, 2000.

[27] Bechhofer, Sean, Ian Horrocks, Carole Goble, Robert Stevens. "OilEd: a Reason-able Ontology Editor for the Semantic Web", in *Proceedings of KI2001, Joint German/Austrian Conference on Artificial Intelligence*, September 19-21, Vienna, Springer-Verlag LNAI Vol. 2174, pp. 396--408. 2001.

[28] Domingue, J., Motta, E. and Corcho Garcia, O. *Knowledge Modeling in WebOnto and OCML: A User Guide*. The Knowledge Media Institute, The Open University, Milton Keynes, UK, 1999.

[29] Anna Formica and Michele Missikoff. "Concept Similarity in SymOntos: An Enterprise Ontology Management Tool", in *The Computer Journal*, Volume 45, Issue 6, pp. 583-594, Oxford University Press, 2002.

[30] R. M. MacGregor, "A Description Classifier for the Predicate Calculus", in *Proceedings of the Twelfth National Conference on Artificial Intelligence*, *(AAAI 94)*, pp. 213-220, 1994.

[31] R. M. MacGregor, "Using a Description Classifier to Enhance Deductive Inference", in *Proceedings of the Seventh IEEE Conference on AI Applications*, Miami, Florida, February 1991.

[32] R. M. MacGregor, "The Evolving Technology of Classification-based Knowledge Representation Systems", in *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, Chapter 13, John Sowa, Ed., Morgan-Kaufman, 1990.

[33] Michael Denny. *Ontology Tools Survey, Revisited*. O'Reilly Media, Inc., July 14, 2004. Available at: http://www.xml.com/pub/a/2004/07/14/onto.html.

[34] See http://www.projecthalo.com/halotempl.asp?cid=20.

[35] Ontology Definition Metamodel Request for Proposal, OMG Document ad/2003-03-40.

[36] Lewis Hart, Patrick Emery, Robert Colomb, Kerry Raymond, Dan Chang, Yiming Ye, Elisa Kendall, and Mark Dutra, "Usage Scenarios and Goals for Ontology Definition Metamodel", to appear in *Proceedings of the The Fifth International Conference on Web Information Systems Engineering (WISE'04)*, November 22 – 24, 2004.

[37] Ceccaroni, Luigi and Elisa Kendall. "A Graphical Environment for Collaborative Ontology Development", in *Proceedings, The Second International Joint Conference on Autonomous Agents & Multi Agent Systems*, Poster Session, July 14-18, 2003, Melbourne, Australia, ACM Digital Library.

[38] See http://codip.grci.com/odm/draft/.

[39] MOF 2.0 Query/View/Transformation Revised Submission, ad/04-04-01, April 5, 2004.