Extracting Ontologies from Standards: Experiences and Issues

Ken Baclawski*, Yuwang Yin, Sumit Purohit
College of Computer and Information Science
Northeastern University

Eric S. Chan*
Oracle

*Co-chair, OASIS ICOM TC
Abstract

- We describe our experiences with extracting formal representations from an OASIS draft standard.
- Representations include OWL-DL, OWL2 (all profiles), RDF/S, XSD, SQL and Java (annotated with JWS, JAXB and JPA).
- Procedural (behavioral) capabilities include persistence (with ORM), data representation transformers (such as XML to RDF), wiki forms, and web services (both SOAP and REST).
- Multiple representations were found to have many advantages
  - Correcting inconsistencies
  - Enriching the standard with more semantics
  - Enhancing interoperability
- Most of the extractions and transformations were automated, and similar techniques may be applicable to other standards that are specified by documents that are reasonably well structured.
The Integrated Collaboration Object Model (ICOM) for Interoperable Collaboration Services

- A Framework for Integrated Collaborative Work Environments
- OASIS Technical Committee
- Currently finalizing Committee Specification
- Builds on existing standards and technologies
- Core module models actors, groups, roles, artifacts, access control and metadata.
- ICOM Technical Committee Draft

Representation Languages

- Standards that enable interoperability between systems should also enable interoperability between languages.
- There are a great many ways to represent (express) a standard (some examples are in parentheses):
  - Natural language (English)
  - Controlled natural language (SBVR)
  - Graphical language (UML)
  - OO programming language (Java, C#)
  - Data language (SQL, XML Schema, OData)
  - Ontology language (RDF, OWL, CL)
- These have different advantages and purposes so it makes sense for a standard to be expressed using more than one or even all of them.
Conceptual Levels

• Metalevels
  • M0: Data (Abox)
  • M1: Schema (Tbox)
  • M2: Language Metadata

• Transformations
  • Normally on one metalevel
  • Sometimes cross metalevels

• Directionality
  • Single directional (may lose information)
  • Two directional (usually requires annotations)
1405 OASIS Standards Documents were processed. Of these 1405 documents, 5 of them are the ICOM drafts.
Schema (M1) Mappings

- RDF/S
- OWL Profiles
- Common Logic
- Podcast Ontology
- ICOM Standard In English
- XSLT
- Purple Semantic MediaWiki (PSMW)
- XSD
- Java
- SQL
- UML
- WSDL
- REST
- xjc
- jpa
- soaprest
- wsgen
-箔

- Fully developed
- Prototype only
Data (M0) Transformations
Integrated Collaboration Object Model (ICOM) for Interoperable Collaboration Services Version 1.0

Committee Specification Draft 04 /
Public Review Draft 04

26 June 2012

Specification URIs
This version:
http://docs.oasis-open.org/icom/icom-ics/v1.0/csprd03/icom-ics-v1.0-csprd03.doc
(Authoritative)
http://docs.oasis-open.org/icom/icom-ics/v1.0/csprd03/icom-ics-v1.0-csprd03.html
http://docs.oasis-open.org/icom/icom-ics/v1.0/csprd03/icom-ics-v1.0-csprd03.pdf

Previous version:
http://www.oasis-open.org/committees/download.php/44405/icom-ics-v1.0-csprd02.zip

Latest version:
http://docs.oasis-open.org/icom/icom-ics/v1.0/icom-ics-v1.0.doc (Authoritative)
http://docs.oasis-open.org/icom/icom-ics/v1.0/icom-ics-v1.0.html
http://docs.oasis-open.org/icom/icom-ics/v1.0/icom-ics-v1.0.pdf

Technical Committee:
OASIS Integrated Collaboration Object Model for Interoperable Collaboration Services (ICOM) TC

Chair:
Eric S. Chan (eric.s.chan@oracle.com), Oracle
Kenneth P. Baclawski (kenb@ccs.neu.edu), Northeastern University

Editors:
Eric S. Chan (eric.s.chan@oracle.com), Oracle
Patrick Durusau, (patrick@durusau.net), Individual
ICOM Metadata in XML format

<standard>
<title>Integrated Collaboration Object Model (ICOM) for Interoperable Collaboration Services Version 1.0</title>
<version>Committee Specification Draft 04 / Public Review Draft 04</version>
<url>http://docs.oasis-open.org/icom/icom-ics/v1.0/csd04/icom-ics-v1.0-csd04.html</url>
<date>26 June 2012</date>
<abstract>The Integrated Collaboration Object Model (ICOM) for Interoperable Collaboration Services defines a framework for integrating collaboration activities, by encompassing and improving on a range of models which are part of e.g.:
</abstract>
<br>
<chair>Eric S. Chan (eric.s.chan@oracle.com), Oracle</chair>
<chair>Kenneth P. Baclawski (kenb@ccs.neu.edu), Northeastern University</chair>
<citeAs>[ICOM-ics-v1.0]</citeAs>
</integratedCollaborationObjectModel>
<editor>Eric S. Chan (eric.s.chan@oracle.com), Oracle</editor>
<editor>Patrick Durusau, (patrick@durusau.net), Individual</editor>
<latestVersion>http://docs.oasis-open.org/icom/icom-ics/v1.0/icom-ics-v1.0.html</latestVersion>
<latestVersion>http://docs.oasis-open.org/icom/icom-ics/v1.0/icom-ics-v1.0.pdf</latestVersion>
/latestVersionAuthoritative>http://docs.oasis-open.org/icom/icom-ics/v1.0/icom-ics-v1.0.doc</latestVersionAuthoritative>
<namespace>http://docs.oasis-open.org/ns/icom</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/core/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/accesscontrol/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/metadata/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/content/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/document/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/message/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/presence/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/contact/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/calendar/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/task/201008</namespace>
<namespace>http://docs.oasis-open.org/ns/icom/forum/201008</namespace>
<previousVersionAuthoritative>http://www.oasis-open.org/committees/download.php/44405/icom-ics-v1.0-csprod02.zip</previousVersionAuthoritative>
<status>This document was last revised or approved by the OASIS Integrated Collaboration Object Model for Interoperable Collaboration Technical Committee members should send comments on this specification to the Technical Committee’s email list. Others should send comments for information on whether any patents have been disclosed that may be essential to implementing this specification, and any offers of:
<br>
technicalCommitteeMember>OASIS Integrated Collaboration Object Model for Interoperable Collaboration Services (ICOM) TC</technicalCommitteeMember>
</status>
</standard>
4.5.7 AccessControlList

4.5.7.1 Description
An access control list (ACL) is an object attached to an entity to specify a list of permissions to access the entity.

4.5.7.2 Class Definition
The AccessControlList class has attribute values:

- **localNamespace**
  - Value: icom_ac

- **localName**
  - Value: AccessControlList

- **extendsFrom**
  - Value:

- **stereotype**
  - Value: primary

- **description**
  - Value: An access control list (ACL) is an object attached to an entity to specify a list of permissions to access the entity.
### Examples of ICOM Properties

#### 3.5.7.3 Property Definitions

The **AccessControlList** class MUST have the property definitions:

<table>
<thead>
<tr>
<th><strong>icom_ac:object</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>Associated object.</td>
</tr>
<tr>
<td>Required:</td>
<td>True</td>
</tr>
<tr>
<td>Inherited:</td>
<td>False</td>
</tr>
<tr>
<td>Property Type:</td>
<td><strong>icom_core:Entity</strong></td>
</tr>
<tr>
<td>Cardinality:</td>
<td>Single</td>
</tr>
<tr>
<td>Updatability:</td>
<td>On Create</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>icom_ac:accessControlEntry</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Description:</td>
<td>One or more access control entries.</td>
</tr>
<tr>
<td>Required:</td>
<td>True</td>
</tr>
<tr>
<td>Inherited:</td>
<td>False</td>
</tr>
<tr>
<td>Property Type:</td>
<td><strong>icom_ac:AccessControlEntry</strong></td>
</tr>
<tr>
<td>Cardinality:</td>
<td>Multi</td>
</tr>
<tr>
<td>Updatability:</td>
<td>Read Write</td>
</tr>
</tbody>
</table>
An access control list (ACL) is an object attached to an entity to specify a list of permissions to access the entity.
ICOM Properties in OWL

<owl:ObjectProperty xml:base='&icom_ac;#' rdf:ID='accessControlEntry'>
  <rdfs:label xml:lang='en'
    rdf:datatype='&xsd;#string'>accessControlEntry</rdfs:label>
  <rdfs:domain rdf:resource='&icom_ac;#AccessControlList'/>
  <rdfs:range rdf:resource='&icom_ac;#AccessControlEntry'/>
</owl:ObjectProperty>

<owl:ObjectProperty xml:base='&icom_ac;#' rdf:ID='object'>
  <rdfs:label xml:lang='en'
    rdf:datatype='&xsd;#string'>object</rdfs:label>
  <rdfs:domain rdf:resource='&icom_ac;#AccessControlList'/>
  <rdfs:range rdf:resource='&icom_core;#Entity'/>
</owl:ObjectProperty>
Issues Raised and Lessons Learned

- A standard document was first converted to HTML.
  - HTML errors are commonplace
- Subsumption (superclass) inferences and transitive closures were performed.
- Metadata was generated.
- 25 consistency checks were performed.
  - Most were errors; some only warnings.
- Inverse properties can be problematic because of property scopes.
  - In principle, only a naming issue.
  - Requirement for interoperability makes it significant.
More Issues and Lessons

- Name conflicts occurred between some enumeration instances and classes
  - This is not allowed in OWL1 but is allowed in OWL2.
  - The names were changed to avoid the conflicts.
- Chimeric properties
  - Whether object or datatype property depends on the domain.
  - Range and inverse depend on the domain.
- Abstract categories
  - Are they covered by (union of) their immediate subcategories? The intention varied.
- Each OWL2 profile imposes restrictions.
Purple Semantic MediaWiki (PSMW)

- A bundle of MediaWiki plugins that add semantic and fine-grained access.
- Mapping ontologies to PSMW:
  - Classes map to Categories
  - Properties are specified with Templates
  - Forms use one or more templates for data entry and display
  - Applications use SPARQL queries
- We are automating schema mapping and data import/export for ICOM and other ontologies.
Generated PSMW Pages

- These were generated for the ICOM AccessControlList class (style information was omitted for clarity):
  - Template (for forms)
  - Form (for data entry)
  - Category (to classify data)

Template:AccessControlList

```plaintext
{|-
! accessControlEntry
| {{#arraymap:{{{AccessControlEntry|}}}|,|x|[AccessControlEntry::x]}}
|-
! object
| [[Entity::{{{Entity|}}}] ]
|
```
An access control list (ACL) is an object attached to an entity to specify a list of permissions to access the entity.

[[Create a new access control list::AccessControlList]]
Generated Data Entry Form

Create AccessControlList: Acl1

accessControlEntry: 

object: 

Summary: 

☐ This is a minor edit  ☐ Watch this page 

Save page  Show preview  Show changes  Cancel

Privacy policy  About PSMW Calendar  Disclaimers
Example Applications

- Podcast Ontology
  - Annotation import/export from/to PSMW
- Calendar
  For more documentation and a demo see:
  http://project.cim3.net/wiki/PSMW-CALENDAR_Presentation-02
- The Ontolog Wiki content is being migrated to PSMW.
PSMW Calendar

Sparql query:

```sql
{{#ask: 
   [[Category:Occurrence]]
   [[StartDate::>Oct 01 2012]]
   ? StartDate
   |format=calendar
}}
```

October 2012

<table>
<thead>
<tr>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>19</td>
<td>20</td>
</tr>
<tr>
<td>21 PSMW Special call 2012 10 22</td>
<td>22 PSMW meeting call 2012 10 23</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Conclusion

- Automated extraction of a variety of representations for standards shows great promise.
  - Improving standards
  - Rapidly developing applications
  - Enhancing interoperability
- Representations can be both declarative and procedural.
- One can automate transformations and mappings at all metalevels and also between metalevels.
- OOR fits naturally as a basis for developing and supporting standards as well as ontologies.