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.
- Extension modules: Content, Document, Message, Presence, AddressBook, Calendar, FreeBusy, TaskList, Forum, Conference
- ICOM Technical Committee Draft

https://www.oasis-open.org/apps/org/workgroup/icom/download.php/47246/CSPRD05-Oct-15-2012.zip

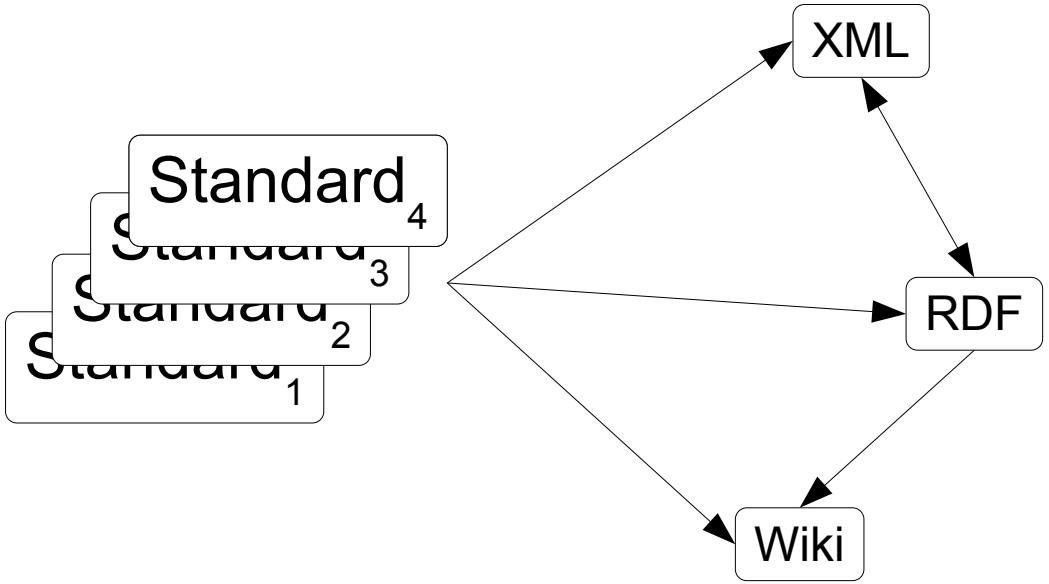
Representation Languages

- 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

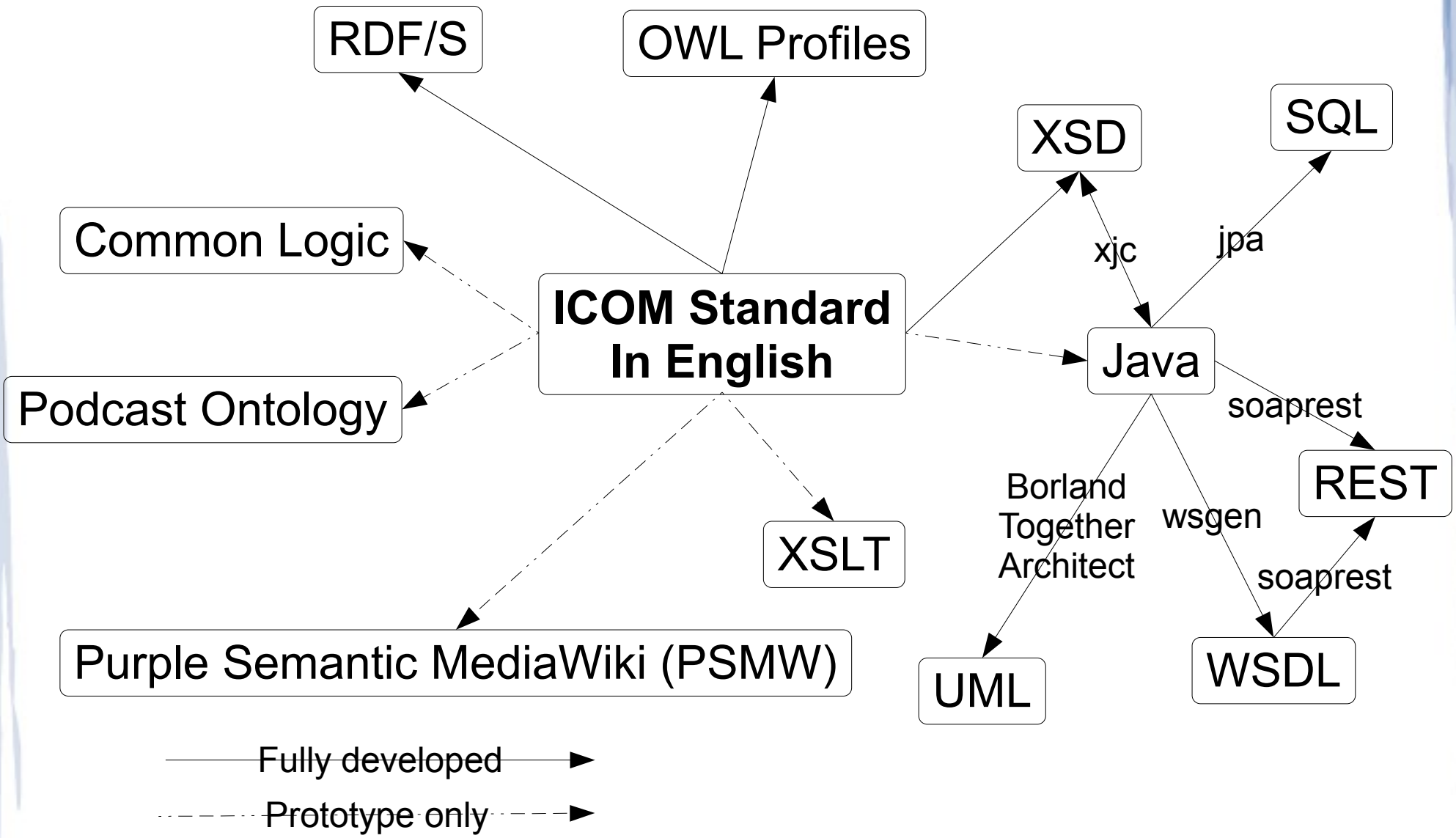
- Metalevels
  - M0: Data (Abox)
  - M1: Schema (Tbox)
  - M2: Language Metadata
- Transformations
  - Normally on one metalevel
  - Sometimes cross metalevels
- Directionality
  - Single directional (may lose information)
  - Two directional (usually requires annotations)

Metadata (M2) Extraction

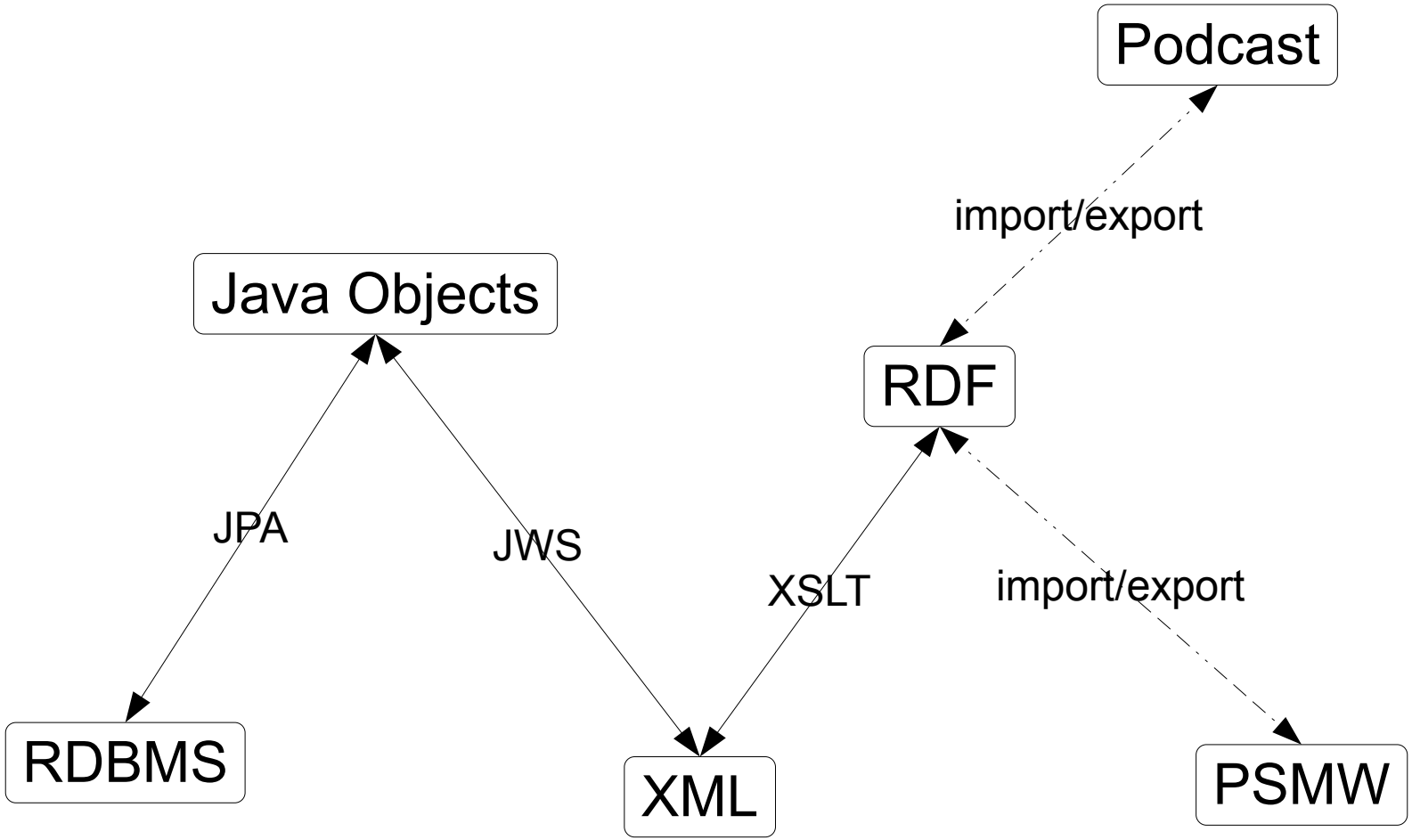


1405 OASIS Standards Documents were processed.
Of these 1405 documents, 5 of them are the ICOM drafts.

Schema (M1) Mappings



Data (M0) Transformations



OASIS 0.1

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 integrati
The framework is not intended to prescribe how applications or services conforming to its model implement, store, or transport the data
The model integrates a broad range of collaboration activities, by encompassing and improving on a range of models which are part of ex
  <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]
Integrated Collaboration Object Model (ICOM) for Interoperable Collaboration Services Version 1.0. 28 March 2012. OASIS Committee Speci
  <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/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-csprd02.zip</previousVersionAutho
  <status>This document was last revised or approved by the OASIS Integrated Collaboration Object Model for Interoperable Collaboratior
Technical Committee members should send comments on this specification to the Technical Committee's email list. Others should send comm
For information on whether any patents have been disclosed that may be essential to implementing this specification, and any offers of
  <technicalCommitteeMember>OASIS Integrated Collaboration Object Model for Interoperable Collaboration Services (ICOM) TC</technicalCo
  <thisVersion>http://docs.oasis-open.org/icom/icom-ics/v1.0/csprd03/icom-ics-v1.0-csprd03.html</thisVersion>
  <thisVersion>http://docs.oasis-open.org/icom/icom-ics/v1.0/csprd03/icom-ics-v1.0-csprd03.pdf</thisVersion>
  <thisVersionAuthoritative>http://docs.oasis-open.org/icom/icom-ics/v1.0/csprd03/icom-ics-v1.0-csprd03.doc</thisVersionAuthoritative>
 </standard>
```

Example of an ICOM Class

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:

icom_ac:object

Description:	Associated object.
Required:	True
Inherited:	False
Property Type:	icom_core:Entity
Cardinality:	Single
Updatability:	On Create

icom_ac:accessControlEntry

Description:	One or more access control entries.
Required:	True
Inherited:	False
Property Type:	icom_ac:AccessControlEntry
Cardinality:	Multi
Updatability:	Read Write

An ICOM Class in OWL

```
<owl:Class xml:base='&icom_ac;#' rdf:ID='AccessControlList'>
  <icom_meta:categoryMetadata>
    <icom_meta:Category rdf:about='&icom_ac;#AccessControlListMetadata'/>
  </icom_meta:categoryMetadata>
  <rdfs:label xml:lang='en'
    rdf:datatype='&xsd;#string'>AccessControlList</rdfs:label>
  <rdfs:comment xml:lang='en' rdf:datatype='&xsd;#string'>An
access control list (ACL) is an object attached to an entity to
specify a list of permissions to access the entity.</rdfs:comment>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource='&icom_ac;#accessControlEntry'/>
      <owl:minCardinality
        rdf:datatype='&xsd;#integer'>1</owl:minCardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource='&icom_ac;#object'/>
      <owl:cardinality
        rdf:datatype='&xsd;#integer'>1</owl:cardinality>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
<icom_meta:Category xml:base='&icom_ac;#' rdf:ID='AccessControlListMetadata'>
  <icom_core:name xml:lang='en'
    rdf:datatype='&xsd;#string'>AccessControlList</icom_core:name>
  <icom_core:description xml:lang='en' rdf:datatype='&xsd;#string'>An
access control list (ACL) is an object attached to an entity to
specify a list of permissions to access the entity.</icom_core:description>
```

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

```
{|-
! accessControlEntry
| {{#arraymap:{{{AccessControlEntry|}}}|,|x|[[AccessControlEntry::x]]}}
|-
! object
| [[Entity::{{{Entity|}}}]]
|}
```

Generated PSMW Pages

Form:AccessControlList

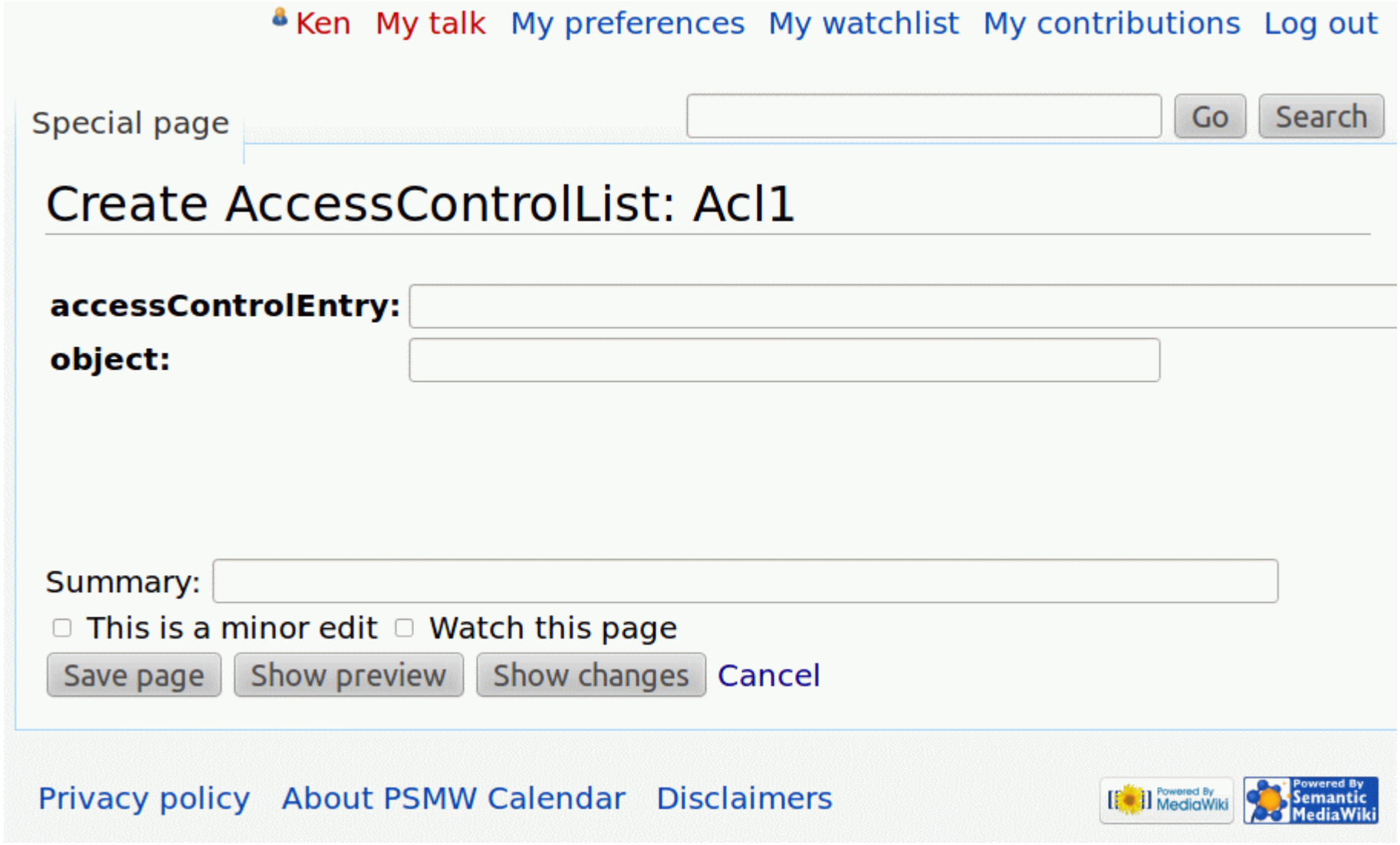
```
{{{for template|AccessControlList}}}
{| class="formtable"
! accessControlEntry:
| {{{field|AccessControlEntry}}}
|-
! object:
| {{{field|Entity}}}
|}
{{{end template}}}
```

Category:AccessControlList

```
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



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.

Page Discussion Read Edit View history Go Search

PSMW Calendar

Sparql query:

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

October 2012

◀ Today ▶

October 2012 Go to month

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

Podcast Ontology

```
<?xml version="1.0"?>
<rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:icom="http://docs.oasis-open.org/ns/icom/core/201008#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns="http://podcast.ontolog.cim3.net#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xml:base="http://podcast.ontolog.cim3.net">
  <owl:Ontology rdf:about="">
    <owl:imports rdf:resource="http://docs.oasis-open.org/ns/icom/conference/201008"/>
    <owl:imports rdf:resource="http://docs.oasis-open.org/ns/icom/core/201008"/>
    <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Created By Sumit Purohit using TopBraid Free Version</owl:versionInfo>
  </owl:Ontology>
  <rdf:Property rdf:ID="createBy">
    <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
    >Author of  podcast</rdfs:label>
  </rdf:Property>
  <rdf:Property rdf:ID="iconImage">
    <rdfs:domain>
      <icom:Folder rdf:ID="podcast">
        <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string"
        >podcast</rdfs:label>
        <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource"/>
      </icom:Folder>
    </rdfs:domain>
    <rdfs:range rdf:resource="http://docs.oasis-open.org/ns/icom/core/201008#Artifact"/>
  </rdf:Property>
  <rdf:Property rdf:ID="publishedDate">
    <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#dateTime"/>
  </rdf:Property>
  <rdf:Property rdf:ID="URLLinkedTo">
    <rdfs:range rdf:resource="http://docs.oasis-open.org/ns/icom/core/201008#Identifiable"/>
    <rdfs:domain rdf:resource="#podcast"/>
```

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.