MOD: Metadata for Ontology Description and publication

BISWANATH DUTTA (DRTC, INDIAN STATISTICAL INSTITUTE, BANGALORE, INDIA)
DURGESH NANDINI (B.C. ROY ENG. COLLEGE, DURGAPUR, INDIA)
GAUTAM KISHORE SHAHI (INST. OF TECH, BIRSA, INDIA)
Documentation Research and Training Centre (DRTC),
Indian Statistical Institute (ISI), Bangalore Centre, INDIA
Outline

- Introduction
- Ontology Metadata: Issues
- Ontology Metadata in Practice: the Current State of the Ontology Libraries
- Guiding Principles
- Methodology
- Top-level Facets
- MOD Metadata Model
- MOD Ontology View
- Conclusion and Future Work
Introduction

- Ontology construction is a **costly** affair
- The fundamental idea is **reuse** the existing ontologies before creating a new one
- Where do we look for an ontology?
- How do we find the **Mr. Right** ontology?

- Ontology **metadata**!!!
Ontology Metadata: Issues

- Ontology Metadata Vocabulary (OMV), the only metadata vocabulary available for ontology
  - Mainly deals with provenance information

- The metadata should also provide the provisions to describe the other important aspects of an ontology, such as,
  - development perspective (e.g., competency questions, ontological commitments, design decisions)
  - implementation perspective (e.g., information for reasoning support, languages, rules, conformance to external standards)
  - etc.
### Ontology Metadata in Practice: the current state of the ontology libraries

<table>
<thead>
<tr>
<th>Ontology Library</th>
<th>Number of Elements</th>
<th>Example Elements</th>
<th>Metadata Followed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bio-Portal (<a href="https://bioportal.bioontology.org/">https://bioportal.bioontology.org/</a>)</td>
<td>30</td>
<td>Acronym, People, Number Of Properties, Status, Description</td>
<td>Partially OMV plus own defined elements</td>
</tr>
<tr>
<td>Colore (<a href="https://code.google.com/p/colore/source/browse/trunk/ontologies/approximate_point">https://code.google.com/p/colore/source/browse/trunk/ontologies/approximate_point</a>)</td>
<td>7</td>
<td>Source Path, File Name, Size, Rev, Author</td>
<td>None</td>
</tr>
<tr>
<td>DAML (<a href="http://www.daml.org/ontologies/">http://www.daml.org/ontologies/</a>)</td>
<td>12</td>
<td>Link, Description, Submitter, Point of contact, Submitter</td>
<td>None</td>
</tr>
<tr>
<td>DERI (<a href="http://vocab.deri.ie/">http://vocab.deri.ie/</a>)</td>
<td>4</td>
<td>Author, Terms, Last Update, Namespace URI</td>
<td>None</td>
</tr>
<tr>
<td>Maven (<a href="http://mvnrepository.com/artifact/edu.stanford.protege">http://mvnrepository.com/artifact/edu.stanford.protege</a>)</td>
<td>4</td>
<td>Artifact, Last Version, Popularity, Description</td>
<td>None</td>
</tr>
<tr>
<td>MISO (<a href="http://www.sequenceontology.org/">http://www.sequenceontology.org/</a>)</td>
<td>6</td>
<td>Definition, Synonyms, DB Xref, Parent, Children</td>
<td>None</td>
</tr>
<tr>
<td>MMI (<a href="http://mmisw.org/">http://mmisw.org/</a>)</td>
<td>22</td>
<td>Full Title, Contact Role, Syntax Format, Authority abbreviation, Contributor, Keywords</td>
<td>None</td>
</tr>
<tr>
<td>OBO Foundry (<a href="http://www.obofoundry.org">http://www.obofoundry.org</a>)</td>
<td>12</td>
<td>Namespace, Current Activity, Help, Home, Documentation, Contact</td>
<td>None</td>
</tr>
<tr>
<td>ONKI (<a href="http://onki.fi/en/browser/">http://onki.fi/en/browser/</a>)</td>
<td>11</td>
<td>Type, URI, Share, superordinate concepts, Coordinate concepts</td>
<td>None</td>
</tr>
<tr>
<td>Ontohub (<a href="https://ontohub.org/ontologies">https://ontohub.org/ontologies</a>)</td>
<td>24</td>
<td>Project Name, Description, Institution, URL, task</td>
<td>Partially OMV plus own defined elements</td>
</tr>
<tr>
<td>ROMULUS (<a href="http://www.thezfiles.co.za/ROMULUS/">http://www.thezfiles.co.za/ROMULUS/</a>)</td>
<td>35</td>
<td>Ontology Name, License Description, Project Domain, Creation date, DL expressivity, Number of classes, Number of individuals</td>
<td>Partially OMV plus own defined elements</td>
</tr>
<tr>
<td>Schemapedia (<a href="http://datahub.io/dataset/schemapedia">http://datahub.io/dataset/schemapedia</a>)</td>
<td>4</td>
<td>Subject, Property, Source</td>
<td>None</td>
</tr>
<tr>
<td>SHOE (<a href="http://www.cs.umd.edu/projects/plus/SHOE/onts/">http://www.cs.umd.edu/projects/plus/SHOE/onts/</a>)</td>
<td>4</td>
<td>Id, Version, Description, Contact</td>
<td>None</td>
</tr>
</tbody>
</table>

- The majority of the above libraries (70%) are found to be using 15 or fewer than 15 elements.
- Different terms are used in describing similar information in different libraries (e.g., author and creator).
MOD Approach

- Two major components:
  - Guiding principles
  - Methodology
    - A two-way approach: Top-down and Bottom-up
Guiding Principles

- Principle of brevity
- Principle of clarity
- Principle of simplicity
- Principle of authority
- Principle of standardization
- Principle of extensibility
- Principle of usability
- Principle of interoperability
Methodology: Top-down approach

- It involves in looking at the **big picture** of the metadata vocabulary.
  - This is accomplished by defining the **top-level facets** conceiving the various aspects of the resource to be described (here the resource is an **Ontology**).

- Each aspects are further **analyzed and narrowed down to define the various classes**.

- The top-down approach proceeds from **an abstract level to a concrete level**.
Methodology: Bottom-up approach

- It involves studying and identifying the properties of a resource for search and discovery to facilitate their effective reuse.
  - This is accomplished by analyzing users’ ontology search behavior, search criteria and parameters.

- The extracted properties are further associated with the classes defined in the top-down approach.

- The bottom-up approach proceeds from a concrete level to an abstract level.
Methodology: Bottom-up approach (contd...2)

- Conducted a survey to understand users’ search behavior, search criteria and parameters.

- Open ended questionnaire is used to conduct the survey.

- Two questions were asked to the participants:
  - How do you search an ontology on the Web or in an ontology library?
  - When you search for an ontology, what is the information you look for before deciding to refer/consult/download it?

- Total participants were 18, of which 12 responded.
Methodology: Bottom-up approach (contd...3)

Some responses:

- **Statement 1:** look at the ontology descriptors like *domain* details, *number of classes*, *properties*, *tools used*.
- **Statement 2:** I look for *representations languages* while downloading an ontology.
- **Statement 3:** I look for SPARQL query file, if any.
- **Statement 4:** I would like to see ‘*user reviews*’ with these ontologies, so that I can save a lot of time in understanding the quality of the ontology.
- **Statement 5:** I prefer to have a *documentation/ information about the methodology* followed to develop an ontology, it will be an additional advantage.
- **Statement 6:** I remain curious about the following facts: *top classes*, *number of classes* and *class definitions*.
- **Statement 7:** I look for *types* and *number of relations*.
- **Statement 8:** I look for *number of entities* and *description* about each of them.
- ....
Top-level Facets

Seven top-level facets (aka aspects) of an ontology are identified and are defined within MOD. These are:

- **General** - an abstraction of the general aspects of an ontology, for instance, the ontologies, ontology type, etc.

- **Ontology Coverage** - an aspect that defines the domain *(a domain is any area of knowledge or field of study that we are interested in or that we are communicating about that deals with specific kinds of entities and scope of an ontology.)*

- **Authority** - describes the agents, like organizations, that own and are responsible for the ontology.

- **Rights** - describes the rights and licenses of an ontology.

- **Environment** - defines the environment in which an ontology has been built, for instance, the tool that is used to build an ontology, the level of formality, and the syntax followed.

- **Action** - an aspect highlighting the applications where an ontology is being applied or used, such as in a project.

- **Preservation** - describes the low level-features of an ontology, for instance, ontology storage, file format, etc.
MOD Metadata Model

- MOD Components:
  - Classes: 15 classes
  - Object property: 18 object properties
  - Data property: 31 data properties
## MOD Classes

<table>
<thead>
<tr>
<th>Top-level facet</th>
<th>Class Name</th>
<th>Example of Class Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>Ontology</td>
<td>Space ontology, Food ontology, Fishery ontology,</td>
</tr>
<tr>
<td>Authority</td>
<td>Agent</td>
<td>Organization related with the ontology and the person associated with it.</td>
</tr>
<tr>
<td></td>
<td>Subclass: Organization</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subclass: Person</td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>License</td>
<td>Creative Commons, GNU Free Documentation License, GNU General Public License</td>
</tr>
<tr>
<td>Scope/Coverage</td>
<td>Domain</td>
<td>Genes, Space, Medicine, Protein</td>
</tr>
<tr>
<td>Action</td>
<td>Project</td>
<td>Smart city, Mobility</td>
</tr>
<tr>
<td>Environment</td>
<td>Ontology design tool</td>
<td>OntoEdit, Protégé, TopBraid composer</td>
</tr>
<tr>
<td></td>
<td>Ontology design language</td>
<td>RDFS, OWL</td>
</tr>
<tr>
<td></td>
<td>Ontology design syntax</td>
<td>Notation3, Turtle, RDF/XML</td>
</tr>
<tr>
<td>Preservation</td>
<td>File Format</td>
<td>.rdf, .gaf</td>
</tr>
<tr>
<td></td>
<td>Level Of Formality</td>
<td>Dictionary, Glossary</td>
</tr>
<tr>
<td></td>
<td>Knowledge Representation Formalism</td>
<td>Frame, Description Logics, First Order Logic.</td>
</tr>
</tbody>
</table>
Conclusion and Future Work

- MOD is a well-guided, refined, easy-to-use standard ontology metadata vocabulary.
- MOD consists of minimal and well-defined set of metadata elements.
- The elements are mapped and standardised with the other Semantic Web metadata standards.
  - In other words, MOD reuses the terminologies of the existing metadata vocabularies.
- We plan to pursue the use of MOD in the context of ontology libraries.
THANK YOU VERY MUCH for YOUR KIND ATTENTION!!!

Questions???

Contact: bisu@drtc.isibang.ac.in/ dutta2005@gmail.com