



# MOD: Metadata for Ontology Description and publication and future plan

---

**BISWANATH DUTTA**

INDIAN STATISTICAL INSTITUTE

DOCUMENTATION RESEARCH AND TRAINING CENTRE

BANGALORE, INDIA

# Indian Statistical Institute (ISI), Bangalore Centre, INDIA

---



# Area of Interest

---

Knowledge representation

Information/ knowledge classification and systems

Ontology

Metadata

Linked Data

CMS

# Outline

---

- Introduction
- Why Metadata?
- Ontology Metadata: Issues
- Ontology Metadata in Practice: the Current State of the Ontology Libraries
- Approach
- Top-level Facets
- MOD Metadata and Overview (MOD 1.0)
- MOD 1.2
- Proposal
- Summary and Our Plan

# Introduction

---

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

# Why Metadata?

---

- Find
- Discover
- Select
- Reuse
- Administer
- Preserve

# Ontology Metadata: Issue

---

- Ontology Metadata Vocabulary (OMV), the only metadata vocabulary available for describing the ontologies
  - Fundamentally deals with provenance information (e.g., name, creator) (Obrst, et al., 2014)
- 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)
  - **usability perspective** (e.g., quality, rights)
  - etc.

**Source:** Obrst, et al. (2014). Semantic web and big data meets applied ontology. *Applied Ontology*, 9, 155-170.



# Ontology Metadata in Practice: the current state of the ontology libraries

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

- The majority of the above libraries (70%) are found to be using 15 or fewer than 15 elements.
- Different words are used for describing similar information in different libraries (e.g., {author, creator}, (name, title)).



# MOD Approach

---

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

# Guiding Principles

---

- Principle of permanence
- Principle of ascertainability
- Principle of exclusiveness
- Principle of exhaustiveness
- Principle of standardization

- Principle of brevity
- Principle of clarity
- Principle of simplicity
- Principle of authority
- 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 (in our case, 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 1.0 Model

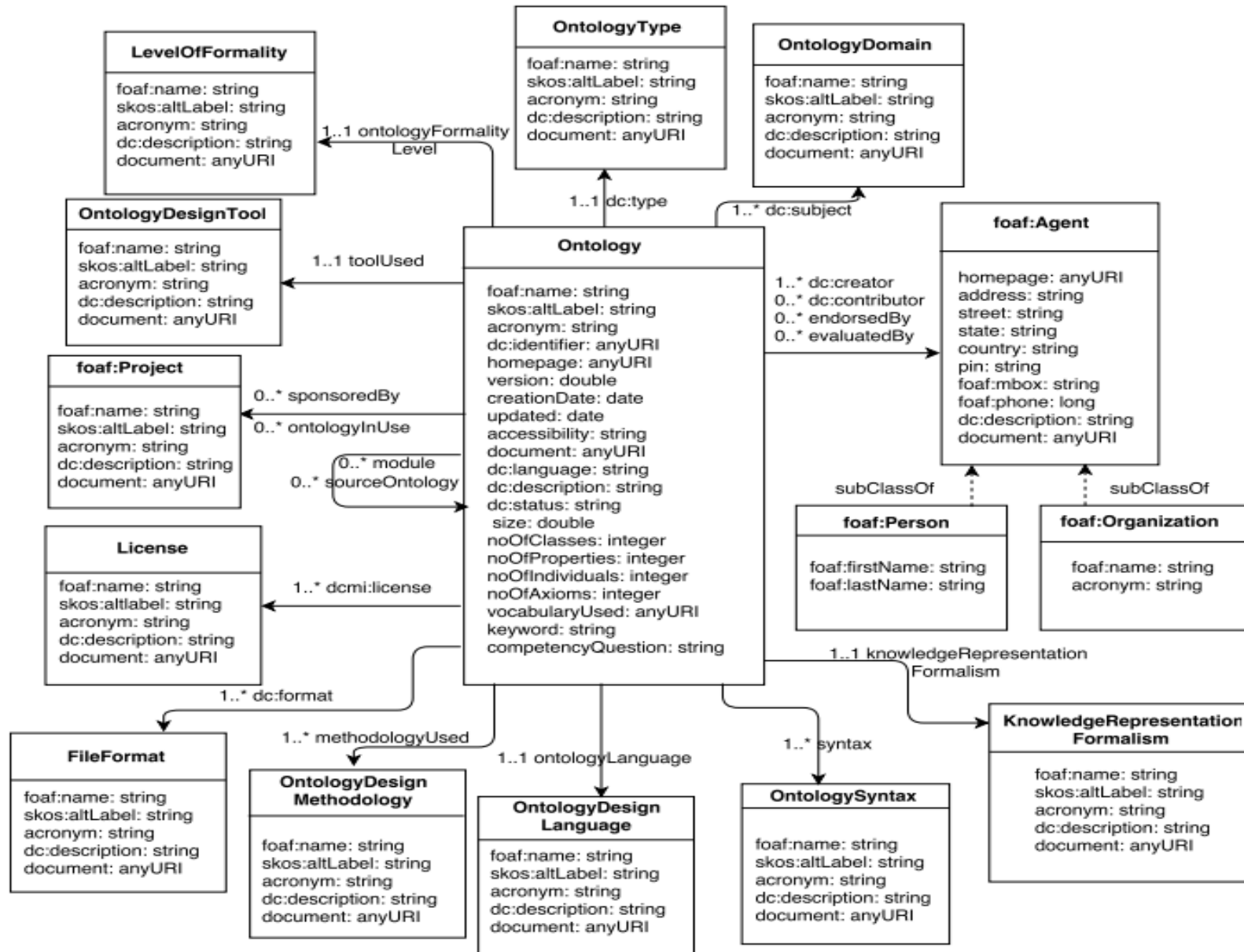
---

- MOD Components:
  - **Classes: 15**
  - **Object property: 18**
  - **Data property: 31**

# MOD 1.0 Classes

Top-level facets	Class Names	Example of Class Instances
General	Ontology	Space ontology, Food ontology, Fishery ontology,
Authority	Agent Subclass : Organization Subclass : Person	Organization related with the ontology and the person associated with it.
Right	License	Creative Commons, GNU Free Documentation License, GNU General Public License
Scope/Coverage	Domain	Genes, Space, Medicine, Protein
	Ontology type	Application Ontology, General Ontology, Core Reference Ontology
Action	Project	Smart city, Mobility
	Methodology	METHONTOLOGY, YAMO
Environment	Ontology design tool	OntoEdit, Protégé, TopBraid composer
	Ontology design language	RDFS, OWL
	Ontology design syntax	Notation3, Turtle, RDF/XML
Preservation	File Format	.rdf, .gaf
	Level Of Formality	Dictionary, Glossary
	Knowledge Representation Formalism	Frame, Description Logics, First Order Logic.

# MOD 1.0 Overview



[dc: http://purl.org/dc/elements/1.1/](http://purl.org/dc/elements/1.1/)  
[dcmi: http://purl.org/dc/terms/](http://purl.org/dc/terms/)  
[foaf: http://xmlns.com/foaf/0.1/](http://xmlns.com/foaf/0.1/) [skos: http://www.w3.org/2004/02/skos/core#](http://www.w3.org/2004/02/skos/core#)

# Some useful links

---

**Specification document is available here:** <http://www.isibang.ac.in/ns/mod.html>

**OWL file is available here:** <http://www.isibang.ac.in/~bisu/ontology/>

**Work published:** In Proceedings of DCMI International Conference on Dublin Core and Metadata Applications (DC-2015), Sao Paulo, Brazil, 1-4 September 2015, pp. 1-9.

---

# MOD 1.2

# MOD 1.2

---

- It is a follow up of MOD 1.0
- Incorporates a set of new elements and also further refines the existing MOD 1.0 elements
- Incorporated more number of **relevant existing metadata vocabularies**
- Has become a collaborative effort

# MOD 1.2 Data

---

- **Classes:** 23 (15)
- **Object Properties:** 34 (18)
- **Data Properties:** 58 (31)
  
- **Project page:** <https://github.com/sifrproject/MOD-Ontology>



# Proposal

---

- Promote the creation of **metadata@source**
  - Ontology editing tools got to play a key role here
  
- Publish metadata like a FOAF file
  - Ontology metadata harvesters can harvest and allow to do analytics

# Why metadata@source?

---

- Creator understands better his/her creation
- Lots of information can be **auto-compiled** by the ontology editors
  - E.g., creator, creation date, byte size, language, syntax, ontology metrics
- Fresh idea

# Summary

---

- MOD is a well-guided, refined, easy-to-use standard ontology metadata vocabulary.
- MOD consists of a 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.

# Our plan

---

## Short term

- Publish MOD 1.2 vocabulary
- Demonstrate the use of MOD with true ontologies (say, from BioPortal, AgroPortal)
- Promote its use as a vocabulary for ontology description and publication
  - Ontology editing tools
  - Ontology repositories
- Evaluate the work

## Long term

- Engage a bigger community
- Develop it as a standard

---

THANK YOU VERY MUCH for YOUR KIND ATTENTION!!!

Questions???

Contact: [bisu@drtc.isibang.ac.in/](mailto:bisu@drtc.isibang.ac.in) [dutta2005@gmail.com](mailto:dutta2005@gmail.com)