

NEW GENERATION METADATA VOCABULARY FOR ONTOLOGY DESCRIPTION AND PUBLICATION

¹**Biswanath Dutta**, ²Anne Toulet, ²Vincent Emonet and ^{2,3}Clement Jonquet

¹DRTC, Indian Statistical Institute, Bangalore, India

²LIRMM, CNRS & University of Montpellier, France

³BMIR, Stanford University School of Medicine, USA



MTSR 2017 (Tallinn, Estonia, 28 Nov. - 1 Dec., 2017)

Introduction

- Ontology is an intelligent knowledge artifact
- Ontology construction is a **costly** affair
- Thumb of rule 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?

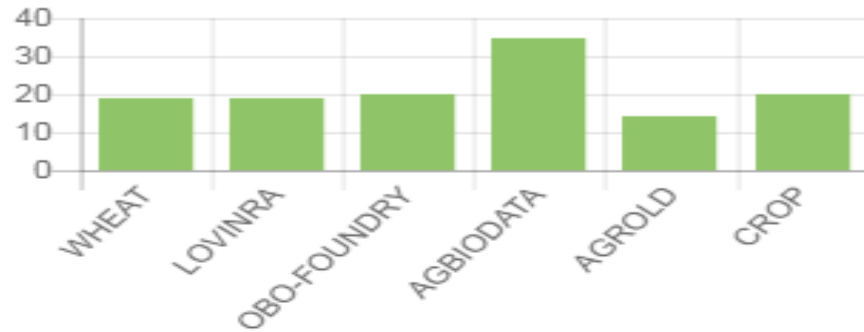
- **Metadata!!!**

Why Metadata?

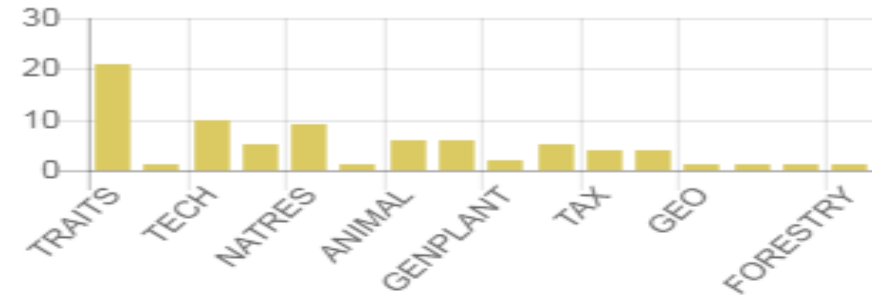
- Find
 - Discover
 - Select
 - Reuse
 - Administer
 - Preserve
-
- Will allow to ask interesting questions
 - Will enable analytics
 -

Some simple analytics (<http://agroportal.lirmm.fr/landscape>)

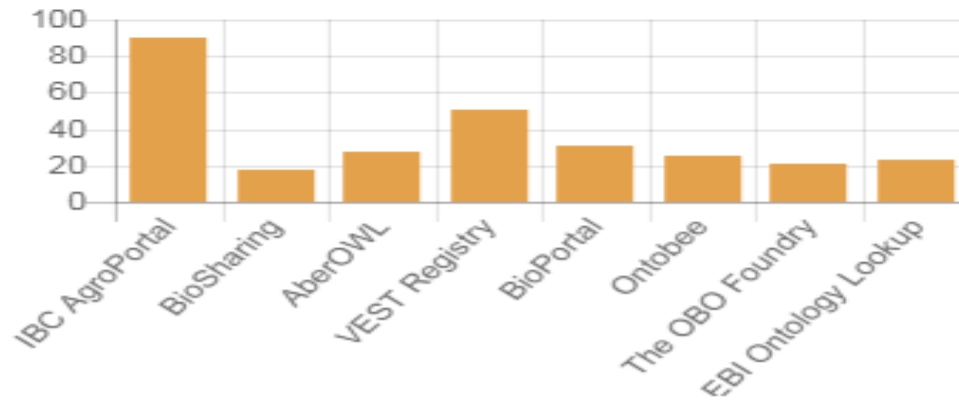
Ontologies by group



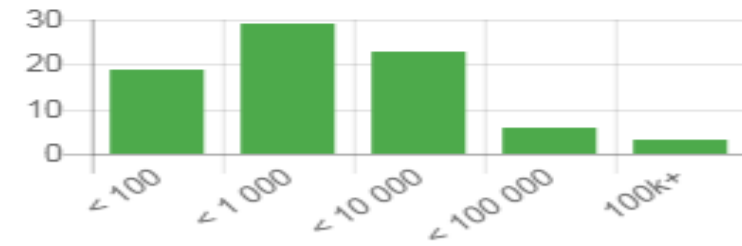
Ontologies by category



Ontologies count in each data catalog

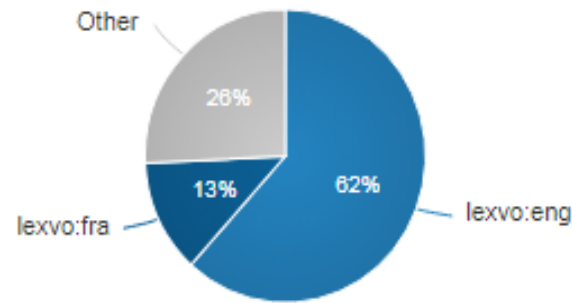


Ontologies by size

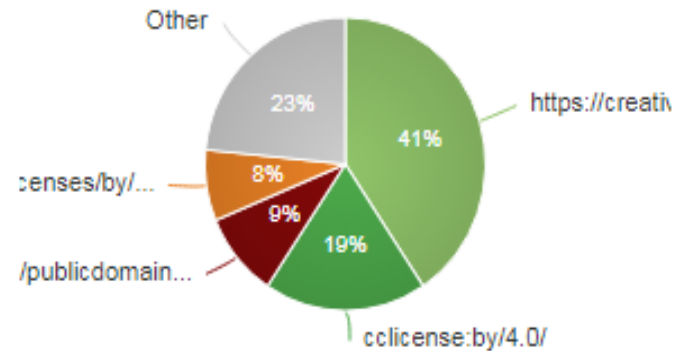


Some simple analytics (<http://agroportal.lirmm.fr/landscape>)

Ontologies natural languages



Licenses used by the ontologies



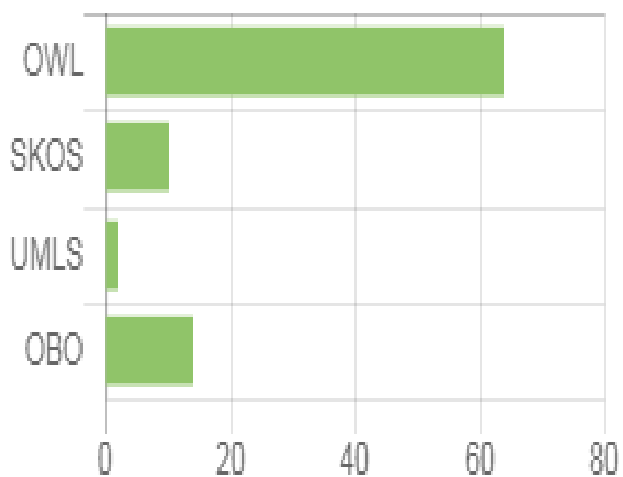
Most used tools to build ontologies



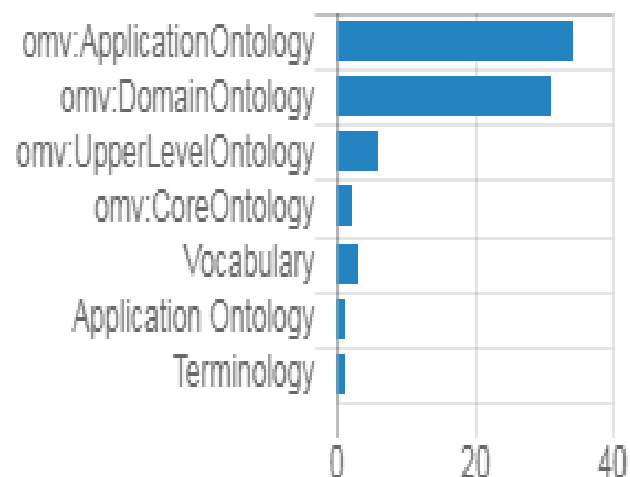
Some simple analytics

(<http://agroportal.lirmm.fr/landscape>)

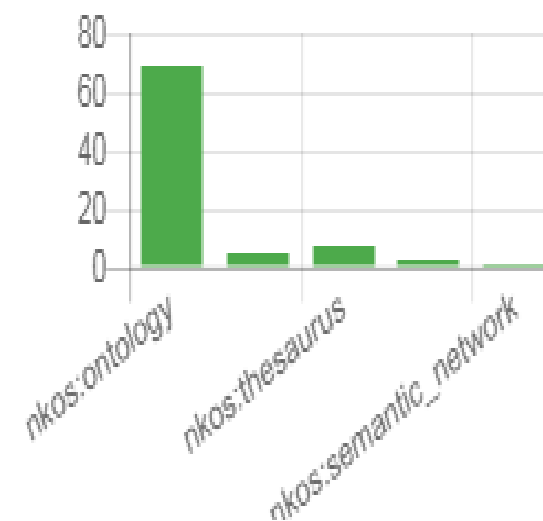
Format used



Ontology types



Ontology formality levels



Some simple analytics (<http://agroportal.lirmm.fr/landscape>)

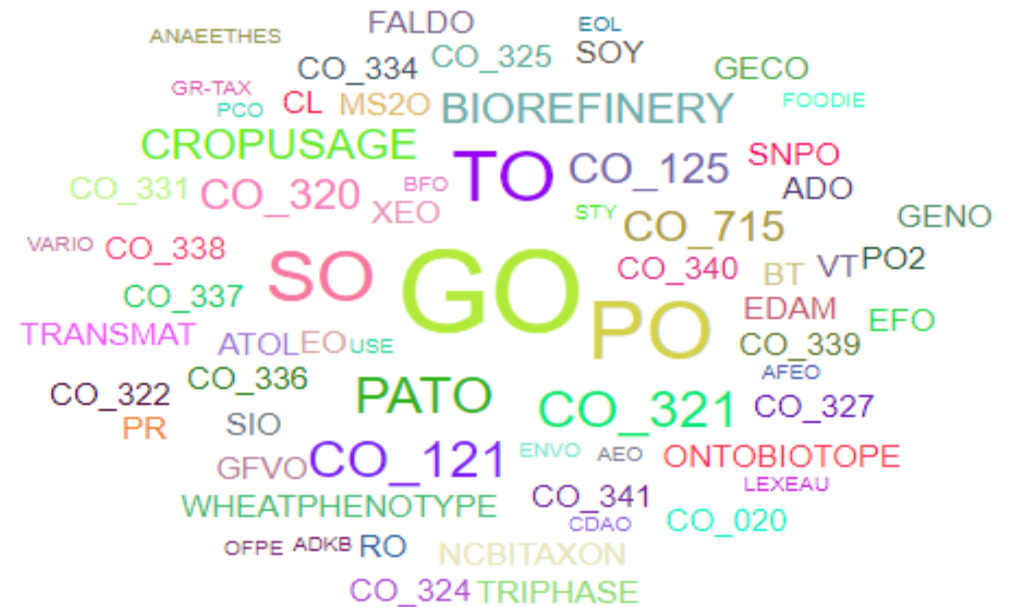
Most active people as reviewer

People that posted notes, review and projects



Most active ontologies

Ontologies that have notes, review and projects



State of the art in ontology metadata

- Conducted **three** different studies:
 1. Analysis of the **existing metadata vocabularies** for describing ontologies
 2. Analysis of the **uses of metadata vocabularies** in describing the ontologies (by the **ontology developers**)
 3. Analysis of the **uses of metadata vocabularies** in describing ontologies in various **ontology repositories/ registries**
- Besides we have also done an extensive literature survey

State of the art in ontology metadata: Findings

- **Ontology developers use a variety of metadata vocabularies** (e.g., DC, DCT, PROV, VOID, DCAT, SCHEMA)
 - **Interestingly:** the only ontology specific metadata OMV (first published in 2005) is found to be hardly used by the community
- Despite a few exceptions, **metadata vocabularies do not rely on one another** although there is a strong overlap observed
 - multiple properties to capture similar information (e.g., dc:license, and cc:license)
- Each of the reviewed libraries uses, to some extent, some metadata elements but **do not always use standard metadata vocabularies**
- The general purpose elements (e.g., rdfs:comment, owl:versionOf and owl:imports) are found to be the **most frequently used elements**

The issue

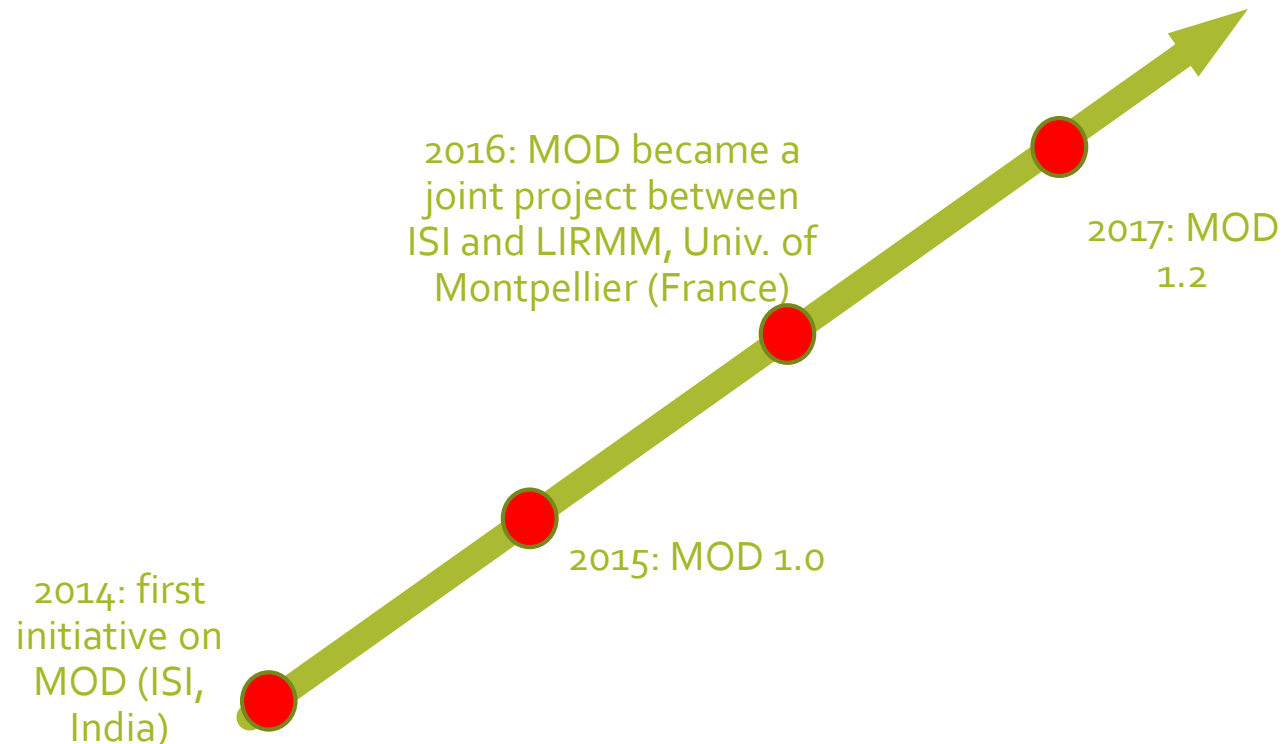
- There are many metadata vocabularies out there, but majority of them are generic in nature
- It is hard to find any of the existing vocabularies covering enough aspects of ontologies (including the most specific one) to support the new type of queries on ontologies

Objective

- Propose a new generation metadata vocabulary for ontology description and publication
 - Of course, the agenda is to integrate and harmonize the previous metadata vocabularies rather than **adding a new one to the list**

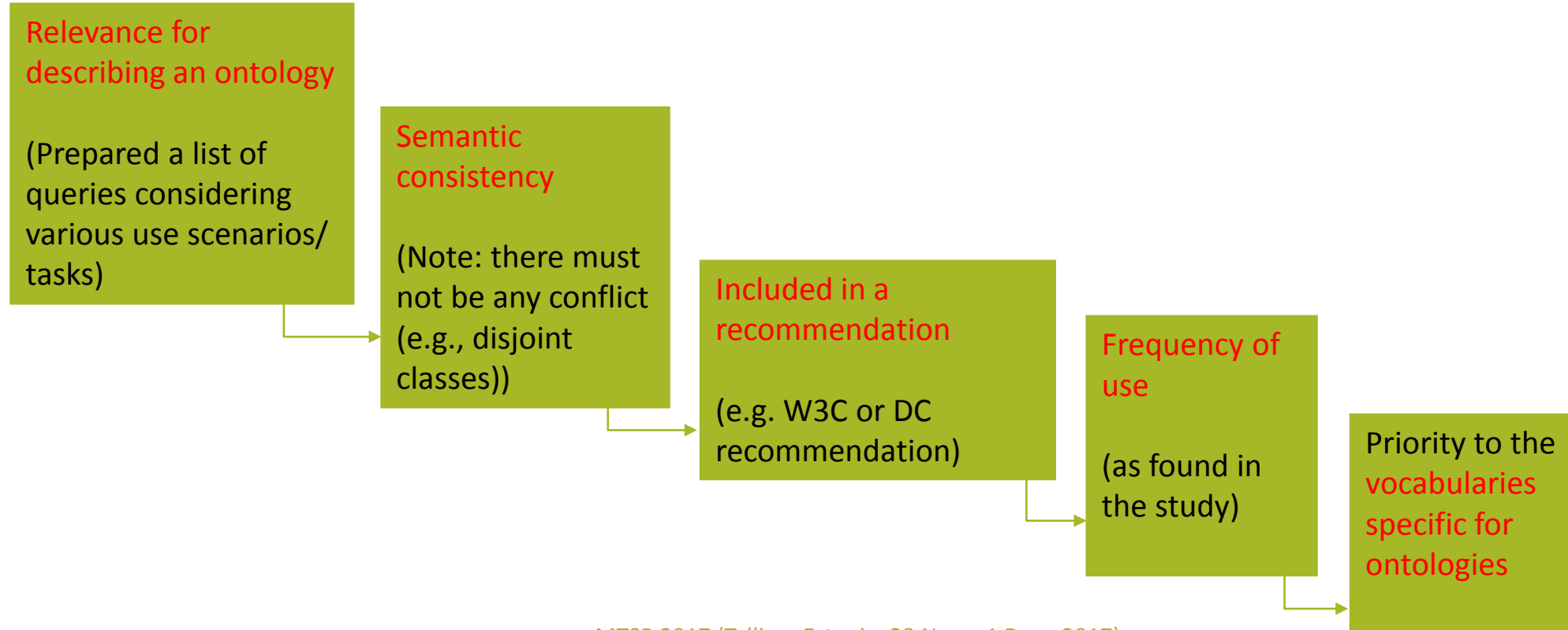


Background of MOD (Metadata for Ontology Description and publication)



MOD 1.2 design methodology

- **Principle:** rely on the existing metadata vocabularies (preferably official recommendations) and also proposes to fusion (and simplify) with the vocabularies that are specific to ontologies (e.g. OMV, MOD 1.0, VANN)



Vocabularies selected/considered in MOD 1.2: General purpose

Prefi x	Namespace	Name	Resource	R	#S	#C	Example (Selected)
dc	http://purl.org/dc/elements/1.1/	Dublin Core	NA	R	0	15	
skos	http://www.w3.org/2004/02/skos/core#	Simple Knowledge Organization System	skos:ConceptScheme	R	0	4	
rdfs	http://www.w3.org/2000/01/rdf-schema#	RDF Schema	rdfs:Resource	R	1	3	rdfs:comment
owl	http://www.w3.org/2002/07/owl#	Web Ontology Language	owl:Ontology	R	7	10	owl:priorVersion, owl:incompatibleWith, owl:priorVersion

Vocabularies selected/considered in MOD 1.2: Vocabulary specific

Prefix	Namespace	Name	Resource	R	#S	#C	Example
omv	http://omv.ontoware.org/2005/05/ontology#	Ontology Metadata Vocabulary	omv:Ontology		20	37	omv:acronym, omv:hasOntologyLanguage, omv:hasOntologySyntax, omv:designedForOntologyTask
mod	http://www.isibang.ac.in/ns/mod#	Metadata for Ontology Description & Publication 1.0	mod:Ontology		13	25	mod:competencyQuestion, mod:group, omv:keyClasses, mod:ontologyInUse
voaf	http://purl.org/vocommons/voaf#	Vocabulary of a Friend	voaf:Vocabulary		0	6	
vann	http://purl.org/vocab/vann/	Vocabulary for annotating vocabulary descriptions	rdfs:Resource		4	6	vann:preferredNamespacePrefix, vann:preferredNamespaceUri, vann:example
nkos	http://w3id.org/nkos#	Networked KOS Application Profile	rdfs:Resource		0	4	
door	http://kannel.open.ac.uk/ontology#	Descriptive Ontology of Ontology Relations	owl:Ontology		0	6	

Vocabularies selected/considered in MOD 1.2: Dataset

Prefix	Namespace	Name	Resource	R	#S	#C	Example
void	http://rdfs.org/ns/void#	Vocabulary of Interlinked Datasets	void:Dataset		0	9	
dct	http://purl.org/dc/terms/	DCMI Metadata Terms	dct:Dataset	R	29	34	dct:title, dct:identifier, dct:relation, dct:hasFormat, dct:isPartOf
dcat	http://www.w3.org/ns/dcat#	Data Catalog Vocabulary	dcat:Dataset	R	0	4	
adms	http://www.w3.org/ns/adms#	Asset Description Metadata Schema	adms:Asset		0	9	
schema	http://schema.org/	schema.org	schema:Dataset		0	33	
idot	http://identifiers.org/idot/	Identifiers.org	dct:Dataset		0	4	

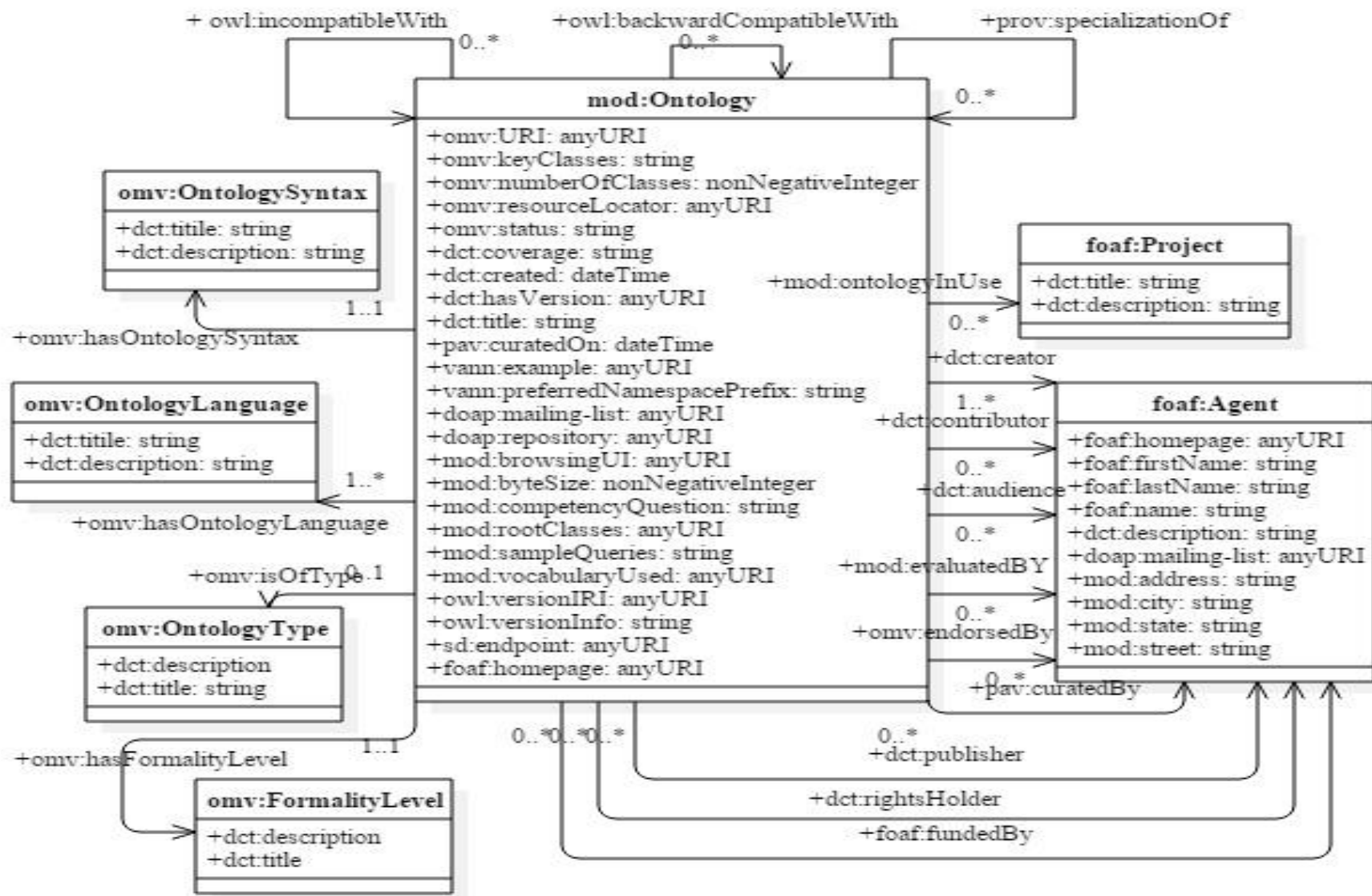
Vocabularies selected/considered in MOD 1.2: other kind of resources

Prefix	Namespace	Name	Resource	R	#S	#C	Example
foaf	http://xmlns.com/foaf/0.1/	Friend Of A Friend vocabulary	foaf:Document		5	10	foaf:fundedBy, foaf:homepage, foaf:depiction
doap	http://usefulinc.com/ns/doap#	Description of a Project	doap:Project		3	11	doap:mailing-list, doap:bugDatabase, doap:repository
cc	http://creativecommons.org/ns#	Creative Commons Rights Expression Language	cc:Work		0	3	
sd	http://www.w3.org/ns/sparql-service-description#	SPARQL 1.1 Service Description	sd:Service	R	1	1	sd:endpoint
prov	http://www.w3.org/ns/prov#	Provenance Ontology	prov:Entity	R	3	9	prov:specializationOf, prov:wasGeneratedBy
pav	http://purl.org/pav/	Provenance, Authoring and Versioning	prov:Entity		2	10	pav:curatedOn, pav:curatedBy
oboInOwl	http://www.geneontology.org/formats/oboInOwl#	oboInOwl Mappings	owl:Ontology		0	6	

MOD 1.2 vocabulary

- MOD 1.2 consists of total 19 classes, 28 object properties and 60 data properties
- It is expressed in OWL
- Out of the total 88 properties, 13 properties are newly created in “mod” namespace <http://www.isibang.ac.in/ns/mod#>

A partial view of MOD 1.2



Full view is available here: <https://github.com/sifirproject/MOD-Ontology>

Newly added properties in MOD 1.2

Properties	Definition
<code>mod:competencyQuestion</code>	A set of questions asked at design time to explain why the ontology is needed and explain its design.
<code>mod:group</code>	A group of ontologies that the ontology is usually considered into.
<code>mod:translation</code>	A pointer to the translated ontology(ies) for an existing ontology.
<code>mod:rootClasses</code>	The root class(es) of an ontology. This could be automatically populated by taking the direct subclasses of owl:Thing. If the ontology is also defined as a unique skos:ConceptScheme, then this property becomes the equivalent of skos:hasTopConcept
<code>mod:browsingUI</code>	The user interface (URL) where the ontology may be browsed or searched.
<code>mod:vocabularyUsed</code>	The vocabularies that are used and/or referred to create the current ontology.
<code>mod:sampleQueries</code>	A set of queries (may be SPARQL, DL Queries) that are provided along with an ontology to illustrate use cases.
<code>mod:ontologyInUse</code>	An ontology that is used in a project.
<code>mod:evaluation</code>	An ontology that has been evaluated by an agent.
<code>mod:numberOfObjectProperties</code>	The total number of object properties in an ontology. Refines omv:numberOfProperties.
<code>mod:numberOfDataProperties</code>	The total number of data properties in an ontology. Refines omv:numberOfProperties.
<code>mod:numberOfLabels</code>	Number of defined labels for any resources in an ontology (classes, properties, etc).
<code>mod:byteSize</code>	The byte size of an ontology file.

Experimental results

- Created a Knowledge Base (KB) of eight Agronomical ontologies (e.g. AGROVOC, Gene Ontology, National Agricultural Library Thesaurus)
 - The knowledge base consists of in total 1962 axioms, 20 classes, 33 objects and 69 data properties, and 217 individuals (available here <https://github.com/sifrproject/MOD-Ontolog>)
- Most of the metadata extracted from AgroPortal (in some cases consulted with the original source of the ontologies, and/or, the other online sources)
- **A common principle:**
- Reuse, wherever available, the existing URIs of the resources instead of creating them in “mod” namespace
 - E.g., for creating the organizational resources, we preferred to use DBPedia defined URIs.
 - In the case of unavailability, decided to use the organizational homepage URL as the resource URI
 - For other resources, used the respective standards
 - E.g. for language, used Lexvo vocabulary (www.lexvo.org), licensing (<https://creativecommons.org>)

Experimental results (2)

- The knowledge base supports the varieties of queries
 - E.g. which is the most popular ontology editing tool? Who are the key contributors in a domain? How many ontologies are produced by OBO Foundry group? What are the projects using the Protein Ontology? **What are the ontologies endorsed by the RDA Wheat Data Interoperability Group (RDA WDI) and the National Science Foundation (NSF)?**
- The queries were expressed in SPARQL and successfully run over the knowledge base.

```
SELECT DISTINCT ?Ontology ?Author
WHERE {
    {?x a mod:Ontology; omv:endorsedBy <https://www.rd-alliance.org/groups/wheat-data-
interoperability-wg.html> ; dct:title ?Ontology .}
    UNION
    {?x a mod:Ontology; omv:endorsedBy
<http://dbpedia.org/resource/Category:National_Science_Foundation> ; dct:title ?Ontology .}
    OPTIONAL {?x dct:creator ?Author .}}
```

Current state

- We are working on MOD 1.3
- MOD is made as an open project on GitHub and ResearchGate
- Recently a sub-task group on “ontology metadata” has been created under the Research Data Alliance Vocabulary and Semantic Services Interest Group (VSSIG - <https://www.rd-alliance.org/groups/vocabulary-services-interest-group.html>)

Our Future Goal (primary)

- To turn the current MOD 1.2 to a **collaborative extended** version MOD 2.0
- To produce an “application profile” for the description of ontologies
- To discuss with the various ontology editor development teams (e.g. Protégé) on integration of MOD in the software

Our Future Goal (Secondary)

- Automate the process of creating mod:Ontology instances of the ontology libraries (e.g., BioPortal, AgroPortal, OBO Foundry)
 - This will enable exporting the content of these libraries without changing their internal data models
- Release the KB as Linked Open Data
- Offer a SPARQL endpoint to provide local and remote advanced queries on ontologies

Our proposal

- Promote the creation of metadata@source
 - Ontology editing tools got to play a key role here
- Publish ontology metadata similar like a FOAF file

Thank you for listening!



- 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.

Biswanath Dutta

E-mail: bisu@drtc.isibang.ac.in