Summary Report for Ontology Metadata
task group of the
Vocabulary and Semantic Services Interest Group

Biswanath Dutta, Clement Jonquet, Barbara Magagna, Anne Toulet
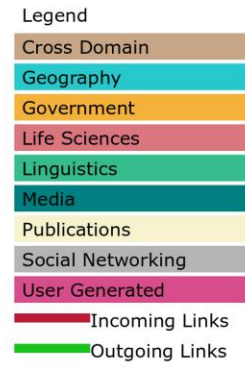
RDA P11 – Berlin, March 2018

Task group interested in studying ontology metadata practices to discuss and provide recommendations
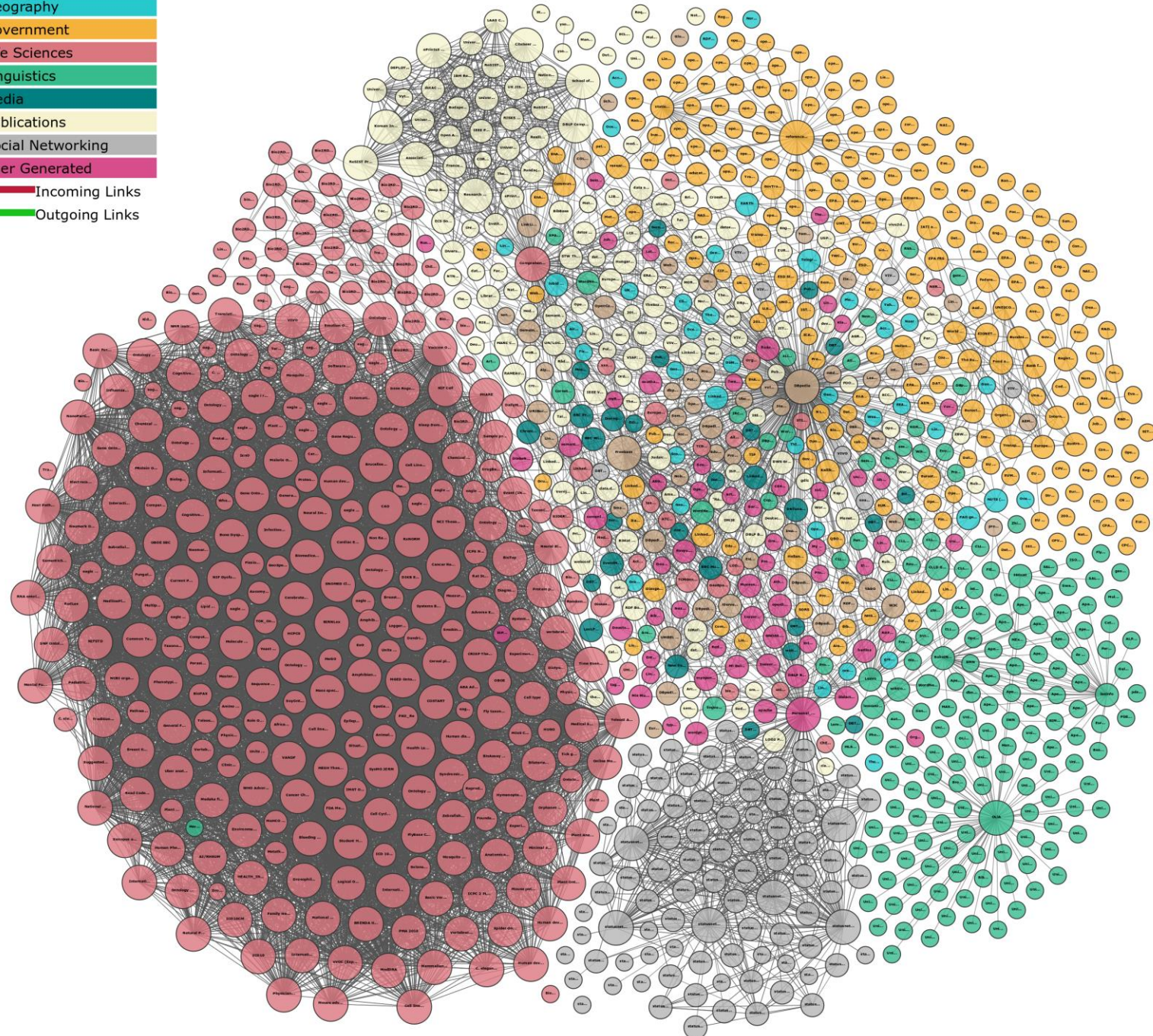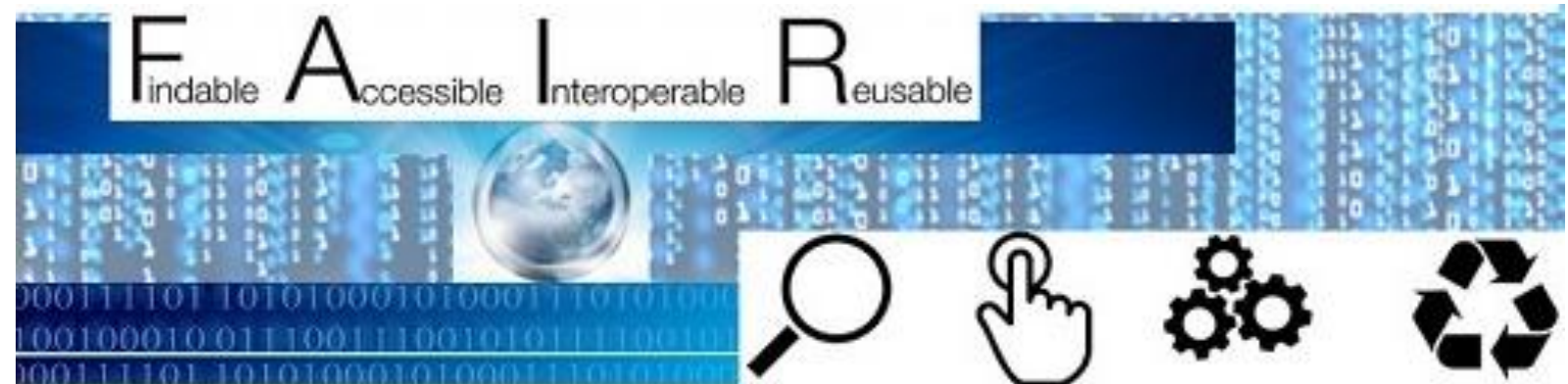
Linked Open Data cloud in 2017 (http://lod-cloud.net)

Legend

- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated
- Incoming Links
- Outgoing Links

NCBO BioPortal
data as of 2013

As any data, ontologies, vocabularies, thesaurus, terminologies…. need to be FAIR

Before the task group:
Review of ontology metadata practices: Methods

1. Analysis of the existing metadata vocabularies for describing ontologies & literature survey
   - More than 23 vocabularies, around 450 properties reviewed

2. Analysis of the uses of metadata vocabularies in describing the ontologies (by the ontology developers)
   - 202 ontologies analyzed (then 805 more recently)

3. Analysis of the uses of metadata vocabularies in various ontology libraries & repositories
   - 12 libraries

Dutta, B., … Jonquet, C.: **New Generation Metadata vocabulary for Ontolog yDescription and Publication**. *11th Metadata and Semantics Research Conference, MTSR'17.* , Tallinn, Estonia (2017).
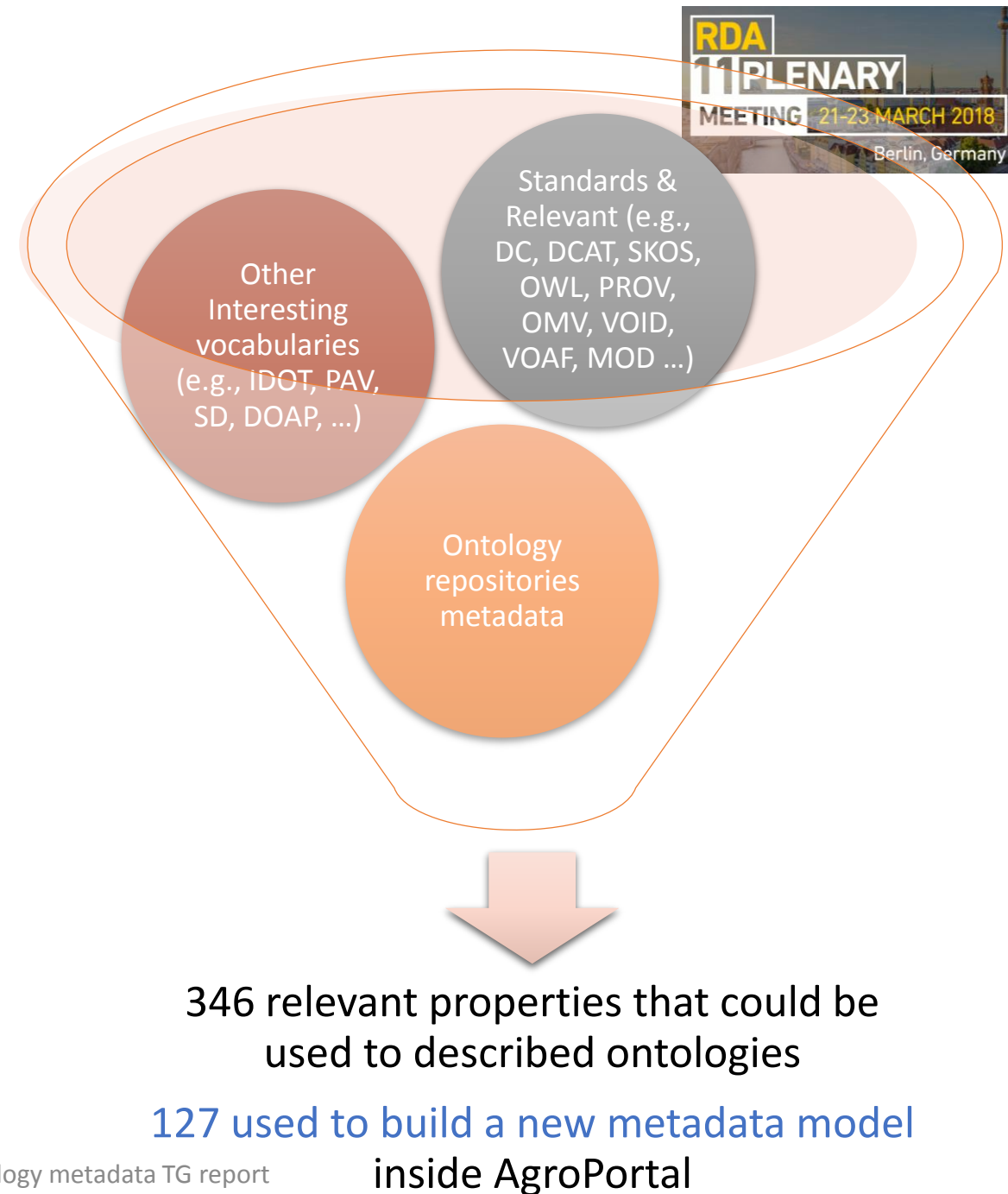
Review of ontology metadata practices: Findings

- Developers use a variety of metadata vocabularies (e.g., DC, DCT, PROV, VOiD, DCAT, Schema.org)
    - Interestingly: the only ontology specific metadata OMV (first published in 2005) is found to be hardly used by the community
    - No existing vocabularies really covers enough aspects of ontologies to be used solely

- 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)
    - For instance 25 properties available for dates

- Reviewed libraries uses, to some extent, some metadata elements but do not always use standard metadata vocabularies

- 16% of ontologies did not use any metadata properties, 43% use less than 10 properties
    - Properties facilitated by ontology editors are more frequent
    - Confusion of use: DC/DC Term or SKOS documentation properties used to describe ontologies
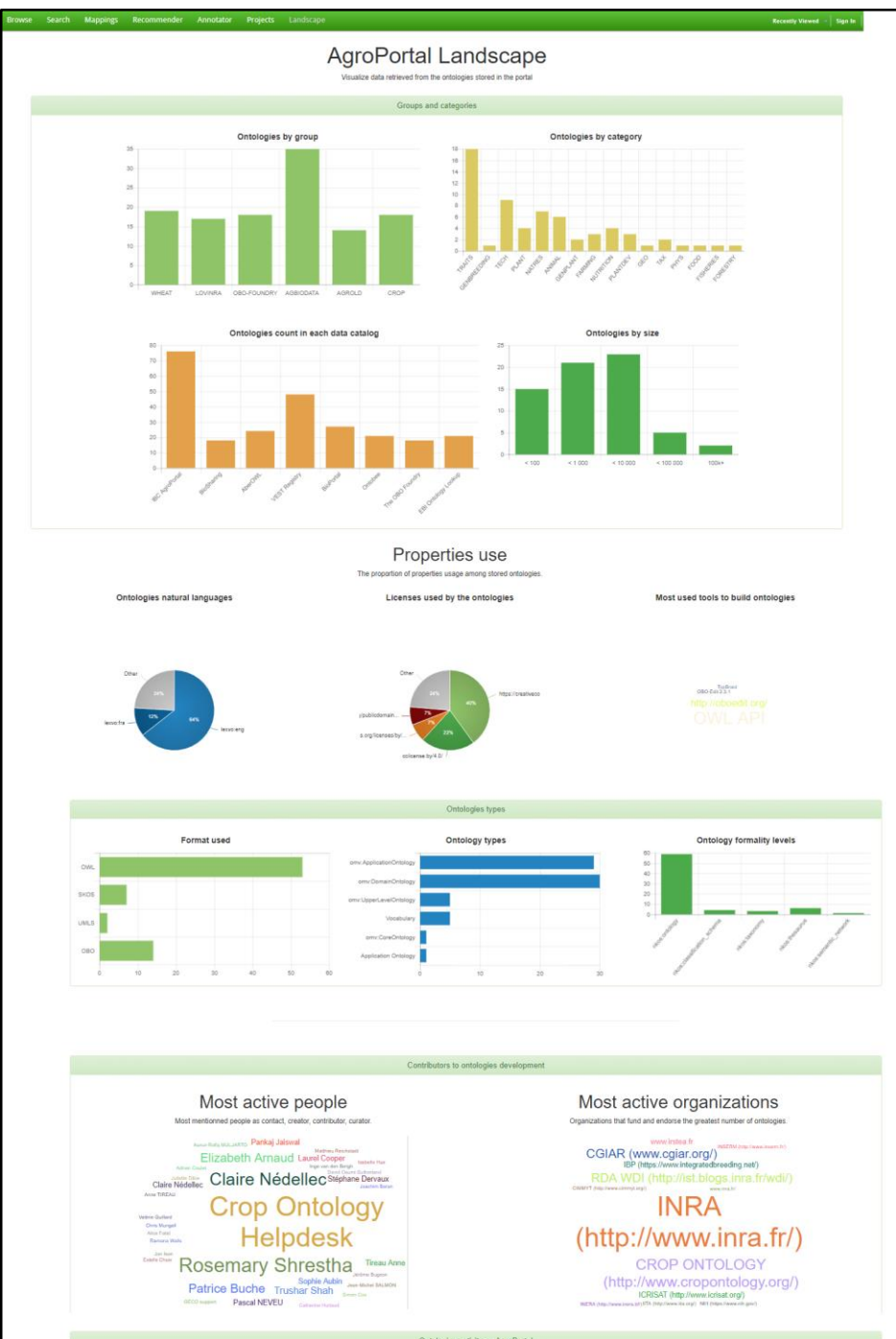
Ontology repositories help to make ontologies FAIR
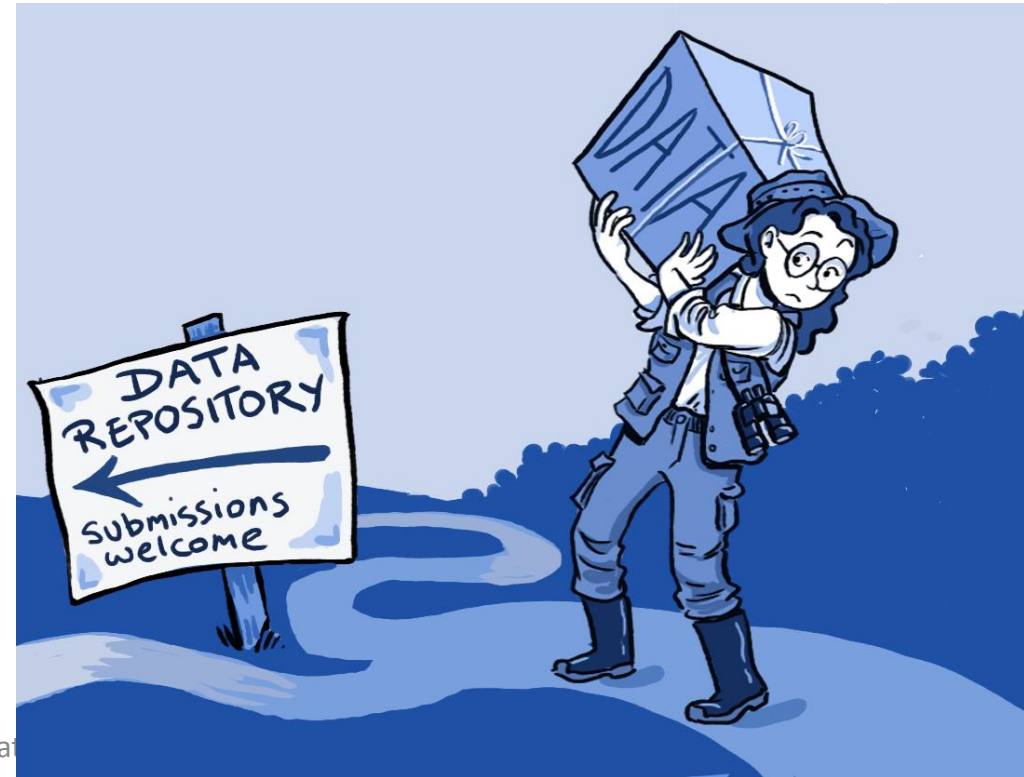
A new metadata model to better support description of ontologies and their relations

- Building a list of properties to describe ontologies

- Pickup properties and relations from 23 existing vocabularies

- Existing properties in ontology repositories (especially BioPortal)
  - Non specific properties that "belong to the ontology"

Other Interesting vocabularies (e.g., IDOT, PAV, SD, DOAP, …)

Standards & Relevant (e.g., DC, DCAT, SKOS, OWL, PROV, OMV, VOID, VOAF, MOD …)

Ontology repositories metadata

346 relevant properties that could be used to described ontologies

127 used to build a new metadata model inside AgroPortal

AgroPortal Landscape page

⚑ Display "per property"
- Global presentation of the properties
- Synthesis diagrams & listing

- Metadata automatically extracted from the files and authored by us and the ontology developers

- Explore the agronomical ontology landscape by automatically aggregating the metadata fields of each ontologies in explicit vizualizations (charts, term cloud and graphs).
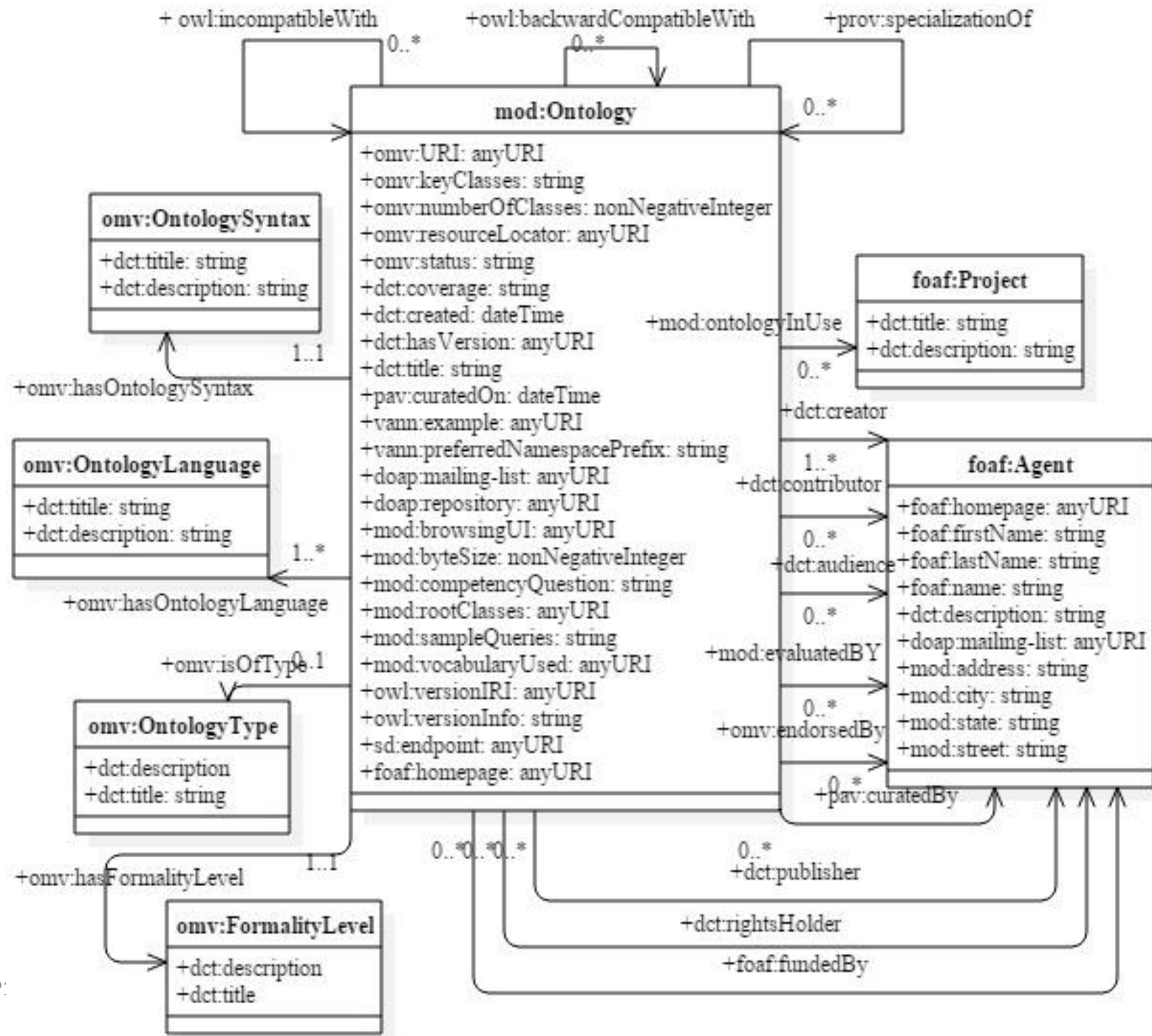
Jonquet, C., Toulet, A., Dutta, B., Emonet, V.: **Harnessing the power of unified metadata in an ontology repository: the case of AgroPortal**. *Data Semant.* UNDER REVIEW.

Generalizing this with MOD

- Metadata vocabulary for Ontology Description and publication (v.1.2)

- 88 properties, only 13 new ones

- https://github.com/sifrproject/MOD-Ontology

Dutta, B., ... Jonquet, C.: **New Generation Metadata vocabulary for Ontolog yDescription and Publication**. *11th Metadata and Semantics Research Conference, MTSR'17.* , Tallinn, Estonia (2017).

RDA P

Beginning of the task group:
Survey of ontology metadata practices

Based on the total 142 responses received until 19th March 2018

https://goo.gl/tXaaMf
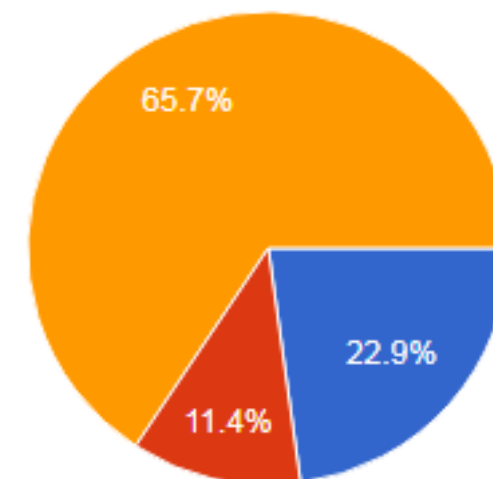
- Ontology developer
- Ontology user
- Both the above

How do you author ontology metadata?

- *I use annotation properties (e.g. dc:creator, foaf:homepage, owl:versionInfo) to describe my ontology (the owl:Ontology or equivalent object* **(61%)**

- *I formally use metadata vocabularies by importing them within my ontology* **(27%)**

- *I rely on the ontology editor and do not go beyond what the user interface allows/suggests me to do* **(21%)**

- *I do not use any metadata to describe the ontology* **(12%)**

Top 5 things you would like to know when searching and selecting an ontology
(*Selected*)

- How complex the ontology (with lots of relations)?
- Update frequency
- Credibility
- Uses and user base
- Subject coverage and comprehensiveness
- Community support
- Expressivity level

- Actively maintained?
- Natural language description
- Depth
- Code source location and issue tracker
- Any standard/ nomenclature applied
- How the ontology evolved (research project, industrial application need)?
- Underline use case, scope, publications

If you have to describe a property, especially the entities like person, organization, location, etc. what do you generally do?

- *I try to always (re)use an existing URI when authoring my metadata*   **(47%)**
  - `dc:creator` https://orcid.org/0000-0003-3059-8202
  - `foaf:organization` https://www.w3.org

- *I do prefer the string values* **(21%)**
  - `dc:creator "John Smith"`

- *I do use URIs but I define them in my own namespace* **(18%)**
  - `dc:creator mynamespace:jsmith`

Do you know or use the following metadata standards?

unknown (u)	NKOS (104), IDOT (102) DOOR (100), VANN (95), ADMS (91), MOD (91), OMV (81), OboInOwl(80), DCT (48),
known but never used (k)	CC (45), SD (42), FOAF(38), OMV (33), VOID (29), SKOS (27), SPARQL (25), OMV (33), MOD (24)
sometimes used (s)	SPARQL (36), MOD (7), OMV (4),
often used (o)	DC(42), DCT(25), DCAT (16), OMV (6), MOD(2)
always used/mandatory (m)	OWL (59), RDFS (54), SPARQL (41), FOAF (16), DCT (15), OMV(2)

Metadata information that you think are missing in the existing metadata vocabularies that you are aware of?
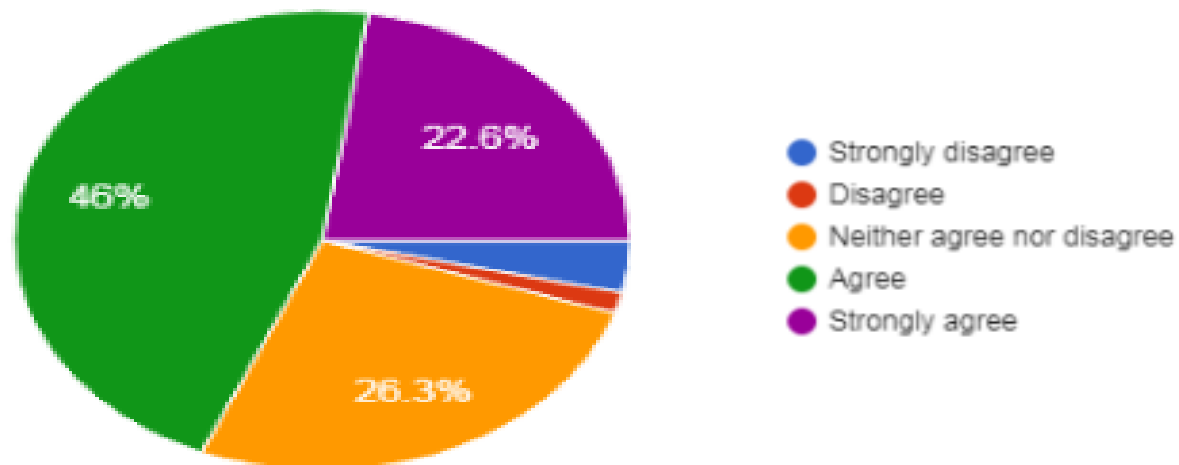
- Complexity, (Human) labelling languages
- Entity similarity measure (for mapping ontologies)
- Data quality metrics
- typical examples
- Privacy constraints
- security information
- none for my current use
- number of synonyms, number of hidden labels, number of hierarchical levels

...

What will make you focus more on ontology metadata authoring?

- *Guidelines and recommendations* **(58%)**

- *Unique community standard* **(49%)**

- *Better user interface within an ontology editor* **(47%)**

- *Simple template to copy/paste and quickly edit in the ontology file* **(29%)**

- *Incentives in terms of ontology citation, reuse, etc.* **(27%)**

- *Mandatory to publish metadata to a library or repository* **(25%)**

Would you find it useful to be supported by a tool to author ontology metadata?

Legend:
- Strongly disagree
- Disagree
- Neither agree nor disagree
- Agree
- Strongly agree

Pie chart values: 22.6%, 46%, 26.3%

Starting as a group….

- A new task group starting… workplan to decide…
  - On going work on AgroPortal metadata model
  - Ongoing work on MOD… we shall do that collaboratively within the group… maybe a DCAT profile ?

- Participate into the survey: https://goo.gl/tXaaMf (until end of March)
  - Use the consolidated outputs of the survey as a starting point

- Online meetings to be organized (starting in April)

- Join us to discuss these subjects on the Slack channel #tg-ontology-metadata