Semantics WG

How FAIR are our vocabularies?

B. Cecconi

IVOA May 2024, Sydney

What is a FAIR vocabulary

Going back to the FAIR principles in Wilkinson et al 2016

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary

How to apply this to vocabularies?

• A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

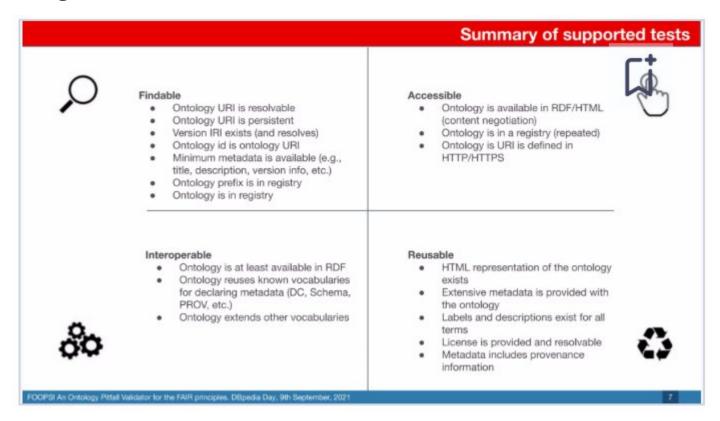
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

What is FAIR Vocabulary

 FOOPS! team (https://foops.linkeddata.es/about.html) proposes the following FAIR criteria for vocabularies



Why should our vocabulary be FAIR?

- Many data infrastructures are currently built with semantic web foundations.
- Examples :
 - Most data repository landing pages include JSON-LD data
 - Search interfaces like OpenCitation or Scholexplorer
- If we want the IVOA entities (services, catalogues, terms) visible and used in those frameworks, we need to slightly improve our practices.
- The good news is: it's easy and we're almost there.

OntoPortal protoype

- ontoportal.org (Stanford) is developing a portal framework for ontology management.
- Many communities already in : Biology, Earth Science, Agriculture, Biodiverity, Environmental sciences, Material sciences...
- Astronomy prototype: http://voparis-ontoportal-dev.obspm.fr
 - All IVOA vocabularies imported, not successfully for some
 - Scope: celestial astronomy, planetary sciences, heliophysics, particle physics, atomic and molecular data
 - Goal: ease matching, reuse and curation of vocabularies/ontologies

Testing our vocabularies

- Test interface : <u>https://foops.linkeddata.es/FAIR_validator.html</u>
- Example results

http://ivoa.net/rdf/uat https://voparis-ns.obspm.fr/rdf/epn/2.0/target-class 57%

- In the 2nd example, there are more metadata, including version, prefered namespace prefix...
- What is still missing: using base URI including version, citation information, attribution on each term, persistent identifier (we could argue that ivoa.net should be persistent...), etc

Evolution of « Vocabularies in the VO »

- Possible pathways:
 - More formal SKOS vocabularies (IVOA-UAT hes been fixed)
 - More formal OWL ontologies (e.g., for property vocabularies)
 - Add extra metadata for version management
 - Add extra metadata for attribution (in vocabulary, in term)
 - Add extra metadata for preferred namespace
 - Register our vocabularies in Vocabulary registries
 - e.g., https://lov.linkeddata.es/ (Linked Open Vocabularies)
 - Set up a persistent URL (but what does this really mean...)