

Beyond Simple FAIR Principles for Ontologies and Semantic Resources

Gary Berg-Cross, ESIP Semantic Harmonization co-lead
(gbergcross@gmail.com)

1. There are important limits of FAIR Principles applied to Ontologies
 - For example, what are limits to Mature, Objective Metrics?
2. What it Looks Like Simply applying FAIR to Ontologies
 - We see the Centrality of Metadata Semantics
3. How big a problem is it that some ontologies fail simple FAIR tests?
4. How important is the Community Role in Quality Ontologies?
 - OBO Foundry & COVID Examples
5. Metadata Vocabularies for Ontologies – 3 examples
6. What is Interoperability in Semantic Spectrum of Resources? Harmonization?

Partially based on *FAIR ontology harmonization and TRUST data interoperability (FOHTI)* workshop at ICBO 2022, (Lin, Berg-Cross & Harris)

<https://icbo-conference.github.io/icbo2022/workshops-and-tutorials/#fair-ontology-harmonization-and-trust-data-interoperability-fohti>

Limits of FAIR Principles - Objective Metrics or Maturity?

- Self-assessment instruments and validators generate a global FAIRness “metric” score.

But surveying “Metrics” involved a degree of subjective interpretation.

- More objective measures called “Maturity Indicators” (MIs) describe components in operational ways that can be objectively evaluated by semantic technologies tools (RDA, 2020).
- MI for Findability checks that a repository implements accessible, structured metadata to enhance its discoverability on the Web.

Essentially an existential check for a data element GUID and that a link is resolved.

- The presence of any Linked Data that can be found for a data element is considered an objective success (FAIR Maturity Indicator, 2019).
- A similar idea is adopted for MD MI - an existential check that metadata follows some “community-defined model.” - but the quality of this standard, especially older standards that are not actively maintained, may vary in important ways.

Finding a meaning? There is no unique definition for “soil” or “soil density” and so there may be many different, but unique GUIDs for a named term in a vocabulary, or a definition in such a vocabulary and/or data using the vocabulary as the label “soil”.

What it Looks Like Simply Applying FAIR Principles to Ontologies

F01 Ontologies (and their elements) are assigned a GUID(s)

F02. Ontologies (& their elements) are described with “rich” metadata (meta-ontologies). Richness can be subjective & thus hard for computer processing but complexity criteria of relations of information in a schema might provide a useful index.

F03. Ontologies (and their element's) metadata clearly and explicitly include the identifier for the ontology/elements they describe.

F04. Ontologies (& their elements) are registered or indexed in a searchable resource.

IO 1. Ontology metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation across different ontologies - conforms to the major, available logic-based knowledge representation languages.

IO 2. Ontology metadata use controlled vocabularies that follow FAIR principles.

IO 3. Ontology metadata includes qualified references to other Ontology metadata.

Centrality of Community, Metadata Semantics & Meaning

FAIR Data for Humans and Machines

Ontology

START

Human-understandable,
ambiguous

concept

A concept

Ontology

Common
Open
Semantics?

SR-Mappings?
Various Tools

Findability, Accessibility
Publish and get persistent
resolvable identifier

Step 1

DOI 10.5281/zenodo.2598799

Citable SR/Ontology
Repositories GUIDs

Interoperability
Transformation to an
open syntax



Step 4

Reusability
Creation and querying
of RDF Linked Data

concepts

Findability, Accessibility
Publish and get persistent
resolvable identifier

DOI 10.5281/zenodo.2640873



Step 5

DOI 10.5281/zenodo.3560778

RESULT

Human & Machine understandable,
unambiguous data

concepts

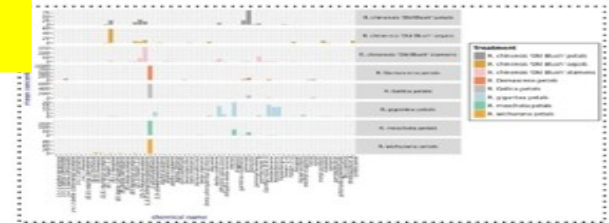
fully opened flowers: 1 g petals in 2 ml hexane containing 5 mg/L camphor
injection of 2 µl into split 25 m DB-5 gas chromatography column for GC-MS (Agilent 6890 with helium)
2 min 40°C, 1.5°C/min until 240°C, 4°C/min until 240°C
ionizing voltage 70eV, mass scan rate 35-450 m/z, mass scan rate 2.94/s
Data analysis made with a threshold of 1.0 to detect major peaks between 4 and 80 min (except for any) acetates after 80 min)
Long chain hydrocarbons are not presented
Compounds determined with Wiley database, NIST online database, purified standard and bibliography
Results in microg/g fresh weight of petals
For Rosa chinensis 'Old Blue', separate stem stamens were analyzed in addition to petals.

	R. chinensis 'Old Blue' sepals		R. chinensis 'Old Blue' stamens	
Compound	Average	Standard error	Average	Standard error
hexan-2-ol	4.95	0.59	0.90	
hexanol	57.62	7.34	9.41	
(E)-2-hexenal	7.04	0.63	5.39	
(E)-2-hexen-1-ol	1.79	0.38		
hexan-1-ol	2.09	0.07		
nonane				
alpha-pinene				
beta-cadinol				

Interoperability
Semantic mapping



ChEBI NCBI
Taxonomy
Plant Ontology



Big picture of standard methods for FAIR and Trust goals for semantic resources/ ontologies

What Happens When We Replace Data with Semantic Resources?

We move to issues of semantics, meaning and richer knowledge representations.

FAIRdata Efforts are Pervasive BUT...

Problems from one Study on the FAIRness of Disaster Ontologies

- Maybe not FAIR stewards of semantic resources meeting the best community interests
 - survey of disaster management ontologies - *FAIR principles are seldom followed*
- Mazimwe et al performed FAIRness tests using the info available as part of a review.
 - Average Findability level was only 1.8%, the average Accessibility level was only 5.8%, and only 4.3% of the retrieved ontologies detail explicit mapping/correspondences between ontologies.
 - 90.9% of URIs to artifacts do not conform to the principle of uniqueness and persistence
 - There is a problem with resolvable via URIs
(*Maybe this is an older publishing issue. Maybe it represents a fragmented community*)

Mazimwe, Allan, Imed Hammouda, and Anthony Gidudu. "Implementation of FAIR Principles for Ontologies in the Disaster Domain: A Systematic Literature Review." ISPRS International Journal of Geo-Information 10.5 (2021): 324.

Standardizing Metadata for Ontologies Ontology Metadata Vocabulary (OMV)

Hartmann et al (2005) (before FAIR & more than FAIR)

- 1) **Accessibility**: Ontology MD must be accessible & processable for machines and humans
- 2) **Usability**: the metadata should be understandable so that a majority of domain users should be able to readily apply metadata.
- 3) **Reuse**: Because ontologies are a (or the) core technology for the Semantic Web, its MD should reflect key issues of the Semantic Web such as reuse and sharing of knowledge.
- 4) **Conceptualization vs. Realization**: MD must reflect a distinction between the underlying semantic conceptualization & a particular realization as a concrete ontology digital artifact.
- 5) **Interoperability**: MD interoperability is key & was defined as conforming to the major representation languages being used for Semantic Web applications in 2005 (OWL etc.).
- 6) **Documentation**: metadata documentation should provide at minimum information about technical, statistical, accessibility, management information, etc.
- 7) **Extensibility**: Reflecting special user needs and the reality that ontologies are build incrementally as knowledge requirements and issues unfold, it is required that beyond such standard metadata, facts can be added and extended easily.
- 8) **Expressiveness**: Metadata must be expressive enough to represent all desired aspects, (as presented above).

OMV had limited vocabulary and so replaced by:
Metadata for Ontology Description and publication (MOD)
(Dutta, Nandini, and Shahi. 2015)

- Make ontologies easily identifiable and reusable for various KE tasks.
- reflected more mature experiences for more consistent & effective documentation to find and share relevant ontologies such as the Bio Portal.
- Balanced requirements perspectives:
 - top-down view of the big picture &
 - bottom-up survey of repository searches - what do users find important?

8 MOD requirement that partially overlapped with OMV's metadata requirements (such as Extensibility & Usability)

1. Brevity: The vocabulary should consist of a minimal set of elements maintaining balance between necessity and sufficiency. (recognizable to humans, BUT lack guidance on machine operability for ontology metadata)
2. Clarity: The MD elements must be well defined & clear descriptions should be provided. (NOT automated)
3. Simplicity: The vocabulary should be easy to use. (NOT automated)
4. Authority: The vocabulary design should be based on a sound methodology in the sense that the inclusion of terms in the vocabulary are justified. (again a human documentation & understandability issue.)
5. Standardization: The element names should be standardized. To confirm the standardization, the individual elements should be mapped with the *existing standard vocabularies*. (But definitions need to be standardized)
6. Extensibility: The vocabulary should be extensible.
7. Usability: The vocabulary should support the reuse of the described resources. That is, the vocabulary should allow the creators/developers to highlight the usage and the quality of the resources in a well-defined manner. (Selection??)
8. Interoperability: The *vocabulary* should be interoperable. That is, it should conform to the major knowledge representation languages... **but what about compatibility to other vocabularies!!!!**

Domain with Ag (like farming) Agro-Portal now recognizes 346 properties from existing metadata vocabularies that could be used to describe different aspects of ontologies (and other semantic resources)

Minimum Information for Reporting an Ontology (MIRO)

initiative associated with Open Biology Ontology (OBO) Foundry

MIRO defines 34 information items, “ontology name”, “ontology license”, “ontology URL” are content-independent metadata don’t depend on the content or it’s associated resource and thus are not really rich metadata, although they will have resolvable GUIDs.

Two important guidelines go deeper than FAIR:

- information about ontology relationships and about axiom patterns.

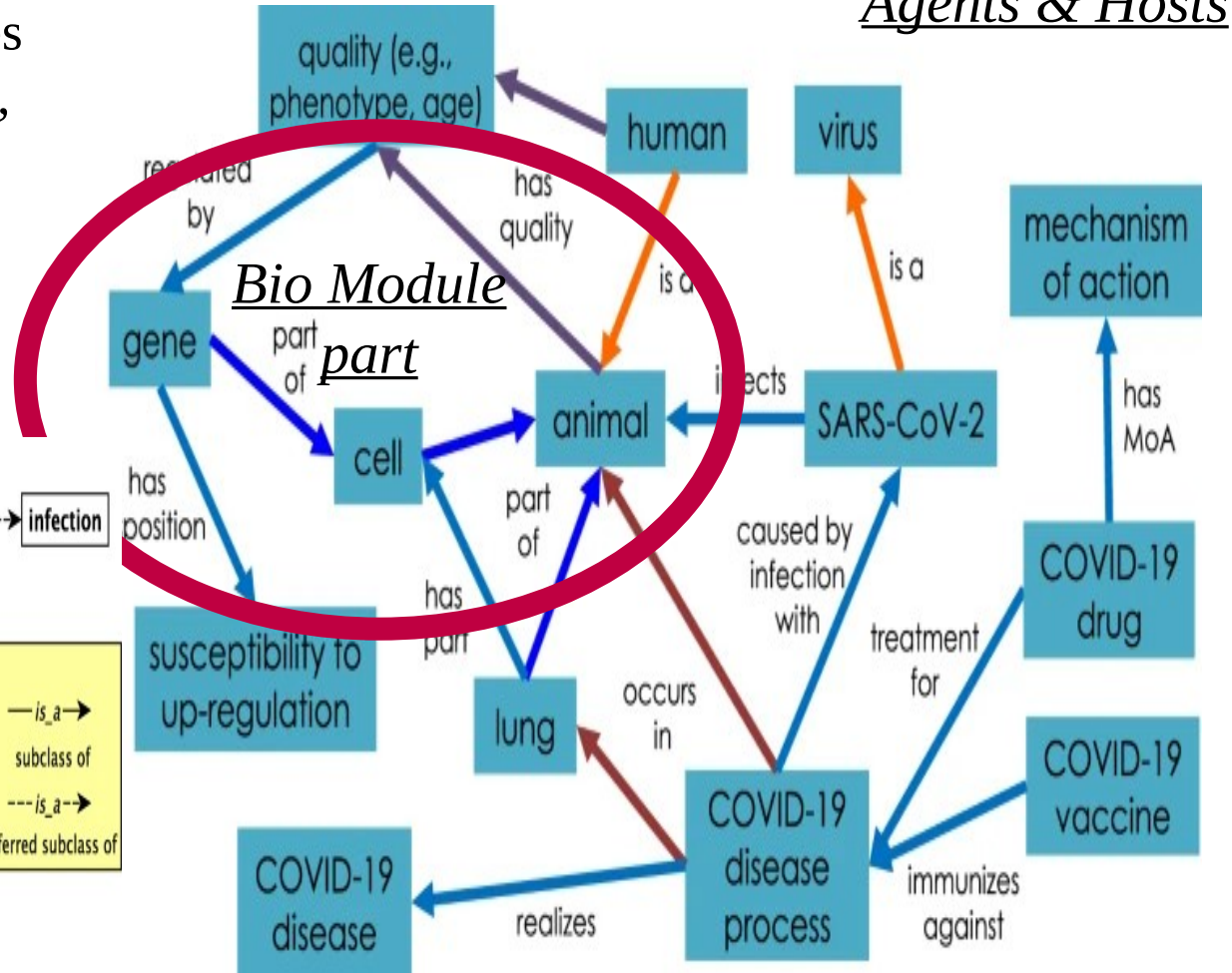
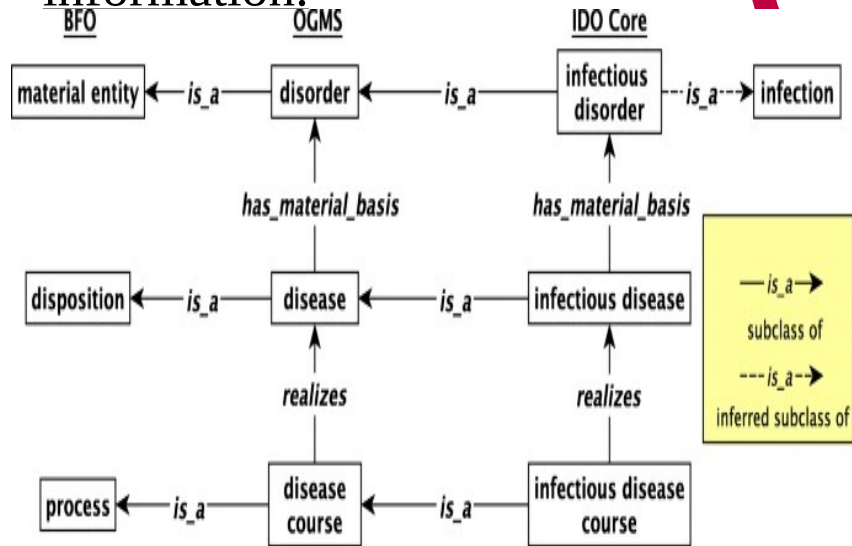
E.9 Ontology relationships (which MUST be documented): The relationships or properties used in the ontology, which were used and why (not automated?).

E.10 Axiom pattern (is also a MUST): An axiom pattern is a regular design of axioms or a template for axioms used to represent a category of entities or common aspects of a variety of types of entities. (How is quality of a pattern assessed?)

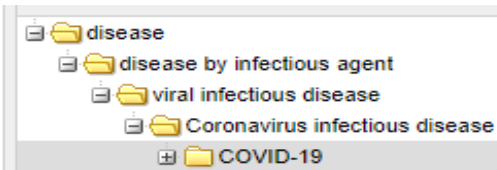
Axiom pattern used to define Soil (super-type: environmental material, sub-types: Leptosols..., relations: *has quality some porous, has part some sand....*) .

Community Role in Quality Ontologies: The OBO Foundry & COVID Examples

- Large accumulation of resources (vocabularies, taxonomies, KGs, linked data) across the disease domain(s) provide a general semantic spectrum context for the Covid-19 pandemic information.



What is Interoperability for SR? Similar or aligned meanings, honest effort on Synonyms & Xrefs? More Complex?



Metadata

ID	DOID:0080600
Name	COVID-19
Definition	A Coronavirus infection that is characterized by fever, cough and shortness of breath and that has_material_basis_in SARS-CoV-2. https://www.cdc.gov/coronavirus/2019-ncov/about/index.html, https://www.ncbi.nlm.nih.gov/pubmed/?term=32007143, https://www.ncbi.nlm.nih.gov/pubmed/?term=32007145, https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=2697049, https://www.who.int/emergencies/diseases/novel-coronavirus-2019
Xrefs	ICD10CM:U07.1 MESH:D000086382 NCI:C171133 SNOMEDCT_US_2021_09_01:840539006 UMLS_CUI:C5203670
Synonyms	2019 Novel Coronavirus (2019-nCoV) [EXACT] 2019-nCoV infection [EXACT] COVID19 [EXACT] SARS-CoV-2 infection [EXACT] Wuhan coronavirus infection [EXACT] Wuhan seafood market pneumonia virus infection [EXACT]
Parent Relationships	is_a Coronavirus infectious disease

[Add an item to the term tracker](#)

<https://disease-ontology.org/?id=DOID:0080600>

Many Linked
Objects

Severe acute respiratory
syndrome coronavirus 2
Taxonomy ID: 2697049
(for references in articles
please use
NCBI:txid2697049)
.....SARS coronavirus

Do we need an (meta) ontology to align/relate ontologies?

- Alignment & matching involves the correspondences (e.g. equivalence ($\equiv$), subsumption ($A < B$), and Disjointness ($A \& B < \perp$) between entities (e.g., classes, objects, properties)
- There are good reviews on alignment
 - Pour, Nikoosaei, et al. "Results of the ontology alignment evaluation initiative 2020." Proceedings of the 15th International Workshop on Ontology Matching (OM 2020). Vol. 2788. CEUR-WS, 2020.
 - And many alignment tools, BUT how good are the semantics of relations?
- There is some work showing relations to connect ODPs:
 - “Extensions to the Ontology Design Pattern Representation Language”
- The Ontology Pattern Language (OPLa) was developed to facilitate annotating ontologies with useful metadata, as well as supporting tooling infrastructure.