



First FAIR Ontology Harmonization and TRUST Data Interoperability Workshop (FOHTI-22) US2TS Panel Discussion

Asiyah Lin
1/18/2023

Panel discussion at US2TS

- September 30, 4:30PM -6PM @ Lansing, Michigan State University
- **Towards FAIR, Trustworthy and Harmonized Semantic Resources**
- Moderators: Asiyah Lin, Gary Berg-Cross
- Panelists: **Pascal Hitzler** (Kansas State University), **Chris Mungall** (Lawrence Berkeley National Laboratory), **Ramona Walls** (Critical Path Institute)
- Discussants: online and in-person (appx. 25)





- A semantic model that includes organism or populations or other entities, their characteristics, data about them all, and relations among them
- Minimization of and interoperability among the various observation ontologies
- Interoperable terminologies: One vocabulary to rule them all? Mappings? (With IRIs, even better with semantics.)
- Open access with sustainable business models for all the pieces
- People and community

Ramona Walls' talk

Ramona's points:

- Need fewer ontologies.
- Mapping interoperable terminologies are complex. I need terms with IRIs and semantics.
- I need open access.
- I need people – can't hire ontologist, successfully trained from internal. Research funding is hard to get for training ontologist.
- Community - it is all about consensus. there should be **one ontology**. This is a big challenge.
- This is more people & community issues not technical issues.

- Fair (Findable, Accessible, Interoperable, and Reusable)
- TRUST (Transparency, Responsibility, User focus, Sustainability and Technology)

Hardly controversial. The question is “how”?

Pascal Hitzler

Let's focus on Interoperable and Reuse:

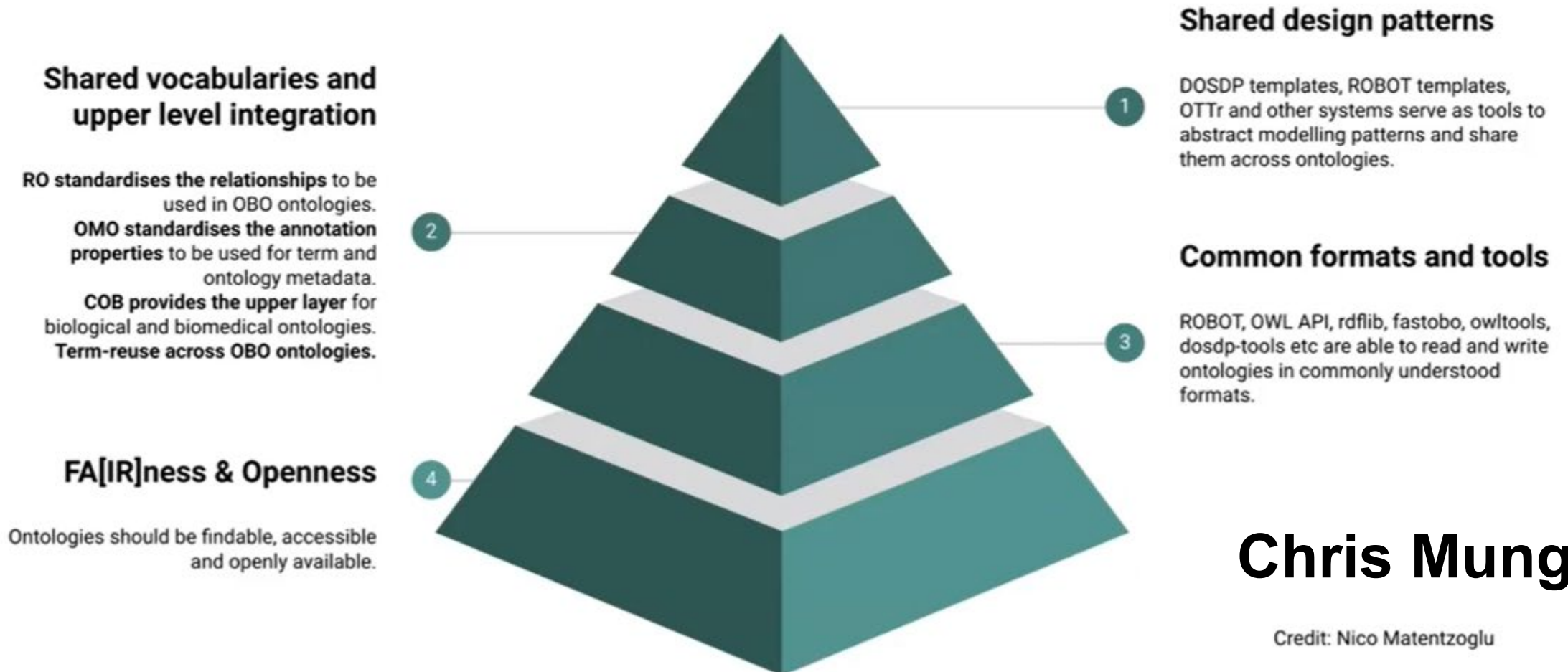
Most ontologies are very hard to re-use and/or solve only very narrow use cases (notable exceptions exist). [also see discussion yesterday!]

In most cases, it's less work to make a new ontology than try to re-use an existing one. **So what to do about this?**

Pascal's points:

- US and EU differences: EU has trainings. US seems concentrating on biomed domain.
- By definition, an ontology should be FAIR and TRUST. The question is how.
- How to change the way we develop ontologies?
- Ontology Design Patterns! sets of interconnected modules - LEGO!
- Reuse modules and patterns as templates

Layers of OBO interoperability



Chris Mungall

Credit: Nico Matentzoglu

How do I trust an ontology?

- Is it active and responsive?
- Is it used by professional data/knowledge curators?
- Do domain experts contribute to its development
- Is it adopted by the community?

Reuse is key!!

Chris Mungall

Chris Mungall's points:

- Ontology use in life science stemming from GO, extremely widely used.
- In GO, definition of a gene is not fully agreed to this day! Hard to reach consensus.
- The need to **bring ontology to data** - LinkML as data exchange standards.

Fantastic Group Discussions

How to bring ontology to domain experts?

What does it mean by using ontology?

CDE, data, and ontology, how does it work?

How to bring ontology to domain experts?

- The best data management is that people don't know it is there. Same as Ontology. System does not need to tell that ontology is used.
- Funding and policy problem. We need an interface. Such as KG front -- industry comes into this corner. Schema diagrams is a good vehicle for communications to domain experts.
- Ontology scares primary researchers. Ontology is not just hierarchy, but also semantics. All scientists in the lab should receive the training on good data generation practice and ontologies.

How to bring ontology to domain experts?

- Need to make ontology visible to end users.
- Domain experts benefits from the intellectual process of modeling a domain ontologically.
- It helps them to go through and think about their field and expand their minds in the way that they haven't thought before.
- **We won the world when ontology is taught in high schools.**

What does it mean by using ontology?

- Hierarchy of ontology is very useful to end users.
- Community uses ontology as standard. Work with community is important.
- Ontology is very helpful to integrate data from diverse sources.
- Different fields use it differently:
 - 1. standardize definition. - not useful for all.
 - 2. Use as schema for KG
 - 3. Type of the use cases determine the types of ontologies and the approaches to develop ontology

Are use cases necessary for ontology?

- Internet does not have use cases, so does ontology.
- Ontology needs to go with data. Ontology - mostly schema for data.
- For data analysis purposes, analysis relevant data integration and curation are intensive.
- The machinery of ontology-KG-RDF is to reduce the cost of the data prep and more effective.

CDEs problem at NIH

- Is NIH creating a monster that each big initiative creates their own CDEs and later needs huge effort to harmonize?
- RADx example:
 - Data centers were quickly established to respond to COVID.
 - Datahub was funded later had to harmonize 500 CDEs/dictionaries.
 - Process: normalize their representations, NLM tools helped to find matchable concepts, and then human identify the rest.
- Repeated processes everywhere, who will turn them into tools?

CDEs and ontologies

- CDE similar with controlled vocab. We should advocate : a little tech added to Controlled Vocab will be a more useful ontology later.
- OBO shows a prospective way rather than retrospectively mapping. It pays massive.
- Ontology constrain meanings and vocabularies.
 - Harmonize is to use ontology axioms, formal logics only constrain meanings but not capture definitions.
- Ontology is OWL file + documentation.
 - Ontology alignment is not successful because of little documentations.

Other thoughts

- Ontology and CDM are similar things.
- **Micro or little semantic model** to help using ontology to translate the data schema that does not have semantics.
- We need to identify the bottom level semantics for the data elements, so we can have a sweet spot.

Ontology fabric building use good principles

- No one model is absolutely correct, but there are number of views can be used.
- We should encourage building ontologies following **good principles** because no one absolutely correct ontology exists.
- It will build amount of fabric ontologies that we will need to develop a master ontology that is closer to the truth.
- Democratization of data and model.

Comparison of industry data capabilities

Capability	Pharma	US Federal Government	Academia
Internal Data Generation	Dark Green	Dark Green	Dark Green
Internal Data Storage Capacity	Light Green	Light Red	Dark Green
Internal Computational Capacity	Light Red	Light Red	Dark Green
Cloud Capability	Dark Green	Light Green	Light Green
Data Transfer	Light Red	Light Green	Dark Green
Data Management	Light Red	Red	Light Red
Data Standards	Light Red	Light Red	Light Red
Data Analytics	Blue	Light Green	Dark Green
AI/ML	Dark Green	Light Green	Blue
Data Sharing	Red	Light Green	Light Green
Overall Data Strategy	Light Red	Light Red	Light Green

This is a heat map

- Innovating
- Sustaining
- Lagging/Average
- Behind/Restricting

Data management issues

- Data management is primarily missing in biomed research.
- FAIR is intending to solve that problem. It is complex.
- Analog with internet that pointing the address to computers, the "data oriented network architecture", where each piece of data contains address and pointing to each other.
- The ontology works as the conceptual mapping of the data and pointing to where the data belongs within the domain.

Risk of NIH policy

- Risk of the NIH data management and sharing policy.
- Researchers are not experienced at all.
- NIH should not make the same mistakes as in clinical: quickly mandate adoption of standards when people are not properly trained.
- NIH Should focus on interoperability, and should not dictate exact data and models to use, but should provide principles and guidance.

Findable

Metadata and data should be findable for both humans and computers

F

A

I

R

Interoperable

Data needs to work with applications or workflows for analysis, storage and processing

Accessible

Once found, users need to know how the data can be accessed

Reusable

The goal of **FAIR** is to optimise data reuse via comprehensive well-described metadata



TRUST principle



Transparency To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.

Responsibility To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.

User Focus To ensure that the data management norms and expectations of target user communities are met.

Sustainability To sustain services and preserve data holdings for the long-term.

Technology To provide infrastructure and capabilities to support secure, persistent, and reliable services.

Thank you!

Please check out the 1.5 full recording of the panel discussion on the

YouTube: <https://youtube.com/watch?v=-Hc95RRovXg&feature=shares>