

Shades of conceptual meaning (snow cover vs. snowpack "mg/l" synonym of "milligrams per liter")

Tack B:
Harmonizing Definitions and Improving Ontologies

Gary Berg-Cross Ontolog Board Member

ESIP Contributors: Ruth Duerr, **Pier Luigi Buttigieg**, Mark Schildhauer, **Kai Blumberg**

Session Outline

- Brief recap of Launch (Gary)
- Introduction to harmonizing “definitions” (Gary)
- The Environmental ontology (EnvO) – scope, topics, philosophy….
- Examples from the Cryosphere/EnvO harmonization efforts (Pier)
- Tools (Kai)
- Issues and next steps (Pier)
- Planned and unplanned discussion

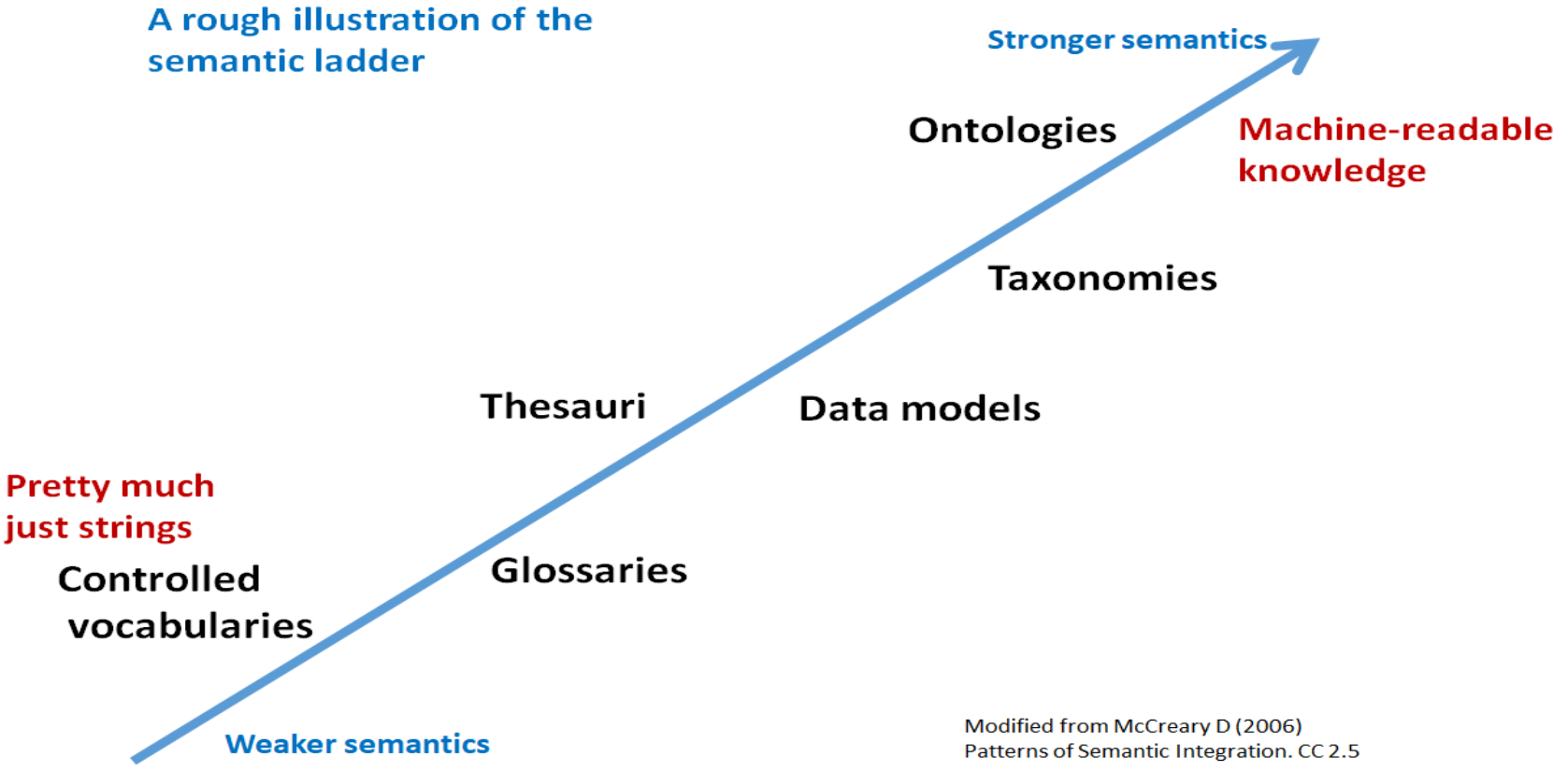
Our Speakers

Dr. Pier Luigi Buttigieg is Digital Knowledge Steward & Senior Data Scientist Researching and implementing a new digital ecosystem for Earth and Environment, as part of the inter-institute Helmholtz Metadata Collaboration Platform.

- He worked a number of years as a Data Scientist at Alfred Wegener Institute Helmholtz centre for polar and marine research,where he was also a post-doctoral fellow
- PhD (Bioinformatics) - Microbial Genomics Group, Max Planck Institute for Marine Microbiology & an MS (Marine Microbiology) - Max Planck Institute for Marine Microbiology
- His Main (of many) Research Interests: Microbial ecology, Bioinformatics, Ecoinformatics, Multivariate analysis and Ontology development (EnvO)

Kai Blumberg works for the Biological and Chemical Oceanography Data Management Office (BCO-DMO) and is a PhD student in the University of Arizona Biosystems Engineering department where he integrates and provides analytical tools to analyze key marine omics and biogeochemical datasets. He is an important contributor to the Environment Ontology (ENVO) and has been an ESIP fellow.

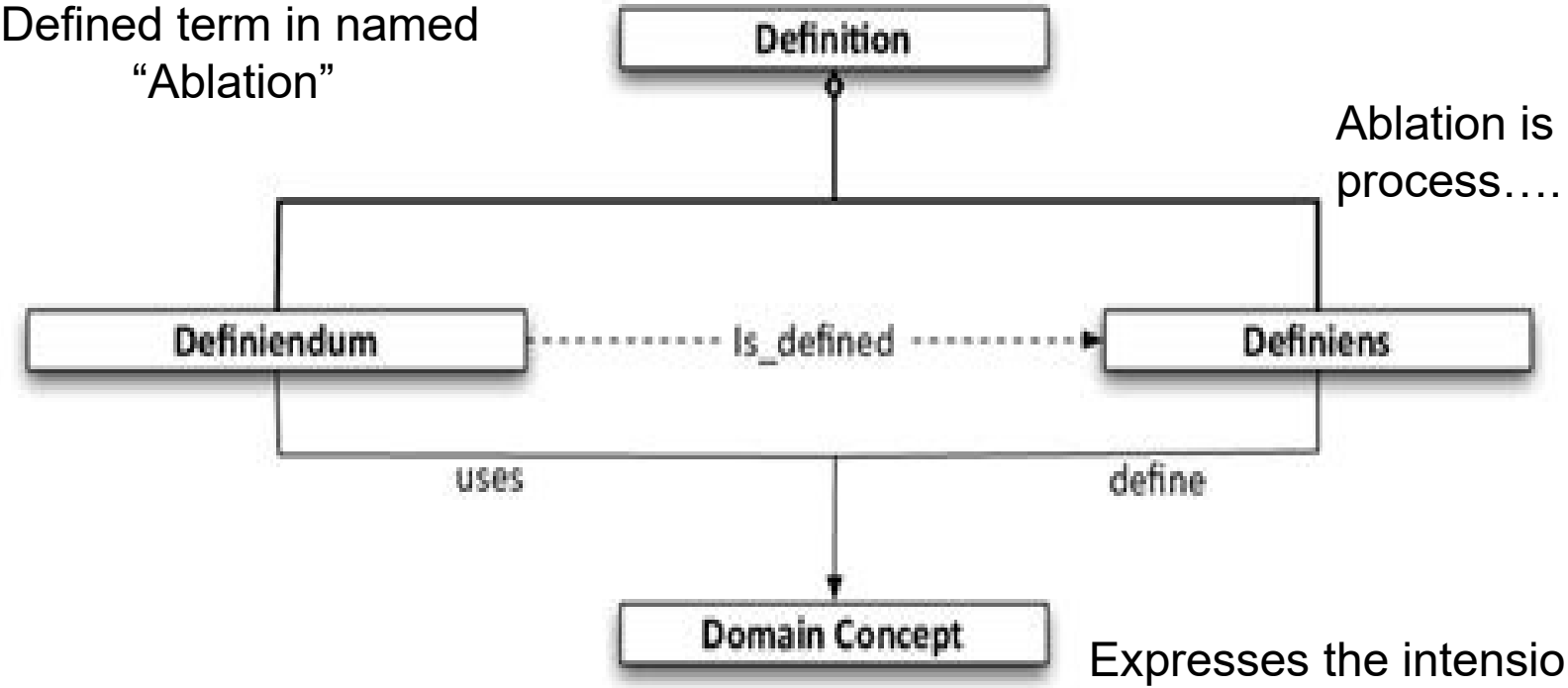
Harmonizations from all levels:
Cryosphere harmonizations starts at the bottom



Modified from McCreary D (2006)
Patterns of Semantic Integration. CC 2.5

Classic Definition Model

Defined term in named "Ablation"



Ablation is a process….

Make distinctions & conveying, converging reduction of the ambiguity for an intended meaning (in context & for communication) including by axioms in a logical language for apt application such as inferences & recognition.

Expresses the intension/ extension definition *content* that we call a 'definition' in semantic resources like ontologies and dictionaries

Some Mixed Requirements and Guiding Principles

Eliminate (conceptual) ambiguity - principles for defining termed concepts
Show relations and structure to help understanding

Control synonyms - term equivalence & similar meanings beyond simple formulation

Establish relations among termed concepts where appropriate e.g. Sub-type & part-whole

Consider and systematize the role of lexical modifiers (e.g. alpine or liquid)

For a final vocabulary try some ISO terminological rules (ISO General Vocabulary) of:
Be brief, use information sufficient to indicate the position in a concept system;
Using the nearest generically superordinate concept, adding one or a few delimiting characteristics and putting other information in notes.....

Some Lessons (still being) learned the hard way:

Semantic Harmonization:

- ***Vocabularies vary significantly in quality and scope***
  - Hard to access: in PDF, deep in website, etc.
  - “Naked terms”-- just a label, no additional description
  - Errors-- in spelling, organization/arrangement
  - Missing terms, and bits & pieces distributed
  - Much redundancy of effort



- ***<u>Post hoc</u> vocabulary alignment IS challenging***
  - Arduous review **necessary--** term by term
  - Granularity of terms **frequently** differ (broader, narrower, ambiguous)
  - Text definitions/descriptions **usually** differ (related? how reconcile?)
  - Formal axioms **always** differ (implied, RDF/SKOS/OWL, other)

Quality Definitions – Applies to Ontologies

- Definitions vary according to context of use and target audience.
- They are to be made relevant for each context to fulfill their cognitive and linguistic goals.
- This involves *adapting* text to some formal/*logical structure*, *type of content*, and *form* for each context of use.

- Adjusting to a receiving agent's semantic inferential competence (knowing)

- Definitions (and supporting axioms) have to be *competent* so an Ontology can answer Competency Questions.

More in *DEFINITIONS IN ONTOLOGIES*
Selja Seppälä, Alan Ruttenberg, Yonatan Schreiber, Barry Smith and "Guidelines for writing definitions in ontologies"

“Snow Cover” Example of Term Harmonization

1. The areal extent of snow-covered ground, usually expressed as percent of total area in a given region.

2. In general, a layer of snow on the ground surface. Compare snowfield, snowpack.

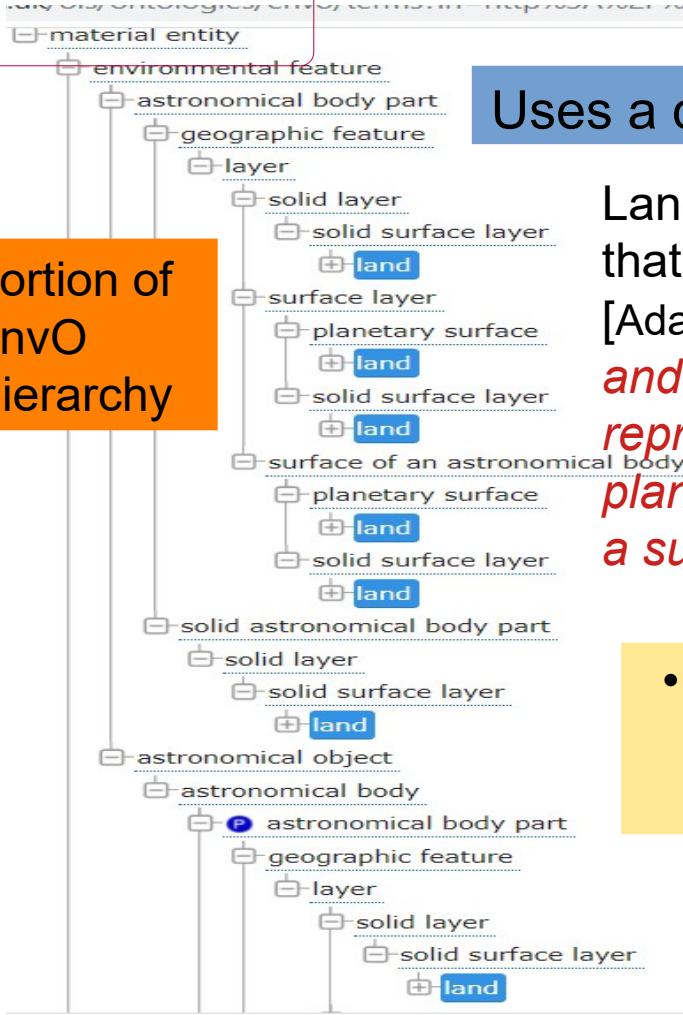
3. The depth of snow on the ground, usually expressed in inches or centimeters.

GCW:
An area density which inheres in snow distributed over an area of a landmass or other substrate.

Portion of ESIP-Cryo hackathon-robot-template

label	parent class	GCW definition	Definition's database cross reference	comment
freezing rain	rain	Rain that freezes on impact to form a coating of clear ice (glaze) on the earth's surface, exposed objects including aircraft in flight, or after percolating below the surface of a glacier or snow field.	https://orcid.org/0000-0003-4808-4736	A surface temperature inversion is usually required. In such an inversion, the surface mu a temperature below freezing, while the temperature of the atmosphere where the precip forms is above freezing. Alternately, if the temperature of the ground surface and glazed is typically near or below freezing (0C or 32F), it is necessary that the water drops be supercooled before striking. Freezing rain can sometimes occur on surfaces exposed to (such as tree limbs) with air temperatures slightly above freezing in strong winds. Local evaporational cooling may result in freezing. Freezing rain frequently occurs, therefore, a transient condition between the occurrence of rain and ice pellets (sleet). When encounte an aircraft in flight, freezing rain can cause a dangerous accretion of clear icing. In aviati weather observations, this hydrometeor is encoded ZR.
hail	ice mass	Precipitation in the form of balls or irregular lumps of ice with a diameter ranging from 5 to 50 millimeters (0.2 to 2.0 inches) or more, always produced by convective clouds, nearly always cumulonimbus.	https://orcid.org/0000-0003-4808-4736	Individual units of hail are called hailstones. Thunderstorms that are characterized by str updrafts, large liquid water contents, large cloud-drop sizes, and great vertical height are favorable to hail formation. The destructive effects of hailstorms upon plant and animal li buildings and property, and aircraft in flight render them a prime object of weather modifi studies. In aviation weather observations, hail is encoded A. Smaller particles of similar c formerly called small hail, may be classed as either ice pellets or snow pellets.
		Any storm that produces hailstones that fall to the ground; usually used when the amount or size of the hail is considered significant.	https://glossary.ametsoc.org/wiki/Hailstorm	
		This generally means... snowfall accumulating to 4" or more in depth in 12 hours or less; or snowfall accumulating to 6" or more in depth in 24 hours or less		

What do Concepts Look Like in an Ontologies?



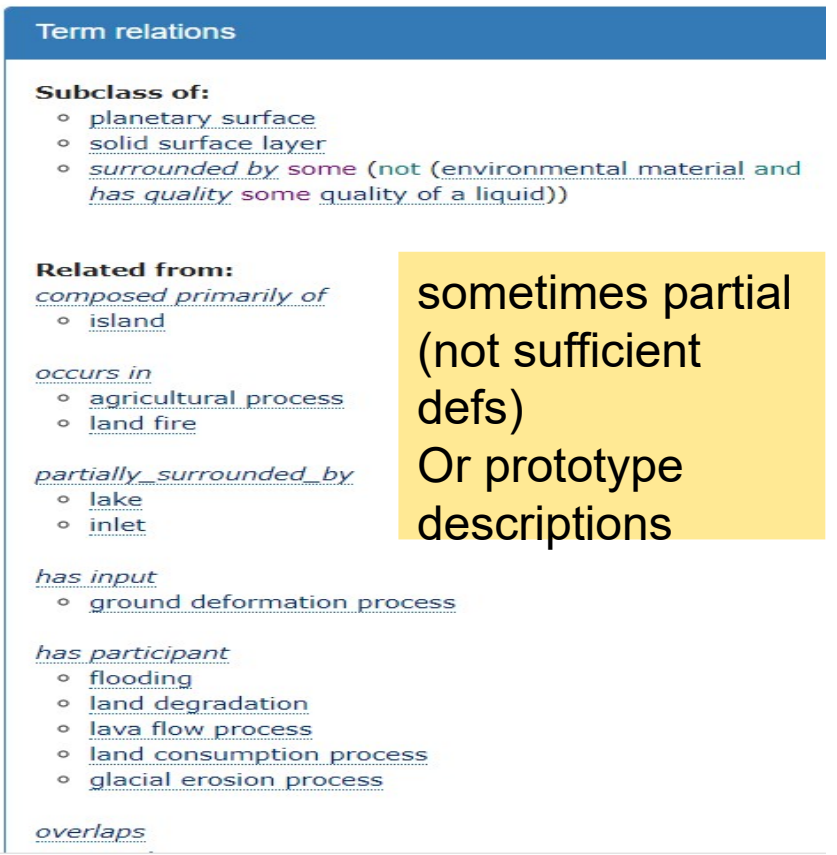
Portion of EnvO Hierarchy

Uses a consensus controlled Vocabulary

Land is a planetary surface that is not covered by liquid. [Adapted from /wiki/]

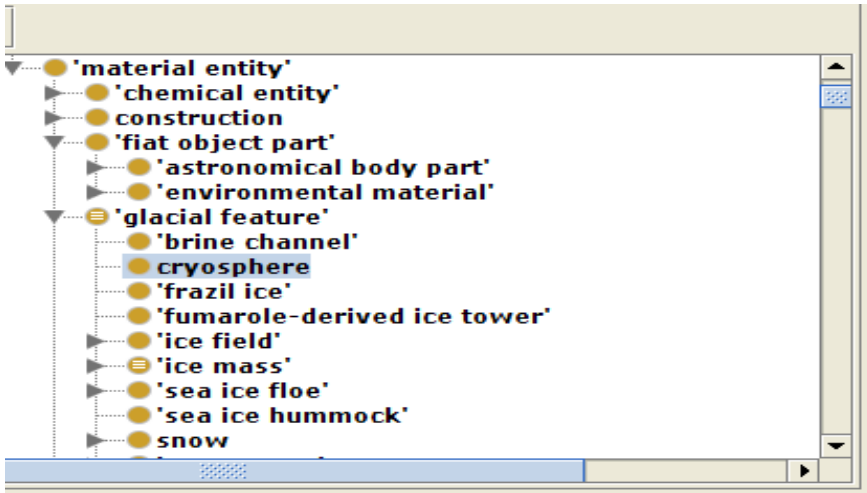
and may more properly be represented as part of a planetary surface rather than a sublclass.

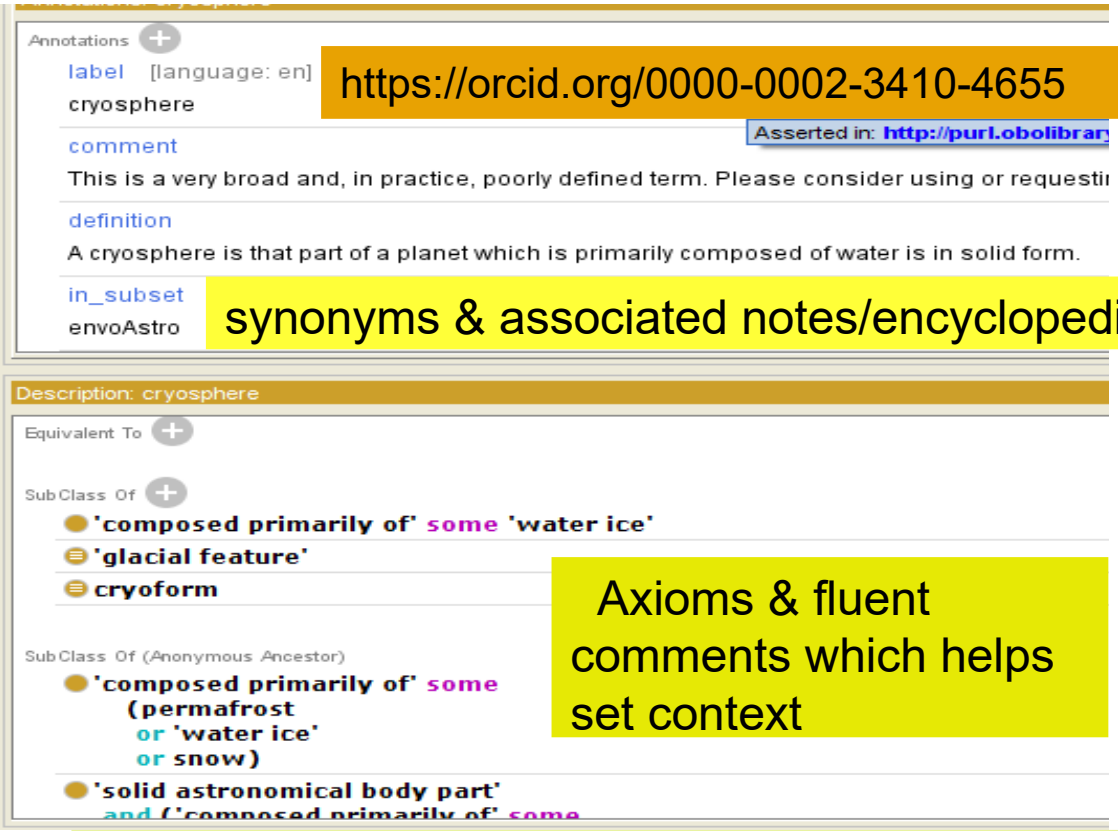
- necessary and, whenever possible, sufficient conditions



sometimes partial (not sufficient defs)
Or prototype descriptions

Protege-Ontological Elements – termed classes & relations+





we see EnvO's axioms & related definitions with linguistic type information for humans, complementing each other

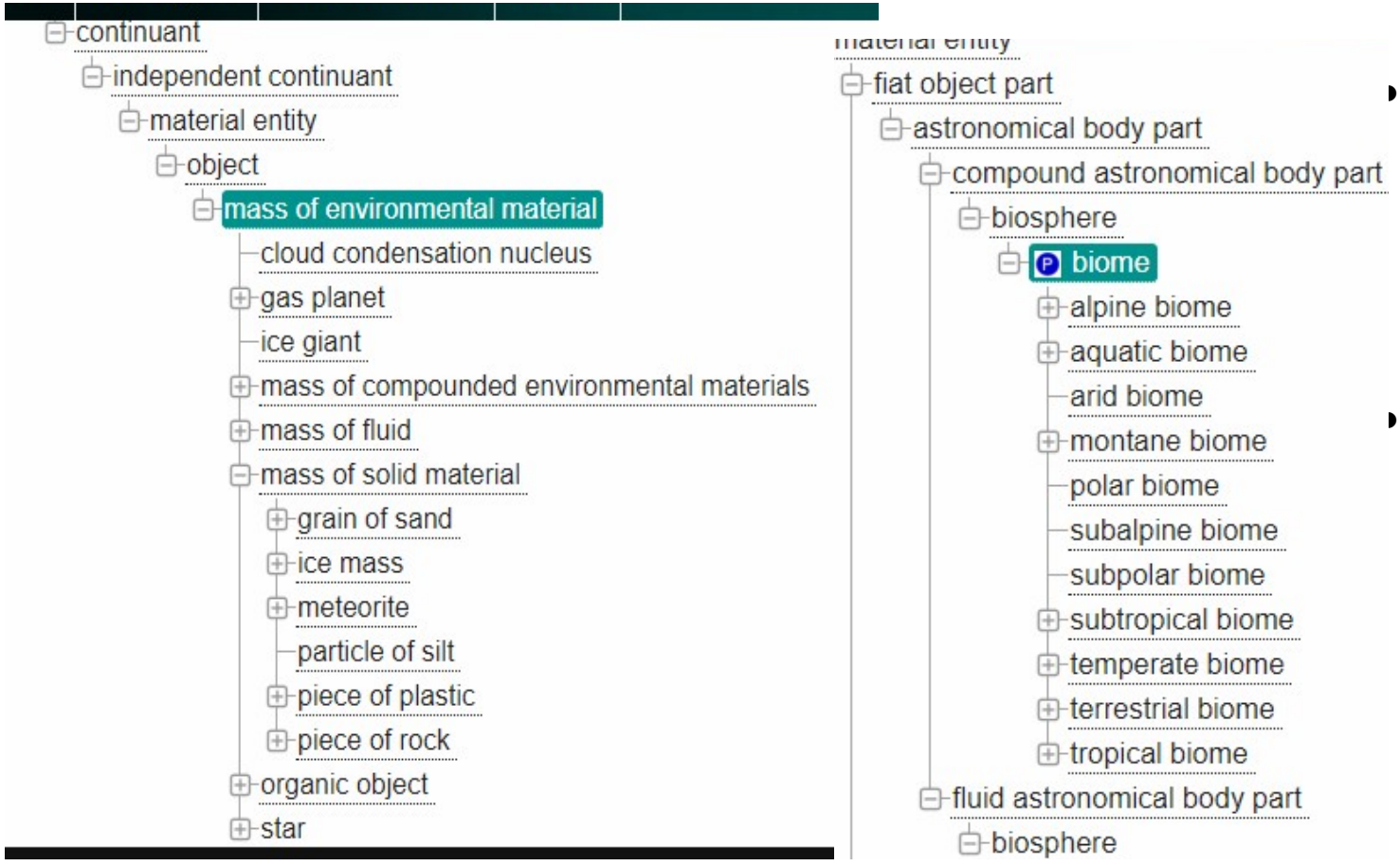
See "Minimum Information for the Reporting of an Ontology (MIRO) guidelines."

- The Environment Ontology (ENVO; http://www.environmentontology.org/) (2013-now)

A semantic resource and research target for the semantically controlled description of environmental entities. (Widely cited - Darwin Core glossary uses ENVO in habitat descriptions)

- Initial aim a representation of the *biomes*, *environmental features*, and environmental *materials* pertinent to genomic and microbiome-related investigations.
  - However, the need for environmental semantics is common to a multitude of fields, and ENVO's use has steadily grown since its initial description.
- Scope has expanded, enhanced, and generalized the ontology to support its increasingly diverse applications – like Cryo & Marine.
- (Source- Buttigieg, Pier Luigi, et al. "The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation." Journal of biomedical semantics 7.1 (2016): 1-12.)

EnvO and Obo Foundary



- “habitat” http://purl.obolibrary.org/obo/ENVO_01000739
- “An environmental system which can sustain and allow the growth of an ecological population.”

• Evolving Methods...now includes harmonization

- Updating to development suite to promote
  - expressivity, consistency, and speed: OWL, templating methods to accelerate class creation.
- Aligning ENVO with SWEET & the Open Biological and Biomedical Ontologies (OBO) Foundryusing principles and interoperate with existing OBO ontologies.
  - So leverages concepts of spatial environment (niche), boundaries, occupants...
- Applying text-mining approaches to extract habitat information from the Encyclopedia of Life and automatically create experimental habitat classes within ENVO.

Example of expanding Scope

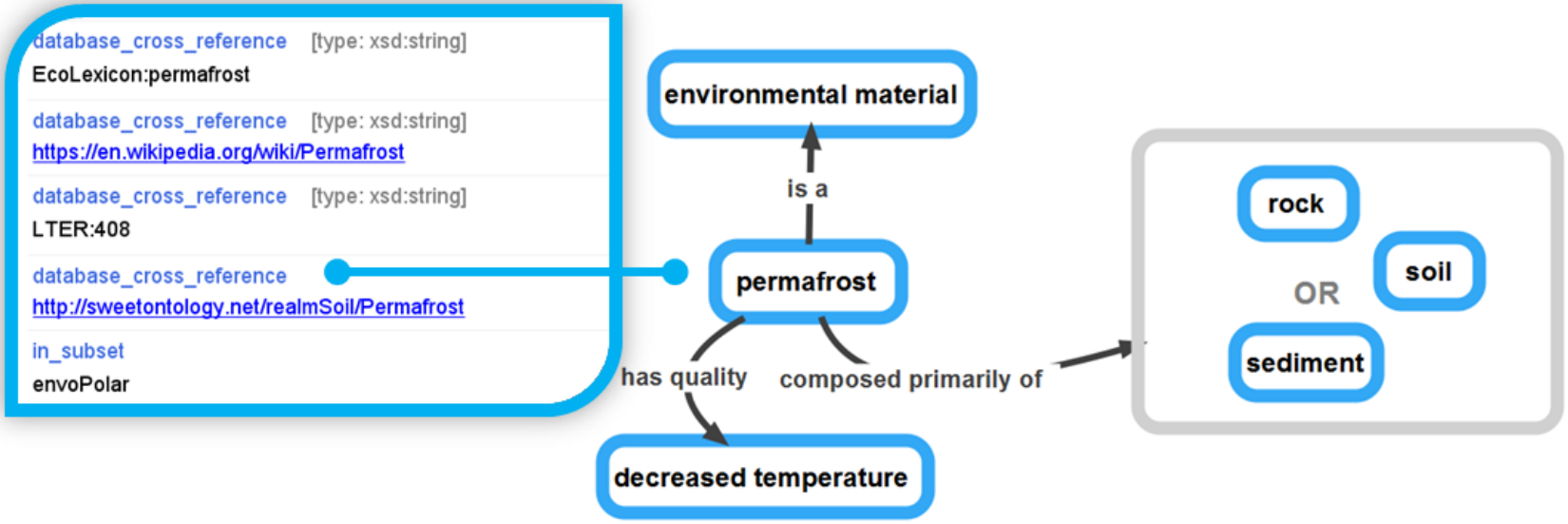
- ENVO now offers representations of habitats,
  - environmental processes,
  - anthropogenic environments, and
  - entities relevant to environmental health initiatives and
  - the global Sustainable Development Agenda for 2030.
- Several branches of ENVO have been used to incubate and seed new ontologies in previously unrepresented domains such as food and agronomy.
- Cryo, marine, soil, landforms…..

So what has the Semantic Harmonization Cluster been doing with Cryospheric terms?

Using the GCW glossary analysis results to:

1. Provide definitions for SWEET terms
2. Link those SWEET terms to ENVO and other OBO Foundry ontology terms
3. Build out the envoPolar subset of ENVO with relevant terms and axioms
4. Attribute all of these updates to the people or groups responsible for these changes

(part of) the current state of permafrost semantics

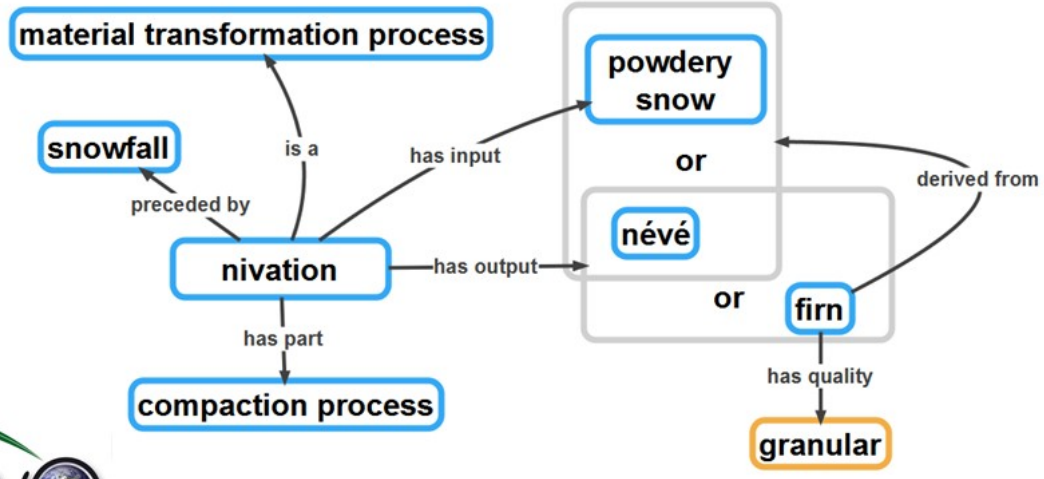


Definition: Soil or rock and included ice or organic material at or below the freezing point of water (0 degrees Celsius or 32 degrees Fahrenheit) for two or more years.

Comment: Ice is not always present, as may be in the case of nonporous bedrock, but it frequently occurs and it may be in amounts exceeding the potential hydraulic saturation of the ground material.

Prototype description

ESIP CryoHackathon series!



PATO

Each node links to many others in ENVO and other interoperating ontologies

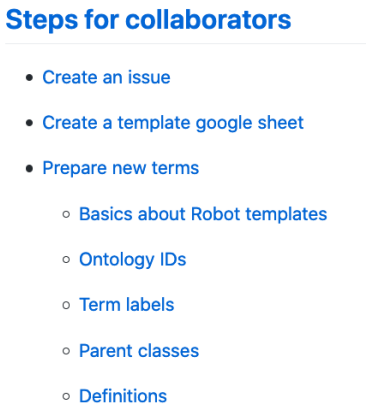
• Some post 2016 Conclusions about EnvO Status/Future

- ENVO bridges multiple domains including biomedicine, natural and anthropogenic ecology, ‘omics, and socioeconomic development.
- Interactions with our users, partners, related ontologies, domain vocabularies accelerating
- Invites further contributions and collaboration to advance the semantic representation of the environment, ranging from geographic features and environmental materials, across habitats and ecosystems, to everyday objects in household settings.

Example Tooling

[ENVO-Robot-template-and-merge-workflow]

- New ENVO collaborative workflow using OBO tools (Robot)
  - Guides users through the process of creating new terms
  - Intended for non-ontologists!



- Overview:
  - Organize new term requests in standardized google sheet template
  - Step by step guidance to fill out the appropriate columns
  - Protocol for ontologists to add request to ENVO (not covered in presentation)

[Example template]

Kai Blumberg Univeristy of Arizona & BCODMO kblumberg@email.arizona.edu

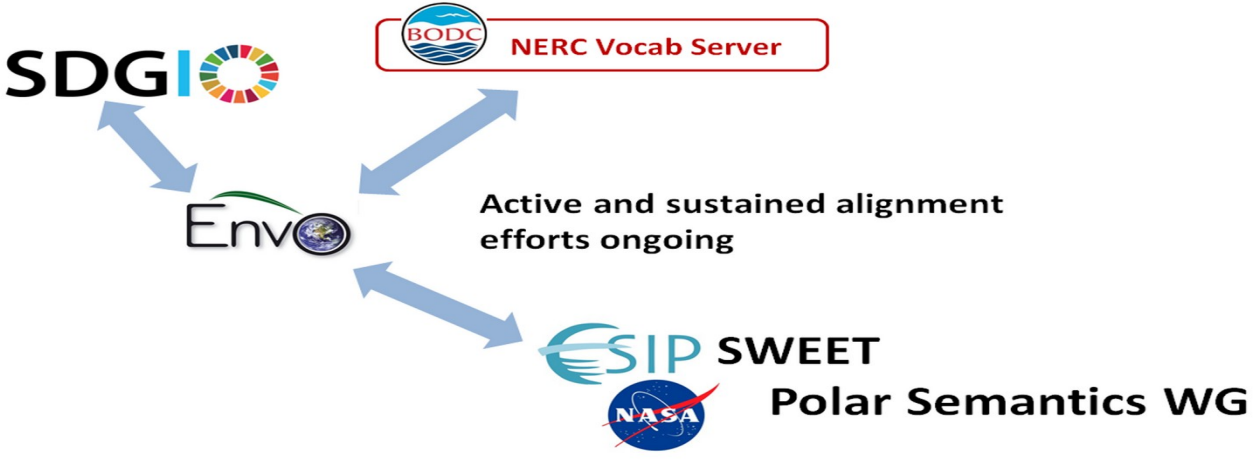
Align, Re-Use, Extend, and Revise not Re-Invent!

Semantic Harmonization:

- Established Earth Science vocabularies exist, e.g.
- SWEET, EnvO, GCMD, CF Metadata, SVO,more...
- Modernize and harmonize
- Extend and revise

- Established General vocabularies exist, e.g.
- Schema.org, Dublin Core, DCAT2, VIVO, DDI, more…
- Re-Use!





Vocabularies & Ontologies (like Nations!) need Governance

- Community development and vetting
  - Using Slack, issues on github…
- Mechanisms for maintaining access to the vocabulary, and extending/revising it (visualizing it)
- Long-term sustainability (***ESIP can help! COR***)
- Partner agreements
- Operations
- Open, collaborative
- Support?



Thanks!

ESIP Semantic Harmonization Cluster Info at:

Mailing List: https://lists.esipfed.org/mailman/listinfo/esip-semanticharmonization

Wiki Page: https://wiki.esipfed.org/SemanticHarmonization

Join us:

Telecons via **GoToMeeting**: https://global.gotomeeting.com/join/800617765

- **Third Wednesday of the month at 2pm EST**

Some References/Links

- Ontology Tools and Resources
  - http://www.obofoundry.org/resources
- Matentzoglu, N., Malone, J., Mungall, C. et al. MIRO: guidelines for minimum information for the reporting of an ontology. J Biomed Semant 9, 6 (2018). https://doi.org/10.1186/s13326-017-0172-7
- Seppälä, Selja, Alan Ruttenberg, and Barry Smith. "Guidelines for writing definitions in ontologies." Ciência da informação 46.1 (2017).
- *DEFINITIONS IN ONTOLOGIES*
  - Selja Seppälä, Alan Ruttenberg, Yonatan Schreiber, Barry Smith
- Michie, Susan, Robert West, and Janna Hastings. "Creating ontological definitions for use in science." Qeios (2019).
  - https://pdfs.semanticscholar.org/139c/3e8cd4d82b4520300404da5f6001e71d6006.pdf
- All things EnvO see https://github.com/EnvironmentOntology/envo
- Buttigieg, P. L., Pafilis, E., Lewis, S. E., Schildhauer, M. P., Walls, R. L., & Mungall, C. J. (2016). The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. Journal of biomedical semantics, 7(1), 1-12.
- Advances in Semantic Harmonization: from the Cryosphere, to the Earth System (ESIP 2021)
  - https://docs.google.com/presentation/d/1bmP8mC_VKpfIUXkejEBtNTqOqrf73NJ5gljeWTFtic0/edit#slide=id.gb44943ed1f_0_9