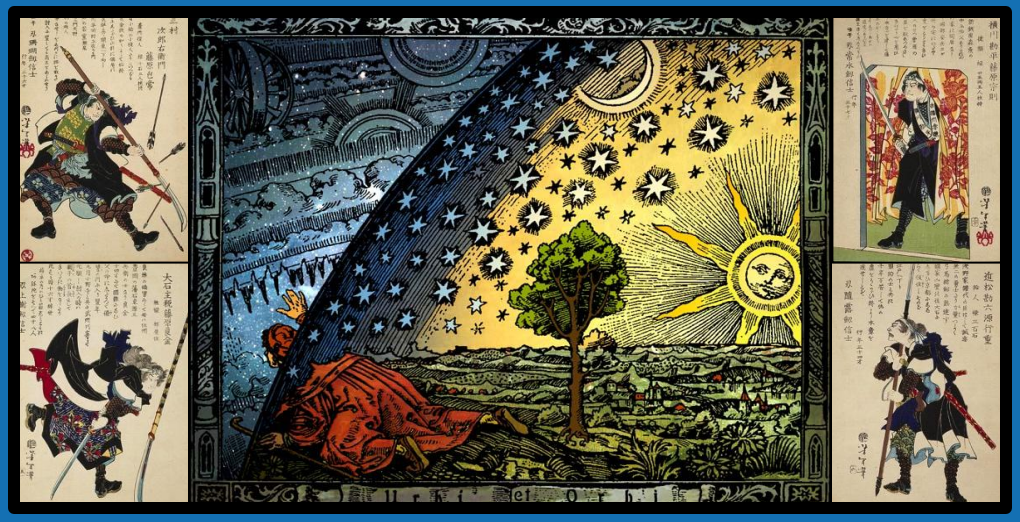


Polar Data Activities

Ruth Duerr

Outline

- GCW glossary
- Arctic Data Committee
- Vocabularies and Semantics Working Group
- Other activities

GCW Glossary

Currently a database of terms and their definitions from 26 sources

Goals:

- To include and be consistent with the recommended and desired variables for CRYONET
- To include and be consistent with the GCW best practices for cryospheric measurements.
- The GCW glossary terms will ultimately be included in WMO's METEOTERM.

See https://globalcryospherewatch.org/reference/glossary.php

GCW Glossary

Two reports were commissioned:

1. An assessment of existing terminology in the Global Cryosphere Watch (GCW) Glossary indicating terms:
   - that are **not problematic** from a semantic standpoint,
   - where multiple definitions can be **coalesced** into a single definition,
   - where the terminology is **inconsistent** and therefore problematic from a semantic standpoint, and
   - where **community resolution** is needed to either agree on a definition or to split the terms up into separate entities.
2. An assessment of how these terms align with concepts in a set of leading semantic resources including the SWEET, ENVO, and Geoscience ontologies and makes recommendations on addressing overlaps and discrepancies

Activities - GCW Glossary - First report

Type of Term	Number of Items	Number of Unique Items
Original spreadsheet	4,141	2,230
Not problematic terms	1,076	1,076
Terms to coalesce	2,033	728
Inconsistent terms	502	131
Community resolution needed	582	269
New Totals	4,193	2,204

Terminology varied by subdiscipline, time period, and geographic source

GCW Glossary: Report 2

Resources assessed:

- GCMD science keywords
- CF Metadata conventions
- GeoSciML
- Geoscience ontology
- SWEET (Semantic Web for Earth and Environmental Terminology) ontology
- ENVO (Environment Ontology)

GCW Glossary - Report 2 Recommendations

- Monitor developments and contribute Cryospheric input to the relevant community as warranted (all).
- Review current status (GCMD, SWEET, ENVO).
- Suggested updates:
  - Add GCW Glossary definitions to GCMD, CF conventions, Geoscience ontologies, SWEET, ENVO,
  - Adding instruments and projects to GCMD,
  - Possibly suggesting tweaks to the hierarchy of the GCMD Science Keywords, and
  - Suggesting modifications to the hierarchies and axioms in SWEET and ENVO (examples, adding/updating properties and relationships).
- Working with the GeoSciML communities to develop a series of controlled vocabularies for subsets of the GCW Glossary (e.g., a sea ice subset, a glacier subset, etc.)

Arctic Data Committee

- A combination of the data committees of the International Arctic Science Committee (IASC) and the Sustaining Arctic Observing Networks
- Series of meetings over the past couple of years

ADC Activities

- Matrix of metadata standards and aggregation protocols fleshed out at meeting in Geneva last fall
- Matrix currently has information from 81 polar repositories, aggregators, etc.
  - However, many records are as yet incomplete
- In addition, a white paper summarizing the matrix and adding 20 recommendations for data centers, funders, is in process.
  - 13 technical recommendations
  - 7 social and organizational recommendations

Main Recommendations re federated search

- Implement schema.org data set metadata on landing pages
  - Spurred by Google's dataset search (https://toolbox.google.com/datasetsearch)
  - bioschema.org extensions for the bio community
  - https://github.com/ESIPFed/science-on-schema.org for geosciences
- schema.org extensions can facilitate federated search
  - NSF's EarthCube program is planning on crawling known repositories to accumulate a federated discovery system

ADC Next Steps

- Develop a federated search capability across all polar data centers

Vocabularies and Semantics Working Group

Vocabularies and Semantics Working Group

- A joint effort between the Arctic Data Committee and the Arctic Data Sub Team of the U.S. Interagency Arctic Research Policy Committee
- Open to any interested party
- Virtual meetings every third Tuesday via Zoom
- https://arcticdc.org/activities/core-projects/vocabularies-and-semantics-wg
- Send email to polarsemantics@gmail.com to join email list

Overview – Purpose of the Group

- Promote awareness of existing vocabularies and semantics initiatives to increase effectiveness and reduce or eliminate redundancy
- Coordinate vocabularies and semantics development activities across the polar information community
- Enable and organize regular communication within the community
- Help members of the community connect to useful and interoperable vocabularies
- Inform the polar community about broader activities (e.g. ESIP, RDA), and act as ambassadors from the polar community to other initiatives

Activities

- Compiling a list of polar related semantic resources in the form of a questionnaire
- Compiling information on the wide variety of related activities
- Developing educational materials for data managers, scientists on semantic topics
- Monthly status and update telecons

Other Activities

Research Data Alliance

- Participating in schema.org work for any kind of research
- Working on semantically standard naming for measurement parameters

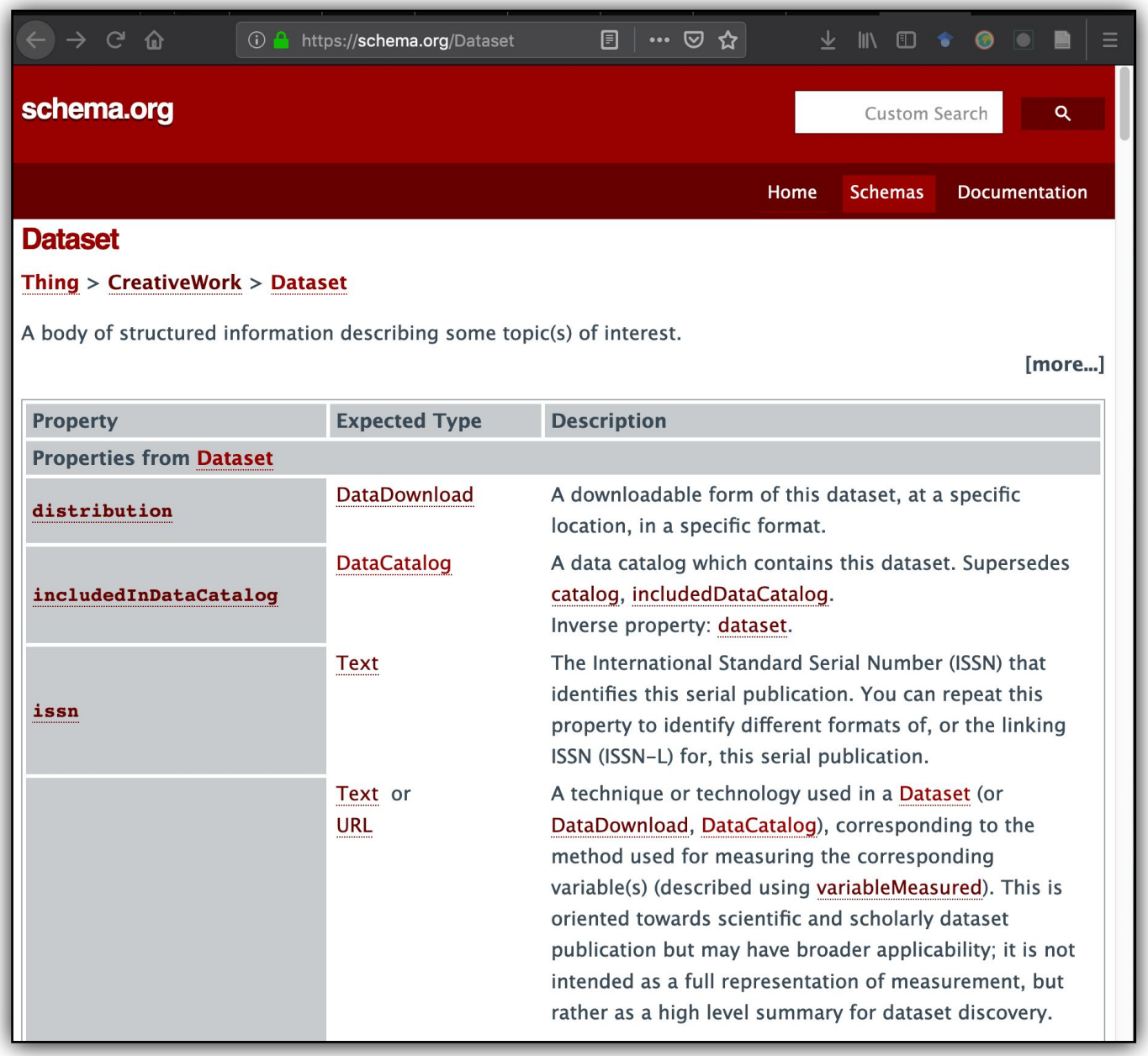
ESIP Semantic Technologies Committee

- Ontology repository for Earth science ontologies, vocabularies, etc. (http://cor.esipfed.org)
- Semantic harmonization cluster is working to harmonize SWEET & the GCMD keywords with ENVO and the OBO Foundry ontologies, while using GCW definitions for cryospheric terms
- A new geoscience on schema.org cluster has been spun up to provide recommendations for implementation for all geoscience repositories

Questions?

Background

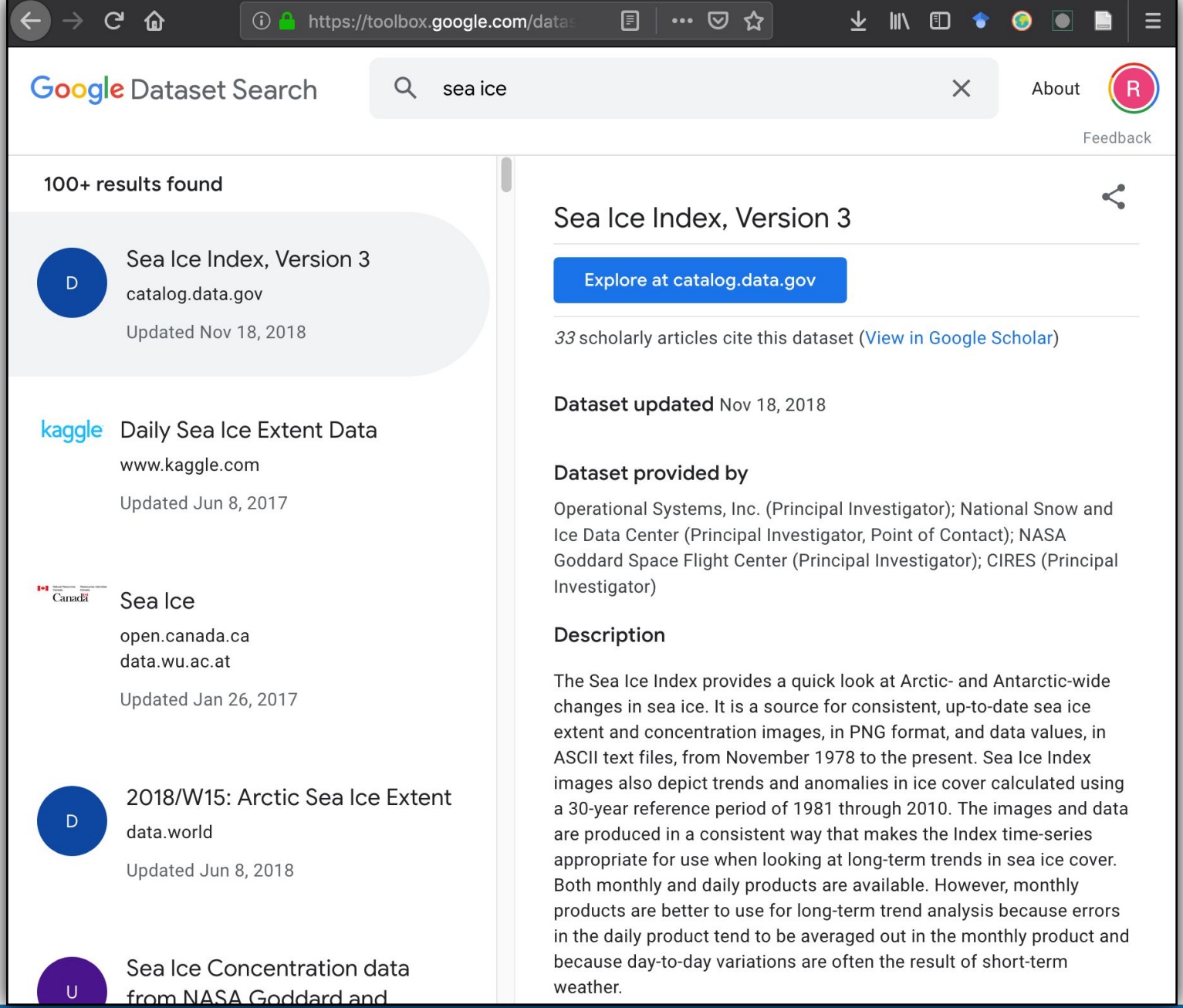
- schema.org metadata has been used on web pages by online vendors for years
  - Why?
    - It is indexed by all of the major search engines (e.g., Google, Bing, etc.)
    - Using it makes your page show up higher in search results!
- A few years ago the data community started working to get data related terms into schema.org


https://schema.org/Dataset
schema.org
Custom Search
Home
Schemas
Documentation
Dataset
Thing > CreativeWork > Dataset
A body of structured information describing some topic(s) of interest.
[more...]
Property
Expected Type
Description
Properties from Dataset
distribution
DataDownload
A downloadable form of this dataset, at a specific location, in a specific format.
includedInDataCatalog
DataCatalog
A data catalog which contains this dataset. Supersedes catalog, includedDataCatalog.
Inverse property: dataset.
issn
Text
The International Standard Serial Number (ISSN) that identifies this serial publication. You can repeat this property to identify different formats of, or the linking ISSN (ISSN-L) for, this serial publication.
Text or URL
A technique or technology used in a Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s) (described using variableMeasured). This is oriented towards scientific and scholarly dataset publication but may have broader applicability; it is not intended as a full representation of measurement, but rather as a high level summary for dataset discovery.

More Background

Last year Google stood up a dataset search engine somewhat analogous to scholar.google.com

https://toolbox.google.com/datasetsearch


https://toolbox.google.com/datas
Google Dataset Search
sea ice
About
Feedback
100+ results found
Sea Ice Index, Version 3
catalog.data.gov
Updated Nov 18, 2018
Daily Sea Ice Extent Data
www.kaggle.com
Updated Jun 8, 2017
Sea Ice
open.canada.ca
data.wu.ac.at
Updated Jan 26, 2017
2018/W15: Arctic Sea Ice Extent
data.world
Updated Jun 8, 2018
Sea Ice Concentration data from NASA Goddard and
Sea Ice Index, Version 3
Explore at catalog.data.gov
33 scholarly articles cite this dataset (View in Google Scholar)
Dataset updated Nov 18, 2018
Dataset provided by
Operational Systems, Inc. (Principal Investigator); National Snow and Ice Data Center (Principal Investigator, Point of Contact); NASA Goddard Space Flight Center (Principal Investigator); CIRES (Principal Investigator)
Description
The Sea Ice Index provides a quick look at Arctic- and Antarctic-wide changes in sea ice. It is a source for consistent, up-to-date sea ice extent and concentration images, in PNG format, and data values, in ASCII text files, from November 1978 to the present. Sea Ice Index images also depict trends and anomalies in ice cover calculated using a 30-year reference period of 1981 through 2010. The images and data are produced in a consistent way that makes the Index time-series appropriate for use when looking at long-term trends in sea ice cover. Both monthly and daily products are available. However, monthly products are better to use for long-term trend analysis because errors in the daily product tend to be averaged out in the monthly product and because day-to-day variations are often the result of short-term weather.

ENVO

PL Buttigieg
@ AWI

includes

An OWL-based community ontology handling the semantics of environmental entities using OBO best practices

environmentontology.org

envoCryo

bundling cryosphere semantics

partly seeded by the NSIDC sea ice ontologies

is concerned with…

…systems

- biomes
- ecoregions
- ecozones
- habitats
- …

…planetary parts

- coral reefs
- cryconite holes
- holothurian guts
- melt ponds…

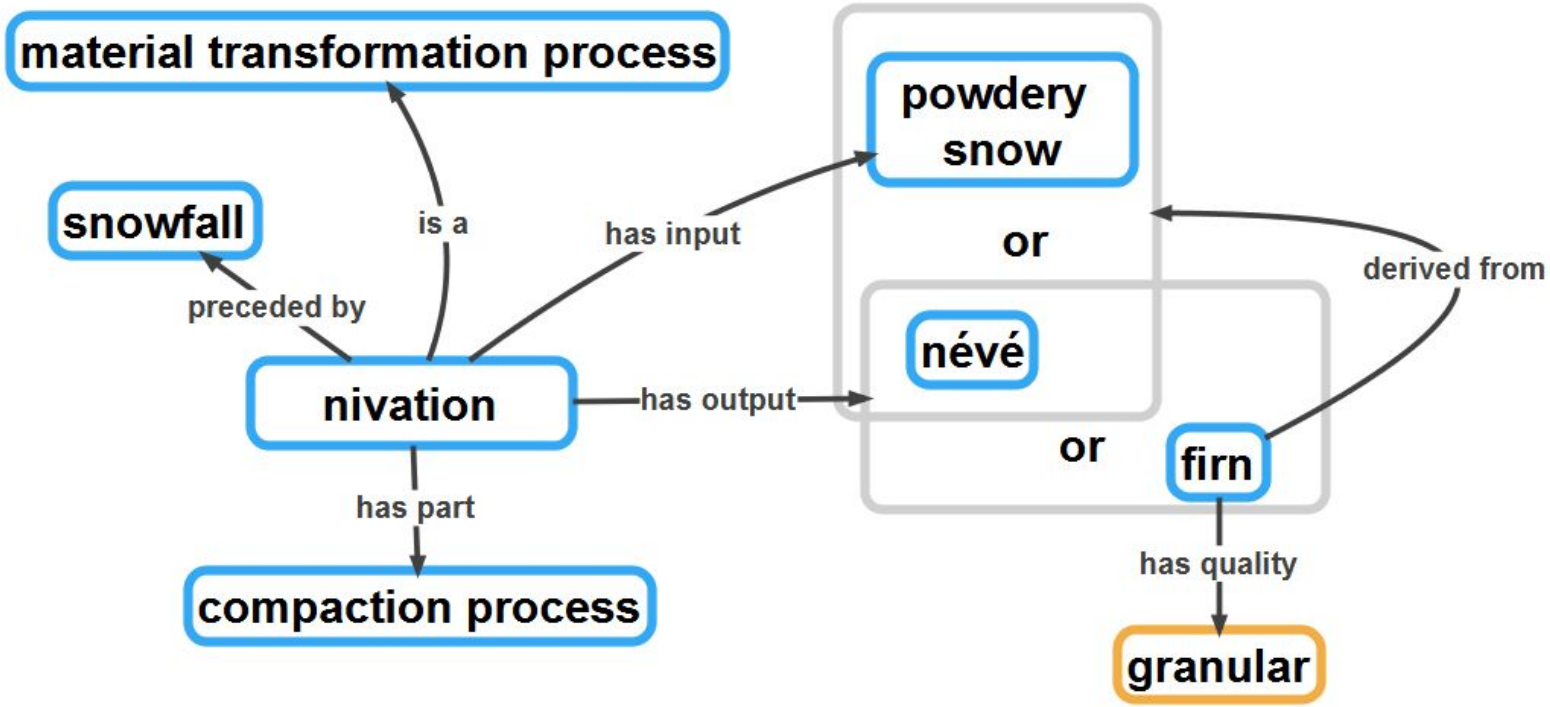
…processes

- glacial ablation
- nivation…

…materials

- névé
- brine
- permafrost
- …

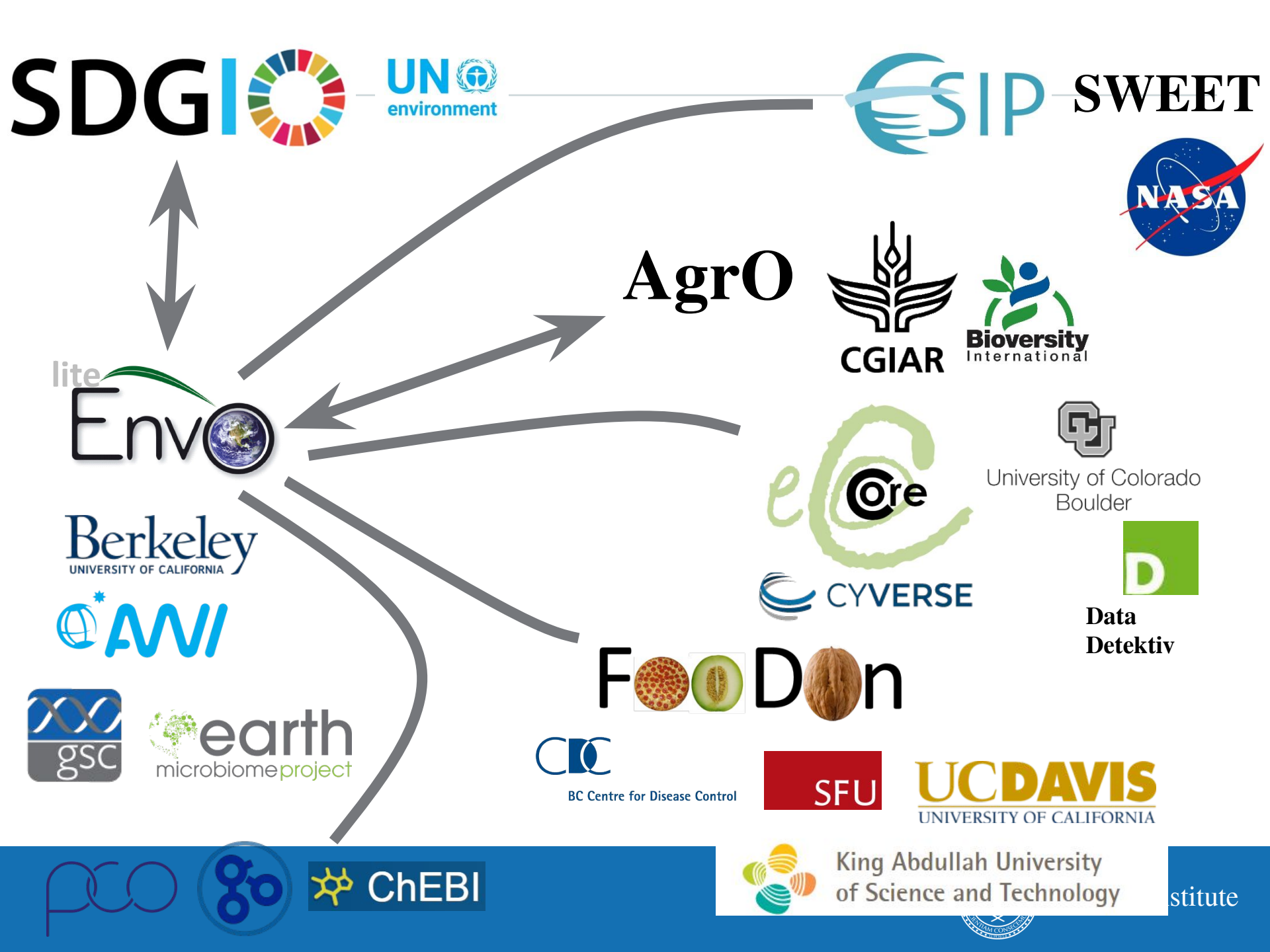
A fragment of envoCryo's nivation semantics



envoCryo

PATO

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


SDGIO
UN environment
ESIP
SWEET
NASA
AgrO
CGIAR
Bioversity International
lite
EnvO
eCore
CYVERSE
University of Colorado Boulder
Berkeley
UNIVERSITY OF CALIFORNIA
AWI
Data Detektiv
FoodOn
gsc
earth microbiome project
BC Centre for Disease Control
SFU
UCDAVIS
UNIVERSITY OF CALIFORNIA
ChEBI
King Abdullah University of Science and Technology
stitute