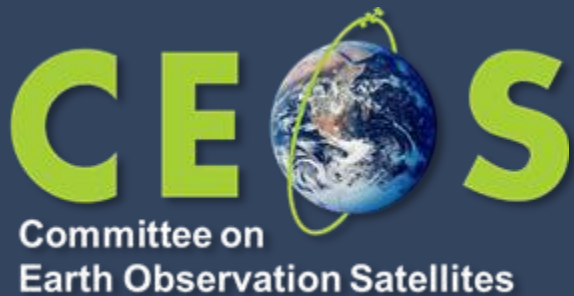


CEOS-WGCV/WGISS

Common Dictionary/ Vocabulary



Peter Strobl, EC-JRC
Katrin Molch, DLR
Emma Woolliams, NPL
Agenda Item E.3
WGISS-58/WGCV-54
15&18 October 2024

Sioux Falls, South Dakota, USA

How it started

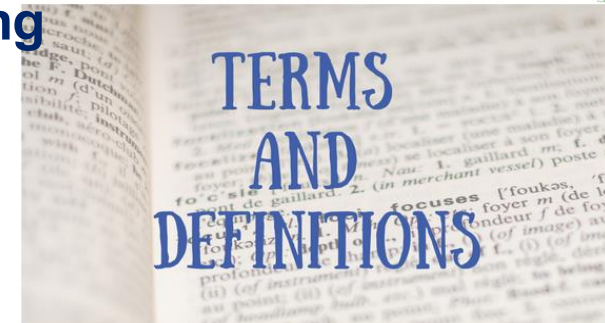


Background

- Joint initiative of CEOS WGCV, CEOS WGISS, and CEOS LSI-VC
- Mandate: CEOS WGCV Action Item 49-06 (June 2021) “[...] to advance the idea of a CEOS common online dictionary, with a view to eventually reaching out to launch a broader community effort”

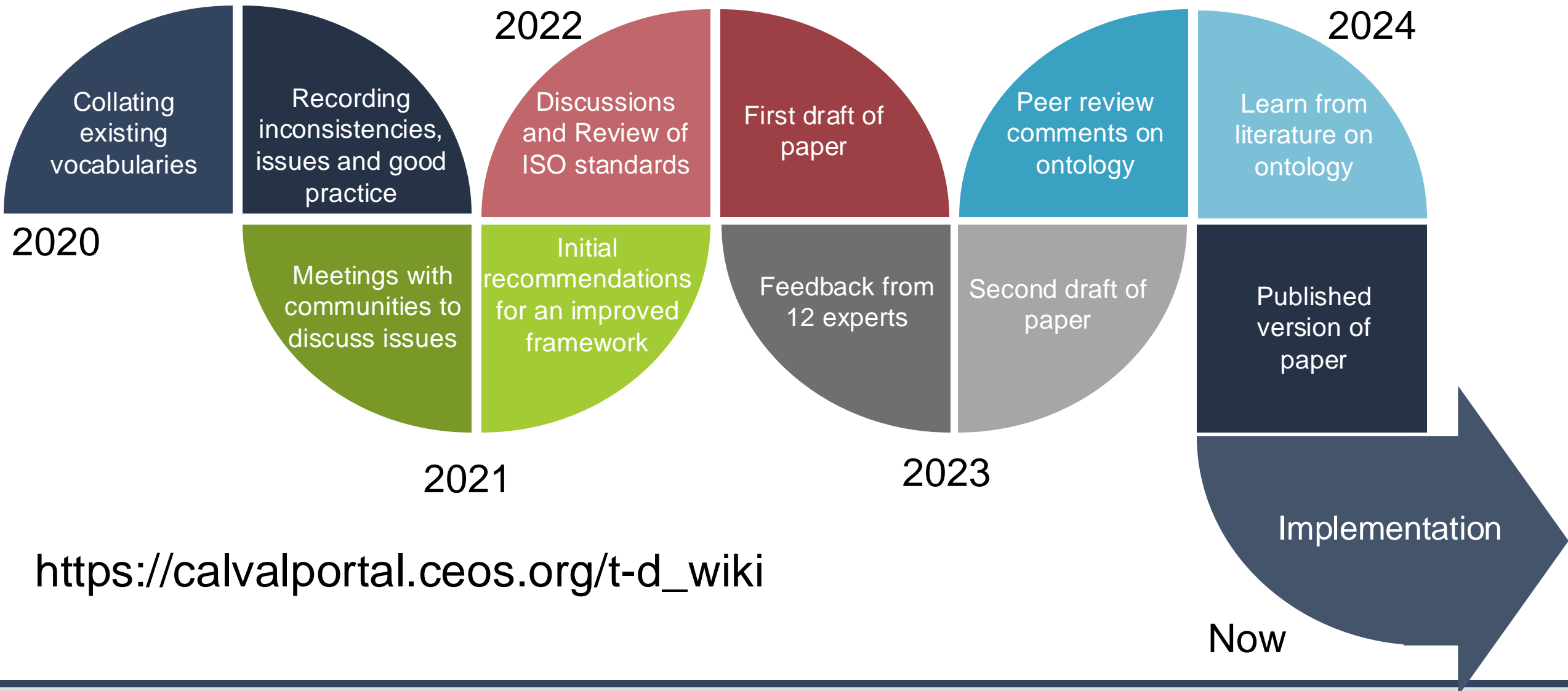
Status of activities

- CEOS Common Terminology group established August 2021
- Workplan drafted
- Integrated WGISS Data Stewardship Glossary (pdf) and NOAA NESDIS Data Management Lexicon (pdf) into existing WGCV ‘Terms and Definitions’ Wiki
- Discussing categorization of terms- prototyping approach using base terms and high impact terms – ongoing
- Discussing governance
- Reaching out to promote initiative internally with focus on WGISS, WGCV, LSI-VC externally with ISO/OGC
- Summarize finding - publish



http://calvalportal.ceos.org/ca/t-d_wiki

Process



https://calvalportal.ceos.org/t-d_wiki

Format of terminologies



TC 211 Geographic information/Geomatics

ISO/TC 211 Multi-Lingual Glossary of Terms (MLGT)

NEWS & ARTICLES REGISTERS STATISTICS ABOUT FEEDBACK SEMANTIC SEARCH

All Concepts

The concepts below are sorted by term name, alphabetically, in a case-insensitive way.

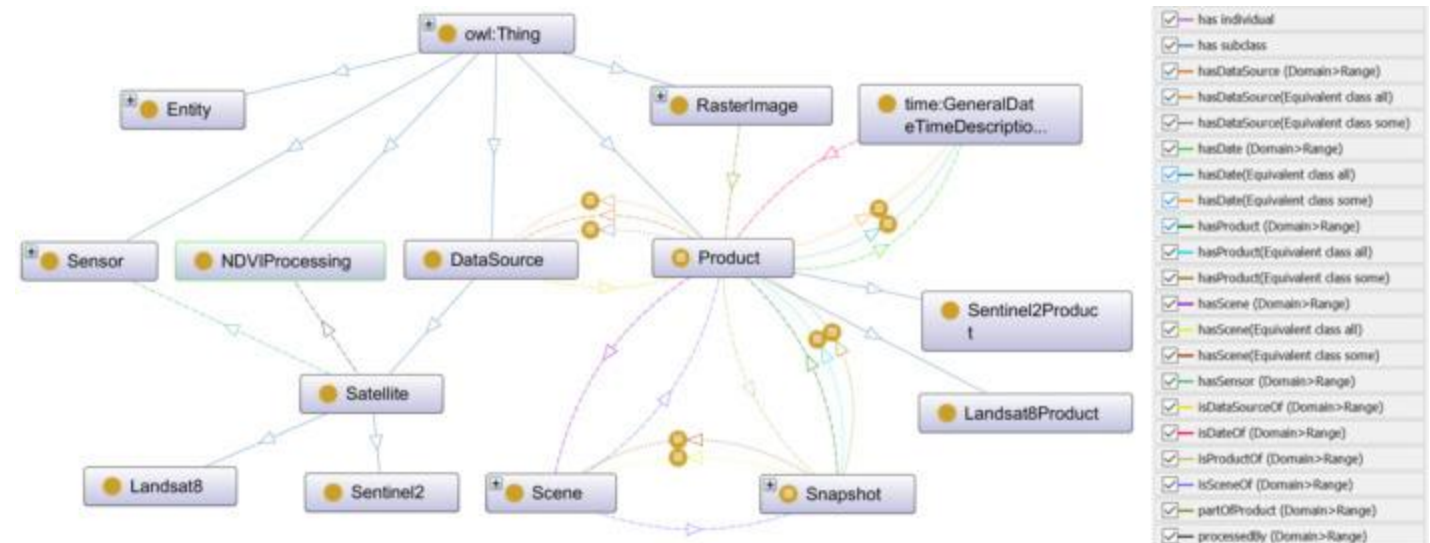
Term ID	Term
815	a posteriori classification
816	a priori classification
2	abbreviation
1928	absolute accuracy
2528	abstract
2028	abstract root
2471	abstract test case
3	abstract test case
4	abstract test method
5	abstract test module

❖ Mostly alphabetical lists

- Often in PDFs
- Hard to 'explore'

❖ Or formal ontology

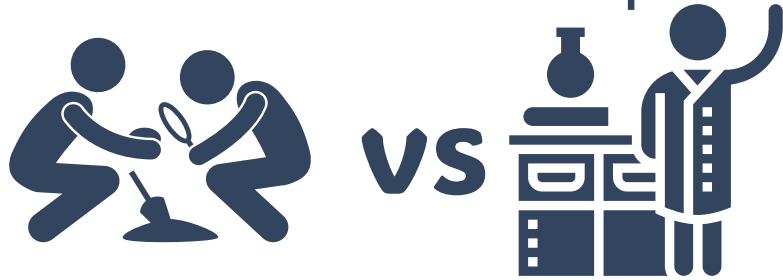
- Meant for computers, not people



- ❖ Simple lists of words are not used often – structure is important
- ❖ Rare to have versioning (or keep older definitions) – and definitions change
- ❖ Very large number of definitions for some terms – e.g. ISO online browsing platform has 77 versions of ‘observation’
- ❖ Inconsistent definitions (e.g. in-situ, observation, sample, ...)
- ❖ Superficial definitions (e.g. interoperability) lacking full framework
- ❖ Circular definitions – and poor use of the foundational ‘base terms’
- ❖ Development process – isolated efforts – creates these problems

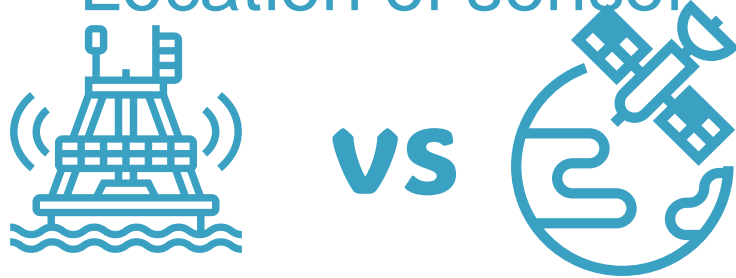
In-situ disambiguation

Location of sample



- ❖ In-situ / Ex-situ
Observed in its original location vs brought back to laboratory for testing

Location of sensor



- ❖ In-situ / satellite
Observed e.g. ground-based, sea- or airborne vs from an orbiting platform

Distance sensor to sample



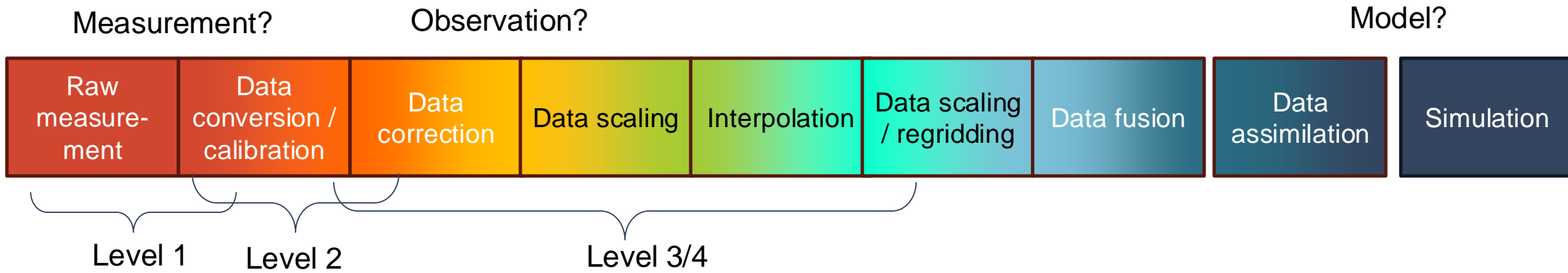
- ❖ In-situ / remote
Observed close to the location of the sample vs from a (significant) distance

Various communities use the term 'in-situ' in different ways – in contrast to different alternatives.

It is very important to be clear about which notion of 'in-situ' is referred to.

Sometimes it relates more to location and sometimes to fidelity.

- ❖ Process of observing / observation value
- ❖ Continuum from measurement to simulation
 - Bokulich 2020 (philosopher) defined increasing ‘model ladenness’ of data – as below
 - (but other splits possible!)
- ❖ Distinction hard – but we disagree with ISO approach of considering simulation as a form of observation based only on an algorithm as observer!



interoperability is a 'characteristic of a product or system to work with other products or systems'

Easy to agree, unhelpful to use

- ❖ Need a framework for each community to define more specific aspects of this.
- ❖ Building on **same** core and base terms. Linked through **single** top-level definition.

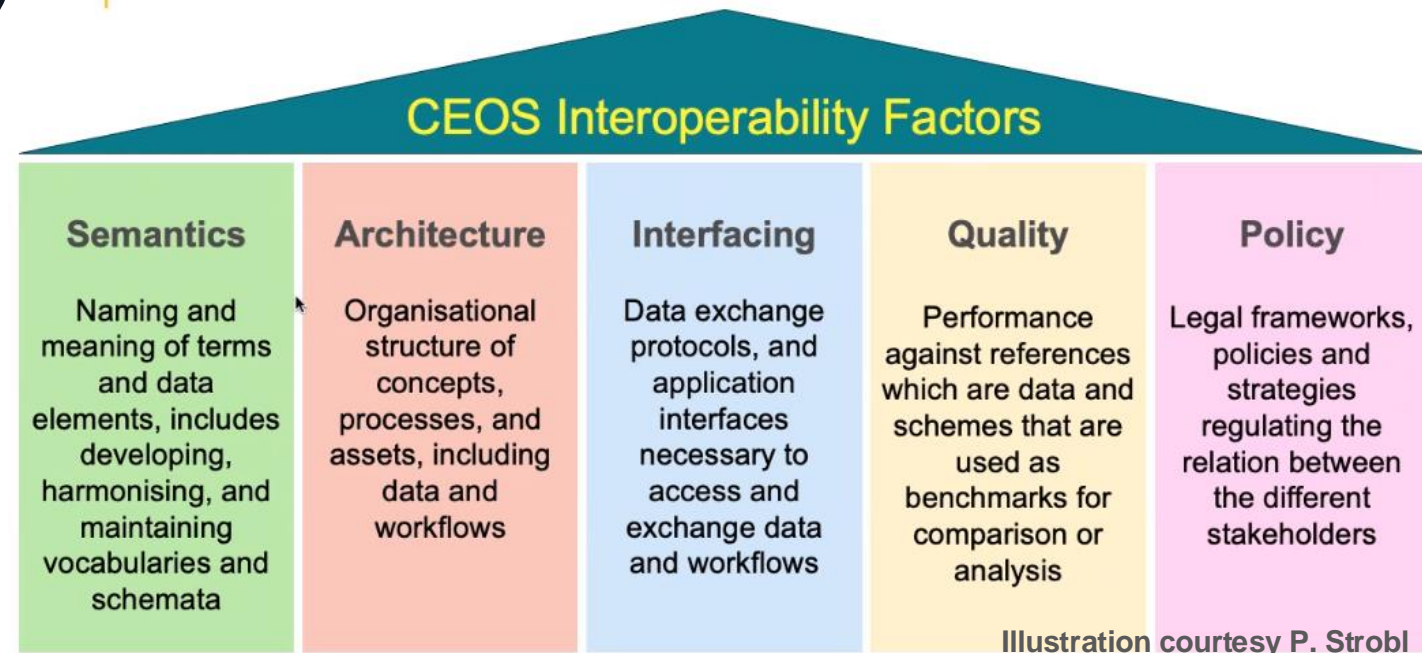


Illustration courtesy P. Strobl

Lost in translation: The need for common vocabularies and an interoperable thesaurus in Earth observation sciences

P.A. Strobl¹, E.R. Woolliams² and K. Molch³
¹EC-JRC, ²NPL, ³DLR

Journal: Surveys in Geophysics

Online: 1 October 2024



<https://doi.org/10.1007/s10712-024-09854-8>

- ❖ First draft (July 2023) reviewed by 12 observation experts
- ❖ Journal peer review included significant challenge from ontology expert
- ❖ Challenging review procedure (8 months)
- Final paper has benefited from the insights of all these perspectives.

Surveys in Geophysics
<https://doi.org/10.1007/s10712-024-09854-8>



Lost in Translation: The Need for Common Vocabularies and an Interoperable Thesaurus in Earth Observation Sciences

P. A. Strobl¹  · E. R. Woolliams²  · K. Molch³

Received: 15 December 2023 / Accepted: 17 July 2024
© The Author(s) 2024

Abstract

The Earth Observation sciences are highly multidisciplinary with long value chains from the development, characterisation and deployment of sensors, through data processing and modelling, to the information services provided to decision makers in, for example, governments, companies and non-governmental organisations. A prerequisite to any multidisciplinary collaboration is effective communication and many communities involved in the value chains have developed vocabularies or terminologies to define terms from a particu-

- ❖ Formal ‘thesaurus’ (controlled and structured vocabulary, with explicit relationship between terms)
- ❖ Collaboration between standard-setting and good-practice setting organisations to develop a collective-effort wiki-style vocabulary
 - Long-term commitment to collaborative working
 - Oversight editorial teams
- ❖ Terminology in documents linked to central reference (with version control at the individual definition level)
 - Conflicts handled through separate ‘branches’
 - More complete frameworks (e.g. CEOS interoperability / CEOS FRM etc) connected to the terminology alongside higher level definitions

A good thesaurus is:



Consistent

On a foundation of base terms. No ambiguity. No circular or overlapping definitions. Clear preferred terms. Alternative definitions explained clearly.

Interrelated

Clearly shows relationships between terms – in a definitional, contextual and ontological way. Mutually exclusive parent / sibling / child relationships

Understandable

Clear definitions, with differentiation/disambiguation of controversial terms. Checked for understandability by multidisciplinary teams

Educational

Not just for computers! Promotes a conceptual framework, shows linkages, satisfies curiosity, helps communication of concepts. Links to detailed frameworks for high-impact terms.

Updateable

Unified thesaurus with version control at the level of individual terms, methods for providing disambiguation links, and adding new terms.

Categories of terms



	Category of term	Examples	
	Base terms	Data, entity, phenomenon, property	Underpinning terms – small committee to create list
	Core terms	Uncertainty, leaf area index, data centre, spectral resolution, accuracy, precision, measurement	Basic vocabulary for Earth sciences – small committee to create list and collate existing definitions
	Controversial terms	Sampling, observation, in-situ, processing level, model, confidence	Terms different communities use differently. Facilitated discussions or disambiguation pages
	High impact terms	Interoperability, analysis ready data, data assimilation, real-time, operational	Terms that need a framework document – developed in the right committee. Link to simple, high-level definition

Suggested practical approach



- ❖ Build on all the existing, excellent work and invite all stakeholders
 - But someone must start this!
- ❖ Treat it as a proper project, led by a ‘thesaurus curator’ and using professional tools to link definitions (glossarist?, github?)
 - Produce and maintain guidelines and principles for definitions from the start
 - Get a small group to build the ‘base’ and ‘core’ terms first, so other definitions build on those
- ❖ Invite the right technical experts to build other definitions on those base and core terms, using existing terminologies
 - Checked against new principles
 - Make disagreements open and public
- ❖ Link to more complete framework documents for high-impact terms

Someone just started!



- ❖ The 'Knowledge Centre on Earth Observation' of the European Commission has started implementing these concepts in the [KCEO Glossary](#)
- ❖ Opportunity to participate once initial set up completed
- ❖ Do we want to have a formal CEOS connection?

This glossary is currently at an early stage and still work in progress. It is not 100% reviewed yet and some things might change over time. Stay tuned in our [blog](#) or on [GitHub](#)!

European Commission | KCEO Glossary

Home Glossary Blog Contact

Glossary

- Laboratory observation
- Latency
- Lead time
- Lithosphere
- Location
- Location error
- Measurement**
- Measurement uncertainty
- Measurand
- Minimum Mapping Unit
- Minimum Mapping Width
- Model

Measurement

1 Definition

A measurement is an **Observation** of a **Quantity**.

Notes

- The process of collecting a measurement is called measuring.

Examples

Table of Contents

- 1 Definition
- Notes
- Examples
- Sources



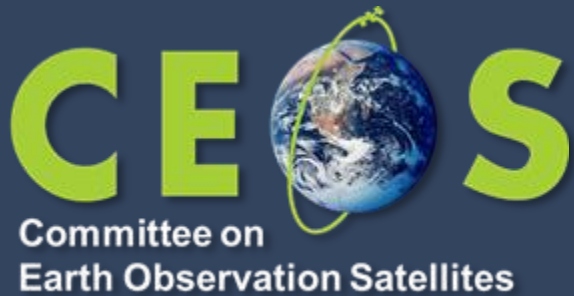
KNOWLEDGE CENTRE ON
EARTH OBSERVATION

Looking for an
EÓ GLOSSARY?



ec-jrc.github.io/KCEO-Glossary/

You found it!



Thank you!

any questions?