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. Strobl[1] · E. R. Woolliams[2] · K. Molch[3]

Received: 15 December 2023 / Accepted: 17 July 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 particular viewpoint or legacy. However, these vocabularies are often inconsistent, with circular definitions, contradictions and using technical terms that are not defined. Here, three case studies from Earth Observation disciplines are considered involving challenges in the definition and use of the terms 'observation', 'in-situ' and 'interoperable'. An approach is suggested for an initiative, starting in Earth Observation, to build a consistent thesaurus taking inspiration from the ISO 25964:2011 standard.

Plain Language Summary
A wide community of engineers, scientists and data quality experts engages in collecting, processing, analysing, and distributing data about the state of our planet and how its environment is changing. These people have various backgrounds and specialised jargons that assist in communicating with others in their own area, but confusion often results when different communities interact. These differences can limit interdisciplinary work. Here we discuss some of the problems with existing formal vocabularies and propose a project structure for developing a more consistent vocabulary for monitoring and understanding planet Earth.

Keywords Vocabulary · Thesaurus · Earth Observation · Observation · Interoperability

✉ P. A. Strobl
peter.strobl@ec.europa.eu

[1] European Commission, Joint Research Centre (JRC), 21027 Ispra, Italy

[2] National Physical Laboratory (NPL), Hampton Road, Teddington TW11 0LW, UK

[3] German Remote Sensing Data Center, German Aerospace Center DLR, Muenchner Str. 20, 82234 Wessling-Oberpfaffenhofen, Germany

Article Highlights

- Current vocabularies published for geospatial data collate and list definitions in use but have inconsistencies, especially in how they use what we call ‘base terms’
- A good vocabulary that is presented as a formal ‘thesaurus’ and is consistent, educational, structured, relational and updateable will improve interdisciplinary communication and research
- Establishing such a thesaurus requires a new initiative endorsed by all the stakeholders (standardisation communities, agencies and authorities) and open to all interested parties working collaboratively on this structured, decentralised and formally managed project

1 Introduction

The Earth systems (the interacting physical, chemical, and biological processes that shape the planet) are inherently complex and interconnected. Observations of the Earth require processing through long ‘value chains’ from the development, characterisation and deployment of sensors, through data processing and modelling, to the information services provided to decision makers. Effectively understanding and managing the inherent complexity and enabling all those involved in these ‘value chains’ to share data and information, requires an interdisciplinary, holistic systems-thinking approach. Increasingly scientists in the Earth sciences are working in interdisciplinary teams.

A prerequisite to any interdisciplinary collaboration is communication. Precise use of language supports the dissemination of information and assists in the interpretation of data. However, it is common for interdisciplinary activities to struggle with terminology, for example when communities use the same term in different ways or use various terms to mean the same thing. Such miscommunications can take considerable time to discover and correct. Sometimes, miscommunication has an even deeper root—a different conceptual framework to think about the Earth systems and our attempts to understand them through instruments and models. Conceptual (epistemological) differences easily become philosophical and are difficult to resolve (MacLeod 2018). Furthermore, in the Earth sciences, scientists are increasingly using machine learning and data mining approaches to automate aspects of data exploration and analysis. These computer-based methods require unambiguous terminology to ensure that data are interpreted consistently. Therefore, it is crucial that scientists communicate clearly not only with one another, but also with computers. By using clear and precise language, scientists can support the dissemination of information and ensure that data are interpreted dependably by both humans and machines.

To counter miscommunication and to provide consistent terminology, data and metadata that both machines and humans can understand, many communities create glossaries, formal vocabularies, or terminologies to define how terms are used within their own communities (Sect. 2.1 discusses these terms). However, even within a single community, these vocabularies can be difficult to use. This difficulty is in part due to the different perspectives and expertise from which the task is approached. While generalists or non-experts will seek understandable definitions, specialists might want to narrow down the overall scope and particular meaning, and an ontologist will put an emphasis on the relations between words and on overall topic structure with less emphasis on individual definitions.

Accordingly, as described in more detail in Sect. 3, the results are often presented in ways that are not helpful for interdisciplinary research and are often within individual standards and guidelines that are not easily traceable online. Alternatively, they can be presented formally for data curation (machine-readable) purposes using an ontological structure that is unfamiliar to many Earth scientists and that in many cases does not yield human-readable definitions. Where online documents exist, they are difficult to cite, especially at item level, and often lack a formal persistent identifier. This variety of perspective and presentation of the vocabularies reduces the usability of such resources, making it difficult to obtain consistency between communities, or even within one community.

To address these difficulties within satellite Earth Observation, a terminology task force was set up by the Committee on Earth Observation Satellites (CEOS) (https://ceos.org/). CEOS is the space arm of the Intergovernmental Group on Earth Observations (GEO) (https://www.earthobservations.org) and is the primary forum for international coordination of space-based Earth observations. Its members—mostly national and international space agencies—cooperate in multi-agency initiatives with a strong focus on harmonisation and interoperability. CEOS works towards the GEO vision of a Global Earth Observation System of Systems to 'better integrate observing systems and share data by connecting existing infrastructures using common standards' (https://www.earthobservations.org).

The CEOS terminology task force performed its activities between 2021 and 2023. Its aim was to investigate the issue of terminology primarily from the perspective of an interested user, who is not necessarily an expert in all the fields touched by CEOS work and is perhaps engaged in multidisciplinary collaboration with scientists from other fields. Such users look to terminologies not only to understand how words relate, but also to learn about scientific concepts from the definitions (for example, to understand what a term such as 'interoperability' or 'metrological traceability' means). They are interested in understanding how the different concepts involved are named, explained, and connected to each other.

This paper describes the output of the work of the task group, which was conducted in three phases that define the structure of this paper. Section 2, as a methods section, presents how the task group reviewed existing vocabularies. Section 3, as a results section, describes the identified strengths and weaknesses of existing vocabularies and consistencies and inconsistencies between vocabularies. Section 4, as a discussion section, makes recommendations for establishing a consistent, common vocabulary. Finally, Sect. 5, as a conclusion, summarises the proposed criteria of a good vocabulary. While our examples are based on the Earth Observation topics covered by CEOS, we are convinced that the concepts and approaches discussed here have a broader applicability in the Earth sciences and should therefore be brought to the attention of a wider audience beyond just CEOS.

2 Method: Investigation of Existing Vocabularies

2.1 Definition of Terms: Vocabularies, Terminologies, and Thesauri

Several expressions are in use to describe the concept of a collection of terms that may or may not include definitions. In everyday English, the term 'thesaurus' is used for a list of synonyms, while a 'dictionary' also includes definitions and may include information on usage, pronunciation, and etymology. The terms 'glossary' and 'terminology' are used for word lists that are specialised to specific technical fields.

In information retrieval communities, however, these everyday words are used more specifically. A 'structured vocabulary' is an 'organized set of terms, headings or codes representing concepts and their inter-relationships, which can be used to support information retrieval' (ISO 25964-1 2011) (where the underlined printed words are also defined), while a 'controlled vocabulary' is more simply a 'prescribed list of terms, headings or codes, each representing a concept'. The term 'thesaurus' is defined very specifically as a 'controlled and structured vocabulary in which concepts are represented by terms, organised so that relationships between concepts are made explicit, and preferred terms are accompanied by lead-in entries for synonyms or quasi-synonyms'. Finally, and using the definition on Wikipedia (https://en.wikipedia.org/wiki/Ontology), an 'ontology' is a way of showing the properties of a subject area and how they are related, by defining a set of terms and relational expressions that represent the entities in that subject area.

Within this paper, we will use the term 'vocabulary', which is the most generic one and thus will not limit the scope of the present discussion. In Sect. 4.2, we discuss the value of a thesaurus, according to the information retrieval definition. However, the collections of terms that were investigated are best described as 'vocabularies'; in several cases, they fit the definition of a 'controlled vocabulary', 'glossary', or also 'terminology'.

2.2 Collating and Reviewing Existing Vocabularies

The initial objective of the CEOS terminology task force was to collate and combine the concepts and definitions in use in the various CEOS subgroups and member agencies. Most of these vocabularies are available on openly accessible sites, obtained through an internet search. Some are based on wider community glossaries published formally as standards. Others were private and made available to us by the working group responsible for them. All the vocabularies considered were in English.

A key authority providing a standard terminology with definitions is ISO/TC 211, a technical committee within the International Organization for Standardisation (ISO) that plays a crucial role in developing standards for geographic information and geomatics (www.iso.org/committee/54904.html). The standards developed by ISO/TC 211 are used by a wide range of industries and sectors, including environmental management, navigation, infrastructure planning and emergency services, among others. A pertinent example for such a standard is ISO 19156 entitled 'Observations, Measurements and Samples' (ISO 19156 2023). These standards are critical for facilitating global data sharing and integration. While each standard comes with its own 'terms and definitions' section, many terms are being re-used in several standards and ISO/TC 211 established a collective online vocabulary of common terms called the 'Geolexica' (https://isotc211.geolexica.org/concepts). Unlike the full standard documents, the 'terms and definitions' and the Geolexica are freely available.

ISO/TC 211 is complemented by the Open Geospatial Consortium (OGC), an international industry consortium that develops free and publicly available geospatial standards to ensure interoperability in geo information systems (www.ogc.org). Unlike ISO/TC 211, which is a formal standards organisation that develops standards through a consensus process among national authorities, OGC's standards are developed through a consensus process among industry, government, and academic members, often resulting in more rapid standards development. OGC frequently collaborates with ISO/TC 211 to ensure compatibility and sometimes joint adoption of standards. OGC maintains a working group

Table 1 Vocabularies collated in this study (not an exhaustive list of available vocabularies)

Vocabulary title	Source	Availability/link	Content
CEOS WGCV	CEOS Working Group on Calibration and Validation (WGCV)	Informal material was already collated on http://calvalportal.ceos.org/t-d_wiki (supplemented in this work)	Online collection of terms and definitions in the scope of Earth Observation, collated from various resources
EO Data Stewardship Glossary	CEOS Working Group on Information Systems and Services (WGISS)	http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/EO-DataStewardshipGlossary.pdf	Document defining acronyms and terms related to Earth Observation data stewardship, collated from various sources, and primarily aimed at aligning definitions between space agencies
Geolexica	ISO/TC 211 'terminology management group'	https://isotc211.geolexica.org/concepts/	Multi-lingual glossary of terms related to geographic information and geomatics, collating terms from the different underlying standards
I-Adopt	Research Data Alliance's RDA Vocabulary Services Interest Group (VSSIG)	https://w3id.org/iadopt/ont/1.0.3	Ontology designed to facilitate interoperability between existing vocabularies. Produced by I-ADOPT working group
IEC Electropedia	International Electrotechnical Commission (IEC)	www.electropedia.org/	Multi-lingual comprehensive online terminology database related to 'electrotechnology'. Used for 'base terms'
Inspire glossary	European Union Directive for an Infrastructure for Spatial Information in Europe (INSPIRE)	http://inspire.ec.europa.eu/glossary	Terms and definitions that specify the common terminology used in the European INSPIRE Directive and in the corresponding Implementing Rules
NASA Earth Observatory Vocabulary	NASA	http://earthobservatory.nasa.gov/glossary	Online glossary of terms related to Earth Observation
NASA GCMD Keywords	NASA Global Change Master Directory (GCMD) Keywords	http://wiki.earthdata.nasa.gov/display/CMR/NASA+GCMD+Keywords	Searchable hierarchical set of controlled Earth Science vocabularies
NERC Vocabulary Server	UK Natural Environment Research Council (NERC) Environmental Data Service, managed by the British Oceanographic Data Centre's National Oceanography Centre (NOC)	http://vocab.nerc.ac.uk/	Central server providing access to standardised and ontologically organised vocabularies targeted at the marine science community

Table 1 (continued)

Vocabulary title	Source	Availability/link	Content
NESDIS Data Management Lexicon and Related Terms	National Environmental Satellite, Data, and Information Service (NESDIS) of the National Oceanic and Atmospheric Administration (NOAA)	Not available online	Document with terms and definitions associated with data management activities aimed at harmonising understanding and acceptance within NOAA
OGC Rainbow	Open Space Geospatial Consortium (OGC)	www.opengis.net/def	Interoperable glossary providing definitions in human and machine-readable form, cross-linking terms, visualising connections in knowledge graphs, and intended to serve as a node in a broader ecosystem of resources
SWEET	Semantic Web for Earth and Environment Technology under governance of Earth Science Information Partners (ESIP)	https://github.com/ESIPFed/sweet	A foundational ontology that contains over 6000 concepts organised in 200 linked ontologies in the Earth and Environmental sciences
VIM: International Vocabulary of Metrology	Joint Committee for Guides in Metrology (JCGM)	www.bipm.org/en/committees/jc/jcgm/publications	Document in English and French for a system of basic and general concepts used in metrology, together with concept diagrams to demonstrate their relationships. Online version with linked terms also available
WIGOS Metadata Representation (CODES repository)	World Meteorological Organization (WMO) Integrated Global Observing System	http://codes.wmo.int/wmdr	Controlled list of terms defining key concepts relating to Earth sciences in observation metadata

responsible for terminology coordination called the 'OGC naming authority' which also operates the OGC online vocabularies (OGC Rainbow) (http://defs.opengis.net/vocprez/vocab).

The set of vocabularies reviewed as part of this study are listed in Table 1.

2.3 Guidelines of *ISO* 25964

Additionally, we reviewed guidelines provided in the ISO standard on 'Information and documentation – Thesauri and interoperability with other vocabularies, Part 1: Thesauri for information retrieval' (ISO 25964-1 2011) and 'Part 2: Interoperability with other vocabularies' (ISO 25964–2 2013). These standards contain comprehensive recommendations related to equivalence of terms and concepts, relationships between terms, as well as thesaurus/vocabulary implementation and management including machine readability among many others. They address and formalise several critical issues we have identified in our vocabulary reviews. Hence, the principles for developing a good vocabulary we propose in this paper takes inspiration in part from this ISO standard.

2.4 Discussions with the Relevant Communities

An important part of this study was the formal and informal presentation of these issues to scientists working in the Earth Observation and standards communities. As our emphasis was on explanatory and educational aspects of vocabulary where we identified the largest deficits, and the terms we could tackle in that respect were rather limited, we refrained from engaging with the ontological community, which we found focussing rather on machine readability and relational aspects of far more complex sets of terms. Formal presentations were made to the working groups listed in Table 2 and, in most cases, these also were followed with discussions with those groups. Additionally, our proposal was discussed with individual scientists working in a wide variety of relevant Earth science fields and an early draft of this paper was shared with twelve individuals with a variety of expertise from different disciplines within the Earth sciences, ranging from climate and atmosphere to land and water applications, satellite operations, metrology and geospatial standardisation, who provided detailed comments on the concepts considered here. The term 'respondent' is used in this paper for anyone who responded to these discussions: members of working groups, conversation partners in informal and formal discussions or individuals who provided detailed comments.

3 Results of Review of Existing Vocabularies

3.1 Usability of Existing Vocabularies

Many of the vocabularies listed in Table 1 would not meet the definition of 'structured vocabulary' as they are presented as simple lists, often in an alphabetical order in a document, with no cross-links between definitions. Some, such as the Geolexica (https://isotc211.geolexica.org/concepts), are presented as online lists with search features and links to the source of the definition, and some provided hierarchical or ontological relationships

Table 2 Presentations made to relevant working groups. Links provided where presentations are publicly available

Date	To	Title/hyperlink
June 2021	CEOS WGCV #49.	CEOS Terminology and online glossary
October 2021	CEOS WGISS #52	CEOS Common terminology
November 2021	CEOS Plenary	CEOS Common terminology (single slide within CEOS WGCV report)
December 2021	OGC Naming Authority 121st Member Meeting	Harmonising Geospatial Glossaries. Do we know what we are talking about?
February 2022	ESA data access and preservation working group meeting #9	CEOS Common Terminology Initiative
March 2022	CEOS WGISS #53	CEOS Common Online Dictionary Initiative
May 2022	Living Planet Symposium	Enabling interoperability across cloud based EO platforms: Open standards and protocols
June 2022	ESA DAP-WG Meeting #10	CEOS Common terminology initiative
August 2022	CEOS LSI-VC #11	An (LSI) Interoperability Framework for CEOS
August 2022	Task Team on WIGOS Metadata (TT-WIGOSMD)	CEOS Common Terminology Initiative
October 2022	WGISS-54/WGCV-51	CEOS Interoperability Framework Initiative
October 2022	ESA DAP-WG Meeting #11	CEOS Common online dictionary

(see Sect. 3.2). Several vocabularies provided cross-references to other vocabularies, where term definitions had been copied from. In almost all cases, these cross-references were presented as a static (one way) reference and would not be able to account for changes in the source vocabulary definition.

Most of the vocabularies provided version control at the vocabulary level, that is, they provided a version number for the complete set of definitions. No access was provided to earlier versions, and some of our conversations with scientists working in these fields had described confusions when they did not realise that they were using an outdated definition. This effect is particularly problematic with derivative materials such as training courses, scientific papers and other materials quoting old definitions. As an example, the VIM has redefined the meaning of 'uncertainty' in each of its versions; this is arguably one of the most important words in a metrology vocabulary, so these changes are significant, but this change is not immediately obvious.

The vocabulary-level versioning provides some insight into the workflow for creating and maintaining such vocabularies. In general, except for ISO/TC 211, these committees tend also to be run by groups of technical experts, rather than thesaurus experts. These small teams collate terms from other vocabularies and add new terms where previous definitions do not exist. Participating in such working groups is often not regarded as a research grade task and receives little appreciation, which tends to limit the number of scientists who actively engage in such processes. Sometimes, these working groups may alter the wording of definitions from previous versions of the vocabulary, or that have been adopted from other vocabularies. During this process, many discussions are held, but the final published vocabulary does not usually represent the level of discussion involved in its creation and remains static until the next revision. Exceptionally ISO/TC 211 does make such discussions public (https://github.com/opengeospatial/om-swg/issues/175) but it does not link them to the published definitions in the Geolexica.

The ISO Online Browsing Platform (https://ttbs.isolutions.iso.org/obp/ui/) collates terms from many ISO standards even beyond ISO/TC 211. Standards' development groups are encouraged to re-use, in new and revised standards, definitions that have already been agreed. However, and as an example, the 77 results from a search for the term 'observation' show the legacy of older approaches.

3.2 Structure of Existing Vocabularies

Most of the vocabularies that we studied were presented as lists of definitions arranged alphabetically by term in vocabulary documents. This presentation is also common within other documents – for example, in formal standards documents from organisations such as ISO, or in good practice guidance documents developed within technical committees, there is often an alphabetical terminology section at the beginning of the document. ISO has provided a tool for searching for the definitions within its documents, but these approaches require active searching of a term – they are hard to browse and explore when beginning an interdisciplinary collaboration.

There were, however, three alternative approaches to linking terms. The online version (JCGM 2012) of the third edition of the International Vocabulary of Metrology (VIM; (BIPM et al. 2012)) is the only vocabulary that we studied that provided cross-links to words used within the definition, i.e. it highlighted words in a definition that the vocabulary also defines.

The GCMD provided terms in a contextual hierarchy. For example, 'freeboard' is a sub-term to 'sea ice', which is a sub-term to 'cryosphere'. This hierarchy requires a human interpretation. Freeboard (the thickness of ice protruding above the water) is a property of sea ice, and sea ice exists in the cryosphere; however, the relationship is not the same as that for the VIM (the definition of freeboard does not require the use of the words sea ice and cryosphere), and it requires interpretation to provide formal links of an 'is_a' or 'has_a' form. Such links are provided in ontologies, which describe the relationships between terms in the form of a 'graph'.

Parsons et al. (2022a, b) provides an 'instructive tale' of the development of the GCMD keywords. They describe how an attempt by NASA to formalise a catalogue of keywords led to a de facto standard whose uses went far further than the original purpose—with both benefits and challenges. One of the issues identified by Parsons et al. (2022a, b) is the hierarchy of the original GCMD. Different users presented different parts of the hierarchy, and it was hard to link them. As other vocabularies developed independently, mapping between them became a challenge. More significantly, the field of informatics was moving towards the concept of 'linked data' and the semantic-web-based approach.

Today, terms are connected in ontologies that show formal relationships. The Semantic Web for Earth and Environment Technology (SWEET) ontology (Raskin and Pan 2005), under the governance of the Earth Science Information Partners (ESIP) foundation, built up from the GCMD with a graph-based hierarchy. In our study, however, we found SWEET difficult to explore as a human-reader. In many cases, the terms were provided without definitions (an issue that is also present in the GCMD). Parsons et al. (2022a, b) explains that SWEET has 'never been broadly adopted in Earth science data systems'. They explain that this is because of the difficulty of translating the hierarchical GCMD into a graph-based ontology, and because of the way the structure was established without iterative close collaboration with the intended users (see Sect. 4.5).

While SWEET focuses on the terms of the Earth and environmental sciences, another ontology (I-ADOPT) developed a graph-based structure for the more underpinning concepts relating to observations: terms such as 'entity', 'phenomenon' and 'property'. The 'InteroperAble Description of Observable Property Terminologies' (I-ADOPT) Working Group provided a set of definitions and recommendations in 2022 (Magagna et al. 2022) under the auspices of the Research Data Alliance's RDA Vocabulary Services Interest Group (VSSIG). The terms within this have definitions considered 'Aristotelian definitions' (Arp et al. 2015), in that the definition is of the form: 'a G that is D' (where G is the immediate parent in the graph-structure, and D is the differentiating property). This type of definition supports an explanation of the ontology, and works well within a field, but, in our opinion, is not very enlightening in supporting interdisciplinary research.

Our focus was therefore on those vocabularies that provided definitions developed for human-readability and to support humans in understanding the concept. Such vocabularies were the ones usually presented as alphabetical lists. The vocabularies that were reviewed in this study had carefully considered definitions of important terms. There were clear efforts to link the various vocabularies to each other (see Table 1), with terms adopted or adapted from one community to another. However, there were also some inconsistencies between, and in some cases, within individual vocabularies. These are discussed in the following section.

3.3 Example Terms Showing Definitional Inconsistencies

The definitions in the vocabularies considered included inconsistent use of what we will call ‘base terms’ in Sect. 4.3. That is, fundamental concepts such as ‘property’ are used in various definitions in diverse ways, and sometimes as a ‘circular definition’, where a term used in the definition of one term, uses that term in its own definition. Between vocabularies there were also examples where different communities defined the same concept in contradictory ways; and examples of broad terms that covered different concepts for each community and so had an ambiguous definition.

As an example of a circular definition, consider the Geolexica (https://isotc211.geolexica.org/concepts). Here, the term ‘property’ is defined as a ‘facet or attribute of an object referenced by a name’, while ‘attribute’ is defined as a ‘named property of an entity’. Such circular definitions are predominantly found in ‘base terms’ rather than in the specialist terms used in a particular scientific field and have arisen from the way vocabularies such as the Geolexica have collated separately defined vocabulary lists. Within the Geolexica, the definition for ‘property’ comes from ISO 19143:2010 (ISO 19143 2010) and the definition of ‘attribute’ from ISO/IEC 2382:2015 (ISO/IEC 2382 2015). Inconsistencies in these underpinning vocabularies can thus create circular and/or contradictory definitions in vocabularies derived from them.

Divergent meanings are found, for example, for the term ‘sample’, which is interpreted differently by the field measurement community, where a scientist would ‘sample’ the physical surface (such as soil, rock, water), and collect ‘a sample’ to take back to the laboratory for further analysis, and by the remote sensing community, where a continuous phenomenon is ‘sampled’ by discrete individual measurements (‘samples’). Note that the use of ‘sample’ as a verb and noun is like the use of ‘measurement’ to describe both a process and the output of that process.

In this section, we will focus on three terms in more detail: ‘observation’, ‘in-situ’, and ‘interoperable’. These terms are given as examples, and other terms could have been chosen. We selected them from a list of problematic terms that were identified in our review of vocabularies. Each of these terms has difficulties in its definition and is presented as an example of a common problem in existing vocabularies. ‘Observation’ is a term that is defined ambiguously. Different communities use this word in separate ways, and the distinctions show up epistemological variations between those communities. ‘In-situ’ is a more straightforward term but is also used in contradictory ways by separate communities. The term ‘interoperability’ is an increasingly important qualitative concept in use in a wide range of communities but is so broad that it acts more like an overarching paradigm, and it is difficult to define in a practical way.

3.3.1 Observation

In the Earth sciences, there is no term more fundamental than ‘observation’, and yet the difficulty in properly defining this concept starts with how ‘observation’ is ambiguously used both to describe a process and to describe its result. The recent (2023) revision of the ISO 19156 standard (ISO 19156 2023) (identical to OGC Abstract Specification Topic 20 (OGC 2023)) resolves this issue by reserving ‘observation’ for the act and introducing ‘observation result’ for its outcome. In this, it mirrors the JCGM VIM that distinguishes ‘measurement’ and ‘measurement result’ in the same way. After defining ‘observation’ as an ‘act carried out by an observer to determine the value of an observable property…’, ISO

19156 goes on to define an observer as ‘an instance of a sensor, instrument, implementation of an algorithm or a being such as a person’. In doing so, the same concepts apply to observations, simulations, and even opinions, rendering them functionally compatible but circumventing a clear distinction between them. Distinguishing sensor observations from modelling would therefore require additional criteria that do not have explicitly defined terms in the standard. A similar decision was made by the I-ADOPT team (Magagna et al. 2022) where ‘observed quantities’ are described by ‘i.e. measured, simulated, counted quantities, or qualitative observations’.

In the philosophy of science, the term ‘observation’ was originally linked to a human sensory response (see discussion in (Boyd and Bogen 2021)). In the very early days of science, philosophers felt that what they could themselves ‘observe’ provided a superior knowledge to other kinds of knowledge. The invention of telescopes and microscopes to augment what the human sensory system could observe, and of instruments with dials and scales that could make quantitative measurements less subjective, changed such a viewpoint. One of our respondents, who worked in meteorology, still considered ‘observations’ linked to human perception (e.g. human estimates of oktas of cloud cover) in opposition to instrumented measurements, although now such observations were considered less robust than those of instruments. These different meanings of ‘observation’ show not only a distinction in vocabulary, but also in an epistemological framework.

Today, imaging instruments go further than the telescope or microscope and often require modelling in deriving a value for the intended measurand. Even a simple thermometer does not measure air temperature but measures the expansion of a liquid in a tube, or the change in resistance of a platinum wire. Physical models mathematically convert those measured values to the temperature of the thermometer, and then further physical models are often needed to relate the temperature of the thermometer accurately to the temperature of the air (Podesta et al. 2018; WMO-No.8, 2021).

Satellite Earth observations often go further still. For example, in satellite measurements of sea surface temperature, a large proportion of the measured top-of-atmosphere radiance comes not from the sea, but from the atmosphere. Further models, often making assumptions about the nature of the atmosphere and relying on measurements in several spectral bands, are used to retrieve the sea surface temperature from the top-of-atmosphere signal. In our discussions, some respondents working in space agencies wanted to distinguish ‘measurement’ from ‘observation’. These experts, wanted to use ‘measurement’ for what they call the ‘level 1 product’: the physical quantity measured by the satellite (for the sea surface temperature example this level 1 product is the ‘top-of-atmosphere radiance’). They left ‘observation’ for more highly processed data that provides retrieved quantities such as sea surface temperature (usually called ‘higher level products’). This distinction is difficult even within the space-based observation communities because the ‘level 1/level 2’ concept is used differently for active (e.g. radar) and passive (e.g. radiometric) sensors (https://copernicus.eu/user-guides/sentinel-3-altimetry/processing-levels), (Strobl 2023; Weaver 2014). This discussion makes clear that modern ‘observations’ involve considerable data processing. However, in all these cases, the origin of the data is in some objective phenomenon in the real world, some property of which is determined by a sensor.

A scientific philosopher, Bokulich (Bokulich 2020) has outlined a taxonomy of various ways in which data can be ‘model-laden’ (also known by philosophers as ‘theory-laden’) to increase their usefulness. She reserves the term ‘synthetic data’, in contrast with ‘real data’, to virtual or simulated data that are not produced by physical interaction with worldly phenomenon. For ‘real data’, she defines six categories: data conversion, data correction, data interpolation, data scaling, data fusion and data assimilation. Data conversion is the use of a physical instrument model to relate a raw signal (analogue current or digital counts) into a physical quantity. Data 'correction' includes processes that correct for environmental factors, natural variability, seasonal effects, dark readings, etc. From a metrological perspective, these two together would be part of establishing a measurement model (BIPM et al. 2020), which may also include 'data scaling' from Bokulich's taxonomy. Data scaling, interpolation and fusion use models to scale between the spatial or temporal scales of separate observations or between observations and models, fill in (impute) missing data or to combine separate data sets. Such processes are common in generating 'level 4' data in radiometric satellite products, which often include regridding and filling missing data. 'Data assimilation' is used for both historical reconstructions of meteorological conditions (e.g. reanalyses) and in short term forecasting of weather. Bokulich's taxonomy covers what can be considered a smooth scale from the 'purest' measurements to data assimilation, with increasing use of modelling to interpret the measured results. It does, however, have a gap in that scale between those processes and synthetic data.

Given the complexity, it is perhaps unsurprising that the ISO 19156 working group, in developing its vocabulary, reached a compromise that did not make a formal distinction between types of 'observers' and included algorithms as observers. From a functional perspective such a blending is desirable to achieve compatibility of results of purely algorithm-derived simulations with those of sensor-based observations. However, from literature and personal communication with a wide range of people involved in measurements, their processing and in predictive modelling, we conclude that many respondents do want to see 'simulation' and 'observation' clearly separated. The distinction between what originates from interacting with an objective part of reality and what is born solely by an algorithm is, in our opinion, fundamental to science at large, even if modern techniques make it in certain cases difficult to define and distinguish.

In Sect. 4.3, we propose how by considering 'observation' as a 'controversial term', we can bring such distinctions much more clearly into the consciousness of scientists using the term and help improve the descriptions used by all the relevant communities.

3.3.2 In-Situ and its Contrasts

Another example of a confusing definition in Earth sciences is 'in-situ' and how it is used across domains. It is primarily intended to delineate a subgroup of observations (or measurements) by highlighting a specific aspect of them. 'In situ' is a Latin phrase meaning 'in the place' and has been used since the nineteenth century in many contexts to refer to an observation of a phenomenon in its 'natural' environment without any disturbances. At the time, the fact that the observation had to take place where the phenomenon occurred usually implied that the observer would travel to the 'field' to study the phenomenon. Originally, it was applied with geological, archaeological and biological fieldwork but meanwhile it has been used to describe a wide range of observational activities.

The term was chosen to contrast to 'ex-situ' methods, which would denote observations of a phenomenon outside (or 'away from') its natural environment. In the original scientific domains, an ex-situ method would usually entail removing a specimen from its 'natural' place and relocating it for analysis or preservation in a laboratory or archive. Because the removal, transfer and storage of the specimen could have an impact on the properties that were to be observed, in these early days, 'in-situ' methods were often regarded more 'original' and 'truthful' compared to 'ex-situ' ones. Extensive resources were mobilised to allow respective fieldwork in the frame of large-scale expeditions all around the globe. This way in-situ and '(in-) fieldwork' became largely synonymous with high quality and reliability (trustworthiness). However, in an analogous way to the changing priority of 'observation' relating to human senses, changing experimental techniques and the availability of more capable instruments in a laboratory has partially changed this value judgement.

When remote sensing technologies became available which could observe objects from a large distance, including from satellites in orbit, the gold standard for verifying the results remained the 'in-situ' fieldwork which was and is indispensable for calibration and validation of remote sensing instruments (Slater et al. 1996). Early remote sensing communities went as far as calling these observations 'ground-truth', expressing the authenticity that was assumed to be associated with this approach, in part because the ground instruments, unlike the satellite instruments, could be brought back to a laboratory for recalibration. While the term 'ground-truth' should be retired for the reasons described in (Woodhouse 2021) (mostly because they are neither always 'ground', nor can be considered 'truth'), the concept remains relevant. Large parts of the remote sensing community use the term 'in situ' or 'in-situ' for any high-fidelity data collected close to the surface ('ground-based'). The use of 'in-situ' has, in turn, led to two new interpretations of the term, one as measuring 'at the actual location' (NASA, https://earthobservatory.nasa.gov/glossary) without a significant distance to the phenomenon and one, more instrument orientated, as 'any sub-orbital measurement' (CEOS, https://calvalportal.ceos.org/t-d_wiki). In both these interpretations, 'in-situ' is meant as opposite to 'remote' or 'satellite', so that an observation could be either one or the other.

It is worth noting that new terms such as 'fiducial reference measurements' (Goryl et al. 2023) are now being used to represent a subset of (mostly but not exclusively) non-satellite observations (previously 'in-situ') that are used for satellite calibration and validation and that are considered of significantly lower uncertainty due to regular SI-traceable calibrations and robust uncertainty assessments.

Authors using the concept 'in-situ' are anticipating that their audience has a shared understanding of the term. This assumption becomes critical when the term is not used just in a vague (common language) way, but also implies the inclusion or exclusion of certain types of observations, e.g. in cataloguing, for specifying metadata, or even for funding. In all cases where a distinction matters, it should be made clear which is the distinctive criterion (location, distance, quality) and name it, rather than using the unqualified term 'in-situ'. This suggestion is considered further in Sect. 4.4.

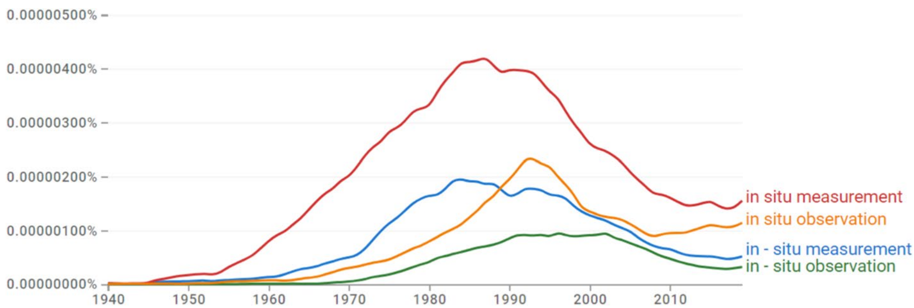

Fig. 1 Google Ngram Viewer (https://books.google.com/ngrams/) for use of different forms of 'in-situ' or 'in situ' over time. The *x*-axis is the year of publication, the *y*-axis the percentage of publications using the term

Finally, further complexity (particularly for machine readability) arises because of the mixed spelling of 'in situ' and 'in-situ', and the choice whether to italicise the term, which depends on if it is considered a Latin phrase or a scientific English term. Figure 1 shows the results of the Google Books Ngram, showing the frequency of 'in-situ' and 'in situ'. The mixed choice makes it more difficult for search engines to identify it uniquely as a keyword in large text repositories. Indeed, for searching terms in large text repositories, there are potential advantages of 'in-situ' precisely because it keeps the two parts together for the search.

3.3.3 Interoperability

The word 'interoperability' has gained widespread use in recent years as a desirable attribute enabling data sets from various sources to be combined and compared. The term is used to cover a broad set of concepts. In data curation, it is one of the key principles of FAIR (Wilkinson et al. 2016) data (the acronym represents data which meet principles of findability, accessibility, interoperability, and reusability, where those terms are defined in the referenced paper). In FAIR, 'interoperability' is defined in terms of a common and formal standard format for data and metadata. In Earth Observation, it can also include the use of a common geospatial grid and reference for data sets, the provision of coefficients to correct biases between data sets, and the homogenisation of data from one satellite sensor so that data from it can be used seamlessly with that from another sensor. In the information technology (IT) world, it expresses the compatibility of different components of hardware and the protocols allowing them to communicate with each other and with humans. Data centres define data sets as 'interoperable' when they exploit common protocols and metadata to provide the capability to query and use data from any source without the need to duplicate files physically. In other applications, there may be additional legal, semantic and organisational aspects that may be decisive for the interoperability of entities or their outputs, such as the use of interface documents.

At its core, 'interoperability' is about enabling entities to work together seamlessly; a very high-level definition given, at the time of writing, on Wikipedia is that 'interoperability is a characteristic of a product or system to work with other products or systems' (Wikipedia interoperability 2023). While this definition applies across the broad range of meanings of 'interoperability', it is not particularly helpful because almost all communities have something more specific in mind when they use the term. For this reason, rather than a definition of 'interoperability', communities need an interoperability framework that defines more specifically which factors they consider relevant in this context.

Such a framework would need to answer some key questions. The first is whether, for that community, 'interoperability' is about homogenising data sets to match each other or a common reference, or about providing the information needed to enable different data sets to be corrected to a common reference. Here, the concept 'homogenisation' is used as it was in the FIDUCEO project glossary (https://research.reading.ac.uk/fiduceo/glossary) as a way of making data sets look the same so that they can be used together without the user having to consider the origin of the data. For example, for those involved in 'data curation', interoperability can relate to accounting for data sets having different data formats. Ensuring interoperability could involve describing those two data formats so that users can produce appropriate readers (software packages to input data in each format and bring them into code in a common way). Alternatively, interoperability as homogenisation would involve reformatting one or both data sets to a common data format. Similarly, if a community's understanding of 'interoperability' includes aspects relating to the spatial collocation of data, that community could interpret this in two ways. One way would be by carefully documenting the spatial representation and reference used by each data set so that users can undertake what is necessary to superimpose them. Alternatively, interoperability could require one or both datasets to be (re-)sampled into a common spatial reference scheme such as a grid (homogenisation). For radiometric satellite sensors, operating in spectral bands, interoperability could include carefully documenting each satellite's spectral response function. Alternatively, it could involve applying (homogenisation) 'spectral band adjustment factors' (Chander et al. 2013) that convert one satellite's observations as though they were taken by the other.

The second question to address in such a framework relates to which factors of interoperability are to be considered. For example, the interoperability of visible and short-wave infrared optical satellite sensors can consider interoperability to include spatial, geometric, temporal, spectral, angular and radiometric aspects. In all these dimensions, the sensor 'samples' a phenomenon by discrete observations (see Sect. 3.3 for a discussion of 'sample', here used in a satellite observation sense). Compatibility in this 'sampling' is a major ingredient to interoperability of the respective data. Other aspects could include data formats, metadata definitions, variable and quantity names, choice of units (for example whether wavelengths are defined in nanometres or microns) and other references (e.g. solar spectra), in addition to the provision of a common set of uncertainty information presented in a consistent format. CEOS has recently embarked on defining for itself an interoperability framework in which many of the above factors are reflected (WGISS 2008). An already established, albeit more generic instance, is the European Open Science Cloud (EOSC) interoperability framework (DG-RI (EC) et al., 2021).

A third question to consider in any interoperability framework is whether 'interoperability' is considered a Boolean (data sets are or are not 'interoperable'), or on a qualitative or quantitative scale. A good example of a qualitative scale is that provided by the EDAP framework (https://earth.esa.int/eogateway/activities/edap) of the Earthnet Data Assessment Project, which is about creating quality metrics for satellite data products, and which builds on earlier maturity matrices for example (Peng et al. 2015; WGISS 2023). The EDAP framework grades a dataset, through a maturity matrix, on the nature of the quality assurance considered, the quality of the uncertainty evaluation and the robustness of the calibration. A quantitative scale, would provide a 'degree of interoperability' of two datasets, perhaps calculated by quantifying differences between the datasets in the various aspects of interest.

The discussion in this section has been about the word 'interoperability'. There are other prominent examples of terms that are like this and need a similar treatment. Concepts such as 'continuity' (e.g. between satellite missions in a series) share many of the same issues.

4 Discussion

4.1 Overview

In the sections above, we have reviewed existing vocabularies and their advantages and disadvantages in terms of structure, content, internal consistency, consistency with other vocabularies, usability by people and machines, and sustainability. All discussion points would merit a deeper consideration than is possible within an overview paper, and we recognise that in some cases, our views as authors may be considered controversial. We present the discussion below to indicate a base for a future vocabulary and to encourage debate within the communities working in the Earth sciences. The first-person pronoun ‘we’ is used to emphasise where viewpoints belong to the authors.

Here, we focus on a vocabulary as a means to support interdisciplinary collaboration between people by helping people become aware of when they may be using the same word in different ways, different words in the same way, and to understand the specialist vocabulary of a new community.

This discussion considers four aspects of what we think is needed. First, in terms of **approach**, it considers why a formal ‘thesaurus’ (using the term as defined in Sect. 2.1) is required, and the benefits of a single thesaurus to be used by the various communities involved in the interdisciplinary Earth sciences. Second, in terms of **structure**, a possible structure and categorisation of the terms is presented, describing what we call ‘base’, ‘core’, ‘controversial’ and ‘high-impact’ terms. Third, in terms of **content**, we comment on the terms that were used as examples in Sect. 3.3 and how the issues identified there may be resolved. Finally, in terms of **practicality**, we suggest the practical steps required to start working towards a new approach.

In this discussion, we are well aware of the well-known joke that in some community there were originally 12 vocabularies, so a group set out to make one combined vocabulary, and now there are 13 vocabularies. We recognise the danger that any new efforts may increase rather than resolve the confusion. We also recognise the concerns that many of our respondents expressed that vocabulary conversations can quickly become entrenched in never-ending discussions between opposing views. Some people we spoke with described work on vocabularies as both ‘pointless’ and ‘painful’ because of such entrenchments. Furthermore, some people we have engaged with have suggested that what we are trying to do is ‘impossible’. We understand all, and sympathise with, these concerns. However, we also believe that the current approach is not suitable for modern interdisciplinary research in the Earth sciences.

It is also important to acknowledge the changing nature of scientific research. The use of computing clouds to store and share data is almost ubiquitous in the scientific community today and yet the shift from paper documents to online documents, and then to active cloud-based repositories is still relatively recent and scientific work processes have not caught up with these changes. Concerns about reproducibility (Baker 2016) have led to journals requiring far more information than was traditionally required, with online data repositories, extensive metadata and in some cases formally structured method statements (Nature editorial 2018). Standards, good practice guides and similar documents are now much more easily accessed online than on paper. These changes exaggerate the impact of the problems that we identified in Sect. 2. Yet, they also provide an opportunity for doing things another way.

In attempting to suggest a new approach to a thesaurus, we realise that our suggestion will be incomplete, flawed and perhaps in places, naïve. We make these suggestions anyway, and encourage readers to engage with these ideas, and bring their own perspectives and expertise to this discussion.

4.2 Approach to a Thesaurus

As discussed in Sect. 3.1, many of the issues we identified with existing vocabularies come from the ways in which most were developed by relatively small, relatively isolated groups of scientists often with limited knowledge of formal vocabulary development techniques.

The only promising way out of the problem of isolated vocabularies is convergence across all stakeholders, ideally through one single entry point following well-established rules (albeit allowing for disambiguation pages to enable different definitions of the same term to both coexist and be linked). While achieving such a collaboration between all the organisations currently involved in vocabularies for the Earth sciences will be difficult, we believe it is needed (see Sect. 4.5 for a discussion on which institutes this would involve).

The first and most important step would be to convince these partners to abandon individual authority over their vocabularies and to transfer it to a common undertaking. It is, however, important to understand that this is not about relinquishing independence in developing suitable vocabularies for their own needs. As Parsons et al. (2022a, b) describes with the GCMD, while it began as a NASA initiative: 'Maintaining a balance of centralized control and distributed adoption/adaptation is an ongoing effort. Over time, the focus of computing vacillates between centralised and the distributed. Power dynamics are inherent in standardization. Sometimes NASA can lead, even dictate. Sometimes NASA must follow, or at least accommodate, other approaches'. As a model for balancing centralised and distributed approaches, we consider Wikipedia as a good conceptual example, although it is likely that vocabularies for the Earth sciences would need some formal approach to approval and acceptance of terms that is based around the existing committees–acting with full sovereignty in their own domains, while collaborating on common terms and structure.

It is in the interest of all parties that no one organisation gains control over the vocabulary alone. This could be best achieved through a collective effort, supported and jointly governed by the main stakeholders, who at the same time commit to using it for several years as their authoritative source vocabulary. As it would free considerable resources currently spent on the maintenance of decentralised vocabularies, redirecting these to a common approach would not only increase quality and consistency but likely also lead to synergies as work could be distributed and results shared. Moreover, a single point of reference of that sort would attract the wider user community to consult and engage in the discussion, promising a much more accepted and updated repository then any currently in use.

Once the kernel of such an approach is in place, another crucial element would be in how it is used. We propose that rather than documentary standards providing their own lists of references, active links to the centralised definition would reinforce the benefits of centralisation. It may even be desirable for other documents such as community good practice guides and even peer reviewed papers to link to formal definitions. Artificial intelligence tools could alert authors to formal definitions, help them choose between similar terms, and provide automatic linkage for readers. They must be integrated into the platform used to build and host the thesaurus, as well as in formal normative standards development, and ideally made available in editing software such as Microsoft Word or LaTeX.

Of course, creating a single repository with automatic links would have to manage conscious choices of different definitions by separate communities (perhaps through 'disambiguation' options as in Wikipedia, https://en.wikipedia.org/w/index.php?title=Wikipedia:Disambiguation). It would need to follow changes that happen over time and through separate communities formally approving refinements of top-level definitions with details that apply to their field only (such as a 'framework for interoperability' as suggested in

Sect. 3.3.3). Version control at the individual term level, with historical and alternative definitions linked to current definitions would be mandatory so that there is complete transparency of the history of terms, as well as of the perspectives of the different contributing communities. Through such a framework, a document (e.g. a standard, training material or good practice guidance developed by a committee or perhaps even a scientific publication) would link words to the term definition at the time of writing, with a clear note describing whether new definitions have been agreed later. This would follow the types of good practice in software development under Git systems. For terms such as 'interoperability' a top-level definition could be linked to more specific definitions in use by various communities. Expanding the remit of a thesaurus further than its current role, in principle links could be made to framework documents as well as to what are strictly definitions. As with Wikipedia, 'controversial terms' (see next section for how we use this term) could link a currently accepted definition with a behind-the-scenes discussion on the controversies to which any user could contribute.

4.3 Possible Structure for a Thesaurus

We consider that a good thesaurus would clearly show the relations between terms. As discussed in Sect. 3.2, there are currently three different types of links that are created within existing vocabularies: links between terms used in the definitions of other terms, contextual hierarchies, and ontology relationships between terms. For the reasons described earlier, the contextual hierarchy is problematic. However, providing some contextual information may be of considerable value to human users of the thesaurus from an exploratory perspective, and different solutions to this should be considered. In the other cases, however, highlighting terms used in the definition of other terms, and ontology relationships is valuable. We feel that such links should be bidirectional, that is that users would be able to follow definitions both 'up' and 'down' in definitional use or ontology inheritance.

We also believe another type of classification of terms is useful (here, we use the term 'classification' loosely, and do not expect the classification to be unique). The different classes relate to the workflow of the different communities that may develop a thesaurus. The class choice would affect how terms are discussed within committees, and would, in some cases be temporary. These classes are presented in Table 3.

The first set of terms we class are our 'base terms' that are commonly used in the definition of other terms while normally not referring to defined terms themselves. Such terms are also sometimes called 'root terms' (Arp et al. 2015). Only these foundational terms would need careful cross-community agreement and collation to ensure they do not have circular or ambiguous definitions.

Building on such a set of base terms, the 'core terms' would define the core vocabulary in use in the (observational) Earth sciences. It is likely that such a core vocabulary would be built up through different expert teams from the different communities and include terms that have already been collated in existing vocabularies and definitions are likely to need only minor changes to link explanatorily to the base terms avoiding the circular and inconsistent definitions discussed previously. A subset of core terms would need more careful handling. These are terms that are used differently by separate communities, or where there is considerable debate even within one community. In considering these 'controversial terms', a discussion of the various existing uses may be important to include within the thesaurus, along with disambiguation pages.

Table 3 Classes of terms that with representative examples. The classification is based on the workflow of the different communities relating to these terms, and may be temporary or non-unique

	Class of term	Examples	Activity required
	Base terms–terms that underpin all other terms	Data, entity, phenomenon, property	Small committee to redefine in consistent way. Automatic software to identify the use of these terms in existing or new definitions
	Core terms–terms that are basic vocabulary for the Earth sciences, where there are broadly common definitions in existing vocabularies	Uncertainty, leaf area index, data centre, spectral resolution, accuracy, precision, measurement	Small committee to check existing definitions for consistent use of base terms and make any necessary minor tweaks. New core terms identified by standards committees, vocabulary task groups, etc. will be needed to define added terms
	Controversial terms–terms that would be core terms except that different existing vocabularies define them differently or various communities use them in separate ways	Sampling, observation, in-situ, processing level, model, **c**onfidence	Facilitated discussions between appropriate communities to obtain mutually acceptable definitions and/or to agree qualifiers that will distinguish the different uses. Discussion made public and linked to terms
	High-impact terms–terms that can only usefully be defined in comprehensive framework descriptions, which will be highly dependent on application	Interoperability, analysis ready data, data assimilation, real-time, operational	Work in specialist communities to define the framework documents. Thesaurus to link to these documents via a high-level definition

Finally, terms such as 'interoperability', 'fiducial reference measurement', 'essential variable', 'data cube', or 'analysis ready data' require more than a simplistic interpretable definition if they are meant to become more than buzzwords in practical implementation. They are considered 'high impact terms' as a correct formulation will be of high value to the communities. A thesaurus definition that could be agreed upon by all communities would only define these terms in such a high-level way that the definition would be unhelpful when it comes to implementation. We propose two solutions to this. The first is to create, alongside the high-level definition, application-specific definitions that can be more precise in their definition for that community. The second is to link, within the thesaurus, to application-specific framework documents. A thesaurus that links to such frameworks would go beyond the traditional role of a thesaurus, but in a way that is enlightening and useful for the scientific community. Application-specific framework documents can describe in detail what any one community requires for a particular data set to meet its requirements. Examples include the 'CEOS Analysis Ready Data' (CEOS-ARD) framework (https://ceos.org/ard) and the 'CEOS fiducial reference measurement' framework (Goryl et al. 2023).

In Table 3, we provide some representative examples of each class of terms. We also indicate what kind of activity would be required to consolidate existing vocabularies and to move towards the framework described in Sect. 4.2. We expect that a small joint committee of particularly interested experts from the different communities that are working together could check base terms for consistency and check existing core terms for correct use of base terms, taking any corrections back to the committee that made the original definition. The base terms committee's work would be supported by software that could automatically identify the use of such base terms. New core terms would be identified by specialist task groups during the process of developing standards and good practice guides (those that currently list terms in the introduction to their documents). Such groups would develop new definitions with the support of software that automatically linked to other core and base terms or highlighted conflicting definitions. 'Controversial term' is a term where communities use the term in diverse ways. These would be automatically identified in a linked thesaurus, and the task groups would have the option of creating a branched disambiguation definition.

It is desirable for human readers of a thesaurus to orient themselves in that thesaurus through more than one structural view of the thesaurus. Providing links between terms in different ways enables readers to 'explore' the definitions as part of learning about a new field. The classes of term proposed above are created through the extent by which terms are used in the definition of other terms and whether or not they have a unique definition. For 'core terms' and 'controversial terms', it may also be helpful to present them with context-based information, avoiding the hierarchical concerns of vocabularies such as the GCMD, perhaps through tags. A third type of relational structure would originate from connecting terms through an ontology. The six key recommendations of the I-ADOPT working group (Magagna et al. 2022) can form the basis of an ontological structure, providing a machine-actionable format to complement human-readable formats. In practice, it would be helpful to connect terms in all these ways, and to allow users of the thesaurus to choose between separate ways of presenting the structure.

4.4 Initial Considerations on Content Issues

In Sect. 3.3, we reviewed three terms that are problematic in the details of their varying definitions: observation, in-situ and interoperability, as examples of broader content issues

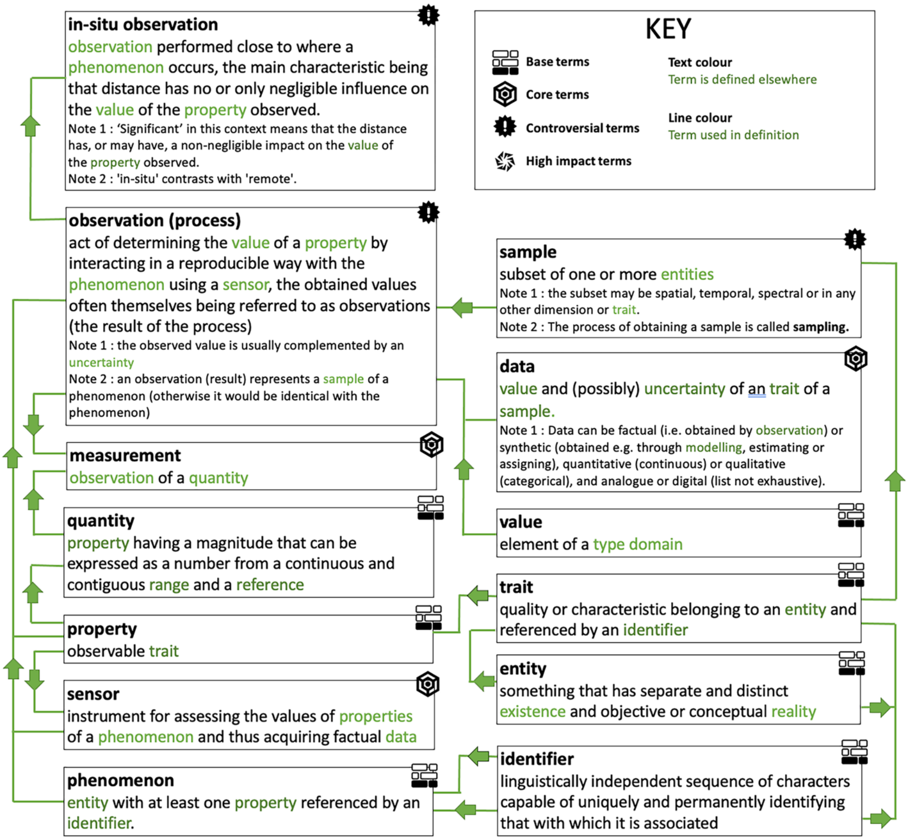

Fig. 2 Initial suggestion for an interrelated set of terms built on carefully defined base terms. Symbols as defined in Table 3. For accessibility reasons, the terms within this diagram are also listed in an Appendix, rather than providing all definitions in an accessible alternative text

in existing vocabularies. In Fig. 2, we present a draft set of definitions assigning them tentatively to the different classes we proposed. These should be considered a starting point, representing the framework outlined here. The diagram contains only a small number of the terms considered by the CEOS Terminology Task Group, an overall set of ~30 linked terms has so far been developed.

In these definitions, we follow the conventions of the VIM (BIPM et al. 2012) which apply the 'substitution principle'. The VIM describes this in its convention as 'it is possible in any definition to replace a term referring to a concept defined elsewhere in the [vocabulary] by the definition corresponding to that term, without introducing contradiction or circularity'. For ontological purposes it may also be appropriate to provide a second Aristotelian (see Sect. 3.1) definition that describes the formal relationships.

Observation is a 'controversial term', for the reasons given in Sect. 3.3.1. Our definition that it is 'the act of determining the value of a property by interacting in a reproducible way with the phenomenon with a sensor' emphasises 'observation' as an act (rather than its result) in the same way as the Geolexica definition. It also emphasises 'observation' as linked to a sensor interaction with a phenomenon, thus not including purely synthetic data within the same definition (and in that way being at odds with the Geolexica definition). It is possible that notes, and branched definitions could allow for different communities to split up the continuous scale from 'pure' measurement to 'pure' simulation at distinct points, perhaps using the Bokulich (Bokulich 2020) categorisation.

In our proposal, we define 'in-situ' as 'observations performed close to where the phenomenon occurs. The main characteristic of such observations is that distance has no or only negligible (within uncertainties) influence on the (obtained) value of the property'. In this way, we are distinguishing 'in-situ' from 'remote'. An alternative would be to create a disambiguation page that would allow for 'in-situ (vs satellite)', 'in-situ (vs ex-situ)' and 'in-situ (vs remote)' to be defined differently.

In the definitions in Fig. 2, we make no suggestion for 'interoperability'. As a high-impact term, it is best defined superficially as in the Wikipedia definition given in Sect. 3.3.3, along with active links to full framework documents for different communities' approaches to interoperability.

4.5 Practical Steps Towards Such a Thesaurus Initiative

Establishing a thesaurus that meets these requirements will be very difficult, but we believe not impossible. We recognise many of the good practices already existing in the communities whose published vocabularies we considered (Sect. 2.2) and we recognise that people working in the Earth sciences have taken care to define their terms diligently. ISO/TC 211, which maintains the Geolexica (https://isotc211.geolexica.org/concepts) has particularly helpful practices, including the already mentioned public discussion forums (e.g. https://github.com/opengeospatial/om-swg/issues/175) on controversial terms, and the use of the GitHub based tool 'glossarist' (www.glossarist.org) to provide the necessary links and version control. Similarly, the work in SWEET and I-ADOPT to develop ontology graphs for these kinds of terms provides another building block of what is needed. Practical steps towards a new initiative must build on (and link) these existing efforts and involve those teams, as well as engaging new participants from the wider Earth science communities.

ISO 25964 part 1 (ISO 25964–1 2011) provides, in its Sect. 13, guidance for how a thesaurus could be developed, warning that 'a thesaurus is a labour-intensive job, requiring commitment for many years if it is to prove worthwhile'. It recommends the development be treated as a project with planning, compilation, construction, dissemination and updating/maintenance phases. Such a project should be overseen by a project manager and include senior champions, a community of interested and informed users and professionals knowledgeable in the subject area, and IT professionals who can support the thesaurus management software. Also implicit is the assumption that such a project involves a thesaurus curator and an ontologist who understand the importance and challenges of constructing formal ontologies and good practices in definition development and can facilitate the discussions between the subject experts to create definitions and links that follow the constructed ontology. This endeavour should be supported by using dedicated thesaurus management software to ensure that each term is individually addressable, all relationships (parent, sibling, child) are identified and maintained, and circular definitions avoided. Such a thesaurus will be human- and machine-readable alike.

In Sect. 4.2, we recommended that the only promising way to avoid creating further isolated vocabularies would be convergence across all stakeholders, ideally to a single thesaurus instance following common rules, and where the partners forego their individual authorities over their vocabularies and transfer these to the common undertaking. We have been warned several times including during the review phase of this manuscript that what we try to accomplish will be difficult if not impossible. Still, in a comment, Parsons et al. (2022a, b) come to similar conclusions as we did at a different, even wider scope and called for building a 'community-driven glossary' for psychology, economics, neuroscience, information science, social science, biology, ecology, public health and linguistics. A respective initiative called the Framework for Open and Reproducible Research Training (FORRT) has started a glossary (https://forrt.org/glossary/) implementing many, though not all, of the principles proposed here, thus demonstrating their feasibility. We are convinced that with Earth Observation as a crucial and horizontal issue in Earth sciences we have identified a very suitable 'soft spot' to start a similar attempt.

We believe that CEOS is in an advantageous position to initiate the development of a joint open Earth Observation thesaurus, most usefully in formal collaboration with ISO/TC 211 (who may perhaps be asked to lead), OGC, the WMO task team on WIGOS metadata. If successful it could be easily expanded by any other stakeholder in Earth System Sciences interested to join. We recognise that it will need significant resources, and that most of these organisations are not placed to fund an additional initiative directly. However, we note that many individuals within the standards curators and across many other stakeholders already spend quite some time on collating and managing separate vocabularies (see Table 1), and it is highly likely that a collaboration will save them considerable effort in the mid to long term. Even larger will be the benefit to the Earth sciences community at large which could save significant resources in searching for and developing shared definitions resulting in faster and deeper interoperability.

Until such a vocabulary collaboration is established, to support cross-disciplinary understanding, we encourage communities and authors to make their meaning of terms explicit, wherever possible reusing existing definitions and referencing their sources, and only as a last resort by assembling new 'terms and definitions' sections for individual documents.

5 Conclusions

This paper has reviewed several selected vocabularies in the Earth (Observation) sciences and identified that although they contain well-considered and consistent definitions in their core terms, their 'base terms' are inconsistent with circular definitions. We believe this inconsistency is a consequence of an inefficient approach to vocabulary development, where decentralised vocabulary efforts are later combined and reconciled through larger committees, rather than one where communities derive their vocabularies from within a common framework that facilitates consistency from the start. Furthermore, there are several fundamental and widely used terms in existing vocabularies, e.g. 'in-situ', 'sample', and 'observation', whose definitions are contradictory for different communities that are relevant across the Earth sciences. There are other terms, e.g. 'interoperability', 'analysis ready data' that describe concepts that cannot meaningfully be captured by a dictionary-style definition but would greatly benefit from a central authoritative source for clarification. We have presented some attempts of defining respective base and core terms in line with our criteria and proposed creating a collaborative initiative to work towards building a thesaurus based on project principles given in the ISO 25964 guidelines.

We believe that such an effort would lead to a thesaurus that is:

5.1 Consistent

Based on a firm foundation of clear base terms. Ambiguity and interpretability would be avoided or made explicit. This objective can be achieved by basing definitions on a single set of preferred terms which are unambiguously and natively defined, and to which additional terms are referenced (ISO 25964–1 clause 15.2.3 acknowledges this aspect).

5.2 Interrelated

With more complex terms building on base terms and establishing clear relationships between terms (parent, sibling, child) avoiding, in particular, circular (child becomes parent) relations. Overlaps between terms that are supposed to delineate more generic concepts (siblings) would be avoided or minimised by clearly defining criteria of differentiation/disambiguation. This concept is described in ISO 25964–1 clause 10.2.

5.3 Understandable

With definitions being made centrally and agreed by all collaborating communities (or through differentiation/disambiguation), the thesaurus will improve the understandability of terminology in interdisciplinary teams. ISO 25964–2 clause 5.1 describes the value of understandability by stating that that a good thesaurus 'enable[s] an expression formulated using one vocabulary to be converted to a corresponding expression in one or more other vocabularies'.

5.4 Educational

Addressing a human audience, a well-presented thesaurus promotes its adoption as a common conceptual framework by a broader community (e.g. across all Earth sciences). Clear and explanatory definitions, and linkages between words expressed in multiple structural ways, as well as highlighted discussions between communities would satisfy the curiosity of scientists at all career stages to gain knowledge, help communicate more efficiently and encourage productive across-community discussions.

5.5 Updateable

A unified thesaurus with version control at the level of individual terms, opportunities for public comment and discussion, disambiguation links and options for adding new terms, will add significant additional value to the community by providing a persistent while current source of reference. Links for 'high impact terms' to framework documents that allow for far more refined definitions and check lists, will also help keep the thesaurus updated.

We strongly encourage the communities engaged in the Earth sciences first and foremost those engaged in multidisciplinary Earth Observation to embark and, where possible, sponsor such an initiative.

Appendix of Terms

The terms given in Fig. 2 are also given here to assist with readability and accessibility. Links are not described here. The Knowledge Centre on Earth Observation (KCEO) of the European Commission has started a pilot implementation of a glossary based on the principles and concepts laid out in this paper. You can follow its proceedings at https://ec-jrc.github.io/KCEO-Glossary/.

in-situ Observation [Controversial Term]

Observation performed in the same place where a phenomenon occurs, normally without isolating it from other systems (its environment) or altering its pre-observation state. The main characteristic of such observations is that the distance has no or only negligible (within *uncertainty*) influence on the *value* of the *property* observed. In-situ *observations* therefore often require either direct physical contact or small distances between a *sensor* and the observed phenomenon.

Note 1: *Observations* not fulfilling these conditions are considered *Remote Sensing*.

observation (Process) [Controversial Term]

Act of determining the value of a property by interacting in a reproducible way with the phenomenon using a sensor, the obtained values often themselves being referred to as observations (the result of the process).

Note 1: the observed value is usually complemented by an uncertainty.

Note 2: an observation (result) represents a sample of a phenomenon (otherwise it would be identical with the phenomenon).

measurement [Core Term]

Observation of a quantity.

quantity [Base Term]

Property having a magnitude that can be expressed as a number from a continuous and contiguous range and a reference.

property [Base Term]

Observable trait.

sensor [Core Term]

Instrument for assessing the values of properties of a phenomenon and thus acquiring factual data.

phenomenon [Base Term]

Entity with at least one property referenced by an identifier.

sample [Controversial Term]

Subset of one or more entities.

Note 1: the subset may be spatial, temporal, spectral or in any other dimension or trait.

Note 2: The process of obtaining a sample is called **sampling.**

data [Core Term]

Value and (possibly) uncertainty of a trait of a sample.

Note 1: Data can be factual (i.e. obtained by observation) or synthetic (obtained e.g. through modelling, estimating or assigning), quantitative (continuous) or qualitative (categorical), and analogue or digital (list not exhaustive).

value [Base Term]

Element of a type domain.

trait [Base Term]

Quality or characteristic belonging to an entity and referenced by an identifier.

entity [Base Term]

Something that has separate and distinct existence and objective or conceptual reality.

identifier [Base Term]

Linguistically independent sequence of characters capable of uniquely and permanently identifying that with which it is associated.

Acknowledgements We thank all members of the CEOS Technical Task Group who were able to provide early guidance on the themes discussed in this paper, as well as to the twelve scientists who provided extremely helpful comments on an early draft of this paper which led to considerable improvements and amendments. We also thank the three anonymous peer reviewers who provided valuable challenges and perspectives to the submitted paper.

Author Contributions All authors contributed to all stages of the paper development. PAS led the conceptualisation and the committee activity that led to this paper. ERW led the drafting of the paper. However, all aspects of the work (conceptualisation, drafting, reviewing, editing, finalising) involved PAS, ERW and KM.

Funding Work by ERW was carried out in the framework of the Instrument Data quality Evaluation and Assessment Service—Quality Assurance for Earth Observation (IDEAS-QA4EO) contract funded by the European Space Agency, European Space Research Institute: ESA-ESRIN (n. 4000128960/19/I-NS).

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article. The authors were participants of the CEOS terminology task force described within this paper.

References

Arp R, Smith B, Spear AD (2015) Building ontologies with basic formal ontology. The MIT Press. https://doi.org/10.7551/mitpress/9780262527811.001.0001

Baker M (2016) 1,500 scientists lift the lid on reproducibility. Nature. https://doi.org/10.1038/533452a

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, & OIML (2012) International vocabulary of metrology—Basic and general concepts and associated terms (VIM) https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf/f0e1ad45-d337-bbeb-53a6-15fe649d0ff1

BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP, & OIML (2020) Guide to the expression of uncertainty in measurement—Part 6: Developing and using measurement models. https://www.bipm.org/documents/20126/2071204/JCGM_GUM_6_2020.pdf/d4e77d99-3870-0908-ff37-c1b6a230a337

Bokulich A (2020) Towards a taxonomy of the model-ladenness of data. Philosophy Sci 87(5):793–806. https://doi.org/10.1086/710516

Boyd N. M, Bogen J (2021) Theory and observation in science. In E. N. Zalta (Ed.) The Stanford Encyclopedia of Philosophy (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/science-theory-observation/

Chander G, Mishra N, Helder DL, Aaron DB, Angal A, Choi T, Xiong X, Doelling DR (2013) Applications of spectral band adjustment factors (SBAF) for cross-calibration. IEEE Trans Geosci Remote Sens 51(3):1267–1281. https://doi.org/10.1109/TGRS.2012.2228007

DG-RI (EC), EOSC Executive Board, Corcho O, Eriksson M, Kurowski K, Ojsteršek M, Choirat C, Sanden M. van de, Coppens F. (2021). EOSC interoperability framework: report from the EOSC executive board working groups fair and architecture. Publications office of the European Union. https://data.europa.eu/doi/https://doi.org/10.2777/620649

Goryl P, Fox N, Donlon C, Castracane P (2023) Fiducial reference measurements (FRMs): What Are They? Remote Sensing. https://doi.org/10.3390/rs15205017

ISO 19143 (2010) ISO 19143:2010 Geographic information—Filter encoding (19143:2010) ISO https://www.iso.org/standard/42137.html

ISO 19156 (2023) ISO 19156:2023 Geographic information—Observations, measurements and samples (19156:2023). ISO. https://www.iso.org/standard/82463.html

ISO 25964–1 (2011) ISO 25964–1:2011: Thesauri and interoperability with other vocabularies—Part 1: Thesauri for information retrieval (26964–1:2011). ISO. https://www.iso.org/standard/53658.html

ISO 25964–2 (2013) ISO 25964–1:2013: Thesauri and interoperability with other vocabularies—Part 2: Interoperability with other vocabularies (26964–2:2013). ISO. https://www.iso.org/standard/53657.html

ISO/IEC 2382 (2015) ISO/IEC 2382:2015 Information technology—Vocabulary (2382:2015). ISO/IEC. https://www.iso.org/standard/63598.html

JCGM 200:2012 (2012) International Vocabulary of Metrology (VIM) [Text.Article]. NASA Earth Observatory. https://jcgm.bipm.org/vim/en/

MacLeod M (2018) What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. Synthese 195(2):697–720. https://doi.org/10.1007/s11229-016-1236-4

Magagna B, Moncoiffe G, Devaraju A, Stoica M, Schindler S, Pamment A. (2022). Interoperable descriptions of observable property terminologies (I-ADOPT) WG outputs and recommendations. RDA endorsed Recommendations https://doi.org/10.15497/RDA00071

Nature editorial (2018) Checklists work to improve science. Nature 556(7701):273–274. https://doi.org/10.1038/d41586-018-04590-7

OGC (2023) OGC Abstract specification Topic 20: observations, measurements and samples (Topic 20) OGC. https://www.ogc.org/standards/as/

Parsons MA, Duerr R, Godøy Ø (2022a) The evolution of a geoscience standard: an instructive tale of science keyword development and adoption. Geosci Front 14(5):101400. https://doi.org/10.1016/j.gsf.2022.101400

Parsons S, Azevedo F, Elsherif MM, Guay S, Shahim ON, Govaart GH, Norris E, O'Mahony A, Parker AJ, Todorovic A, Pennington CR, Garcia-Pelegrin E, Lazić A, Robertson O, Middleton SL, Valentini B, McCuaig J, Baker BJ, Collins E, Aczel B (2022b) A community-sourced glossary of open scholarship terms. Nat Hum Behav. https://doi.org/10.1038/s41562-021-01269-4

Peng G, Privette JL, Kearns EJ, Ritchey NA, Ansari S (2015) A unified framework for measuring stewardship practices applied to digital environmental datasets. Data Sci J. https://doi.org/10.2481/dsj.14-049

de Podesta M, Bell S, Underwood R (2018) Air temperature sensors: dependence of radiative errors on sensor diameter in precision metrology and meteorology. Metrologia 55(2):229. https://doi.org/10.1088/1681-7575/aaaa52

Raskin RG, Pan MJ (2005) Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Comput Geosci 31(9):1119–1125. https://doi.org/10.1016/j.cageo.2004.12.004

Slater PN, Biggar SF, Thome KJ, Gellman DI, Spyak PR (1996) Vicarious radiometric calibrations of EOS sensors. J Atmos Oceanic Tech 13(2):349–359. https://doi.org/10.1175/1520-0426(1996)013%3c0349:VRCOES%3e2.0.CO;2

Strobl P (2023) A revised processing level scheme for Earth Observation data. Big Data from Space, Vienna. https://data.europa.eu/doi/10.2760/46796

Weaver R (2014) Processing Levels. In: Njoku EG (ed) Encyclopedia of Remote Sensing. Springer, pp 517–520

WGISS (2008) CEOS WGISS Interoperability Handbook. https://ceos.org/ourwork/workinggroups/wgiss/documents/

WGISS (2023) CEOS WGISS data management and stewardship maturity matrix. https://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/White_Papers/WGISS%20Data%20Management%20and%20Stewardship%20Maturity%20Matrix.pdf

Wikipedia interoperability (2023) Interoperability. In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Interoperability&oldid=1180639404

Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Mons B (2016) The FAIR guiding principles for scientific data management and stewardship. Scientific Data. https://doi.org/10.1038/sdata.2016.18

WMO-No.8 (2021) Guide to instruments and methods of observation (WMO-No. 8). Volume I–Measurement of Meteorological Variables. WMO. https://community.wmo.int/en/activity-areas/imop/wmo-no_8

Woodhouse IH (2021) On 'ground' truth and why we should abandon the term. J Appl Remote Sens. https://doi.org/10.1117/1.JRS.15.041501