

The I-ADOPT Interoperability Framework for FAIRer Data Descriptions of Biodiversity

Barbara Magagna¹, Iliaria Rosati², Maria Stoica³, Sirko Schindler⁴,
Gwenaelle Moncoiffe⁵, Anusuriya Devaraju⁶, Johannes Peterseil¹ and Robert Huber⁷

¹Environment Agency Austria, Spittelauer Laende 5, 1090 Vienna, Austria

²Institute of Research on Terrestrial Ecosystems, National Research Council (CNR), Via Salaria Km 29.300, 00015 Monterotondo Stazione, Rome, Italy

³Institute of Arctic and Alpine Research University of Colorado 4001 Discovery Drive Boulder, CO 80303 USA

⁴Institute of Data Science, German Aerospace Center, Mälzerstraße 3-5, 07745 Jena, Germany

⁵British Oceanographic Data Centre National Oceanography Centre Liverpool, Joseph Proudman Building, 6 Brownlow Street, Liverpool L3 5DA, United Kingdom

⁶Terrestrial Ecosystem Research Network (TERN), University of Queensland, Long Pocket Precinct, Level 5 Foxtail Building 1019, 80 Meiers Road, Indooroopilly, QLD 4068, Australia

⁷MARUM - Center for Marine Environmental Sciences, University of Bremen, Leobener Str. 8, 28359 Bremen, Germany

Abstract

Biodiversity, the variation within and between species and ecosystems, is essential for human well-being and the equilibrium of the planet. It is critical for the sustainable development of human society and is an important global challenge. Biodiversity research has become increasingly data-intensive and it deals with heterogeneous and distributed data made available by global and regional initiatives, such as GBIF, ILTER, LifeWatch, BODC, PANGAEA, and TERN, that apply different data management practices. In particular, a variety of metadata and semantic resources have been produced by these initiatives to describe biodiversity observations, introducing interoperability issues across data management systems. To address these challenges, the Interoperable Descriptions of Observable Property Terminology WG (I-ADOPT WG) was formed by a group of international terminology providers and data center managers in 2019 with the aim to build a common approach to describe what is observed, measured, calculated, or derived. Based on an extensive analysis of existing semantic representations of variables, the WG has recently published the I-ADOPT framework ontology to facilitate interoperability between existing semantic resources and support the provision of machine-readable variable descriptions whose components are mapped to FAIR vocabulary terms. This contribution will focus on how the I-ADOPT framework can be applied to represent variables commonly used in the biodiversity domain.

Keywords

observable properties, variables, biodiversity, observations, semantic interoperability


S4BioDiv 2021: 3rd International Workshop on Semantics for Biodiversity, held at JWOW 2021: Episode VII The Bolzano Summer of Knowledge, September 11–18, 2021, Bolzano, Italy

✉ barbara.magagna@umweltbundesamt.at (B. Magagna); ilia.rosati@cnr.it (I. Rosati); maria.stoica@colorado.edu (M. Stoica); sirko.schindler@dlr.de (S. Schindler); gmon@bodc.ac.uk (G. Moncoiffe); a.devaraju@uq.edu.au (A. Devaraju); johannes.peterseil@umweltbundesamt.at (J. Peterseil); rhuber@uni-bremen.de (R. Huber)

ORCID 0000-0003-2195-3997 (B. Magagna); 0000-0003-3422-7230 (I. Rosati); 0000-0002-6612-3439 (M. Stoica); 0000-0002-0964-4457 (S. Schindler); 0000-0001-6559-4178 (G. Moncoiffe); 0000-0003-0870-3192 (A. Devaraju); 0000-0003-0631-8231 (J. Peterseil); 0000-0003-3000-0020 (R. Huber)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

1. Introduction

The interactions and diversity of organisms within and across the Earth's ecosystems play critical roles in the coevolution of the biosphere and the broader Earth system. Human societies have evolved with and are part of this complex system [1]. Societal development and well-being depend on ecosystem services, or nature's contributions to people [2]. Global biodiversity, however, is declining faster than ever in human history [3]. Biodiversity knowledge integration is urgently required to support responses towards its conservation and a sustainable future [4].

Over the last decades, the need to share data and knowledge on biodiversity has led to numerous local, regional, and global initiatives resulting in an unprecedented biodiversity data mobilization [5]. Like many other domains, biodiversity research has been transformed by a big data revolution. As the creation of large volumes of complex data becomes routine, new approaches using automated systems are required to find, access, combine, and interpret their meaning [6]. The variety in data management practices and lack of common standards led to an increase of heterogeneity at multiple levels and severely impedes the integration of disparate data. Semantic approaches promise to overcome this challenge by capturing rich representations of biodiversity data to facilitate maximum interoperability and provide detailed descriptions for re-use. This has led to a large collection of independent terminology resources and tools across research infrastructures and communities as explored below. Their complexity and diversity often overwhelm data managers and users, ironically maintaining barriers to interoperability.

LifeWatch Italy is one of the national branches of LifeWatch ERIC, the European e-science infrastructure for biodiversity and ecosystem research (LifeWatch ERIC¹) [7]. Coordinated by the National Research Council, LifeWatch Italy includes 35 partner institutions covering a wide range of scientific disciplines (terrestrial, marine and freshwater ecology, biology, zoology, botany, archaeobotany, geography, forestry, agriculture). It seeks to reinforce integrated scientific research into biodiversity by the development of a National Hub as the main access point to data, apps, and eScience services for the management, aggregation, analysis, and re-use of biodiversity and ecosystem research data. LifeWatch Italy has already developed a Data Management System which can receive biodiversity (e.g., checklists on geo-spatial distribution and abundance in Italy of vegetal and animal species) and ecosystem data (e.g., abiotic variables), including species traits data with grain size from individual to species (e.g., morphological traits of phytoplankton). The system consists of two major components, the LifeWatch profile for the dataset description, based on the metadata standard EML 2.2.0² and a data schema based on the Darwin Core standard³ and controlled vocabularies (e.g., LifeWatch Italy thesauri, EnvThes; in EcoPortal⁴), which provide a unified framework for the information management of LifeWatch Italy. Thanks to a national infrastructure project, new developments and upgrades are already planned aiming specifically to improve the FAIRness of biodiversity and ecosystem data.

The emerging European Long-Term Ecosystem, critical zone and socio-ecological Research Infrastructure (eLTER RI⁵ [8]) is currently in the preparatory phase building on LTER Europe, the

¹<https://www.lifewatch.eu/>

²<https://eml.ecoinformatics.org/>

³<https://dwc.tdwg.org/>

⁴<http://ecportal.lifewatch.eu/>

⁵<https://www.lter-europe.net/elter-esfri>

regional network ofILTER. eLTER RI will adopt a fundamentally systemic approach to observe and analyze the environmental system, encompassing biological, geological, hydrological and socio-ecological perspectives. It will allow in-situ, co-located acquisition and gathering of Essential Variables ranging from bio-physicochemical to biodiversity and socio-ecological data. Ecosystem change caused by long-term pressures and short-term pulses will be investigated in a nested design across the scales covered by the eLTER RI site network. Biodiversity data for a number of species groups are collected by the contributing sites. Pilotto et al. [9] compiled 161 biodiversity time series mainly from European eLTER Sites focusing on freshwater ecosystems taking the effort to harmonize and standardize the data. A common infrastructure to harmonize species names still needs to be implemented. Currently, e.g. the R-package TaxonStand⁶ is used to harmonize species data based on ThePlantList⁷. This will be further elaborated in the near future focusing on a harmonized reporting of biodiversity data. For reporting of species information a common data specification for eLTER [10] has been developed following the recommendation of, e.g., the ICP Integrated Monitoring programme.⁸ For the future, biodiversity research and monitoring will be able to profit from a fully standardized and harmonized set of biodiversity standard observations. With this, many more sites could be used for various scientific analyses. The same applies to a suitable set of environmental standard observations [11] that help interpreting changes in biodiversity.

The British Oceanographic Data Centre (BODC) is managed and operated by the UK's National Oceanography Centre (NOC), an independent self-governing organization with a status of charitable company limited by guarantee. BODC is the marine component of the Natural Environment Research Council's (NERC's) Environmental Data Centre network. Its mission is to develop, coordinate, and provide specialist data services for marine science communities; to enable innovative use and re-use of data; to ensure long-term curation of valuable and unique marine data resources; and to champion Open Data. BODC runs the NERC Vocabulary Server (NVS)⁹, an internationally known and globally used infrastructure for the management of and access to controlled vocabularies related to the marine science and associated domains. Vocabularies served by the NVS underpin the SeaDataNet and EMODnet infrastructure. They have been adopted by the Ocean Biodiversity Information System (OBIS) to support data harmonization within and interoperability of the OBIS Darwin Core MeasurementOrFact Extension [12]. These vocabularies, describing and identifying sensor and platform types and models, variables measured or derived, units, methods, and many other essential metadata elements, are used by an increasing number of scientific marine data networks across the globe.

The information system PANGAEA¹⁰ is jointly managed by the Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI) and the Centre for Marine Environmental Sciences (MARUM) at the University of Bremen. PANGAEA is a certified (CoreTrustSeal), trustworthy long-term operating repository providing continued access to more than 400,000 datasets from various sub-disciplines of Environmental Sciences. A large part of PANGAEA's dataset consists of biodiversity data, ranging from fossil records to modern faunal and floral

⁶<https://cran.r-project.org/web/packages/Taxonstand/>

⁷<http://www.theplantlist.org/>

⁸<https://unece.org/integrated-monitoring>

⁹https://www.bodc.ac.uk/resources/products/web_services/vocab/

¹⁰<https://www.pangaea.de/>

observations collected through research infrastructures, research projects, and programs as well as individual researchers. The required observed variables are managed in a controlled vocabulary that is routinely undergoing an automated semantic annotation process [13] to link, e.g., taxonomic vocabularies such as WORMS or ITIS to harmonize species-related data. Access to biodiversity datasets which are routinely tagged by a DOI is enabled through support for a number of community-specific and cross-domain standards.

The Terrestrial Ecosystem Research Network (TERN)¹¹ collects and preserves biodiversity data together with other critical terrestrial ecosystem data from the continental scale to field sites across Australia to support researchers and policymakers' immediate and long-term needs, including state and regional governments. TERN has developed several valuable data services to support the discovery, analysis, and re-use of data. One of the core services is enabling data harmonization and integration through semantic data model¹² (i.e., ontology) and machine-interpretable vocabularies. The TERN data model is developed based on the SOSA (Sensor, Observation, Sample, and Actuator)¹³ and comprises vocabularies for representing ecological survey data, including biodiversity, vegetation, and soil. The datasets are enriched with machine-interpretable vocabularies representing variables measured, features, units, methods, instruments, platforms, organization, and people. The vocabularies are available publicly through the TERN Linked Data Services¹⁴. They are developed either in-house based on community-endorsed specifications (e.g., Australian Soil and Land Survey Field Handbook and AusPlots Rangelands Survey Protocols Manual) or imported from external semantic resources (e.g., QUDT¹⁵, GCMD¹⁶, and NERC). The extensible data model and rich collections of vocabularies are building blocks for improving TERN data curation and discovery.

These are just a few regional implementations demonstrating a highly diverse landscape of data management practices, even when dealing with the same type of resources. This makes it clear that there are still some difficult hurdles to overcome in order to achieve interoperability and convergence between infrastructures in the same domain, and even more challenging across domains. Initial progress has been made in providing machine-readable descriptions of sensors and their observation types through the OGC's Sensor Web Enablement SensorML, Observations and Measurements (O&M), the Sensor Observation Service (SOS) as well as the W3C's/OGC's Semantic Sensor Network (SSN) ontology¹⁷. This work has been further refined within the OGC SensorThings API specification, as well as the upcoming version 3 of O&M. However, "deep metadata" that further contextualizes observations (e.g., methodology or observable properties) is typically represented as coarsely qualified classes. What exactly falls into these classes is currently unconstrained and could be anything ranging from unstandardized free-text to standardized descriptions accessible via fully resolvable URIs.

To address the challenge to properly model and describe observable properties (also called variables), representative members of the above initiatives decided to collaborate under the

¹¹<https://www.tern.org.au/>

¹²https://ternaustralia.github.io/ontology_tern/

¹³<https://www.w3.org/TR/vocab-ssn/>

¹⁴<https://linkeddata.tern.org.au/>

¹⁵<https://www.qudt.org/>

¹⁶<https://earthdata.nasa.gov/earth-observation-data/find-data/idn/gcmd-keywords>

¹⁷<https://www.ogc.org/docs/is>

umbrella of the Research Data Alliance. The Working Group Interoperable Descriptions of Observable Property Terminology (RDA I-ADOPT WG)¹⁸ was born in 2019 and is now in its final phase. This group of international terminology providers and data center managers set itself the objective to produce an Interoperability Framework paving the way for seamless terminology alignment for variable descriptions.

2. Methodology

This I-ADOPT WG has a strong focus on **variables observed in environmental research** because it leverages existing efforts to accurately encode what was measured, observed, derived, or computed in relation to Earth's systems. However, many of the principles it leans on are relevant to or connected with other domains. The construction of the framework has been informed by a review of current practices used in the community. The working group is also iteratively testing and refining the framework through a set of in-depth use cases. Much like a generic blueprint, the refined conceptual framework will be a basis upon which terminology developers can formulate or refine their local design patterns to more easily align with others and to avoid insurmountable inconsistencies. By utilizing the recommendations of the I-ADOPT framework, terminology developers may expand the applicability and interconnectivity of their local resources in a collective attempt to uniformly represent complex properties observed across the environmental sciences, from marine, atmospheric, and terrestrial Earth sciences to biodiversity.

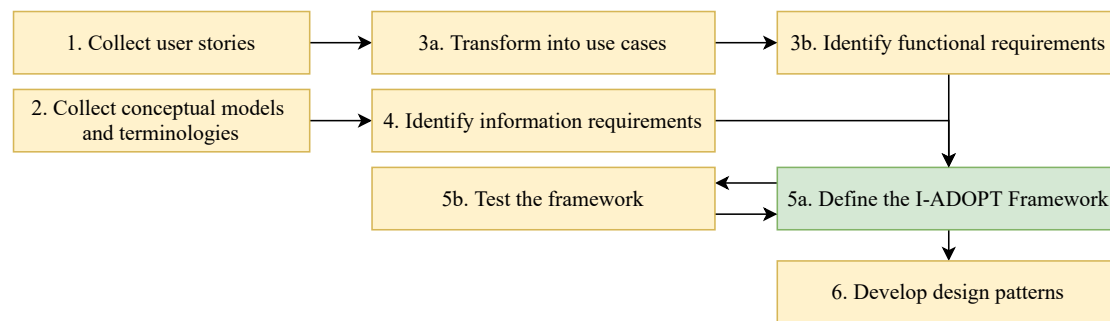


Figure 1: Tasks of the RDA WG I-ADOPT.

The I-ADOPT WG has been working for about two years to develop the common framework for agreed-upon components to represent different aspects of observable properties. The timeline followed over the last few years involved completing six different tasks (see Figure 1):

1. Collect user stories from the community about needs related to semantic representation of observable properties involving nitrogen-related observable properties, following a submission template¹⁹.

¹⁸<https://www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt>

¹⁹https://github.com/i-adopt/users_stories

2. Collect semantic representations from the community, including terminologies, standard vocabularies, ontologies, and conceptual models²⁰.
3. Derive framework requirements by transforming user stories into concrete use cases, categorizing use cases, and defining functional needs within each category²¹.
4. Analyze cataloged semantic artefacts to derive the framework's information requirements – a set of conceptual classes and relationships common to the varying representation mechanisms, whether explicit or implicit.
5. Develop the I-ADOPT Framework, including core variable components, their relationships and definitions, and iterate on feedback from the community.
6. (*ongoing*) Use the I-ADOPT Framework to develop aligned design patterns across several participating terminology resources for selected use cases.

During the framework development, many potential framework components were considered, discussed, and iterated upon, but only a small subset of those are included in the current recommendation. Discussions of the different possible components that were in the end not considered essential to the framework, such as statistical measures, denominations of objects of interest, and property dimensions and units, are documented through the group's GitHub page and supplemental documents. These efforts resulted in an ontology for providing a machine-readable representation of the framework.

3. Results

The I-ADOPT Framework ontology²² is a concise ontology (see Figure 2) made of four classes (Variable, Property, Entity, Constraint) and six object properties (hasProperty, hasObjectOfInterest, hasContextObject, hasMatrix, hasConstraint, constrains). It is designed to facilitate interoperability between existing variable description schemes (including domain-specific ontologies, semantic models, and structured controlled vocabularies) and to support further development of machine-readable variable descriptions that re-use components mapped to FAIR vocabulary terms. This first official version of the ontology has been developed by a core group of terminology experts and users from the I-ADOPT WG.

The framework is variable-centric and can be used to describe result values of any type of data acquisition events, be it a human-based observation, a sensor-based measurement, a calculation, or a simulation. The Variable describes WHAT has been observed, measured, simulated, or calculated independently of WHERE (site description, geographical coordinates), HOW (procedure, protocol) and WHEN (measurement time, time resolution) the data acquisition has taken place, although in some cases the HOW might be important to include in the variable description, as it might pose certain constraints on it. This makes the application of this concept reusable in different settings giving meaning to the determined value. Variables exist as concepts in many vocabularies (see CF Standard Names²³ or CSDMS Standard Names²⁴) to

²⁰<https://github.com/i-adopt/terminologies>

²¹https://github.com/i-adopt/usecase_analysis

²²<https://w3id.org/iadopt/ont/>

²³<https://cfconventions.org/Data/cf-standard-names/47/build/cf-standard-name-table.html>

²⁴https://csdms.colorado.edu/wiki/CSDMS_Standard_Names

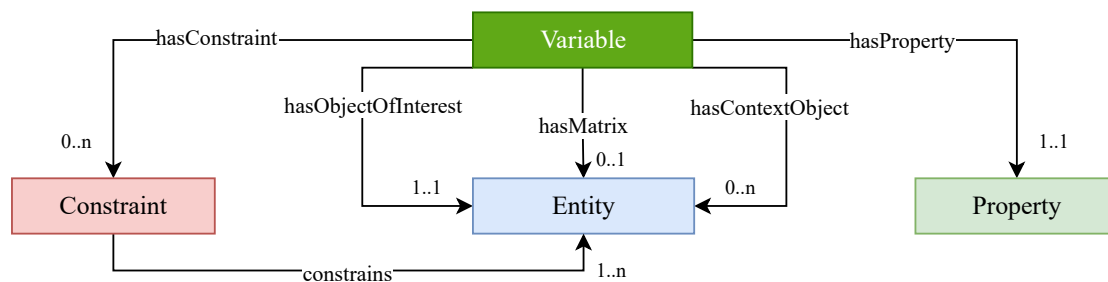


Figure 2: The I-ADOPT Framework.

provide metadata for the values provided in databases. They are also known as observable properties (see Observable Property Vocabulary²⁵) or parameters (see BODC Parameter Usage Vocabulary (P01)²⁶ and EnvThes²⁷). This variety of approaches and the lack of a common representation strategy lead to a situation where variable concepts from different terminologies are still not directly comparable. Consequently, the data that is annotated with these concepts are not interoperable and requires substantial manual efforts to integrate.

In the following paragraphs, we explain the implementation of the framework by the use of an example²⁸, describing a complex variable that requires all of the components of the framework. The variable description is “concentration of endosulfan sulfate in wet flesh of ostrea edulis” (see Figure 3) and refers to the quantitative result (i.e., requiring a magnitude and unit) of a measurement.

The I-ADOPT ontology, inspired by the atomisation approach of the Complex Property Model [14] and The Scientific Variables Ontology²⁹ [15, 16], conceives the **Variable** as a compound concept consisting of at least one entity (the ObjectOfInterest) and its Property, but very often includes further entities contextualizing the target object of observation. Splitting the Variable into constituent concepts enables the reuse and the mapping of these components in the context of other variables. The ObjectOfInterest (*endosulfan sulfate*) can be identified independently of an observable characteristic (*concentration*). By concatenating the constituent elements with prepositions the preferred label of a variable can be automatically created (compare with the upper side of the table in Figure 3). However, the I-ADOPT Framework does not provide any recommendation how the label of the variable should be composed or which order to follow as its main contribution targets machine-readability of the concept. Not included in the variable description are units, statistical measures (unlike in the Complex Property Model), measurement methods, and time-related or geographical location information. Units are essential information for describing measures, but a quantitative variable might be expressed in different units which requires units be modeled independently of variables. An I-ADOPT variable associated with a

²⁵http://registry.it.csiro.au/def/environment/_property

²⁶https://www.bodc.ac.uk/resources/vocabularies/parameter_codes/

²⁷<http://vocabs.lter-europe.net/EnvThes/>

²⁸The following prefixes will be used throughout the examples: obo: <<http://purl.obolibrary.org/obo/>> ; lifewatch: <<http://thesauri.lifewatchitaly.eu/PhytoTraits/index.php?>>> ; nerc: <<http://vocab.nerc.ac.uk/collection/>>> ; worms: <<http://marinespecies.org/aphia.php?p=taxdetails&id=>>>

²⁹<http://www.scientificvariablesontology.org/>

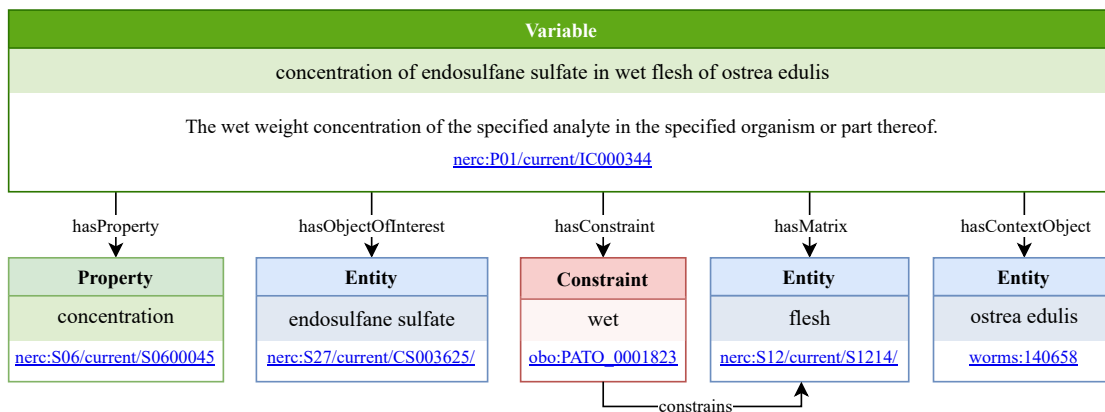


Figure 3: Quantitative example: Concentration of endosulfan sulfate in *Ostrea edulis*.

unit can, however, be interpreted as a more specific variable. By isolating the Variable concept from these supplementary information elements, it is possible to use the I-ADOPT framework in a broader context interoperating with other representations of observations. In this way the Variable with its explicit constituents can be conceived as an extended and machine-readable conceptualization of the respective WHAT concept of these models (e.g. ObservableProperty in SOSA³⁰, Characteristic in OBOE³¹).

The **Property** (*concentration*) is a type of characteristic of the ObjectOfInterest. The cardinality of the object property `hasProperty` is 1..1, meaning that a Variable has to have exactly one Property. The Property is in most cases a more generalized concept than the variable itself, as the variable adds more detail to the property. This has the advantage that the property term can be reused in other variable descriptions and further on allows to identify similar measurements across different variables (e.g., *concentration* regardless of the substance observed).

The **Entity** is any object (*endosulfan sulfate*) or process (e.g., erosion) that has a role in an observation. An Entity may play one of the following roles: ObjectOfInterest, ContextObject, or Matrix. The roles of the entities in the variable description are implemented as object properties in the I-ADOPT ontology. Whether the involvement of a particular entity is meaningful enough to include in the variable description, depends on the specific context.

- **hasObjectOfInterest:** The ObjectOfInterest (*endosulfan sulfate*) is the Entity whose property is observed. A Variable requires exactly one ObjectOfInterest (cardinality: 1..1).
- **hasContextObject:** The ContextObject (*Ostrea edulis*) is the Entity that provides additional background information regarding the ObjectOfInterest. A Variable can have more than one ContextObject (cardinality: 0..n).
- **hasMatrix:** The Matrix (*flesh*) is the Entity in which the ObjectOfInterest is contained. A Variable might only have one Matrix (cardinality: 0..1).

The **Constraint** (*wet*) limits the scope of the observation and confines the context to a particular state. It describes properties of the involved entities (*flesh*) that are relevant to the

³⁰<https://www.w3.org/TR/vocab-ssn/>

³¹<https://github.com/NCEAS/oboe/>

particular observation. Oftentimes Constraints take the form of observations themselves, fixing other properties that influence the main observation considered. It may constrain (**constrains**) multiple entities playing the role of ObjectOfInterest, Matrix, or ContextObject. Constraints are optional, so the cardinality of the object property **hasConstraint** is $0..n$.

The I-ADOPT ontology provides a framework to adequately address the data interoperability requirements emphasized by the FAIR Guiding Principles [6]. Appropriate terms from FAIR semantic artefacts (like SKOS vocabularies or OWL ontologies) should be used not only for the variable itself, but also for each of the components of the variable description. This provides qualified references to other metadata (FAIR Principle I3) where these exist or can be easily created. In order to model a variable accurately, domain-specific knowledge is required. This knowledge should be adequately reflected as human-readable definition of the variable concept. Applying the I-ADOPT framework at the metadata level means using the persistent and resolvable identifier of the variable concept (e.g., Internationalized Resource Identifier - IRI) that links to its components instead of plain text information in the metadata field addressing the observed property (FAIR Principle I2). This results in semantically rich, machine-readable descriptions (metadata) of the dataset. At the dataset level, the provision of FAIR data instead depends very much on the distribution format of the datasets. In the case of spreadsheets, where the quantitative values are described by the column headings, variable concept IRIs should be used instead of labels (see Figure 3). In an ideal FAIR world, data providers would distribute their datasets in a formal, accessible, shared, and broadly applicable language for knowledge representation like RDF (Resource Description Language) (FAIR Principle I1) as a way describing resources on the web. Using RDF allows to use IRIs of the variable concept directly in the triple statement describing the measurement result. Leadbetter and Vodden [14] demonstrated that by FAIRifying metadata it is possible to produce RDF files of the semantically enriched datasets. Because the variable itself links to further descriptive elements, making implicit knowledge about the measurement explicit, it is possible to apply faceted semantic data search via SPARQL queries on data platforms, allowing the selection of datasets involving a substance (like *endosulfan sulfate*) or a specific ObjectOfInterest (like *Ostrea edulis*).

Many variables in the biodiversity domain are human-based observations with a qualitative result value selected from a controlled list. Some elements for the Occurrence, Organism, and Taxon metadata components of the Darwin Core standard require qualitative values for which a few recommended SKOS vocabularies were created by the TDWG Darwin Core Maintenance Group³². LifeWatch Italy vocabularies provide specific terminology for functional traits of several groups of aquatic organism and alien species[17]. An example of how the I-ADOPT framework can be applied to qualitative results of human-based observations is provided in Figure 4 using a variable from the LifeWatch Phytotraits thesaurus. The variable is called *shape* implicitly targeting cells, which is explained in the definition of this specific variable. It is implemented as a SKOS concept and has as more specific concepts qualitative values such as *sphere*, *cube*, or *ellipsoid*. In dataset spreadsheets these values would normally populate the cells of the variable column, but we recommend to use instead the IRIs of these narrower concepts as manifestations of the observed value. However, a more convenient way to store this type of information would be to use RDF for the entire dataset. The constituent components

³²<https://dwc.tdwg.org/list/>

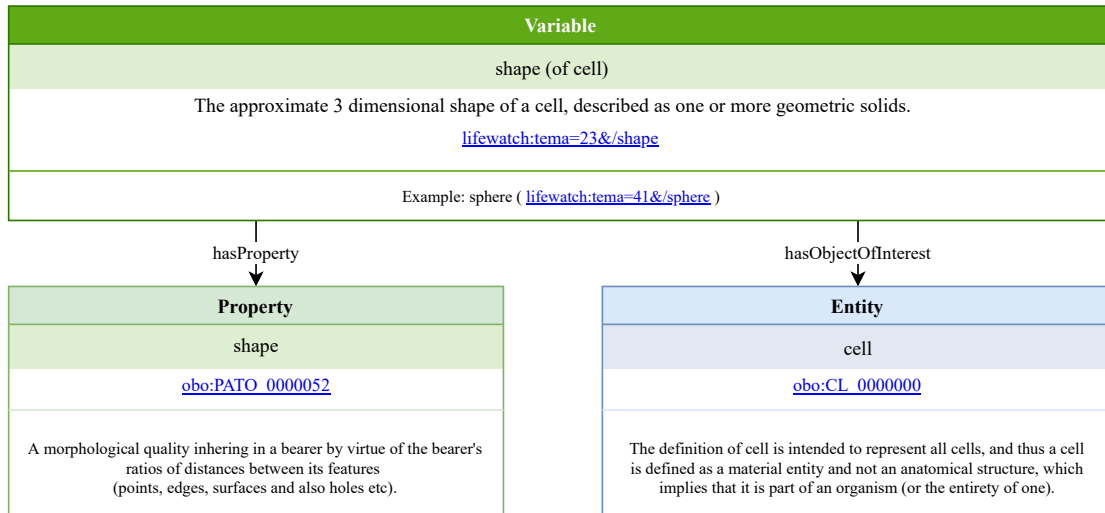


Figure 4: Qualitative example: Shape of cell.

of the variable *shape (of cell)* are the more general concept of *shape* (the **Property**) and the **ObjectOfInterest** the concept *cell*. It is not always possible to find more general properties which could be reused in the context of other variables. For example, the Darwin Core metadata element *degreeOfEstablishment* (degree to which an Organism survives, reproduces, and expands its range at the given place and time) would be represented by the variable with the IRI³³ and composed by the Property *degreeOfEstablishment* with the same IRI and the ObjectOfInterest *organism*³⁴.

4. Outlook

The RDA WG I-ADOPT will in its final phase produce guidelines for the community on how to apply the framework. These will include design patterns to generalize variable descriptions and their modelling for their reuse in different settings (e.g., *concentration of substance in biota*). The authors of this paper will evaluate the adoption of the I-ADOPT approach in their own infrastructures. One first proof-of-concept will be the SKOS representation of the eLTER Standard Observation variables based on the I-ADOPT Framework in EnvThes. An important task will also be to work on mappings to existing observation models (e.g., OBOE) and to foster the ongoing collaboration with initiatives like OGC Observations and measurements Standard Working Group³⁵ and DDI-CDI³⁶ to evaluate how I-ADOPT might increase semantic interoperability of the variable concept within those approaches.

³³<https://dwc.tdwg.org/doi/>

³⁴http://purl.obolibrary.org/obo/OBI_0100026

³⁵<https://github.com/opengeospatial/om-swg>

³⁶<https://ddi-alliance.atlassian.net/wiki/spaces/DDI4/pages/860815393/DDI+Cross+Domain+Integration+DDI-CDI+Review>

Acknowledgments

The authors like to thank the Research Data Alliance, eLTERplus, ENVRI-FAIR (grant agreements No 871128, No 824068) and Alison Pamment (NCAS, UK) for their support.

References

- [1] C. Folke, S. Polasky, J. Rockström, V. Galaz, F. Westley, M. Lamont, M. Scheffer, H. Österblom, S. R. Carpenter, F. S. Chapin, K. C. Seto, E. U. Weber, B. I. Crona, G. C. Daily, P. Dasgupta, O. Gaffney, L. J. Gordon, H. Hoff, S. A. Levin, J. Lubchenco, W. Steffen, B. H. Walker, Our future in the anthropocene biosphere, *Ambio* 50 (2021) 834–869. doi:10.1007/s13280-021-01544-8.
- [2] S. Díaz, U. Pascual, M. Stenseke, B. Martín-López, R. T. Watson, Z. Molnár, R. Hill, K. M. A. Chan, I. A. Baste, K. A. Brauman, S. Polasky, A. Church, M. Lonsdale, A. Larigauderie, P. W. Leadley, A. P. E. van Oudenhoven, F. van der Plaats, M. Schröter, S. Lavorel, Y. Aumeeruddy-Thomas, E. Bukvareva, K. Davies, S. Demissew, G. Erpul, P. Failler, C. A. Guerra, C. L. Hewitt, H. Keune, S. Lindley, Y. Shirayama, Assessing nature's contributions to people, *Science* 359 (2018) 270–272. doi:10.1126/science.aap8826.
- [3] J. Bongaarts, IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, *Population and Development Review* 45 (2019) 680–681. doi:10.1111/padr.12283.
- [4] J. M. Heberling, J. T. Miller, D. Noesgaard, S. B. Weingart, D. Schigel, Data integration enables global biodiversity synthesis, *Proceedings of the National Academy of Sciences* 118 (2021) e2018093118. doi:10.1073/pnas.2018093118.
- [5] R. O. Wüest, N. E. Zimmermann, D. Zurell, J. M. Alexander, S. A. Fritz, C. Hof, H. Kreft, S. Normand, J. S. Cabral, E. Szekely, W. Thuiller, M. Wikelski, D. N. Karger, Macroecology in the age of big data – where to go from here?, *Journal of Biogeography* 47 (2019) 1–12. doi:10.1111/jbi.13633.
- [6] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR guiding principles for scientific data management and stewardship, *Scientific Data* 3 (2016). doi:10.1038/sdata.2016.18.
- [7] A. Basset, W. Los, Biodiversity e-science: LifeWatch, the european infrastructure on biodiversity and ecosystem research, *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* 146 (2012) 780–782. doi:10.1080/11263504.2012.740091.
- [8] M. Mirtl, eLTER, european long-term ecosystem and socio-ecological research infrastructure, *h2020, Impact 2018* (2018) 30–32. doi:10.21820/23987073.2018.8.30.

- [9] F. Pilotto, I. Kühn, R. Adrian, R. Alber, A. Alignier, C. Andrews, J. Bäck, L. Barbaro, D. Beaumont, N. Beenaerts, S. Benham, D. S. Boukal, V. Bretagnolle, E. Camatti, R. Canullo, P. G. Cardoso, B. J. Ens, G. Everaert, V. Evtimova, H. Feuchtmayr, R. García-González, D. G. García, U. Grandin, J. M. Gutowski, L. Hadar, L. Halada, M. Halassy, H. Hummel, K.-L. Huttunen, B. Jaroszewicz, T. C. Jensen, H. Kalivoda, I. K. Schmidt, I. Kröncke, R. Leinonen, F. Martinho, H. Meesenburg, J. Meyer, S. Minerbi, D. Monteith, B. P. Nikolov, D. Oro, D. Ozoliņš, B. M. Padedda, D. Pallett, M. Pansera, M. Á. Pardal, B. Petriccione, T. Pipan, J. Pöyry, S. M. Schäfer, M. Schaub, S. C. Schneider, A. Skuja, K. Soetaert, G. Sprinĝe, R. Stanchev, J. A. Stockan, S. Stoll, L. Sundqvist, A. Thimonier, G. V. Hoey, G. V. Ryckegem, M. E. Visser, S. Vorhauser, P. Haase, Meta-analysis of multidecadal biodiversity trends in europe, *Nature Communications* 11 (2020). doi:10.1038/s41467-020-17171-y.
- [10] J. Peterseil, S. Geiger, Field Specification for data reporting, Technical Report, European long-term ecosystem, critical zone and socio-ecological systems research infrastructure PLUS, 2020. URL: https://www.lter-europe.net/projects/PLUS/eLTER_TechDoc_TemplateSpecification.pdf.
- [11] S. Zacharias, S. Anttila, J. Bäck, K. Böttcher, U. Mallast, M. Mirtl, M. Schaub, V. Trotsiuk, Discussion paper on eLTER Standard Observations, Technical Report, Deliverable D3.1 EU Horizon 2020 eLTER PLUS Project, Grant agreement No. 871128., 2021.
- [12] D. D. Pooter, W. Appeltans, N. Bailly, S. Bristol, K. Deneudt, M. Eliezer, E. Fujioka, A. Giorgetti, P. Goldstein, M. Lewis, M. Lipizer, K. Mackay, M. Marin, G. Moncoiffé, S. Nikolopoulou, P. Provoost, S. Rauch, A. Roubicek, C. Torres, A. van de Putte, L. Vandepitte, B. Vanhoorne, M. Vinci, N. Wambiji, D. Watts, E. K. Salas, F. Hernandez, Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences, *Biodiversity Data Journal* 5 (2017) e10989. doi:10.3897/bdj.5.e10989.
- [13] A. Devaraju, U. Schindler, T. Dohna, R. Huber, M. Diepenbroek, K. Koop-Jakobsen, FAIRification of PANGAEA datasets: Recent Developments and Lessons Learned, *ESSI2.7: Establishing A Comprehensive Open and FAIR Ecosystem for Solid Earth and Environmental Researchers, Repositories, Publishers, Policy Makers and Funders* 21 (2019).
- [14] A. M. Leadbetter, P. N. Vodden, Semantic linking of complex properties, monitoring processes and facilities in web-based representations of the environment, *International Journal of Digital Earth* 9 (2015) 300–324. doi:10.1080/17538947.2015.1033483.
- [15] M. Stoica, S. Peckham, An ontology blueprint for constructing qualitative and quantitative scientific variables, in: *Proceedings of the ISWC 2018 Posters & Demonstrations, Industry and Blue Sky Ideas Tracks colocated with 17th International Semantic Web Conference (ISWC 2018)*, 2018.
- [16] M. Stoica, S. Peckham, The scientific variables ontology: A blueprint for custom manual and automated creation and alignment of machine-interpretable qualitative and quantitative variable concepts, in: *Modeling the World's Systems Conference*, 2019.
- [17] I. Rosati, C. Bergami, E. Stanca, L. Roselli, P. Tagliolato, A. Oggioni, N. Fiore, A. Pugnetti, A. Zingone, A. Boggero, A. Basset, A thesaurus for phytoplankton trait-based approaches: Development and applicability, *Ecological Informatics* 42 (2017) 129–138. doi:10.1016/j.ecoinf.2017.10.014.