InteroperAble Descriptions of Observable Property Terminology WG (I-ADOPT WG)

Introduction

The research community is creating ever-larger volumes of data to understand phenomena via their observable properties at every scale. Our ability to exploit these data as a common resource is hampered by a lack of interoperability in how we describe observable properties. A large collection of independent terminology[1] resources and tools across research domains and communities has emerged. Their complexity and diversity often overwhelm data managers and users, ironically maintaining barriers to interoperability.

Great progress has already been made in providing machine-readable descriptions of sensors and their observation types through the OGC's Sensor Web Enablement SensorML[2], Observations and Measurements[3] [1] or the W3C's/OGC's Semantic Sensor Network (SSN) ontology[4]. However, "deep metadata" that further contextualizes observations (e.g. methodology, variables, parameters[5]) is typically represented as coarsely qualified classes (e.g. "Procedure" or "Observed property"). 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.

WG CHARTER

The I-ADOPT WG will produce an Interoperability Framework, co-developed by a wide community of terminology experts, for representing observable properties. This effort will have a strong focus on ***observable properties in environmental research*** because it leverages existing efforts to accurately encode what was measured, observed, derived, or computed in relation to the earth systems. But many of the principles it will lean on will be relevant to or connected with other domains. The construction of the framework will be informed by a review of current practices used in the community. Furthermore, the working group will iteratively test and refine 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, in alignment with others. With these, they may leverage their local resources in a collective attempt to represent complex properties observed across the environmental sciences (from marine, atmospheric, and terrestrial Earth sciences, as well as biodiversity). The WG will then seek to synthesize these approaches and share outcomes with other research communities in

[1] Terminology is used as the overarching name for any set of fixed denotations that are used to describe something with the goal to reduce ambiguity and facilitate interoperability. A terminology can range from a simple controlled vocabulary (a simple list of terms) to a complex ontology (formal definitions of terms and their relations semantically expressed in a machine-readable way). This term may also include taxonomies, thesauri, or any other kinds of knowledge organization sources.

[2] https://www.opengeospatial.org/standards/sensorml

[3] https://www.opengeospatial.org/standards/om

[4] https://www.w3.org/TR/vocab-ssn/

[5] An information content entity which is about a property or set of properties that determine the state or behavior of some entity.

order to develop a set of best practice recommendations. Furthermore, it will help mediate between generic observation standards (O&M, OBOE, SOSA/SSNO, SensorML, ..) and current community-led resources, fostering harmonized implementations. Through this effort, FAIRer observable property terminologies will be created, the global effectiveness of tools operating upon them will be improved and their impact increased. The WG will thus strengthen existing collaborations and build new connections between terminology developers and providers, disciplinary experts, and representatives of scientific data user groups.

Core Activities

- Collecting user stories from different stakeholder views, which should be domain-agnostic and rely on real applications of data where an interoperable description of observation and measurement data is needed
- Gathering requirements to enhance the interoperability of terminology used to describe observational data pursued through a series of use cases, which will be developed through the formalisation of the collected user stories into more generic and classified descriptions
- Surveying existing, semantic representations and design patterns of observable properties including descriptions of methodologies, instrumentation, and environmental context used in different environmental domains, focusing on a few widely used variables
- Analysing the overlaps and gaps between these representations on the basis of the requirements of each use case (leveraging the list composed by the W3C review of Sensor and Observations Ontology[6] but also extending to those represented by members of the group, such as SVO, Complex Properties Model, variable treatment in schema.org, including those which we will additionally discover during the survey)
- Building on the above, the development and iterative refinement of a conceptual framework to support interoperability between representations of observable properties across participating terminologies.
- Deploy the interoperability framework in selected linguistic settings and derive recommendations for translations
- Developing local, aligned design patterns across participating terminologies

Value Proposition

Addressing the "I" of FAIR Data Management

The activity proposed by this group will address the I (interoperability) of the FAIR data principles. The Interoperability Framework, a conceptual framework for describing observable properties, will foster the interoperability among cross domain terminologies by providing recommendations to support the systematic expression of the observable properties, thus paving the way for seamless terminology alignment.

[6] https://www.w3.org/2005/Incubator/ssn/wiki/Review_of_Sensor_and_Observations_Ontologies

User benefits

Research communities (e.g., domain scientists, ontology engineers, data providers, data centers, etc.) may gain the following benefits through the deliverables produced by the WG (i.e., a review of existing observation-centric semantic representations and recommendations for terminology harmonization):

- a basis to align semantic products,
- guidance in identifying and adapting relevant observation representations and/or further developing their own standards for data publication and ingestion,
- recommendations for a minimum viable set of elements and relationships to encode complex semantic descriptions of observable properties (footnote: when adopted, these elements will allow data aggregators, such as the Australian National Data Service, to integrate data and metadata from diverse sources, provide a common interface for data discovery, and thus enable data reuse),
- a means for identifying relevant terminologies for the annotation of datasets through the systematic comparison of observation-centric models, and
- access to terminologies that can be leveraged to support semantic searches and improve data curation and dissemination activities on data portals. (footnote: for example, the recommendations may help data curators to decide on the 'mandatory' metadata elements of datasets (concerning observable properties) that should be submitted by data authors).

Engagement with existing work in the area

I-ADOPT WG, which is a task group within the RDA Interest Group on Vocabulary and Semantic Services (VSSIG), engages a number of parties, including Research Infrastructures (RI), data centers, terminology providers, which are already addressing the challenge of representing observable properties (see list of external supporting projects in the appendix) and relevant RDA groups. The WG will ensure that these parties are connecting their domain-specific activities and keep their efforts more synchronised by identifying and adopting measures that enable them to grow more efficiently, sustainably, and collaboratively. Such experts will be directly involved in providing their own use cases and co-developing the interoperability framework. It is primarily the task of the current WG members to actively reach out to their communities and to represent I-ADOPT WG and its work, e.g. in presentations or direct communication. If not engaged directly, the WG will maintain bilateral discussions with these parties.

Directly involved parties:

The **eLTER RI (Integrated European Long-Term Ecosystem, Critical Zone & Socio-Ecological Research Infrastructure)**[7] comprises a wide range of highly instrumented sites focusing on terrestrial, freshwater and transitional water ecosystems on a European scale. eLTER RI is building on the site network of LTER Europe and is currently in the phase of its operational

[7] https://www.lter-europe.net/elter-esfri

implementation. It enables the in-situ and co-located acquisition and gathering of ecosystem characteristics and Essential Variables ranging from biochemistry to biodiversity as well as socio-ecological characteristics of LTSER regions. In addition to providing well documented and maintained datasets, the provision of a semantically well-defined set of terms is the basis for later analysis. It serves as harmonized specification of parameters in the observation and measurement of ecosystem processes. Therefore, the EnvThes (Environmental Thesaurus)[8] a controlled vocabulary, was developed for the LTER-Europe community using the Complex Properties Model [2]. It serves as a unified semantic backbone for the provision of data in the eLTER context. EnvThes is based on the core ontology of SERONTO[9], developed for the AlterNet FP6 project, as well as building on the core concepts of US LTER CV[10].

The **International Long Term Ecological Research (ILTER)[11]** Network, comprised of 44 member networks, inter-alia seeks improved interoperability and consistency of the data archived in its multiple repositories. Toward this goal, members of the ILTER information management community are participating in the development of a multilingual ecological thesaurus to harmonize terms applied to whole datasets. Another challenge facing the ILTER Network in the global context is to annotate ILTER datasets with observable properties to enhance interoperability and discoverability. Currently, only a small set of ILTER datasets show such annotations, and no best practices have been established for this process outside DEIMS-SDR. ILTER semantics specialists will participate in this WG to help articulate best practices for an interoperability framework for observable properties.

As a multidisciplinary ICSU World Data Center, **PANGAEA[12]** comprises more than 140.000 measurement and observation types. To improve the consistency, interoperability, and findability of archived data, PANGAEA links measurement and observation types to existing terminologies like WoRMS, ICES, ChEBI, QUDT, ENVO etc. [3]. For this purpose PANGAEA has developed a terminology catalog as part of its editorial system. More recently, in the context of the GFBio[13] project, a terminology service has been implemented[14], significantly reducing the effort necessary to maintain the bilateral interfaces with the various terminologies.

The British Oceanographic Data Centre - UK Research and Innovation (BODC) hosts the **NERC Vocabulary Server (NVS)[15]** which has successfully served controlled vocabularies to the marine domain for more than 10 years. NVS provides access to lists of standardized terms and thesauri that cover a broad spectrum of disciplines relevant to the oceanographic and wider community. It is published as Linked Data using the World Wide Web Consortium's Simple Knowledge Organization System (SKOS) to represent knowledge in a format understandable by computers. All of the vocabularies are fully versioned and assured (governed) by a dedicated group of experts before publication. The largest of these vocabularies is the ‘BODC Parameter Usage Vocabulary’

8 http://vocabs.ceh.ac.uk/edg/tbl/EnvThes.editor
9 https://www.umweltbundesamt.at/fileadmin/site/daten/Ontologien/SERONTO/SERONTOCore20090205.owl
10 http://vocab.lternet.edu
11 https://www.ilter.network/
12 https://www.pangaea.de/
13 https://www.gfbio.org/
14 https://terminologies.gfbio.org/
15 https://www.bodc.ac.uk/resources/products/web_services/vocab/

(P01)[16] which is used to annotate the fields of data files with the physical properties that the numbers represent. P01 is used to identify what the data value represents, describing precisely what was measured, derived, predicted or estimated. P01 concepts are built from a rigorous, exposed, semantic model that is fed by either internal or external controlled vocabularies. The use of the NVS has been widely adopted in the context of European (SeaDataNet I/SeaDataNet II/SeaDataCloud2, EUROFLEETSv2, and EMODnet projects) and international initiatives (e.g. EU H2020 AtlantOS, Ocean Data Interoperability Project (ODIP), etc.)

The **Scientific Variables Ontology (SVO)**[17] is a framework that enables modular, principled expression, manipulation, and creation of scientific variable concepts [4]. It provides a formal, domain-agnostic upper ontology, a set of machine-readable rules that define how elementary concepts can be combined to create more complex concepts, an extensible lower ontology that declares all of the instances of the concept category classes defined in the upper ontology, a simple search interface that allows casual users to browse the current database using exact term match of both preferred and alternate (synonym) labels stored in the ontology, a SPARQL endpoint, an API for matching freeform strings to concepts in the ontology using information from public linked data catalogs, a concept-based graph search, and the ontology blueprint, and a GUI tool for constructing new variables. SVO was motivated by the need to map a selection of over 13,000 CF Standard Names and NWIS SRS names into the CSDMS standard names (CSN) object-process-quantity (O-P-Q) format for use within the CSDMS earth systems modelling framework. During this effort, the O-P-Q data model was significantly augmented and revised, and ontological design patterns were developed that would allow more flexible, modular, and descriptive encoding of scientific variable concepts. SVO is currently used for semantic mediation within the interdisciplinary, collaborative MINT[18] framework, a system that integrates data and models through automated workflows to assist analysts in predictive modelling of complex systems.

LifeWatch ERIC[19] (LW ERIC) is a European Infrastructure Consortium providing e-Science facilities to scientists seeking to increase our knowledge and deepen our understanding on Biodiversity organization and Ecosystem functions and services in order to address key planetary challenges. LW ERIC provides data tools and services for a large and broad user community dealing with a variety of data types (taxonomic, functional traits, genetic, ecological, biogeographic, environmental and various types of observation data). Essential services supplied by LW ERIC are data curation, long-term data storage and preservation, data analysis and data publication. Data curation includes quality control of data, semantic annotation, data cataloging and the development of terminologies according to international protocols and standards. LW ERIC developed and is extending domain-specific thesauri with a particular focus on functional biodiversity and an ontology (LifeWatch Core Ontology) as a semantic framework for data harmonization and integration. Another important asset is EcoPortal[20], a repository for terminology

[16] http://vocab.nerc.ac.uk/collection/P01/current/
[17] geoscienceonotlogy.org
[18] http://mint-project.info/
[19] http://www.lifewatch.eu
[20] http://ecoportal.lifewatchitaly.eu

resources in the ecological domain supporting the community in the management and alignment of their semantics.

AnaEE[21] is the RI dedicated to experimentation on continental ecosystems and their biodiversity. It provides facilities to experiment in controlled, semi-controlled or open-air environments as well as analytical platforms and tools for data management and modeling. Due to the large heterogeneity of the data produced, the harmonized management and sharing of data is one of the main AnaEE challenges. Based on semantic interoperability of its components and the use of common vocabularies (AnaeeThes thesaurus and an OBOE-based ontology), a distributed Information System (IS) is under development. Discovery (ISO19115/19139 standard) and access portals are fed by triples produced by the semantic annotation of the AnaEE distributed resources (relational databases and modeling platforms). Two pipelines developed within the ENVRIplus context automate the semantic annotation of the distributed IS and the generation of metadata records and data sets. The AnaEE IS team aims at contributing to the fully strategic RDA WG and will take benefit from the expected results for the implementation of data semantic interoperability within and among RIs.

The **DLR-Institute of Data Science** is currently active in standardization efforts for observational data models in the context of citizen science. This includes activities dedicated to enabling interoperability between observed parameters in the context of Citizen Science, e.g. in the OGC Interoperability Experiment[22] and a community of practice for citizen observatories recently established within the EU project **WeObserve**[23] as well as in data standardization activities, e.g. within the Citizen Science COST Action CA15212[24]or the OGC (description of Earth Observation Products).

AquaDiva[25] is a large CRC aiming to study the effect of biodiversity on the subsurface levels in the German Hainich forest. To this end, a number of scientists from different domains, including ecology, biology, and geosciences, are collecting samples covering different aspects and thus generating a large number of heterogeneous data sets. Furthermore, scientists may use unstructured data, such as text to store datasets related information. In order to answer the research questions, datasets generated by different groups of scientists should be integrated and then analyzed. In order to deal with the heterogeneity issue, the AquaDiva CRC is building a semantic layer on top of datasets stored in a common repository. These datasets cover a wide range of aspects and a large number of ontologies is required to cover those aspects. An essential step to building such a conceptual layer is to allow interoperability across this set of ontologies.

21 https://www.anaee.com/

22 https://www.opengeospatial.org/projects/initiatives/citsci-ie

23 http://www.iiasa.ac.at/web/home/research/researchPrograms/EcosystemsServicesandManagement/event/180603-WeObserveCOPsLaunch.html

24 https://www.cs-eu.net/wgs/wg5

25 http://www.aquadiva.uni-jena.de/

SIOS[26] is a regional observing system for long-term measurements in and around Svalbard addressing Earth System Science questions. SIOS integrates the existing distributed observational infrastructure, exchanging discovery metadata via the OAI-PMH protocol serving either GCMD DIF or ISO19115. Harmonization of data across the contributing data centers is ongoing pushing for standardization and is more mature in the physical domain (meteorology and oceanography) than in the biological domain. Controlled vocabularies which are in use include GCMD DIF, OSGeo Cat-Interop, Climate, and Forecast Convention. But development is ongoing for internal vocabularies using SKOS, including collections, institutions, and platforms. SIOS is also working on WIGOS metadata standards for building a station catalog.

Representatives of the **OBO Foundry and Library**[27]**,** a federation of hundreds of production and research ontologies, will contribute refined best practices in highly-expressive semantic research and engineering [5]. The widely-adopted Environment Ontology (ENVO)[28], with its focus on Earth and environmental phenomena, will be the hub of this activity. Its standing interoperation with other OBO ontologies used to represent observations, collection activities, and the representation of information artifacts, as well as its links to United Nations semantic interoperability efforts and data systems, will contribute to the WGs pool of expertise and stakeholders. ENVO [6] will also serve as a link to the UNESCO Global Ocean Observing System's (GOOS) Essential Ocean Variables, the Earth Science Information Partners (ESIP) Federation's Semantic Web for Earth and Environment Technology (SWEET) Ontology[29], and the UN Environment's Sustainable Development Goals Interface Ontology. These, and other linked efforts, all seek property-level interoperability which may be catalyzed through this WG.

CF conventions[30] for climate and forecast metadata were designed to promote the processing and sharing of files created with the netCDF Application Programmer Interface [NetCDF]. The conventions define metadata that provide a definitive description of what the data in each variable represents, and of the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, regridding, and display capabilities. The CF conventions generalize and extend the COARDS conventions. The extensions include metadata that provides a precise definition of each variable via specification of a standard name, describes the vertical locations corresponding to dimensionless vertical coordinate values, and provides the spatial coordinates of non-rectilinear gridded data.

The **International Bio-logging Society's Working Group on Data Standardization**[31] is tasked with addressing the Society's goal "to progress standardization of data protocols used within the bio-logging community, with a view to making databases interoperable." This WG has identified that vocabularies (i.e. the terms used to describe on-animal sensor data that originate from both

26 https://sios-svalbard.org/
27 http://obofoundry.org
28 https://github.com/EnvironmentOntology/envo
29 http://www.sweetontology.org/
30 http://cfconventions.org/standard-names.html
31 https://www.bio-logging.net/#news

sensors—e.g. location coordinates—and biologists—e.g. species names, and their definitions and relationships) as a critical issue we must address to enable data standardization and interoperability. This is a challenge due to numerous sets of terms in wide use by existing bio-logging databases and equipment manufacturers. Another challenge is the interdisciplinary nature of bio-logging, which can encompass marine and terrestrial ecology, earth sciences, biodiversity and all regions of the earth, making most discipline- or region-specific standards unavailable for use by the field as a whole.

The **USA NSF's Arctic Data Center** (ADC)[32] is responsible for archiving all data products produced by Arctic researchers funded under NSF's Polar Programs.

The **USA NSF's DataONE[33]** cyberinfrastructure is a major metadata/data aggregator, offering services that include replication, and federated search across over 3 dozen major earth and environmental science data repositories. DataONE is developing together with ADC a semantic search and annotation framework to improve discoverability and reusability of its data holdings, based on existing best-of-class ontologies such as EnvO, and using a semantic design pattern for observational data, OBOÉ, that is harmonized with the OGC/ISO O&M and OBO Foundry's EQ.

The **Australian Research Data Commons (ARDC)** hosts the Research Vocabularies Australia (RVA), a terminology server that enables users to publish, manage and discover vocabulary from various domains. CSIRO played a major role (through Simon Cox) in developing international standards such as Geography Markup Language and Observations & Measurements. The organization also developed SISSVoc, a RESTful API that provides persistence access to SKOS-based vocabularies.

University of Amsterdam (UvA) leads the development of the ontology of Open Information Linking for environmental science Research Infrastructure (OIL-e[34]) based on the ENVRI Reference Model[35], and the community knowledge base for ENVRI research infrastructures in the context of ENVRI, ENVRIPLUS, and ENVRI-FAIR[36] projects. UvA has been active in modeling network and virtual infrastructures using semantic web technologies, and in applying those semantic models in Virtual Research Environment to enable applications to across different infrastructures [7].

GO FAIR International Support and Coordination Office[37] is a bottom-up initiative sponsored by Dutch, German and French ministries promoting the coordinated reuse of existing standards and technologies as implementations of the FAIR Principles [8] . The mission of GO FAIR is to accelerate community convergence toward a widely used and domain agnostic Internet of FAIR Data and Services (IFDS).

---

[32] https:/arcticdata.io
[33] https://dataone.org
[34] http://oil-e.net/ontology/
[35] https://confluence.egi.eu/display/EC/ENVRI+Reference+Model
[36] http://envri.eu/envri-fair/
[37] https://www.go-fair.org

Interaction with existing RDA groups

I-ADOPT will interact with the following groups, identified during the BoF session at the 13th PM, concerning specific items in its work plan and scope:

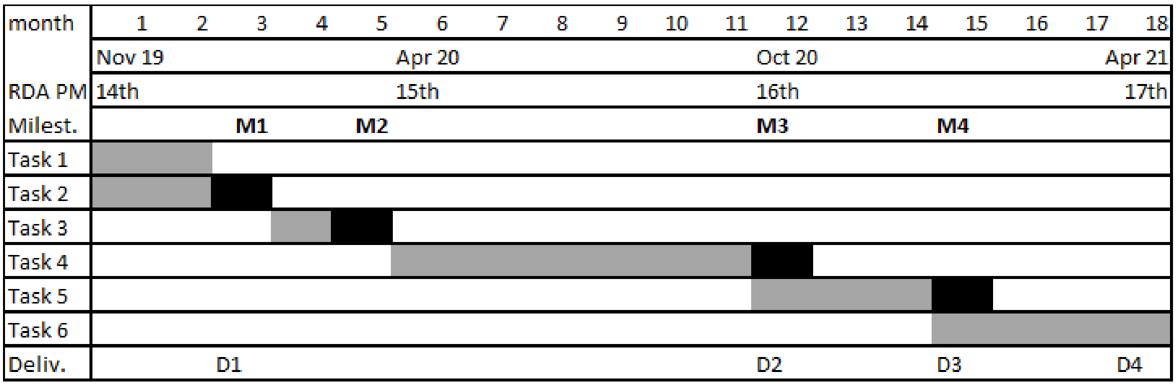
- **Data Fabric IG:** I-ADOPT will evaluate the framework of software components the Data Fabric group is concerned with, and contribute feedback regarding those components that may be essential to deliver property management across disciplines and legal/technical boundaries. I-ADOPT will also contribute to Data Fabric discussions on the cross-disciplinary and global governance of e-infrastructures and their components with respect to observable property management.
- **Data Type Registries WG:** The properties under discussion within I-ADOPT can act as a customer case for the DTR group's foreseen system or federation of registries. The I-ADOPT group may contribute key requirements from its use cases to the DTR WG discussions, and, in particular, feedback on registry governance schemas under discussion in DTR.
- **Data Foundations and Terminology IG:** I-ADOPT WG will consider the DFT IG's maintained vocabularies as further input for the analysis of existing observation representation strategies (Task 4), and then evaluate against the use cases and requirements.
- **Biodiversity Data Integration IG:** This group deals with morphological, genomic, biochemical and anatomical traits, environmental measurements and taxonomic and nomenclatural information of species within the DiSSCo[38] research infrastructure. Like I-ADOPT it seeks to identify existing key resources that can be part of a trans-disciplinary commons it is clear that there is a considerable overlap which has to be aligned between the two groups.
- **Small Unmanned Aircraft Systems' Data IG:** Introducing a new technology, this group's main objective is to develop and adopt standards and best practices as early as possible. I-ADOPT's input will be valuable to steer this group towards interoperability.
- **Marine Data Harmonization IG:** This group is focusing on marine data harmonization developing a common global framework for the management of marine data and I-ADOPT will make sure our efforts are aligned.
- **IG Physical Samples and Collections in the Research Data Ecosystem:** This group focuses on persistence identification, representation and discovery of physical samples and sample collections. The review and recommendations that will be produced by I-ADOPT are relevant to the group, as one of the activities of interest the group is to identify common concepts describing samples and related resources.
- **Education and Training on handling of research data IG:** To increase the interoperability and adaptability of training materials the group wants to collaborate with I-ADOPT to achieve a broader standardization of terms.

38 https://dissco.eu/

- **Research Data Repository Interoperability WG:** The WG is interested in the outcome of I-ADOPT because harmonizing parameter descriptions in data packages could increase re-usability, raise acceptance and consolidate data package exchange within domains.
- **Using Schema.org and enriching metadata to enable/boost FAIRness on research resource:** As this group is looking into discoverability of research data via schema.org, I-ADOPT's Interoperability Framework may be an important contribution to the group's requirements. Additionally, their outcomes will help align our framework to schema.org extensions.

Workplan

The WG's work plan constitutes a progressively more refined process of alignment between the representation of observable properties across participating environmental terminologies. During this process, new perspectives, synthesis, and recommendations will be formulated and delivered to the wider community. At the close of the WG, a new level of interoperability between distinct, but aligned, resources will be realized, and form the basis for more bold initiatives in the future.



Tasks:

Task 1: Collecting user stories

Collect user stories of observational research from environmental parties involved as input to Task 3. These should refer to applications of data where an interoperable description of observation and measurement data is needed. Formalize user stories into a set of use cases (November–December 2019) - leverage P14 for community input.

Task 2: Survey observation-centric terminologies

Conduct a survey, and synthesize the results of related surveys conducted by other groups (like the Polar Vocabulary survey[39]), of observation-centric top-level representations (in most cases ontologies) and domain-specific terminologies (mostly taxonomies, thesauri, controlled lists for

[39] https://tinyurl.com/PolarVocabs

units, methods, instruments, etc.) in use. If necessary, prepare a questionnaire to be sent out Q1 2020, (6 weeks of circulation).

This will contribute to building a fine-grained collection list of semantic representations of observable properties (as input for task 4) and to a catalog of domain specific terminologies.

Task 3: Deriving use case requirements

Derive requirements from the use cases that have been formalized in Task 1 as input to Tasks 4 and 5 (February–April 2020) - finalize with a discussion at P15

Task 4: Analysis of existing representation strategies

The application of existing representation strategies to the observable properties in each use case will be analyzed in order to identify gaps, overlaps, strengths and weaknesses. The survey and a brief literature review[40] will be used as source material. The findings will be used to instantiate the use cases and define/refine the conceptual framework. (May–October 2020)

Task 5: Developing the Interoperability Framework

Define the Interoperability Framework for observable properties based on the outcomes of Tasks 3 and 4 (see Value proposition). As a subtask, potential issues in deploying the framework in selected linguistic settings (e.g., Chinese) will be tested to derive recommendations for translations (November 2010–January 2021).

Task 6: Development of Design Patterns

In line with the Interoperability Framework, WG participants will develop local, aligned design patterns across several participating terminology resources for selected use cases (February–April 2021)

Milestones:

M1 User stories and terminologies collected (2020-01-31)

M2 Collection of requirements completed (2020-03-31)

M3 Overlap and gap analysis completed (2020-10-31)

M4 Consensus achieved on the Interoperability Framework (2021-01-31)

Deliverables:

D1: Catalog of domain-specific terminologies in use in environmental research. This inventory will also help identifying missing terminologies. (2020-01-31)

[40] Including known resources like ODM 2, SDMX/DDI [9], DATS#, OM#, EFO#, SWEET#, QUDT# and others

D2: A review and synthesis report on the expressiveness of existing top-level representation strategies based on the comparison of descriptions of the selected use cases. (2020-12-31)

D3: A report on the Interoperability Framework of observable property terminology in environmental research (2021-01-31)

D4: Guidelines on best practices on the implementation and use of the Interoperability Framework by terminology developers outside the WG (2021-04-30)

WGs mode:

- meet twice a month (one US-friendly and one Australian-friendly telco) to coordinate actions, one extra telco for WG chairs and technical group
- organize remote hackathons to work on common tasks and accelerate progress
- at least 4 physical meetings at the RDA Plenary Meetings (14th, 15th, 16th, 17th) with one (or more) co-located VoCamps at the Plenary Meetings
- use of GitHub for issues tracking and solving

Consensus and conflict management:

I-ADOPT WG plans to develop consensus by encouraging and ensuring participation in meetings, both virtual and physical. Collaborative work will be done ahead of conference calls and active discussion will be encouraged during calls in order to draw from a wide range of ideas. This should ensure that team members are heard and acknowledged and that the development is not biased toward the thinking of a few members. For critical issues the WG will vote using a weighted multi-vote approach whereby each member can cast up to three votes, from most to least preferred. This approach ensures members think thoroughly about the available options and decreases chances of stalemate. The co-chairs are committed to keeping the development on track and within scope. The WG will adopt a framework for ranking priorities that helps the team deciding which option will best qualify priorities. We will also undertake regular retrospectives (such as after completing a milestone) where we assess our performance and feedback improvements into the next phase.

Adoption plan:

Use Case providers:

The parties involved in the WG's use cases (eLTER, ILTER, PANGAEA, LifeWatch, AnaEE, BODC, SIOS, AquaDiva, WeObserve and potentially some others from the ENVRI-FAIR context) are expected adopters of the WG recommendations. These are either institutions, terminology providers or scientists (essential variables - ocean, biodiversity) that will initially evaluate their existing implementations according to the evaluation criteria/metrics developed by this WG. Accordingly, they will ensure they develop plans to implement the recommendations produced to enable interoperability of measurements of observable properties across the participants.

LW ERIC

Developers and science domain experts of LW ERIC will participate in this RDA WG to provide but also acquire insights as terminology providers and users, test the applicability of the WG results using a use case on Phytoplankton, and potentially adopt the best practices and recommendations defined from this WG in the EcoPortal and Catalogue of Data of LW RI.

SVO (MINT)

The SVO team will share their ontology blueprint and custom variable construction tools with the working group. We will also provide input based on our previous experience with our efforts to date in generalizing design patterns of scientific variable concepts. We will work with the group to restructure and adapt our upper ontology as necessary to make it compatible with the output from this WG.

International Bio-logging Society Working Group

At least one member of the International Bio-logging Society's (IBioLS) Working Group on Data Standardization's steering committee will participate in this RDA WG to provide input based on the IBioLS WG's experience, learn from the expertise of the RDA WG and use output and recommendations from the RDA WG in the development of proposed bio-logging standards. This WG member can also share both draft and final WG results with database managers and data providers within IBioLS membership to encourage feedback and adoption in the field of bio-logging.

ADC & DataONE

Developers of the ADC and the DataONE platform will participate in this RDA WG to ensure greater interoperability of their platform's approaches relative to vocabulary construction and semantic expression.

GO FAIR International Support and Coordination Office

Although the primary vehicle by which GO FAIR operates has been voluntary Implementation Networks[41], formal IN status is not a prerequisite in order to leverage GO FAIR support and coordination. As such, the I-ADOPT WG can already consider tapping ongoing GO FAIR resources and activities as needed, including (but not limited to): (1) participating with 20-30 INs in the FAIR Implementation Matrix, an Open registry of community "declarations" around the FAIR implementation choices; Matrix development working group; (2) ongoing Metadata for Machines (M4M) workshops; (3) The FAIR Funder Implementation Study that makes a registry of reusable community-specific, machine-actionable metadata available for funding agencies when composing requirements for new research calls and for data stewards when composing data stewardship plans (but also for any other organization like I-ADOPT who wish to harmonize reuse of FAIR metadata standards and terminologies). Likewise, the inventory, assessment (overlaps and gaps), and roadmapping of standards in I-ADOPT to achieve more automated forms of data interoperability are of fundamental interest to GO FAIR. Thus, GFISCO will be keen observers / participants of I-ADOPT activities to better inform our own methods of coordination and support

[41] https://www.go-fair.org/implementation-networks/overview/

for communities aiming to increase FAIRness in general, and data/metadata Interoperation in particular. Of special interest to GO FAIR will be keeping overview among the broad spectrum of domains (including humanities, economics, government and policy), and targeting applications that most efficiently cross conceptual and disciplinary boundaries not only between but also outside Earth and environmental topics.

University of Amsterdam (UvA)

UvA will actively join the working group and exploit the output to improve the OIL-e ontology and the development of VRE solutions.

All members:

Members of the I-ADOPT WG will use their networks and involvement in projects to extend the community engagement and promote the use and adoption of the WG's output (e.g. ESIP, GOOS, ENVRI-FAIR, FAIRsFAIR, GoFAIR, FAIRmetrics, MBON, GBON, GODAN). The WG also seeks to extend its network towards African research organisations.

Literature:

1. S. J. D. Cox (2017): Ontology for observations and sampling features, with alignments to existing models, Semantic Web, vol. 8, no. 3, pp. 453–470.
2. A. M. Leadbetter & P. N. Vodden (2016): Semantic linking of complex properties, monitoring processes and facilities in web-based representations of the environment, International Journal of Digital Earth, 9:3, 300-324, DOI: 10.1080/17538947.2015.1033483
3. M. Diepenbroek et al (2017): Terminology supported archiving and publication of environmental science data in PANGAEA, Journal of Biotechnology 261 (2017) 177-186
4. M. Stoica & S.Peckham (2018): An Ontology Blueprint for Constructing Qualitative and Quantitative Scientific Variables, ISWC 2018, Monterey CA.
5. D. Osumi-Sutherland et al. (2017): Dead simple OWL design patterns, Journal of Biomedical Semantics (2017) 8: 18, https://doi.org/10.1186/s13326-017-0126-0
6. P. L. Buttigieg, E. Pafilis, S. E. Lewis, M. P. Schildhauer, R. L. Walls, and C. J. Mungall (2016): The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation, Journal of Biomedical Semantics, vol. 7, p. 57, Sep. 2016.
7. Liao, X., Bottelier, J. and Zhao, Z., 2019. A Column Styled Composable Schema Matcher for Semantic Data-Types. Data Science Journal, 18(1), p.25. DOI:http://doi.org/10.5334/dsj-2019-025
8. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).
9. T. Bosch, R. Cyganiak, A. Gregory, und J. Wackerow, „DDI-RDF Discovery Vocabulary: A Metadata Vocabulary for Documenting Research and Survey Data.“, in *LDOW*, 2013.

Initial Membership:

Co-chairs:

Name	*Gender*	*Associated Network*	*Project/Service/Terminology*	*Domain*	*Institution*	*Country*
Barbara Magagna	female	eLTER, ENVRI	EnvThes ENVRI-FAIR	ecosystem biodiversity	Environment Agency Austria	Austria
Michael Diepenbroek	male		PANGAEA	environmental sciences	MARUM, University of Bremen	Germany
Gwenaelle Moncoiffe	female	SeaDataNet, ODIP	NVS	marine	BODC, National Oceanography Centre	UK
Maria Stoica	female		SVO (MINT)	geosciences	University of Colorado	US

Appendix

Members:

Name	*Gender*	*Associated Network*	*Project/Service/Terminology*	*Domain*	*Institution*	*Country*
Alexandra Kokkinaki	female	SeaDataNet, ODIP	NVS	marine	BODC, National Oceanography Centre	UK
Louise Darroch	female	SeaDataNet, ODIP	NVS	marine	BODC, National Oceanography Centre	UK
Adam Leadbetter	male	SeaDataNet, ODIP	NVS Complex Properties Model	marine	Marine Institute, Oranmore	Ireland
John Watkins	male	eLTER	CEH semantic web editor	environmental informatics	Centre for Ecology & Hydrolgy	UK
Mike Brown	male	eLTER	CEH semantic web editor	environmental informatics	Centre for Ecology & Hydrolgy	UK
Pier Luigi Buttigieg	male	OBO Foundry ESIP GOOS	ENVO, ECOCORE, SDGIO	environmental sciences and societal goals	Alfred - Wegener-Institut (AWI)	Germany
Kristin Vanderbilt	female	ILTER US-LTER Environmental Data Initiative		data science	Florida International University	USA
Margaret O'Brien	female	US-LTER	EDI	data science	University of California	US

Appendix

Name	*Gender*	*Associated Network*	*Project/Service/Terminology*	*Domain*	*Institution*	*Country*
Alison Pamment	female	CF conventions	CF conventions	Environmental data analysis	Science and Technology Facilities Council	UK
Melanie Sattler	female		PANGAEA Data center	marine biochemistry, ontologies	MARUM, University of Bremen	Germany
Anusuriya Devaraju	female		PANGAEA Data center	data science	MARUM, University of Bremen	Germany
Robert Huber	male	ENVRI	PANGAEA Data center	marine and environmental sciences	MARUM, University of Bremen	Germany
Nicola Fiore	male	LifeWatch ERIC	EcoPortal LifeWatch thesauri	biodiversity, ecosystem	University of Salento	Italy
Ilaria Rosati	female	LifeWatch ERIC	LifeWatch thesauri	biodiversity, ecosystem	University of Salento	Italy
Christian Pichot	male	AnaEE	AnaEE Thesurus	ecosystem		France
Mark Schildhauer	male	NSF DataONE, Arctic Data Center (ADC)	OBOE	environmental informatics	NCEAS, UCSB	USA
Friederike Klan	female	OGC Interoperability Experiment Citizen Science COST Action AquaDiva	WeObserve	data science	DLR-Institute of Data Science	Germany

Appendix

Name	*Gender*	*Associated Network*	*Project/Service/Terminology*	*Domain*	*Institution*	*Country*
Sirko Schindler	male	OGC Interoperability Experiment Citizen Science COST Action	WeObserve	data science	DLR-Institute of Data Science	Germany
Alsayed Algergawy	male	AquaDiva		data science	University Jena	Germany
Sarah Davidson	female	International Bio-logging Society's WG on Data Standardization	Movebank	biology	Max Planck Institute of Animal Behavior	Germany
Simon Cox	male	ARDC	O&M	environmental informatics	CSIRO	Australia
Øystein Godøy	male	SIOS ADC Vocabularies and Semantics Working Group Global Cryosphere Watch (GCW)		environmental sciences	Norwegian Meterological Insitute	Norway
Javad Chamanara	male				L3S Research Center	Germany
Markus Stocker	male	ENVRI			Technische Informationsbibliothek (TIB)	Germany
Ruth Duerr	female	ADC Vocabularies and Semantics Working Group GCW		data science, earth sciences	Ronin Institute	USA

Name	*Gender*	*Associated Network*	*Project/Service/Terminology*	*Domain*	*Institution*	*Country*
Alsayed Algergawy	male	AquaDiva		data science	FSU Jena	Germany
Adam Shepherd	male	Biological and Chemical Oceanography Data Management Office (BCO-DMO)		marine	Woods Hole Oceanographic Institution WHOI)	USA
John Graybeal	male	BioPortal	BioPortal	environmental informatics	Stanford	USA
Francesco Benincasa	male	ACTRIS	data center	data science	Barcelona Supercomputing Center	Spain
Sara Basart	female	ACTRIS	data center	data science	Barcelona Supercomputing Center	Spain
Zhiming Zhao	male	ENVRI	OIL-E	data science	University of Amsterdam	Netherlands
Tobias Weigel	male	CODATA Data Science Journal EUDAT B2HANDLE DTR WG		data science	Detutsches Klimarechenzentrum (DRZ)	Germany
Naoel Karam	female	GFBio	GFBio Terminology Server	data science	Frauenhofer-Gesellschaft	Germany
Erik Schultes	male	GO FAIR	FAIR Matrix	Genomics	GO FAIR International Support and Coordination Office	Netherlands

Appendix

Name	*Gender*	*Associated Network*	*Project/Service/Terminology*	*Domain*	*Institution*	*Country*
Gerard Coen	male	FAIRsFAIR		data science	Data Archiving and Networked Services	Netherlands
Meng Ji	female			linguist (Chinese, Japanese, Spanish)	School of Languages and Cultures, University of Syndey	Australia
Adriana Pagano	female			linguist (Spanish, Portuguese)	Universidade Federal de Minas Gerais, Brasil	Brazil
Isa Elegbede	male	Marine Biodiversity Observation Network World (GBON) Meteorological Organization (MBON)		data science	SAEIO global Brandenburg University of Technology	Nigeria Germany
Falilu Adekumbi	male	Marine Biodiversity Observation Network World (GBON)		marine	Nigerian Institute of Oceanography and Marine Research	Nigeria

On the web page of the I-ADOPT WG[42] are listed more members which have not been participating in the teleconferences so far, therefore they are not yet included in the table above.

[42] https://rd-alliance.org/groups/harmonizing-fair-descriptions-observational-data-wg

Supporting external projects

Group member	*Institution*	*Project or Service name*	*Project type or objectives*	*Project duration*	*Comments*
Alison Pamment	National Centre for Atmospheric Science (UK)	Vocabulary management for CF Metadata Conventions	Facilitating community agreement of terminology (and associated definitions) within climate and earth system science communities	Open ended	CF conventions are particularly important to the climate modelling community, e.g. for CMIP project, but are also becoming more widely used for observations (e.g. satellite EO, radar, oceanography)
Tom Gulbransen	Battelle	National Ecological Observatory Network (NEON)	Operational observatory with efforts dedicated to increasing data products' FAIRness	decades	
Maria Stoica	University of Colorado	Scientific Variables Ontology	Creating an ontology blueprint, including universal categories and design patterns, and interpretation tools for creating machine-readable, atomistic representations of scientific variables.	at least 2 more years	SVO provides design patterns for physical observables and mathematical abstractions, including models and operations. Terminology is currently mostly covering the geosciences, but ongoing research efforts are focused on sourcing more terminology from open knowledge bases such as Wiktionary and Wikidata.
Barbara Magagna	Environmental Agency Austria	ENVRIfair	The high-impact ambition of ENVRI-FAIR is to establish the technical preconditions for the successful implementation of a virtual, federated machine-to-machine interface to access environmental data and services provided by the	more than 3 years	harmonising vocabularies for observations across the domains represented in ENVRIfair

			contributing ENVRIs, called the ENVRI-hub.		
Group member	*Institution*	*Project or Service name*	*Project type or objectives*	*Project duration*	*Comments*
Barbara Magagna	Environmental Agency Austria	eLTER Plus	Integrating the operations of advanced cross-disciplinary and transdisciplinary research as reflected by the eLTER Sites and eLTSER Platforms experimental design and Standard Observation framework	at least next 5 years	A major aim is to foster the development and facilitate the uptake of harmonised observation protocols and methodologies as basis for Findable, Accessible, Interoperable & Re-usable (FAIR) data on Essential Ecosystem Variables (EEVs). Barbara Magagna is the coordinator of EnvThes, the thesaurus of eLTER, where observable properties are described semantically
Barbara Magagna	Environmental Agency Austria	eLTER PPP	The eLTER Preparatory Phase Project (eLTER PPP) represents a decisive phase in the process towards achieving the long-term strategic goals of the eLTER RI in the mid- and long term	at least next 5 years	eLTER PPP will develop and disseminate novel methods for integrating ecosystem, critical zone, and socio-ecological research at pan-European, in situ research sites, and to integrate in situ measurements and remote sensing to develop and disseminate novel methods for integrating ecosystem, critical zone, and socio-ecological observations.
Barbara Magagna	Environmental Agency Austria	PhD	Interoperability of observable properies	more than 3 years	Main objective is to prove how interoperable terminologies can leverage data discovery and integraion
Nicola Fiore, Ilaria Rosati	University of Salento	Lifewatch Plus		At least 3 years	Semantic uptake into the LifeWatch community
Gwen Moncoiffé, Alexandra Kokkinaki, Lou Darroch	British Oceanographic Data Centre	NERC Vocabulary Server and Services governance	Provision of vocabulary services to the oceanographic community.	open ended	Supported by multiple sources of funding for maintenance, growth and collaboration (including UK NERC national capability core funding, SeaDataCloud, Envri-FAIR,

					EMODnet chemistry, Ocean Data Interoperability Program ODIP)
Group member	*Institution*	*Project or Service name*	*Project type or objectives*	*Project duration*	*Comments*
Gwen Moncoiffé, Alexandra Kokkinaki, Lou Darroch, Mike Brown, John Watkins, Alison Pamment, Petra Ten Hoopen	NERC Data Centres (BODC, CEH, CEDA, BAS, BGS) - UK	NERC Environmental Data System integration	Develop greater integration and semantic interoperability between the NERC's domain specific data centres: marine (NOC BODC), atmospheric and earth observations (CEDA), terrestrial and freshwater (CEH EIDC), geoscience (BGS NGDC), and polar and cryosphere (BAS PDC)	at least next two years	
Gwen Moncoiffé, Alexandra Kokkinaki, Lou Darroch, Mike Brown, John Watkins	BODC+CEH - UK	eLTER and Envri-FAIR	Collaborate on developing shared or interoperable sematic resources	next 5 years	
Alsayed Algergawy	Informatics, FSU Jena	AquaDiva	create the AquaDiva ontology to cover the domain of participating projects in CRC. The ontology is used to support semantic integration of collected datasets.	at least next two years	
Michael Diepenbroek, Anusuriya Devaraju,	MARUM, University of Bremen	PANGAEA data center	PANGAEA links measurement and observation types to existing terminologies like		Ensure semantic interoperability with the planned Interopoerability Framework

Melanie Sattler, Robert Huber			WoRMS, ICES, ChEBI, QUDT, ENVO		
Group member	*Institution*	*Project or Service name*	*Project type or objectives*	*Project duration*	*Comments*
Lucia Mona, Markus Stocker	CNR - Italy, TIB	ACTRIS DC (Aerosol Clouds Trace gases Reserach Infrastructure datacentre)	Node of the data centre of a big and diverse European research infrastructure working on increasing data products' FAIRness	3 decades	Lucia: Working on lidar and aerosol remote sensing observations. Currently using CF convention but not all our needs are accomadated by CF so that we would need in the future to apply for new names into those vocabulary. Markus: Happy to contribute the new particle formation event use case, likely a description and summary of this: https://github.com/markusstocker/carbon-workshop
John Graybeal	Stanford University	OntoPortal Alliance	Increase semantic uptake in science and social research	open-ended (new!)	Use Case: A better model of variable representation could be applicable across multiple ontology repository types.
John Graybeal	Stanford University	CEDAR: Center for Expanded Data Annotation and Retrieval	Improve metadata management in research science, with particular attention to semantically meaningful metadata.	open-ended	Use Case: Generic model for variable representation could be powerful and interoperable cross-domain metadata representation solution.
John Graybeal	Stanford University	Stanford Center for Biomedical Informatics Research	Semantic tools (Protégé, BioPortal, CEDAR) supporting biomedical and general research	open-ended (>15 years)	
John Graybeal	Stanford University	ESIP Federation, Marine Metadata Interoperability project	Community Ontology repository containing semantic resources for earth science	open-ended (>12 years)	Important to recognize other variable representation models and leverage them for this application.

Appendix

Pier Luigi Buttigieg	Alfred Wegener Institute	ESIP Semantic Harmonization Cluster	Interest group on promoting cross resource alignment for federate operations	open-ended (>5 years)	Work in this cluster overlaps the objectives of the WG, and participants intend to align activities
Group member	*Institution*	*Project or Service name*	*Project type or objectives*	*Project duration*	*Comments*
Pier Luigi Buttigieg	Alfred Wegener Institute	OBO Operations Committee	Federated ontology research and development in a production-grade context	open-ended (>10 years)	OBO has a long-standing interest and willingness to provide contributions to alignmnet activities such as i-ADOPT
Simon Cox	CSIRO	TERN	TERN is Australia’s land ecosystem observatory. This information is standardised, integrated and transformed into model-ready data, enabling researchers to discern and interpret changes in land ecosystems.	open-ended (>10 years)	As consultant ensure semantic interoperability between the approaches
Kristin Vanderbilt	US LTER Network	US LTER Controlled Vocabulary	Create thesaurus for annotating datasets generated by the US LTER	open ended	
Frederike Klan, Sirko Schindler	DLR-Institute of Data Science	WeObserve	H2020 Coordination and Support Action (CSA) which tackles three key challenges that Citizens Observatories (COs) face: awareness, acceptabili ty and sustainability.	At least one more year	Ensure semantic interoperability with the Interoperability Framework