





## Common framework to integrate the environmental dimension into SDG indicators

The multi-disciplinary nature of large-scale monitoring creates a complex collaborative environment characterised by a broad and varied knowledge-base. Ensuring that entities in this environment are clearly represented on a semantic level can greatly enhance the gathering, retrieval, querying, handling, sharing, analysis, and reuse of data by diverse systems and communities. The discipline of ontology can be used to achieve this goal.

**An ontology** attempts to systematically identify, in simple and precise terms, what the component entities in a domain of interest are and how they relate to one another. This is done by creating a defined and logically-structured vocabulary comprising classes and the relations between them. (See an example in **Figure 1**).

**A fully realised ontology** differs from a glossary, vocabulary (controlled, structured, or otherwise), taxonomy, or thesaurus in several). For example, **classes in ontology represent conceptual rather than textual entities**: the textual representation of a given class is merely a label and alternative labels can be added as synonyms. Class definitions and logical relations to other classes take precedence in identifying their meaning. As long as collaborators agree on the class' position in the conceptual map (see *Figure 1*), they can add and use their own labels while availing of homogenous semantics. Further, **every sub-class inherits all the properties of its super-class**. For example, given a class 'rainforest', the subclass 'tropical rainforest' inherits all the properties of its super-class; however, it is differentiated from other types of rainforests by some property, 'tropical'. This formalism is among several which impose logical constraints on ontological classes which contribute to clear communication both between human and machine agents.

As it would be overly ambitious and vastly cumbersome to model the diverse knowledge underpinning any one of the SDGs targets and indicators with a single ontology managed, there is a need to **distribute the task of modelling** each "orthogonal" (i.e. largely unrelated) domain to several domain-specific expert groups. A workable template for this approach has been established in the life sciences in the form of the OBO Foundry.

**Well-aligned domain ontologies can easily import portions of one another** to create compound concepts that are, instantaneously, linked to all knowledge models involved. To illustrate, consider the environment class 'gut environment'. A class such as 'digestive tract' can be imported from an anatomy ontology such as UBERON and combined with an environment ontology's (e.g. ENVO) concept of an environment determined by a specific material entity to create a new class, 'digestive tract environment'. The knowledge represented in both ontologies would then be linked and exploitable while the concept stands adequately represented. Similarly, concepts such as 'contaminated soil' or 'heavy metal enriched wastewater' can be constructed using ENVO and CHEBI. *Table 3.1* lists a few OBO-Foundry-linked ontologies that are likely to provide good starting points in the development of an application ontology for environmental monitoring. (See the OBO Foundry homepage for more: <http://www.obofoundry.org>).





One key benefit of **ontologies is that they can assist in developing coherent and robust standards which are poised for conversion to machine-readable representations**. Casting knowledge in an ontological form encourages the ‘teasing apart’ of concepts into their empirical parts, which prevents unstructured debate over poorly-defined, inter-domain inconsistencies when they arise. Further, existing standards can be linked to an appropriate ontology and provide the raw material to extend that ontology. Thus, ontology projects with open membership and development models offer official entities an opportunity to embed their standards into future development. In conclusion, ontologies have great potential to enhance multiple facets of monitoring endeavours by clarifying the semantics of these complex undertakings both for human and machine agents.

#### Examples of domain ontologies primarily used in the biomedical sciences

Domain	Ontology	Citation or URI
Chemical entities of biological interest	CHEBI	(Degtyarenko <i>et al.</i> , 2008)
Human disease	DOID	<a href="http://purl.obolibrary.org/obo/doid.owl">http://purl.obolibrary.org/obo/doid.owl</a>
Environments and ecosystems	ENVO	(Buttigieg <i>et al.</i> , 2013)
Phenotypic qualities	PATO	<a href="http://purl.obolibrary.org/obo/pato.owl">http://purl.obolibrary.org/obo/pato.owl</a>
Populations and communities	PCO	(Walls <i>et al.</i> , 2014)
Cross-species anatomy	UBERON	(Mungall <i>et al.</i> , 2012)

#### Examples of candidate vocabularies

Domain	Instance	Concepts
Biodiversity	Global names architecture GBIF	Institutions, Networks Country nodes, Datasets Search and Metrics
	eCat name parser	Taxonomic names
Ecosystem characterisation	LTER	Organizational units, disciplines, events measurements, methods, processes substances, substrates ecosystems, organisms
Environmental law	ECOLEX/FAOLEX	
Hydrology and inland water sciences	CUHASI	Observations Data Model (ODM) Controlled Vocabulary Registry
	Water ML OGC	OG
Oceanography	Rolling Deck to Repository (R2R)	Controlled vocabulary and ontology
Pollution control	US-EPA Terminology Reference System	
Socio-economics	SEDLAC	





### Building integrated indicators

There are numerous inter-linkages and relationships between the various targets and goals and between different environmental, social and economic domains. To arrive at a common understanding and meaning of what the indicators are actually measuring requires a clear analysis of terms, roles, classes and processes as well as a clear description of data flows and statistics.

**Air quality**, especially in cities, is important to the achievement of all 17 SDGs; and in particular to four SDGs (see Figure 2). The overarching SDG objective for air quality can best be achieved through up-to-date assessments of urban emissions, including the estimation of exposures in urban populations and vulnerable groups, and assessments of the short and long-term health impacts. Existing indirect and direct indicators, plus a new design for a global indicator based on an ontology for urban air quality health has been developed. The integrated indicator is based on new global data sources derived from satellites and sensor-web enablement to provide air pollution exposure maps for vulnerable groups in cities.

### Urban air quality health indicator



