UNEP

A Common Approach for developing SDG integrated indicators

At the first intergovernmental negotiations on the post-2015 agenda in January 2015, UN Member States expressed broad support for the 17 goals and multiple targets proposed by the OWG in 2014.

The goals and targets reflect the complexity and variety of the contemporary challenges of sustainable development. The overall package represents an opportunity to deliver the four major innovations of the post-2015 vision, i.e. universality; a full and balanced integration of the three dimensions of sustainable development including governance and peaceful societies; legitimacy, reinforced by an intergovernmental process and the involvement of civil society; and a transformative approach. However, many targets are imprecisely defined, some are less ambitious than existing internationally agreed targets and integration is still not sufficiently captured.

The next step in establishing the post-2015 agenda, is the development and assembly of SDG indicators which can support the level of ambition and transformative nature of the targets and goals, capture the integration of the three dimensions of sustainable development and retain the political balance of the current proposal.

Experience from the MDGs and other processes shows that to be successful, indicators must be SMART, avoid duplication, and be consistent with existing standards and agreements. They should be meaningful, scientifically credible, statistically sound, consistent over time, and sensitive to root causes, drivers and underlying phenomena. They should allow international comparison and be universally applicable. They should be intuitive, intelligible to both negotiators and technical experts and compelling.

The current set of goals and targets include many imprecise terms and definitions. The development, assembly and technical proofing of indicators will require a universal understanding of what precisely each indicator is intended to measure. The definition of such broad terms as *access*, *build*, *ensure and promote* are non-trivial terms and context dependent. They conceal a diversity of processes which need to be rendered distinctly to allow coherent data gathering, analysis, and interpretation. Such an understanding will need to be delivered in a transparent manner through a common framework of definitions (semantics) and relationships (ontologies).

Institutional capacities to collect data from different sources, compile evidence from the relevant knowledge domains and deliver indicators for reporting at the national and global level need to be aligned to avoid duplication and enhance streamlining. Based on the 2014 IEAG Report on the Data Revolution, there is a clear need to align and integrate the data derived from statistics, earth observation, in situ monitoring, laboratory testing, social and economic surveys and big-data from citizen science and social media. This can be achieved through common data-related standards and business processes.

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.

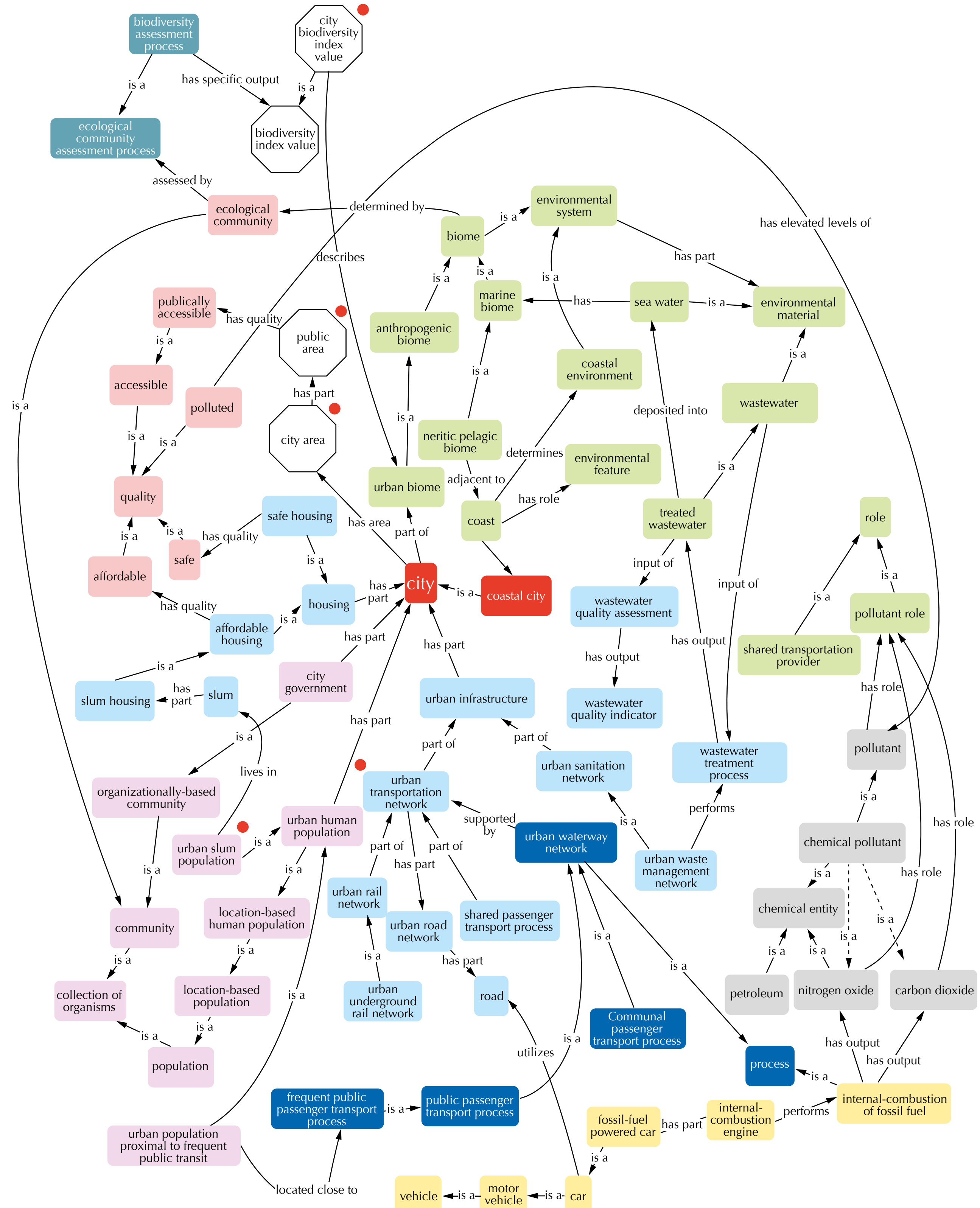
An ontology attempts to systematically identify, in simple and precise terms, what the component entities in a domain of interest are and how they relate to one another. This is done by creating a defined and logically-structured vocabulary comprising classes and the relations between them. (See an example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

Figure 1: A conceptual map of urban entities which pertain to SGD 11 indicators.



Red, filled circles indicate concepts that are directly relevant to one or more indicators linked with SDG 11. Octagons indicate data-level (rather than semantic-level) entities. Urban continuants discussed – pale blue; processes – dark blue; environmental concepts – pale green; qualities – pink; communities and populations – pale purple; roles – pale yellow; chemicals – grey. Subclasses of "city", such as "coastal city", may be linked to the relevant environmental concepts (as shown) and then onto measures of ecosystem status and change relevant to other SGDs (such as those related to marine ecosystems). Dashed arrows indicate inferred relations. Note that this is an illustrative example. Many classes and relations are not in common usage and may not be available in existing ontologies. Further, several relations are simply used for convenience (e.g. "car" utilises "road") and are likely to require deconvolution into a set of more informative classes and relations. These are shown simply to show that classes are readily linkable. Several relations have been omitted in aid of visual clarity.

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 *et al.*, 2008)
Human disease	DOID	http://purl.obolibrary.org/obo/doid.owl
Environments and ecosystems	ENVO	(Buttigieg *et al.*, 2013)
Phenotypic qualities	PATO	http://purl.obolibrary.org/obo/pato.owl
Populations and communities	PCO	(Walls *et al.*, 2014)
Cross-species anatomy	UBERON	(Mungall *et al.*, 2012)

Examples of candidate vocabularies

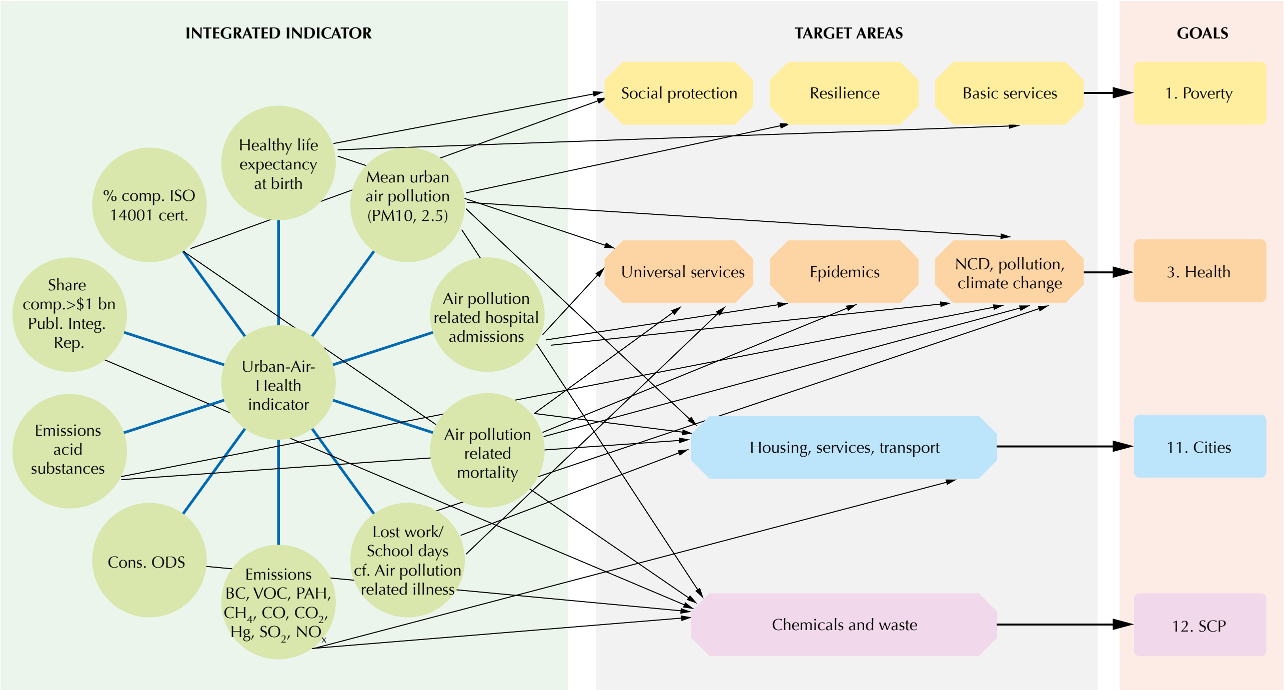
Domain	Instance	Concepts
Biodiversity	Global names architecture GBIF	Institutions, Networks Country nodes, Datasets Search and Metrics
	eCat name parser	Taxonomic names
Ecosystem characterisation	LTER	Organizational units, disciplines, events measurements, methods, processes substances, substrates ecosystems, 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.

A**ir 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



Water quality is relevant to social, environmental and environmental aspects of sustainable development and is closely linked to many of the SDGs. These links are partially reflected by the proposed targets, e.g. the sound management of chemicals proposed in target 12.4 that directly relates to eliminating dumping and minimizing release of hazardous chemicals stated in target 6.3. The development of integrated indicators for water will benefit from using a causal systems framework taking into account functional and contextual relations defined through well-aligned ontologies such as environments (ENVO), location (GAZ), and populations and communities (PCO). A closer collaboration especially with the biodiversity and chemicals and waste communities is necessary, matched in some cases by additional efforts in terms of monitoring coordination. Ontologies could augment global monitoring systems and indicator application and would help underpin the considerable amount of harmonization work that will be needed to underpin the SDGs. Large-scale water quality modelling will also help to bridge the data gaps and support indicator application but requires careful analysis and clear communication of model-related uncertainties.

In addressing **oceans**, a number of issues need to be taken into account more broadly, including the ontology of rights, and benefit sharing. Ocean problems are linked to land-based problems, and experts on both themes need to work together to ensure that these inter-linkages are property reflected in any integrated approach. Connectivity of ecosystem services should also be reflected in the indicators, as well as mainstreaming the value of ocean ecosystem services in national level measures of progress and outcomes. For the ocean goal, there is clearly a need for integrated indicators to monitor Small-Scale Fisheries; Industrial fisheries (capture fisheries and aquaculture); Coastal and marine Development and Areas beyond national jurisdiction (ABNJ), using ontologies to address Decent work - Food security - Profit and income - Inclusion in decision making - Ecosystem health ("ecological foundation").

Current indicators for **common land and natural resources**, pertaining to rangelands, forests, wetlands, and the natural resources above and below ground, often do not adequately capture the complexity of diverse, flexible and periodic tenure rights and regimes, of the important role that reciprocity and non-marketed goods, services and relationships play, or the voice of users themselves. Data are generally patchy, and definitions and methodologies vary across countries. But the sustainable management of common lands and natural resources can provide substantial benefits to indigenous peoples and local communities (IPLC), to the poor in rural areas, to the health of ecosystems, and downstream benefits such as the water supply of cities. It is therefore urgent to measure progress on this issue in a more systematic manner. There are two types of indicators that can be considered; a) those that focus on the existence of IPLC rights, governance, and equitable distribution of benefits, as expressed either in area of land or percentage of people, and disaggregated by gender, ethnicity, age group, land-user group, or other parameters of inequality, both within communities and in comparison with national averages; and b) those that focus on how the rights are exercised and practiced, on the extent of loss or gain of common lands and natural resources, and on how the land and natural resources are used and managed.

The **design of integrated indicators** based on casual linkages captured through the use of ontologies and the semantic web avoids the risk of extensive redundancy in data gathering and ensures that different data and statistics standards can be combined across varying time and spatial scales.