A brief encounter with ontology in the context of environmental monitoring

PL Buttigieg (ORCID: 0000-0002-4366-3088)

Introduction

Large-scale monitoring endeavours generate a suite of informatics challenges ranging from data acquisition and handling to analysis and synthesis. In this complex space, there is a need to ensure meaningful communication between both human and machine agents which must be addressed early on in the design of a monitoring system in order to ensure accuracy and efficiency. Ontology offers a means to address this need and provide semantic clarity in service of a wide range of stakeholders (e.g. Pundt, 2002). This informal note aims to grant the reader some bearing on the nature and potential of ontology in information-rich monitoring environments. No attempt is made to present a complete overview of this field and this text should be considered nothing more than an invitation to delve deeper into the use of ontology in monitoring. Accessible reviews such as that by Madin et al. (2008) are excellent points to start further reading.

While several contemporary definitions of ontology exist (e.g. Smith, 2004), the discipline is often summarised as the "specification of a conceptualization" (Gruber, 1993, 2009). Roughly restated, an ontology attempts to systematically identify, in simple[1] and precise terms, what the component entities in some domain of interest are and how they relate to one another (for illustration, see **Figure 1**). This is done by creating a defined and logically structured vocabulary comprising terms (or, more correctly, classes) and the relations[2] between them. Ontology's systematic character and reliance on logical constructs are highly transferable to the exercise of knowledge modelling in the information sciences and artificial intelligence (see Smith, 2014). In these domains, it can be used to construct a conceptual map of the physical world alongside a semantic representation of information entities (e.g. database attributes and records) which can be recognised and reasoned over[3] by diverse, ontology-enabled systems, regardless of their underlying architecture.

A fully realised ontology differs from a glossary, vocabulary (controlled, structured, or otherwise), taxonomy, or thesaurus in several aspects (see Smith and Welty, 2001 for a more developed overview). Classes in an 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[4]. In stricter

---

[1] Here, "simple" does not imply that the components of an ontology are not complex. It suggests that the representation of the domain should be as 'low-level' (or empirical) as possible.

[2] See Smith et al., 2005 and Hoehndorf et al., 2010 for more detail.

[3] See software such as the ELK reasoner (Kazakov et al., 2012), for more detail.

[4] This is useful in collaborative environments where differing and, at times, conflicting term usage is common. 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.

implementations, every sub-class inherits all the properties of its super-class (unlike, e.g., a taxonomy where properties can be lost towards the ‘leaves’ of a dendrogram). 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 (see Kohler et al., 2011 for more perspective).

In developing an ontology, it is generally efficient to use a ‘bottom-up’ approach by identifying a set of use cases and developing the necessary class structure and relations to model them in a collaborative manner. Grounding development in application is perhaps the most straightforward method to engage a multi-disciplinary community (such as that involved in monitoring) and produce practicable results. However, developers can also draw from the philosophical bases of, for example, realist ontologies (Smith and Ceusters, 2010; Lord and Stevens, 2010) to provide more general forms of guidance.

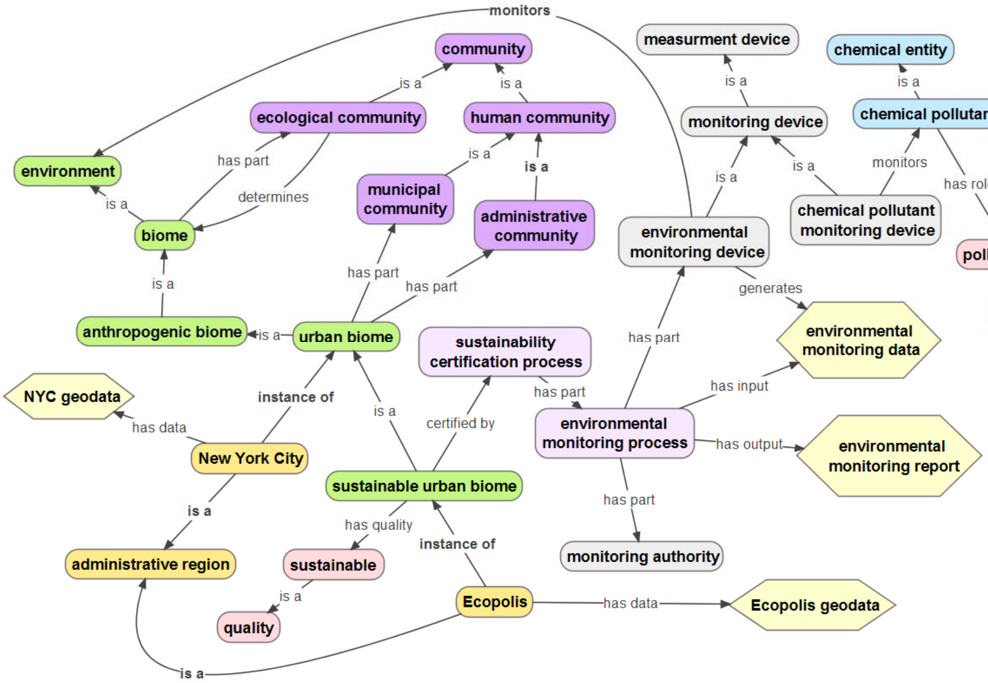

Figure 1: The beginnings of an application ontology for urban environmental monitoring. Classes from environmental (green), chemical (blue), gazetteer (yellow), and community (purple) ontologies have been called upon and other classes created as needed (grey). Both instance-level (e.g. New York City and other objects present in the real world) and class-level (e.g. an urban biome and other categories into which instances can be grouped according to their common properties) entities are shown[5]. The easily-extensible, structured web of classes and relations provide a basis for coherent informatics. Data, documents, or other informational entities (pale yellow hexagons) can be linked into this web for semantically-aware mobilisation by, e.g., database systems. Note that this is merely an illustrative example, despite some of the classes being present in existing ontologies.

---

[5] Note that all classes must have at least one instance to be valid in a practicable ontology. Figure 1 only shows instances of a few classes for illustration.

A federated and collaborative approach to knowledge modelling

As noted above, the multi-disciplinarity of large-scale monitoring endeavours necessitates a collaborative environment due to the breadth of essential expertise involved. It would be overly ambitious and vastly cumbersome to model the diverse knowledge in this environment with a single, monolithic ontology managed by a single authority. The solution is to distribute the tasks of modelling each orthogonal[6] domain to several domain-specific expert groups which follow the same development model and interoperate both on the theoretical and technical level. A workable template for this model has been established in the life sciences and is introduced below.

Following the development of the highly successful Gene Ontology (Ashburner et al., 2000) a family of biologically- and biomedically-oriented ontologies – each focused on a specific domain of knowledge – were federated under the umbrella of the OBO Foundry (Smith et al., 2007) and are developed under the guidance of the Foundry principles (The Principles of the OBO Foundry). OBO Foundry ontologies are linked at their 'upper levels' (i.e. the level at which general conceptualisations of what a "material entity", "process", "role", or "spatial region" is) by the Basic Formal Ontology (Smith, 2013). This philosophical-technical alignment may seem abstract at times, but sets a practicable common framework to develop 'lower level' domain ontologies and facilitates interoperability in numerous ways. For example, 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 'gastrointestinal tract' can be imported from an anatomy ontology such as UBERON (Mungall et al., 2012) and combined with an environment ontology's (e.g. ENVO; Buttigieg et al., 2013) concept of environments determined by a specific material entity to create a new class, 'gastrointestinal 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 (Degtyarenko et al., 2008). **Table 1** lists a few Foundry-linked ontologies that are likely to provide good starting points for an application ontology suite for environmental monitoring.

Table 1: Examples of domain ontologies in the biomedical sciences. See the OBO Foundry homepage for more: http://www.obofoundry.org

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)

---

[6] Here, "orthogonal" suggests that there is minimal conceptual overlap between domain models. Proactively seeking orthogonality prevents duplication of labour and de-standardisation.

Interaction with official standards

Standardised reporting is a cornerstone of collaborative exercises and the design of efficient information systems. For reasons touched on above, ontologies – or, at the very least, an ontologically-flavoured development approach – can assist in developing coherent and robust standards which are poised for conversion to machine-readable representations. Further, casting knowledge in an ontological form encourages the 'teasing apart' of concepts into their (more or less) empirical parts, which prevents unstructured debate over nebulously-defined, inter-domain inconsistencies when they arise.

Existing standards can be linked to an appropriate ontology and provide the raw material to extend that ontology. This may be initiated by automatically mapping ontological classes to terms in existing standards, thereby allowing adopters of that standard to benefit from an underlying semantic model. To illustrate, the Environment Ontology maps to NASA's SWEET resources (Raskin and Pan, 2005) as well as the Alexandria Digital Library's Feature Type Thesaurus (The Alexandria Digital Library Feature Type Thesaurus). Following an automated mapping, several rounds of curation by domain experts, ontology developers, and custodians of the standards in question are required to refine and align the knowledge model. Long-term synchronicity is then ensured when this initial contact matures into a standing collaboration followed by the development of software (e.g. editing tools such as the biologist-centric OBO Edit [Day-Richter et al., 2007]) and services alongside periodic training and workshop events (e.g. Katayama et al., 2014).

Ontologies with open membership and development models, such as those associated with the OBO Foundry and its principles, offer official entities an opportunity to embed experts in their development processes. Guided by the ontology developers, the needs of official organisations can be integrated into their chosen ontology and cross-linkages to other ontologies may be created as needed. While it is almost certain that the chosen domain ontology will contain more classes and relations than needed by any specific organisation, officially-sanctioned and relatively static subsets of a given ontology, nestled within an actively developed superset, may readily be delivered to defined user communities. Multiple solutions to this issue are possible and may be tailored to stakeholder needs. While not trivial to construct, this setting provides an exciting opportunity for cross-fertilisation and co-development and an abundance of development driven by use cases and the needs of diverse communities.

Conclusion & Outlook

Ontology has much to contribute to information-rich monitoring systems. Emphasis on the semantic clarity of concepts and entities can greatly enhance the gathering, retrieval, querying, handling, sharing, analysis, and reuse of data by diverse systems and communities. Significant work remains to be done in integrating ontologies from the social (e.g. Lawson, 2014; Searle, 2006) and political (e.g. Hay, 2009; Jessop, 2014; http://aims.fao.org/aos/geopolitical.owl) sciences as well as the law domain (e.g. Wyner, 2008) with those from the natural and physical sciences. A key step in realising this aim is the facilitation of sustainable collaboration between the organisations which the monitoring system will service and the developers of the ontologies needed for its semantic enhancement. However, successful integration will allow the deep modelling of concepts such as an ‘indigenously regulated ecosystem’ or a ‘multi-nationally administered biodiversity reserve’. This, in turn, will allow efficient mobilisation and delivery of tailored information to users of ontology-enabled monitoring systems which is likely to greatly enhance their reactivity and impact.

References

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25–9. doi:10.1038/75556.

Buttigieg, P. L., Morrison, N., Smith, B., Mungall, C. J., and Lewis, S. E. (2013). The environment ontology: contextualising biological and biomedical entities. *J Biomed Semant* 4, 43. doi:10.1186/2041-1480-4-43.

Day-Richter, J., Harris, M. A., Haendel, M., and Lewis, S. (2007). OBO-Edit--an ontology editor for biologists. *Bioinformatics* 23, 2198–2200. doi:10.1093/bioinformatics/btm112.

Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., and Ashburner, M. (2008). ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 36, D344–50. doi:10.1093/nar/gkm791.

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowl Creat Diff Util* 5, 199–220.

Gruber, T. R. (2009). Ontology. *Encyclopedia of Database Systems*.

Hay, C. (2009). “Political Ontology,” in *The Oxford Handbook of Contextual Political Analysis*, eds. R. E. Goodin and C. Tilly (Oxford University Press). doi:10.1093/oxfordhb/9780199270439.003.0004.

Hoehndorf, R., Oellrich, A., Dumontier, M., Kelso, J., Rebholz-Schuhmann, D., and Herre, H. (2010). Relations as patterns: bridging the gap between OBO and OWL. *BMC bioinformatics* 11, 441. doi:10.1186/1471-2105-11-441.

Jessop, B. (2014). Towards a political ontology of state power: a comment on Colin Hay’s article. *The British journal of sociology* 65, 481–6. doi:10.1111/1468-4446.12087.

Katayama, T., Wilkinson, M. D., Aoki-Kinoshita, K. F., Kawashima, S., Yamamoto, Y., Yamaguchi, A., Okamoto, S., Kawano, S., Kim, J.-D., Wang, Y., et al. (2014). *BioHackathon series in 2011 and 2012: penetration of ontology and linked data in life science domains*. doi:10.1186/2041-1480-5-5.

Kazakov, Y., Krötzsch, M., and Simancík, F. (2012). ELK Reasoner: Architecture and Evaluation. in *Proceedings of the {OWL} Reasoner Evaluation Workshop (ORE’12)*, eds. I. Horrocks, M. Yatskevich, and E. Jimenez-Ruiz (CEUR-WS.org).

Kohler, S., Bauer, S., Mungall, C. J., Carletti, G., Smith, C. L., Schofield, P., Gkoutos, G. V, and Robinson, P. N. (2011). Improving ontologies by automatic reasoning and evaluation of logical definitions. *BMC Bioinformatics* 12, 418.

Lawson, T. (2014). “A Conception of Social Ontology,” in *Social Ontology and Modern Economics*, ed. S. Pratten (Oxon: Routledge), 1–28.

Lord, P., and Stevens, R. (2010). Adding a little reality to building ontologies for biology. *PLoS ONE* 5, e12258. doi:10.1371/journal.pone.0012258.

Madin, J. S., Bowers, S., Schildhauer, M. P., and Jones, M. B. (2008). Advancing ecological research with ontologies. *Trends in ecology & evolution* 23, 159–68. doi:10.1016/j.tree.2007.11.007.

Mungall, C. J., Torniai, C., Gkoutos, G. V, Lewis, S. E., and Haendel, M. A. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome Biol* 13, R5. doi:10.1186/gb-2012-13-1-r5.

Pundt, H. (2002). Domain ontologies for data sharing–an example from environmental monitoring using field GIS. *Comput Geosci* 28, 95–102. doi:10.1016/S0098-3004(01)00018-8.

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

Searle, J. R. (2006). Social ontology: Some basic principles. *Anthropological Theory* 6, 12–29. doi:10.1177/1463499606061731.

Smith, B. (2013). Basic Formal Ontology 2.0: Draft Specification and User's Guide. Available at: http://bfo.googlecode.com/svn/trunk/docs/bfo2-reference/BFO2-Reference.docx.

Smith, B. (2004). "Ontology," in *The Blackwell Guide to the Philosophy of Computing and Information*, ed. L. Floridi (Oxford: Blackwell Publishing Ltd.), 155–66.

Smith, B. (2014). "The Relevance of Philosophical Ontology to Information and Computer Science," in *Philosophy, Computing and Information Science*, eds. R. Hagengruber and U. Riss (Chatto and Pickering), 75–83.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., et al. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 25, 1251–5. doi:10.1038/nbt1346.

Smith, B., and Ceusters, W. (2010). Ontological realism : A methodology for coordinated evolution of scientific ontologies. *Applied Ontology* 5, 139–188. doi:10.3233/AO-2010-0079.

Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. L., and Rosse, C. (2005). Relations in biomedical ontologies. *Genome Biol* 6, R46. doi:10.1186/gb-2005-6-5-r46.

Smith, B., and Welty, C. (2001). Ontology : Towards a New Synthesis. in *Proceedings of the International Conference on Formal Ontology in Information Systems (FOIS2001)* (ACM Press).

The Alexandria Digital Library Feature Type Thesaurus Available at: http://www.alexandria.ucsb.edu/gazetteer//FeatureTypes/FTT2HTM/.

The Principles of the OBO Foundry Available at: http://www.obofoundry.org/crit.shtml.

Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., Bowers, S., Buttigieg, P. L., Davies, N., Endresen, D., et al. (2014). Semantics in Support of Biodiversity Knowledge Discovery: An Introduction to the Biological Collections Ontology and Related Ontologies. *PLoS ONE* 9, e89606. doi:10.1371/journal.pone.0089606.

Wyner, A. (2008). An ontology in OWL for legal case-based reasoning. *Artificial Intelligence and Law* 16, 361–387. doi:10.1007/s10506-008-9070-8.