

Article (refereed) - postprint

Leadbetter, Adam M.; Vodden, Peter N. 2016. **Semantic linking of complex properties, monitoring processes and facilities in web-based representations of the environment.**

© 2015 Taylor & Francis

This version available <http://nora.nerc.ac.uk/510770/>

NERC has developed NORA to enable users to access research outputs wholly or partially funded by NERC. Copyright and other rights for material on this site are retained by the rights owners. Users should read the terms and conditions of use of this material at <http://nora.nerc.ac.uk/policies.html#access>

This is an Accepted Manuscript of an article published by Taylor & Francis Group in *International Journal of Digital Earth*, March 2016, available online:
<http://www.tandfonline.com/10.1080/17538947.2015.1033483>

Contact CEH NORA team at
noraceh@ceh.ac.uk

The NERC and CEH trademarks and logos ('the Trademarks') are registered trademarks of NERC in the UK and other countries, and may not be used without the prior written consent of the Trademark owner.

Semantic linking of complex properties, monitoring processes and facilities in Web-based representations of the environment.

Abstract: Where a virtual representation of the Earth must contain data values observed within the physical Earth system, data models are required which allow the integration of data across the silos of various Earth and environmental sciences domains. Creating a mapping between the well-defined terminologies of these silos is a stubborn problem. This paper presents a generalised ontology for use within Web 3.0 services, which builds on European Commission spatial data infrastructure models. The presented ontology acknowledges that there are many complexities to the description of environmental properties which can be observed within the physical Earth system. The ontology is shown to be flexible and robust enough to describe concepts drawn from a range of Earth science disciplines, including ecology, geochemistry, hydrology and oceanography. This paper also demonstrates the alignment and compatibility of the ontology with existing systems and shows applications in which the ontology may be deployed.

Keywords: Observable properties; Monitoring properties; Environmental monitoring facilities; Semantic web; Geospatial data integration

Introduction

Within the realm of Digital Earth research, Craglia *et al.* (2012) outlined challenges for the next generation of Digital Earth systems on, amongst others, a scientific basis and on the basis of a web of sensors. Where sensors proliferate, deployed for scientific research applications, producing a coherent and comprehensive virtual representation of the Earth system relies on methods and models to link across the data silos of the various environmental and Earth Science domains. At the same time these cross cutting models must retain sufficient representation of the concepts that are meaningful to the data originator, be that professional research scientist, autonomous machine or “citizen scientist”. In many cases, those domains have well developed local terminology and concepts specific to their domain - but mapping from one domain to another remains a stubborn problem (Diviacco, et al., 2014).

Attempts to directly map between two sets of complex concepts often involve significant effort and create a large maintenance overhead. There are also the risks associated with an individual projecting their own understanding, or meaning, on to a set of relationships created between such complex concepts (Diviacco, 2015). The effort/cost escalates geometrically when three or more domains are added to the mapping, and therefore a framework to guide such efforts should be sought.

In their analysis, Craglia *et al.* note that Spatial Data Infrastructure promotes the knowledge base of spatially referenced scientific data and policy making on regional levels. To achieve this, frameworks such as the INSPIRE Directive (EC, 2007) make it possible to encode and transmit data using agreed concepts and vocabularies. However, the INSPIRE data specifications do not provide a strong governance over the content of domain-specific vocabularies. For example, the Land Cover Data Specification (INSPIRE, 2013a) states:

“The data specification does not prescribe or recommend any particular land cover nomenclature for use in INSPIRE. There is a multitude of different ways to describe land cover. This is partly due to the wide range of aspects of the environment embraced by land cover, but also due to the many different uses of land cover data. There is only one "real world" but many different descriptions of this world depending on the aims, methodology and terminology of the observer.”

Indeed, at the time of writing the INSPIRE Registry (<http://inspire.ec.europa.eu/registry/>) contains only terms useful for very high-level discovery of data, and not for the full description of data from, say, a real-time stream from an instrument deployed in a coastal ocean. As an alternative to direct mapping, the approach considered within this paper is based upon breaking down the complex concepts into “atomic concepts” and identifying where the same atomic concepts are present in different domains, following the approach in Weinberger’s (2002) Unified Theory of the Web of “small pieces loosely joined”.

The intent with this approach is not to build “yet another ontology of everything” - rather, it is to reuse and adapt extant models and instances from existing resources to define a minimal core ontology that can form the basis of many solutions. The ontologies described within this paper are freely available on the World Wide Web, for the use of other parties.

The Environmental Monitoring Facilities Data Specification (INSPIRE, 2013b) provides an important pattern which can be described as “The observation of a phenomenon at a feature of interest by a specified process”. Or as the data specification states:

“The class Observing Capability is modelled to serve the need that a measurement regime can be described without providing the observed or measured value itself.”

Both Observation and Observing Capability are consequently associated with this pattern.

Building around this pattern, and the existing resources of the ISO Observation and Measurements conceptual schema (O&M) (Cox, 2010) and the complex properties extension to O&M (INSPIRE, 2013c) this paper will show that it is possible to create a model for the atomisation of complex environmental properties which is practical for implementation. This paper presents an application of the approach, using a simple “overlay” of classes and predicates from the established World Wide Web Consortium PROV ontology (Lebo et al., 2013) to semantically connect the complex properties recorded in datasets to the environmental monitoring facilities which generated the data.

Software demonstrating the approach gives discovery-level functionality (discovering datasets which contain relevant observations) across the domains of terrestrial and freshwater ecology and hydrology, geochemistry, and oceanography. Such software offers the potential to inform environmental issues such as cycles of nitrogen or carbon within the environment, or acidification of the oceans.

The complex property model can also be used to describe actual data, or expose an underlying semantic model in well-known vocabularies such as the SeaDataNet Parameter

Usage Vocabulary used in oceanography (Schaap and Lowry, 2010), or the CEH Analytical Services Thesaurus used in chemical analysis (Wright et al., 2014).

Conceptual Models

The conceptual models created, used and adapted in this approach are self-contained, yet work together to offer a powerful framework for exploiting environmental data and metadata resources. The models comprise Complex Properties, Monitoring Properties and Environmental Monitoring Facilities.

Diagram Notation

Diagrammatic representations of the conceptual models used in this paper are based on:

- Unified Modelling Language (UML, ISO/IEC 19505-1:2012)
- Graph-based Data Modelling (RDF 1.1 Concepts and Abstract Syntax W3C Recommendation 25 February 2014)

The semiotics of these diagrams is summarised in Figure 1.

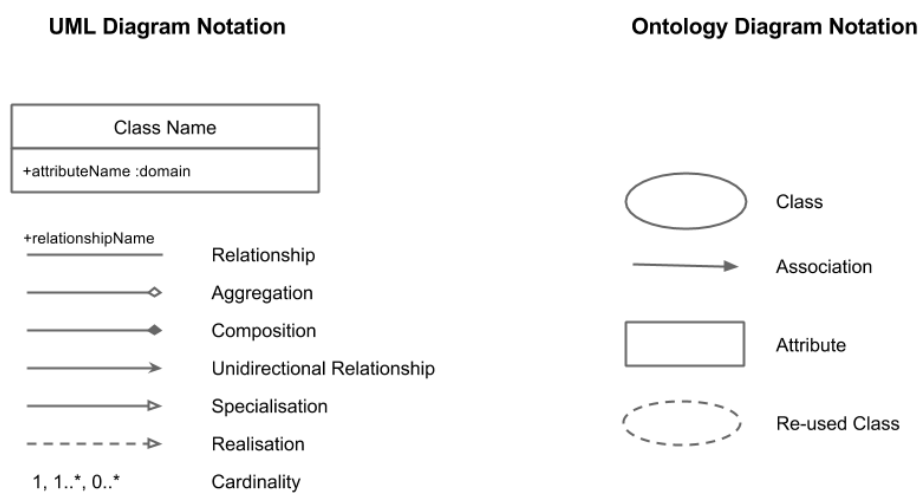


Figure 1. A key to the semiotics of the Unified Modelling Language (UML) diagram notation and ontology diagram notation used in the figures throughout this paper.

The Complex Properties Model

The Complex Properties model is based upon INSPIRE extensions to Observations & Measurements (O&M) (INSPIRE, 2013c) providing a logical model represented graphically in Figure 2.

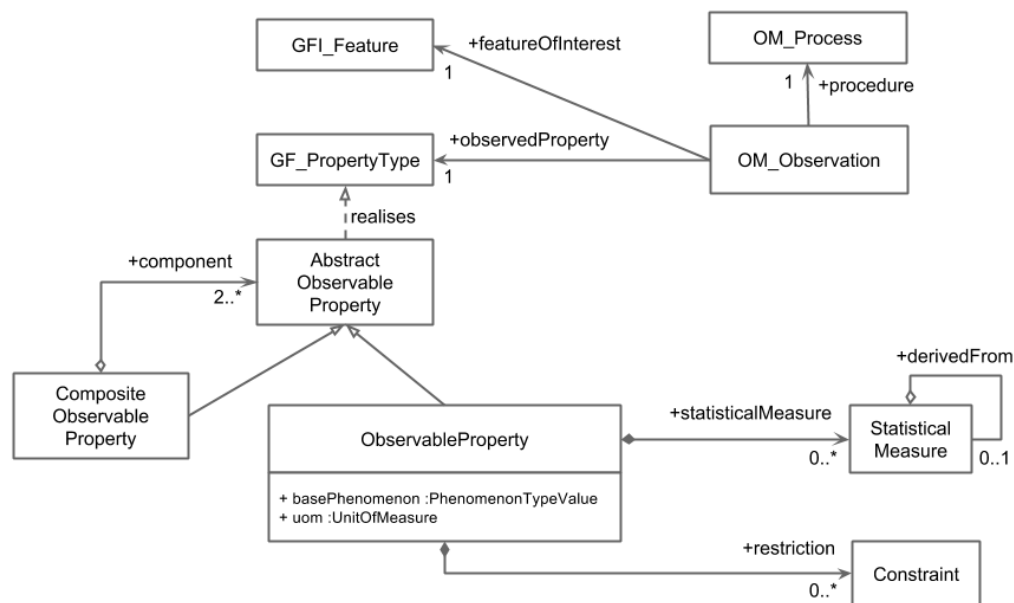


Figure 2. Extract from the UML model of the INSPIRE extensions to the ISO Observations & Measurements model.

The Complex Properties Model (CPM) presented here has been evolved from this original INSPIRE specification through iterative testing involving analysis within this model of pre-existing simple single-string lexical representations of complex real-world environmental properties.

Analysis

In this original UML specification, the domain of the attribute Base Phenomenon is a code list (Phenomenon Type Value) which is intended to be extended and specified for a particular

domain (INSPIRE, 2013c). The allowed values are defined by data providers, and can be at any level, as indicated by the code lists quoted in Table 1.

Table 1. Example Phenomenon Type Value code lists showing the range of granularity which exists in the original INSPIRE specification.

| Code List | From |
|---|---|
| Climate and Forecast Standard Names Value | INSPIRE Generic Conceptual Model, version 3.4 [DS-D2.5] |
| Profile Element Parameter Name Value | INSPIRE Data specification on Soil [DS-D2.8.III.3] |
| Soil Derived Object Parameter Name Value | INSPIRE Data specification on Soil [DS-D2.8.III.3] |
| Soil Profile Parameter Name Value | INSPIRE Data specification on Soil [DS-D2.8.III.3] |
| Soil Site Parameter Name Value | INSPIRE Data specification on Soil [DS-D2.8.III.3] |
| EU Air Quality Reference Component Value | INSPIRE Data specification on Atmospheric Conditions and Meteorological Geographical Features [DS-D2.8.III.13-14] |
| British Oceanographic Data Centre (BODC) Parameter Usage Vocabulary Value | INSPIRE Data specification on Oceanographic Geographical Features [DS-D2.8.III.15] |

A limitation of this approach is that the phenomenon types can themselves be complex concepts which are not easily mapped across domains, as shown in Table 2.

Table 2. Example Phenomenon Types showing the complexity of the lexical labels in the code lists from the INSPIRE specifications, and highlighting the difficulty in cross-domain mapping inherent in this approach.

| Source Vocabulary | Term |
|---|--|
| BODC Parameter Usage Vocabulary | “Concentration of phosphate {PO ₄ } per unit volume of the water body [dissolved plus reactive particulate phase] by colorimetric autoanalysis” |
| Climate Forecasting (CF) Standard Names Table | “surface upwelling radiance per unit wavelength in air emerging from sea water” |
| Centre for Ecology & Hydrology (CEH) Analytical Services Thesaurus (CAST) | “area weighted stream flow” |

For practical purposes, Base Phenomenon from the UML must be mapped to at least two OWL classes: Object Of Interest and Property.

Splitting the Base Phenomenon (e.g. “concentration of carbon”) enables identification of the constituent concepts. This enables mapping of a property such as “concentration” regardless of whether it is measured of carbon or nitrogen. The Object Of Interest (e.g. the substance carbon) can be identified authoritatively (see Ontology Alignment section below) and is independent of the observability of a characteristic (i.e. concentration).

Object Of Interest and Property are mandatory. These basic components may optionally be supplemented by three concepts from existing O&M, and an additional concept, Matrix. The Matrix is needed because in many situations the Object Of Interest is embedded, dissolved or otherwise entailed within something else, and must first be extracted or separated - for example the Matrix might be “stream sediment”, the actual Object Of Interest might be “Nitrogen”. Both these facts represent relevant environmental context to the observation. An Object Of Interest which has direct observability (e.g. a water body) does not have an associated Matrix.

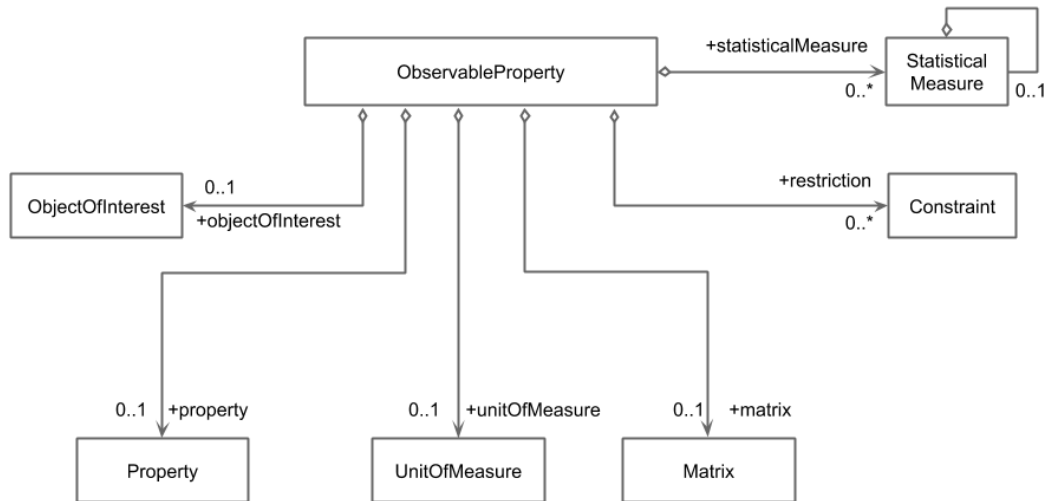


Figure 3. The adapted UML model highlighting the class Observable Property.

The rationale for this adaptation (Figure 3) is based upon the authors' belief that the existing INSPIRE complex properties model remains in some respects too abstract for the purposes of implementation. Other models (Observable Properties, <http://environment.data.gov.au/def/op>; Quantities, Units, Dimensions and Types, <http://qudt.org/>; Long-Term Ecosystem Research Network, <http://www.lter-europe.net>; SeaDataNet, <http://www.seadatanet.org/>) exist too close to individual domains to be easily mapped, or are not mature enough to be re-used - therefore there is a need for semantic mediation between the abstract and the specific which will lend itself to re-use across many domains, whilst allowing users the amenity of their preferred domain-specific terminology (as recommended by Leadbetter et al., 2015).

The Monitoring Properties model

Monitoring Properties is a model developed for the purpose of associating instances in the domain of Complex Properties (as described in the previous section) to instances in the

domain of Environmental Monitoring Facilities (INSPIRE, 2013b).

The purpose of such linking is to enable specific environmental issues to be addressed by exploiting metadata rather than data. The case for exploiting metadata rather than the actual data derives from the fact that terrestrial environmental data is often highly heterogeneous, with few agreed schemas upon which to base wide-reaching analysis.

Monitoring Properties evolved as a response to the challenge of the use case "Where have we measured X?" - a problem which can be approached at a data discovery level by asking questions of the metadata. The storage volumes involved in metadata are typically orders of magnitude smaller than the volumes of the data they describe, and consequently it is relatively easier and less costly to carry out semantic mark-up of the metadata rather than the data. In this sense, the Monitoring Properties model is an application of Complex Properties in the context of the discovery of environmental monitoring datasets.

Although Monitoring Properties was developed directly as an OWL ontology, it has been reverse-engineered into UML for consistency of documentation within this paper (Figure 4).

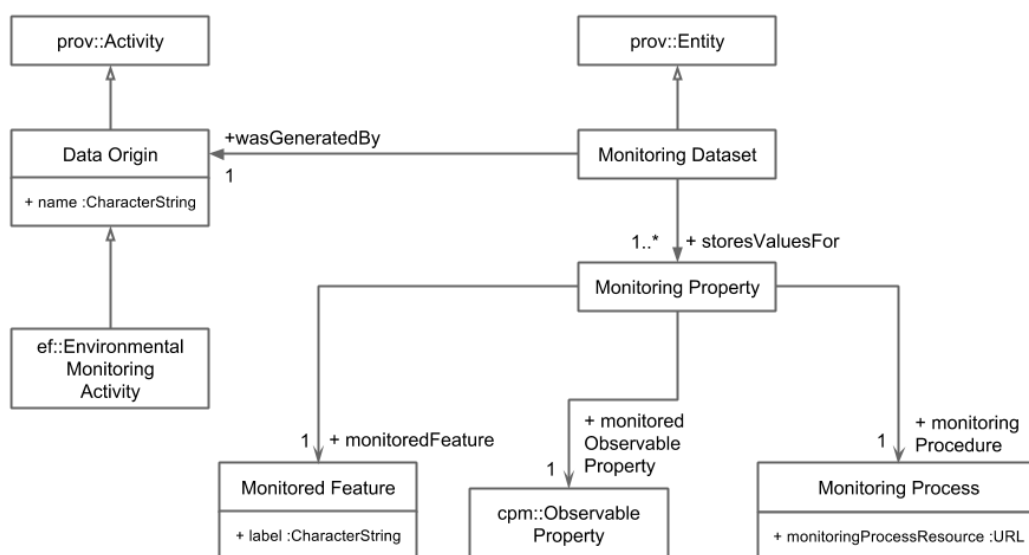


Figure 4. A UML model for the Monitoring Properties model

Monitoring Datasets represent a subset of all datasets described in some metadata catalogue (to which the model is being applied) and consequently are semantically equivalent to dataset records as described in DCAT (Maali and Erickson, 2014). Such equivalence is however, not specified in the Monitoring Properties ontology and the identifier of a Monitoring Dataset can be any URI. Instances will typically be identified by the URL of a dataset record in a metadata catalogue. This class is a subtype of Entity as defined in the PROV-O ontology (Lebo, Sahoo and McGuinness, 2013).

An association is required to make the connection between a dataset and its originating activity. The well-established PROV data model (Moreau and Missier, 2013) conveniently describes just such an association between a document and its origin as the "was generated by" relationship. In order for Monitoring Properties to inherit and use this concept, it is necessary to regard the Monitoring Dataset as a PROV "Entity", and to regard the Data Origin as a PROV "Activity". Instances can then be connected meaningfully via "Monitoring Dataset was generated by Environmental Monitoring Activity".

A Monitored Feature is not a device, but is focusses observability, for example:

- “United Kingdom”.
- “Plynlimon Catchment, Wales”.
- “Irish Sea”.

Note that Monitored Features can form a hierarchy, for example:

- “Site #37”
 - “on the River Thames”

Monitored Features are usually recognizable because they carry Observable Properties and have associated geometry (point location, polygon boundary, bounding box, etc.) (INSPIRE, 2013c).

When linking for the purposes of dataset discovery, we are most often describing reasonably high level features such as "the UK river network" or "the UK land surface".

Monitoring Processes refer to the circumstances of the Environmental Monitoring Activity, examples instances of this class are:

- “The Environmental Change Network protocol for surface water chemistry and quality”.
- “G-BASE sample collection and analytical techniques”
- “AMT18 Above water radiance measurement processing procedures”

If a laboratory was involved in deriving the measurements, the laboratory methodology should be included with the actual data, rather than here with the monitoring metadata (INSPIRE, 2013b).

The Environmental Monitoring Facilities Model

The INSPIRE Environmental Monitoring Facilities (EMF) data specification (INSPIRE, 2013b) provides a generic model which can be used across various domains. It is designed to integrate with Observations & Measurements, and describes (inter alia) the core classes that are employed in environmental monitoring as shown in Figure 5.

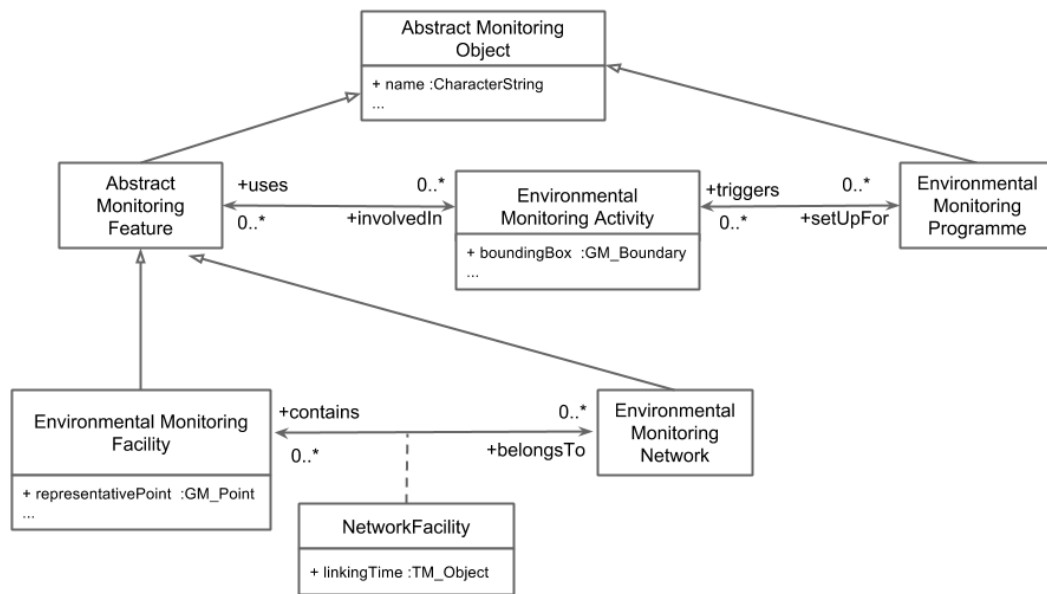


Figure 5. Extract from the INSPIRE Environmental Monitoring Facilities UML Model.

Analysis

In this context the EMF model provides all of the requisite semantics - with the notable exception being that activities in the EMF model are treated as a linking concept between programmes and facilities. Activities have some attribution, but those attributes do not include anything that could be treated as a name or otherwise human-readable label for the activity. In a discovery level software application, such names are essential to the utility of any human interface. We overcome this weakness in the EMF model by affording a name attribute to the Monitoring Properties concept of Data Origin, which becomes a supertype of Environmental Monitoring Activity. All of our activities inherit the attributes of the supertype, and can therefore have names attached.

It should be noted that activities recorded in the UK Environmental Observation Framework (EOF) catalogue (<http://www.ukeof.org.uk/>) have the same problem, which its designers resolved by a similar method - i.e. by implementing a single supertype "above" the

concepts of programme, activity, facility and network. Users of the UK EOF catalogue can consequently identify and differentiate activities by name.

RDF/OWL Implementation

For the concepts derived from existing Unified Modelling Language (UML) models, the UML descriptions of individual conceptual models outlined above have been translated into Web Ontology Language (OWL) documents following the approach outlined by Cox (2013). The following basic rules apply in this translation:

- A UML Package becomes an OWL Ontology
- A UML Class becomes an OWL Class with UML Specialisations modelled as OWL Sub-classes
- UML attributes and association properties become Resource Description Framework (RDF) Properties. In UML, property names have a local scope (i.e. only the class to which they are assigned is aware of them). Within these UML-to-OWL translation rules there is the distinct possibility of RDF Properties becoming homonyms; therefore care must be with naming the RDF Properties.

The detailed mappings between UML and OWL are presented as Appendix A, see the online supplementary material to this paper. As noted by Cox (2013), there are two philosophical approaches available to these sorts of translation. The first is to assume that the UML represents a canonical worldview and the UML is strictly represented in the resulting OWL. In this case the RDF properties are well defined to reflect the classes to which the attributes in the UML are bound. The second approach is to assume that the UML represents a worldview which is but an analysis of some deeper, underlying model. In this approach the use of open world assumptions familiar to traditional RDF/OWL modelling activities may be used in place of the closed world assumptions of the strongly coupled UML model.

In the OWL implementations presented below, we have taken a mixture of the two approaches. In some instances through applying the logical models to pre-existing real-world definitions of observations made in the field we have refined the logical models (e.g. the Complex Properties Model and the Environmental Monitoring Facilities classes). In the case of the Monitoring Properties class we have allowed the RDF/OWL open world assumptions to be the primary focus of our model. A discussion of the appropriateness of this is presented in the Ontology Alignment section below.

In the following sections, the following XML namespaces shown in Table 3 are used extensively.

It is desirable to re-use pre-existing controlled vocabularies published using the Simple Knowledge Organisation System (SKOS) as instances of the classes defined in the Complex Properties Model, see the Ontology Alignment section below. In order to achieve this, classes defined in the Complex Property Model ontology are also defined as instances of SKOS Concepts (Figure 6). For example,

```
cpm:ObservableProperty rdf:type owl:Class.
```

```
cpm:ObservableProperty rdfs:subClassOf skos:Concept.
```

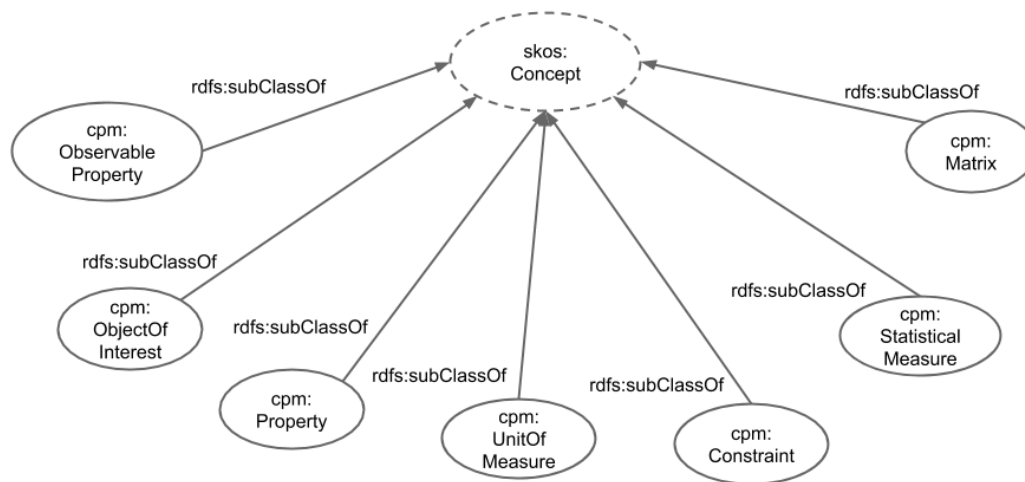


Figure 6. An overview of the Complex Properties Model class showing use of the SKOS Concept class in order to allow the re-use of terms from controlled vocabularies.

This allows, for example, an instance of the `cpm:UnitOfMeasure` class for “degrees Kelvin” to simply be a reference to the SKOS Concept at <http://vocab.nerc.ac.uk/collection/P06/current/UPKA/> which represents “degrees Kelvin” within a controlled vocabulary. This has the benefit of SKOS-aware software being able to navigate the Concept definitions without needing to be aware of the ontology definitions, and that OWL-aware applications can navigate the ontology without needing to be aware of the hierarchies of SKOS Concepts. Similarly, each property has been defined with its domain and range in order to constrain the classes to which a property can be attached, and each property has also been modelled as a subproperty of the “related” property from SKOS in order to facilitate the use of SKOS Concepts as instances of the Complex Properties Model classes. An example of this is:

`cpm:statisticalMeasure rdfs:type owl:ObjectProperty.`

cpm:statisticalMeasure rdfs:subPropertyOf skos:related.

cpm:statisticalMeasure rdfs:domain cpm:ObservableProperty.

cpm:statisticalMeasure rdfs:range cpm:StatisticalMeasure.

Complex Properties

The base of the Complex Properties ontology (<http://purl.org/voc/cpm>) consists of the following classes (Figure 7):

- Abstract Observable Property
- Constraint
- Matrix
- Object of Interest
- Property
- Statistical Measure.

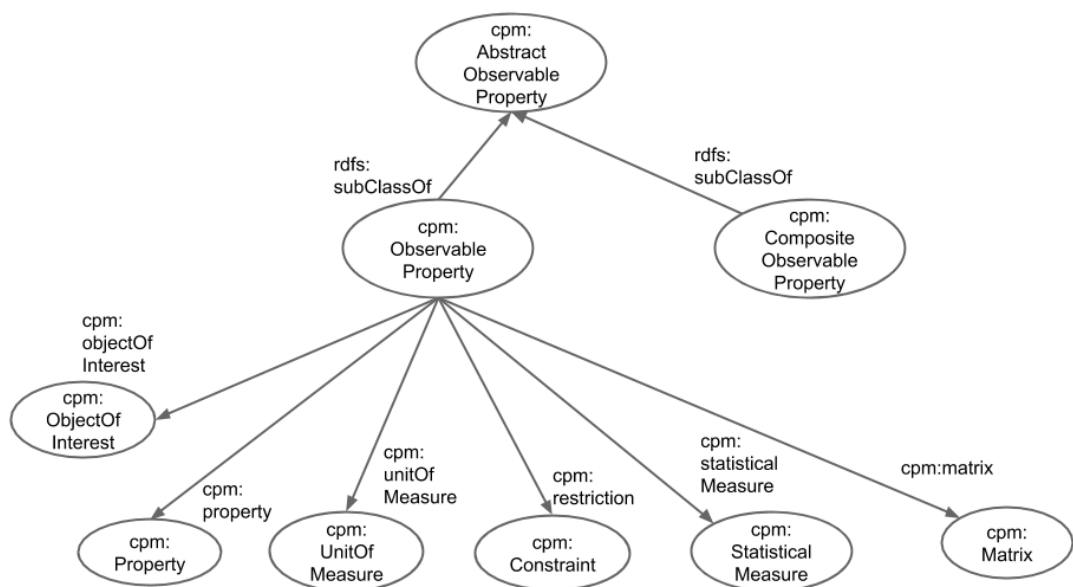


Figure 7. Overview of the Complex Properties Ontology (<http://purl.org/voc/cpm>)

The UML stereotypes of “Composite Observable Property” and “Observable Property” are modelled within the OWL ontology as subclasses of “Abstract Observable Property” such that

```
cpm:CompositeObservableProperty rdfs:subClassOf
    cpm:AbstractObservableProperty.
```

and

```
cpm:ObservableProperty rdfs:subClassOf cpm:AbstractObservableProperty.
```

Similarly, substance, taxon and phenomenon, which are specialisations of “Object of Interest”, are modelled thus:

```
cpm:Substance rdfs:subClassOf cpm:ObjectOfInterest.
```

```
cpm:Taxon rdfs:subClassOf cpm:ObjectOfInterest.
```

and

```
cpm:Phenomenon rdfs:subClassOf cpm:ObjectOfInterest.
```

Monitoring Properties

The Monitoring Properties ontology (<http://purl.org/voc/mp>) consists of the following classes (Figure 8):

- Monitoring Dataset
- Data Origin
- Monitored Property
- Monitored Feature

- Monitored Observable Property
- Monitoring Procedure

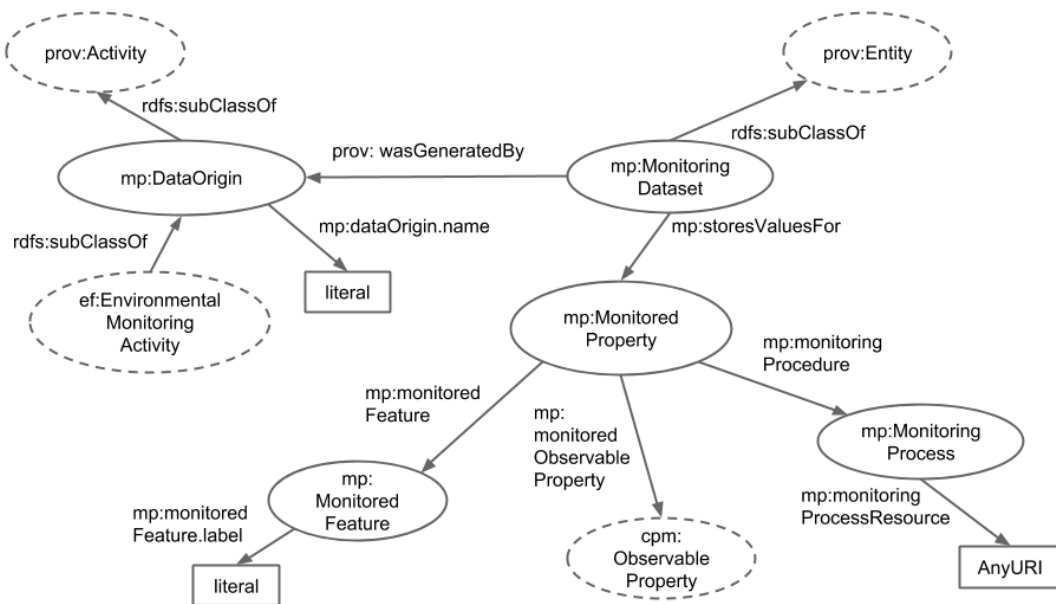


Figure 8. Overview of the Monitoring Properties Ontology (<http://purl.org/voc/mp>)

Monitoring Dataset is abstract class which can be implemented in a variety of ways, but typically, its role would be fulfilled by any URI that represents a dataset - for example, the URI of a standard spatial metadata record in a catalogue. It is modelled as a subclass of `prov:Entity` from the PROV-O ontology, placing any instances of Monitoring Dataset within the domain of the predicate `prov:wasGeneratedBy` which is employed in order to relate a dataset to its originating activity.

```
mp:MonitoringDataset rdf:type prov:Entity.
```

```
mp:MonitoringDataset prov:wasGeneratedBy mp:DataOrigin.
```

```
mp:MonitoringDataset mp:storesValuesFor mp:MonitoredProperty.
```

Data Origin is modelled as a subclass of `prov:Activity`, placing the class Data Origin within the range of the predicate `prov:wasGeneratedBy`. By also declaring DataOrigin to be a

superclass of `ef:EnvironmentalMonitoringActivity`, we enable activities from the Environmental Monitoring Facilities ontology to act as the origin of a dataset:

```
mp:DataOrigin rdfs:subClass prov:Activity.
```

```
ef:EnvironmentalMonitoringFacility rdfs:subClass mp:DataOrigin.
```

Individuals of the class `Monitored Property` represent a unique combination of an `Observable Property` being measured of a specific `Monitored Feature` by means of a specific `Monitoring Process`. The range of the predicate `mp:monitoredObservableProperty` is modelled as `cpm:ObservableProperty`. In this way, `Monitoring Properties` becomes an implementation pattern for the `Complex Properties` ontology which can be used to create a semantic link between `Complex Properties` and `Environmental Monitoring Facilities`. The RDF of this modelling is:

```
mp:MonitoredProperty mp:monitoredFeature mp:MonitoredFeature.
```

```
mp:MonitoredProperty mp:monitoredObservableProperty  
cpm:ObservableProperty
```

```
mp:MonitoredProperty mp:monitoringProcedure mp:MonitoringProcess.
```

Environmental Monitoring Facilities

The key classes of the Environmental Monitoring Facilities ontology (<http://purl.org/voc/ef>) are (Figure 9):

- Abstract Monitoring Object
- Environmental Monitoring Programme
- Environmental Monitoring Activity

- Abstract Monitoring Feature
- Environmental Monitoring Facility
- Environmental Monitoring Network

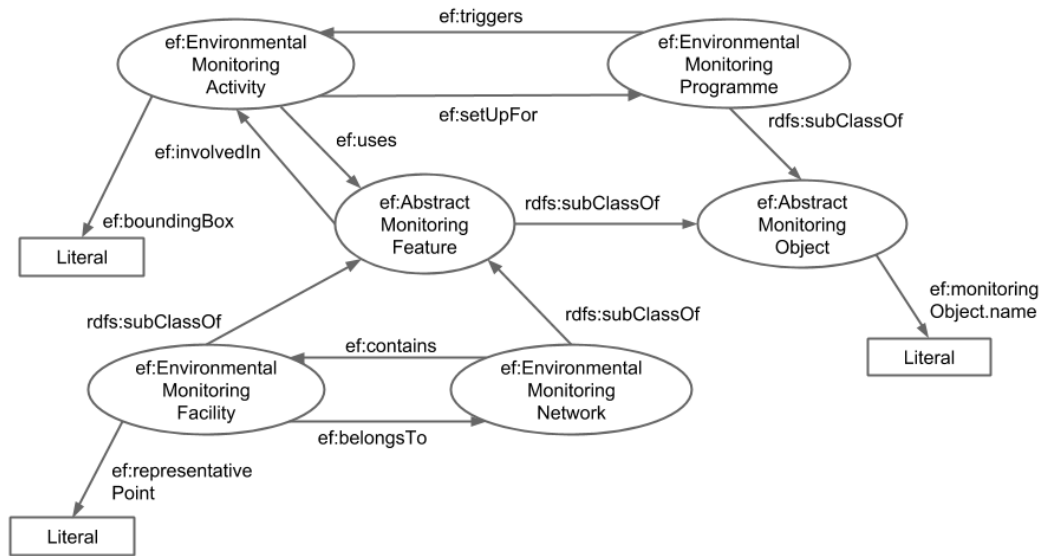


Figure 9. Overview of the Environmental Monitoring Facilities Ontology

(<http://purl.org/voc/ef>)

The UML stereotypes of “Environmental Monitoring Programme” and “Abstract Monitoring Feature” are modelled within the OWL ontology as subclasses of “Abstract Monitoring Object” such that:

```
ef:EnvironmentalMonitoringProgramme rdfs:subClassOf
    ef:AbstractMonitoringObject.
```

and

```
ef:AbstractMonitoringFeature rdfs:subClassOf ef:AbstractMonitoringObject.
```

Similarly, “Environmental Monitoring Facility” and “Environmental Monitoring Network”, which are specialisations of “Abstract Monitoring Feature” are modelled thus:

ef:EnvironmentalMonitoringFacility rdfs:subClassOf
ef:AbstractMonitoringFeature.

and

ef:EnvironmentalMonitoringNetwork rdfs:subClassOf
ef:AbstractMonitoringFeature.

An “Environmental Monitoring Activity” is associated with its programme governance, and with the facilities (or networks) that it uses:

ef:EnvironmentalMonitoringActivity ef:setUpFor
ef:EnvironmentalMonitoringProgramme.

ef:EnvironmentalMonitoringProgramme ef:triggers
ef:EnvironmentalMonitoringActivity.

ef:EnvironmentalMonitoringActivity ef:uses ef:AbstractMonitoringFeature.

The association between “Environmental Monitoring Facility” and “Environmental Monitoring Network” demands more attention in the modelling because in the original UML model the relationships “belongsTo” and “contains” can have attribution. In a simple mapping from UML to OWL, attributes become data properties, and relationships become object properties. In the OWL schema, the domain of data property does not include object property (only class).

We solve this by modelling the domain of the object property “contains” as the union of “Environmental Monitoring Network” and “NetworkFacility”, and its range as the union of “Environmental Monitoring Facility” and “NetworkFacility”. And similarly for the object property “belongsTo”.

Application

As noted in the Introduction to this paper, the INSPIRE Registry provides high level terminology which the ontologies presented above may be mapped to. This mapping is presented in Appendix B, see the supplementary material to this paper. It is also important to emphasise that Monitoring Properties is only one of many contexts to which Complex Properties may be applied.

The ontologies (Complex Properties, Monitoring Properties & Environmental Monitoring Facilities) were specifically designed to work together in order to provide interoperability between existing metadata resources. In order to demonstrate the potential benefits of the approach to scientists, a prototype application was created linking information from two existing UK national-level metadata catalogues:

- (1) The spatial data resources catalogued via the portal at www.data.gov.uk
- (2) The UK Environmental Observation Framework catalogue at www.ukeof.org.uk

Prototyping Approach

Craglia *et al.* (2012) considered that a future Digital Earth would allow for visualisation on a globe of what lies underground, and what lies under the water. In order to show how the ontologies described in this paper further that aim, terms defined in SKOS vocabularies were first “atomised” according to the Complex Properties ontology. The constituent concepts of a complex property were associated with the property itself utilising the `skos:related` predicate.

Dataset metadata was then extracted from the source repositories which feed the data.gov.uk portal. Specific datasets were chosen to represent typical data published by three different NERC Data Centres - British Geological Survey (BGS), British Oceanographic Data Centre (BODC), and the Centre for Ecology & Hydrology (CEH).

Using Open Refine (<http://www.openrefine.org>) with the RDF extension (<http://www.refine.deri.ie>), raw metadata was loaded in spreadsheet format and semantically marked up by referencing the afore-mentioned SKOS vocabularies. A mapping from the spreadsheet columns to the OWL ontologies was then declared, and the resulting triples exported as a set of RDF/OWL files.

The RDF files containing the now semantically richer metadata were then loaded into a Sesame (<http://www.rdf4j.org>) triple store, and SPARQL Protocol and RDF Query Language (SPARQL) queries coded to answer typical use cases relating to discovery of datasets in an environmental monitoring context.

A content management system (PLONE, <http://www.plone.org>) was utilised to generate interactive web pages targeted at end users. This involved caching SPARQL query results, and so the approach is limited in terms of data volume scalability - however, Content Management System capabilities proved more than adequate for quickly generating standard visualisations such as charts, maps and plots with both free-text and faceted search functionality (Figures 10-12).

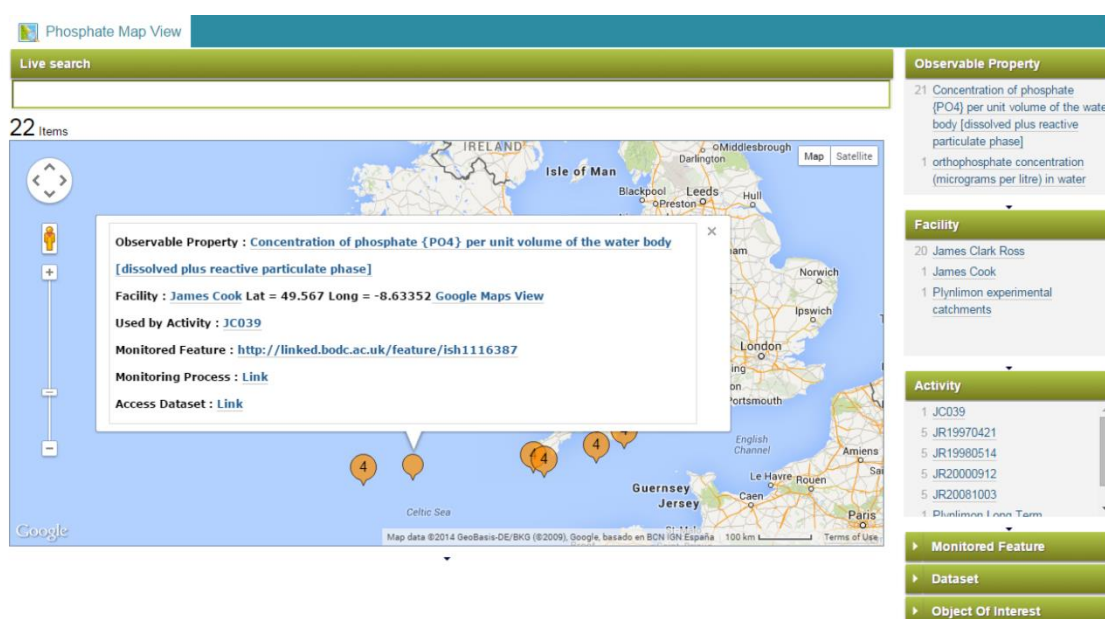


Figure 10. Searching for “phosphate” measurement reveals a cruise report held at the British Oceanographic Data Centre.

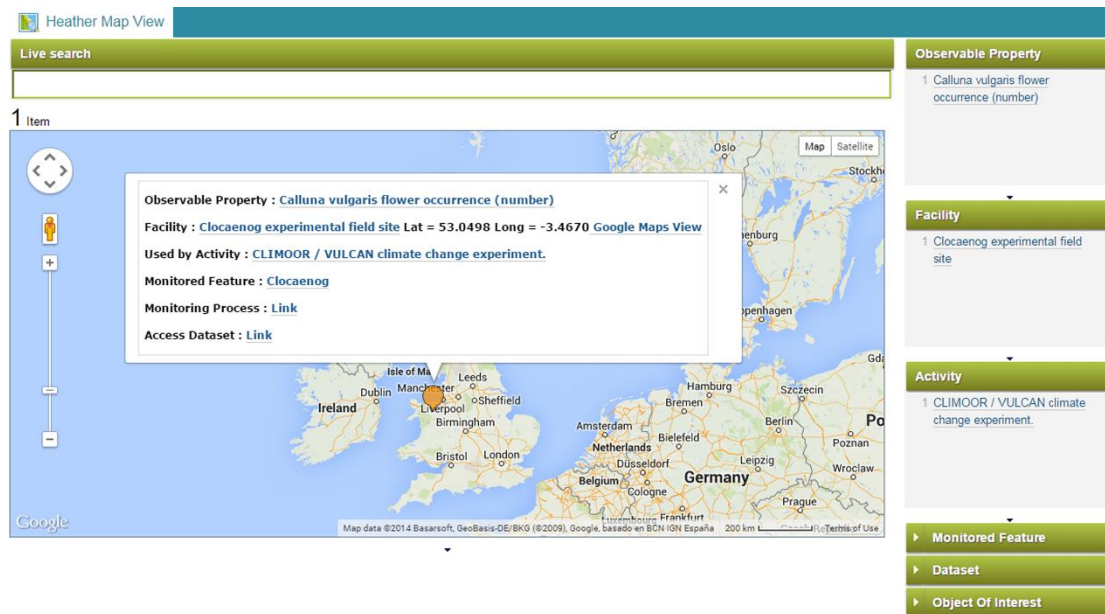


Figure 11. Searching for “heather” identifies catchment monitoring activity carried out in Wales.

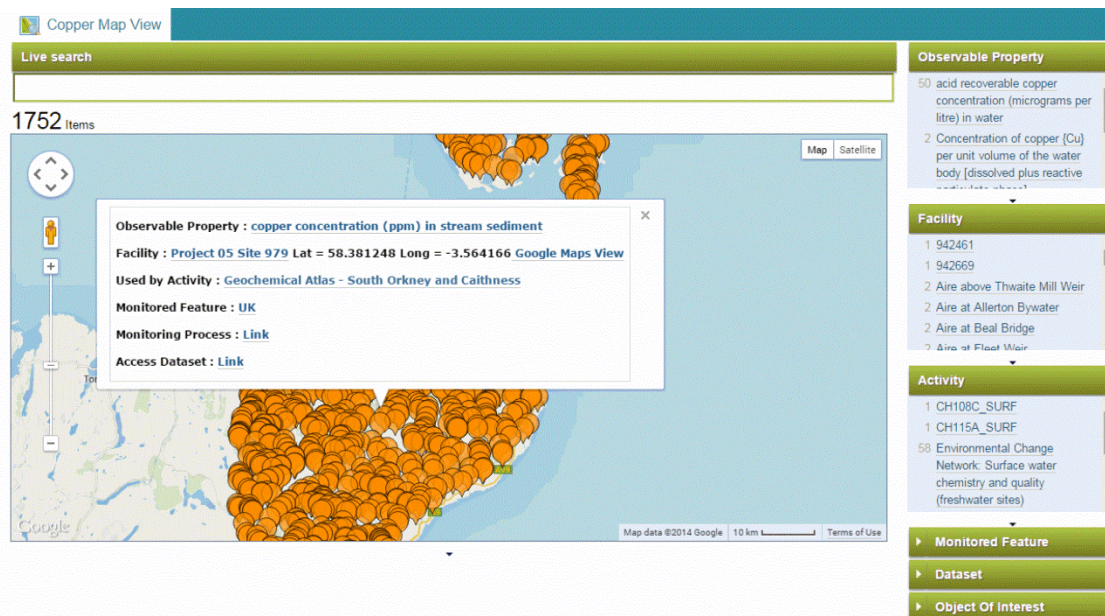


Figure 12. A relatively high density of locations reflecting some of the Geochemical survey data in northern Scotland.

It was possible to demonstrate capabilities including:

- (1) Selecting a substance (e.g. “phosphate”) and requesting a map of locations where an Observable Property semantically related to the concept of phosphate was monitored.
- (2) Selecting one such location and discovering details such as its name, its governance/managing body, and relevant contact information.
- (3) Being satisfied that the governance and monitoring protocols at that site are appropriate to the user’s requirements, following links to resources providing access services for the specific dataset.

Ontology Alignment

Ontology alignment is the process of determining the commonality between classes and concepts from different ontologies. The ontologies presented in this paper have common ground with a number of other resources. By aligning with these ontologies, it is possible to use the Complex Properties, Monitoring Properties and Environmental Facilities ontologies in a broader context, to interoperate with other representations of observations made the “real world” and incorporated into Digital Earth systems. In particular, from a Digital Earth perspective, integration with these ontologies allows a flow of heterogeneous data from sensors to be represented on a three-dimensional virtual globe, and can allow the representation of both current and historical observations (Craglia, *et al.*, 2012). This meshes with the philosophical approach above that there is an “open-world” and that the representations of the Complex Properties, Monitoring Properties and Environmental Facilities data models are not the only valid worldviews. Following this argument reinforces the decision to be flexible in the approach taken to the translation of the initial UML models to OWL models as the alignment with these other semantic resources was not necessarily

envisaged within the original INSPIRE specifications.

Observations & Measurements and Observed Properties

The ISO / Open Geospatial Consortium standard Observations and Measurements (O&M) model is used to:

“.. determine values of properties, though application of some procedure at a particular time and place. The result of an observation is strictly an estimate of the true value, conditioned by procedure and circumstances, so description of the latter are important in the assessment of the reliability of the estimate.” Cox (submitted).

O&M overlaps with the Complex Properties Model as they share the Observed Property and Feature of Interest concepts. This should be unsurprising given that the source of the Complex Property model comes from a specific extension to the O&M model. The Monitoring Properties model also overlaps with O&M, specifically as the O&M Process concept can be instantiated by the Monitoring Properties class of Monitoring Process. There also exists an ontological model to describe the Observed Property of an Observation from an O&M instance. This Observable Properties (<http://environment.data.gov.au/def/op>) ontology covers many of the same concepts as the Complex Properties Model (Constraint, Matrix, Object Of Interest, Substance, Taxon). However, there are limitations of the Observable Properties ontology which the Complex Properties ontology addresses (Finney, *per. comm.*). One of the identified difficulties with using the Observable Properties ontology is that in many cases observations are a combination of a quantity and a mathematical operator (like “Average Height”). The classes of Constraint and Statistical Measure from the Complex Properties model help address some of the mathematical qualifying issues in this mapping. Full alignment could be achieved by declaring the Observable Properties ontology Property Kind class to be equivalent to the Observable Property class of the Complex Properties model.

Domain Ontologies

Within various disciplines there are existing ontologies with which any generic data model should be seen to align. Here we demonstrate the applicability of the model presented in this paper to a number of domain specific resources.

Ocean Data Ontology

The Ocean Data Ontology (ODO, <http://www.ocean-data.org/>) represents a logical data model for storing metadata and data related to oceanographic field activities, in particular research vessel cruises. As such one of the key classes within ODO is that of “Vehicle”, which describes the aircraft, towed vehicle, submarine or water-surface vessel from which the environmental monitoring takes place. It can therefore be seen that an ODO Vehicle is an Environmental Monitoring Facility, its deployment on a cruise or mission is an Environmental Monitoring Activity which is therefore a Monitoring Properties Data Origin. ODO models data collected on a specific deployment from a specific instrument as a Deployment Dataset, which can be seen to be analogous to a Monitoring Dataset, and the ODO Parameter (what is measured and how) is aligned to Monitoring Property.

Chemical Entities of Biological Interest

Chemical Entities of Biological Interest (ChEBI, <http://www.ebi.ac.uk/chebi/>) is a freely available dictionary of molecular entities focused on ‘small’ chemical compounds. The term ‘molecular entity’ refers to any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, etc., identifiable as a separately distinguishable entity. The molecular entities in question are either products of nature or synthetic products used to intervene in the processes of living organisms. Each molecular entity within ChEBI is available as an instance of an OWL class, and therefore ChEBI is a

powerful register of Substance instances for use with the Complex Properties model and allowing a standard point of interoperability between various domains.

World Register of Marine Species

Similar to ChEBI but direct at biological entities rather than molecular entities, the World Register of Marine Species (WoRMS, <http://www.marinespecies.org>) provides an authoritative and comprehensive list of names of marine organisms, including information on synonymy. While highest priority goes to valid names, other names in use are included so that this register can serve as a guide to interpret taxonomic literature.

The content of WoRMS is controlled by taxonomic experts, not by database managers. WoRMS has an editorial management system where each taxonomic group is represented by an expert who has the authority over the content, and is responsible for controlling the quality of the information. As with ChEBI, WoRMS offers each instance as an RDF resource and it is therefore a register which offers instances for the Taxon class of the Complex Properties model in a marine setting.

Light-weight Ontologies

In the spectrum of semantic resources (McGuinness, 2003), at the least complex end there exist a range of resources that can be considered as controlled vocabularies, thesauri or lightweight ontologies.

NERC Vocabulary Server

Within the oceanographic domain, the vocabularies served from the NERC Vocabulary Server (Leadbetter, Lowry and Clements, 2014). The largest of these vocabularies is the “BODC Parameter Usage Vocabulary” (P01, <http://vocab.nerc.ac.uk/collection/P01/current/>)

which is used within the SeaDataNet project to annotate the fields of data files with the physical properties that the numbers represent. P01 concepts are built from a rigorous, but largely hidden, semantic model (Lowry and Leadbetter, 2014) describing object of interest, phenomenon of interest species of interest, substance of interest, matrix and analysis procedure.. Thus it can be seen that the P01 vocabulary is a target candidate for being atomised using the Complex Properties Model. Target concepts from the sub-disciplines of physical, chemical and biological oceanography are shown within the Complex Properties Ontology model in Table 4.

Table 4. A breakdown of lexical terms published established controlled vocabularies to the Complex Properties model. Controlled vocabularies represented within this table include the BODC Parameter Usage Vocabulary.

| Web address of base concept | Statistical Measure | Property | Object of Interest | Matrix | Constraint | Unit of Measure |
|---|---------------------|-------------------|---------------------------------|---------------------------|---|-----------------------|
| http://vocab.nerc.ac.uk/collection/P01/current/MMUSDTBT/ | | Concentration | tributyltin cation | Mytilus galloprovincialis | flesh of Mytilus | $\mu\text{g kg}^{-1}$ |
| http://vocab.nerc.ac.uk/collection/P01/current/AI18GCD1/ | | Concentration | 2,3',4,4',5-pentachlorobiphenyl | water body | dissolved plus reactive particulate <GF/F phase | ng l^{-1} |
| http://vocab.nerc.ac.uk/collection/P01/current/ASLVMNDY/ | Daily mean | Surface elevation | | water body | unspecified datum | m |

However, there are several other vocabularies served by the NERC Vocabulary Server, such as the International Council for the Exploration of the Seas Platform Code List (<http://vocab.nerc.ac.uk/collection/C17/current/>) and the BODC data storage units (<http://vocab.nerc.ac.uk/collection/P06/current/>) vocabulary which may be used as instances of Environmental Monitoring Facilities and Complex Properties' Units of Measure respectively. This allows the data model from this paper to be interoperable with work already conducted using the NERC Vocabulary Server's lexical concepts, such as the European Commission's SeaDataNet and EMODnet projects.

CEH Analytical Services Thesaurus

The CEH Analytical Services Thesaurus (CAST, <http://onto.nerc.ac.uk/CAST>) is used within the environmental chemistry domain to automatically tag environmental data within a Laboratory Information Management System (Wright, Harrison and Watkins, 2014). Terms from CAST cover aspects of a sample or feature which are measured and assigned a value from an agreed domain ("determinands"); various processes of filtration, analytical methodology and preservation; and units of measure. As with the NERC Vocabulary Server, there are various parts of the data model presented in this paper covered by these terms. For instance "determinands" covers both Constraint and Property classes from the Complex Properties model, while filtration, analytical methodology and preservation may all be associated with the Monitoring Properties class of Monitoring Procedure. However, this does highlight one of the weaknesses in the model that these are actually analytical processes and not monitoring process, and a fourth class indicating *post hoc* analyses may be required in future versions of the data model presented here.

Environmental Change Network

The Environmental Change Network (<http://www.ecn.ac.uk>) is a long term monitoring and research programme in the UK, which operates to defined protocols. It collects, analyses and interprets long-term data from a network of sites. ECN data aims to improve our understanding of how and why environments change. The variables recorded are candidate Observable Properties which can be “atomised” according to the Complex Properties model (Table 5).

Table 5. A breakdown of some example lexical terms used in Environmental Change Network protocols.

| Statistical Measure | Property | Object of Interest | Matrix | Constraint | Unit of Measure |
|----------------------------|-----------------|---------------------------|---------------|-------------------|------------------------|
| | weight | suspended solids | river water | ash-free dry | mg l ⁻¹ |
| | level | water | | | m |
| | concentration | carbon | lake water | total organic | mg l ⁻¹ |

G-BASE

The British Geological Survey Geochemical Baseline Survey of the Environment (G-BASE, <http://www.bgs.ac.uk/gbase>) is an annual campaign of geochemical sampling within many parts of the UK. Beginning in the late 1960s, it is a high-resolution geochemical survey producing baseline data relevant to many environmental issues (Table 6).

Table 6. A breakdown of some example lexical terms used by the British Geological Survey Geochemical Baseline Survey of the Environment.

| Statistical Measure | Property | Object of Interest | Matrix | Constraint | Unit of Measure |
|---------------------|-------------------------|--------------------|-----------------|------------|----------------------|
| | concentration | copper | stream sediment | | parts per million |
| | mass loss on ignition | carbon | soil | | % |
| | electrical conductivity | | stream water | | $\mu\text{S m}^{-1}$ |

Use of the Provenance Ontology

The Provenance Ontology (PROV-O, Lebo, Sahoo and McGuinness, 2013) of the World Wide Web Consortium provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts. Compton, Corsar and Taylor (2014) have shown the applicability of this ontology to sensor data in the context of the Semantic Sensor Network, and Ma *et al.* (2014) have demonstrated the use of PROV to track use of datasets, roles in report generation and organisational contributions to the Global Change Information system. There is a similar connection to PROV-O as an overlay here to connect the Monitoring Properties classes and the Environmental Monitoring Facilities classes. Explicitly, a Monitoring Dataset is a digital entity with some fixed parameters and therefore can be seen as an instance of PROV-O Entity. An Entity within PROV-O is generated by a PROV-O Activity, and the two are associated by the “was generated by” association. PROV-O defines an Activity as “something that occurs over a period of time and acts upon or with entities; it may include ...

generating entities.” In light of this the Data Origin of a Monitoring Property can be seen as a PROV-O Activity, and as a Monitoring Property Data Origin is modelled as a subclass of an Environmental Monitoring Activity, PROV-O provides a bridge from the Environmental Facilities, to the Monitoring Properties and therefore through to the Complex Properties observed.

Future Work

The model presented in this paper could be further developed in the future, in particular as the method of acquisition of a Complex Property is currently defined only through the collection or acquisition process at the Monitoring Properties class level. This means that the detailed usage level information regarding the observation is one step removed from the detailed description of how that observation was obtained. This could be resolved by offering an “analysis” process at the level of the Complex Property.

The instances of the data model used in the applications which have been described in this paper have been created by hand, and as such there is future work in making the creation of these instances an automated procedure to enhance the coverage of data in these demonstration applications. Other potential applications include: the aggregation of data from a “raw” data portal such as SeaDataNet into agreed, scientifically meaningful products for reporting to policy makers in frameworks such as the European Marine Observation and Data Network; and automatically creating systematic reviews of literature in ecological studies. The presented ontologies are designed to be OWL-DL compatible. However, computational completeness and decidability was not a focus of the initial development. Future work could include enhancement of the ontologies, paying particular attention to necessary and sufficient conditions, with a view to testing the application of reasoners. Typical reasoning applications would include inferencing, classification, and the detection of inconsistencies in existing

environmental data. Finally, a system with intelligent semantic annotation, from the point of collection on an instrument (“born semantic”, Leadbetter and Fredericks 2014; Buck, Leadbetter and Williams 2015) through to processing at the desk or in the laboratory and finally to data archive could be achieved by beginning with instrument data file headers carrying information from this data model. This kind of system could the complex and monitored properties concepts within the instrument firmware, output an RDF representation of the data and transmit it in real-time using a binary-encoding of the RDF data (e.g. HDT see Gallego *et al.* 2011; or NetCDF-LD, see Yu *et al.*, 2015) or store the RDF on the data logger for later retrieval. The platform on which the instrument is deployed could stamp the environmental monitoring facilities concepts onto the raw data as an additional header piece. Semantically-aware data processing tools could then be used to update the analysis information in the monitored properties model and all provenance information as the data are worked up in the laboratory.

Conclusion

This paper has shown the development of a model which allows the description of complex environmental observing systems, and complex environmental observations, in a robust manner which allows the easy exchange of data across the range of disciplines which comprise environmental and Earth sciences. It has been developed from the point of view of coherently and comprehensively integrating data about the Earth system from a range of sources into digital representations of the Earth system. Specific applications of this model have been shown, including answering cross-domain questions regarding the location of measurements of specific chemical entities. These prototype applications have been shown to further the Digital Earth 2020 vision of Craglia *et al.* (2012). These cross-domain applications provide significant benefits in breaking down the barriers between disciplinary

data silos, which gives better public access to complex environmental data, and allows easier integration of these complex data across domains and disciplines in virtual representations of the Earth system. This again aligns with the Digital Earth 2020 vision, where Craglia *et al.* state that the future Digital Earth will “synthesise heterogeneous information.” Indeed, the use of the W3C’s Provenance ontology in the environmental ontologies presented in this paper also provides a link to Craglia *et al.*’s next statement, that future Digital Earth will provide “metrics of quality and trust of both data inputs and outputs” as provenance is one aspect of a measure of trust in a collaborative research network (Leadbetter, 2015).

References

- Justin Buck, Adam Leadbetter, Charlotte Williams, 2015. Born semantic: Linking data from sensors to users and balancing hardware limitations with data standards. Association for the Sciences of Limnology and Oceanography 2015 Aquatic Sciences Meeting, 22nd - 27th February 2015 Granada, Spain.
- Michael Compton, David Corsar and Kerry Taylor, 2014. Sensor Data Provenance: SSNO and PROV-O Together at Last. In: *7th International Workshop on Semantic Sensor Networks, 20th October 2014, Trentino, Italy*.
- Simon Cox, 2010. Observations and Measurements. *OGC Abstract Specification Topic 20. Open Geospatial Consortium document 10-004r3*
<http://portal.opengeospatial.org/files/41579> Accessed 8th December 2014
- Simon Cox, 2013. An explicit OWL representation of ISO/OGC Observations and Measurements. In *SSN 2013 Semantic Sensor Networks Proceedings of the 6th International Workshop on Semantic Sensor Networks co-located with the 12th International Semantic Web Conference (ISWC 2013) Sydney, Australia, October 22nd, 2013*, Oscar Corcho, Cory Hanson and Payam Barnaghi Eds, CEUR Workshop Proceedings 1063: 1-18.
- Simon Cox, submitted. Basic Observations and Sampling Feature Ontology. *Semantic Web Journal*. Retrieved from <http://www.semantic-web-journal.net/content/basic-observations-and-sampling-feature-ontology> 8th December 2014.
- Max Craglia , Kees de Bie , Davina Jackson , Martino Pesaresi , Gábor RemeteyFülöpp , Changlin Wang , Alessandro Annoni , Ling Bian , Fred Campbell , Manfred Ehlers ,

- John van Genderen , Michael Goodchild , Huadong Guo , Anthony Lewis , Richard Simpson , Andrew Skidmore and Peter Woodgate, 2012. Digital Earth 2020: towards the vision for the next decade, *International Journal of Digital Earth*, 5:1, 4-21, DOI: 10.1080/17538947.2011.638500
- Paolo Diviacco, Cyril Pshenichny, Roberto Carniel, Zinaida Khrabrykh, Victoria Shterkhun, Dmitry Mouromtsev, Silvina Guzmán and Paolo Pascolo, 2014. Organization of a geophysical information space by using an event-bush-based collaborative tool. *Earth Science Informatics*. 10.1007/s12145-014-0182-2
- Paolo Diviacco, 2015. Maps, graphs and annotations as Boundary Objects in Knowledge Networks, Distributed Cognition and Collaborative e-Research. In *Collaborative Knowledge in Scientific Research Networks*, P. Diviacco, P. Fox, C. Pshenichy, and A. Leadbetter, Eds. Hershey, PA:IGI Global, 2015.
- EC, 2007. Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). European Commission, retrieved 11th December 2014 from http://eur-lex.europa.eu/legal-content/EN/ALL/;ELX_SESSIONID=hSHpJJ3P7QQvzP47G7PQGbbqDQyjD1zrvQwN9yvy4kTzv1YLvMTFC!-1737207740?uri=CELEX:32007L0002
- Mario Arias Gallego, Javier D. Fernandez, Miguel A. Martinez-Prieto, and Claudio Gutierrez, 2011. HDT-it: Storing, Sharing and Visualizing Huge RDF Datasets. In *Proceedings of the 10th International Semantic Web Conference*. INSPIRE, 2013a.
- D2.8.II.2 INSPIRE Data Specification on Land Cover – Technical Guidelines European Commission Joint Research Centre. Accessed from http://inspire.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_LC_v3.0.pdf 17th October 2014.
- INSPIRE, 2013b. D2.8.II/III.7 Data Specification on Environmental Monitoring Facilities – Technical Guidelines 10th December 2013
- INSPIRE, 2013c. D2.9 Guidelines for the use of Observations & Measurements and Sensor Web Enablement - related standards in INSPIRE Annex II and III data specification development. 8th Apr 2014.
- Adam Leadbetter, Roy Lowry and Oliver Clements, 2014. Putting meaning into NETMAR - the Open Service Network for Marine Environmental Data. *International Journal of Digital Earth* 7(10): 811-828.

- Adam Leadbetter and Janet Fredericks, 2014. We have "born digital" - now what about "born semantic?". EGU General Assembly Conference Abstracts 16: 1707.
- Adam Leadbetter, 2015. Examining trust in collaborative research networks. In Collaborative Knowledge in Scientific Research Networks, P. Diviacco, P. Fox, C. Pshenichy, and A. Leadbetter, Eds. Hershey, PA:IGI Global, 2015.
- Adam Leadbetter, Robert Arko, Cynthia Chandler, Adam Shepherd and Roy Lowry, 2015. Loose integration of local information to generate collaborative marine science knowledge. In Collaborative Knowledge in Scientific Research Networks, P. Diviacco, P. Fox, C. Pshenichy, and A. Leadbetter, Eds. Hershey, PA:IGI Global, 2015.
- Timothy Lebo, Satya Sahoo and Deborah McGuinness, editors. (2013). PROV-O: The PROV Ontology. World Wide Web Consortium. Accessed from <http://www.w3.org/TR/2013/REC-prov-o-20130430/> 17th October 2014.
- Roy Lowry and Adam Leadbetter, 2014. Semantically supporting data discovery, markup and aggregation in the European Marine Observation and Data Network (EMODnet). Geophysical Research Abstracts 16: EGU2014-13693.
- Deborah McGuinness, 2003. "Ontologies Come of Age." In Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential, edited by Dieter Fensel, James Hendler, Henry Lieberman, and Wolfgang Wahlster, 171–196. Cambridge, MA: MIT Press.
- Xiaogang Ma, Jin Guang Zheng, Justin Goldstein, Stephan Zednik, Linyun Fu, Brian Duggan, Steven Aulenbach, Patrick West, Curt Tilmes and Peter Fox, 2014. Ontology engineering in provenance enablement for the National Climate Assessment. Environmental Modelling & Software 61, 191-205.
<http://dx.doi.org/10.1016/j.envsoft.2014.08.002>
- Fadi Maali and John Erickson, 2014. Data Catalog Vocabulary (DCAT), W3C Recommendation 16 January 2014. World Wide Web Consortium, Retrieved from <http://www.w3.org/TR/vocab-dcat/> 11th December 2014.
- Luc Moreau and Paolo Missier, 2013. PROV-DM: The PROV Data Model, W3C Recommendation 30 April 2013. World Wide Web Consortium, Retrieved from <http://www.w3.org/TR/prov-dm/> 10th February 2015.
- Dick M. A. Schaap, Roy K. Lowry, 2010. SeaDataNet - Pan-European infrastructure for marine and ocean data management: unified access to distributed data sets. Int. J. Digital Earth 3(sup1): 50-69

Daniel G. Wright, Kathryn A, Harrison and John Watkins, 2014. Automated tagging of environmental data using a novel SKOS formatted environmental thesaurus. *Earth Science Informatics*. 10.1007/s12145-014-0183-1

David Weinberger, 2002. *Small Pieces Loosely Joined: A Unified Theory of the Web*. Perseus Publishing. ISBN 0-7382-0543-5

Jonathan Yu, Nicolas Car, Adam Leadbetter, Bruce Simons, and Simon Cox, 2015. Towards Linked Data Conventions for Delivery of Environmental Data Using netCDF. In *Environmental Software Systems. Infrastructures, Services and Applications* (pp. 102-112). Springer International Publishing.

Appendix A – Detailed Mapping between UML and OWL.

| Ontology | UML Type | UML Name | Description | OWL Type | OWL Name | Mapping Notes |
|--------------------|-------------|--------------------|--|-----------------|--------------------|---|
| Complex Properties | Association | restriction | Relates an Observable Property with one or more Constraints | Object Property | restriction | Equivalent |
| Complex Properties | Association | statisticalMeasure | Relates an Observable Property with one or more Statistical Measures. | Object Property | statisticalMeasure | Equivalent |
| Complex Properties | Attribute | basePhenomemon | A value from a codelist of property types (temperature, wind, speed). | Class | Property | <p>A Property is a characteristic of the Object Of Interest that is observed during the act of observation, for example: if the temperature of the water is measured, then the Property is Temperature.</p> <p>Object Of Interest is the substance, taxon or other physical/chemical phenomenon of the feature that is being observed, for example: Waves, Rainfall, Calluna Vulgaris, Aluminium.</p> |
| Complex Properties | Attribute | uom | O&M does not explain its use of Unit Of Measure, other than referencing Geographic Information - | Class | UnitOfMeasure | Promoting the UML attribute to an OWL class allows it become a subclass of skos:Concept, so that we can manage units in a SKOS thesaurus. |

| | | | | | | |
|--------------------|-------|-----------------------------|--|-------|-----------------------------|------------|
| | | | Conceptual schema language [ISO/TS 19101:2005]. | | | |
| Complex Properties | Class | AbstractObservableProperty | The complex properties extension to O&M allows the abstract class Abstract Observable Property to be implemented by two specialisations, see “Class: Observable Property” and “Class: Composite Observable Property”. | Class | AbstractObservableProperty | Equivalent |
| Complex Properties | Class | CompositeObservableProperty | Composite Observable Property is a class grouping together multiple Phenomena Observable Properties into one Composite Observable Property element e.g. a strongly linked pair of phenomena such wind speed and wind direction). | Class | CompositeObservableProperty | Equivalent |

| | | | | | | |
|--------------------|-------|---------------------|---|--------------|--------------------|--|
| Complex Properties | Class | Constraint | A restriction acting to shrink the domain of a specific Observable Property to certain circumstances or restrictions. For example, if the Observable Property is “attenuance of light” then “of red wavelengths” might be a Constraint. | Class | Constraint | Equivalent |
| Complex Properties | Class | ObservableProperty | Observable Property is a class representing a reference to a phenomenon definition in a codelist with optional units of measure, which may be augmented using Constraints and/or Statistical Measures. | Class | ObservableProperty | Equivalent |
| Complex Properties | Class | Observation | Observation is a class representing the act of measuring or otherwise determining the value of a property | Not in scope | | Note that in this mapping, we are not directly concerned with actual observations – only with identifying the Datasets that contain Observation values. |
| Complex Properties | Class | Statistical Measure | Some function over time or space which aggregates the values associated with Observable | Class | StatisticalMeasure | Equivalent |

| | | | | | | |
|-------------------------------------|-------------|-----------|--|-----------------|-----------|---|
| | | | Properties, e.g. "daily maximum" | | | |
| Complex Properties | n/a | n/a | This concept is not included in the O&M model. | Class | Matrix | Where a feature is a complex entity, the Matrix may be identified in order to clarify the particular component or aspect of the feature at, in, or on which the Object Of Interest was observed - for example: "Vegetation", "Soil", "Water", "Stream Sediment". A sample taken from a river could be either of the water or of the sediment. |
| Complex Properties | n/a | n/a | This concept is not included in the O&M model. | Object Property | matrix | Observable Properties may be observed within a Matrix. |
| Environmental Monitoring Facilities | Association | belongsto | A link pointing to the Environmental Monitoring Networks this the Environmental Monitoring Facility pertains to. | Object Property | belongsTo | owl:domain = UNION(EnvironmentalMonitoringFacility AND NetworkFacility) owl:range=UNION(EnvironmentalMonitoringNetwork AND NetworkFacility) |
| Environmental Monitoring Facilities | Association | contains | A link pointing to the Environmental Monitoring Facilities included in the Environmental Monitoring Network. | Object Property | contains | owl:domain = UNION(EnvironmentalMonitoringNetwork AND NetworkFacility) owl:range=UNION(Environmen |

| | | | | | | |
|--|-------------|---------------------|---|--------------------|-----------------------|---|
| | | | | | | talMonitoringFacility AND NetworkFacility) |
| Environm ental Monitorin g Facilities | Association | triggers | Environmental Monitoring Activities triggered by the Environmental Monitoring Programme. | Object Property | triggers | Equivalent |
| Environm ental Monitorin g Facilities | Association | uses | The specific set of Abstract Monitoring Features involved in the Environmental Monitoring Activity. | Object Property | uses | Equivalent |
| Environm ental Monitorin g Facilities | Attribute | boundingBox | A representation of the area in which the Environmental Monitoring Activity takes place. | Data Property | boundingBox | Equivalent |
| Environm ental Monitorin g Facilities | Attribute | name | A plain text denotation of the Abstract Monitoring Object. | Data Property | monitoringObject.name | Note that because names are local in UML, we must qualify it for it to be unique in OWL. |
| Environm ental Monitorin g Facilities | Attribute | representativePoint | A representative point location for the Environmental Monitoring Facility. | Data Property | representativePoint | Equivalent |

| | | | | | | |
|--|-------|-------------------------------------|---|-------|-------------------------------------|------------|
| Environm ental Monitorin g Facilities | Class | AbstractMonitoringFea ture | An abstract base class for the environmental monitoring features in the real world (Environmental Monitoring Facility, and Environmental Monitoring Network) | Class | AbstractMonitoringFeature | Equivalent |
| Environm ental Monitorin g Facilities | Class | AbstractMonitoringObj ect | An abstract base class for environmental monitoring objects. | Class | AbstractMonitoringObject | Equivalent |
| Environm ental Monitorin g Facilities | Class | EnvironmentalMonitor ingActivity | The specific set of Abstract Monitoring Features used for a given domain in a coherent and concise timeframe, area and purpose. Examples of instances of this class include: “A research vessel cruise such as Royal Research Ship James Clark Ross 20080221”, “Geochemical Baseline Survey of the Environment (G- BASE)”, “Long-term Monitoring e.g. the Cumbrian lakes environmental research (CLEAR)”. | Class | EnvironmentalMonitoringActi vity | Equivalent |

| | | | | | | |
|--|-------|-------------------------------------|---|-------|------------------------------------|------------|
| Environm ental Monitorin g Facilities | Class | EnvironmentalMonitor ingFacility | A geo-referenced object directly collecting or processing data about objects whose properties (e.g. physical, chemical, biological or other aspects of environmental conditions) are repeatedly observed or measured. Examples include: "RRS James Clark Ross", "Rothamsted Agricultural Research Station", "BILSAT-1 Earth Observation Satellite". | Class | Environmental Monitoring Facility | Equivalent |
| Environm ental Monitorin g Facilities | Class | EnvironmentalMonitor ingNetwork | An administrative or organisational grouping of Environmental Monitoring Facilities managed the same way for a specific purpose, targeting a specific area. Each network respects common measurement protocols. | Class | EnvironmentalMonitoringNet work | Equivalent |

| | | | | | | |
|--|-------------|--------------------------------------|--|--------------------|--------------------------------------|------------|
| Environm ental Monitorin g Facilities | Class | EnvironmentalMonitor ingProgramme | A framework based on policy relevant documents defining the target of a collection of observations and/or the deployment of Abstract Monitoring Features on the field. | Class | EnvironmentalMonitoringProg ramme | Equivalent |
| Monitorin g Properties | Association | monitoredFeature | Relates a Monitored Property to the geographic feature (the class Monitored Feature) which manifests an Observable Property. | Object Property | monitoredFeature | Equivalent |
| Monitorin g Properties | Association | monitoredObservableP roperty | Relates a Monitored Property to the Observable Property (from the Complex Properties ontology) which describes the result values stored in the Monitoring Dataset. | Object Property | monitoredObservableProperty | Equivalent |
| Monitorin g Properties | Association | monitoringProcedure | Relates a Monitored Property to the Process that was involved in gathering the data values which are stored within the Monitoring Dataset. | Object Property | monitoringProceure | Equivalent |

| | | | | | | |
|-----------------------|-------------|-----------------|---|-----------------|------------------------|---|
| Monitoring Properties | Association | storesValuesfor | A Monitoring Dataset stores the result values arising from observations of Monitored Properties. | Object Property | storesValuesFor | Equivalent |
| Monitoring Properties | Association | wasGeneratedBy | A Monitoring Dataset is related to the Data Origin (typically instantiated as an Environmental Monitoring Activity) that caused it come into existence. An association is required to make the connection between a dataset and its originating activity. | Object Property | wasGeneratedBy | Equivalent |
| Monitoring Properties | Attribute | label | Ideally Monitored Features would be already described in a feature registry, but this may not always be the case. This attribute is included for convenience in the situation where feature URIs and labels cannot be readily imported from another source. | Data Property | monitoredFeature.label | Note that because names are local in UML, we must qualify it for it to be unique in OWL. |

| | | | | | | |
|-----------------------|-----------|---------------------------|---|---------------|---------------------------|---|
| Monitoring Properties | Attribute | MonitoringProcessResource | It is unlikely that the detail of a specific Monitoring Process is available in a structured, linked-data format, since there exists no common logical model for such things. This attribute facilitates linking to an external web resource (such as an HTML or PDF document) that can provide a narrative description of the process. | Data Property | monitoringProcessResource | Equivalent |
| Monitoring Properties | Attribute | name | A Data Origin has a label by which the activity is commonly known. This data property is required because Environmental Monitoring Activity as defined in the INSPIRE Environmental Monitoring Facilities context is primarily an associative concept linking programmes | Data Property | dataOrigin.name | Note that because names are local in UML, we must qualify it for it to be unique in OWL. |

| | | | | | | |
|------------------------------|-------|------------|--|-------|------------|------------|
| | | | with the facilities or networks of facilities that those programmes utilise - it does not carry any human meaningful label. Becasue | | | |
| Monitorin g Properties | Class | DataOrigin | Data Origins are activity documentation objects. Instances are the subset of Environmental Monitoring Activities which have actually generated data - and where that data has been collected together and catalogued as one or more Monitoring Datasets. This class is a subtype of Activity as defined in the PROV-O ontology (Lebo, Sahoo and McGuinness, 2013). | Class | DataOrigin | Equivalent |

| | | | | | | |
|------------------------------|-------|-------------------|--|-------|-------------------|------------|
| Monitorin g Properties | Class | MonitoredFeature | A Monitored Feature is a representation of a real world thing which carries the property that is being observed. It is the thing from which the observation/measurement is taken. It may be a sampling feature. | Class | MonitoredFeature | Equivalent |
| Monitorin g Properties | Class | MonitoredProperty | A Monitored Property is an Observable Property that was observed/measured/estimated of a Monitoring Feature using a specified Monitoring Process. The class Monitored Property is related to the INSPIRE EF class "Observing Capability" - however, whereas an Observing Capability represents a period of time during which it was, is, or will be, possible to observe something, a Monitored Property classifies a set of | Class | MonitoredProperty | Equivalent |

| | | | | | | |
|-----------------------|-------|-------------------|---|-------|-------------------|------------|
| | | | observations which have definitely happened, and for which there exists a metadata record describing the associated collection of result values. | | | |
| Monitoring Properties | Class | MonitoringDataset | Monitoring Datasets are dataset documentation objects. Instances are those datasets which are known to have originated from a particular environmental monitoring activity. Example instances include: "Chemical contaminants in White-tailed sea eagle eggs - Predatory Bird Monitoring" | Class | MonitoringDataset | Equivalent |

| | | | | | | |
|------------------------------|-------|-------------------|--|-------|-------------------|------------|
| | | | Scheme (PBMS)", "United Kingdom Butterfly Monitoring Scheme: collated indices 2011", "Regional geochemistry of north-east England (1996)". | | | |
| Monitorin g Properties | Class | MonitoringProcess | The Monitoring Process is the process or methodology used during the monitoring which generated the dataset. | Class | MonitoringProcess | Equivalent |

Appendix B – Mapping to INSPIRE Feature Concepts

| Model | Concept | Mapping | INSPIRE Feature Concept |
|-------------------------------------|------------------------------------|--------------|---|
| Environmental Monitoring Facilities | Abstract Monitoring Feature | Same As | http://inspire.ec.europa.eu/featureconcept/AbstractMonitoringFeature/ |
| Environmental Monitoring Facilities | Abstract Monitoring Object | Same As | http://inspire.ec.europa.eu/featureconcept/AbstractMonitoringObject |
| Environmental Monitoring Facilities | Environmental Monitoring Activity | Same As | http://inspire.ec.europa.eu/featureconcept/EnvironmentalMonitoringActivity |
| Environmental Monitoring Facilities | Environmental Monitoring Facility | Same As | http://inspire.ec.europa.eu/featureconcept/EnvironmentalMonitoringFacility |
| Environmental Monitoring Facilities | Environmental Monitoring Network | Same As | http://inspire.ec.europa.eu/featureconcept/EnvironmentalMonitoringNetwork |
| Environmental Monitoring Facilities | Environmental Monitoring Programme | Same As | http://inspire.ec.europa.eu/featureconcept/EnvironmentalMonitoringProgramme |
| Complex Properties | Application Schema | Adapted From | http://inspire.ec.europa.eu/featureconcept/ObservableProperties/ |
| Monitoring Properties | Process | Subclass Of | http://inspire.ec.europa.eu/featureconcept/Process/ |
| Monitoring Properties | Monitoring Dataset | Subclass Of | http://inspire.ec.europa.eu/glossary/DataSet/ |
| Monitoring Properties | Monitored Feature | Subclass Of | http://inspire.ec.europa.eu/glossary/Feature/ |

| | | | |
|--------------------------|--------------------|-----------------|---|
| Monitoring Properties | Monitored Property | Patterned On | http://inspire.ec.europa.eu/featureconcept/ObservingCapability/ |
|--------------------------|--------------------|-----------------|---|