

PROV ontology supports alignment of observational data (models)

Simon J D Cox

*^a CSIRO Land and Water, Clayton, Vic 3168
Email: simon.cox@csiro.au*

Abstract: The W3C PROV ontology provides a flexible process-flow model that can capture many specific applications. A provenance trace is the retrospective view of a workflow, with specific instance data added. Thus it provides a basis for the description of any chain of activities which generate interesting outputs, such as observations, actuations, or acts of sampling. Furthermore, its relatively generic structure and naming allows it to be used as an alignment bridge with other ontologies that have previously challenged simple mappings. In this paper, we will show a harmonization of a number of important ontology patterns that can be linked through the PROV-O OWL implementation of PROV.

The alignments stack is as follows:

- PROV-O aligned to W3C OWL-Time
- PROV-O aligned to BFO
- W3C SSN/SOSA aligned to PROV-O
- OBOE, OBI and BCO (from the obo foundation) aligned to SOSA/SSN and thus PROV-O

Some of the alignments have been proposed previously, but the set described here both augments them and is larger in aggregate than previous work.

The availability of these alignments supports the fusion of data from a range of disciplines such as earth and environmental sciences, in particular observational data where the act of sampling and observation is understood in a provenance context.

Keywords: *Ontology alignment, observations, sensors, provenance*

1. INTRODUCTION

A number of models for observation metadata have been developed in the earth and environmental science communities. These include the Observations and Measurements (O&M) (Cox, 2007a, 2007b, 2011; ISO/TC 211, 2011) from Open Geospatial Consortium (OGC); the Extensible Observation Ontology (OBOE) (Madin et al., 2007; Schildhauer et al., 2016), the Ontology for Biomedical Investigations (OBI) (Brinkman et al., 2010), and the Biological Collections Ontology (BCO) (Walls et al., 2014) from the life-sciences and ecosystems community; the Semantic Sensor Network Ontology (SSN/SOSA) (Compton et al., 2012; Haller et al., 2017) from the World Wide Web Consortium (W3C); and the Observations Data Model v2 (ODM2) (Horsburgh et al., 2016) from the Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) and NSF collaborators. While all of the models mentioned are *in principle* capable of formalizing almost any earth and environmental science data, each is optimized for particular applications, and usage tends to be concentrated in a particular application domain or science community. This introduces subtle and no-so-subtle variations in both model topology and terminology. Thus, in order to combine data expressed using the various models, mappings between these must be developed.

In some cases mappings are straightforward. For example, because ODM2 took O&M as its starting point, the terminology is almost completely aligned. On the other hand, in the biodiversity and ecological monitoring domains OBOE focuses on groupings and context, and BCO focuses largely on specimens, much more than O&M and SSN. While SSN also adopted terminology from O&M, a direct mapping was confounded by the primary alignment of SSN to the DOLCE foundational ontology (Gangemi, 2010) in which ‘Situation’ is disjoint with ‘Event’. Foundational ontologies such as DOLCE (Gangemi, 2010), BFO (Grenon & Smith, 2004), GFO (Herre, 2010) and UFO (Guizzardi & Wagner, n.d., 2010) can provide a framework for alignment. However, different upper ontologies can be based in different worldviews, and use of incommensurate frameworks may confound rather than help (Mascardi, Cordi, & Rosso, 2007).

A potential resolution is suggested by comparing recent studies that align the original SSN and O&M, respectively, with the PROV-O ontology (Compton, Corsar, & Taylor, 2014; Cox, 2017). PROV-O has just three base classes: Entity, Activity and Agent (Lebo, Sahoo, & McGuinness, 2013). These can effectively serve as a foundational ontology for any process modeling application. Then if ‘observation’ is conceived of as an event or activity, that has stimuli as inputs and generates a value or result as its output, this nicely matches a standard process model. We can thus use PROV-O as a framework for aligning the various observation ontologies.

In this paper we gather together a number of relevant alignments. As presented here they all involve PROV-O, which therefore acts as a bridge between the various ontologies. In this paper, only the key alignments are presented – many more are possible, some of which have been described in other places. The alignments are presented axiomatically using RDFS and OWL (Brickley & Guha, 2014; W3C OWL Working Group, 2012), and serialized using the W3C Turtle syntax (Beckett, Berners-Lee, Prud’hommeaux, & Carothers, 2014). Some of the key alignments are also presents diagrammatically, though it should be noted that the figures only show a subset of the classes and relationships in each case.

2. CORE PROV CLASSES ALIGNED TO SOME FUNDAMENTAL ONTOLOGIES

First we establish some key alignments between PROV-O and widely-used fundamental ontologies.

For reference, Figure 1 summarizes the core classes and relationships from PROV-O. The most important distinction is between the mutually disjoint classes Activity and Entity. An Activity is a temporally bounded thing which is of interest when it generates a (persistent) Entity.

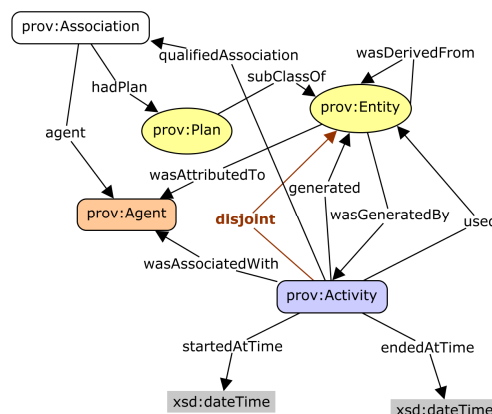


Figure 1. The core classes and relationships in PROV-O

2.1. Alignment to OWL-Time

The Activity class can be related to the TemporalEntity class from OWL-Time (Cox & Little, 2017; Hobbs & Pan, 2006).

The basic axiomatization (Figure 2) is:

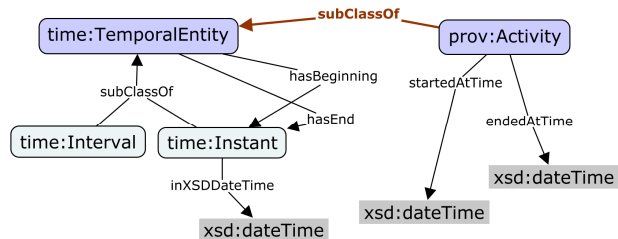


Figure 2. Alignment of Activity with TemporalEntity

```
prov:Activity      rdfs:subClassOf      time:TemporalEntity .
prov:endedAtTime  owl:propertyChainAxiom ( time:hasEnd time:inXSDDateTime ) .
prov:startedAtTime owl:propertyChainAxiom ( time:hasBeginning time:inXSDDateTime ) .
```

Some additional relationships are described in the OWL-Time recommendation (Cox & Little, 2017). While this alignment is not used further in the paper, it provides a useful grounding that emphasizes a key behavior of Activities.

2.2. Alignment to BFO

PROV Activity and Entity are closely related to the concepts Occurrent and Continuant, respectively (or Perdurant and Endurant) which appear in some well-known upper ontologies. Since the Basic Formal Ontology (BFO) (Grenon & Smith, 2004) is the basis for the ontologies from the life sciences community maintained in the OBO Foundry, and thus some of the ontologies considered below, we provide an initial alignment of PROV-O with BFO. The provisional alignment (Figure 3) is:

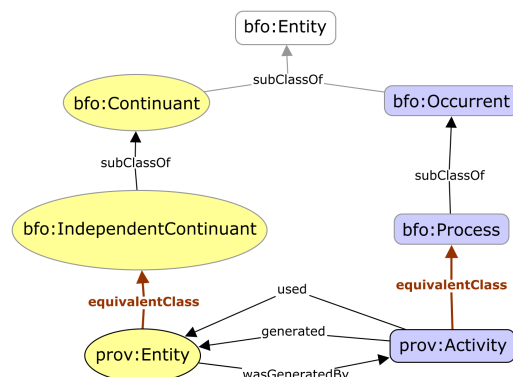


Figure 3. Alignment of PROV core classes and BFO

```
prov:Activity      owl:equivalentClass      obo:BFO_0000015 .      # bfo:Process
prov:Entity        owl:equivalentClass      obo:BFO_0000004 .      # bfo:IndependentContinuant
```

Since the URIs for OBO Foundry classes and properties use opaque codes, a more familiar name taken from the resource label is provided in a comment string.

The relationships between Activity and Entity (used, generated, etc.), may also be aligned with BFO but this requires some elaborate constraints and property chains which we cannot present in the space available.

3. SSN/SOSA AND PROV-O

The original Semantic Sensor Network Ontology (SSN) was developed through a W3C incubator group (Compton et al., 2012; Lefort et al., 2011). It built on a survey of pre-existing information models, particularly the OGC's Observations and Measurement (O&M) (Cox, 2017; ISO/TC 211, 2011), also adding a significant original contribution in the Stimulus-Sensor-Observation pattern (Janowicz & Compton, 2010)

The alignment of the original SSN ontology to PROV-O was the subject of a study by authors of both specifications (Compton et al., 2014), and was also discussed by (Cox, 2017) in the context of reconciliation of SSN with the original O&M model. A particular tension was that the original SSN was aligned with DOLCE-Ultralite in a way that meant that the `ssn:Observation` could not be interpreted as an activity or event. This meant that the term Observation in SSN was used differently to the precedent in O&M. Where Compton *et al.* proposed a new class 'Activity of sensing' to fill this hole, Cox questioned the choices made in the original DOLCE alignment which seemed to cause an unnecessary problem.

Taking these analyses into account, in the revised SSN ontology, developed through a joint W3C/OGC working group (Haller et al., 2017), the semantics of the term Observation have been clarified, so that it refers to an 'act of sensing'. Note that 'sensing' in this context also includes computation, simulation, interpretation, and any other process that applies a procedure to obtain an estimate of a property value. The scope of the revised SSN

was also expanded, to cover and Sampling alongside Observation. The ontology is modularized and the core terms are provided in a lightweight module ‘SOSA’ for Sensor, Observation, Sample, and Actuator.

A Procedure for observation, actuation or sampling is a kind of recipe or re-usable protocol thus:

```
sosa:Procedure rdfs:subClassOf prov:Plan .
```

Observation (which uses a Sensor to generate a Result – see Figure 4) is paralleled by the related acts of Actuation (which uses an Actuator to change an ActuatableProperty) and Sampling (which uses a Sampler to produce a Sample), thus

```
sosa:Observation      rdfs:subClassOf      prov:Activity .
sosa:Actuation        rdfs:subClassOf      prov:Activity .
sosa:Sampling         rdfs:subClassOf      prov:Activity .
```

A Sensor, Actuator or Sampler are entities that implement a procedure, and thus act as Agents in the context of a corresponding activity:

```
sosa:Sensor           rdfs:subClassOf      prov:Agent , prov:Entity .
sosa:Actuator         rdfs:subClassOf      prov:Agent , prov:Entity .
sosa:Sampler          rdfs:subClassOf      prov:Agent , prov:Entity .
```

The observation properties (see Figure 3) can be aligned to PROV-O (Figure 1) as follows:

```
sosa:hasFeatureOfInterest rdfs:subPropertyOf prov:used .
sosa:hasResult            rdfs:subPropertyOf prov:generated .
sosa:isResultOf           rdfs:subPropertyOf prov:wasGeneratedBy .
sosa:madeBySensor         rdfs:subPropertyOf prov:wasAssociatedWith .
sosa:resultTime           rdfs:subPropertyOf prov:endedAtTime .
```

where

```
sosa:FeatureOfInterest  rdfs:subClassOf      prov:Entity .
sosa:Sample             rdfs:subClassOf      prov:Entity .
sosa:ObservableProperty rdfs:subClassOf      prov:Entity .
sosa:Result             rdfs:subClassOf      prov:Entity .
```

The final property alignment first requires some sub-properties of PROV properties to be defined:

```
sp:eventAssociation      rdfs:subPropertyOf      prov:qualifiedAssociation ;
                        rdfs:domain              sosa:Observation .
sp:hadProcedure          rdfs:subPropertyOf      prov:hadPlan ;
                        rdfs:range              sosa:Procedure .
```

Then sosa:usedProcedure is given by a property chain axiom:

```
sosa:usedProcedure      owl:propertyChainAxiom ( sp:eventAssociation sp:hadProcedure ) .
```

A graph containing these axioms is included as a ‘vertical alignment module’ in the SSN/SOSA specification.

4. OBOE, SOSA AND PROV

OBOE, the Extensible Observation Ontology, is used within the biodiversity community for semantic representation of observation data (Schildhauer et al., 2016). An oboe:Observation is composed of a collection of oboe:Measurements with the same feature of interest. Each Measurement concerns a distinct characteristic and uses a distinct protocol (Figure 5).

We align sosa:Observation with oboe:Measurement, and the other classes and properties as follows:

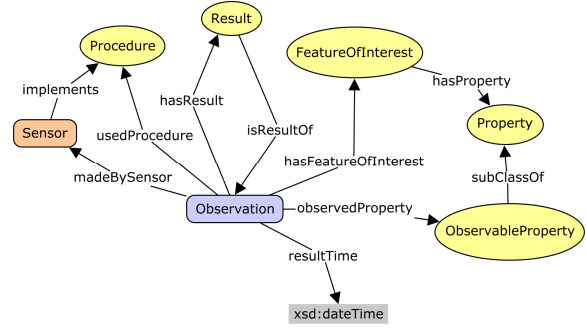


Figure 4. The core observation model from SSN, arranged for comparison with Figure 1

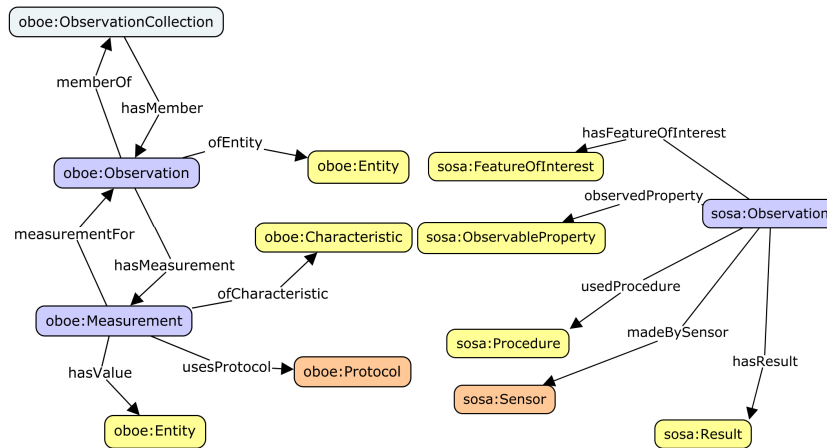


Figure 5. OBOE core aligned with SOSA.

```

oboe-core:Characteristic owl:equivalentClass      sos:ObservableProperty .
oboe-core:Measurement   owl:equivalentClass      sos:Observation .
oboe-core:Protocol       owl:equivalentClass      sos:Procedure .
oboe-core:hasMeasurement rdfs:subPropertyOf        rdfs:member .
oboe-core:hasValue       rdfs:subPropertyOf        sos:hasResult .
oboe-core:measurementFor rdfs:subPropertyOf        [ owl:inverseOf rdfs:member ; ] .
oboe-core:ofCharacteristic owl:equivalentProperty sos:observedProperty .
oboe-core:usesMethod      rdfs:subPropertyOf        sos:usedProcedure .
oboe-core:usesProtocol    rdfs:subPropertyOf        sos:usedProcedure .
sos:FeatureOfInterest     rdfs:subClassOf           oboe-core:Entity .
sos:Result                rdfs:subClassOf           oboe-core:Entity .
sos:hasFeatureOfInterest  owl:propertyChainAxiom ( oboe-core:measurementFor oboe-core:ofEntity ) .

```

Combining these axioms with those from section 3, we get the following entailments directly:

```

oboe-core:Characteristic rdfs:subClassOf      prov:Entity .
oboe-core:Measurement    rdfs:subClassOf      prov:Activity .
oboe-core:Protocol        rdfs:subClassOf      prov:Plan .
oboe-core:hasValue        rdfs:subPropertyOf   prov:generated .
oboe-core:usesMethod      rdfs:subPropertyOf   sos:usedProcedure .

```

and we can also suggest that

```

oboe-core:Entity          owl:equivalentClass      prov:Entity .
oboe-core:ofEntity        rdfs:subPropertyOf        prov:used .

```

Further property alignments require more elaborate restrictions and property chain axioms that are beyond the scope of this paper.

5. BCO AND OBI

The Biological Collections Ontology (BCO) (Walls et al., 2014) provides an alternative to OBOE, with greater emphasis on samples and specimens. Note that BCO uses a number of classes from the Ontology for Biomedical Investigations (OBI) (Brinkman et al., 2010) as well as BFO. We propose the following alignment between SSN/SOSA and BCO:

```

obo:BCO_0000003      rdfs:subClassOf      sos:Observation .      # bco:ObservingProcess
obo:BCO_0000044      rdfs:subClassOf      sos:FeatureOfInterest .
                                                              # bco:MaterialTargetOfObservation
obo:OBI_0000659      rdfs:subClassOf      sos:Sampling .          # obi:SpecimenCollectionProcess
obo:OBI_0100051      rdfs:subClassOf      sos:Sample .            # obi:Specimen
sos:Actuation         rdfs:subClassOf      obo:BFO_0000015 .      # bfo:Process
sos:Sampling          rdfs:subClassOf      obo:BFO_0000015 .      # bfo:Process
sos:Result            rdfs:subClassOf      obo:BFO_0000002 .      # bfo:Continuant
sos:Actuator          rdfs:subClassOf      obo:BFO_0000004 .      # bfo:IndependentContinuant
sos:Sensor            rdfs:subClassOf      obo:BFO_0000004 .      # bfo:IndependentContinuant
sos:Sampler           rdfs:subClassOf      obo:BFO_0000004 .      # bfo:IndependentContinuant
sos:Platform          rdfs:subClassOf      obo:BFO_0000004 .      # bfo:IndependentContinuant
sos:System            rdfs:subClassOf      obo:BFO_0000004 .      # bfo:IndependentContinuant
sos:Sample            rdfs:subClassOf      obo:BFO_0000004 .      # bfo:IndependentContinuant

```

```
ssn:Property          owl:equivalentClass    obo:BFO_0000019 . # bfo:Quality
sosa:ObservableProperty rdfs:subClassOf      obo:BFO_0000019 . # bfo:Quality
```

Since `sosa:Observation` also covers computation, simulation, interpretation it is more general than `bco:ObservingProcess`, which is therefore a subclass. Similarly, `sosa:Sample` and `sosa:Sampling` involve more kinds of sample than just material samples or specimens, so the OBI classes are subclasses of the corresponding SOSA classes. Combining with the SSN-PROV alignment from section 3, we get these initial entailments:

```
obo:BCO_0000003    rdfs:subClassOf    prov:Activity . # bco:ObservingProcess
obo:OBI_0000659    rdfs:subClassOf    prov:Activity . # obi:SpecimenCollectionProcess
obo:BFO_0000019    rdfs:subClassOf    prov:Entity . # bfo:Quality
obo:BCO_0000044    rdfs:subClassOf    prov:Entity . # bfo:MaterialTargetOfObservation
obo:OBI_0100051    rdfs:subClassOf    prov:Entity . # obi:Specimen
```

This is consistent with the PROV-O – BFO alignment described in section 2.2.

6. CONCLUSION AND FUTURE WORK

The PROV model and PROV-O ontology provide a convenient framework for analyzing and aligning observation models and ontologies. The core elements of PROV capture the key parts of a generic process-flow model, which matches sensing and observation, as well as the related activities of actuation and sampling. The simple 3-class base model from PROV appears to provide just enough logic to serve as a lightweight upper ontology, particularly for workflow or process-based information. Alignment with the simple PROV-O classes clarifies the relationships between observation models and ontologies in a way that had previously proved difficult to resolve. The axiomatization presented here is provisional and incomplete. While we have shown that the key classes in the observation ontologies can be tied back to the core PROV classes, further investigations will be required to verify that other inconsistencies are not introduced.

ACKNOWLEDGEMENTS

This work was supported primarily by CSIRO's OzNome initiative.

REFERENCES

- Beckett, D., Berners-Lee, T., Prud'hommeaux, E., & Carothers, G. (2014). RDF 1.1 Turtle. *W3C Recommendation*. World Wide Web Consortium. Retrieved from <https://www.w3.org/TR/turtle/>
- Brickley, D., & Guha, R. V. (2014). RDF Schema 1.1. Cambridge, Mass. USA: World Wide Web Consortium. Retrieved from <http://www.w3.org/TR/rdf-schema/>
- Brinkman, R. R., Courtot, M., Derom, D., Fostel, J. M., He, Y., Lord, P., ... Zheng, J. (2010). Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*, 1(Suppl 1), S7. <https://doi.org/10.1186/2041-1480-1-S1-S7>
- Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S. J. D., ... Taylor, K. (2012). The SSN ontology of the W3C semantic sensor network incubator group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17, 25–32. <https://doi.org/10.1016/j.websem.2012.05.003>
- Compton, M., Corsar, D., & Taylor, K. (2014). Sensor Data Provenance: SSNO and PROV-O Together at Last. In M. Compton, K. Janowicz, & K. Taylor (Eds.), *7th International Workshop on Semantic Sensor Networks* (p. 16). Riva del Garda, Trentino Italy: Kno.e.sis, Wright State University, Dayton, OH, USA. Retrieved from <http://ceur-ws.org/Vol-1401/paper-05.pdf>
- Cox, S. J. D. (2007a). Observations and Measurements - Part 1 - Observation schema. *OGC 07-022r1*. Wayland, Mass.: Open Geospatial Consortium. Retrieved from <http://portal.opengeospatial.org/files/22466>
- Cox, S. J. D. (2007b). Observations and Measurements – Part 2 - Sampling Features. *OGC 07-002r3*. Wayland, Mass.: Open Geospatial Consortium. Retrieved from <http://portal.opengeospatial.org/files/22467>
- Cox, S. J. D. (2011). Topic 20 - Geographic Information - Observations and Measurements (same as ISO 19156:2011). *OGC Abstract Specification*, 10–004r3, 54. <https://doi.org/10.13140/2.1.1142.3042>
- Cox, S. J. D. (2017). Ontology for observations and sampling features, with alignments to existing models. *Semantic Web*, 8(3), 453–470. <https://doi.org/10.3233/SW-160214>
- Cox, S. J. D., & Little, C. (2017). Time Ontology in OWL. World Wide Web Consortium. Retrieved from <https://www.w3.org/TR/owl-time/>
- Gangemi, A. (2010). Ontology: DOLCE+DnS Ultralite - Ontology Design Pattern. Retrieved February 13, 2014, from http://ontologydesignpatterns.org/wiki/Ontology:DOLCE+DnS_Ultralite

- Grenon, P., & Smith, B. (2004). SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation*, 4(1), 69–103. https://doi.org/10.1207/s15427633scc0401_5
- Guizzardi, G., & Wagner, G. (n.d.). A Unified Foundational Ontology and some Applications of it in Business Modeling. Retrieved from <http://tmiswww.tm.tue.nl/staff/gwagner>
- Guizzardi, G., & Wagner, G. (2010). Using the Unified Foundational Ontology (UFO) as a Foundation for General Conceptual Modeling Languages. In *Theory and Applications of Ontology: Computer Applications* (pp. 175–196). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-8847-5_8
- Haller, A., Janowicz, K., Cox, S. J. D., Le Phuoc, D., Taylor, K., & LeFrancois, M. (2017). Semantic Sensor Network Ontology. World Wide Web Consortium. Retrieved from <https://www.w3.org/TR/vocab-ssn/>
- Herre, H. (2010). General Formal Ontology (GFO): A Foundational Ontology for Conceptual Modelling. In *Theory and Applications of Ontology: Computer Applications* (pp. 297–345). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-90-481-8847-5_14
- Hobbs, J. R., & Pan, F. (2006). Time Ontology in OWL. *W3C Working Draft*. Cambridge, Mass. USA: World Wide Web Consortium. Retrieved from <http://www.w3.org/TR/owl-time/>
- Horsburgh, J. S., Aufdenkampe, A. K., Mayorga, E., Lehnert, K. A., Hsu, L., Song, L., ... Whitenack, T. (2016). Observations Data Model 2: A community information model for spatially discrete Earth observations. *Environmental Modelling & Software*, 79, 55–74. <https://doi.org/10.1016/j.envsoft.2016.01.010>
- ISO/TC 211. (2011). *ISO 19156:2011 Geographic Information – Observations and Measurements*. (S. J. D. Cox, Ed.). Geneva: International Organization for Standardization. Retrieved from <https://www.iso.org/standard/32574>
- Janowicz, K., & Compton, M. (2010). The Stimulus-Sensor-Observation Ontology Design Pattern and its Integration into the Semantic Sensor Network Ontology. In K. Taylor, A. Ayyagari, & D. de Roure (Eds.), *3rd International Workshop on Semantic Sensor Networks (SSN10)* (p. 12pp). Shanghai, China: CEUR Workshop Proceedings. Retrieved from http://geog.ucsb.edu/~jano/Semantic_Sensor_Ontology_2010.pdf
- Lebo, T., Sahoo, S., & McGuinness, D. L. (2013). PROV-O: The PROV Ontology. *W3C Recommendation*. Cambridge, Mass. USA: World Wide Web Consortium. Retrieved from <http://www.w3.org/TR/prov-o/>
- Lefort, L., Henson, C., Taylor, K., Barnaghi, P., Compton, M., Corcho, O., ... Page, K. (2011). *Incubator Report - Semantic Sensor Network Incubator Group. W3C Incubator Report*. W3C. Retrieved from http://www.w3.org/2005/Incubator/ssn/wiki/Incubator_Report
- Madin, J., Bowers, S., Schildhauer, M., Krivov, S., Pennington, D., & Villa, F. (2007). An ontology for describing and synthesizing ecological observation data. *Ecological Informatics*, 2, 279–296. <https://doi.org/10.1016/j.ecoinf.2007.05.004>
- Mascardi, V., Cordi, V., & Rosso, P. (2007). A Comparison of Upper Ontologies. (M. Baldoni, A. Boccalatte, F. De Paoli, M. Martelli, & V. Mascardi, Eds.), *Proceedings of WOA, Workshop Dagli Oggetti Agli Agenti, September 24-25*. Genova, Italy: Seneca Edizioni Torino. <https://doi.org/10.1.1.107.1689>
- Schildhauer, M., Jones, M. B., Bowers, S., Madin, J., Krivov, S., Pennington, D., ... O'Brien, M. (2016, January 1). OBOE: the Extensible Observation Ontology, version 1.1. <https://doi.org/10.5063/F11C1TTM>
- W3C OWL Working Group. (2012). OWL 2 Web Ontology Language Document Overview (Second Edition). *W3C Recommendation*. Cambridge, Mass. USA: World Wide Web Consortium. Retrieved from <http://www.w3.org/TR/owl2-overview/>
- Walls, R. L., Deck, J., Guralnick, R., Baskauf, S., Beaman, R., Blum, S., ... Wooley, J. (2014). Semantics in support of biodiversity knowledge discovery: an introduction to the biological collections ontology and related ontologies. *PloS One*, 9(3). <https://doi.org/10.1371/journal.pone.0089606>