An RDF guide for the Darwin Core standard

Editor(s):
Solicited review(s):
Open review(s):

Steven J. Baskauf[a,*], John Wieczorek[b], John Deck[c], and Campbell O. Webb[d]
[a]*Department of Biological Sciences, PMB 351634, Vanderbilt University, Nashville, TN 37146, USA*
E-mail: steve.baskauf@vanderbilt.edu
[b]*Museum of Vertebrate Zoology, University of California at Berkeley, 3101 Valley Life Sciences Building, Berkeley, CA 94720, USA*
E-mail: tuco@berkeley.edu
[c] *Berkeley Natural History Museums, University of California at Berkeley, 1001 Valley Life Sciences Building, Berkeley, CA 94720, USA*
Email: jdeck@berkeley.edu
[d]*Arnold Arboretum of Harvard University, 1300 Centre St., Boston 02131, USA*
E-mail: cw@camwebb.info

Abstract. The Darwin Core vocabulary is widely used to transmit biodiversity data in the form of simple text files. In order to support expression of biodiversity data in the Resource Description Framework (RDF), a guide was created as a non-normative addition to the Darwin Core standard. The guide resolves a number of issues that arise from adapting terms designed to have literal values for use with URI references. Although there are some problems that are beyond the scope of the guide, the guide is an important step towards enabling the biodiversity informatics community to participate in broader Linked Data and Semantic Web efforts.

Keywords: biodiversity, informatics, standards, semantics, interoperability

1. Introduction

Darwin Core (DwC) [1] is a technical standard[1] of Biodiversity Information Standards[2] (TDWG). Since its ratification in 2009, Darwin Core has been widely used to publish and transmit data. For example, as of 29 Jan 2014 the Global Biodiversity Information Facility (GBIF) network has aggregated over 428 million Darwin Core records.[3]

The Darwin Core standard contains a general-purpose vocabulary for describing biodiversity resources. It includes several guides that describe how the vocabulary terms should be used to transmit data in various formats such as simple text and XML. Darwin Core roughly follows the Dublin Core Abstract Model,[4] which was built on Resource Description Framework[5] (RDF) concepts. A DwC term is designated as a URI, which can be dereferenced to access its normative RDF/XML metadata. Despite these connections to RDF, there has previously been no guide to describe how DwC vocabulary terms should be used to describe biodiversity resources as RDF.

The Darwin Core vocabulary was primarily designed to facilitate the sharing of data in simple text files containing a single table of rows and columns. Extensive discussion on the TDWG mailing list[6] between 2009 and 2011 showed that there was significant interest in expressing Darwin Core data in RDF

[*] Corresponding author.
[1] http://www.tdwg.org/standards/450/; enter reference website at http://rs.tdwg.org/dwc/terms/
[2] http://www.tdwg.org/
[3] http://www.gbif.org/
[4] http://dublincore.org/documents/abstract-model/
[5] http://www.w3.org/TR/rdf-concepts/
[6] http://lists.tdwg.org/pipermail/tdwg-content/

and identified several issues that impeded the effective use of Darwin Core terms in RDF:

- Text files often use name strings as values of properties, whereas it is often advantageous in RDF to have URI reference objects.
- Simple fielded text files (where a line in the file contains the data for a single record) are by their nature “flat”, while RDF is often highly normalized. (We use the term "flat" to denote data which is expressed in a single table.)
- Darwin Core defines a number of “ID” terms, which are used in flat files to specify identifiers and record type, whereas RDF provides different standard methods for expressing identity and class membership.

In 2011, an RDF/OWL Task Group[7] was chartered by TDWG, and in 2012 a team of writers began work on a Darwin Core RDF guide to address the issues listed above by providing a set of best practices and by creating some new Darwin Core terms intended specifically for use in RDF. The guide[8] was completed in 2013 and reviewed by the Task Group, which recommended it for adoption. Upon adoption by TDWG, the RDF guide will become a non-normative part of the Darwin Core standard.

In this paper, URIs are sometimes abbreviated as QNames using standard namespace prefixes, e.g. `dwc:recordedBy`. The prefixes are defined in footnotes or legends. URIs, RDF serializations, and SPARQL queries are written in `Courier` font. In many cases, URIs identify real resources in the wild, although example triples containing those URIs are not necessarily asserted there.

Section 2 of this paper lays out how each major issue was addressed by the guide. Section 3 describes several challenges to the implementation of Darwin Core as RDF that are beyond the scope of the guide. Section 4 concludes the paper.

2. Issues and their resolution in the guide

2.1. Terms intended for use with non-literal objects

Because the Darwin Core vocabulary was designed primarily to facilitate the transfer of text-based records from relatively flat database tables, definitions and comments for terms in the general namespace `dwc:`[9] suggest using text strings to refer to physical and conceptual entities, e.g., names to represent people, citations to represent articles, codes to represent institutions, etc. When a record has multiple values for a property, these term definitions specify that the multiple strings be concatenated and delineated in a single field to avoid forcing the creation of a more normalized data structure. These specifications are at odds with best practices for RDF, where it is preferable in triples to identify non-literal objects using URI references. Those URIs can then be associated with additional properties that describe the non-literal resource.

The conflicting demands of flat, string-based tables and normalized, graph-based RDF creates a problem when terms that were originally designed for use with text strings are co-opted for use with non-literal objects in RDF. This is a longstanding problem[10] that is not unique to Darwin Core. The Dublin Core RDF Guidelines[11] provide a mechanism to permit legacy string literal data to be associated with terms that were not intended for use with literal objects. This mechanism, which involves use of the `rdf:value`[12] property, has not been widely implemented. A more widely used dual-term alternative allows Dublin Core terms in the legacy `dc:`[13] namespace to be used with literal values,[14] while reserving terms in the `dcterms:`[15] namespace that have declared non-literal ranges for use with URI reference or blank node objects.

2.1.1. New Darwin Core terms in the `dwcuri:` *namespace*

The Darwin Core RDF guide adopts the dual-term approach by creating a new Darwin Core namespace, `dwcuri:`,[16] whose terms are intended for use only with non-literal objects. For example, the existing Darwin Core term `dwc:recordedBy` would continue to be used with a value that consisted of a name string for agents who recorded an occurrence, whereas the new term `dwcuri:recordedBy` would relate the subject to a non-literal object (URI reference or blank node) that denotes the actual agent.

[7] http://code.google.com/p/tdwg-rdf/

[8] The draft guide is at http://code.google.com/p/tdwg-rdf/wiki/DwcRdf with an eventual permanent URL of http://rs.tdwg.org/dwc/terms/guides/rdf/

[9] `http://rs.tdwg.org/dwc/terms/`

[10] http://wiki.foaf-project.org/w/UsingDublinCoreCreator

[11] http://dublincore.org/documents/dc-rdf/#sect-4

[12] `rdf: = http://www.w3.org/1999/02/22-rdf-syntax-ns#`

[13] `http://purl.org/dc/elements/1.1/`

[14] http://wiki.dublincore.org/index.php/User_Guide/Publishing_Metadata#Legacy_namespace

[15] `http://purl.org/dc/terms/`

[16] `http://rs.tdwg.org/dwc/uri/`

```
<http:// arctos.database.museum/guid/MVZ:Mamm:115956>
     dwc:recordedBy "Oliver P. Pearson; Anita K. Pearson";
     dwcuri:recordedBy <http://viaf.org/viaf/263074474>,
                       <http://museum-x.org/personnel/akp>.
```

Fig. 1. Recorders of an occurrence (serialized as RDF/Turtle)

```
<http://bioimages.vanderbilt.edu/baskauf/00001#loc>
     a dcterms:Location;
     dwc:continent "North America";
     dwc:country "United States";
     dwc:stateProvince "Tennessee";
     dwc:county "Robertson".
```

Fig. 2. Darwin Core convenience terms describing the political subdivisions of a location (serialized as RDF/Turtle)

```
<http://bioimages.vanderbilt.edu/baskauf/00001#loc>
     a dcterms:Location;
     dwcuri:inDescribedPlace <http://sws.geonames.org/4653638/>.
```

Fig. 3. Using a `dwcuri:` term to link a location to its lowest level political subdivision (serialized as RDF/Turtle)

The guide allows legacy string name data in the form of concatenated lists to continue to be exposed in RDF as a literal object of a `dwc:` namespace term. However, the guide specifies that if a record using a term from the general `dwc:` namespace is serialized as RDF using a `dwcuri:` namespace term, each non-literal resource in a concatenated list of names should be the object of separate triple. This is illustrated in Fig. 1.

An advantage of the dual-term approach is that it allows large databases consisting of legacy string name data to be exposed immediately as RDF without imposing a requirement that the provider immediately implement the use of URI identifiers for non-literal resources.

2.1.2. Convenience terms

Darwin Core contains several collections of hierarchical terms designed to provide a set of text-based property/value pairs that will unambiguously specify a resource. These sets describe: ownership of a collection item, a taxonomic entity, names of geographic subdivisions, chronostratographic descriptors, and lithostratigraphic descriptors. The example in Fig. 2 shows how the terms `dwc:county`, `dwc:stateProvince`, `dwc:country`, and `dwc:continent` allow a location to be placed in its geographical context.

No single term value is sufficient to unambiguously place the location in its lowest level political subdivision because there may be several low level political subdivisions having the same name that are contained within different upper level political subdivisions. Thus each location record must provide the entire set. In the context of a flat database structure, it is convenient to expose the full set of property/value pairs for a location since that would allow a user to query for locations in the database by specifying the particular values of interest for certain properties in the set (hence the name "convenience terms" for properties that are included in such sets to make searching convenient).

It would be possible to define `dwcuri:` analogues for all convenience properties included in Darwin Core. However, this does not make sense in the context of RDF. Requiring a data provider to specify a URI value for every resource in the hierarchy essentially requires that provider to define the hierarchy in every record of the dataset. It should be possible to specify particular hierarchical sets of property/value pairs and the relationships among levels in the hierarchy in a standardized external database. In that case, a provider need only specify a URI for the lowest level in an administrative subdivision hierarchy, and clients consuming the RDF could

A. Asserted RDF (serialized as RDF/Turtle):

```
<http://guid.mvz.org/collectingEvents/23459>
     a dwc:Event;
     dwc:locationID <http://locations.mvz.org/493056921>.
```

intending `http://locations.mvz.org/493056921` to be the URI of the Location at which the Event happened.

B. Triple entailed by `rdfs:subPropertyOf` relationship:

```
<http://guid.mvz.org/collectingEvents/23459>
     dcterms:identifier <http://locations.mvz.org/493056921>.
```

implying (incorrectly) that `http://locations.mvz.org/493056921` is the identifier of the Event.

C. Triple entailed by range declaration of `dcterms:identifier`:

```
<http://locations.mvz.org/493056921> a rdfs:Literal.
```

implying (incorrectly) that the URI reference `http://locations.mvz.org/493056921` is a literal value.

Fig. 4. Unintended effects of using a Darwin Core ID term as an object property

discover the higher levels by retrieving information from the external database. Following this approach would alleviate the need for data providers to update their database each time there is a change in the upper levels of the hierarchy (change in spelling, reassignment of lower level resources to different upper level resources, reorganization of upper levels, etc.).

For each of the kinds of convenience terms in Darwin Core, a term has been defined in the `dwcuri:` namespace that is intended to be used as a property that links a resource to the lowest known level in the hierarchy described by that kind of term. In Fig. 3, the term `dwcuri:inDescribedPlace` links the location of Fig. 2 to the URI of its lowest known level in a standard reference of geographic subdivisions.

By dereferencing the GeoNames[17] URI, a client can discover all of the higher levels in the hierarchy of geographic subdivisions. Implementing this means that a SPARQL[18] query using the Robertson County, Tennessee, USA URI would not depend on consistent spelling of "Robertson", "Robertson Co.", "Robertson County", "United States", "United States of America", "U.S.", "USA", "États-Unis", etc.

2.2. Identity and type

2.2.1. Darwin Core ID terms

The Darwin Core vocabulary includes a number of terms whose local name ends in "ID" (e.g., `dwc:occurrenceID`, `dwc:locationID`, `dwc:identificationID`, etc.), known collectively as "ID terms". The ID terms were designed to perform a particular function in the context of a record in a flat database. A particular record might contain identifiers for the resource that was the subject of that record as well as identifiers for other resources related to the subject record (i.e., identifiers to serve as foreign keys). Because a particular table might contain several identifiers, using a generic term such as `dcterms:identifier` as a field header would be ambiguous. The approach taken by Darwin Core was to define ID terms that would serve the dual purpose of indicating that a field contained an identifier and to indicate the type of the resource referenced by the identifier.

[17] http://www.geonames.org/

[18] http://www.w3.org/TR/sparql11-overview/

```
<http://biocol.org/urn:lsid:biocol.org:col:15685>
     owl:sameAs <urn:lsid:biocol.org:col:15685>.

<http://zoobank.org/8FE313DD-BB2C-47CB-805E-B87E320D1864>
     owl:sameAs <urn:uuid:8FE313DD-BB2C-47CB-805E-B87E320D1864>.
```

Fig. 5. Examples declaring equivalence of URNs to HTTP URI proxies (serialized as RDF/Turtle)

It is problematic to convert flat Darwin Core records to RDF by using their ID terms and values as predicates and objects in triples. The first problem is that a pre-existing understanding between the data provider and consumer is required to know which of several ID term fields represents the identifier for the record (i.e., provides the identifier for the subject resource). A second problem stems from a property assigned to all ID terms in their normative RDF. Each ID term is declared to be `rdfs:subPropertyOf dcterms:identifier`. If a data provider were to use an ID term in a triple to provide the value of a foreign key string that references a resource that is related to the subject (Fig. 4.A), a client performing reasoning on that triple would incorrectly infer that that the object of the triple was the identifier of the subject resource and not the identifier of the object resource as the provider intended (Fig. 4.B).

It has also been suggested that the ID terms be used as object properties that can link subject resources to the URIs of related resources of other classes. In addition to the incorrect inference mentioned above, there is an additional problem caused by the subproperty declaration. Since the range of `dcterms:identifier` is `rdfs:Literal`, using an ID term as an object property could cause a client to reason that the URI-identified non-literal resource is a literal (Fig. 4.C).

Because of these problems associated with the use of ID terms in RDF, the Darwin Core RDF guide states that ID terms should not be used as predicates in RDF triples.

2.2.2. Associating an identifier with a subject resource

Darwin Core is not strict about the identifiers that are used as values of its properties. Although globally unique identifiers are recommended, identifiers specific to a data set are allowed. There is also no requirement that globally unique identifiers be URIs. Thus the RDF guide provides some guidelines for translating the various kinds of values of ID terms into RDF.

If the subject resource identifier is a URI, that URI is simply asserted as the subject of triples describing the subject resource. If the subject resource identifier is a non-URI string, the string is presented as the literal value of a `dcterms:identifier` property.

In the past, TDWG has recommended the use of Life Science Identifiers (LSIDs).[19] As URIs, LSIDs may be the subjects of RDF triples. However, Recommendation 30 of the TDWG LSID Applicability Statement standard[20] requires that "The description of all objects identified by an LSID must contain an owl:sameAs, owl:equivalentProperty or owl:equivalentClass statement expressing the equivalence between the object identifier in its standard form and its proxy version". The Darwin Core RDF guide extends this recommendation to any non-HTTP URI (i.e., including other varieties of URNs such as ARK, UUID, ISBN, etc.) by specifying that the subject resource be identified by an HTTP-proxy version of the non-HTTP URI, and that the non-HTTP URI be the object of an `owl:sameAs`[21] property of a triple having the HTTP URI as the subject (Figs. 5 and 7).

2.2.3. Specifying the type of a resource

The TDWG GUID Applicability Statement standard[22] specifies that an object in the biodiversity domain that is identified by a GUID should be typed using a well-known vocabulary. As a well-known standard for biodiversity resources, Darwin Core is a logical source of classes to be used as values for `rdf:type` statements. Historically, there has been considerable confusion about the definitions of Darwin Core classes that were present in two different namespaces. In parallel to the creation of the RDF guide, there has been an effort to create a single set

[19] http://www.omg.org/cgi-bin/doc?dtc/04-05-01.pdf
[20] http://www.tdwg.org/standards/150/
[21] `http://www.w3.org/2002/07/owl#sameAs`
[22] http://www.tdwg.org/standards/150/

Occurrence database table

dwc:occurrenceID	dwc:recordedBy	dwc:eventDate	dwc:locationID
urn:catalog:MVZ:Mamm:115987	Oliver P. Pearson; Anita K. Pearson	1952-04-13	http://guid.mvz.org/sites/per/127

Location database table

dwc:locationID	dwc:country	dwc:stateProvince	dwc:locality
http://guid.mvz.org/sites/per/127	Peru	Puno	Pampa de Titre, 29 km NE Tarata

Fig. 6. Example records from database tables with field names based on Darwin Core terms from the namespace `dwc:` = `http://rs.tdwg.org/dwc/terms/`

of clearly defined classes that can be used to identify the type of biodiversity-related resources.

Generally, each ID term has a corresponding class term, so when flat databases that use ID terms are translated to RDF, records can be typed according to the parent class of their subject ID term (illustrated in Figs. 6 and 7).

Because Darwin Core also imports terms from Dublin Core that have range or domain declarations, the guide draws attention to the fact that use of those terms also entails type relationships that may not be explicitly declared.

3. Limitations of the Darwin Core RDF guide

3.1. Terms having complex functions

Even though the guide provides guidance for using most Darwin Core properties as RDF predicates, there are several categories of DwC properties that are not easily translated to RDF. Auxiliary terms listed under the `dwc:ResourceRelationship` and `dwc:MeasurementOrFact` classes, and the `dwc:dynamicProperties` term are too complex to suggest prescriptive translations. In these cases, users are directed to ancillary web pages outside the standard where suggestions are being developed to help users generate RDF based on these terms.

3.2. Lack of object properties

Although the creation of properties within the `dwcuri:` namespace allows biodiversity resources to be linked to some types of related resources, there are no object properties in Darwin Core that are suitable for linking instances of the main Darwin Core classes (that is, those classes not considered auxiliary terms). Lack of standards in this regard is a serious impediment to efforts to expose Darwin Core records as RDF. Figs. 6 and 7 illustrate this problem.

The database records in Fig. 6 are found in two tables: one representing Occurrences and the other representing Locations. The DwC ID property `dwc:locationID` is used to link a record in the Occurrence table to a record in the Location table.

Although the two tables imply the existence of instances of two classes, the tables could be considered to contain information about instances of five classes: `dwc:Occurrence`, `dwc:Event`, `dcterms:Location`, `dcterms:Agent` (or `foaf:Agent`[23]), and `gn:Feature`.[24] When the data in the tables are expressed as RDF according to the DwC RDF guide, `dwcuri:recordedBy` is used to link the Occurrence to the Agent recording it, and `dwcuri:inDescribedPlace` is used to link the Location to a standardized geographic Feature. However, there are currently no terms in Darwin Core that can be used to link the Occurrence, Event, and Location classes.

Fig. 7 shows how the data in Fig. 6 can be serialized as RDF under several non-Darwin Core models.[25] The TDWG Ontology[26] does not include the notion of separate classes for Event and Location, so properties related to those classes are grouped as properties of the Occurrence instance in Fig. 7.A. The TaxonConcept ontology[27] includes the notion of both Occurrence and Location (in the form of the `txn:Area`[28] class) but does not recognize a separate

[23] `http://xmlns.com/foaf/0.1/Agent`

[24] `http://www.geonames.org/ontology#Feature`

[25] The models are compared at http://code.google.com/p/tdwg-rdf/wiki/BiodiversityOntologies

[26] http://wiki.tdwg.org/twiki/bin/view/TAG/TDWGOntology

[27] http://www.taxonconcept.org/

[28] `txn:` = `http://lod.taxonconcept.org/ontology/txn.owl#`

A. Serialized as RDF/Turtle using the TDWG Ontology model:

```
<http://arctos.database.museum/guid/MVZ:Mamm:115987>
     a dwc:Occurrence;
     owl:sameAs <urn:catalog:MVZ:Mamm:115987>;
     dwc:recordedBy "Oliver P. Pearson; Anita K. Pearson";
     dwcuri:recordedBy <http://viaf.org/viaf/263074474>,
                       <http://museum-x.org/personnel/akp>;
     dwc:eventDate "1952-04-13"^^xsd:date;
     dwc:locality "Pampa de Titre, 29 km NE Tarata";
     dwc:country "Peru";
     dwc:stateProvince "Puno";
     dwcuri:inDescribedPlace <http://sws.geonames.org/3931274/>.
```

B. Serialized as RDF/Turtle using TaxonConcept object properties:

```
<http://arctos.database.museum/guid/MVZ:Mamm:115987>
     a dwc:Occurrence;
     owl:sameAs <urn:catalog:MVZ:Mamm:115987>;
     dwc:recordedBy "Oliver P. Pearson; Anita K. Pearson";
     dwcuri:recordedBy <http://viaf.org/viaf/263074474>,
                       <http://museum-x.org/personnel/akp>;
     dwc:eventDate "1952-04-13"^^xsd:date;
     txn:occurrenceHasArea <http://guid.mvz.org/sites/per/127>.

<http://guid.mvz.org/sites/per/127>
     a dcterms:Location;
     dwc:locality "Pampa de Titre, 29 km NE Tarata";
     dwc:country "Peru";
     dwc:stateProvince "Puno";
     dwcuri:inDescribedPlace <http://sws.geonames.org/3931274/>.
```

C. Serialized as RDF/Turtle using Darwin-SW object properties:

```
<http://arctos.database.museum/guid/MVZ:Mamm:115987>
     a dwc:Occurrence;
     owl:sameAs <urn:catalog:MVZ:Mamm:115987>;
     dwc:recordedBy "Oliver P. Pearson; Anita K. Pearson";
     dwcuri:recordedBy <http://viaf.org/viaf/263074474>,
                       <http://museum-x.org/personnel/akp>;
     dsw:atEvent <http:// arctos.database.museum/guid/MVZ:Mamm:14523#event>.

<http://arctos.database.museum/guid/MVZ:Mamm:14523#event>
     a dwc:Event;
     dwc:eventDate "1952-04-13"^^xsd:date;
     dsw:locatedAt <http://guid.mvz.org/sites/per/127>.

<http://guid.mvz.org/sites/per/127>
     a dcterms:Location;
     dwc:locality "Pampa de Titre, 29 km NE Tarata";
     dwc:country "Peru";
     dwc:stateProvince "Puno";
     dwcuri:inDescribedPlace <http://sws.geonames.org/3931274/>.
```

Fig. 7. RDF/Turtle serialization of the data in Fig. 6 using three ontologies outside of Darwin Core. Namespace abbreviations used are: dwc: = http://rs.tdwg.org/dwc/terms/, dwcuri: = http://rs.tdwg.org/dwc/uri/, txn: = http://lod.taxonconcept.org/ontology/txn.owl#, dsw: = http://purl.org/dsw/, dcterms: = http://purl.org/dc/terms/, owl: = http://www.w3.org/2002/07/owl#, and xsd: = http://www.w3.org/2001/XMLSchema#

```
PREFIX dwcuri: <http://rs.tdwg.org/dwc/uri/>
PREFIX dsw: <http://purl.org/dsw/>

SELECT ?occurrence WHERE
  {
  ?location dwcuri:inDescribedPlace <http://sws.geonames.org/3931274/>.
  ?event dsw:locatedAt ?location.
  ?occurrence dsw:atEvent ?event.
  }
```

Fig. 8. SPARQL query based on Darwin-SW object properties.

class for Event. Therefore the serialization in Fig. 7.B uses a single object property `txn:occurrenceHasArea` to link the occurrence and location instance. The Darwin-SW ontology[29] adopts all of the main Darwin Core classes and therefore includes Occurrence, Event, and Location classes. The serialization in Fig. 7.C is more normalized than the original database, requiring either a placeholder URI (created using an `#event` fragment identifier in the example) or blank node to represent the Event instance. Two object properties, `dsw:atEvent` and `dsw:locatedAt`,[30] are used to link the three Darwin Core classes.

Fig. 8 shows a SPARQL query designed to find occurrences recorded for Puno Province, Peru by querying for its GeoNames URI. Because the object properties used in the query are the Darwin-SW properties that link the Occurrence, Event, and Location classes, the query would be successful in finding the desired occurrences in any data serialized as in Fig. 7.C. However, it would not find those same Occurrences if the data were serialized using the classes and object properties included in the TDWG Ontology (Fig. 7.A) or TaxonConcept ontology (Fig. 7.B).

It would be possible to merge graphs from providers that used different models and object properties and then to adjust by creating complex queries. However, standardization and consistent use of object properties among providers would make data integration and querying much simpler. Creating a uniform set of object properties to link Darwin Core classes is contingent on the development of a consensus model for the biodiversity informatics domain and that is an effort beyond the scope of the Darwin Core RDF guide.

3.3. Need for guidance on the assignment of properties

The Darwin Core Quick Reference Guide[31] lists most terms under headings which are Darwin Core classes. In general, this implies that those terms should be used as properties of instances of the class under which they are listed. In many cases, the type of resource with which a particular property should be associated is apparent. However, in other cases, there is confusion about the type of resource with which a particular property should be associated. This is particularly true for Darwin Core classes which currently lack clear definitions, such as `dwc:Occurrence` and `dwc:Taxon`. Current efforts to clarify the definitions of the DwC classes will help to alleviate this problem. But this problem is also somewhat contingent on the development of a consensus domain model. For example, in Fig. 7 whether the property `dwc:eventDate` was assigned to an Occurrence or an Event depended on whether the underlying model included an Event class or not.

4. Conclusions

Obstacles to expressing Darwin Core data as RDF include:

- unclear or inconsistent use of identifiers
- reliance on strings to identify non-literal resources
- single literal values that consist of concatenated and separated lists
- unclear or inconsistent typing of resources
- lack of object properties to link instances of the main DwC classes

The Darwin Core RDF guide addresses the first four of these obstacles. The ability to consistently express most Darwin Core properties as RDF will facilitate

[29] http://code.google.com/p/darwin-sw/

[30] `dsw: = http://purl.org/dsw/`

[31] http://rs.tdwg.org/dwc/terms/index.htm

the development and testing of a consensus domain model, which will enable the TDWG community to define the missing object properties and open the door for the integration of traditionally "flat" biodiversity data with the vast and growing body of semantically-enabled information.

Acknowledgements

The authors would like to thank the members of the TDWG RDF/OWL Task Group for feedback on the draft Darwin Core RDF guide.

Steve Baskauf's participation in the Semantics of Biodiversity Symposium at TDWG 2013 was supported by the Research Coordination Network for the Genomic Standards Consortium (RCN4GSC, NSF DBI-0840989) and the Scientific Observations Network (SONet, NSF #0753144, OCI-Interop).

References

[1] J. Wieczorek, D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, D. Vieglais, Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLOS ONE **7**(1):e29715, 2012. http://dx.doi.org/10.1371/journal.pone.0029715