Lessons Learned from Adapting the Darwin Core Vocabulary Standard for Use in RDF

Editors: Pascal Hitzler, Wright State University, USA; Krzysztof Janowicz, University of California, Santa Barbara, CA
Solicited reviews: two anonymous reviewers

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. This paper describes the major issues that were addressed in the creation of the guide, particularly problems related to adapting terms designed to have literal values for use with IRI references. By making it possible to express millions of existing records as RDF, 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

Terms that are commonly used as predicates in Resource Description Framework[1] (RDF) often come from vocabularies that were specifically designed for use in RDF. For example, two vocabularies commonly used as a source of RDF predicates, Friend of a Friend (FOAF) [1] and Simple Knowledge Organization System (SKOS) [2], were defined as OWL ontologies[2] [3] and have always been intended to be used as a source of RDF predicates. Another well-known vocabulary, Dublin Core [3], contains terms[4] built upon an abstract model [4] designed to be compatible with RDF. The Biodiversity Information Standards (TDWG)[5] Darwin Core (DwC) standard [5] defines the most widely used vocabulary for describing biodiversity resources[6]. Over 467 million occurrence records of the 533 million aggregated by the Global Biodiversity Information Facility (GBIF)[7] have been described using Darwin Core terms. In contrast to the vocabularies mentioned above, the DwC vocabulary was primarily designed to facilitate

* Corresponding author.
1 http://www.w3.org/TR/rdf-concepts/
2 http://xmlns.com/foaf/spec/20140114.rdf
3 http://www.w3.org/2009/08/skos-reference/skos.rdf
4 http://dublincore.org/2012/06/14/dcterms.rdf
5 http://www.tdwg.org/
6 http://www.tdwg.org/standards/450/; enter reference website at http://rs.tdwg.org/dwc/terms/
7 http://www.gbif.org/ accessed 17 May 2015

Box 1. Major components of the Darwin Core RDF Guide and their sections in the Guide:

- Rationale for the Guide and review of RDF (Sections 1.1 - 1.3)
- Summary of major issues (Sections 1.4 - 1.5)
- Handling identifiers (Section 2.2)
- Handling type (Section 2.3)
- Distinguishing between literal and non-literal values (Sections 2.4 - 2.5)
- Handling idiosyncratic categories of Darwin Core terms (Sections 2.6 - 2.9)
- Categorized term reference tables (3.1 - 3.8)

the sharing of data in simple single-table text files and was not specifically intended for use in RDF.

Because so many data are already described using Darwin Core terms, there has been significant interest in adapting the DwC terms to describe biodiversity resources in RDF. Since the DwC terms are designated as IRIs, and because the normative term definitions are expressed in RDF/XML, it would seem trivial to use Darwin Core property terms as RDF predicates. However, the results of experimentation reported on the TDWG mailing list[8] between 2009 and 2011 identified a number of issues that impeded the effective use of Darwin Core terms in RDF.

In 2011, an RDF/OWL Task Group was chartered by TDWG. In 2012 a team of writers began work on a Darwin Core RDF Guide to address the identified issues by providing a set of best practices and by creating some new Darwin Core terms intended specifically for use in RDF. The Guide [6] was completed in 2013 and reviewed by the Task Group, which recommended it for adoption. When adopted by TDWG in 2015, the RDF Guide became a non-normative part of the Darwin Core standard and joined existing guides that describe how to use Darwin Core terms in simple text files and XML.

Adapting existing metadata vocabularies and datasets for use in the Semantic Web is a current challenge. [7] [8] [9] This paper describes how the Task Group adapted a vocabulary that was not designed specifically for use in RDF so that its terms could be used as RDF predicates in a consistent manner. In Section 2 of this paper, we describe each of the major issues (Box 1) and how they were resolved. Section 3 describes future challenges and prospects for integrating Darwin Core-described data into the broader Semantic Web.

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

2. Issues and their resolution in the Guide

2.1. Explaining the rationale to new users of RDF

The TDWG constituency consists primarily of biologists and data managers. Relatively few members of the organization are familiar with RDF, Linked Data, and the Semantic Web. Therefore, an important component of the RDF Guide is an explanation of important ways in which RDF differs from more traditional data transfer systems with which data managers may be familiar. The introduction of the Guide (Guide section 1) highlights several important features of RDF that data managers need to consider when adapting their data for output as RDF. It discusses the importance of IRIs as resource identifiers (Guide section 1.3.2) and references the best-practices specified in the TDWG GUID Applicability Statement standard[9] [10]. It explains the necessity of considering unintended entailments caused by inappropriate term use (Guide section 1.4.2). It also outlines the variety of circumstances under which string values are used in non-RDF implementations of DwC (Guide section 1.5) and how those circumstances must be handled differently when converted to RDF.

Each of these issues mentioned in the introduction of the Guide were identified as important points of confusion in threads on the TDWG email list prior to the formation of the Task Group. Although brief, the summary of these issues in the introduction provides links to more extensive reference information for implementers who are not familiar with those issues.

[8] http://lists.tdwg.org/pipermail/tdwg-content/

[9] http://www.tdwg.org/standards/150/

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

A. Record in an Occurrence 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

B. Record in a Location database table

Fig. 1. Example Darwin Core records. Actual records would have additional fields. `dwc:` = `http://rs.tdwg.org/dwc/terms/`

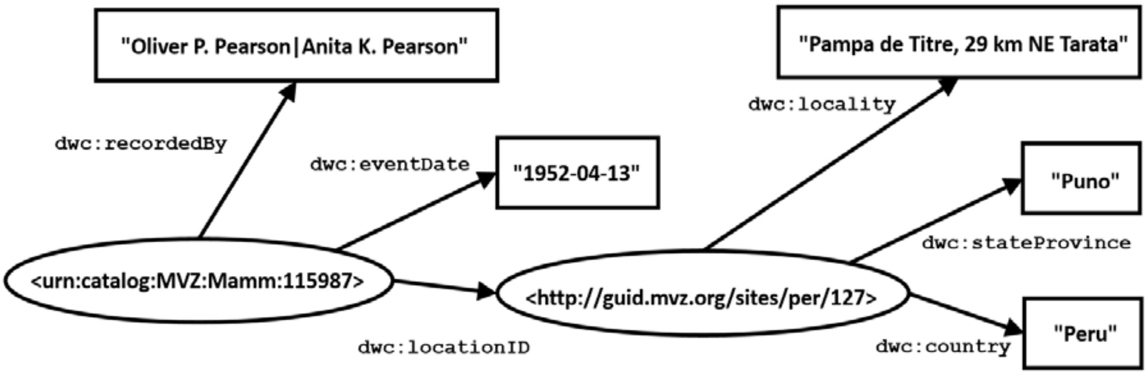

Fig. 2. Attempt to represent data from Fig. 1 as RDF. Compare with Fig. 4, which follows the guidelines of the DwC RDF Guide.

2.2. Resolution of issues related to string values

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:`[10] 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. This practice is illustrated in Fig. 1 where string values are used to refer to non-literal resources such as people and places.

Fig. 2 shows an attempt to represent these data as RDF using only `dwc:` terms as predicates. Although it is possible to identify people and places using strings as shown in Fig. 2, it is preferable to identify non-literal objects using IRI references, which can then be associated with additional properties that describe the non-literal resource.

2.2.1. Terms intended for use with non-literal objects

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 long-standing problem[11] 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]

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

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

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

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

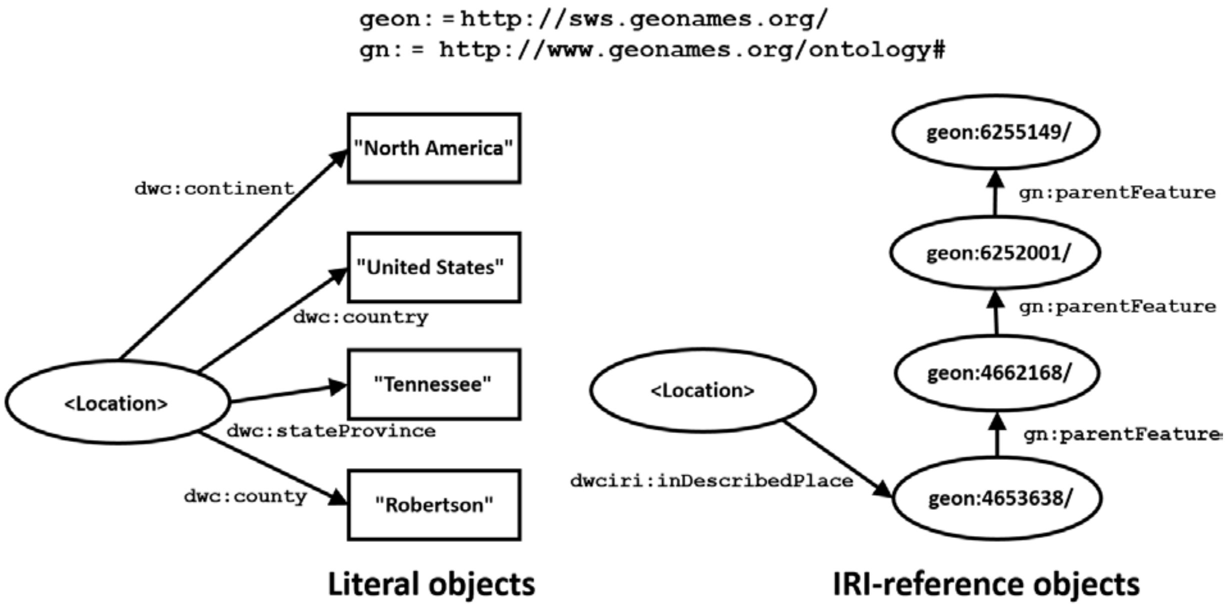

Fig. 3. Use of Darwin Core geographic "convenience" properties and their corresponding `dwciri:` analog

namespace to be used with literal values [12], while reserving terms in the `dcterms:`[14] namespace that have declared non-literal ranges for use with IRI reference or blank node objects.

The Darwin Core RDF Guide (Guide section 2.5) adopts the dual-term approach by creating a new Darwin Core namespace, `dwciri:`,[15] 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 `dwciri:recordedBy` would relate the subject to a non-literal object (IRI reference or blank node) that denotes the actual agent.

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 `dwciri:` namespace term, each non-literal resource in a concatenated list of names should be the object of separate triple. This is illustrated in Example 1 in the Appendix.

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 using `dwc:` namespace terms without imposing a requirement that the provider immediately implement the use of IRI identifiers for non-literal resources. Although there is some ambiguity about which real-world entities are represented by the strings, this is the same degree of ambiguity that is already present in the non-RDF string-based records. The availability of `dwciri:` namespace terms makes it possible to eliminate the ambiguity by linking to IRI identified objects.

2.2.2. "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. For example, the terms `dwc:county`, `dwc:stateProvince`, `dwc:country`, and `dwc:continent` allow a location to be placed in its geographical context (Fig. 3 and Appendix Example 2 in this paper). Other sets of terms describe ownership of a collection item, a taxonomic entity, chronostratographic descriptors, and lithostratigraphic descriptors.

In the location term set, 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

[14] `dcterms: = http://purl.org/dc/terms/`
[15] `dwciri: = http://rs.tdwg.org/dwc/iri/`

same name that are contained within different upper level political subdivisions. Thus, each location record must provide values for the entire set of terms. 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 `dwciri:` 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 an IRI 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 describe particular hierarchical sets of resources and the relationships among levels in the hierarchy in a standardized external database. In that case, a provider need only link to an IRI for the lowest level in the hierarchy, and clients consuming the RDF could discover the higher levels by retrieving information from the external database. In Fig. 3 and Appendix, Example 3, the property `dwciri:inDescribedPlace` links a location to the IRI for Robertson County, Tennessee, USA: `http://sws.geonames.org/4653638/`, a lowest-level administrative geographic subdivision in a standard reference. By dereferencing the GeoNames[16] IRI, a client could discover all of the higher levels in the hierarchy of geographic subdivisions. If this method were used, a SPARQL[17] query using the Robertson County, Tennessee, USA IRI would not depend on consistent spelling of "Robertson", "Robertson Co.", "Robertson County", "United States", "United States of America", "U.S.", "USA", "États-Unis", etc.

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 `dwciri:` 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.

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

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

2.3. Identity and type

The general 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 an identifier for the resource that was the subject of that record as well as identifiers for other resources linked 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. However, difficulties arise if ID terms are used when flat Darwin Core records are converted to RDF triples (Guide section 2.6).

The first 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 resource (effectively intending the ID term to serve as an object property), 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. For example, in Fig. 2 `dwc:locationID` was used as a predicate in an attempt to link the Occurrence and Location resources. However, because of the subproperty declaration, this use entails that `http://guid.mvz.org/sites/per/127` is the identifier of the Occurrence, not the Location, which is incorrect.

A second problem is that a pre-existing understanding between the data provider and consumer is required to know which of several ID term fields that might be present in the record represents the identifier for the record (i.e., provides the identifier for the subject resource) and which ID term fields represent identifiers of linked resources (i.e., of object resources). For example, In Fig. 1.A. it is not possible to know whether `dwc:occurrenceID` or `dwc:locationID` is the identifier for the record without knowing in advance that this is a record of an Occurrrence.

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. Instead, RDF best practices should be followed for specification of identifiers (discussed in section 2.3.1. of this paper) and for the assignment of type (discussed in section 2.3.2. of this paper). The linking function of ID terms must be served by object properties not defined by Darwin Core as discussed in section 2.4.2 of this paper.

2.3.1. Associating an identifier with a subject resource

In non-RDF uses, 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 IRIs. Thus, section 2.2 the RDF Guide provides some guidelines for translating the various kinds of ID term values into RDF.

If the subject resource identifier is an IRI, that IRI is simply asserted as the subject of triples describing the subject resource. If the subject resource identifier is a non-IRI 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) [13]. As IRIs, LSIDs may be the subjects of RDF triples. However, Recommendation 30 of the TDWG LSID Applicability Statement standard 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" [10]. The Darwin Core RDF Guide extends this recommendation to any non-HTTP IRI (i.e., including other varieties of URNs such as ARK, UUID, ISBN, etc.) by specifying that if possible, the subject resource should be identified by an HTTP-proxy version of the non-HTTP IRI, and that the non-HTTP IRI be the object of an `owl:sameAs` [18] property of a triple having the HTTP IRI as the subject (illustrated in Appendix Example 4).

[18] `owl: = http://www.w3.org/2002/07/owl#`

2.3.2. Specifying the type of a resource

Specifying the types of resources in a database record using the general Darwin Core vocabulary is complex and involves using the terms `dwc:basisOfRecord`, `dcterms:type`, and the various ID terms. Section 2.3.1 of the RDF Guide simplifies this situation by clarifying that `rdf:type` should be the primary property for indicating the type of all resources in RDF.

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.

2.4. Linking to related resources

Because RDF is a graph-based model, one of its primary concerns is linking non-literal, IRI-identified or anonymous nodes using object properties. Although there are a variety of ways that Darwin Core properties in the `dwc:` namespace use string values to establish links between the subject resource of a database record and other related resources, there is no consistent way to translate all of those relationships to RDF. The RDF Guide provides strategies for translating distinct categories of string-based linking properties into object properties suitable for use in RDF.

The first category corresponds to relationships defined by literal value terms in the general Darwin Core namespace `dwc:`. The solution, creating terms in the `dwciri:` namespace that can act as corresponding object properties, was already described in Sections 2.2.1. and 2.2.2. of this paper.

2.4.1. Association terms

In the second category, generic relationships were indicated by existing literal value Darwin Core terms. The general Darwin Core vocabulary included a number of "association terms" (terms whose local names begin with "associated", e.g. `dwc:associatedTaxa`, `dwc:associatedMedia`, `dwc:associatedOrganisms`, etc.) that indicated that a subject resource was linked to an instance of a related resource of some type. Often, the nature of that relationship was unspecified. To create these links in RDF, an existing well-known (non-Darwin Core) object property may be substituted for the association term to link to a non-literal object. For example, foaf:depiction can be used in place of dwc:associatedMedia if the object is an image. However, in many cases, there may not be a specific object property to substitute for the association term. In that case, section 2.8 of the Guide establishes that the well-known term dcterms:relation should be used along with an appropriate type declaration for the object.

2.4.2. Linking instances of the Darwin Core classes

The other major category of Darwin Core terms whose purpose is to establish links to other non-literal resources is the category of ID terms (discussed in section 2.3 of this paper). Since the ID terms cannot be used as RDF predicates, it seems as though it would have been a relatively simple task for the Guide to mint a set of object properties that could have been be used instead. However, minting such terms was hampered by the lack of a standard biodiversity domain model.

Example 4 in the Appendix is based on the data in Fig. 1 and illustrates the difficulty. Although the two tables in Fig. 1 imply the existence of instances of two classes (Occurrence and Location), the tables could actually be considered to contain information about instances of five classes: dwc:Occurrence, dwc:Event, dcterms:Location, dcterms:Agent (or foaf:Agent [19]), and gn:Feature. [20] When the data in the tables are expressed as RDF according to the DwC RDF Guide, dwciri:recordedBy is used to link the Occurrence to the Agents recording it, and dwciri: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.

Parts A through C of Example 4 show how the data in the tables can be serialized as RDF under several non-Darwin Core models.[21] The TDWG Ontology[22] 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 part A. The TaxonConcept ontology[23] includes the notion of both Occurrence and Location (in the form of the txn:Area[24] class) but does not recognize a separate class for Event. Therefore the serialization in part B uses a single object property txn:occurrenceHasArea to link the occurrence and location instance. The Darwin-SW ontology[25] adopts all of the main Darwin Core classes and therefore includes Occurrence, Event, and Location classes. The serialization in Example 4, part C (Fig. 4) is more normalized than the original database, requiring either a placeholder IRI (created using a #event fragment identifier in the example) or a blank node to represent the Event instance. Two object properties, dsw:atEvent and dsw:locatedAt,[26] are used to link the three Darwin Core classes.

Appendix Example 5 shows a SPARQL query designed to find occurrences recorded in Departamento de Puno, Peru by querying for its GeoNames IRI. 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 Example 4 part 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 (part A) or TaxonConcept ontology (part B).

It would be possible to merge graphs from providers that used different models and object properties, 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 was an effort beyond the scope of the Darwin Core RDF Guide. Work in this area is being actively pursued in the context of the Biological Collections Ontology (BCO) [14].

[19] foaf: = http://xmlns.com/foaf/0.1/
[20] gn: = http://www.geonames.org/ontology#
[21] The models are compared at http://code.google.com/p/tdwg-rdf/wiki/BiodiversityOntologies
[22] http://wiki.tdwg.org/twiki/bin/view/TAG/TDWGOntology
[23] http://www.taxonconcept.org/
[24] txn: = http://lod.taxonconcept.org/ontology/txn.owl#
[25] https://github.com/darwin-sw
[26] dsw: = http://purl.org/dsw/

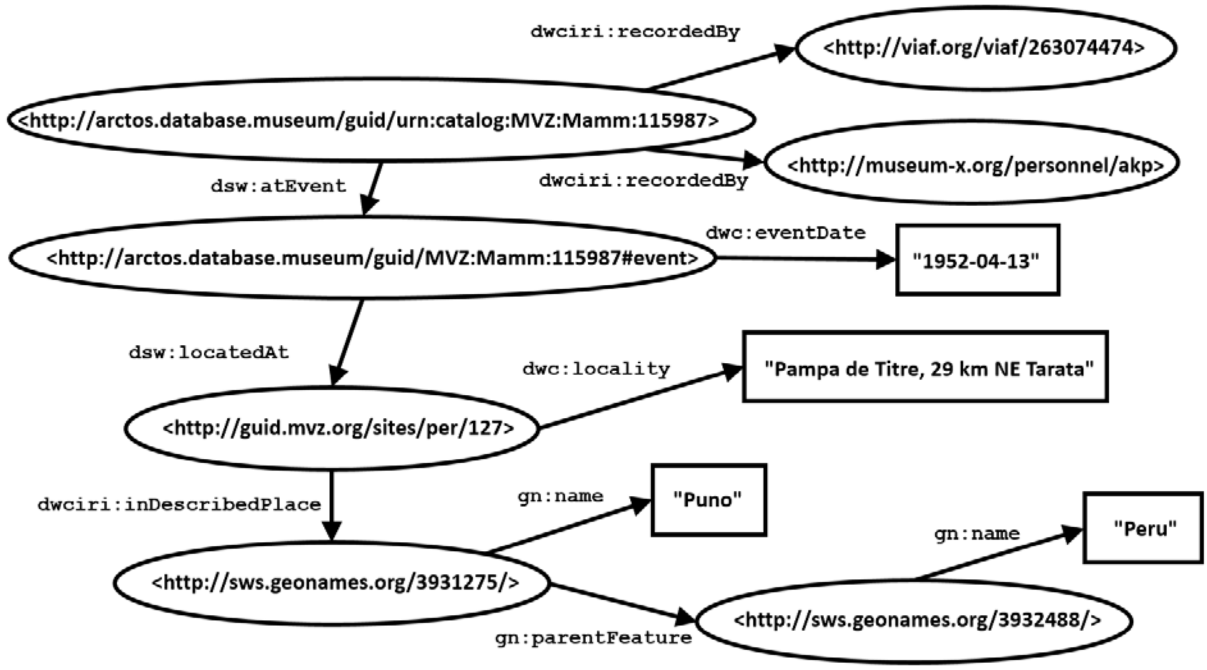

Fig. 4. Data from Fig. 1 serialized as RDF using object properties to link IRI-identified resources. See Example 4.C. for Turtle serialization.

3. Integrating Darwin Core-described data into the Semantic Web

Because of its wide acceptance, the Darwin Core vocabulary is an obvious source of predicates for description of biodiversity resources in RDF. However, because the original Darwin Core vocabulary was a general-purpose vocabulary intended to enable data transmission based primarily on simple tables of literal values, some of its properties could not be used as predicates in unmodified form, while the values of other properties did not unambiguously specify the real-world entities to which they refer. The Darwin Core RDF Guide defines new `dwciri:` namespace terms that can be used in most circumstances where the original terms were inadequate.

Comparison of Figs. 2 and 4 illustrates the advantages provided by using new `dwciri:` terms. In Fig. 4, using `dwciri:recordedBy` to link to agent IRIs removes any ambiguity about the identity of the agents and facilitates discovery of other resources that are also linked to those agents. Using `dwciri:inDescribedPlace` makes it possible to link to the extensive resources of the GeoNames database with little effort on the part of the provider of the Darwin Core records. In contrast, the approach taken in Fig. 2 places a significant burden on the data consumer, who must parse and process the literals and perform string-matching to attempt to link the Darwin Core records to related resources.

The lack of object properties to link instances of the main Darwin Core classes remains a major obstacle to effective expression of biodiversity data as RDF. However, the ability to consistently express many other relationships as RDF using Darwin Core properties will facilitate the development and testing of a consensus domain model for the biodiversity informatics community. This will in turn enable the community to add those missing object properties to the standard.[27] In tandem with efforts to define classes in a more semantically robust manner in the BCO, these developments will open the door for the integration of traditionally "flat" biodiversity data with the vast and growing body of semantically-enabled information.

[27] Darwin Core is a continuously evolving standard. As such this paper reflects the state of Darwin Core at the time of the ratification of the RDF Guide. To examine the current status of the standard, visit the Darwin Core documentation at http://rs.tdwg.org/dwc/.

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. Anonymous reviewers of a previous draft of this paper provided very helpful suggestions for its improvement.

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] D. Brickley, L. Miller. *FOAF Vocabulary Specification 0.99 Namespace Document.* Jan. 2014. http://xmlns.com/foaf/spec/

[2] A. Miles, S. Bechhofer. *SKOS Simple Knowledge Organization system Namespace Document - HTML Variant*: *Recommendation Edition*. World Wide Web Consortium, Aug. 2009. http://www.w3.org/2009/08/skos-reference/skos.html

[3] DCMI Usage Board. *DCMI Metadata Terms*. DCMI, Jun. 2012. http://dublincore.org/documents/dcmi-terms/

[4] A. Powell, M. Nilsson, A. Naeve, P. Johnston, T. Baker. *DCMI Abstract Model*. DCMI, Jun. 2007. http://dublincore.org/documents/2007/06/04/abstract-model/

[5] 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

[6] S. Baskauf, J. Wieczorek, J. Deck, C. Webb, P. J. Morris, M. Schildhauer. *Darwin Core RDF Guide*. Biodiversity Information Standards, Mar. 2015. http://rs.tdwg.org/dwc/terms/guides/rdf/

[7] G. Martens, R. Verborgh, C. Poppe, R. Van de Walle. Lifting a metadata model to the semantic multimedia world. *Journal of Information Processing Systems* 7(1):199-208, 2011. http://dx.doi.org/10.3745/JIPS.2011.7.1.199

[8] M. Chen, B. Plale. From metadata to ontology representation: A case of converting severe weather forecast metadata to an ontology, In *Proc. Association for Information Science and Technology*, Oct. 2012, ASIST. https://www.asis.org/asist2012/proceedings/Submissions/286.pdf

[9] M. Arenas, A. Bertails, E. Prud'hommeaux, J. Sequeda. *A Direct Mapping of Relational Data to RDF*: *W3C Recommendation*. World Wide Web Consortium, Sep. 2012. http://www.w3.org/TR/2012/REC-rdb-direct-mapping-20120927/

[10] K. Richards. *TDWG GUID Applicability Statement Standard*. Biodiversity Information Standards, Sep. 2010. http://www.tdwg.org/standards/150/

[11] M. Nilsson, A. Powell, P. Johnston, A. Naeve. *Expressing Dublin Core Metadata Using the Resource Description Framework (RDF).* DCMI, Jan. 2008. http://dublincore.org/documents/2008/01/14/dc-rdf/#sect-4

[12] S. Rühle, T. Baker, P. Johnston. *User Guide/Publishing Metadata*. DCMI, Jan. 2015. http://wiki.dublincore.org/index.php/User_Guide/Publishing_Metadata#Legacy_namespace

[13] Object Management Group. *Life Sciences Identifiers Specification*. v1.0, formal/04-12-01. EMBL-EBI, Interoperable Informatics Infrastructure Consortium, International Business Machines Corporation, Object Management Group, Dec. 2004. http://www.omg.org/spec/LIS/1.0/

[14] R. Walls, J. Deck, R. Guralnick, S. Baskauf, R. Beaman, S. Blum, S. Bowers, P. Buttigieg, N. Davies, D. Endresen, M. Gandolfo, R. Hanner, A. Janning, L. Krishtalka, A. Matsunaga, P. Midford, M. Morrision, É. Ó Tuama, M. Schildhauer, B. Smith, B. Stucky, A. Thomer, J. Wieczorek, J. Whitacre, J. Wooley. Semantics in support of biodiversity knowledge discovery: An introduction to the Biological Collections Ontology and related ontologies. *PLOS One* 9(3):e89606, 2014. http://dx.doi.org/10.1371/jorunal.pone.0089606

Appendix

Example. 1. Recorders of an Occurrence (serialized as RDF/Turtle)

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

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

```
<http://bioimages.vanderbilt.edu/ind-baskauf/00000#2001-03-14loc>
       a dcterms:Location;
       dwc:continent "North America";
       dwc:country "United States";
       dwc:stateProvince "Tennessee";
       dwc:county "Robertson".
```

Example 3. Using a `dwciri:` term to link a Location to its lowest level political subdivision (serialized as RDF/Turtle)

```
<http://bioimages.vanderbilt.edu/ind-baskauf/00000#2001-03-14loc>
        a dcterms:Location;
        dwciri:inDescribedPlace <http://sws.geonames.org/4653638/>.
```

Example 4. Database records and their RDF/Turtle serialization using three ontologies outside of Darwin Core. Namespace abbreviations used are:

`dwc:` = `http://rs.tdwg.org/dwc/terms/`, `dwciri:` = `http://rs.tdwg.org/dwc/iri/`, `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#`

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";
     dwciri: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";
     dwciri:inDescribedPlace <http://sws.geonames.org/3931275/>.
```

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";
     dwciri: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";
     dwciri:inDescribedPlace <http://sws.geonames.org/3931275/>.
```

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";
     dwciri:recordedBy <http://viaf.org/viaf/263074474>,
                       <http://museum-x.org/personnel/akp>;
     dsw:atEvent <http:// arctos.database.museum/guid/MVZ:Mamm:115987#event>.
<http://arctos.database.museum/guid/MVZ:Mamm:115987#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";
     dwciri:inDescribedPlace <http://sws.geonames.org/3931275/>.
```

Example 5. SPARQL query based on Darwin-SW object properties.

```
PREFIX dwciri: <http://rs.tdwg.org/dwc/iri/>
PREFIX dsw: <http://purl.org/dsw/>

SELECT ?occurrence WHERE
  {
  ?location dwciri:inDescribedPlace <http://sws.geonames.org/3931275/>.
  ?event dsw:locatedAt ?location.
  ?occurrence dsw:atEvent ?event.
```