An RDF Guide for the TDWG
Darwin Core Standard

(http://code.google.com/p/tdwg-rdf/wiki/DwcRdf)

Steve Baskauf, John Wieczorek, John Deck, Campbell Webb, and Mark Schildhauer
Why express Darwin Core as RDF?

1. Darwin Core is designed primarily for flat records; RDF allows more complex interconnections between resources.

2. Many Darwin Core terms specify literal name string values; but RDF uses URIs to disambiguate which resources are intended.

3. Lack of clarity in Darwin Core about how typing should be done and what classes should be used.
1. URIs are less ambiguous than literals

\[ \text{dwc}: = \text{http://rs.tdwg.org/dwc/terms/} \]
\[ \text{dwcuri}: = \text{http://rs.tdwg.org/dwc/uri/} \]

**literal object:**

<http://museum.org/id/123>

\[ \text{dwc:identifiedBy} \ "\text{John Smith}". \]

(allows existing text-based data to be exposed, but which John Smith is intended?)

**URI-reference object:**

<http://museum.org/id/123>

\[ \text{dwcuri:identifiedBy} \ <\text{http://museum.org/people/smithj}>. \]

(persistent, unique URIs are unambiguous and permit discovery of further information)
2. Imported Dublin Core terms used inconsistently


**literal object allowed:**

```xml
<dx.doi.org/10.1098/rsbl.2011.0228>
  dc:language "en".
</dx.doi.org/10.1098/rsbl.2011.0228>
```

**correct usage with URI-reference object:**

```xml
<dx.doi.org/10.1098/rsbl.2011.0228>
</dx.doi.org/10.1098/rsbl.2011.0228>
```

This makes DwC recommendations consistent with Dublin Core RDF guidelines.
3. Dealing with string lists

**Problem:** Concatenated and delimited lists of literal values must be parsed and matched with standard values.

**Solution:** use multiple triples containing `dwc:recordedBy` properties.

**literal object:**

```
  *dwc:recordedBy* "Oliver P. Pearson|Anita K. Pearson".
```

**URI-reference objects:**

```
  *dwc:recordedBy* <http://viaf.org/viaf/263074474>,
  <http://museum-x.org/personnel/akp>.
```

Note: the URI-reference example denotes two separate triples – one for each documenter.
4. Use of DwC "ID" terms and expression of identifiers

Example: `dwc:identificationID`

Darwin Core "ID" terms have three issues relevant to RDF:

- They combine a typing function with specification of an identifier.
- DwC documentation suggests they can refer to either the subject or the object.
- They are declared `rdfs:subPropertyOf dcterms:identifier` which implies they have literal values and that they express the identifier of the subject.

Solution: **DON’T USE THEM.** Use standard RDF practices:

```xml
<http://museum.org/occurrences/23459>
  rdf:type dwctype:Occurrence;
```
5. Classes to use for \texttt{rdf:type}

\textbf{Which to use???
}\texttt{dwc:Occurrence} vs. \texttt{dwctype:Occurrence}
\texttt{dcmitype:Event} vs. \texttt{dwctype:Event}
\texttt{dwctype:Location} vs. \texttt{dcterms:Location}
etc.

\textbf{Solution, part 1 - Change the DwC Type Vocabulary}
1. Use the Dublin Core term for a class if there is one.
2. Make sure all DwC classes are represented in the Type Vocabulary.

\textbf{Solution, part 2 - Users follow this convention:}
Use the class in the DwC Type Vocabulary. Don't use \texttt{dwc}: class terms with \texttt{rdf:type}.
6. "Convenience terms"

geon: = http://sws.geonames.org/

gn: = http://www.geonames.org/ontology#

<Location>

"North America"

dwc:continent

"United States"

dwc:country

"Tennessee"

dwc:stateProvince

"Cheatham"

dwc:county

<Location>

dwcuri:inDescribedPlace

geon:6255149

gn:parentFeature

geon:6252001

gn:parentFeature

geon:4662168

gn:parentFeature

geon:4612891

7. The problem of *dwc:Taxon*

- Unclear, a mixture of terms for describing taxon names and taxon concepts.
- RDF guide can't fix this, but provides a URI-reference linking term to a solution outside the guide.

**Solution:**
DON'T USE *dwc:Taxon* or *dwctype:Taxon* in RDF.

Link to a taxon concept URI
State of Darwin Core as RDF

- Some DwC terms are too complex to translate directly into RDF (e.g. `dwc:ResourceRelationship` class terms).
- Darwin Core classes still don't have ontological definitions.
- There is no prescribed model for the biodiversity domain.
- There are no object properties to connect the main DwC classes.

Fixing these problems is beyond the scope of the Guide.
Acknowledgements

Thanks to members of TDWG RDF/OWL Task Group who provided comments and suggestions on the guide.

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

URI of draft Darwin Core RDF Guide:
http://code.google.com/p/tdwg-rdf/wiki/DwcRdf