A Brief Primer on Data Standards

iDigBio Vertebrate Digitization Workshop Two
UC Berkeley, CA  04Apr2016
David Bloom and Laura Russell
Communication is Key
# Biodiversity Data Standards

**TDWG Standards**

TDWG's standards are listed below. For information about current work within TDWG, we would suggest that you look at the [TDWG Wiki](#). For further information, please contact us.

<table>
<thead>
<tr>
<th>Title</th>
<th>Activity</th>
<th>Category</th>
<th>Status</th>
<th>Date Published</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUID and Life Sciences Identifiers Applicability Statements</td>
<td></td>
<td>Applicability Statements</td>
<td>Current</td>
<td>30-Jan-2011</td>
</tr>
<tr>
<td>Access to Biological Collection Data - version 2.06</td>
<td>Access to Biological Collections Data</td>
<td>Technical Specification</td>
<td>Current</td>
<td>16-Sep-2005</td>
</tr>
<tr>
<td>Structured Descriptive Data</td>
<td>Biological Descriptions Interest Group</td>
<td>Technical Specification</td>
<td>Current</td>
<td>16-Sep-2005</td>
</tr>
<tr>
<td>TDWG Standards Documentation Specification</td>
<td>TDWG Infrastructure Project</td>
<td>Technical Specification</td>
<td>Draft</td>
<td></td>
</tr>
<tr>
<td>Natural Collections Descriptions (NCD): A data standard for exchanging data describing natural history collections</td>
<td>Natural Collections Descriptions Interest Group</td>
<td>Technical Specification</td>
<td>Draft</td>
<td></td>
</tr>
<tr>
<td>ABCDDNA - DNA extension for Access to Biological Collection Data</td>
<td>Access to Biological Collections Data</td>
<td>Technical Specification</td>
<td>Draft</td>
<td></td>
</tr>
<tr>
<td>Plant Occurrence and Status Scheme</td>
<td></td>
<td>Status and Categories</td>
<td>Prior</td>
<td>01-Oct-1995</td>
</tr>
<tr>
<td>Plant Names in Botanical Databases</td>
<td></td>
<td>Best Current Practice</td>
<td>Prior</td>
<td>01-Oct-1995</td>
</tr>
<tr>
<td>Authors of Plant Names</td>
<td></td>
<td>Status and Categories</td>
<td>Prior</td>
<td>01-Oct-1992</td>
</tr>
<tr>
<td>World Geographical Scheme for Recording Plant Distributions</td>
<td></td>
<td>Status and Categories</td>
<td>Prior</td>
<td>01-Oct-1992</td>
</tr>
</tbody>
</table>
The Natural History Community Developed Darwin Core.
Biodiversity Data Standards: DwC

collectionCode: The name, acronym, coden, or initialism identifying the collection or data set from which the record was derived. Examples: "Mammals", "Hildebrandt", "eBird".
Biodiversity Data Standards: DwC-A

A DwC Archive is a compressed file containing a minimum of three files.

1. A data file (occurrence.txt) conforming to the SIMPLEDWC in a CSV format where the first row includes Darwin Core standard term names.
2. A meta file (meta.xml) in XML format containing technical details to instruct a computer in how to use the data file.
3. A meta file (eml.xml) in an XML format containing explanatory details about the records contained within the data file to instruct a user if the data will be fit for their use.

DwC Archives are becoming the preferred method for sharing biodiversity occurrence data.

http://rs.tdwg.org/dwc/terms-guides/text/index.htm
New types of data

Extensions

Audubon Media Description

http://rs.tdwg.org/adc/terms/Multimedia

The Audubon Core is a set of vocabularies designed to represent metadata for biodiversity multimedia resources and collections. These vocabularies aim to represent information that will help to determine whether a particular resource or collection will be fit for some particular biodiversity science application before acquiring the media. Among others, the vocabularies address such concerns as the management of the media and collections, descriptions of their content, their taxonomic, geographic, and temporal coverage, and the appropriate ways to retrieve, attribute and reproduce them.

Link: http://terms.tdwg.org/wiki/Audubon_Core_Term_List

back  download

Properties

Management Vocabulary

<table>
<thead>
<tr>
<th>Vocabularies</th>
<th>Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribution Vocabulary</td>
<td><a href="http://terms.tdwg.org/wiki/Audubon_Core_Attribution_Vocabulary">http://terms.tdwg.org/wiki/Audubon_Core_Attribution_Vocabulary</a></td>
</tr>
<tr>
<td>Agents Vocabulary</td>
<td><a href="http://terms.tdwg.org/wiki/Audubon_Core_Agents_Vocabulary">http://terms.tdwg.org/wiki/Audubon_Core_Agents_Vocabulary</a></td>
</tr>
<tr>
<td>Content Coverage Vocabulary</td>
<td><a href="http://terms.tdwg.org/wiki/Audubon_Core_Content_Coverage_Vocabulary">http://terms.tdwg.org/wiki/Audubon_Core_Content_Coverage_Vocabulary</a></td>
</tr>
<tr>
<td>Taxonomic Coverage Vocabulary</td>
<td><a href="http://terms.tdwg.org/wiki/Audubon_Core_Taxonomic_Coverage_Vocabulary">http://terms.tdwg.org/wiki/Audubon_Core_Taxonomic_Coverage_Vocabulary</a></td>
</tr>
<tr>
<td>Resource Creation Vocabulary</td>
<td><a href="http://terms.tdwg.org/wiki/Audubon_Core_Resource_Creation_Vocabulary">http://terms.tdwg.org/wiki/Audubon_Core_Resource_Creation_Vocabulary</a></td>
</tr>
<tr>
<td>Related Resources</td>
<td><a href="http://terms.tdwg.org/wiki/Audubon_Core_Related_Resources">http://terms.tdwg.org/wiki/Audubon_Core_Related_Resources</a></td>
</tr>
</tbody>
</table>

identifier

An arbitrary code that is unique for the resource. See also http://terms.tdwg.org/wiki/Audubon_Core_identifier

Examples: "http://www.morphbank.net/840303"

Qualified Name: http://purl.org/dc/terms/identifier
Namespace: http://purl.org/dc/terms/
Group: Management Vocabulary
Data Type: string
Required: true

measurements

datatype may take as value any type term from the MovingImage, InteractiveResource, Text. Values may be found at http://purl.org/dc/dcmitype/SItiImage) from a content type and use dcterms.type when you can supply the URI of one of dctype and dcterms/type must be supplied, values of each should designate the same type, or be xi:

Examples: "image", "MovingImage", "Sound"

Qualified Name: http://purl.org/dc/elements/1.1/type
Namespace: http://purl.org/dc/elements/1.1/
Data Type: string
Required: false

measurementType

The nature of the measurement, fact, characteristic or assertion. Recommended best practice is to use a controlled vocabulary. See also http://terms.tdwg.org/wiki/Audubon_Core_measurementType

Examples: "tail length", "temperature", "top line length", "survey area", "top type"

Qualified Name: http://purl.org/dc/dcmitype/Measure
Namespace: http://purl.org/dc/dcmitype/
Data Type: string
Required: false

measurementValue

The value of the measurement, fact, characteristic or assertion. See also http://terms.tdwg.org/wiki/Audubon_Core_measurementValue

Examples: "45", "20", "1", "14.5", "UV-light"

Qualified Name: http://purl.org/dc/dcmitype/MeasureValue
Namespace: http://purl.org/dc/dcmitype/
Data Type: string
Required: false

DARWIN CORE MEASUREMENT OR FACTS

http://rs.tdwg.org/dwc/terms/index.html#measurementindex

Link: http://rs.tdwg.org/dwc/terms/index.html#measurementindex

back  download
Biodiversity Data Standards: Simple DwC

Flat table

Few restrictions

http://rs.tdwg.org/dwc/terms/simple/index.htm