Collections Description (CD) Standard

What it is, why it’s cool, and whatcha do with it

Sharon Grant, FMNH  |  Janeen Jones, FMNH  |  Rebecca Snyder, NMNH  |  Kate Webbink, FMNH
Collections Description

Collections Description (CD) is an upcoming standard from TDWG - the group who formalized and maintains DarwinCore (DwC).

While the DwC standard focuses on collection object based fields, CD focuses on fields to describe collections as a whole. High level information about a collection, versus specific catalog record level information.

Especially useful for creating a searchable record for currently undigitized collections.
Use Cases

Researchers: find non-digitized collections of material, population studies of groups of species.

Governmental Agencies: find data on populations over time for policy making.

Museum Admin: aid decisions, like collections building, funding for underrepresented groups, etc.

Boards: track progress on digitization, areas of interest not being addressed.

Collection Staff: answer questions about: missing data, missing specimens, digitization efforts, Board quarterly and annual reporting, disaster mitigation.

Public/Media: find an institution with likely answers to specific questions.

Others: Suggestions?
Collections Description Charter

We urgently need a core vocabulary that describes the collections themselves, not the published datasets. We need to be able to use the published dataset metrics together with collections-level metadata and metrics in order to make meaningful comparisons across collections (e.g. what's unique about a particular collection, or where are the gaps). These data are especially critical for discovering collections that are not yet publishing their specimen data anywhere. There are valuable collections that have no information stored in a database and many do not have a presence on the Internet. Some **current stakeholders** asking for such a standard include GBIF, DiSSCo, SYNTHESYS+, ICEDIG, MOBILISE, and iDigBio and collections around the world (MfN, MOBOT, RBGE, CMN, Naturalis, The Field Museum, The Smithsonian, The NHM, and others, to name a few). Members in the group bring in relevant perspectives from other standards organizations such as the [Research Data Alliance (RDA)](https://rd-alliance.org/) as well. Other potential stakeholders include funding agencies (like conservation companies, and foundations) who are looking for certain data that are needed for addressing their missions. This metadata has the potential to expose what's available and help with prioritizing funding.
High-Level Data Classes

- **Collections and Collections Descriptions:** The term ‘collection’ is used here in the common sense as the holdings of an institution, and a Collection Description is any part of that collection. Within this model, a collection may not be an entity in itself, but rather an aggregation of all of the Collection Descriptions grouped by a particular property e.g. the holding institution.

- **Dimensions:** Provide dynamic descriptions of different parts of the collection, using common terms which can be used to aggregate and group across collections.

- **Metrics:** A ‘metric’ is defined here as a numeric property of a Collection Description that can be aggregated across multiple Collection Descriptions.

- **Narratives:** Provide richer narrative descriptions of Collection Descriptions.

- **Institutions:** Provide information about collection-holding institutions, and custody of the collections.

- **People:** Provide information about custodianship, expertise, attribution etc for Collection Descriptions.
1 CD record for The Field Museum where:
- basisOfCollection = Institution
- dateFormed = 2000
- currentStatus = Active
- title = The Field Museum
- collectionID = 1234
- hasParent = No
- ParentID = Null

1 CD record for The Field Museum’s Zoology Collections
- basisOfCollection = Collection
- hasParent = Yes
- title = Zoology Collections
- collectionID = 456
- ParentID = 1234

1 CD record for The Field Museum’s Invertebrate Zoology Collections
- basisOfCollection = Collection
- hasParent = Yes
- title = Invertebrate Zoology Collections
- collectionID = 789
- ParentID = 456

1 CD record
- basisOfCollection = Accession
- hasParent = Yes
- parentID = 789
- dateFormed = 1895
- cataloguedCount = 4,450
- name = Philip Carpenter Collection
- recordedBy = Philip P. Carpenter
- region = Mazatlan
- phylum = Mollusca
- collectionID = 5432
Interesting in Learning More?

CD Working Group:
https://github.com/tdwg/cd

TDWG Conference, Oct 20-25:
https://biodiversitynext.org/