

Rationale for an Extended CD to Accommodate Grouped Collections

Description	2
Grouped Collection Fields by Class	3
Class: TBD	3
Class: Institution	4
hasParent	4
parentID	4
immediateParentID	5
Class: GeographicOrigin	5
region	5
Class: Collection	6
basisOfCollection	6
typeOfAcquisition	6
acquisitionID	6
namedCollections	7
namedSubCollections	7
subsumedCollections	8
dateFormed	8
dateRetired	8
reasonRetired	9
Sub-class: Metric	9
cataloguedQuantity formerly cataloguedCount	9
cataloguedQuantityType	10
digitizedCount	10
digitizedCountType	10
georeferencedCount	11
associatedMediaCount	11
Class: Classification	12
materials	12
Class: Provenance	12
culturalAffiliation	12

Description

The Collection Descriptions Interest Group of TDWG has undertaken the effort to develop a data standard to describe entire collections that contain objects of and information about natural history. A detailed description of the CD effort, as well as the previous Natural Collections Descriptions (NCD) effort, can be read on the TDWG web site (<https://www.tdwg.org/community/cd/>, <https://www.tdwg.org/standards/ncd/>).

Of the many collections structures to be addressed by CD, grouped and/or hierarchical collections (e.g., accessions) are perhaps the type with the greatest possibility to address the growing quantities of data gathered within collections that are not yet shared with scientists or the general public. Grouped records, and sub-collections, are collections that have been acquired by an institution, but that have not necessarily been fully catalogued and processed, rendering them unsearchable or undiscoverable by the people who may benefit from their content.

We propose here to develop and promote a standardized structure for grouped collections (e.g. accession records) so that their contents and metadata may be shared via local, regional and global biodiversity data portals. The extended CD record for grouped collections will offer a description of a given collection's physical or undigitized objects, notes, specimens, media, and other content. The goal is to develop a structure that will link this standard seamlessly with the institutional levels of the CD structure, as well as the soon-to-be-revived collections registry, GRBio (now under the purview of GBIF). In addition to direct linkages with institution-level CD and GRBio, grouped collection records will include,

- metadata of the institutional owner, well as the host institution if the collection resides in a physical location that is not the same as the owner institution.
- metadata that links a specific record to a parent sub-collection or collection, as well as parts of the parent collection that have been incorporated into accessible collections, including the purpose of the collection.
- the ability to look at counts (e.g., # of objects within collection and the # that have been catalogued, digitized, or georeferenced) across collections.
- overlap with other biodiversity data-sharing standards (e.g., Darwin Core) to permit publication in biodiversity data portals.
- accommodations for anthropological and ethnographic objects and data.

Grouped Collection Fields by Class

Class(Dimension) Definitions:

- Collection : The fields in this class describe the contents of the record. What things does the record represent. (what Matt calls CollectionType. We think Collection is a better phrase)
 - Metric is a sub-class of the Class Collection
- RecordLevel : The fields in this class are used to define resource itself and allow for sharing between records of the same type. They are the metadata for the record.
- NEW History :
- GeographicalOrigin
- Provenance

The following fields are recommended as a means to describe grouped collections. These fields are not represented in other ratified standard and standards extensions. Each field is presented with:

- Field name
- Level
- Definition
- Comments
- Examples/Vocabulary

Use:

<https://docs.google.com/document/d/1KPOJM0WnZGyXCP--k-IKHiPpXIBZNHV6lfob9DuuG3g/edit#>

Add "Components"/Classes to each table below to show symbiosis with the document above.

Class: TBD

Field	otherOccurrenceIDs or relatedOccurrenceIDs (formerly occurrenceIDs)
Level	Record-level
Class:(Dimension)	TBD

Definition	A list (concatenated and separated) of previous or alternate fully qualified occurrenceIDs or other human-used identifiers for objects within the same collection, whether in the current or any other data set or collection.
Comments	Recommended best practice is to separate the values in a list with space vertical bar space (" ").
Examples/Vocab	FMNH:Mammal:1234, NPS YELL06778 MBG 33424

Field	otherDatasetIDs
Level	Record-level
Class:(Dimension)	TBD / new dimension? [Collection:History]
Definition	A list (concatenated and separated) of other published data sets that contains objects or information from within the collection.
Comments	Recommended best practice is to separate the values in a list with space vertical bar space (" ").
Examples/Vocab	FMNH:Mammal:1234, NPS YELL06778 MBG 33424

Class:Institution

Field	hasParent
Level	Record-level
Class:(Dimension)	Institution
Definition	Confirmation that the collection, object(s) or information is, or was, a part of a collection when that collection name and location are unknown.
Comments	Recommended best practice that cdac:hasParent should have a corresponding cdac:parentRemarks
Examples/Vocab	Yes, No

Field	parentID
-------	-----------------

Level	Record-level
Class:(Dimension)	Institution
Definition	An identifier for the institution having custody of the object(s) or information referred to in the record.
Comments	Recommended best practice is to use an identifier from a known collections registry such as GRBio ([url when up]).
Examples/Vocab	http://biocol.org/urn:lsid:biocol.org:col:34795

Field	immediateParentID
Level	Record-level
Class:(Dimension)	Institution
Definition	An identifier for the collection or sub-collection of which the object(s) or information is a part.
Comments	Recommended best practice is to use an identifier from a known collections registry, such as GRBio ([url when up]).
Examples/Vocab	http://grbio.org/cool/05pf-h6mh

Class:GeographicOrigin

Field	region
Level	Location
Class:(Dimension)	GeographicOrigin
Definition	The name of a spatial region or named place of any size within an individual or multiple administrative areas.
Comments	Recommended best practice is to use this field in which the location occurs where administrative location fields (country, stateProvince, county, etc.) provide insufficient description.
Examples/Vocab	Multi-country (North European Plain), Multi-stateProvince (Pacific

	Northwest), waterbody (Southern).
--	-----------------------------------

Class:Collection

Field	basisOfCollection
Level	Record-level
Class:(Dimension)	Collection
Definition	The specific nature of the collection.
Comments	Recommended best practice is to use a standard vocabulary.
Examples/Vocab	Accession, Inventory, Expedition, InstitutionalGift, [Transferral of Rights]

Field	typeOfAcquisition
Level	Record-level
Class:(Dimension)	Collection
Definition	Additional information that exists that describes object(s) or information in the context of the basisOfCollection.
Comments	
Examples/Vocab	"Field maps", "field notebooks", other historical or factual descriptions of the collections composition or purpose.

Field	acquisitionID
Level	Record-level
Class:(Dimension)	Collection

Definition	An institutional of collection identifier for the acquisition of which the object(s) or information is a part.
Comments	An identifier for the acquisition. May be a global unique identifier or an identifier specific to a collection of institution.
Examples/Vocab	7b8aff00-a9f8-11d8-944b-b8a03c50a862, accession-FMNH-12345

Field	namedCollections
Level	Record-level
Class:(Dimension)	Collection:History
Definition	A list (concatenated and separated) of previous or alternate collections to which the grouped collection is related, whether at the owner institution or any other institution.
Comments	Recommended best practice is to separate the values in a list with space vertical bar space (" ").
Examples/Vocab	Grant Southwest Ceramics Collection, FMNH Mammals TNHC Vertebrates

Field	namedSubCollections
Level	Record-level
Class:(Dimension)	Collection:History
Definition	A list (concatenated and separated) of previous or alternate collections within the grouped collection, whether in the current or any other data set or collection.
Comments	Recommended best practice is to separate the values in a list with space vertical bar space (" ").
Examples/Vocab	Grant Southwest Ceramics Collection, FMNH Mammals TNHC Vertebrates

Field	subsumedCollections
Level	Record-level
Class:(Dimension)	Collection:History
Definition	A list (concatenated and separated) of previous or alternate fully qualified CD identifiers from collections that have been subsumed into the collection.
Comments	
Examples/Vocab	[examples of CD IDs if different from] FMNH:Mammal:1234, NPS YELL06778 MBG 33424

Field	dateFormed
Level	Record-level
Class:(Dimension)	Collection:History
Definition	The date-time or interval during which the collection was formed. Not suitable for a time in a geological context.
Comments	Recommended best practice is to use a date that conforms to ISO 8601:2004(E).
Examples/Vocab	1963-03-08T14:07-0600 (8 Mar 1963 at 2:07pm in the time zone six hours earlier than UTC). 2009-02-20T08:40Z (20 February 2009 8:40am UTC). 2018-08-29T15:19 (3:19pm local time on 29 August 2018). 1809-02-12 (some time during 12 February 1809). 1906-06 (some time in June 1906). 1971 (some time in the year 1971). 2007-03-01T13:00:00Z/2008-05-11T15:30:00Z(some time during the interval between 1 March 2007 1pm UTC and 11 May 2008 3:30pm UTC). 1900/1909 (some time during the interval between the beginning of the year 1900 and the end of the year 1909). 2007-11-13/15 (some time in the interval between 13 November 2007 and 15 November 2007).

Field	dateRetired
-------	--------------------

Level	Record-level
Class:(Dimension)	Collection:History
Definition	The date or date-time during which the institution or collection ceased operation or was subsumed.
Comments	Recommended best practice is to use a date that conforms to ISO 8601:2004(E).
Examples/Vocab	1963-03-08T14:07-0600 (8 Mar 1963 at 2:07pm in the time zone six hours earlier than UTC). 2009-02-20T08:40Z (20 February 2009 8:40am UTC). 2018-08-29T15:19 (3:19pm local time on 29 August 2018). 1809-02-12 (some time during 12 February 1809). 1906-06 (some time in June 1906). 1971 (some time in the year 1971). 2007-03-01T13:00:00Z/2008-05-11T15:30:00Z(some time during the interval between 1 March 2007 1pm UTC and 11 May 2008 3:30pm UTC). 1900/1909 (some time during the interval between the beginning of the year 1900 and the end of the year 1909). 2007-11-13/15 (some time in the interval between 13 November 2007 and 15 November 2007).

Field	reasonRetired
Level	Record-level
Class:(Dimension)	Collection:History
Definition	If dateRetired is not null, document the reason the record was retired.
Comments	This is free text related to the cd:DevelopmentStatusTypeTerm field.
Examples/Vocab	Destroyed, Transferred

Sub-class:Metric

Field	cataloguedQuantity formerly cataloguedCount
Level	Record-level
Class:(Dimension)	Collection:Metric

Definition	A number or enumeration value for the quantity of objects.
Comments	A cdac:cataloguedQuantity must have a corresponding cdac:cataloguedQuantityType.
Examples/Vocab	27 (cataloguedQuantity) with individuals (cataloguedQuantityType). 12.5 (cataloguedQuantity) with %biomass(cataloguedQuantityType). r (cataloguedQuantity) with BraunBlanquetScale (cataloguedQuantityType).

Field	cataloguedQuantityType
Level	Record-level
Class:(Dimension)	Collection:Metric Description
Definition	The type of the quantification system used for the quantity of objects
Comments	A cdac:cataloguedQuantityType must have a corresponding cdac:cataloguedQuantity.
Examples/Vocab	27 (cataloguedQuantity) with individuals (cataloguedQuantityType). 12.5 (cataloguedQuantity) with %biomass(cataloguedQuantityType). r (cataloguedQuantity) with BraunBlanquetScale (cataloguedQuantityType).

Field	digitizedCount
Level	Record-level
Class:(Dimension)	Collection:Metric
Definition	The number of objects within the collection or sub-collection that have been digitized.
Comments	
Examples/Vocab	0, 8, 27

Field	digitizedCountType
-------	---------------------------

Level	Record-level
Class:(Dimension)	Collection:Metric Description
Definition	The type of objects that have been digitized within the collection or sub-collection.
Comments	Recommended best practice is to use a standard vocabulary.
Examples/Vocab	Collection, StillImage, Sound, MovingImage, InteractiveResource, Text

Field	georeferencedCount
Level	Record-level
Class:(Dimension)	Collection:Metric
Definition	The number of objects within the collection or sub-collection that have been georeferenced.
Comments	Recommended best practice is to include additional known information about the georeferencing process including dwc:georeferencedBy, dwc:georeferencedDate, dwc:georeferenceProtocol, dwc:georeferenceSources, dwc:georeferenceVerificationStatus, and dwc:georeferenceRemarks.
Examples/Vocab	0, 8, 27

Field	associatedMediaCount
Level	Record-level
Class:(Dimension)	Collection:Metric
Definition	The number of media objects within the collection or sub-collection.
Comments	Recommended best practice for multiple media IDs is to use Aubon Core Media Extension.
Examples/Vocab	0, 8, 27

Class:Classification

Field	materials
Level	Record-level
Class:(Dimension)	Classification
Definition	The materials that compose the collection.
Comments	
Examples/Vocab	wood, copper

Class:Provenance

Field	culturalAffiliation
Level	Record-level
Class:(Dimension)	Provenance
Definition	Any cultural groups affiliated (e.g., as owners or creators) with the collection
Comments	
Examples/Vocab	Inuit

Recommended Existing Fields

The following fields are recommended as a part of the core content fields for grouped collections. These have been collated from the existing ratified standards and standards extensions including Dublin Core, Darwin Core, Audubon Core, and the Encoded Archival Description. We anticipate additional fields from NCD/CD at the collection and record level when determined.

Field	Field type (according to DwC)	Class (according to CD Dimension)	Proposed content or link to description
cd:collectionID	Record	Collection	LSID
cd:DevelopmentStatusTerm	Record	Collection	Add standard vocabulary terms: Retired, Transferred, Destroyed, Active
dc:description	Collection	Collection	http://dublincore.org/specifications/dublin-core/dcmi-terms/2012-06-14/?v=terms#description
dwc:recordedBy	Collection	Collection	https://dwc.tdwg.org/terms/#dwc:recordedBy
dc:title	Collection	Collection	http://dublincore.org/specifications/dublin-core/dcmi-terms/2012-06-14/?v=terms#title
ead:language	Collection	Collection	https://www.loc.gov/ead/tglib/elements/language.html
dwc:taxonID	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#taxonID
dwc:scientificNameID	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#scientificNameID
dwc:scientificName	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#scientificName
dwc:higherClassification	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#higherClassification
dwc:kingdom	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#kingdom
dwc:phylum	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#phylum
dwc:class	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#class
dwc:order	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#order

dwc:family	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#family
dwc:genus	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#genus
dwc:subgenus	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#subgenus
dwc:specificEpithet	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#specificEpithet
dwc:infraspecificEpithet	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#infraspecificEpithet
dwc:taxonRank	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#taxonRank
dwc:verbatimTaxonRank	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#verbatimTaxonRank
dwc:vernacularName	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#vernacularName
dwc:nomenclaturalCode	Taxon	Classification	http://rs.tdwg.org/dwc/terms/#nomenclaturalCode
dwc:eventDate	Event	Event	http://rs.tdwg.org/dwc/terms/#eventDate
dwc:verbatimEventDate	Event	Event	http://rs.tdwg.org/dwc/terms/#verbatimEventDate
dwc:habitat	Event	Event	http://rs.tdwg.org/dwc/terms/#habitat
dwc:locationID	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#locationID
dwc:higherGeographyID	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#higherGeographyID
dwc:higherGeography	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#higherGeography
dwc:continent	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#continent
dwc:waterBody	Collection	GeographicOrigin	http://rs.tdwg.org/dwc/terms

	Location		/waterBody
dwc:islandGroup	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#islandGroup
dwc:island	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#island
dwc:country	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#country
dwc:countryCode	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#countryCode
dwc:stateProvince	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#stateProvince
dwc:county	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#county
dwc:municipality	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#municipality
dwc:locality	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#locality
dwc:verbatimLocality	Collection Location	GeographicOrigin	http://rs.tdwg.org/dwc/terms/#verbatimLocality
dc:accessRights	Access Restrictions	Record-Level	https://dwc.tdwg.org/terms/#dcterms:accessRights https://terms.tdwg.org/wiki/Audubon_Core_Term_List#dc:rights ?
xmpRights:UsageTerms	Access Restrictions	Record-Level	https://terms.tdwg.org/wiki/Audubon_Core_Term_List#xmpRights:UsageTerms
dc:rights dc:rightsHolder dc:license	Access Restrictions	Record-Level	https://dwc.tdwg.org/terms/#dcterms:license
[Specific fields TBD from GGBN genomic extensions to Darwin Core]	Genomic Context	[SpecimentType / PreservationMethod]	https://terms.tdwg.org/wiki/GGBN_Amplification_Vocabulary , https://terms.tdwg.org/wiki/GGBN_DNA_Cloning_Vocabulary ,

			https://terms.tdwg.org/wiki/GGBN_Gel_Image_Vocabulary , https://terms.tdwg.org/wiki/GGBN_Loan_Vocabulary , https://terms.tdwg.org/wiki/GGBN_Material_Sample_Vocabulary , https://terms.tdwg.org/wiki/GGBN_Permit_Vocabulary , https://terms.tdwg.org/wiki/GGBN_Preparation_Vocabulary , https://terms.tdwg.org/wiki/GGBN_Preservation_Vocabulary
dwc:geologicalContextID	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#geologicalContextID
dwc:earliestEonOrLowestEonothem	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#earliestEonOrLowestEonothem
dwc:latestEonOrHighestEonothem	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#latestEonOrHighestEonothem
dwc:earliestEraOrLowestErathem	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#earliestEraOrLowestErathem
dwc:latestEraOrHighestErathem	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#latestEraOrHighestErathem
dwc:earliestPeriodOrLowestSystem	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#earliestPeriodOrLowestSystem
dwc:latestPeriodOrHighestSystem	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#latestPeriodOrHighestSystem
dwc:earliestEpochOrLowestSeries	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#earliestEpochOrLowestSeries

			ries
dwc:latestEpochOrHighestSeries	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#latestEpochOrHighestSeries
dwc:earliestAgeOrLowestStage	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#earliestAgeOrLowestStage
dwc:latestAgeOrHighestStage	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#latestAgeOrHighestStage
dwc:lowestBiostratigraphicZone	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#lowestBiostratigraphicZone
dwc:highestBiostratigraphicZone	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#highestBiostratigraphicZone
dwc:lithostratigraphicTerms	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#lithostratigraphicTerms
dwc:group	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#group
dwc:formation	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#formation
dwc:member	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#member
dwc:bed	Geological Context	GeologicalTimeRange	http://rs.tdwg.org/dwc/terms/#bed
dc:creator	Collection	Provenance	http://purl.org/dc/elements/1.1/creator
dwc:basisOfRecord	Record-Level	ObjectType	http://rs.tdwg.org/dwc/terms/basisOfRecord
dwc:preparations		ObjectType	http://rs.tdwg.org/dwc/terms/preparations

Recommended Pending Fields

The following fields are recommended as a part of the core content fields for grouped collections. These have been collated from standards and standards extensions which have been developed but are pending ratification including Chronometric Age.

dwc:verbatimChronometricAge	Chronometric Age	GeologicalTimeRange	https://github.com/VertNet/dwc-chronology/blob/master/standard/vocabularies/term_versions.csv
dwc:maximumChronometricAge	Chronometric Age	GeologicalTimeRange	https://github.com/VertNet/dwc-chronology/blob/master/standard/vocabularies/term_versions.csv
dwc:minimumChronometricAge	Chronometric Age	GeologicalTimeRange	https://github.com/VertNet/dwc-chronology/blob/master/standard/vocabularies/term_versions.csv
dwc:chronometricAgeUncertaintyInYears	Chronometric Age	GeologicalTimeRange	https://github.com/VertNet/dwc-chronology/blob/master/standard/vocabularies/term_versions.csv
dwc:chronometricAgeProtocol	Chronometric Age	GeologicalTimeRange	https://github.com/VertNet/dwc-chronology/blob/master/standard/vocabularies/term_versions.csv
dwc:chronometricAgeRemarks	Chronometric Age	GeologicalTimeRange	https://github.com/VertNet/dwc-chronology/blob/master/standard/vocabularies/term_versions.csv