

# Natural Collections Description (NCD)

## A data standard for exchanging data describing natural history collections

Neil Thomson (Natural History Museum, London), Roger Hyam (Royal Botanic Garden Edinburgh), Constance Rinaldo (Harvard University), Carol Butler (Smithsonian Institution), Doug Holland (Missouri Botanical Gardens), Barbara Mathé (American Museum of Natural History), Günter Waibel (RLG Programs, OCLC), Wouter Addink (ETI Bioinformatics), Ruud Altenburg (ETI), Markus Döring (Berlin Botanic Garden)

## Summary

Natural Collections Description (**NCD**) is a data standard for describing collections of natural history materials at the collection level; one NCD record describes one entire collection.

Collection descriptions are electronic records that document the holdings of an organisation as groups of items, which complement the more traditional item-level records such as are produced for a single specimen or a library book. NCD is tailored to natural history. It lies between general resource discovery standards such as Dublin Core (**DC**) and rich collection description standards such as the Encoded Archival Description (**EAD**). It is possible to extract a Dublin Core record from an NCD record for use with general resource discovery systems, or to use an NCD record as the seed for a richer collection description, like an EAD record.

The NCD standard covers all types of natural history collections, such as specimens, original artwork, archives, observations, library materials, datasets, photographs or mixed collections such as those that result from expeditions and voyages of discovery.

NCD primarily holds information about collections of objects, but can also be used to describe organisations (collections of collections) and networks (collections of organisations). There are many existing sources of information about biodiversity organisations, but they are scattered and in different formats.

## Description

### Collection descriptions

Collection descriptions are electronic records that document the holdings of organisations as groups of items. Such descriptions complement the more traditional item-level records describing a single specimen or a library book. Each collection record describes one entire collection, including narrative information on the collection itself, its extent and purpose, conditions of access and use along with who to contact for more information.

## Natural Collections Descriptions (NCD) v0.9 2008

A collection may be loosely defined as any group of things that have something in common. That "something in common" can be defined by the basic questions that users ask when accessing collections – who, what, where and when.

Examples of collections include:

- items that were collected or made by a particular person
- items that have the same format, such as art on paper
- items that came from the same place
- specimens that belong to the same taxonomic group
- materials collected on a voyage of discovery

In natural history museums, for example, researchers are most familiar with the collections of specimens, the library and the archives but the exhibitions, paintings, sculptures and learning materials are also collections.

Digital collections include images, video, datasets and databases (which are collections of item-level records) and the thematic sections of web sites. Noting the formats and media used to store digital data will be of value in digital sustainability planning, so that the process of migrating data from imminently obsolescent formats may be effectively managed. This will probably be carried out in conjunction with tools that are being developed by the digital sustainability community.

Collections of natural history material can be large. Consequently, detailed item-level descriptions can take a long time to complete. Collection-level records can ensure that knowledge about the richness of collections can be revealed more rapidly. Relating collections that are in museums, libraries, archives or other organisations (cross-domain resources) is a priority for many governments and by adopting the same description standard for collections in each domain, it becomes possible to search across all collections, regardless of management domain or location.

Some organisations divide collections between departments for curatorial purposes. Researchers would need to contact each department individually to assess the complete collection. Similarly, some collections have been dispersed throughout several organisations or even across several countries. These collections may be reunited in a virtual sense, using collection descriptions for each component.

A collection description record can be created for a collection whether the items in that collection have their own records in a database, or not. Where a database containing item-level details exists, a link can be provided to that database for those that need that level of detail. If the collection does not have an item-level database, producing a collection description reduces the chances of that collection being overlooked by researchers using the Web for resource discovery. Collections cannot be protected if they are not known to exist.

Collection descriptions provide a broad perspective and such records can serve a variety of additional purposes for organisations:

- A collections inventory is helpful in protecting against both loss of data and loss of collections and thus serves as a form of audit control and security against unwarranted disposal.
- They can help with the assessment of the strengths and gaps in the organisation as a whole, so that finding collaboration partners that have either the same or complementary strengths is simplified.

- They can help to identify which areas should be a priority for development in strategic plans and to establish priorities for item-level cataloguing. For conservation assessment, the McGinley scale is recommended, details of which can be found at:

McGinley, R. J. 1993. Where's the Management in Collections Management? Planning for Improved Care, Greater Use and Growth of Collections. *In*: Rose, C. L., et al. (eds.). International Symposium and First World Congress on the preservation and conservation of Natural History Collections 3. Comunidad de Madrid Consejería de Educacion y Cultura and Direccion General de Bellas Artes y Archives Ministerio de C, Madrid. Pages 309-338.

- Collection descriptions can serve to prevent loss of data that is in a physical form or electronic data in a format or medium that is nearing technological obsolescence. Creating collection descriptions for datasets that includes format information will help to act as an early warning so that data can be migrated to a more current format. Such data then becomes part of a digital sustainability programme, rather than a digital archaeology project.
- Collection description records act as a convenient place to store information volunteered by collections managers or visitors, which may otherwise be lost on their departure.

Records can be created *de novo* or from existing resources, such as published finding aids. There are many of these, but they are all in different formats, mainly on paper and cannot easily be searched. Once collection level data exists it can be used for internal projects such as exhibition labels or for external initiatives such as the merging of data from several sources to provide regional coverage of biodiversity collections.

### **NCD records**

An NCD record consists minimally of the 4 mandatory fields (Author, Record created date, Collection name and Description) so that it is easy to set up holding records that may be filled out when resources allow. It is suggested that each record will be serialized in the Resource Description Framework (**RDF**) and its Identifier will be a resolvable Life Sciences Identifier (**LSID**) or Uniform Resource Locator (**URL**) to that RDF file but the standard does not mandate the use of RDF (see *Implementation and Compliance* below). All other fields are considered to be optional, but of course the more information that can be provided about a collection the more useful the record will be.

The normative documentation gives the labels, Uniform Resource Identifiers (**URIs**) and definitions for each NCD class and property. Also provided are the tables of consistent terms for use in pick-lists and an example record.

The standard caters for collections of any type of material, physical or digital and either private or corporate ownership. It may be important to distinguish between physical collections and derived collections. An example of a derived collection record is one that has been produced as the result of a query on a collections management database, such as "all the items from Australia". This contains useful information that the institution may wish to keep, but could cause inaccurate totals if included in a count of collections held at the institution, since it does not exist as a discrete collection.

Records include information about who created the record and when, or the source of the records if they have been harvested from elsewhere. If a record is subsequently edited then

the editor and date of editing may be recorded. NCD only directly addresses the most recent edit, but an edit history could be built up using the <Notes> memo field.

Many of the fields may be repeated, either to accommodate multiple entries, such as the <Associated person> property in the example, or because the entry is in more than one language. Eight of the fields have English-language controlled terms associated with them, to aid searching and sorting.

Other fields may draw terms from existing authorities and it is recommended that an indication is given of the source of those terms along with, if possible, the identifier for the authority record for the term within that source. For an example, see the <Place name coverage> property in the example record, which gives the Getty Thesaurus of Geographic Names (**TGN**) identifiers for several of the place names entered. This service may be used from [http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

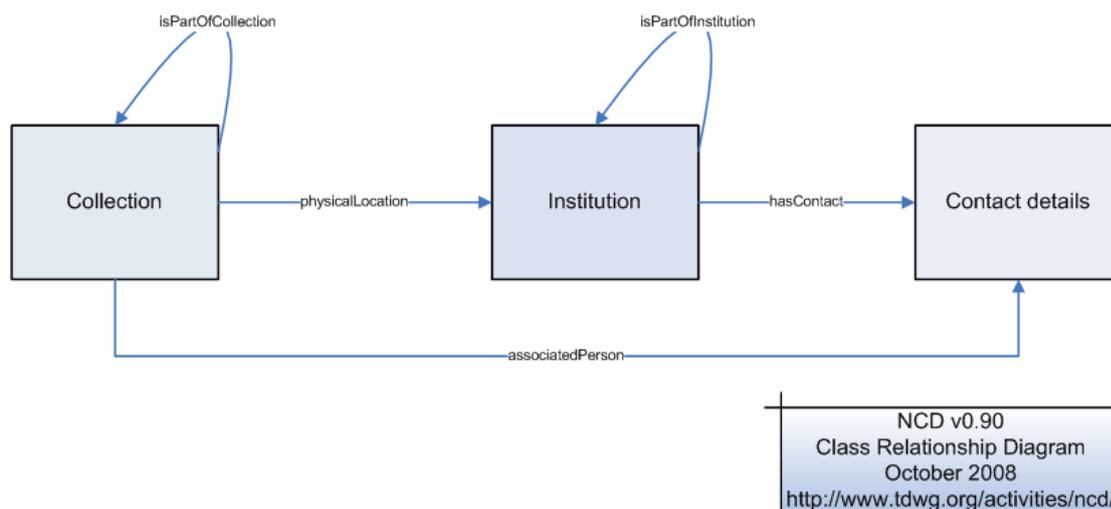


Figure 1 NCD Class Relationship Diagram

Each collection will be associated with one or more persons, through the <Associated person> property, or with an institution, which will typically be the owner and/or location of the collection. The vCard standard (<http://www.w3.org/TR/vcard-rdf>) has been adopted and supplemented for use in recording details about persons and institutions, since one of the main purposes for NCD records will be to find out who to contact for more information about consulting the collection.

An institution may be considered as a “collection of collections” and so has its own description property, along with a property for recording the various acronyms and codes by which it may be known. Similarly a network, such as BioCASE or the European Distributed Institute of Taxonomy **EDIT** (<http://www.e-taxonomy.eu/>) may be considered as a “collection of institutions”.

Collections may be related to their parent collection or institution and institutions may be related to their parent institution or network so that it is possible to build hierarchies. In general, it is easier to relate upwards to a parent than downwards to children. The latter may be achieved by requesting all records that have *this* identifier in their Parent collection identifier field.

## Implementation and Compliance

In a similar spirit to the Dublin Core metadata initiative (**DCMI**), NCD is defined in as a technology-neutral way as possible. It provides natural language definitions of classes, properties and instances that are identified by URIs and it makes recommendations on the use and content of properties from other vocabularies (Dublin Core and vCard).

The URIs defined here may be used across a number of technologies, such as namespaces in XML Schema validated documents and column headings in tab delimited text files.

This approach facilitates:

- Embedding of NCD data within other standards such as descriptions of specimens or literature.
- The extension of NCD records with other data types such as geospatial attributes.
- Cross walking between technologies such as a Comma Separated Value file, an RDF graph, an XML document and a JSON object.

The weakness of this approach is that this standard itself does not provide an off-the-shelf, self validating exchange format. The strength is that multiple such exchange formats meeting different requirements can be defined and this standard allows mapping between them.

The RDF files of the latest version of NCD may be found at: <http://rs.tdwg.org/ontology/voc/> (Note: Use **View | Source** in a Web browser to see the actual RDF).

To implement this standard, consult the NCD Toolkit User Guide. The NCD Toolkit was developed by ETI in Amsterdam and based on NCD v0.8. Individuals and institutions that would like to start managing their collection-level records in NCD are encouraged to make use of the Toolkit, which may be downloaded from Sourceforge at the URL provided below.

The Toolkit allows the export of data in NCD format so that records may be aggregated in to regional or national systems, or into the global Biodiversity Collections Index.

## NCD Normative documentation: Fields and definitions

**Note:**The Cardinality column shows fields that should be considered mandatory (**M**), repeatable (**R**), or may appear in one or more local languages (**L**).

### Header

Label	Definition URL	Description	Cardinality
	At <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a> unless otherwise indicated. See the rdfs:comment at the top of collection.rdf		
Record Source	<a href="http://purl.org/dc/elements/1.1/source">http://purl.org/dc/elements/1.1/source</a>	Source of the record if not created by the author named in Author	
Record Harvest Date	<a href="#">#recordHarvestDate</a>	Date the record was last harvested	
Author	<a href="http://purl.org/dc/elements/1.1/creator">http://purl.org/dc/elements/1.1/creator</a>	Person that created the record	<b>M</b>
Corporate Affiliation	<a href="http://www.w3.org/2001/vcard-rdf/3.0#Orgname">http://www.w3.org/2001/vcard-rdf/3.0#Orgname</a>	Organisational affiliation of the author	R L
Record Created Date	<a href="http://purl.org/dc/terms/created">http://purl.org/dc/terms/created</a>	Date of record creation	<b>M</b>
Editor	<a href="http://purl.org/dc/elements/1.1/contributor">http://purl.org/dc/elements/1.1/contributor</a>	Person that last edited the record	
Record Edited Date	<a href="http://purl.org/dc/terms/modified">http://purl.org/dc/terms/modified</a>	Date the record was last edited	
Record Rights	<a href="http://purl.org/dc/elements/1.1/rights">http://purl.org/dc/elements/1.1/rights</a>	IPR statement about the record	L
Notes	<a href="http://www.w3.org/2001/vcard-rdf/3.0#Note">http://www.w3.org/2001/vcard-rdf/3.0#Note</a>	Notes	L

### Collection

A group of specimens or other natural history objects.

Label	Definition URL	Description	Cardinality
	At <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a> unless otherwise indicated		
Derived Collection	<a href="#">#derivedCollection</a>	A "derived" collection record. The record has been derived from a query on an item-level database e.g. all items from Australia.	
Collection Identifier	<a href="#">#collectionId</a>	The URI (LSID or URL) of the collection. In RDF,	

Natural Collections Descriptions (NCD) v0.9 2008

		used as URI of the collection resource.	
Alternative Identifier	<a href="#">#alternativeld</a>	Alternative identifier for the collection with an indication of the source e.g. ISCW.	R
Parent Collection Identifier	<a href="#">#isPartOfCollection</a>	Identifier for the parent collection for this sub-collection. Enables a hierarchy of collections and sub collections to be built.	
Acronym or Coden	<a href="#">#acronymOrCoden</a>	Acronym, coden or initialism by which the collection is generally known indicating the source, e.g. Index Herbariorum.	R
Collection Name	<a href="http://purl.org/dc/elements/1.1/title">http://purl.org/dc/elements/1.1/title</a>	Official name of the Collection in the local language.	<b>M</b> R L
Alternative Name	<a href="http://purl.org/dc/terms/alternative">http://purl.org/dc/terms/alternative</a>	Alternative name for the collection. In English language where <i>Collection name</i> is not in English.	R L
Associated Person	<a href="http://www.w3.org/2001/vcard-rdf/3.0#FN">http://www.w3.org/2001/vcard-rdf/3.0#FN</a>	Name(s) of person(s) associated with the collection.	R
Description	<a href="http://purl.org/dc/elements/1.1/description">http://purl.org/dc/elements/1.1/description</a>	Main description element may be anything from a few words to an essay, suitable for a general audience.	<b>M</b> L
Description for Specialists	<a href="#">#descriptionForSpecialists</a>	Optional additional descriptive text using terms that are more suited to a specialist audience. Attribute indicates the language of the text.	L
Extent	<a href="#">#collectionExtent</a>	A free-text indication of the size or extent of the collection.	L
Collection Type	<a href="#">#collectionType</a>	Picklist keyword to indicate the typical content or unifying characteristic of the collection. Expected to contain an instance for the Collection Type Term vocabulary.	R
Common Name Coverage	<a href="#">#commonNameCoverage</a>	Common names or group names equivalent to the taxon names, in the local language.	R L

Natural Collections Descriptions (NCD) v0.9 2008

Conservation Status	<a href="#">#conservationStatus</a>	Picklist keyword indicating the conservation status of the collection using the McGinley scale. Link to the date of assessment and repeat to show sequence of snapshots.	R
Conservation Status Date	<a href="#">#conservationStatusDate</a>	The date that the conservation status was assessed.	R
Digital Format	<a href="http://purl.org/dc/elements/1.1/format">http://purl.org/dc/elements/1.1/format</a>	Use for digital collections, recording MIME Types or PUIDs.	R
Digital Medium	<a href="http://purl.org/dc/terms/medium">http://purl.org/dc/terms/medium</a>	Use for digital collections, recording the material or physical carrier of the resource e.g. DVD-R.	R L
Development Status	<a href="#">#developmentStatus</a>	Picklist keyword indicating potential for change to the scope of the collection.	
Expedition Name	<a href="#">#expeditionNameCoverage</a>	Uniform name for an expedition or voyage of discovery during which the material was collected.	R
Formation Period	<a href="#">#formationPeriod</a>	Text description of the time period during which the collection was assembled e.g. "Victorian", or "1922 - 1932", or "c. 1750".	R L
Place Name Coverage	<a href="#">#geospatialCoverage</a>	Place names from which material in the collection originated.	R L
Geospatial Coordinates	<a href="#">#geospatialCoordinates</a>	Decimal latitude(s) and longitude(s) of places of origin of the material in the collection.	R
Item-Level Access	<a href="#">#itemLevelAccess</a>	URI of an item-level data source. This can be a Web service or a Webpage designed for humans.	R
Kingdom Coverage	<a href="#">#kingdomCoverage</a>	Picklist keyword to indicate biological kingdom.	R
Known to Contain Types	<a href="#">#knownToContainTypes</a>	Flag property to indicate that the collection is known to include type specimens	
Living Time Period	<a href="#">#livingTimePeriodCoverage</a>	Time period during which biological material was alive. Includes palaeontological time periods or other text phrases.	R L

Natural Collections Descriptions (NCD) v0.9 2008

Physical Location	<a href="#">#physicalLocation</a>	Link to the contact details for the host institution, or person for a private collection.	L
Primary Grouping Principle	<a href="#">#primaryGroupingPrinciple</a>	Picklist keyword to indicate what the collection is about. The intellectual basis for grouping the collection rather than grouping based on physical characteristics. Expected to contain an instance from the Primary Grouping Principle Type Term vocabulary.	
Primary Purpose	<a href="#">#primaryPurpose</a>	Picklist key word to indicate the primary reason that the collection was assembled.	
Related Material	<a href="http://purl.org/dc/elements/1.1/relation">http://purl.org/dc/elements/1.1/relation</a>	Indicating resources that are related to the collection, such as printed catalogues, finding aids, Websites, obituaries, etc.	R L
Related Collection	<a href="#">#relatedCollection</a>	Short descriptions of related collections, such as for other components of a large collection that has been dispersed.	R L
Specimen Preservation Method	<a href="#">#specimenPreservationMethod</a>	Picklist keyword indicating the process or technique used to prevent physical deterioration of non-living collections. Expected to contain an instance from the Specimen Preservation Method Type Term vocabulary	R
Taxon Coverage	<a href="#">#taxonCoverage</a>	Taxon or taxa in the collection at Family level or higher.	R
Temporal Coverage	<a href="#">#temporalCoverage</a>	Time period covered by the materials in the collection.	R L
Access Conditions	<a href="http://purl.org/dc/terms/accessRights">http://purl.org/dc/terms/accessRights</a>	Terms and conditions under which the collection may be accessed. May be a URL to standard T&C. Default text for an institution may be overwritten for a particular collection if the restrictions are different.	R L

Natural Collections Descriptions (NCD) v0.9 2008

Usage Conditions	<a href="#">#usageRestrictions</a>	Terms and conditions under which the collection may be used. May be a URL to standard T&C. Default text for an institution may be overwritten for a particular collection if the restrictions are different.	R L
Provenance	<a href="http://purl.org/dc/terms/provenance">http://purl.org/dc/terms/provenance</a>	A statement of any changes in ownership and custody of the collection since its creation that are significant for its authenticity, integrity, and interpretation.	R L
IPR Statements	<a href="http://purl.org/dc/elements/1.1/rights">http://purl.org/dc/elements/1.1/rights</a>	Information about rights held in and over the resource. Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.	R L
Notes	<a href="http://rs.tdwg.org/ontology/voc/Common#note">http://rs.tdwg.org/ontology/voc/Common#note</a>	Notes	L

## Institution

An established organization or foundation, especially one dedicated to education, public service, or culture

Label	Definition URL	Description	Cardinality
	<a href="http://rs.tdwg.org/ontology/voc/Institution">http://rs.tdwg.org/ontology/voc/Institution</a> unless otherwise indicated		
Institution Identifier	<a href="#">#institutionId</a>	The URI (LSID or URL) of the institution. In RDF this will be used as URI of the institution resource.	
Name	<a href="http://www.w3.org/2001/vcard-rdf/3.0#Orgname">http://www.w3.org/2001/vcard-rdf/3.0#Orgname</a>	Official name of the Institution in the local language.	L
Alternative Name	<a href="http://purl.org/dc/terms/alternative">http://purl.org/dc/terms/alternative</a>	Alternative name for the institution in English, if the official name is not in English.	
Unit Name	<a href="http://www.w3.org/2001/vcard-rdf/3.0#OrgUnit">http://www.w3.org/2001/vcard-rdf/3.0#OrgUnit</a>	Name of a component unit, such as a department.	L
Description	<a href="http://purl.org/dc/elements/1.1/description">http://purl.org/dc/elements/1.1/description</a>	Description of institution, suitable for a general audience.	L
Alternative Identifier	<a href="#">#alternativeId</a>	Alternative identifier for the institution.	R
Contact	<a href="#">#hasContact</a>	Link to Contact details (address, email etc.) to get more information about the collection.	R
Type	<a href="#">#institutionType</a>	Picklist keyword describing the primary activity or purpose of an organisation or entity that holds or uses collections.	R
Parent Institution or Network	<a href="#">#isPartOfInstitution</a>	Name, acronym, identifier or coden of parent institution or network(s)	R L

## Contact Details

Information on how to contact an entity including postal and electronic methods.

Label	Definition URL <a href="http://www.w3.org/2001/vcard-rdf/3.0">http://www.w3.org/2001/vcard-rdf/3.0</a> unless otherwise indicated	Description	Cardinal ity
Name	#FN	Person name, as would appear on a name label.	
Family Name	#Family	Family name.	
Given Name	#Given	Given name.	
Other Name	#Other	Other name(s).	R
Prefix	#Prefix	Prefix, e.g. Mr or Dr.	
Suffix	#Suffix	Suffix e.g. Jr or III.	
Job Title	#Title	Job title e.g. Curator.	L
Role	#Role	Role e.g. Collector or Observer.	R L
Institution Name	#Orgname	Institution name.	L
Institution Unit	#Orgunit	Institution unit name, e.g. department name.	L
Post Office Box	#Pobox	Post Office box no.	
Extended Address	#Extadd	Extended address.	R
Street Number and Name	#Street	Street number and name.	
Local Area name	#Locality	Local area name.	
Regional Name	#Region	Regional name.	
Postcode or ZIP code	#Pcode	Postcode or ZIP code.	
Country Name	#Country	Country name.	
Telephone Number	#Tel	Telephone number.	R
Fax Number	#Fax	Fax number.	R
Email Address	#Email	Email address.	R
URL	#URL	URL for person or institution Web page	R
Logo URL	#Logo	URL for a logo.	R
Notes	#Note	Notes.	L

## Consistent Terminology (picklists)

Institution Type	At <a href="http://rs.tdwg.org/ontology/voc/InstitutionType">http://rs.tdwg.org/ontology/voc/InstitutionType</a>	
Institution Type	#institutionTypeTerm	Controlled vocabulary for institution types
Aquarium	#aquarium	An institution where aquatic collections of living organisms are kept and exhibited.
Archive	#archive	An institution where public and/or private records and historical documents are preserved and made accessible.
Botanic garden	#botanicGarden	A garden, sometimes with greenhouse, for the study and exhibition of special plants.
Conservation	#conservation	An organization dedicated to the conservation or preservation of wildlife and/or habitats.
Field station	#fieldStation	An establishment for the study and observation of phenomena that is in close proximity to the object of study.
Government	#government	Local or regional government agency or entity.
Historical society	#historicalSociety	An organization devoted to preserving the historical record of a place, institution, people, activity or thing
Horticultural institution	#horticulturalInstitution	An institution associated with the science and study of propagating plants.
Industry	#industry	A private or public industrial enterprise.
Laboratory	#laboratory	An institution or facility for experimental, observational or analytical studies.
Library	#library	An institution holding books, manuscripts, images and recordings for research and study.

Natural Collections Descriptions (NCD) v0.9 2008

Management	#management	An institution dedicated to preservation and management of natural resources in a specific region.
Museum	#museum	An institution for the procurement, preservation, study and exhibition of specimens and objects of value.
Nature education center	#natureEducationCenter	An institution dedicated to informing audiences on phenomena related to the natural world.
Non university college	#nonUniversityCollege	An institution of higher learning that does not grant advanced degrees beyond the bachelor's.
Park	#park	Managed open space e.g. national park
Repository	#repository	An institution that provides care and accessibility to deposited materials based upon a series of defined formal agreements.
Research institute	#researchInstitute	An institution dedicated to acquiring, developing and communicating knowledge.
School	#school	An institution dedicated to providing basic knowledge and education, typically from the earliest stages to the college or university level.
Science center	#scienceCenter	An institution dedicated to informing audiences on scientific facts, phenomena and theory.
Society	#society	An organization or association of persons engaged in a common profession, activity, or interest
University	#university	A degree-granting institution of higher learning.
Zoo	#zoo	An institution where living animals are maintained and exhibited.

Natural Collections Descriptions (NCD) v0.9 2008

<b>CollectionType</b>	<b>At</b> <a href="http://rs.tdwg.org/ontology/voc/CollectionType">http://rs.tdwg.org/ontology/voc/CollectionType</a>	<b>Definition</b>	<b>Notes</b>
Archival	#archival	Public and/or private records and historical documents in various media which are preserved and accessible.	Papers and records, field logs, photographs, etc.
Art	#art	Representational and impressionistic works from the human hand.	Paintings, drawings, scientific illustrations.
Audio	#audio	Recorded auditory phenomena.	Amphibian mating calls.
Cell Cultures	#cellCultures	Biological material that is intentionally cultured and maintained in a viable state.	cell lines, microorganisms.
Electronic	#electronic	Data or information that has been captured and fixed for storage and manipulation in an automated system and that requires the use of the system to render it intelligible by a person	Electronic documents, web sites, databases.
Facsimiles	#facsimiles	Accurate representations of the original item; may be naturally occurring or manufactured.	Endocasts, moulds.
Fossils	#fossils	Biological specimens that are preserved through geological processes.	Tyrannosaurus Rex.
Genetic	#genetic	Stored or banked biological samples for future use.	Germplasm, seed, blood, DNA.
Geological	#geological	Collections of rocks and gemstones	
Herbarium	#herbarium	Collections of preserved plant material	
Living	#living	Collections that grow and reproduce, including botanic gardens, zoos and aquaria.	Plants, fish.
Manuscripts	#manuscripts	Collections of manuscripts	Letters, field note books
Mineralogical	#mineralogical	Collections of minerals	Rocks, minerals, gemstones, meteorites
Observations	#observations	Recorded visual phenomena, may be observation or measurement.	Avian mating displays.
Preserved	#preserved	Biological specimens that are	Pinned butterfly,

Natural Collections Descriptions (NCD) v0.9 2008

		preserved indefinitely.	herbarium specimen, skeleton, skin, fluid-preserved invertebrate.
Products	#audio	Specimens that result from biological functions or behaviours.	Nests, webs, moults
Specimens	#specimens	Relating to a particular grouping of biological or geological specimens.	North American Mammal collection, lichen collection, arachnid collection.
Texts	#texts	A collection of books or other written material, or a list of works.	Books, journals and other printed publications.
Tissue	#tissue	Plant and/or animal tissues.	Blood, muscle, ovary, leaf.
Visual	#visual	Visual representations of an item or object produced with the aid of a camera or a digital imaging system.	Photograph, SEM image, video.

KingdomType	At <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a>	Controlled vocabulary for taxonomic kingdom.
Animalia	#animalia	Animals
Archaeobacteria	#archaeobacteria	Archaeobacteria
Eubacteria	#eubacteria	Eubacteria
Fungi	#fungi	Fungi
Plantae	#plantae	Plants
Protista	#protista	Protists

Primary Grouping Principle Type	At <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a>	Definition	Notes
Cultural	#cultural	Collections relating to a group of people.	Example: Victorian taxidermy
Ecosystems	#ecosystems	Collections relating to complex community of living organisms and their environment.	Examples: Water column samples or forest plot collections.
Environmental	#environmental	The aggregate of external conditions in which organisms live or phenomena occur; may be defined components only.	Examples: Rainfall measurements, average temperature, specimens from hydrothermal vents.

Natural Collections Descriptions (NCD) v0.9 2008

Events	#events	Collections that were made as a result of an event or that document an event.	Examples: ash samples collected during a specific volcanic eruption, birds collected or killed as a result of suspected H5N1 (bird flu)
Expeditions	#expeditions	Collections resulting from a named journey or trip such as a voyage of discovery.	Examples: Items from Captain Cook's First Voyage or from H.M.S. Beagle
Historical	#historical	relating to a particular time period relevant to human history	Example: Collecting equipment used in the 1700's
National	#national	Collections of material from a specific country c.f. regional	Example: Fossils from Canada
Relationships	#relationships	Organisms sharing an intimate biological relationship.	Examples: Parasitic or Symbiotic
Repository	#repository	Similar or related items that are placed in a facility for long term care as the result of a specific agreement.	Examples: Specimens from US Mineral Management Service, specimens from US National Park Service.
Regional	#spatial	Relating to a physical location or region rather than a country c.f. national.	Examples: Plants from North America
Stage	#stage	Relating to a specific phase in physical development.	"Examples: Larva, fetus.
Taxonomic	#taxonomic	Collections laid out according to a particular taxonomy.	Example: Insect collection grouped by Family
Temporal	#temporal	Relating to a particular time period.	Examples: Cenozoic mollusks.

<b>Primary Purpose</b>	<b>At</b> <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a>	<b>Controlled vocabulary for primary purpose.</b>
Commercial	#commercial	Biological or geological material for sale or other commercial use, such as some culture collections.
Education	#education	Acquired to educate, inform, or convey

Natural Collections Descriptions (NCD) v0.9 2008

		information e.g. for distance learning.
Exhibition	#exhibition	Acquired to educate, inform, or convey information through display.
Monitoring	#monitoring	Acquired as a result of observation, measurement, or sampling.
Ornamental	#ornamental	Acquired for decorative purposes.
Personal	#personal	Acquired for idiosyncratic reasons.
Research	#research	Acquired for current and future study.
Voucher	#voucher	Acquired and maintained to document organisms, observations, or phenomena.

<b>Specimen Preservation Method</b>	<b>At</b> <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a>	<b>Controlled vocabulary for specimen preservation method.</b>
Controlled atmosphere	#controlledAtmosphere	Held in conditions that vary from the typical elemental mix of the atmosphere; usually implies reduced oxygen.
Cryopreserved	#cryopreserved	Held at very low temperatures, typically in liquid Nitrogen.
Dried	#dried	Treated or exposed to remove all internal fluids; may have used silica gel.
Dried and pressed	#driedAndPressed	Treated to remove all internal fluids; flattened, typically mounted on paper.
Embedded	#embedded	Maintained in a solid matrix suitable for making thin sections.
Fluid preserved	#fluidPreserved	Maintained in formalin, ethyl alcohol, or other liquid preservative.
Freeze dried	#freezeDried	Removal of internal liquids using cold, low humidity environment.
Frozen	#frozen	Held at temperatures below freezing (32 Degrees F or 0 degrees C).
Glycerin	#glycerin	Replacement of original internal fluids with glycerine.
No treatment	#noTreatment	Absence of any deliberate preservation action.
Not applicable	#notApplicable	For living collections that grow and reproduce.
Pinned	#pinned	Treated to remove all internal fluids; held on metallic pin.
Recorded (analog)	#recordedAnalog	Observations, measurements, or images that capture phenomena in analog formats.
Recorded (digital)	#recordedDigital	Observations, measurements, or images that capture phenomena in digital formats.
Refrigerated	#refrigerated	Maintained at cool temperatures above freezing.
SEM stub	#semstub	Mounted and coated for scanning electron microscopy.

Natural Collections Descriptions (NCD) v0.9 2008

Skeletonized	#skeletonized	Removal of soft tissues including connective tissues.
Slide mount	#slideMount	Held on a glass slide using fixative and cover slip.
Stasis	#stasis	Collections capable of growing and reproducing that are held in a suspended state.
Surface coating	#surfaceCoating	Application of protective or consolidating surface coating using organic or inorganic materials.
Tanned	#tanned	Chemically treated skins and hides.

<b>Development Status Type</b>	<b>At</b> <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a>	<b>Controlled vocabulary for development status.</b>
Active growth	#activeGrowth	Currently increasing in volume, count, number of taxa represented or other dimension due to intentional activity.
Closed	#closed	Currently closed to acquisitions or deposits.
Consumable	#consumable	Collection may or will be decreasing in volume, count, representation or other dimension due to intentional destructive analytical activity.
Decreasing	#decreasing	Currently diminishing in volume, count, number of taxa represented, or other dimension due to intentional activity.
Lost	#lost	Host institution has concluded that collection is permanently unavailable after appropriate efforts have been made although documentation of the collection exists.
Missing	#missing	Physical collection cannot be located although its documentation exists; the possibility of finding the collection exists.
Passive growth	#passiveGrowth	May increase in volume, count, number of taxa represented or other dimension due to unplanned and unpredictable activity.
Static	#static	Cannot or will not increase or decrease in any manner.

<b>Conservation Status Type Term</b>	<b>At</b> <a href="http://rs.tdwg.org/ontology/voc/Collection">http://rs.tdwg.org/ontology/voc/Collection</a>	<b>Controlled vocabulary for conservation status. Terms are taken from the McGinley levels.</b>
McGinley Level 1	#mcginley1	Specimens deteriorating, potentially cullable, or unprepared. Collection unit in need of immediate attention; museum pests, rusting pins, crystallizing slide media, unringed Hoyer's media,

Natural Collections Descriptions (NCD) v0.9 2008

		evaporated alcohol, fading labels, broken cover slip or slide, etc. Primary types missed in general collection.
McGinley Level 2	#mcginley2	Material properly prepared but not sorted or only rough-sorted; not readily available to specialists.
McGinley Level 3	#mcginley3	Specimens sorted to a level necessary to be efficiently accessible to research specialist for study. All specimens in soft-bottom trays, shell vials in jars or slide boxes, with appropriate levels denoting pertinent taxonomic information. Taxonomic category may vary among different taxa.
McGinley Level 4	#mcginley4	Valuable material that has been identified to the species level but not yet put away effectively inaccessible.
McGinley Level 5	#mcginley5	All specimens identified and integrated. However, this collection unit needs upgrading, e.g., names checked, header labels for unit trays prepared, transfer to soft-bottom trays, etc., box header labels checked against slides. Alcoholic collection with individual exposed vials, jar labels misleading or without detail, overall format heterogenous.
McGinley Level 6	#mcginley6	All specimens in soft-bottom unit trays; all unit-tray header labels, drawer or box, and cabinet labels completed; space left for expansion. When a collection reaches Level 6, the name(s) or who did the curation, the date of curation and listing of catalogues or monographs on which the names are based should be entered in the Smithsonian Collection Inventory File computerized files. Vial within jar system established, labels complete and accurately typed, expansion space allocated, and alcohol levels adequate.
McGinley Level 7	#mcginley7	Species inventory with header or box label generation completed. Alcoholic collection with taxa listed on jar labels and in database, as in Level 6 plus species-level inventory.
McGinley Level 8	#mcginley8	Label data are recorded for systematic, biogeography or natural history studies. The data elements recorded will vary by group, however, a standard core set includes the following: a) a country/province/place/latitude-longitude; b) date of collection; c) collector(s); d) miscellaneous (for example, elevation, host). Miscellaneous voucher specimens, including types, from various studies are also ranked at this level.
McGinley Level 9	#mcginley9	Measurements, graphics, or other pertinent (descriptions) of specimens are captured in a database.
McGinley Level 10	#mcginley10	Groups of specimens included in published monographic, synthetic, or revisionary works,

		including primary type collection if storage is at Level 7 or higher. When groups of specimens serve to voucher published scientific collection-based studies and have undergone conservation and accession requirements of at least Level 7.
--	--	---

## NCD: Normative documentation: Example NCD record

**Note:**The Cardinality column shows fields that should be considered mandatory (**M**), repeatable (**R**), or may appear in one or more local languages (**L**).

### Header

Metadata about the record itself.

Label	Data	Description	Cardinality
Record Source	Test record transferred from NHM Collections Navigator. Originally authored in the NHM Unicorn LMS	Source of the record if not created by the author named in Author	
Record Harvest Date	2003-05-13	Date the record was last harvested	
Author	Susan Bennett	Person that created the record	<b>M</b>
Corporate Affiliation	Natural History Museum, London	Organisational affiliation of the author	R L
Record Created Date	2002-09-09	Date of record creation	<b>M</b>
Editor	Neil Thomson	Person that last edited the record	
Record Edited Date	2008-07-15	Date the record was last edited	
Record Rights	© 1994 - 2002 The Natural History Museum (London)	IPR statement about the record	L
Notes		Notes	L

### Collection

A group of specimens or other natural history objects.

Label	Data	Description	Cardinality
Derived Collection		A "derived" collection record. The record has been derived from a query on an item-level database e.g. all items from Australia.	

Natural Collections Descriptions (NCD) v0.9 2008

Collection Identifier		LSID or URL of the RDF:about record	
Alternative Identifier	Museum Number: uls-a352878	Alternative identifier for the collection with an indication of the source e.g. ISCW.	R

Natural Collections Descriptions (NCD) v0.9 2008

Parent Collection Identifier		Identifier for the parent collection for this sub-collection. Enables a hierarchy of collections and sub collections to be built.	
Acronym or Coden		Acronym, coden or initialism by which the institution is generally known indicating the source, e.g. Index Herbariorum.	R
Collection Name	Manuscript Collection of Hugh Whistler (1889-1943)	Official name of the Collection in the local language.	<b>M R L</b>
Alternative Name		Alternative name for the collection. In English language where <i>Collection name</i> is not in English.	R L
Associated Person	Whistler, Hugh, 1889-1943	Name(s) of person(s) associated with the collection.	R
Associated Person	Whistler, Fuller	[repeat]	
Associated Person	Ticehurst, Claud Buchanan, 1881-1941	[repeat]	
Associated Person	Mason, Kenneth	[repeat]	
Associated Person	Cave, Francis Oswin, 1897-1974	[repeat]	
Description	<p>The manuscript collection of Hugh Whistler (1889-1943) comprises notes on birds collected in 1926 on an expedition into the Shaksgam Valley and the Aghil Range which lie near the frontier between Kashmir and Chinese Turkestan [1926-1927] by K. Mason and F. O. Cave, accompanied by newspaper cuttings and letters to Whistler.</p> <p>The collection also contains notes on birds seen and collected on a trip to Sicily 1933, lists of Indian collecting localities compiled circa 1937, with annotated printed maps, and the following volumes entitled:</p> <p><i>Collection of birds skins (mounted and unmounted) in the possession of Captain F[uller] Whistler and Hugh</i></p>	Main description element may be anything from a few words to an essay, suitable for a general audience.	<b>M L</b>

Natural Collections Descriptions (NCD) v0.9 2008

	<p><i>Whistler of Battle, Sussex</i> [from the British Isles, Egypt and India] 1909-1916.</p> <p><i>Measurements of Indian Birds</i> compiled circa 1940, <i>Measurements of Palearctic Birds</i> compiled circa 1942, and letters, maps, printed cuttings, and notes, including two boxes on the moult of mostly Indian birds.</p> <p><i>Birds of the Himalayas</i>, circa 1942, an incomplete typescript volume.</p>		
Description for Specialists	<p>Hugh Whistler (1889-1943), F.Z.S., M.B.O.U., was born at Mablethorpe in Lincolnshire and educated at Aldenham School. He served with the Indian police in the Punjab province for seventeen years.</p> <p>Whistler studied and collected birds and on retiring to England in 1926 he continued his researches into Indian ornithology. He also made collecting trips to Europe and North Africa, often in the company of Claud Buchanan Ticehurst (1881-1941).</p> <p>Whistler's publications include the <i>Popular handbook of Indian birds</i> (1928) and many papers on the birds of India.</p> <p>In collaboration with C. B. Ticehurst he began work on what was to be a comprehensive handbook on the birds of India, Burma and Ceylon.</p> <p>With the death of Whistler two years after that of Ticehurst the work was never completed.</p>	Optional additional descriptive text using terms that are more suited to a specialist audience. Attribute indicates the language of the text.	L
Extent	26 boxes of manuscript notes, 8 volumes, 2 typescripts, 1 manuscript, 2 notebooks, Letters, 2 newspaper cuttings and maps maps	An indication of the size or extent of the collection.	L
Collection Type	Manuscripts	Picklist keyword to indicate the typical content or unifying characteristic of the collection.	R

Natural Collections Descriptions (NCD) v0.9 2008

Common Name Coverage	Birds (LCSH)	Common names or group names equivalent to the taxon names, in the local language.	R L
Conservation Status		Picklist keyword indicating the conservation status of the collection using the McGinley scale. Link to the date of assessment and repeat to show sequence of snapshots.	R
Conservation Status Date		The date that the conservation status was assessed.	R
Digital Format		Use for digital collections, recording MIME Types or PUIDs.	R
Digital Medium		Use for digital collections, recording the material or physical carrier of the resource e.g. DVD.	R L
Development Status	Closed	Picklist keyword indicating potential for change to the scope of the collection.	
Expedition Name		Uniform name for an expedition or voyage of discovery during which the material was collected.	R
Formation Period	c. 1926	Text description of the time period during which the collection was assembled e.g. "Victorian", or "1922 - 1932", or "c. 1750".	R L
Place Name Coverage	Shaksgam Valley (local)	Place names from which material in the collection originated.	R L
Place Name Coverage	Aghil Range (local)	[repeat]	
Place Name Coverage	Jammu-Kashmir (tgn 7000799)	[repeat]	
Place Name Coverage	Kashmir local Bharat (tgn 7000198)	[repeat]	
Place Name Coverage	Asia (tgn 1000004)	[repeat]	

Natural Collections Descriptions (NCD) v0.9 2008

Place Name Coverage	Turkestan (tgn 7016836)	[repeat]	
Place Name Coverage	Sicilia (tgn 7003122)	[repeat]	
Place Name Coverage	Sicily (local)	[repeat]	
Place Name Coverage	Italia (tgn 1000080)	[repeat]	
Place Name Coverage	Europe (tgn 1000003)	[repeat]	
Place Name Coverage	India (local)	[repeat]	
Place Name Coverage	Great Britain (tgn 7008653)	[repeat]	
Place Name Coverage	United Kingdom (tgn 7008591)	[repeat]	
Place Name Coverage	As Sa`id (tgn 7001437)	[repeat]	
Place Name Coverage	Misr (tgn 7016833)	[repeat]	
Place Name Coverage	Egypt (local)	[repeat]	
Place Name Coverage	Africa (tgn 7001242)	[repeat]	
Place Name Coverage	Himalayas (tgn 7016919)	[repeat]	
Geospatial Coordinates		Decimal latitude(s) and longitude(s) of places of origin of the material in the collection.	R
Item-Level Access		URI of an item-level data source. This can be a Web service or a Webpage designed for humans.	R
Kingdom Coverage	Animalia	Picklist keyword to indicate biological kingdom.	R
Known to Contain Types		Flag element to indicate that the collection is known to include type specimens	
Living Time Period	Early 20 <sup>th</sup> century	Time period during which biological material was alive. Includes palaeontological time periods or other text phrases.	R L

Natural Collections Descriptions (NCD) v0.9 2008

Physical Location	Tring Library MSS. WHISTLER 1	Link to the contact details for the host institution, or person for a private collection.	L
Primary Grouping Principle		Picklist keyword to indicate what the collection is about. The intellectual basis for grouping the collection rather than grouping based on physical characteristics.	
Primary Purpose	Research	Picklist key word to indicate the primary reason that the collection was assembled.	
Related Material	Finding aid by Frances E. Warr. <i>Manuscripts and Drawings in the Ornithology and Rothschild Libraries of The Natural History Museum at Tring</i> . Tring, Hertfordshire: British Ornithologists' Club in association with The Natural History Museum, London, 1996.	Indicating resources that are related to the collection, such as printed catalogues, finding aids, Websites, obituaries, etc.	R L
Related Material	Obituary. <i>Ibis</i> , 1943 pp.524-532.	[repeat]	
Related Collection	Whistler's huge collection of 17,000 birds was presented by the Hon. Mrs Whistler to The Natural History Museum (London)	Short descriptions of related collections, such as for other components of a large collection that has been dispersed.	R L
Related Collection	Letters written by Claud Buchanan Ticehurst (1881-1941) to Hugh Whistler (1889- 1943) are in the Zoology Department of The Natural History Museum (London)	[repeat]	
Specimen Preservation Method		Picklist keyword indicating the process or technique used to prevent physical deterioration of non-living collections.	R
Taxon Coverage	Aves	Taxon or taxa in the collection at Family level or higher.	R
Temporal Coverage	Early 20 <sup>th</sup> century	Time period covered by the materials in the collection.	R L
Access	<a href="http://www.nhm.ac.uk/research-">http://www.nhm.ac.uk/research-</a>	Terms and conditions	R L

Conditions	<a href="#">curation/collections-library/collections-management/collections-navigator/accessrestrict.jsp</a>	under which the collection may be accessed. May be a URL to standard T&C. Default text for an institution may be overwritten for a particular collection if the restrictions are different.	
Usage Conditions	<a href="http://www.nhm.ac.uk/research-curation/collections-library/collections-management/collections-navigator/userrestrict.jsp">http://www.nhm.ac.uk/research-curation/collections-library/collections-management/collections-navigator/userrestrict.jsp</a>	Terms and conditions under which the collection may be used. May be a URL to standard T&C. Default text for an institution may be overwritten for a particular collection if the restrictions are different.	R L
Provenance		A statement of any changes in ownership and custody of the collection since its creation that are significant for its authenticity, integrity, and interpretation.	R L
IPR Statements	Citation: By permission of the Trustees of The Natural History Museum (London).	Information about rights held in and over the resource. Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights.	R L
Notes		Notes	L

## Institution

An established organization or foundation, especially one dedicated to education, public service, or culture

Label	Data	Description	Cardinality
Institution Identifier		LSID or URL of the RDF:about record	
Name	Natural History Museum	Official name of the Institution in the local language.	L
Alternative Name		Alternative name for the institution in	

Natural Collections Descriptions (NCD) v0.9 2008

		English, if the official name is not in English.	
Unit Name	Department of Ornithology Library	Name of a component unit, such as a department.	L
Description	Outstation of the Natural History Museum at Tring in Hertfordshire, previously owned by Lionel Walter Rothschild	Description of institution, suitable for a general audience.	L
Alternative Identifier		Alternative identifier for the institution.	R
Contact	The Librarian, Natural History Museum, Department of Ornithology.	Link to Contact details (address, email etc.) to get more information about the collection.	R
Type	Research Institute Museum	Picklist keyword describing the primary activity or purpose of an organization or entity that holds or uses collections.	R
Parent Institution or Network	<a href="http://urn:lsid:biocol.org:col:34665">urn:lsid:biocol.org:col:34665</a>	Name, acronym or coden of parent institution or network(s)	R L

## Contact Details

Information on how to contact an entity including postal and electronic methods.

Label	Data	Description	Cardinality
Name		Person name, as would appear on a name label.	
Family Name		Family name.	
Given Name		Given name.	
Other Name		Other name(s).	R
Prefix		Prefix, e.g. Mr or Dr.	
Suffix		Suffix e.g. Jr or III.	
Job Title	The Librarian	Job title e.g. Curator.	L
Role		Role e.g. Collector or Observer.	R L
Institution Name	Natural History Museum	Institution name.	L
Institution Unit	Department of Ornithology	Institution unit name, e.g. department name.	L
Post Office Box		Post Office box no.	
Extended Address		Extended address.	R
Street Number	The Walter Rothschild building ,	Street number and	

Natural Collections Descriptions (NCD) v0.9 2008

and Name	Akeman Street	name.	
Local Area name	Tring	Local area name.	
Regional Name	Hertfordshire	Regional name.	
Postcode or ZIP code	HP23 6AP	Postcode or ZIP code.	
Country Name	United Kingdom	Country name.	
Telephone Number	+44 (0)20 7942 6171	Telephone number.	R
Fax Number		Fax number.	R
Email Address		Email address.	R
URL	<a href="http://www.nhm.ac.uk">http://www.nhm.ac.uk</a>	URL for person or institution Web page	R
Logo URL		URL for a logo.	R
Notes		Notes.	L

---

End Normative Document

# Natural Collections Description (NCD)

A data standard for exchanging data describing natural history collections

## NON-NORMATIVE DOCUMENT

Neil Thomson (Natural History Museum, London), Roger Hyam (Royal Botanic Garden Edinburgh), Constance Rinaldo (Harvard University), Carol Butler (Smithsonian Institution), Doug Holland (Missouri Botanical Gardens), Barbara Mathé (American Museum of Natural History), Günter Waibel (RLG Programs, OCLC), Wouter Addink (ETI Bioinformatics), Ruud Altenburg (ETI), Markus Döring (Berlin Botanic Garden)

### 1. Summary

**Note:** This is a non-normative companion document, which provides some background to the aims and uses of the proposed standard. The normative document has been separated and may be found on the TDWG Website at <http://www.tdwg.org/activities/ncd/>

Natural Collections Description (**NCD**)<sup>1</sup> is a data standard for describing collections of natural history materials at the collection level; one NCD record describes one entire collection.

Collection descriptions are electronic records that document the holdings of an organisation as groups of items, which complement the more traditional item-level records such as are produced for a single specimen or a library book. NCD is tailored to natural history. It lies between general resource discovery standards such as Dublin Core (**DC**) and rich collection description standards such as the Encoded Archival Description (**EAD**). It is possible to extract a Dublin Core record from an NCD record for use with general resource discovery systems, or to use an NCD record as the seed for a richer collection description, like an EAD record.

The NCD standard covers all types of natural history collections, such as specimens, original artwork, archives, observations, library materials, datasets, photographs or mixed collections such as those that result from expeditions and voyages of discovery.

NCD primarily holds information about collections of objects, but can also be used to describe organisations (collections of collections) and networks (collections of organisations). There are many existing sources of information about biodiversity organisations, but they are scattered and in different formats.

---

<sup>1</sup> A glossary of acronyms is provided at Appendix 2. First usage of an acronym is in bold font.

## Natural Collections Descriptions (NCD) v0.9 2008

This document accompanies the normative part of the NCD standard. The standard consists of the series of class and property definitions and is presented in a separate document. These definitions are identified by unique TDWG Uniform Resource Identifiers (URI). It assumes that copies of the definitions will be hosted at the specified URIs given in the normative form document. The URL stem is <http://rs.tdwg.org/ontology/voc/>

The standard also contains recommendations on the use of Dublin Core and vCard properties. This avoids duplicating established vocabularies and facilitates interpretation of NCD documents by non NCD aware applications.

It is expected that NCD will develop further as experience is gained in the projects that are making use of it, particularly in the addition of terms to the pick-lists. It has reached a sufficiently mature state that applications may be based on it. Following approval from the TDWG appraisers it will be designated as NCD version 1.0 and made widely known to the biodiversity informatics community.

The NCD standard is the culmination of work on collections descriptions carried out for the European Union Framework VI programme SYNTHESYS and the work performed by the Anglo-American group Resources Available in Natural Sciences (**RAVNS**), which operates under the auspices of RLG Programs, OCLC.

The normative documentation includes an example record and, for those that are developing applications based on NCD a column provides suggested cardinality – that is, for fields that should be considered mandatory (**M**) or may have more than one value – that is, repeatable (**R**) or indicate text in the local language (**L**).

To ensure that the barriers to usage are as low as possible, only four properties of the Collection class are considered to be mandatory:

1. Author of the record
2. Date of record creation
3. Collection name
4. Collection description

An NCD Toolkit has been developed by ETI Bioinformatics in Amsterdam (<http://www.eti.uva.nl/>) with the aid of funding from GBIF and is available for download at <https://sourceforge.net/projects/ncdtoolkit/>.

Version 1.0 of the NCD Toolkit is based on NCD v0.8. It is a cross-platform database which enables natural history organisations to record data about their own collections.

## 2. Motivation and Rationale

Many valuable collections exist that have no information stored in databases, nor do they have a web presence. Such collections are easily overlooked by researchers, so a brief descriptive record as defined by the NCD standard can act as the “business card” for a collection, providing enough information to identify and locate it.

The standard enables the aggregation of collections descriptions from many sources and facilitates resource discovery, including establishing relationships among collections in several locations. NCD records can also be used as an aid for collections management processes, allowing an institution to take a step back and see which collections are most in need of conservation or would benefit from a higher priority for item-level cataloguing.

The standard was developed by the TDWG NCD Interest Group to fit with the suite of data standards being developed on behalf of the Global Biodiversity Information Facility (GBIF) by Biodiversity Information Standards (TDWG).

Diagram 1 gives a very simplified view of the use of these and other standards to exchange information between some of the stakeholders in biodiversity informatics.

The relationships between some of the major components are represented by reading down the left-hand column, which also shows an organisation or project that is developing each. Data interchange standards enable the flow of data between providers, users and communities and one of each is shown bridging the first and second columns, with a key to the acronyms provided at the foot of the diagram. Note that this is a two-way flow. Finally, the influence of name servers and generic data such as globally unique identifiers (**GUID**), dates and geospatial coordinates in providing consistency is shown in the right-hand column with some of the driver organisations that provide guidance on good practice.

One use for NCD can be seen within this overall picture, providing information about the collections that hold specimens. A project based on NCD (which cannot be named in this document) aims to provide a GUID for all such collections throughout the world so that researchers may unambiguously specify the source of their material and contact information for those wishing to visit collections. NCD provides a mechanism for exchanging and aggregating information about collections in a standard format so that applications based on it may share data.

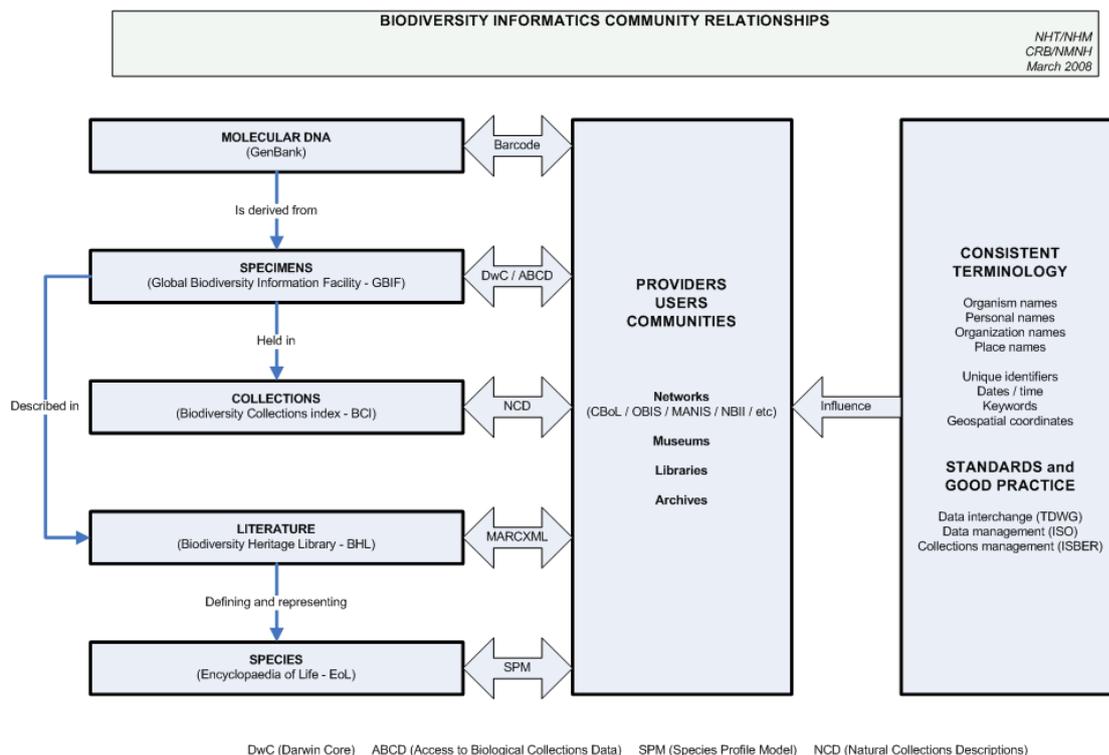


Figure 2: Simplified view of biodiversity community relationships and standards

## Natural Collections Descriptions (NCD) v0.9 2008

Funding has been provided by:

- The European Commission
- Gordon & Betty Moore Foundation
- Natural History Museum, London
- RLG Programs, OCLC
- Smithsonian Institution

Grateful thanks are offered to all of these.

### **3. Further Information**

- NCD Website:  
<http://www.tdwg.org/activities/ncd/>
- To join the mailing list:  
<http://lists.tdwg.org/mailman/listinfo/tdwg-ncd>
- Discussion and documents:  
<http://wiki.tdwg.org/twiki/bin/view/NCD/WebHome>
- NCD Toolkit:  
<https://sourceforge.net/projects/ncdtoolkit/>

## Appendix 1: NCD Development Background

The genesis of the NCD data standard can be traced back to 1999 at the Natural History Museum (**NHM**) in London, UK. At that time, a new building was being created that would allow the visiting public to view something of the extent of the working collections of specimens in addition to those highlight specimens that are normally on view in the exhibitions areas.

This building was to be called the Darwin Centre and it was intended to make the first class scientific research that is carried out at the Museum more visible and better known. The key phrase was “unprecedented access to the collections”, but there was no overall understanding of what collections existed in the Museum to which to give this unprecedented access.

Funding was provided to create collection-level descriptions (**CLDs**) for all the Museum’s collections, whether they were of specimens in the science departments; original artwork depicting those specimens held in the Library; expedition field guides held in the Archives; sculptures scattered through the galleries; learning materials or some other collection. The ability to link related, but separately managed, collections was a bonus.

Over 1,000 records were created over the next couple of years and two things became clear. One was that this was just the tip of a very large iceberg and the other was that the data standard being used was too rich for the purpose. The project made use of the archival standard Encoded Archival Description (**EAD**) since this was specifically designed to describe collections and was already in use in the Museum Archives.

Thoughts turned to creating a simplified metadata standard that would serve to record guide-book style information about collections which could be amplified at a later stage, if necessary.

About this time there were two relevant external developments. One was a rising interest in CLDs, mainly driven by the Research Support Libraries Programme (**RSLP**) in the UK. This was developing a generalised simple collection description standard which would eventually become one of the Dublin Core specialties, DC: Collections. Although this had its attractions, it was believed that natural history collections would need certain specialised fields over and above what was under discussion for this developing standard – an example is the “Known to contain types” field.

The other development was the involvement of the NHM in the BioCASE project, funded by the European Union between 2001 and 2004 and led by Professor Walter Berendsohn of the Berlin Botanic Garden. The current BioCASE service can be found at <http://www.biocase.org/>

The BioCASE partnership of over 30 countries can be viewed as a precursor to GBIF, mobilising and integrating specimen data using new standards. It was recognised that it would be many years before all specimens would have a database record, so a means to alert researchers to collections with no database, but worthy of attention, was required. The data standard for the gathering and exchange of specimen data became Access to Biological Collections Data (**ABCD**) and the collection-level complement to ABCD became the foundation for NCD. Information about ABCD can be found at

Natural Collections Descriptions (NCD) v0.9 2008

<http://www.bgbm.org/tdwg/CODATA/Schema/default.htm> and at <http://www.tdwg.org/standards/115/>

National nodes were set up that would create and organise the data for their country and a central index was created from data harvested on a nightly basis from the national nodes. The Berlin team developed the National Node Data Input Tool (**NoDIT**), which was a Microsoft® Access™ database based on the collection-level data XML Schema, which enabled the system, using a computer-to-computer data transfer protocol that developed into TAPIR. A brief description of the harvesting process into the Core Metadatabase (**CorM**) can be found at [http://www.biocase.org/whats\\_biocase/meta\\_net\\_old.shtml](http://www.biocase.org/whats_biocase/meta_net_old.shtml)

International interest in collections descriptions for natural history extended across the Atlantic with the formation of the RLG Natural History Group. **RLG** (which used to be known as the Research Libraries Group and is now RLG Programs, OCLC) acted as facilitator for several large natural history museums and botanic gardens, providing a mechanism for collaborative working that proved to be very successful under the leadership of Günter Waibel.

The Group became known as the RAVNS – Resources Available in Natural Sciences. This group developed the BioCASE collection-level XML schema from being relevant only to collections of specimens to a schema that would also manage descriptions of materials found in natural history libraries and archives, so making it truly cross-domain. This group was later expanded into the NCD Interest Group as part of the re-developed Taxonomic Databases Working Group (**TDWG**) and are collectively the authors of this standard.

Since TDWG mandated that the Resource Description Framework (**RDF**) would be its preferred technical environment, NCD was converted from an XML Schema into RDF. This proved to be a complex and controversial move and there are still calls for an XML Schema version, but keeping the two synchronised would be more troublesome than useful and so the normative version of NCD is in RDF, as presented here.

## Appendix 2: Glossary

<b>BCI</b>	Biodiversity Collections Project. A central index of biodiversity collections around the world, based on NCD. <a href="http://www.biodiversitycollectionsindex.org/">http://www.biodiversitycollectionsindex.org/</a>
<b>BioCASE</b>	Biological Collection Access Service for Europe. A multi-national specimen information network for Europe. <a href="http://www.biocase.org/">http://www.biocase.org/</a>
<b>CorM</b>	Core Metadatabase. Central database used for harvesting collection descriptions in the BioCASE project. <a href="http://www.biocase.org/whats_biocase/meta_net_old.shtml">http://www.biocase.org/whats_biocase/meta_net_old.shtml</a>
<b>CODENS</b>	Institutional acronyms, abbreviations or other codes. See also <a href="http://circa.gbif.net/irc/Download/kleYAJJ_moGCtjTxGtCbK1qGh-4pYxTs/F-hH-dlxQfm2jlxFJFqGyi2s2wP/codenHowTo-v0.4.1-draft.html">http://circa.gbif.net/irc/Download/kleYAJJ_moGCtjTxGtCbK1qGh-4pYxTs/F-hH-dlxQfm2jlxFJFqGyi2s2wP/codenHowTo-v0.4.1-draft.html</a>
<b>DC</b>	Dublin Core. Metadata element set that is a standard for cross-domain information resource discovery. <a href="http://dublincore.org/documents/1999/07/02/dces/">http://dublincore.org/documents/1999/07/02/dces/</a>
<b>DCMI</b>	Dublin Core Metadata Initiative. The organization engaged in developing Dublin Core metadata standard. <a href="http://dublincore.org/">http://dublincore.org/</a>
<b>EAD</b>	Encoded Archival Description. The rich standard for encoding archival finding aids using XML. <a href="http://www.loc.gov/ead/">http://www.loc.gov/ead/</a>
<b>EDIT</b>	European Distributed Institute of Taxonomy. Consortium to integrate taxonomic research. <a href="http://www.e-taxonomy.eu/">http://www.e-taxonomy.eu/</a>
<b>GBIF</b>	Global Biodiversity Information Facility. Interoperable network of biodiversity databases and information technology tools. <a href="http://www.gbif.org/">http://www.gbif.org/</a>
<b>JSON</b>	JavaScript Object Notation. Lightweight data-interchange format. <a href="http://www.json.org/">http://www.json.org/</a>
<b>NCD</b>	Natural Collections Description is a data standard for describing collections. <a href="http://www.tdwg.org/activities/ncd/">http://www.tdwg.org/activities/ncd/</a>
<b>NoDIT</b>	National Node Data Input Tool. MS Access database used by the BioCASE National Nodes to record collection description data.
<b>OCLC</b>	Online Computer Library Center (previous name). A non-profit, membership, computer library service and research organization dedicated to the public purposes of furthering access to the world's information and reducing information costs. <a href="http://www.oclc.org/">http://www.oclc.org/</a>

<b>PLANETS</b>	Preservation and Long-term Access through Networked Services. Addressing core digital preservation issues. <a href="http://www.planets-project.eu/">http://www.planets-project.eu/</a>
<b>PLATO</b>	Preservation planning tool for digital objects, developed under the PLANETS project. <a href="http://olymp.ifs.tuwien.ac.at:8080/plato/website/intro.html">http://olymp.ifs.tuwien.ac.at:8080/plato/website/intro.html</a>
<b>RAVNS</b>	Resources Available in Natural Sciences. Group that was formed through RLG Programs to work on NCD.
<b>RDF</b>	Resource Description Framework. Lightweight ontology system to support knowledge exchange online. <a href="http://en.wikipedia.org/wiki/Resource_Description_Framework">http://en.wikipedia.org/wiki/Resource_Description_Framework</a>
<b>RLG</b>	RLG Programs. Formerly the Research Libraries Group, now part of OCLC. <a href="http://www.oclc.org/programs/about/default.htm">http://www.oclc.org/programs/about/default.htm</a>
<b>SYNTHESYS</b>	Synthesis of Systematics Resources. Large scale facilities EU Framework VI project. <a href="http://www.synthesys.info/">http://www.synthesys.info/</a>
<b>TDWG</b>	Taxonomic Databases Working Group. Now known as the Biodiversity Information Standards (TDWG) group that develops standards and protocols for sharing biodiversity data. <a href="http://www.tdwg.org/">http://www.tdwg.org/</a>
<b>TGN</b>	Thesaurus of Geographic Names. Developed by the Getty Institution, providing identifiers for place names. <a href="http://www.getty.edu/research/conducting_research/vocabularies/tgn/">http://www.getty.edu/research/conducting_research/vocabularies/tgn/</a>
<b>URI</b>	Unique Resource Identifier. Generic term for linking web resources, includes URLs. <a href="http://en.wikipedia.org/wiki/Uniform_Resource_Identifier">http://en.wikipedia.org/wiki/Uniform_Resource_Identifier</a>
<b>vCard</b>	File format standard for electronic business cards. <a href="http://www.w3.org/TR/vcard-rdf">http://www.w3.org/TR/vcard-rdf</a>
<b>XML</b>	Extensible Markup Language. A simple flexible text format playing an increasingly important role in the exchange of a wide variety of data on the Web. <a href="http://www.w3.org/XML/">http://www.w3.org/XML/</a>