

# Natural Collections Descriptions (NCD) Workshop 18 – 19 June 2007 ETI, Amsterdam

## 1. Summary

The third NCD Workshop was attended by 10 delegates (see Appendix) and funded primarily by a grant from the Gordon & Betty Moore Foundation via GBIF, the Global Biodiversity Information Facility.

In addition to the workshop, GBIF have funded a series of NCD deliverables. Details of these may be found on the NCD wiki at <http://wiki.tdwg.org/twiki/bin/view/NCD/NCDListOfDeliverables> but in brief, these are:

- NCD Toolkit
- RDF version of NCD
- Terms for use with pick-lists
- Documentation set
- Institution descriptions

The workshop itself has a set of deliverables, of which this report is one.

The original agenda for the workshop split the two days into reporting on progress made to date and looking to the future, once the development of the NCD Toolkit has been completed. However, it was agreed that it was essential to produce a stable version of NCD on which all dependent work, such as the toolkit, mappings and documentation, could be based. The agenda was modified accordingly.

Day one started with a presentation from Roger Hyam about the technical architecture that TDWG is creating, into which NCD will fit. This was followed by a presentation by Markus Döring about the Resource Description Framework (RDF), the relationship between RDF and XML Schema and how NCD needs to change.

Some changes to NCD had been made in the week prior to the workshop, resulting in version v0.60. Changes included the move to use the vCard standard for persons and institutions, a simplification of the header and more extensive use of attributes.

Since it was clear that further changes were desirable, the decision was made to pin down a stable version of NCD at the workshop so that other work could proceed with confidence that no more changes would be made. Since it is undesirable to maintain both an XML Schema version and an RDF version, a further decision was agreed that the RDF version will be the normative form of NCD and, if necessary, an XML version could be derived from it.

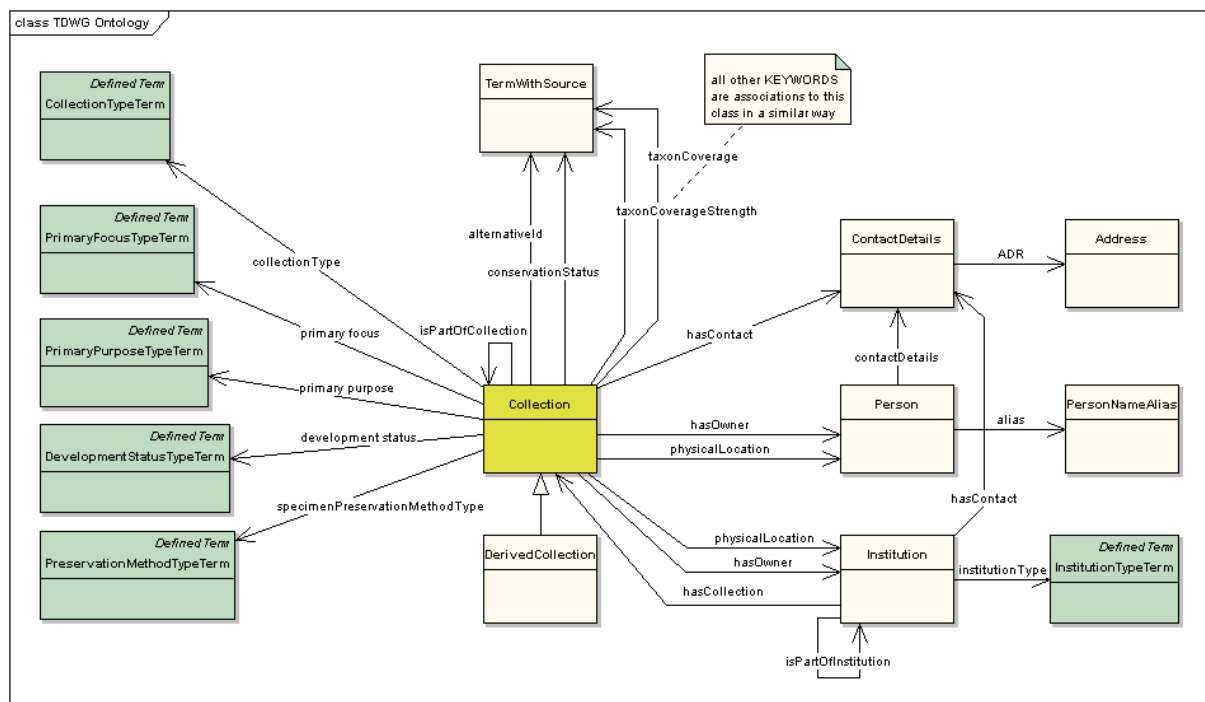
One of the major differences is that cardinality is not specified in the RDF version – it is up to applications to determine what should be mandatory or repeatable, so one reason for deriving an XML version is to give guidance on these aspects, to aid interoperability.

With this as background, day one continued with reports on the GBIF contract deliverables, with much helpful discussion on each. All are on track for the September deadline. It was noted that the contract for the NCD Toolkit was for a record management system and does not include a Web-based front-end for searching - the options for addressing this need to be discussed.

The User Guide will be derived from the TDWG Ontology, where the element definitions will be developed. In addition, a primer will be produced as an introduction to NCD and to collection descriptions, for those new to the concept.

Day two saw the emergence of NCD version v0.7. This is sufficient for work on the toolkit to continue, but lacks the final lists of terms for those elements that will have pick-lists to aid consistency. Version v0.8, which will be in RDF and synchronised with the TDWG ontology, will be made available within a few days of the workshop.

The NCD class diagram is shown here:



Further discussion included the most practicable toolkit implementation model (single central model or GBIF National Node model or purely local model) and extending compatibility with external standards through the incorporation of the vCard and Dublin Core namespaces and the use of identifiers.

The day finished with an events timeline culminating in NCD v0.9, which is intended to be the version of NCD (with supporting documentation) that will be submitted to the TDWG approvals process in the late autumn of 2007.

For more details, including presentations, see the NCD wiki at: <http://wiki.tdwg.org/twiki/bin/view/NCD/WebHome> and the NCD Website at: <http://www.tdwg.org/activities/ncd/>.

NCD files may be found at <http://rs.tdwg.org/ncd/> and the RDF version may be found at: <http://rs.tdwg.org/ontology/voc/Collection.rdf> (Note: Use **View** | **Source** to see the actual RDF)

Please contact Neil Thomson (n.thomson AT nhm.ac.uk) or Markus Döring (m.doering AT BGBM.org) for any other information.

## 2. Goals

- To have an understanding of the TDWG technical environment and how NCD needs to change to mesh with it.
- To review progress, problems and expectations for the work being carried out under the GBIF contracts.
- To agree on the normative version of NCD that will be submitted to the TDWG approvals process.
- To discuss the options for deployment of the NCD Toolkit.

## 3. Outcomes

- The main outcome of the workshop is a version of NCD that is much more tightly integrated with the TDWG technical environment. A sufficiently stable version has been produced to enable other work to proceed and a timeline has been agreed that results in the submission of the NCD standard and supporting documentation to TDWG in the late autumn.
- NCD is now also more compatible with existing data standards through the adoption of vCard for person and institution metadata and the direct use of Dublin Core elements and terms.
- Significantly, agreement was reached to adopt UUIDs (Universally Unique Identifiers) for collections, institutions and persons. UUIDs are TDWG-safe and have the advantage that they can be generated without reliance on a central authority. In the toolkit there will be a single URL for every record. An application URL plus the UUID becomes the identifier for the record as in the “rdf:about ...” statement, thus replacing the old NCDID element.
- The toolkit will allow the use of two languages, the local language and English. Whilst there are distinct advantages in creating a centralised system, the issues raised by multiple languages are significant.
- The most likely implementation scenario is for the toolkit to be offered to the GBIF National Nodes, with an implementation in Berlin for those that do not, or cannot, host their own data. Larry Speers had agreed that GBIF would create a central index for NCD records and this will be raised with his replacement at GBIF at an opportune time.
- If a move to a central system looks desirable this could be achieved by aggregating records from different implementations through the export / import of records in the NCD standard. Further discussion will be held on testing the transfer of records between institutions.

## 4. Conclusions

The 2007 NCD Workshop was very productive, with valuable contributions from each of the 10 delegates who represented museums, libraries, archives and biodiversity informatics in a unique mix of skills. Thanks are due to the Gordon & Betty Moore Foundation for their generous sponsorship and to ETI for hosting the workshop and for the local arrangements.

The enthusiasm and dedication of the delegates has moved NCD from concept to reality. As a result, by the end of the year, the NCD Toolkit will be available for use and NCD itself complete with documentation, mappings and term-lists will be starting the TDWG approvals process with the confidence that it will interoperate with the emerging TDWG technical architecture.

## Appendix A: Attendees

Name	Email	Institution
Wouter Addink	wouter@eti.uva.nl	ETI, Netherlands
Ruud Altenburg	ruud@eti.uva.nl	ETI, Netherlands
Carol Butler	ButlerCR@si.edu	Smithsonian Institution, USA
Markus Döring	m.doering@BGBM.org	Berlin Botanic Gardens, Germany
Doug Holland	doug.holland@mobot.org	Missouri Botanic Gardens, USA
Roger Hyam	roger@tdwg.org	TDWG, UK
Barbara Mathé	bmathe@amnh.org	American Museum of Natural History, USA
Constance Rinaldo	crinaldo@oeb.harvard.edu	Harvard, USA
Neil Thomson	n.thomson@nhm.ac.uk	Natural History Museum, UK
Günter Waibel	waibelg@oclc.org	RLG Programs Group, USA