

DRAFT ABCD Group Meeting Report

Access to Biological Collection Data – Comprehensive Content Definition (TDWG ABCD standard)

26th – 28th July 2006

Royal Museum For Central Africa, Tervuren, BELGIUM

1. Background

The TDWG meeting in St. Petersburg ratified the ABCD Schema v. 2.06 as a TDWG standard. In parallel, the TDWG Infrastructure Project (TIP) started its work with the aim of turning TDWG into an effective organisation for biodiversity informatics and data standards. ABCD is one of the few standards in current use and is thus evolving. Several bugfixes have to be made to the current version. Also, the group has to discuss how to fit most adequately into the new structure proposed by the TIP.

2. Summary

The Access to Biological Collection Data subgroup (ABCD, <http://www.bgbm.org/tdwg/codata/schema/>) of the Taxonomical Database Working Group (TDWG, <http://www.tdwg.org>) held a meeting to review the current state of affairs of the ABCD standard, to analyze the interim ABCD version, to update the ABCD documentation to the TDWG requested standard formats and to identify the priorities to move towards the next major release of ABCD.

The meeting confirmed the important role of this task or interest group in the content definition process for biological collections. The group has clearly stated its purpose, scope and audience in a charter document (see annex). A number of corrections and additions to the schema were discussed and will be made accessible through the group's wiki site. The new version will be called v. 1.06a – it was decided to follow the example of GML and proceed at least for some time with an incremental approach of versions using the same namespace and preserving full backwards compatibility.

3. Goals

The goals of the present meeting were to discuss and agree on the following points:

1. State of affairs
 - a. Reports on current implementations and implementations plans
 - b. Comments from users upgrading from ABCD 1.2 to ABCD 2.06
 - c. Potential purposes of ABCD (access, exchange, repository)
2. Interim ABCD version (bug fixes and necessary modifications)
 - a. Type of modifications and their impact on current usage
 - b. Documentation and follow up of the changes
 - c. Walk through the schema and specification
 - d. GUIDS in ABCD 2.0x
3. Documentation
 - a. Define categories according to TDWG standard documentation types
 - b. Charter, Primer, Introduction and other available documentation on ABCD

- c. Review of currently available documentation and define further tasks
- 4. Toward ABCD 3
 - a. Identification of potential modules for ABCD sub-schemas
 - b. Imported modules
 - c. Modules to be dealt with separately
 - d. ABCD Core business
- 5. Further matters
 - a. TDWG Infrastructure Project (TIP) support

4. Discussion

The meeting identified the role of the task group as a content definition group.

Discussions concerned the following points:

- How to handle bug fixing and minor modifications of ABCD, without breaking the applications using the current approved version. New namespaces are needed to keep the XML schema valid; however the usage of new namespaces may break the applications. If making only additions and keeping the semantics the same, applications should not break.
- ABCD is currently mainly used for access. It has been discussed if ABCD could also be used for data exchange between different systems, for data repository structure or as a data dumping format for digital sustainability. The variable atomization choice (atomized structure or text) for a same content may cause issues if wanting to use ABCD for imports at global level. At local level for groups using routinely the same application with a long term stable data structure imports and exports in ABCD for data exchange is considered feasible.
- The metadata at dataset level need to be adapted to the needs of the UDDI registry of GBIF. The new TAPIR protocol will enable to provide the needed information independently from the Data Standard. Meanwhile several options are discussed
 - Adapt <ContentContact>, <TechnicalContact>, <AdministrativeContact> in the metadata at Dataset level to UDDI needs
 - Get GBIF use a form for UDDI, where the needed information can be filled in when registering to GBIF
 - Keep for the time being the (undesirable) workaround to use the metadata from the first dataset and ensure that this first dataset has all the needed and correct data.
- Best practice recommendations need to be in the primer or ABCD Introductory Documentations like:
 - Atomized *versus* text element for a same content
 - Mandatory elements should not be left empty
 - Highlight override rules between dataset and unit level (ie IPR statements)
 - Clarify terminology used for the fundamental building blocks used in ABCD like Unit, specimen, observation...
 - Clarify the use of Assemblages, Association in the Curational part of ABCD
 - How to handle Information withheld practices
- The Gathering part of ABCD was analyzed and synergies with DwC geospatial extensions and the GML community were discussed. Additional information is needed on how far the geospatial extension of DwC is ready, on how it could be included in ABCD and the effects on versioning and applications.
- In the identification part, the handling of multiple preferred identifications of mixed content units must be clarified. The handling of identification history in parallel to this may be confusing. An additional problem is that some providers bring in not only (the intended) multiple results from identification events but also synonyms of the name identified.

- The potential modularization of ABCD has been discussed and the effects on versioning, namespaces and validation.
- In the TIP support expectations two points were mainly discussed:
 - Which kind of printed documents for distribution may be needed (book, booklet, leaflet?)
 - Is it worthwhile to have an UML version of ABCD to hook up with what is done at higher level in the TDWG TAG group with ontologies?

5. Outcomes

1. State of affairs

a. List of known implementations and implementation plans:

- BioCASE
- GBIF Index now supporting ABCD v. 1.2, 1.49, and 2.06
- Consultative Group on International Agricultural Research (CGIAR, <http://www.cgiar.org>) and Generation Challenge Programme (GCP, <http://www.generationcp.org>) have 6 Providers with 15 databases, using ABCD 1.2 and MCPD (Multi Crop Passport Data) with direct mapping to ABCD 2.06. They are waiting for TAPIR before further upgrades
- Royal Botanic Gardens of Melbourne (Peter Neish) are looking at HISPID/ABCD relations
- Naturalis, Leiden are looking for ABCD in RDF (Steven van der Mije)
- Morphbank, Florida State University (Debbie Paul)

b. Purposes of ABCD (access, exchange, repository)

- Data dump for permanent storage
- Data exchange (export/import), including for version upgrades
- RDF representation as LSID metadata

2. Interim ABCD version (bug fixes and necessary modifications)

a. List of possible types of changes

- adding types to existing elements
- adding elements
- adding attributes
- making elements unbound
- adding an unbound substructure

b. Impact assessment on schema clients

- GBIF Indexer classes (also used by CRIA data cleaning tool)
- Stylesheets
- Barcelona analysis tool
- Parsers used by CGIAR
- XML export programs (BGBASE at Meise Botanical Garden, Laboratory and collection management system of several Microorganisms collections communities)
- ABCD concepts referenced in other programs or documentations.

c. Walk-through of the schema and specification

- Comments and suggestions for changes received since the last revision of ABCD have been analyzed and where judged pertinent implemented.
- The effected changes are documented on the ABCD WIKI under Editorial process at <http://ww3.bgbm.org/abcddocs/SchemaChanges>

d. GUIDs in ABCD 2.06

- GUID elements have been inserted in the schema where references are made to the 3 IDs (SourceInstitutionID, SourceID and UnitID), namely under Associations and SpecimenHistory.
- The GUIDs are referring to the underlying object (the specimen or original observation unit), not to data record.

3. Documentation

- a. Define categories according to TDWG standard documentation types
 - The schema itself without annotations is the Type 1 documentation
 - The documentation of concepts (currently on the ABCD wiki) is Type 2 documentation
 - The primer is considered Type 3 documentation.
- b. Charter, Primer, Introduction and other available documentation on ABCD
 - A charter document has been written and approved by the members present. It defines the name, purpose, scope and target audience and strategy of this group (see document in annex)
 - A primer document has been written, is open for comments and is available on the ABCD wiki at <http://ww3.bgbm.org/abcd/docs/AbcdPrimer>
- c. Review of currently available documentation and define further tasks
 - The time line for future tasks is in the Charter document

4. Toward ABCD 3

- a. Identification of potential modules for ABCD sub-schemas
 - These modules are envisioned as independent sub-schemas as a short-term solution, but long term implementation is left open
- b. Imported modules
 - These modules have their own namespace and result in a new version of the schema.
- c. Modules to be dealt with separately
 - Are considered as doable in a relatively short timeframe:
 - Metadata sections (including IPR, other rights, ownership ...)
 - Names: Agree with TCS and SDD to have a common module for names as a separate structure.
 - Collections and subcollections referenced to Natural Collections Descriptions (NCD Standard under Development), Target March 2007
 - Are considered as desirable
 - Geospatial module, but should include further applications outside the collections community (e.g. species distributions)
 - Gazetteer subgroup
 - Addresses/Agents
 - Citations (Need: Bibliographic Titles including Articles -- BHL may provide a solution)
 - Multimedia objects
 - Observation projects
- d. ABCD Core business
 - Natural history collection curatorial issues
 - Hierarchical / multiple object problems therein (e.g. several specimens in a single collection unit)
 - Collection- and observation event issues (excluding geographic "where")
 - Identification

5. TDWG Infrastructure Project (TIP) support sought

- Create a dedicated mailing list
- Putting together the content documentation (examples, how-to, Q&A, suggested vocabulary). Can be derived from provided data.
- Printed primer for outreach, a book or booklet is important because it can be seen as a publication and kept in libraries as opposed to leaflets.
- UML model of ABCD, idea left open, alternatively have UML model of collections based on relational information models and incorporate the progress made there (ie in the geographical area).

Recommendations

For interim versioning and versioning in XML schema standards:

- Pay attention on any changes that may break applications (recommendation from ABCD group to TDWG TAG, that all apply this as best practice)
- Following the GML example and implement smaller changes under the same namespace indicating the version by a Version attribute in the root.
- To ensure backward compatibility no elements are deleted
- Elements may be marked as deprecated
- No elements are renamed
- The semantics of elements is left unchanged
- Type changes are restricted to assigning types to previously untyped elements
- If no namespace change is required, generalize the ABCD procedure: ABCD 2.06 → 2.06a.

Actions

- Further actions as derived form the charter document:

| Output/outcome | Version | Timeframe |
|---|---------|--|
| ABCD Content Definition Group Charter | 1.0 | Complete: July 28, 2006 |
| ABCD Primer (Type 3) | 1.0 | Complete: September 15, 2006 |
| ABCD Schema (Type 1, normative part plus cover page) | 2.06a | Ready for reference implementation: August 15, 2006 |
| ABCD | 3.0 | To be determined, part of TDWG new biodiversity infrastructure |
| Contribute to TDWG Ontology group | | During TDWG 2006 Meeting and thereafter |
| Update the ABCD Concept documentation (Schema Viewer and WIKI) | | Complete: September 15, 2006 |
| Transfer ABCD schema to TDWG website and create RDDDL page | | Dependent upon TDWG Infrastructure Project. Estimated for September, 2006. |
| Mapping between ABCD and DwC 1.4 core, geospatial and curatorial extensions | | Complete: August 15, 2006 |

- Strategy for further actions as derived from the charter document:

| Output/outcome | Version | Strategy |
|--|---------|---|
| ABCD Content Definition Group Charter | 1.0 | Publish workshop result. |
| ABCD Primer (Type 3) | 1.0 | Develop draft at workshop, review and revise, distribute draft to group, finalize. |
| ABCD Schema (Type 1, normative part plus cover page) | 2.06a | Edit results of workshop and check for consistency, finalize schema. Procedures for ratification of new minor version are under revision by TDWG. |
| ABCD | 3.0 | Participate in discussions of modularized TDWG data infrastructure. Make plans accordingly and adapt ABCD to that structure. |
| Contribute to TDWG Ontology group | | Review ontology definitions with regard to collections data and provide input |
| Update the ABCD Concept documentation (Schema Viewer and WIKI) | | Check the possibility for making a direct link between Schema Viewer and WIKI concept pages and implement if possible. |
| ABCD schema documentation on TDWG website and create RDDDL page | | Wait for transfer to be effected by TDWG Infrastructure project. Learn to use TDWG Infrastructure content management system. |
| Mapping between ABCD and DwC 1.4 core, geospatial and curatorial extension | | Use table provided by DwC group and insert ABCD schema paths and insert comments as necessary. Publish on ABCD website wherever it is. |

6. Conclusions

Important milestones were reached during this meeting:

- Agreeing on a best practice scheme for versioning
- Preparation and approval of a Charter document for the ABCD Content Definition Group
- Preparation of a primer document open for comments
- Agreeing on a timeline and strategy for further tasks to be achieved.

Annexes

- Charter document
- Links
- List of attendees

ABCD CONTENT DEFINITION GROUP CHARTER

1. The NAME of the (Interest or Task) Group
ABCD Content Definition Group – Access to Biological Collections Data

2. VERSION HISTORY of the Charter document
Version 1.0, July 27, 2006, ABCD Workshop, Tervuren, Belgium

3. CONVENER

Walter G. Berendsohn

Botanic Garden and Botanical Museum Berlin-Dahlem
Freie Universität Berlin
Königin-Luise-Straße 6-8,
D-14191 Berlin, Germany.
Phone: (+4930) 838-50143
Fax (+4930) 841729-43
Email: w.berendsohn [at] bgbm.org.

4. CORE MEMBERS

Adrian Rissoné

The Natural History Museum, London
a.rissone [at] nhm.ac.uk

Alex R. Chapman

WA Herbarium - Department of
Conservation and Land Management
Locked Bag 104 Bentley Delivery Centre
Western Australia 6983
alexc[at] calm.wa.gov.au

Anna Weitzman

National Museum of Natural History
Smithsonian Institution
Washington
weitzman.anna [at] nmnh.si.edu

Anton Güntsch

Dept. of Biodiversity Informatics and
Laboratories,
Botanic Garden and Botanical Museum
Berlin-Dahlem
Königin-Luise-Str. 6-8
14191 Berlin, Germany
Phone: +49 30 838 50 166
a.guentsch[at] bgbm.org

Junko Shimura

junko [at] nies.go.jp

Markus Döring

Dept. of Biodiversity Informatics
and Laboratories,
Botanic Garden and Botanical
Museum Berlin-Dahlem
Königin-Luise-Str. 6-8
14191 Berlin, Germany
m.doering[at] bgbm.org

Mickaël Graf

GBIF Sweden
Swedish Museum of Natural History
mickael.graf[at]nrm.se

Neil Thomson

Department of Library and
Information Services
The Natural History Museum
Cromwell Road
London SW7 5BD, United Kingdom
Phone: +44 (0)207 942 5294
n.thomson[at] nhm.ac.uk

Ben Richardson

WA Herbarium - Department of
Conservation and Land Management
Locked Bag 104 Bentley Delivery Centre
Western Australia 6983
benr[at] calm.wa.gov.au

Charles Copp

The Natural History Museum, London
Cromwell Road
London SW7 5BD, United Kingdom
eim [at] globalnet.co.uk

Charles Hussey

Department of Library and Information
Services
The Natural History Museum
Cromwell Road
London SW7 5BD, United Kingdom
Phone: +44 (0) 207 942 5213
c.hussey[at] nhm.ac.uk

Chuck Miller

Missouri Botanical Garden
4344 Shaw Blvd
Saint Louis, Missouri 63110
Chuck.Miller[at]mobot.org

Denis Lepage

dlepage[at]bsc-eoc.org

Donald Hobern

Global Biodiversity Information Facility
Secretariat
Universitetsparken 15, DK-2100
Copenhagen, Denmark
dhobern[at] gbif.org

George Garrity

Bergey's Manual Trust
Department of Microbiology and
Molecular Genetics
152 Giltner Hall, Michigan State
University
East Lansing, MI 48824-1101
garrity[at] pilot.msu.edu

Patricia Mergen

Royal Museum for Central Africa
Department of Zoology
Leuvensesteenweg 13
3080 Tervuren, Belgium
Phone: +32 2 769 5626
patricia.mergen [at] africanmuseum.be

Raul Jimenez Rosenberg

CONABIO
Av. Liga Periférico-Insurgentes Sur
4903
Col. Parques del Pedregal
Delegación Tlalpan
14010 México, Distrito Federal,
Mexico
rjimenez[at] xolo.conabio.gob.mx

Renato De Giovanni

CRIA
Av. Dr. Romeu Tórtima, 388, Barão
Geraldo
13084-520 Campinas SP Brasil
renato [at] cria dot org dot br

Rudolf May

Federal Agency for Nature
Conservation
Konstantinstr. 110
53179 Bonn, Germany
Phone: +49 228 8491-260
mayr[at] bfn.de

Sabine Roscher, ZADI, Bonn

ZADI - Zentralstelle für
Agrardokumentation und -
information
Villichgasse 17
53177 Bonn, Germany
+49 (228) -9548212
roscher[at] zadi.de

Stanley Blum

California Academy of Sciences
Golden Gate Park
San Francisco, CA 94118, USA
Phone: +1 (415) 750-7032
sblum[at] calacademy.org

Greg Whitbread

Australian National Botanic Gardens
GPO Box 1777
Canberra 2601
Phone: +61 2 62509 482
ghw [at] anbg.gov.au

Gregor Hagedorn

G.Hagedorn [at] bba.de

Hideaki Sugawara

Center for Information Biology
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540,
Japan
Phone: +81-(0)559-81-6895
hsugawar[at] genes.nig.ac.jp

Javier de la Torre

Museo Nacional de Ciencias Naturales -
CSIC
C/José Gutiérrez Abascal, 2. 28006
Madrid, Spain
jatorre [at] gmail.com

Jerry Cooper

Landcare Research
PO Box 40, Lincoln 8152
New Zealand
CooperJ[at]LandcareResearch.co.nz

Steve Wilkinson

steve.wilkinson [at] jncc.gov.uk

Walter G. Berendsohn

Botanic Garden and Botanical
Museum Berlin-Dahlem
Königin-Luise-Str. 6-8
14191 Berlin, Germany
Phone: +49 30 838 50 143
wgb[at] zedat.fu-berlin.de

William Ulate Rodríguez

Instituto Nacional de Biodiversidad,
INBio
Apdo. Postal 22-3100 Santo
Domingo, Heredia Costa Rica
Phone:(506) 507-8179
E-mail: wulate[at] inbio.ac.cr

Wouter Addink

ETI BioInformatics
University of Amsterdam
Mauritskade 61, 1092 AD
Amsterdam, The Netherlands
waddink[at]eti.uva.nl

Yde de Jong

Zoological Museum Amsterdam
Mauritskade 61, 1092 AD
Amsterdam, The Netherlands
yjong[at] science.uva.nl

5. HOME ADDRESS

www.bgbm.org/tdwg/codata/schema/

6. PURPOSE

The purpose of the ABCD Content Definition Group is to foster accessibility of existing and emerging biological collection data banks at the international level by developing proposals for data and metadata standards. The group's long-term objectives:

- Foment standardization of the terminology used to model biological collection information.
- Collect and make public documents providing standards used in (or of potential use for) biological collections.
- Provide a general format for data exchange and retrieval for biological collections.

7. BACKGROUND

Biological collections exist in different scientific sub-disciplines: zoological, botanical, and palaeontological natural history collections, living collections like botanical and zoological gardens and microbial strain and tissue collections, and data collections stemming from surveys of objects in the field (like floristic and faunistic mapping, inventories). Research conducted over the past decade has revealed that all these collections have most of their attributes in common, although the terminology used to describe them may differ substantially.

Biological collections represent an immense knowledge base on global biodiversity. Field and research notes contain detailed data on the locality, time, and often appearance of organisms; the collected object itself can be a physical resource for research and industry. The preserved object also presents a falsifiable source of information, i.e. it can be re-observed to verify a scientific hypothesis based on it. Between 2 and 3 billion objects exist in natural history collections alone. Currently, this knowledge base is largely under-utilized, because its highly distributed, heterogeneous, and complex scientific nature obstructs efficient information retrieval.

Databasing and networking is now seen as the key to employ the potential value of biological collections for science, government, education, the public, and businesses, operating in the environmental sector, in biotechnology, or in biodiversity research. Efforts to network the resources exist, but there is little transfer of technology and co-ordination on a global level. International collaboration on the standardization of information models and standard data used in collection databases can enhance the efficiency of this process.

8. SCOPE

What is in scope?

- Collections of any kind of biological object both physical and observations.
- Terminology associated with biological collection data
- Retaining compatibility with prior existing collection standards, where possible
- Relationships between components and attributes of biological collections data

What is not in scope?

- Data interchange protocols
- Non-biological collections, eg. minerals (but see ABCD – Extension For Geosciences)
- Taxonomic data not associated with identification of collection items
- Human biological data

9. AUDIENCE

- Biological collection holders
- Scientists or other practitioners needing collection data
- Developers of collections management systems, eg. Specify, KEmu
- Other TDWG task groups, eg. SDD, TCS, and users of their standards
- Protocol developers, eg. BioCASE, TAPIR
- Collection portal and network developers, eg. GBIF, GBIF Nodes, IABIN
- Archiving initiatives

10. OUTPUTS and OUTCOMES

| Output/outcome | Version | Timeframe |
|---|---------|--|
| ABCD Content Definition Group Charter | 1.0 | Complete: July 28, 2006 |
| ABCD Primer (Type 3) | 1.0 | Complete: September 15, 2006 |
| ABCD Schema (Type 1, normative part plus cover page) | 2.06a | Ready for reference implementation: August 15, 2006 |
| | 3.0 | To be determined, part of TDWG new biodiversity infrastructure |
| Contribute to TDWG Ontology group | | During TDWG 2006 Meeting and thereafter |
| Update the ABCD Concept documentation (Schema Viewer and WIKI) | | Complete: September 15, 2006 |
| Transfer ABCD schema to TDWG website and create RDDDL page | | Dependent upon TDWG Infrastructure Project. Estimated for September, 2006. |
| Mapping between ABCD and DwC 1.4 core, geospatial and curatorial extensions | | Complete: August 15, 2006 |

11. STRATEGY

| Output/outcome | Version | Strategy |
|--|---------|--|
| ABCD Content Definition Group Charter | 1.0 | Publish workshop result. |
| ABCD Primer (Type 3) | 1.0 | Develop draft at workshop, review and revise, distribute draft to group, finalize. |
| ABCD Schema (Type 1, normative part plus cover page) | 2.06a | Edit results of workshop and check for consistency, finalize schema. |
| Procedures for ratification of new minor version are under revision by TDWG. | | |
| | 3.0 | Participate in discussions of modularized TDWG data infrastructure. |
| Make plans accordingly and adapt ABCD to that structure. | | |
| Contribute to TDWG Ontology group | | Review ontology definitions with regard to collections data and provide input |
| Update the ABCD Concept documentation (Schema Viewer and WIKI) | | Check the possibility for making a direct link between Schema Viewer and WIKI concept pages and implement if possible. |
| ABCD schema documentation on TDWG website and create RDDDL page | | Wait for transfer to be effected by TDWG Infrastructure project. |
| Learn to use TDWG Infrastructure content management system. | | |
| Mapping between ABCD and DwC 1.4 core, geospatial and curatorial extension | | Use table provided by DwC group and insert ABCD schema paths and insert comments as necessary. Publish on ABCD website wherever it is. |

Links

<http://www.tdwg.org/>
<http://wiki.tdwg.org/>
<http://www.gbif.org>
<http://www.bgbm.org/TDWG/CODATA/Schema>
<http://ww3.bgbm.org/abddocs/>
<http://ww3.bgbm.org/abddocs/SchemaChanges>
<http://ww3.bgbm.org/abddocs/AbcdPrimer>
<http://www.africamuseum.be/>
<http://www.cgiar.org>
<http://www.generationcp.org>

Attendees

Wouter Addink (ETI, Amsterdam, NL)
Walter Berendsohn (BGBM, Berlin, DE)
Renato De Giovanni (CRIA, Campinas, BR)
Javier de la Torre (MNCN, Madrid, SP)
Markus Döring (BGBM, Berlin, DE)
Anton Güntsch (BGBM, Berlin, DE)
Andrea Hahn, (GBIF, Copenhagen, DK)
Michel Louette (RMCA, Tervuren, BE)
Bart Meganck (RMCA, Tervuren, BE)
Patricia Mergen (RMCA, Tervuren, BE)
Chuck Miller (MBG, St. Louis, USA)
Franck Theeten (RMCA, Tervuren, BE)
Neil Thomson (NHM, London, UK)
An Tombeur (RMCA, Tervuren, BE)
William Ulate (INBIO, San José, CR)