



DNA extension for ABCD - General information & installation manual

Another BioCASE mapping on all connected databases had to be made mandatory to offer DNA samples and its related specimen data on the central webportal (www.dnabank-network.org). Although the [ABCD 2.06 schema](#) (part of the BioCASE wrapper) is currently preferred its existing part for DNA 'Sequences' lacks important features. So, it shouldn't be used any longer. The whole ABCD 2.06 schema is available [here](#). If you want to learn more about the meaning of BioCASE and ABCD please follow the links.

ABCD offers two options to add supplementary contents: 'MeasurementsOrFacts' and 'UnitExtensions'. Since the hierarchical structure of DNA specific features is too complex for 'MeasurementsOrFacts' we decided to use 'UnitExtensions' to integrate an xml schema definition for DNA data similar to the [ABCDEFG extension](#) for geosciences. That new DNA extension for ABCD 2.06 is called ABCDDNA.

The DNA Sample means ABCD Unit and the identifier (Triple ID) for the related specimen is defined by 'UnitAssociation'. Please read the following manual for using the schema within the BioCASE provider software and have a look at a [mapping example](#).

Using ABCDDNA

The basic ABCD 2.06 version is remains unmodified. The new created DNA schema covers more than 35 elements such as "ExtractionDate", "ExtractionMethod" for DNA extraction as well as an "Amplification-Container" for Sequences, GenBankNumbers, CloneStrain etc.

-> View ABCDDNA schema: [html](#) or [xml](#).

-> View DNA part only: [html](#) or [xml](#).

The following requirements have to be met to use the schema:

1. Installation of BioCASE Provider Software (Download section).

The installation is documented [here](#).

Note: *If the software package DiGIR is already implemented on your specimen database it is possible to run both installations in parallel!*

2. Installation of ABCDDNA-Template

The current BioCASE Provider Software package does not include this template! It requires a cmf file which can be downloaded form the website of the DNA Bank Network – [Download section](#). Please copy it into the template folder of your BioCASE installation.

Version 2.4.2: `your_biocase2.4.2_folder/configuration/templates/cmf/`

Version 2.5.0: `your_biocase2.5.0_folder/templates/cmf/`

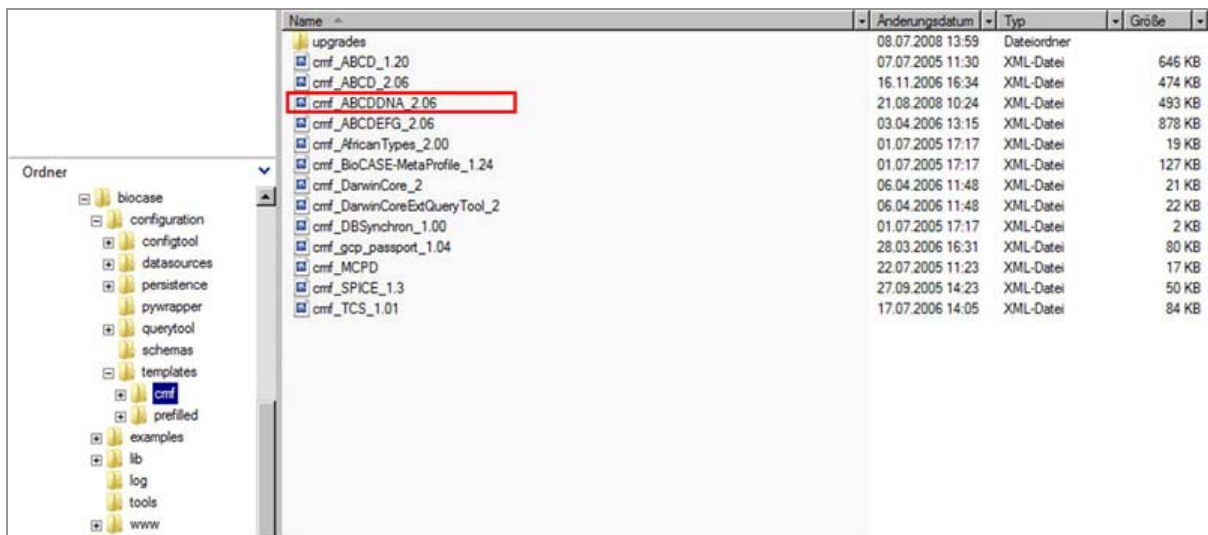


Fig.

4. Copy the cm file into your BioCASE cmf folder.

3. Creating a new datasource

If you want to use the ABCDDNA schema for mapping you have to create a new datasource connection. You can use the BioCASE [documentation](#) for assistance or contact Gabi Droege (g.droege[at]bgbm.org).

After you have created a new datasource you have to declare the database connection and the database structure.

Then you can select a schema (ABCDDNA) and click on “Create”.

4. Mapping the DNA database

The first time you open the mapping page only few fields are visible. Select “Show all concepts” and press “Refresh” to see all available features.

Since the DNA sample means ABCD Unit the Triple Identifier (composed of UnitID, SourceInstitutionID and SourceID) has to map as follows:

UnitID → DNA Bank Number (unique DNA Number in your database but NOT the ID in the DNA table)

SourceInstitutionID → Code or abbreviation of the institution the DNA bank is located, e.g. BGBM for Botanic Garden and Botanical Museum Berlin-Dahlem

SourceID → Designation of the collection, e.g. DNA bank

Note: Please don't map DNA data only. To verify the related specimen it is absolutely essential to map both the identification part (Taxon Name) and the UnitAssociation. It is furthermore advantageous to map few gathering attributes such as CountryName or ISO-Code.

UnitAssociation (Triple Identifier of the related specimen):

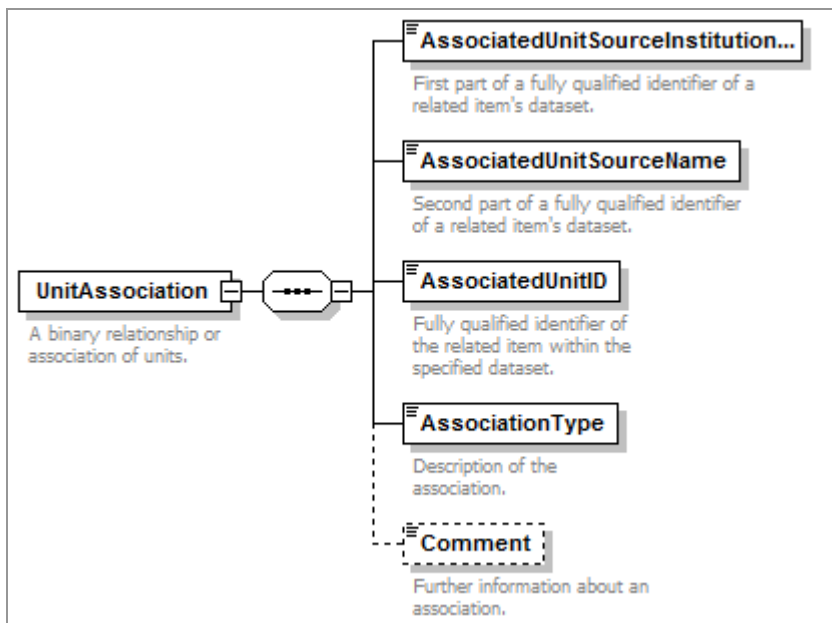


Fig. 5. UnitAssociation in ABCD schema. All attributes are required. It has to be mapped similar to the Triple Identifier of the DNA sample.

AssociatedUnitSourceInstitutionCode → Code or abbreviation of the institution the specimen/voucher of the DNA sample is deposited.

AssociatedUnitSourceName → Designation of the collection the specimen/voucher of the DNA sample is deposited.

AssociatedUnitID → Unique identifier/number of the specimen/voucher of the DNA sample within the specimen database.

Comment → WrapperURL of the specimen database.

These three listed attributes of the 'UnitAssociation' and the Triple Identifier used in the original specimen database should have the same values. They can then be used as GUIDs ([Globally Unique Identifier](#)) since both values describe the same specimen.

Example:

DNA sample taken from *Ballota itegrifolia* deposited in the 'Herbarium Berolinense' at the Botanic Garden and Botanical Museum Berlin-Dahlem with Barcode number 'B 10 0140204'.

AssociatedUnitSourceInstitutionCode →> BGBM

AssociatedUnitSourceName → Herbarium Berolinense

AssociatedUnitID → B 10 0140204

Comment → <http://ww3.bgbm.org/biocase/pywrapper.cgi?dsa=Herbar>

That Triple Identifier equals to the Triple Identifier used in the original specimen database ([verify with BioCASE](#)).

Note: It is essential to map 'AssociationType' describing the relation between a DNA sample and its voucher/specimen, e.g. 'DNA from specimen', 'DNA and specimen from the same population', 'DNA from cultivated offspring of the specimen' or the like.

Note: Please pay attention to special characters like μ , ä, ö, ü, ß! They have to be translated into Unicode. E.g. a common mistake is to complete the field Unit@Concentration with "ng/ μ l" when "ng/ µl" should be used instead.

It is no problem if special characters are used in your database. But applied in the mapping scheme these characters constantly produce "fatal python errors"! To fix such a fatal mapping mistake you have to close the browser window and open it again. So, you will lose your changes if they are not saved. There's no way to annul that error by going backwards in your browser!

Please have a look at a mapping example.

Don't forget to set the 'Root table alias' and the 'Static table alias'. The static table mostly equals the table for metadata and contains only one row.

Please save your mapping to finish the process!

5. Test Mapping

Please press 'Test Mapping!' to execute a capability test as well as a scan and search test for ABCD 2.06. If you get the message. You can start a test search if you receive the message "No errors found!" during all three tests. Therefore, click on "QueryForms" and select "ABCD2 search". The following code should then be visible:

```
<filter>
<like
path='/DataSets/DataSet/Units/Unit/Identifications/Identification/Result/
TaxonIdentified/ScientificName/
FullScientificNameString'>A*</like>
</filter>
```

Please press "Submit" (on the top of the page) to send the query to your database. You will hopefully see some results in xml. Otherwise you should check the search key (A* searches for taxa beginning with A.) or change the debugging level for more detailed error messages.

Mapping Example

The following boxws should lead you through important or required ABCDDNA attributes in mapping. If you have further questions please contact Gabi Droege (g.droege[at]bgbm.org). To see all available attributes select 'Show all concepts' and press 'Refresh'.

Note: *Please don't select 'Add new concepts'! Otherwise only the basic ABCD 2.06 version will appear but not the entire ABCDDNA schema.*

Fig. 6. Mapping with ABCDDNA - Metadata. The root and static table alias should be defined at first. The root table should include the UnitIDs which are unique 'DNA Bank Numbers' in that case. Static table alias applies for all data and is mostly defined as the Metadata table. It is very important to define contact persons as well as legal rights for using the DNA data. Finally, it is

impossible to see the data via wrapper if the Metadata has not been mapped.

BioCAsE Configuration Tool - dnabanktest1

Home | Overview | Settings | DB connection | DB structure | QueryTool | QueryForms | Logout | Help

Mapping - ABCDDNA_2.06.xml
http://www.tdwg.org/schemas/abcd/2.06

Generated	2008-06-19 14:39:18.15 by PyWrapper v2.0.0
Modified	2008-08-21 10:27:21.00
Record definition	/DataSets/DataSet/Units/Unit
Root table alias	dnabanknumbers
Static table aliases	metadata

Show all concepts:
Add new concepts:

DataSets/DataSet

- ContentContacts
 - ContentContact
 - Address
 - Mapping1: metadata.ContentContactAddress(text)
 - Email
 - Mapping1: metadata.ContentContactEmail(text)
 - Name
 - Mapping1: metadata.ContentContactName(text)
- Metadata
 - Description
 - Representation
 - language
 - Mapping1: "en"
 - Details
 - Mapping1: metadata.DatasetDetails(text)
 - Title
 - Mapping1: metadata.DatasetTitle(text)
 - IPRStatements
 - Acknowledgements
 - Acknowledgement
 - language
 - Mapping1: "en"
 - Text
 - Mapping1: metadata.AcknowledgementsText(text)
 - Citations
 - Citation
 - language
 - Mapping1: "en"
 - Text
 - Mapping1: metadata.CitationsText(text)
 - Copyrights
 - Copyright
 - language
 - Mapping1: "en"
 - Text
 - Mapping1: metadata.CopyrightText(text)
 - Disclaimers
 - Disclaimer
 - language
 - Mapping1: "en"
 - Text
 - Mapping1: metadata.DisclaimersText(text)
 - Licenses
 - License
 - language
 - Mapping1: "en"
 - Text
 - Mapping1: metadata.LicenseText(text)
 - TermsOfUseStatements
 - TermsOfUse
 - language
 - Mapping1: "en"
 - Text
 - Mapping1: metadata.TermsOfUseText(text)
 - Owners
 - Owner
 - Addresses
 - Address
 - Mapping1: metadata.OwnerAddress(text)
 - EmailAddresses
 - EmailAddress
 - Mapping1: metadata.OwnerEmail(text)
 - LogoURI
 - LogoURI
 - Mapping1: metadata.OwnerLogoURI(text)
 - Organisation
 - Name
 - Representation
 - language
 - Mapping1: "en"
 - Text
 - Mapping1: metadata.OwnerOrganizationName(text)
 - URIs
 - URL
 - Mapping1: metadata.OwnerURI(text)
 - RevisionData
 - DateModified
 - Mapping1: metadata.DateModified(date)
 - TechnicalContacts
 - TechnicalContact
 - Address
 - Mapping1: metadata.TechnicalContactAddress(text)
 - Email
 - Mapping1: metadata.TechnicalContactEmail(text)
 - Name
 - Mapping1: metadata.TechnicalContactName(text)

Related Specimen data:



Fig. 7. Mapping with ABCDDNA - UnitAssociation. The actual Unit data including all DNA data and some rudimental specimen information are listed following the Metadata. It is absolutely essential to map both the Triple Identifier for the related voucher/specimen (UnitAssociation) and the mandatory feature 'AssociationType'. All UnitAssociation data won't be visible otherwise. For further information please see above.

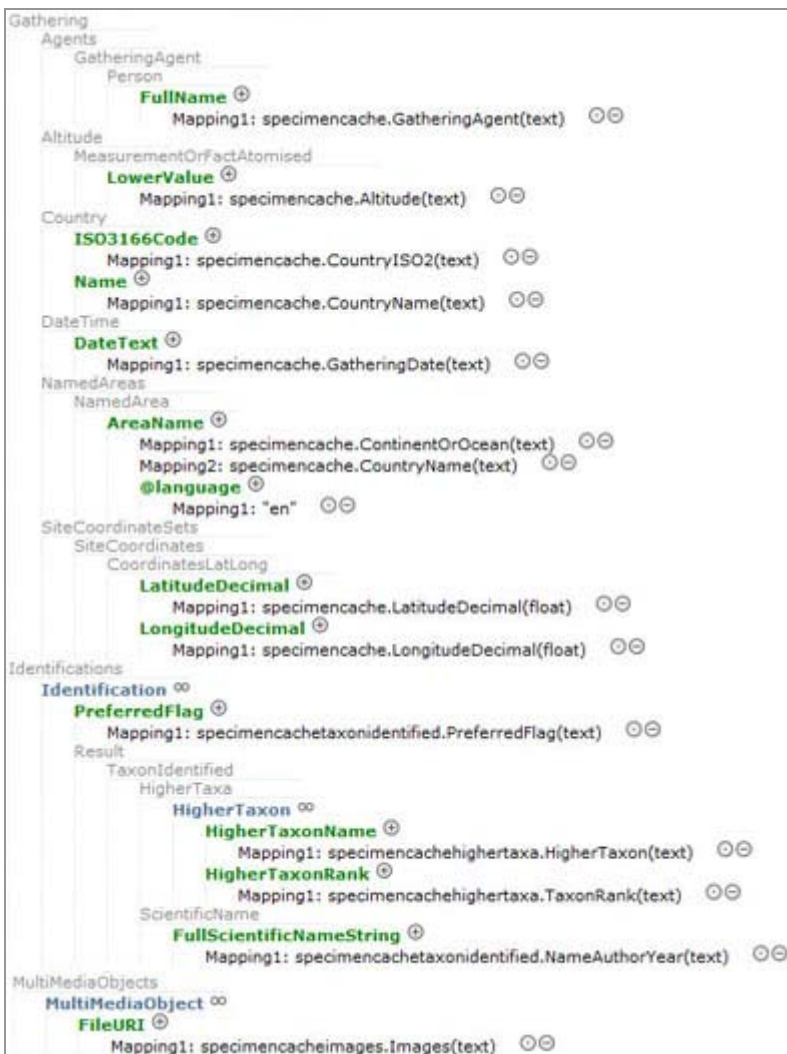


Fig. 8. Mapping with ABCDDNA - Gathering and Identification. The 'Gathering' and 'Identification' part has to be mapped that these data can be loaded into the web index of the DNA Bank Network's web portal. If such data, e.g. 'Taxon Name' or 'Country', are not in the web index searches for them won't be successful. For that matter it would be as well advantageous to map specimen images as 'MultiMediaObject'. Search results in the web portal can then be limited to samples only if a digital image to the related voucher/specimen of a DNA sample exists.

Note: There is another attribute for an image URL within the part 'DNA Sample'. Please pay attention to correctly map the image URL for the specimen images and for the DNA such as gel images by applying the appropriate ABCD attribute!

DNA data:

Some attributes of the basic ABCD 2.06 schema such as 'KindOfUnit', 'RecordBasis', 'UnitID', 'SourceID', 'SourceInstitutionID' and the Metadata (see Fig. 6) have to be mapped in the DNA context.

KindOfUnit ⊕	Mapping1: dnabanknumbers.Type(text) ⊕⊖
RecordBasis ⊕	Mapping1: metadata.RecordBasis(text) ⊕⊖
SourceID ⊕	Mapping1: metadata.SourceID(text) ⊕⊖
SourceInstitutionID ⊕	Mapping1: metadata.SourceInstitutionID(text) ⊕⊖
UnitID ⊕	Mapping1: dnabanknumbers.DNA_Bank_Number(text) ⊕⊖

Fig. 9. Mapping with ABCDDNA - Kind of Unit, Record Basis, SourceID, SourceInstitutionID, UnitID. 'Kind of Unit' indicates the type of DNA isolated, e.g. mitochondrial or genomic DNA. 'RecordBasis' has to be mapped as 'OtherSpecimen', because no other option for 'DNA sample' is provided in the ABCD 2.06 vocabulary ([details](#)).

Since ABCD Unit is synonymous to DNA sample as a certain sample type the UnitID has to be mapped as the (unique) 'DNA Bank Number'. Furthermore 'SourceID' could, for example, be mapped as 'DNA Bank' and 'SourceInstitutionID' as the institution holding the collection.

AmplificationDate ⊕	Mapping1: amplifications.Amplification_Date(date) ⊕⊖
GeneticLocus ⊕	Mapping1: amplifications.Fragment(text) ⊕⊖
GenBankNumber ⊕	Mapping1: amplifications.GenBankNumber(text) ⊕⊖
GenBankNumber-URI ⊕	Mapping1: amplifications.GenBankLink(text) ⊕⊖
BlockedUntil ⊕	Mapping1: dnabanknumbers.Block_Until(date) ⊕⊖
Concentration ⊕	Mapping1: dnabanknumbers.Concentration(float) ⊕⊖
@Unit ⊕	Mapping1: "ng/ul" ⊕⊖
ExtractionDate ⊕	Mapping1: dnabanknumbers.Extraction_Date(date) ⊕⊖
ExtractionMethod ⊕	Mapping1: dnabanknumbers.Extraction_Method(text) + "(" + dnabanknumbers.Extraction_Company(text) + ")" ⊕⊖
Preservation ⊕	Mapping1: dnabanknumbers.Preservation(text) ⊕⊖
ProvidedBy ⊕	Mapping1: dnabanknumbers.Provided_By(text) ⊕⊖
PurificationMethod ⊕	Mapping1: dnabanknumbers.Purification_Method(text) + "(" + dnabanknumbers.Purification_Company(text) + ")" ⊕⊖
Quality ⊕	Mapping1: dnabanknumbers.Quality(text) ⊕⊖
QualityCheckDate ⊕	Mapping1: dnabanknumbers.Quality_Check_Date(date) ⊕⊖
RatioOfAbsorbance260_230 ⊕	Mapping1: dnabanknumbers.Absorbance230(float) ⊕⊖
RatioOfAbsorbance260_280 ⊕	Mapping1: dnabanknumbers.Absorbance280(text) ⊕⊖
Tissue ⊕	Mapping1: dnabanknumbers.Tissue(text) ⊕⊖

Fig. 10. Mapping with ABCDDNA - Detail of the DNA extension. The lately created DNA part of the mapping schema covers more than 35 elements such as 'ExtractionDate', 'ExtractionMethod' for DNA extraction as well as an 'Amplification Container' for sequences, GenBank Numbers, clone strains etc. Note that 'Amplification Container' subsumes data of more than one amplification if necessary. Most of the other ABCDDNA attributes (DNA part) have a one-to-one relation to the UnitID that is every Unit (every UnitID) has, for example, only one 'ExtractionDate'.