



Project no. 018340

Project acronym: EDIT

Project title: Toward the European Distributed Institute of Taxonomy

Instrument: Network of Excellence

Thematic Priority: Sub-Priority 1.1.6.3: “Global Change and Ecosystems”

C5.87 WP6 exemplar group data migrated to CDM v1

Due date of component: Month 26

Actual submission date: Month 26

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this component: 11 RBGK, 9 FUB-BGBM, 10 NHML

Revision: final

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

C5.87 WP6 exemplar group data migrated to CDM v1

For enabling the EDIT WP6 exemplar groups to use their data within the upcoming EDIT Cybertaxonomy tools based on the Common Data Model (CDM) and the CDM Store the exemplar groups' data have been transformed from their original format into CDM directly or into a format that is an import format for the CDM.

The whole progress of transformation is described in detail at <http://dev.e-taxonomy.eu/trac/wiki/SampleDataConversion>

3 exemplar groups' data have been transformed:

1) Diptera: The original Diptera data exist in 2 formats: An MS Word file that includes all taxonomic data and a MS Access database containing all bibliographic information. The word files were linking to the bibliographic references where needed by integer keys. As usual the transformation was fairly complex as MS Word data are poorly structured and missing schema validation.

All the Diptera data were transformed from the original format into a Berlin Model database first, as this allowed to check and correct data at a time when the corresponding CDM functionality was not yet available. The database is located on a server at the BGBM and editable by means of the Berlin Model WebEditor (<http://ww2.bgbm.org/projects/WebEditor/Authent/login.cfm>).

From the Berlin Model the data can be imported directly into the CDM using the import/export functionality for Berlin Model data within the CDM library (C5.57). This is done for most of the data already; only descriptive data transformation from Berlin Model to CDM is still missing.

2) Cichorieae: The Cichorieae data are stored in a Berlin Model database so that they can use the CDM Library import/export functionality for the Berlin Model as well.

3) Monocotyledoneae: The Monocots are stored in a Sybase database and in some other file formats at RBKG. The data have been transformed into TCS/RDF which will be supported by the first version of CDM Library (C5.57). The transformation results can be found at <http://dev.e-taxonomy.eu/trac/wiki/SampleDataConversion/Monocotyledoneae>

The whole Sample Data Conversion is described at <http://dev.e-taxonomy.eu/trac/wiki/SampleDataConversion>.