



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"

## **M5.35 CDM refinement concluded and documented for Platform release 3**

Due date of milestone: Month 50  
Actual submission date: Month 51

Start date of project: 01/03/2006

Duration: 5 years

Organisation name of lead contractor for this milestone: 9 FUB-BGBM

Revision: final

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

## M5.35 CDM refinement concluded and documented for Platform release 3

In January 2008 and in March 2009 the first two versions of the Common Data Model (CDM) were frozen and documented at <http://wp5.e-taxonomy.eu/cdm/v1/> and <http://wp5.e-taxonomy.eu/cdm/v20/> respectively. The first version was primarily based on experiences using existing data models and formats and attempted to cover all data elements relevant for the taxonomic workflow, so that data coming from different applications could be stored and processed with a single database. Version 2 included additions and refinements that came up as requirements from CDM Users and developers of the JAVA-based programming library on top of the CDM (<http://wp5.e-taxonomy.eu/cdmlib/site/index.html>).

During the last year the CDM became increasingly stable because most of its parts have already been tested prior to the release of version 2. Therefore, both the number of model changes and the number of changes in the derived database schema decreased significantly over time. This represents an important step towards stable applications with well defined migration paths from older to newer versions.

However, the CDM is still evolving due to additional requirements from new user communities. Some of these requirements required model changes which are implemented into the new CDM version 3 available at <http://wp5.e-taxonomy.eu/cdm/v30/>.

Some of the changes have already been implemented and integrated into the CDM Library versions 2.2 and 2.4 which explicitly included database schema changes. Some more recent changes will be included into the upcoming CDM Library version 3.0.

The main sources for model changes towards version 3.0 were:

- Further development of the EDIT Taxonomic Editor and the EDIT Data Portals.
- Further development of the CATE Data Portals and Web Editor.
- Refinement of the import procedures for descriptive data via SDD.
- Import of additional data from the WP6 exemplar groups and setup of eFlora initiatives migrating to the EDIT CDM.

The model additions and refinements include:

- 1) Additional packages, classes and attributes: With the increasing usage of the CDM, it became clear that certain information elements had to be added. This includes information elements for fleshing out classes (new attributes) as well as entirely new classes and in some cases packages:
  - Security package (<http://wp5.e-taxonomy.eu/cdm/v30/EARoot/EA11/EA310.htm>)
  - User management (<http://wp5.e-taxonomy.eu/cdm/v30/EARoot/EA2/EA43.htm>)

An example for a new class is the Classification class which used to be modelled using the TaxonRelationship class but turned out to be not expressive enough.

An example for missing attributes is the country attribute in the GatheringEvent class which is valid only for areas of type country. It turned out that this is required in

addition to the more general collectingAreas attribute which holds all different types of area information for gathering events.

- 2) Restructuring: In some cases, the structure of classes had to be revised for a better reflection of the “taxonomic reality” or to make the CDM easier to use. The most significant change of this type was the removal of particular classes for the representation of references. References are now represented with a single class (ReferenceBase) instead of using multiple sub-classes for specific reference types (Article, Book, Database, etc.). This modification was necessary to make reference type changes easier (one of the requirements taxonomic users were asking for. Strict typing is still possible via a set of new interfaces (e.g. IArticle, IBook, IDatabase, etc.) derived from the old ReferenceBase subclasses. The new structure is accessible at <http://wp5.e-taxonomy.eu/cdm/v30/EARoot/EA9/EA264.htm>
- 3) Renaming: Some classes and attributes were renamed for improved consistency with the overall CDM naming strategy and the naming conventions of communities usually handling these data.

The majority of the modifications were applied to the occurrence and description (structured descriptions and key part) packages because most of the package features were not used by the taxonomic editor software as well as the data portals before version 3.

Changes to the Model are reflected in the new version 3.0 of the CDM at <http://wp5.e-taxonomy.eu/cdm/v30/>. A thorough class documentation is the Java documentation of the Model which is part of the library at <http://wp5.e-taxonomy.eu/cdmlib/site/cdmlib-model/apidocs/index.html>.

Starting from version 2.4 it was decided to document all model changes at <http://dev.e-taxonomy.eu/trac/wiki/CdmChanges> and to provide upgrade scripts for MySQL-based CDM databases to migrate older versions to newer ones.