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"

C5.124 Report on technical integration of descriptive tools into the Platform

Due date of component: Month 51
Actual submission date: Month 51

Start date of project: 01/03/2006 Duration: 5 years

Organisation name of lead contractor for this component: 2 MNHN

Revision: final

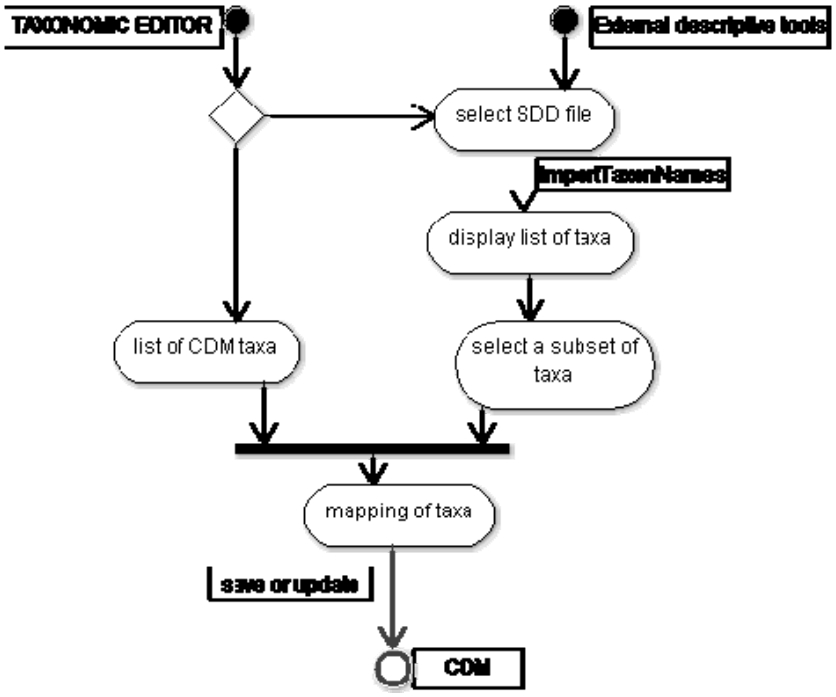
Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)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

Several tools already exist to manage descriptive data or to use and to analyse these data (Xper², DiversityDescriptions, Linnaeus, or other tools). These tools have hence proven their interest and their accessibility for taxonomists even if they have little or no previous computer experience. As EDIT's objective is to facilitate sharing of information and access to resources for the taxonomic community, enforcing the use of a new tool would contradict this spirit and then it was then decided to integrate existing descriptive tools into the Internet platform for Cybertaxonomy following the proposal of component 5.50 and the design drawn up in task 5.2.8. To perform this integration new modules were developed: SDD-CDM import/export and CDM input/output functions to include descriptive data into CDM Data portals.
This C5.124 describes the technical integration related to the taxonomic editor (to import/export or to edit descriptive data) and to the outputs (to display these data).

1. The integration to import/export or to edit descriptive data

SDD import & export

The external descriptive tools are accessible through the exchange format SDD. A button will be done in the taxonomic editor to access import/export SDD-CDM functions. The different steps are described in the following figure.



First, each SDD block is imported thanks to the following functions (*SDDDescriptionIO.java*): when *doInvoke()* is called on a *SDDDescriptionIO* object, after all necessary checks, *importDataset()* calls the rest of the import functions (such as *importTaxonNames*, *importCharacters*, etc) which map and then store the xml elements into java objects of the CDM model.
These objects are then saved in the database using Hibernate DAOs (different methods can be called for saving: *save*, *update*, *create*, etc).
Thus it is possible to select only one object in order to save or upload it in the database, this allows for example, to upload the database if changes have been made on some particular taxa or specimens only.

If only certain taxa of a SDD file have to be imported, it is possible to select them (by importing first the list of taxon names with the corresponding function) and updating the database: either the user chose the taxa for which new data have to be added either it is done automatically by comparing the names.

For exporting a CDM database into a SDD file, the process is the exact opposite:
data from the CDM database are extracted in java objects (*SDDCdmExporter.retrieveData()*) and then *SDDDocumentBuilder* calls the build functions (like *buildTaxonNames*, *buildAgents*, etc) that create an *ElementImpl* (from the Apache Xerces library, a processor for manipulating xml) which can be then written in a SDD file.

Descriptive data editor

The direct use of Xper^2 java code is tested in order to edit data directly on CDM classes. Xper^2 data classes are different from CDM classes but it appears that Xper^2 java code could be easily adapted in a way that Xper^2 works directly on CDM classes. Therefore, the Xper^2 base class *Base* needs to integrate all functionality of the CDM base class *CdmBase*. Also loading of data needs to be adapted in a way that data is not loaded and stored from and into a file anymore but directly from/into the CDM database using the CDM API.
Before Xper^2 can be fully integrated into the Taxonomic Editor in a long term the short term solution will be to integrate Xper^2 into the Taxonomic Editor in such a way that it can be opened via the Editor but runs as a separate application but on the same data as the Taxonomic Editor.

2. The integration of tools for new outputs on descriptive data

Displaying descriptive data as natural language descriptions will be easy now within the editor as well as within the dataportals using the *CDM io functions* developed by MNHN team.
The only open issue is how to mark the description in a way that the editor and the portal know that these are structured and not unstructured data and therefore need to be handled differently.

Several data portals already include a section for keys. From a taxon page dedicated to a taxon T, the tool to compute key will be called. It selects in CDM the descriptions referring to the taxa of a lower rank included in taxon T (working set class). Then the key construction tool works only on this working set class. The produced output is a structured XML key. Each terminal node is associated to a Taxon name and the corresponding taxon page.