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"

C5.50 Report on integration of existing descriptive tools into the Platform

Due date of component: Month 29
Actual submission date: Month 29

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)	

C5.50 Report on integration of existing descriptive tools into the Platform

Several existing tools allow the management of descriptive data, the construction of printable keys and the identification with free access keys. Their inventory and analysis were performed in a former step. Although none of the existing tools stands out against the others, several of them have an extensive user interface, are already adopted by biologists and were validated by numerous applications and taxonomic keys. These tools have hence proven their interest and their accessibility for taxonomists even if they have little or no previous computer experience. The European project KeytoNature, with which a collaboration is in progress, will also contribute to promote the interest in such systems and to distribute applications for multiple audiences.

The improvement of the existing tools is encouraged by the positive stimulation between providers. In addition, the cost in resources (persons and time) to design and implement a new tool for identification in the context of the EDIT Common Data Model (CDM) is considered to exceed the presently available resources. Also, as EDIT's objective is to facilitate sharing of information and access to resources for the taxonomic community, enforcing the use of a new and unique tool would contradict this spirit. We feel that taking advantage of past advances made by the community by increasing their availability and interoperability will be the best option.

Figure 1 describes the global architecture for this integration process and the different steps in its implementation.

Choice of TDWG-SDD

To ensure the future compatibility of the platform with other initiatives in the world, it was decided to integrate the existing tools into the cybertaxonomy platform by using the already-existing public standard exchange format for descriptive data: the TDWG-SDD XML schema. This choice was made in accordance with the latest efforts of the developers of several existing tools to allow interoperability with SDD.

In this context, some preliminary tests were performed to evaluate the exchange possibilities between two existing tools (Lucid and Xper^2) through SDD. These tests underlined some potential ambiguities, but these are mainly caused by the complementary features of the different tools, which offer specific benefits to taxonomists using them.

Relying on our expertise with SDD, some guidelines containing warnings could be formulated to help providers of descriptive tools to implement their own SDD-compliant import and export routines.

Import/export SDD-CDM

The CDM java classes will be able to import descriptions edited and created with any of the existing tools as long as they are compatible with SDD. This task requires the development of an SDD import into the CDM by complementing the CDM input-output library with SDD XML reading and mapping functionalities (fig 1-a). In addition, the implementation will also provide the possibility to export descriptive data stored in the CDM to an SDD XML file. The development will benefit of the Xper^2-SDD import-export which was developed using Java.

Following the architecture of the EDIT Platform, the permanent storage of the data (in a DB) is performed using Hibernate. The current CDM will need be checked and improved if necessary to adapt optimally to descriptive data (fig 1-b).

Displaying descriptive data from the CDM store

Like other data stored in the CDM, descriptive data should be accessible for display in web pages, e.g. in the EDIT Data Portal, for display (not editing) in the EDIT Taxonomic Editor, etc. Accessibility will be achieved using an API developed for the CDM store, as well as the SDD

export functionality, and will make use of tools customized specifically to offer new content visualizations (e.g. by using a tool transforming structured descriptive data into readable texts, i.e. "natural language" output). Accessibility for other applications, e.g. Scratchpads, will be affected using the export functionality.

Adding new tools for descriptive data
The architecture of the platform integration is flexible enough to allow addition of new tools at the level chosen by the provider of the tool: using the specific data format of an already existing tool (fig 1-c) ; - using SDD format (fig 1-d) ; using CDM and the remote API (fig 1- e). Useful features could include tools:

- to compute similarities between taxa,
- to create automatic diagnoses,
- to split continuous data into characters with intervals or to work out new characters by combining other ones,
- etc.

The evolution of some of these tools into web-services could allow taxonomists to access the functionality seamlessly through a web interface.
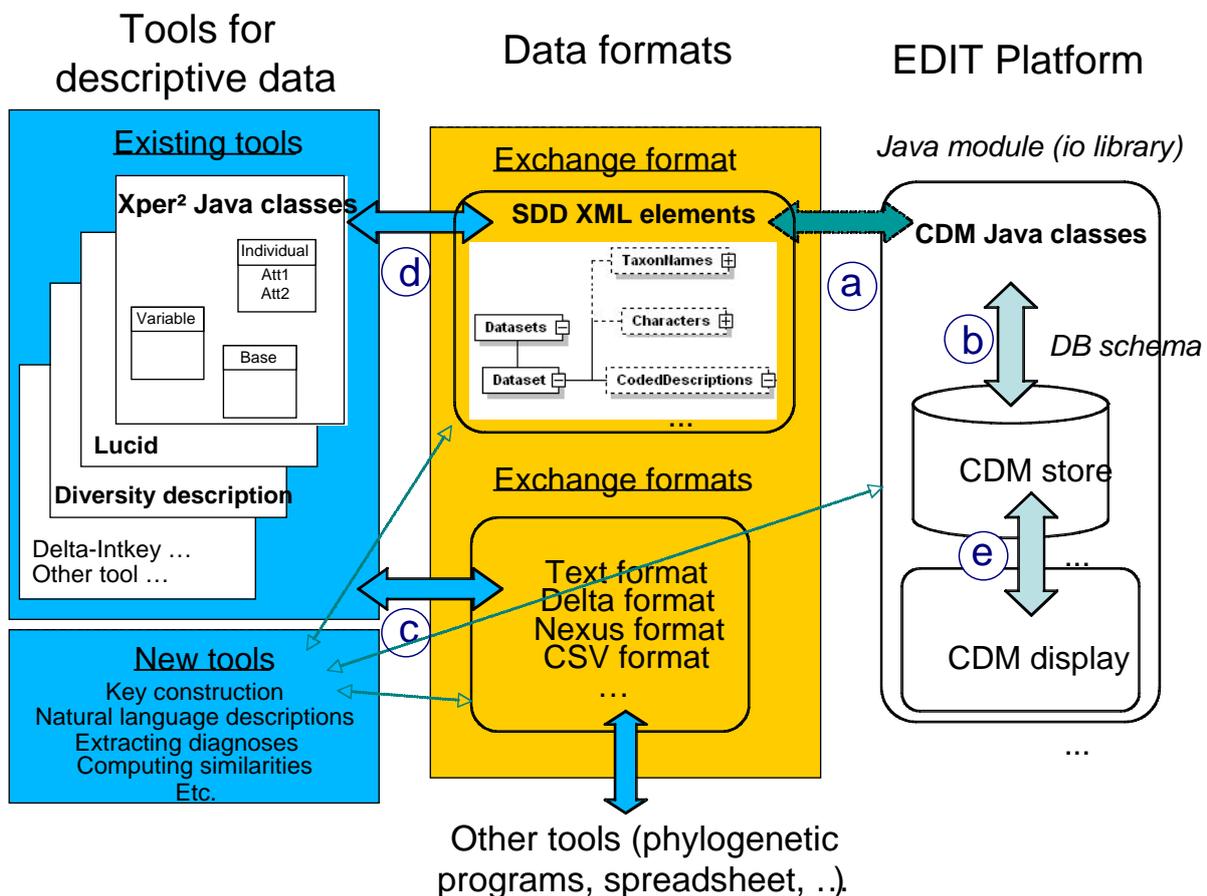

Fig 1: Descriptive tools in the platform of Cybertaxonomy