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"

M5.18 SDD import/export module functional

Due date of milestone: Month 32
Actual submission date: Month 32

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

Organisation name of lead contractor for this milestone: 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)	

M5.18 SDD import/export module functional

TDWG-SDD for ensuring compatibility of the CDM platform with other descriptive systems

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 a unique already-existing public standard exchange format for descriptive data: the TDWG-SDD XML schema. This choice was made in accordance with the late efforts of several existing tools (DiversityDescriptions, Xper², Lucid, FRIDA) to allow interoperability with SDD.

Import/export SDD-CDM

To allow the CDM to communicate with other descriptive systems through the TDWG-SDD XML schema, import functionalities into the CDM java libraries were implemented.

Concerned CDM libraries

The SDD import was developed in keeping with the other import functionalities in the CDM library 'cdmlib-io'. Activators for test examples, SDD file sources and CDM database destinations, were implemented in the 'app-import' library.

Imported SDD elements

For the first version of the import, the following descriptive data can be retrieved from SDD and stored into a CDM database: metadata about the SDD file: how it was generated and when, dataset information: name, author, editor, date of modification, copyright, taxon names, descriptive system with the definition of characters: categorical, quantitative and text characters, taxa descriptions using the descriptive system: associated with a taxon if the description refers to a taxon name, publications and their association with descriptions, images references.

The import of the following elements will be implemented in the next version:

- character dependencies and groups,
- single access keys,
- specimen descriptions.

We established a mapping between SDD XML elements and CDM description model Java objects and modified slightly some previously implemented CDM Java objects. This work enabled the initiation of a reflection on the CDM description model that will fuel the revision of the model planned for February 2009.

The import of SDD data was tested on different data sets: a simple example included with the SDD schema with the description of the taxon *Viola hederacea* (http://wiki.tdwg.org/twiki/bin/view/SDD/Version1dot1), SDD examples available on the SDD wiki site ranging from 54 to more than 2,000 taxa descriptions (http://wiki.tdwg.org/twiki/bin/view/SDD/RealWorldExamples_SDD1dot1), and an SDD export from a Xper² application on freshwater aquatic insects (6 taxa, 14 categorical characters, 35 states, 52 images).

Figure 1 shows the original knowledge base viewed from the taxon interface, while Figure 2 shows part of the export of the base into SDD format.

Both saving descriptive data into an empty base and adding descriptive data into an existing base have been tested.

Figure 1: Freshwater Aquatic Insects knowledge base in Xper²

```
<CodedDescriptions>
    <CodedDescription id="D1">
        <Representation>
            <Label>Odonate (juvénile)</Label>
            <Detail>Les&amp;nbsp&lt;b&gt;odonates&lt;/b&gt;&amp;nbsp (Odor
            <MediaObject ref="m2" role="Primary"/>
            <MediaObject ref="m3" role="Primary"/>
            <MediaObject ref="m4" role="Primary"/>
            <MediaObject ref="m5" role="Primary"/>
        </Representation>
        <SummaryData>
            <Categorical ref="c1">
                <Note>Les odonates respirent au moyen de branchies trachée
                <State ref="s1"/>
                <State ref="s6"/>
            </Categorical>
            <Categorical ref="c2">
                <State ref="s10"/>
            </Categorical>
```

Figure 2: Freshwater Aquatics Insects knowledge base: SDD export from Xper²

Until we have at our disposal a tool to view CDM descriptive data, we are using SQL queries to visualize the content of the descriptive part of the CDM database and test the consistency of the imported information (Figure 3).

DESCRIPTION	ELEMENT	DESCRIPTOR	DESCRIPTOR_NAME	STATE	STATE_NAME
Coléoptère (adulte)	CategoricalData	2049	Extrémité de l'abdomen	2035	sans particularité
Coléoptère (adulte)	CategoricalData	2050	Pattes articulées libres	2028	présentes
Coléoptère (adulte)	CategoricalData	2051	Pièces buccales	2042	sans particularité

Figure 3: Example of CDM storage for a categorical description after import of SDD file

Finally, as we enhance the import functionalities, an export function will be implemented creating SDD XML files from extracted CDM descriptive data, allowing bidirectional communication between tools using SDD and the CDM.