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"

M5.19b Functional description for descriptive user interface

Due date of component: Month 38
Actual submission date: Month 38

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

Organisation name of lead contractor for this component: 2 MNHN

Draft for revision

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)	

Introduction

This report completes the component C5.081 (Use case model and functional description of CDM descriptive data editor) and provides an illustration of the essential functionalities expected from the future CDM descriptive interface. Functionalities that could be made available and how existing tools can be integrated in order to provide such functionalities, are presented through one of the EDIT exemplar group: the Arecaceae family.

As in component C5.081, functionalities are divided into two major categories: edition of descriptive content and analysis/treatment of descriptive data.

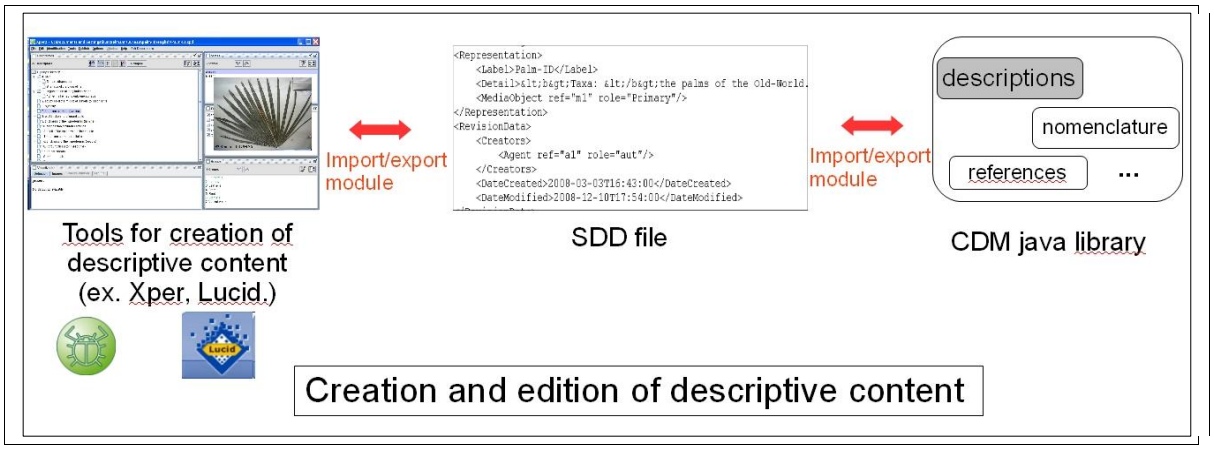
Editing descriptive content

CDM search for descriptive resources

The starting point for the taxonomist in the CDM descriptive system will most often be the search for existing descriptive content about a particular group (knowledge bases, images, bibliography, specimens, distribution, nomenclature). Resources could be displayed into different categories depending on their content, their format and their origin.

Creation and modification of descriptive content

Until a CDM editor for descriptive data can be implemented, the edition of descriptive data can be performed thanks to existing tools such as Xper², Lucid or DiversityDescriptions which support the international standard for structured descriptive data (SDD) developed by the TDWG. The SDD data can be imported and exported into and from the CDM which will allow the creation and modification of CDM descriptive data.



Visualization

The taxonomist will need to be able to visualize the descriptive data he is interested in. His needs in terms of visualization might vary:

- he may need to visualize only a subset of an existing resource, so he may want to filter data;
- he may want to display the descriptive content in a specific format adapted to the support (PDA, paper, HTML pages, etc.);
- he may want to display the descriptive content in a structured format or in natural language;
- he may want to choose the language if available.

CDM Descriptive data editor

The implementation of a CDM descriptive data editor was put aside for now but would be the ideal solution in the future to target seamless interoperability. The editor may be web-based and/or a standalone desktop application that connects to the according database (like the existing Taxonomic Editor) and allows the taxonomist to work anywhere. Based on the CDM descriptive model, it would also support SDD import/export for obvious reasons of opening to other international systems.

Analysing descriptive content

Once the content is created or modified and available in the CDM descriptive model, treatments valuable for the work of the taxonomists will be available and applicable

Building keys

The taxonomists may want to get printed keys from its descriptive data in order to include them to a monography, a work of revision or for popularization of scientific knowledge . Printed keys can be performed through specific functionalities of existing tools such as Lucid or Xper[2]. Free-access keys are also easy to create from structured descriptive data. Both kind of identification keys can be built thanks to SDD files generated from CMD descriptive data as inputs for Xper[2] or Lucid softwares.

Building diagnosis

As for building keys, the creation of diagnosis from CDM descriptive data can be performed with the intervention of external existing tools (ex. Mindescr).

Tool for writing in natural language

This functionality exists in Xper and Lucid but for this kind of specific treatment, we have to keep in mind that it requires an important involvement of the taxonomist in order to define the very precise parameters needed.

What is missing in the existing tools

- **Generalizing concepts**

The possibility of generalizing concepts can be very useful as the taxonomy of a given group can be subjected to frequent modifications. For example : generalization of specimens descriptions can lead to define a taxon description, or generalization of species descriptions can allow to create a genus description. This functionality has to be created in order to be available for the CDM descriptive content.

- **Tool for syntactic reading of descriptive text in natural language**

A taxonomist might want to be assisted in extracting relevant descriptive data from natural language descriptions found in taxonomic papers, monographs, etc. This functionality could be provided by referencing generic existing tools which can structure natural language text at first and then think of adaptations of these tools to the taxonomic context and the CDM descriptive model in particular.

Conclusion

The descriptive user interface of the EDIT cybertaxonomy platform will benefit from former research and existing descriptive tools. The availability of an import/export tool between the EDIT CDM descriptive model and the international TDWG standard for descriptive data SDD makes it possible to capitalize on existing functionalities.

However, the two-step translation from tools to the CDM (via SDD) increases the risks of loss or deformation of the initial information. In order to limit the impacts of that translation, integration of the tools in the CDM library should be encouraged so that treatments can be applied directly on CDM descriptive data. It is also a strong argument in favour of the implementation of an EDIT common tool for the edition of descriptive data.