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"

C5.081 Use case model and functional description of CDM descriptive data editor (Model/Report)

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 provides a summary of a study led on taxonomic activities related with descriptive data as part of the EDIT Activity 5.6 Keys and descriptions. The report presents UML use-case model for handling descriptive data and detailed use case scenarios for major functionalities. The use case scenarios show necessary descriptive functionalities for the work of the taxonomist with descriptive data: some of these functionalities are already available in existing descriptive tools, while others will need to be integrated or implemented. The objective of this study is to define the needed functionalities for the future CDM system for descriptive data.

The reflection was in part fuelled by the report from David Taylor on "Operational functional model for revisionary taxonomy" (WP 5.08 - July 2007).

USE-CASE MODEL

The use-case model presented on **Figure 1** depicts the main taxonomist tasks related to the creation and handling of descriptive data. The building and processing of descriptive data contain the majority of the identified use-cases and include the following main steps:

- acquisition of data,
- representation of descriptive data,
- analysis of descriptive data,
- paper-based or electronic diffusion.

Other activities of the taxonomist such as teaching and delivering or transmitting expertise benefit from this production of knowledge on descriptive data.
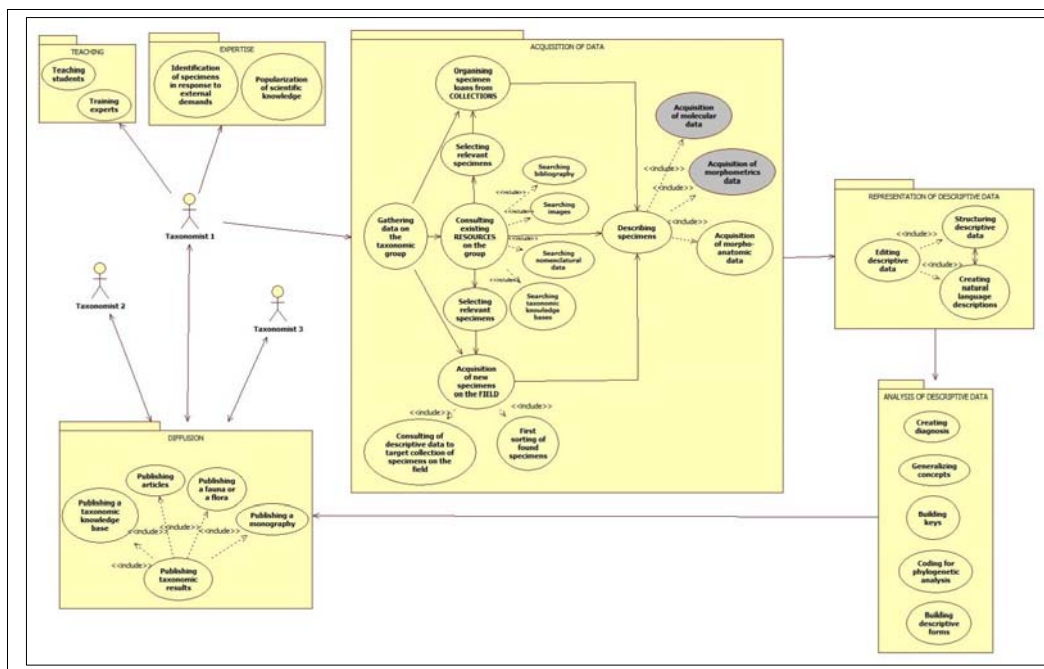

Figure 1: UML use case model of taxonomic activities in relation with descriptive data

USE-CASE SCENARIOS AND RESULTING FUNCTIONALITIES

We will now focus on the use-cases related with the creation and the use of descriptive data and their scenarios which defined the interactions between the users and the system, the system being the EDIT platform for cybertaxonomy.

To refine the use case model of the **Figure 1**, the activity models represented on **Figure 2** & **3** depict the set of actions that can be performed by the taxonomist while interacting with the system. These actions are placed in two main categories: edition of descriptive data and analysis of descriptive content.

Functionalities are prioritized with a colour indicating how important the functionality is for the taxonomist work with descriptive data. The colour scale is defined in the following way:

- red = core: without this functionality, the taxonomist cannot work on descriptive data.
- purple = high priority: this functionality is very important, but it does not prevent the user from entering basic data.
- green = low priority: major tasks for the taxonomist activity can be achieved without this functionality, but could be valuable subsequently.
- blue = high priority for key makers: the functionality is essential for a taxonomist whose aim is to build a key.

Activity diagrams cover the following use cases:

- searching and editing descriptive data,
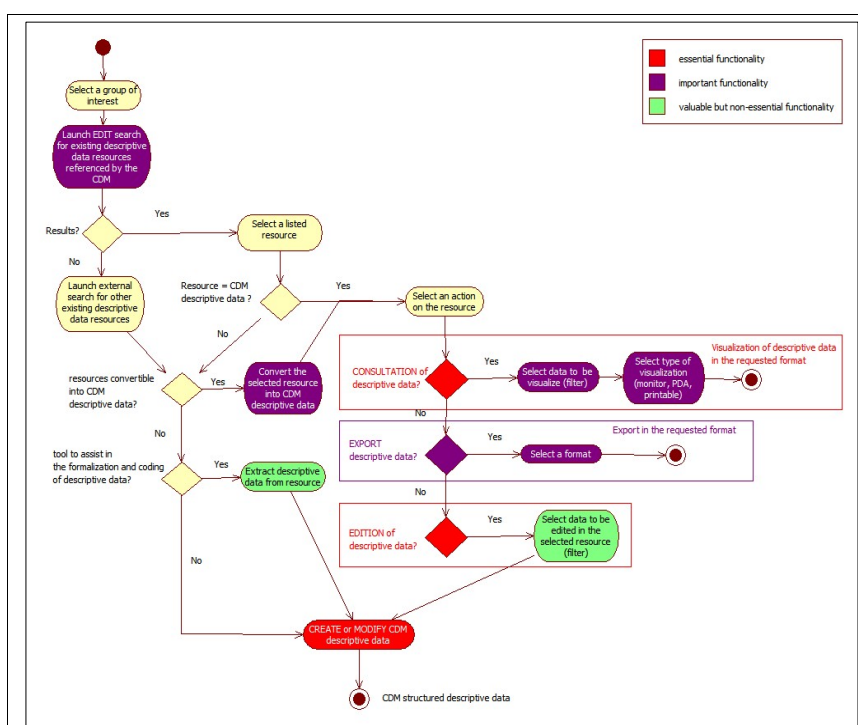- analysing descriptive data.

Figure 2: CDM search for existing resources and edition of descriptive data

AVAILABLE FUNCTIONALITIES IN EXISTING TOOLS

Some of the functionalities highlighted above can be found across existing descriptive tools . A list of such tools can be found in the EDIT Biodiversity Service & Application Tracker (http://www.bdtracker.net/softwareTracker/list/all/domain/2).

Available functionalities are:

Edition

- edition of structured descriptive data to build a taxonomic knowledge base,
- association of images with taxa, descriptions, descriptors,
- visualisation of structured data with html pages,
- handling of languages (partial),
- import and export from and to the international standard for descriptive data from the TDWG (SDD),
- import and export from and to tabulated formats (.txt/.csv),
- export in the Nexus file format widely used in several popular phylogenetic programs such as Paup*, Mesquite and MacClade.
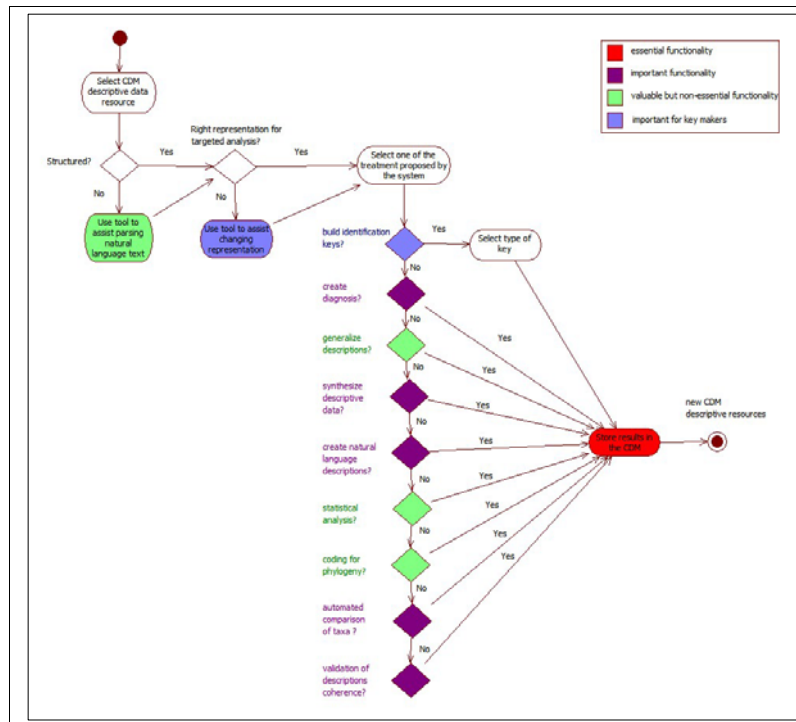
Figure 3: CDM search for existing resources and edition of descriptive data

Analysis
- building of printed keys,
- identification with free-access keys,
- creating natural language from structured descriptive data,
- building diagnosis,
- building forms synthesizing descriptive data on a taxon,
- automated comparison of taxa,
- validation of descriptions coherence.

Therefore, the CDM system for descriptive data could possibly reuse functionalities of some of the existing tools, either through an SDD import/export bridge, or through an adaptation and integration of the application code to the CDM library.

MISSING FUNCTIONALITIES

For some functionalities, alternative solution will need to be found, such as redevelopment within the CDM or selection of appropriate non-descriptive tools.

We selected the following functionalities ranked from higher to lower priority:
- EDIT inventory of descriptive resources and associated search functionality,
- tool for syntactic reading of descriptive text in natural language,
- generalisation of descriptions (e.g. building taxon descriptions based on specimens descriptions),
- complete handling of languages,
- statistical analysis of descriptive data (data summary, ANOVA),
- collaborative edition of descriptive data.

CONCLUSION

By highlighting necessary functionalities for the work of taxonomist with descriptive data, the objective is to detect how to capitalise on existing tools and prioritize actions for the setting up of the CDM system for descriptive data whether by integration of existing functionalities or development of new ones.