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"

C5.116 Inventory of descriptive knowledge bases

Due date of component: Month 40
Actual submission date: Month 40

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 presents the train of reflection on the execution of an inventory of descriptive knowledge bases and interactive keys as part of the EDIT Activity 5.6 Keys and descriptions. The main steps to get this inventory started and efficient are developed here and a proposition including a collaboration with the Key to Nature is provided.

IMPORTANCE OF AN INVENTORY OF KNOWLEDGE BASES

The importance of an inventory of knowledge bases (KB) takes root in the general use-case depicted in a former component (C5.081). As parallel productions for expertise or/and education, or direct standardized storage of taxonomic descriptions, knowledge bases and associated keys represent an important step in the management of descriptive data in the current (and future) context of cyber-taxonomy. When working on its own group, evaluating which group should be taxonomically investigated or simply looking for examples of what can be actually done, the taxonomist must be informed of existing knowledge bases.
Some websites actually provides an overview of knowledge bases or keys. However we noticed that no list is designed to achieve a complete inventory, without any particular focus on a single software, taxonomic group or geographic zone. These are some examples:

- Key to Nature - identifications keys for educational purposes - http://www.keytonature.eu/wiki/Main_Page;
- DELTA website: - knowledge bases and keys under DELTA format - http://delta-intkey.com/www/data.htm;
- Lucid website - free and commercial keys edited with the Lucid software - http://www.lucidcentral.com/Keys173/SearchforaKey/tabid/217/Default.aspx;
- Fauna keys - Lucid keys for some groups of vertebrates and invertebrates of New South Wales and Australia - http://www.faunanet.gov.au/faunakeys/index.htm.

The result is a lack of visibility for a taxonomist looking for an exhaustive list, and a need for centralisation of data on knowledge bases and keys obviously becomes apparent.

The aim of this inventory is to provide a list of descriptive KB that can be easily: **maintained, increased and consulted.**

TECHNICAL PROPOSAL FOR THE INVENTORY

A list of about 40 descriptive datasets and/or identifications keys has been firstly collected by hand in tabular format. When possible, the keys has been at least opened and briefly tested (to evaluate if they were functional). The aim was to evaluate the diversity of metadata that we will have to manage.

Even if a good number of descriptive works is easy to reach through basic web research, the idea of a collaborative inventory arose as the task appeared to be rough because of the heterogeneity of metadata found for each knowledge base/key (authors names, geographical scope, taxonomic group, used software, etc.), and because some knowledge bases/keys are scattered on the web (not belonging to one of the big portals, see above) and then not easily reachable.

A Wikipedia-like website seems to be a smart solution for allowing people to add their own KB and keys, and create an emulation around the use and the production of this type of tools. A project called Species-ID (http://www.species-id.net/wiki/) enables people to create collaborative pages as well as authored pages. See the Ownership and sponsors page for further information on Species-ID (http://www.species-id.net/wiki/Ownership_and_Sponsors) as well as its Charter (http://www.species-id.net/wiki/Charter).

The strategy will be to start to fill the inventory on this wiki and then encourage other users to contribute. If we have to face a growing list, the possibilities of querying have to be thought ahead. We can then plan a supplementary alternative which could consist in harvesting the

collaborative list in order to create a database of metadata that would be interrogated from an other website (e.g EDIT website, BDTracker in particular) with a simple PHP-MySQL tool.

Thanks to the creation of templates for metadata (XHTML tags) on Species-ID, we should meet this requirement. The technical details are being currently discussed with G. Hagedorn (Key to Nature) and the valuable advantage is that Key to Nature has already predefined templates' standards for the storage of metadata related to descriptive datasets and keys. Find in fig.1 a schematic overview of the technical proposal for the inventory.

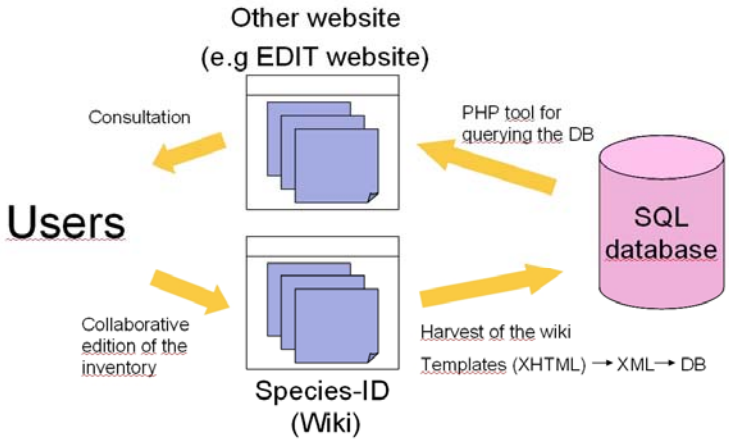

fig 1: Collaborative edition of the inventory and storage in a database for interrogation.

DISCUSSION

Species-ID, even if under construction for now, seems to be the best choice for beginning a collaborative inventory as we will benefit from the yet implemented handling of metadata in Key to Nature. The only risk is if the Species-ID wiki does not win over the taxonomists. We could then encounter a problem of participation which could impaired the collaborative increase of the inventory. That is why we will start the inventory on Species-ID and try to invite some targeted people to participate. We do not want this inventory to be disconnected from the rest of the EDIT project: this implies to have the opportunity to link the inventory to the Cyber-Platform.

CONCLUSION

This solution of a collaborative inventory seems to meet the actual needs for a better visibility of all standardized descriptive works but it has to be developed together with a solution for smart querying.

The effective beginning of the inventory depends on some still in curse discussions with Key to Nature, on the adaptation of metadata templates to the inventory issues. We would like the PHP tool for querying the subsequent database to be implemented by our future developer (H. Fradin's successor) under the supervision of R. Vignes-Lebbe.