

## Xper<sup>2</sup>: introducing e-taxonomy

Visotheary Ung\*, Guillaume Dubus, René Zaragüeta-Bagils and Régine Vignes-Lebbe

UPMC, Université Pierre et Marie Curie, UMR 7207 (CR2P), Centre de Recherche sur la Paléobiodiversité et les Paléoenvironnements, Paris, France

Associate Editor: Alex Bateman

### ABSTRACT

**Motivation:** Computer Aided Identification systems provide users with the resources to relate morpho-anatomic observations with taxa names and to subsequently access other knowledge about the organisms. They have the ability to manage descriptive data and make identifications through interactive keys. They are essential for both authors and users of biodiversity information. Xper<sup>2</sup> version 2.0 is one of the most user-friendly tools in its category and provides a complete environment dedicated to taxonomic management.

**Availability:** Xper<sup>2</sup> software can be freely downloaded at <http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/softwares/xper2>

**Contact:** visotheary.riviere-ung@snv.jussieu.fr

Received on October 15, 2009; revised on November 24, 2009; accepted on December 28, 2009

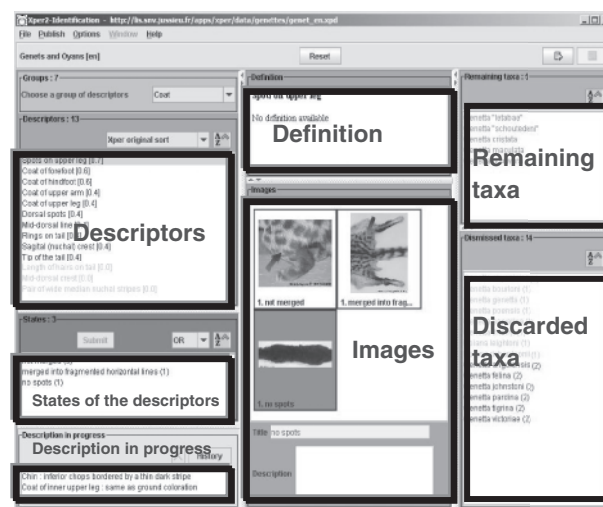
### 1 INTRODUCTION

Linnaeus wrote that naming is the first stage of knowledge. Indeed, names of taxa are the main access keys to biological information (Polaszek, 2005). They link the real world of living organisms and the conceptual world proposed by systematists through the description and characterization of taxa. Identifying life forms is crucial for all activities related to biology and their taxonomic descriptions are more than necessary in the study of Biodiversity (Godfray, 2002; Lebbe and Vignes, 1991, 1998). Systematists inventory, study and structure biological diversity as precisely as possible (Matile *et al.*, 1987). All this information may be summarized into knowledge bases to perform identification of specimens for new inventories and monitoring surveys, phylogenetic analyses, as well as biogeographic and ecological studies. A significant increase of this kind of digitized information should be expected in a near future. The deep social and scientific impact of web 2.0 stimulates sharing of digital data and proposals of 'cybertaxonomy' projects to study global biodiversity and climate changes. The generalization and the integration of knowledge-base management systems into current taxonomic work may be the next revolution in taxonomy. International projects are already opening the way of this new, delocalized taxonomy.

### 2 WHAT IS XPER<sup>2</sup>?

Xper<sup>2</sup> is a management system for storage, editing, analysis and on-line distribution of descriptive data. It dynamically creates interactive keys for identifying specimens (Fig. 1). Xper<sup>2</sup> does

\*To whom correspondence should be addressed.



**Fig. 1.** A screenshot of the identification module (Genets and Olyans knowledge base).

not require any special computer skills. It is written in Java and available on Windows™ Mac™ or Linux in French, English or Spanish versions. It has a user-friendly and intuitive interface aimed at professional taxonomists as well as naturalists who merely want to identify specimens using a ready-made application. Xper<sup>2</sup>'s users are taxonomists, teachers, fauna and flora experts (for biodiversity monitoring or survey) and ecologists. Formalization of knowledge is the main difficulty for taxonomists. How to structure and organize different types of data? How to keep the scientific sources of the data in order to maintain traceability? Xper<sup>2</sup> is designed to manage this rich and heterogeneous information. Divided into four modules, Xper<sup>2</sup> allows taxonomists to either access the descriptive environment, the taxonomic description module, the management tools or the free access key:

- (1) descriptive model: creation and editing of a standardized description (structured into characters, states, groups and logical dependencies);
- (2) taxonomic editor: characteristics of the taxon, i.e. its name, descriptions, comments and pictures;
- (3) management tools (see Section 3.4); and
- (4) free access key: creation of interactive identification keys (IIK).

### 3 NEW FUNCTIONALITIES

#### 3.1 Import/export structural descriptive data

Many international research programs focus on the establishment of standard formats for data and knowledge exchange. Xper<sup>2</sup> implements the structured descriptive data (Hagedorn *et al.*, 2006), the exchange format for taxonomic descriptions that has emerged from the work of the Biodiversity Information Standards (TDWG, <http://www.tdwg.org>, formerly known as taxonomic database working group). Xper<sup>2</sup> can seamless import/export data from/to SDD, following international recommendations.

#### 3.2 Nexus export

Xper<sup>2</sup> extends the descriptive morpho-anatomic work and the taxonomic expertise of taxonomists into phylogenetic studies. A knowledge base can be exported in nexus format and directly be run in most phylogeny programs such as PAUP\*.

#### 3.3 HTML export

Xper<sup>2</sup> provides users with a user-friendly tool to instantly format the knowledge base in a set of HTML files according to a defined cascading style sheet provided by the application. These pages include a form for each taxon displaying its complete description, pictures and any additional comments. By this way, monographies can be easily computed and published in order to serve the entire community. As an example, a website on monitor lizards (<http://lis-upmc.snv.jussieu.fr/varanID>) has been created by a naturalist group from the Paris Museum.

#### 3.4 An accurate knowledge representation and more tools to analyze descriptive data

Xper<sup>2</sup> offers solutions to control and prevent inconsistencies via a check base function. Furthermore, analyzing descriptive data is one of the most important taxonomists' tasks and Xper<sup>2</sup> gives tools to carry it out:

- Xper<sup>2</sup> is able to manage uncertainty by selecting several descriptor-states and using logical operators (logical AND, logical XOR, logical NOT).
- Mismatch threshold: one can define a different threshold according to his expertise level when performing identification.
- Instant comparison of all taxa in a matrix summarizing descriptions.

### 4 FORTHCOMING DEVELOPMENTS

Management of extended descriptor types are under development. This includes management of numerical data, hierarchical or structured descriptor states, computed values, or any other user-declared type (Chalubert and Vignes, 2006). Associated tools allowing writing descriptions or diagnoses in natural language or proposing different kinds of similarity index among taxa will be implemented in future versions.

### 5 AVAILABILITY

Xper<sup>2</sup> is a versatile tool for editing and using taxonomic descriptions. It is free of charge and can be downloaded at <http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/softwares/xper2>. A mailing-list facility provides users with any help or advice that may be needed. Authors choose how they want to use or publish their knowledge base: distributed on CD (i.e. Mascarene Corals identification tool, Conruiy *et al.*, 2000) or online publication (e.g. XperBotanica—a project to identify the flora of France, <http://lis-upmc.snv.jussieu.fr/xperbotanica>).

### 6 CONCLUSION

Many international projects such as GBIF, EDIT, BIOTA, Catalogue of Life, Encyclopedia of Life, Key2Nature, etc. have a common purpose: sharing knowledge and providing free web access to Biodiversity Information (Bisby *et al.*, 2002; Ramsey and Janzen, 2002; Tillier and Roberts, 2006). In this context of collaborative work, Xper<sup>2</sup> fits with current taxonomists' practices and gathers functionalities to perform a suitable *e*-taxonomy!

### ACKNOWLEDGEMENTS

The authors thank B. Pavie, C. Gallut and A. Chalubert for their contribution to the initial Xper<sup>2</sup> version in Java. They thank all the Xper<sup>2</sup> users and especially Fabienne Audebert, Òna Maiocco, Philippe Gaubert and David Caron for their bases and also members of the EDIT project for their comments on the software. They are also very grateful to Dr Kurt Liittschwager for comments and style improvement.

*Funding:* Xper<sup>2</sup> development has been partially funded by the BioInfo 2002 grant of the CNRS and by the French Ministère for Research and New Technology (No. 04L370—Project Xper Botanica, 2005–07).

*Conflict of interest:* none declared.

### REFERENCES

- Bisby, F. *et al.* (2002) Taxonomy at a click of mouse. *Nature*, **418**, 367.
- Chalubert, A. and Vignes, R. (2006) A New Model for Descriptive Knowledge. *Proceedings of TDWG Annual Meeting 2006*. Available at <http://www.tdwg.org/proceedings/article/view/32>.
- Conruiy, N. *et al.* (2000) IKBS, a knowledge base management system for reengineering systematics: application to corals of the Mascarene Archipelago. *ICRS 2000—Nineth International Conference for Reef Studies*. ICRS, Bali, p. 17.
- GBIF, *Global Biodiversity Information Facility*. <http://www.gbif.fr/>
- Godfray, H.C.J. (2002) Challenges for Taxonomy. *Nature*, **417**, 17–19.
- Hagedorn, G. *et al.* (2006). The Structured Descriptive Data (SDD) w3c-xml-schema, version 1.1.
- Lebbe, J. and Vignes, R. (1991) Génération de graphes d'identification à partir de descriptions de concepts. In Kodratoff, Y. and Diday, E. (eds) *Induction Symbolique et Numérique à partir de données*. Cepadues, France, pp. 193–239.
- Lebbe, J. and Vignes, R. (1998). State of the art in computer-aided identification in biology. *Oceanis*, **24**, 305–317.
- Matile, L. *et al.* (1987) Introduction à la systématique zoologique. *Biosystema*, 1–67.
- Polaszek, A. (2005) A universal register for animal names. *Nature*, **437**, 477.
- Ramsey, S. and Janzen, D. (2002) All living things, online. *Nature*, **418**, 262–263.
- Tillier, S. and Roberts, D. (2006) Taxonomy on the fly in a European web project. *Nature*, **440**, 24.