BIODIVERSITY INFORMATION MANAGEMENT

Régine Vignes Lebbe
Université Pierre et Marie Curie, Paris, France

Contents

Summary

The term "biodiversity information system" is a generic name to denote all types of information systems that allow management of biodiversity data. These data involve information on a multitude of living beings, including their physiological and biological characteristics, and a description of the ecosystems where they are observed**.** One very important issue in such systems is the correct identification of these beings, to allow subsequent studies of their habitats, determine spatio-temporal correlations among species, and therefore derive new facts about our planet. This article discusses the problem of naming in the context of these information systems, and the role a proper unambiguous identification plays in biodiversity studies.

1. Naming Living Things

The naming of edible or beautiful plants, useful or dangerous animals, or any living being has always been considered important for humans. The act of naming gives an identity—it means an unambiguous way to designate organisms, and subsequently makes it possible to speak and exchange knowledge about plants and animals. According to their environment and their type of life, human populations have created numerous names for birds, fishes, vegetables, and so on. The vernacular names are often restricted to a country or a region, to a development stage, or to a special use of the

designated species. For example, in French a *faon* is a *Cervidae*'s calf and a *brocard* is a male at least one year old, with branched antlers.

In the middle of the eighteenth century, Carl von Linné proposed a binominal nomenclature that is still used today. Each species is designated by two Latin names: the first one stands for the genus and the second for the species. For example, the wolf is named *Canis lupus*.

Whereas vernacular names are used, more often in a limited region, Latin scientific names are used as worldwide reference. Each time a new species name is created, it is attached to a specific specimen, which has to be stored in a museum. This type specimen serves as an indelible and objective link between the scientific name and the biological reality that this name designates. Despite this, the non-ambiguity of scientific names is utopian: several experts can designate the same species with the same name but with different delimitations or interpretations, and at the other extreme experts have designated the same species with different names. As a matter of fact, a specialist may think that he/she has discovered a new species and thereby create a new name, which can later be shown to be a synonym of the earlier name designating this species. Moreover, for the sake of stability we seek to avoid name multiplication, so when a species is split as a consequence of further studies, one of the new species keeps the old name even if the designated biological reality has changed. Likewise, if two species are grouped into a single one (the observed differences being only geographical variants or due to special ecological or climatic conditions), the new species is designated by one of the old names whereas the other name becomes a synonym. Sometimes the situation is even more complex, since only a part of an old species may be attached to a new one (individuals of only one sex, or even some individuals of each sex, as an example male and female of a species of mosquitoes might initially have been mismatched). In such cases, we have what are called "partial synonymy" (*pro parte*). Nomenclatural codes regulate these naming actions.

Of course the interpretation of the species evolves with expert studies, just like the general classification of life. Nowadays we estimate 1.7 million, with 3 million scientific names, species known to date. In this context, finding information about a specific species is tricky: information is dispersed, written in many languages, stored on different media and related to multiple scientific names. The accumulated knowledge causes many problems of management and of access to the information, with more than a million synthesis books (monographs, faunas, floras) and millions of specimens conserved in many reference collections, especially in natural history museums. The Internet and the web, while providing new and effective ways of diffusion, also contribute to disperse information, whether validated or not. Moreover, contrary to appearances, we are far from achieving a complete species inventory.

The demand for access to information is growing every day; and as a matter of fact, the end of the twentieth century revealed a world preoccupation with biodiversity issues. This feeling was put in concrete form during the Earth Summit held in Rio de Janeiro in 1992, with the following Convention for Biological Diversity. The engagement of the nations is inducing systematists to propose numerous national and international projects to give efficient access to biodiversity information. For more on these issues, see

Advanced Geographic Information Systems.

2. Systematics

The level of organization studied by systematics includes organisms, populations, and taxa (see Figure 1). Of course all scientific domains have strong links, and systematists have to take into account the results of molecular biology or ecology and so on in their own work.
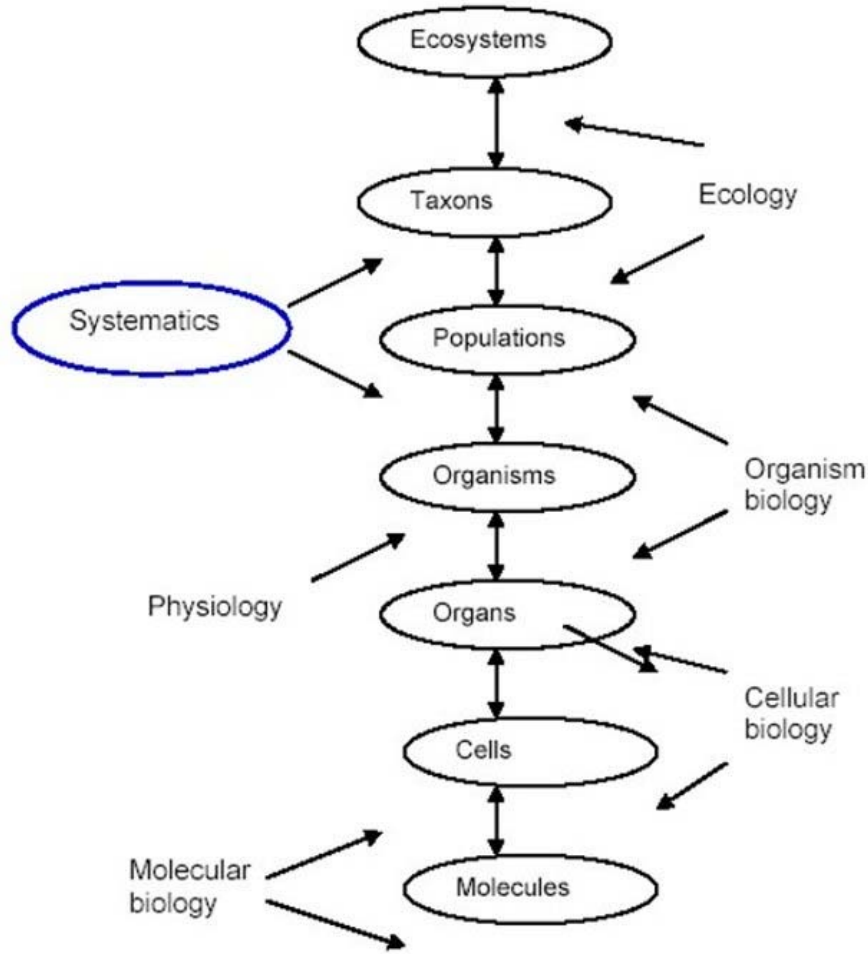

Figure 1. Systematics in biology

For studying biodiversity, the first task of systematics is to observe and describe any living organisms, using different character types (morphology, molecular data etc., and paleontology). Then it identifies, defines and classifies groups of organisms. These groups are named "taxa" and placed at different levels: species, genus, family, order, and so on.. For instance, human beings form a species taxon named *Homo sapiens*; the genus is *Homo*, the family *Hominidae* and the order *Primate*. Today this classification aims to be phylogenetic, which is to say that it is a hypothesis on the genealogical relationships of organisms.

Systematics has also the responsibility of providing efficient tools that help to identify any specimen. The most classical method of identification is named "identification key" or "diagnostic key," and is published as a sort of questionnaire (see Figure 2).

Within general biology, systematics appears between the domains analyzing suborganism structures, such as physiology or cellular and molecular biology, and domains studying the relationships between taxa and their environment as ecology.
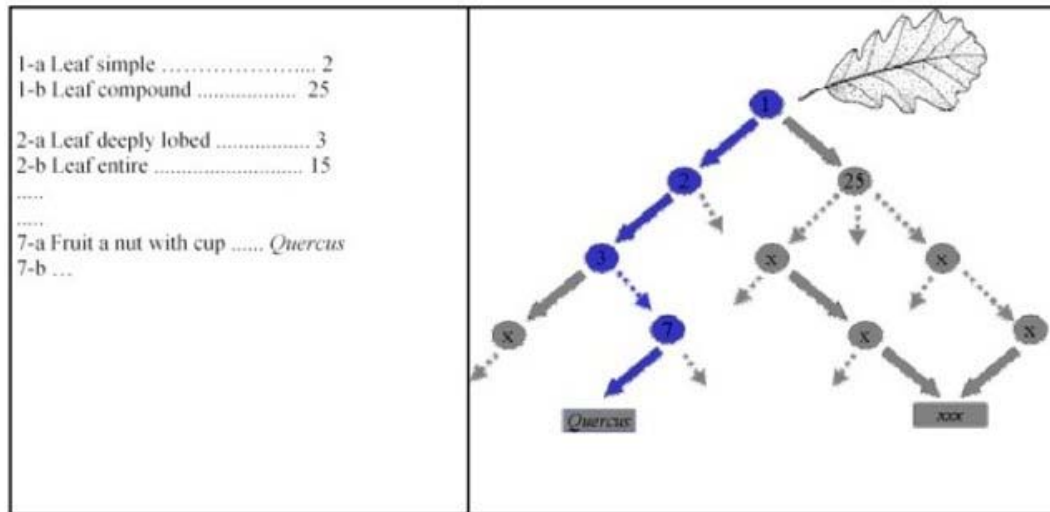

Figure 2. An identification key appears as a questionnaire leading to a taxonomic name. The structure of such a questionnaire is like a special graph.

3. Computer Science and Systematics

Ever since the 1960s, computer softwares have been used to help systematicists in all their activities: database management systems to manage observations in the field or specimens in collections; pattern recognition, morphometry, data analysis, and statistics are used to analyze the data; classification algorithms compute phylogenetic trees; discrimination methods help construct identification keys or computer-aided identification systems, and so on. Some computer programs have general applications, like statistics programs, but others are specific (phylogeny, identification key) or have to deal with the original properties of taxonomic data. Thus some biologists have acquired skill in computer science to develop, with or without the collaboration of computer scientists, specific informatics tools for biological purposes.

Since the 1990s the Internet has become a key element for helping systematists to share information. More than just exchanging electronic mail or transferring files, the use of the Internet gives efficient and global access to the data, and is the starting point of computer-supported collaborative work at the scale of the planet.

Information management systems and the Internet are central elements to respond to the needs of science and society. Many databases, CD-ROMs, and websites exist on different taxonomic groups or on precise geographic areas. But how to interconnect

these systems? How to know where and if information is available? How to evaluate the reliability of this information? Even with modern management systems, a crucial problem is not solved: how to organize and to structure the information about the 1.7 million species already known to give a global efficient data access. For more on structuring data, see *Integrative Data Structures for Collaborative Modeling* and *Visualization in Spatial Decision Support Systems*

-
-
-

TO ACCESS ALL THE **13 PAGES** OF THIS CHAPTER,
Visit: http://www.eolss.net/Eolss-sampleAllChapter.aspx

Bibliography

Fishbase. *A Global Information System on Fishes*. <http://www.fishbase.org/home.htm> [Fishbase on the web contains practically all fish species known to science. A relational database includes information to cater for different professionals such as research scientists, fisheries managers, zoologists, and many more.]

Fortuner R. ed. (1993). *Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision*, 560 pp. Baltimore, Johns Hopkins University Press. [Compiled papers from a large panel of scientists involved in taxonomic computing.]

GBIF. *The Global Biodiversity Information Facility*. <http://www.gbif.org/> [GBIF will be an interoperable network of biodiversity databases and information technology tools that will enable users to navigate and put to use the world's vast quantities of biodiversity information to produce national economic, environmental, and social benefits.]

Lecointre G. and Le Guyader H. (2000). *Classification Phylogénétique du Vivant*. 543 pp. Paris: Belin Ed. [A wonderfully illustrated book to discover the tree of life, with the arguments sustaining each taxon.]

Pankhurst R.J. (1991). *Practical Taxonomic Computing*. 202 pp. Cambridge, UK: Cambridge University Press. [Written by a pioneer in the use of computer in systematics, this book gives a clear overview of identification methods and databases in systematics, 15 years after the beginning of taxonomic computing.]

Species 2000. Species 2000 website: index of the world's known species. <http://www.sp2000.org/>. [Species 2000 has the objective of enumerating all known species of plants, animals, fungi, and microbes on Earth as the baseline dataset for studies of global biodiversity.]

Tillier S. ed. (2000). *Systématique: Ordonner la Diversité du Vivant.* 257 pp. Paris: Rapport sur la Science et la Technologie No.11. Académie des sciences. Technique and Documentation. [Report on systematics in France: research, teaching, national and international projects; description of the different tasks; how to preserve taxonomic expertise.]

TDWG. International Union of Biological Sciences Taxonomic Database Working Group website. <http://www.tdwg.org/> [The missions of this international group are to provide an international forum for biological data projects, to develop and promote the use of standards, and to facilitate data exchange.]

Winston J.E. (1999). *Describing Species*. 518 pp. New York: Columbia University Press. [A very complete book describing all the tasks of taxonomic experts in fields from phylogenetic hypothesis to knowledge diffusion.]

Biographical Sketch

Dr. Régine Vignes-Lebbe was born in Paris in 1957, and became a doctor of medicine in 1983. For some years she worked as a medical doctor and simultaneously prepared a scientific thesis about biological specimen identification methods and help for diagnosis; she worked with API Systems and then bioMérieux SA on identification of bacteria. In 1987, the first computer-aided identification system on phlebotomine sandflies (small insects vectors of diseases) was distributed by the Institute Pasteur with World Health Organization support. In 1996 she entered the University Paris 6 and joined Professor Jacques Lebbe's laboratory, which develops methodological research in systematics and promotes computer-aided systematics. In 2000, she became Professor and, after the death of Jacques Lebbe, Director of the laboratory. Today she is also Codirector of the Classification, Evolution and Biosystematic research team, and responsible for a Master's course on Systematics and for the palaeobotanic collection database in the University Paris 6. She is a partner in different international projects on biodiversity such as the European project BioCASE (Biological Collection Access Service for Europe). Her research includes work on knowledge representation at specimen or taxonomic levels, work on collaborative editions, and exploration of databases on the web. She also works on online identification and taxonomic descriptions, key construction, diagnosis extraction and knowledge analysis, natural language to help data acquisition, and editing monographs. New works in progress include discrimination with morphometric data image indexing.