

An overview on biodiversity information in databases

Piramanayagam Shanmughavel*

DBT Bioinformatics Facility, Department of Bioinformatics, Bharathiar University, Coimbatore - 641046;

E-mail: shanvel_99@yahoo.com; * Corresponding author

received January 05, 2007; revised March 09, 2007; accepted March 20, 2007; published online March 27, 2007

Abstract:

The massive development of biodiversity related information systems over the WWW (World Wide Web) has created much excitement in recent years. These arrays of new data sources are counterbalanced by the difficulty in knowing their location and nature. However, biologists and computer scientists have started to pull together in a rising tide of coherence and organization to address this issue. The fledging field of biodiversity informatics is expected to deliver major advances that could turn the WWW into a giant global biodiversity information system. The present paper briefly reviews the databases in preserving the biodiversity data.

Keywords: biodiversity; informatics; database; internet

Background:

There is a resonance between the needs of biodiversity science and the opportunities for globalization and interoperability provided by the internet. [1-3] Biodiversity scientists are distributed all over the globe and are interdependent. Global biodiversity depends on several parameters such as biomass, ecosystems, phyla, floras and faunas, hot-spots, genetic erosion, the impact of aliens and others. [4] The goal for biodiversity informatics projects is to develop systems that permit data interoperability and knowledge synthesis across wide arrays of local systems, and to embed them in global knowledge architectures. This paper focuses the importance of biodiversity databases in data management and integration of scattered data for biotechnological and agricultural research.

Current status

The term 'biodiversity' comprehended the totality and variability of species, genes and the ecosystems they occupy. Biodiversity is usually used to refer biological diversity at three levels such as (1) genetics, (2) species and (3) ecology. In recent years, the advancement in bioinformatics is mainly accelerated by the study of molecular events using data obtained by high throughput experiments such as whole genome sequencing, functional annotation, expression analysis and others. However, the full value of molecular biological information cannot be realized until it is possible to correlate genetic information with native habitat, neurobiology, physiology, or with genealogical relationships of the species. At the same time, biodiversity informatics would greatly benefit from inter-compatibility with molecular-level datasets.

Scientific needs

In the last 20 years, brain and behavioral research has experienced explosive growth because conceptual links have been made across different species, different levels of biological organization, and different experimental and theoretical approaches. [5] The dramatic increase in the amount of information has caused neuroscientists, of necessity, to increasingly narrow their areas of specialty, just to be able to keep up with publications most relevant to their own research. The cost of such specialization is a decrease in the development of new conceptual linkages. Thus, the amount of information generated by the engine of interlinked

research threatens to choke the engine itself. However, advances in informatics focused on brain and behavioral research information can prevent the stifling of this success. A major scientific consideration in biodiversity science is the need to bring 25 decades worth of accumulated information into an electronically available format. Unlike other sub disciplines of biology, biodiversity (primarily taxonomic and ecological) research results do not rapidly go out of date. In fact, many such results probably cannot be replicated because of anthropogenic habitat modifications that have occurred since the research was done. In addition, new data types are being generated by satellite imagery and other measures of non-biological global phenomena that have significant influence upon biodiversity. Great forward strides could be made in the understanding of the biological world, for instance, if informatics techniques were developed to make it possible to correlate historical information with newly collected satellite data; if molecular genetic datasets could be linked to species-documentation datasets such as those held by natural history collections; and if neurobiological, physiological, chemical, and other datasets could be correlated with taxonomic and ecological ones.

Biodiversity data management

The first step in biodiversity conservation is documentation based on the availability of information about each species with data starting from its systematic position to molecular aspects. [4] In many biodiversity databases, data is held either about species or specimens such as (1) nomenclature-species name, geographical data and status scale; (2) descriptive data-morphology, anatomy, chemistry, ecology etc.; (3) economic importance; (4) conservation status; (5) images; (6) bibliography sources of data used in the database.

Kinds of biodiversity data and databases

Taxonomists create the nomenclature and classification databases. Databases can also contain all kinds of information about organisms, including their characteristics, economic importance, conservation and management (Figure 1). Bioinformatics developed technologies for the management of genomic and proteomic data. In the past decade, electronic storage media, WWW, database technology and digitalization of data creation of public databases are creating a revolution in the way that biodiversity information is created,

maintained, distributed and used. Biodiversity informatics then includes the application of information technologies to the management, algorithmic exploration, analysis and

interpretation of primary data regarding life particularly the species level organization. [1-4]

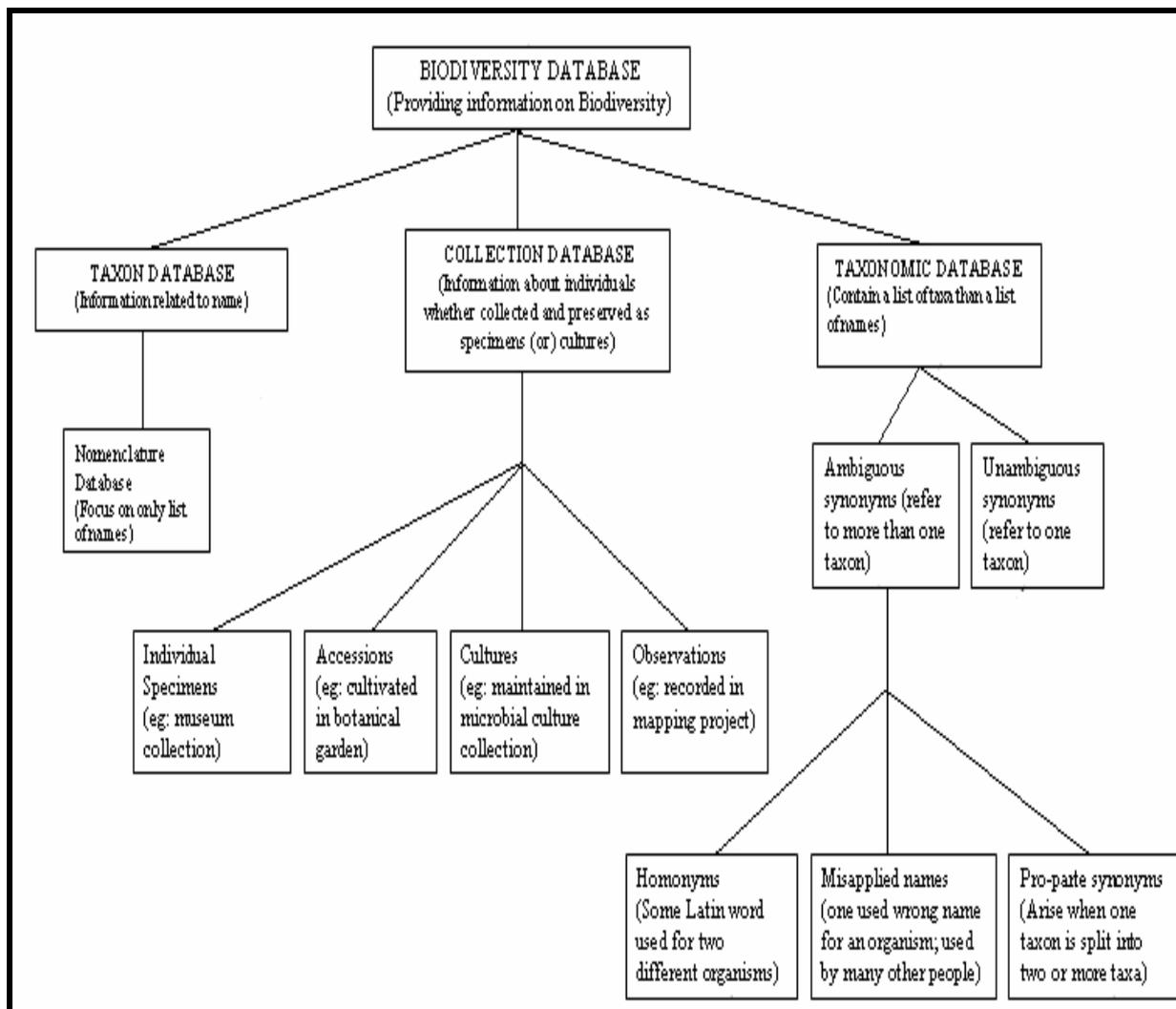


Figure 1: Different types of databases and its importance

Problems in retrieval of biodiversity information

The biological information contained in print media, in outmoded electronic form, and in modern databases constitutes an intellectual wealth produced by decades and centuries of research and considerable societal investment. Further advances in scientific understanding of biodiversity at the gene, organism, population, species, ecological community and landscape at global levels are to be made and the results of the work of the predecessors as well as contemporaries of the world biological scientists should, using the technologies now at our disposal, be made readily and comprehensively available to the current generation of researchers, no matter where they reside. This same information is needed by persons with policy and decision making responsibilities, and there are applications in education, both formal and informal, and industry to which the information could contribute. With the proper investments in infrastructural and software developments, the advantages of modern informatics techniques can be employed to exploit

this intellectual wealth with great benefit not only to biological research, but to decision and policy-makers, educationalists, and society at large. [5] However, advancements in informatics capabilities for biological sciences (in data management, in network connections, and in data content) are still needed.

Conclusion:

Biodiversity is distributed all over the world but the scientific information about biodiversity is largely concentrated in major centers in developed countries, especially in the scientific collections of the world's natural history museums, herbaria, and microorganism repositories. Data about biodiversity are either scattered in many databases or reside on paper or other media not amenable to interactive searching. There is an enormous amount of information already collected about the world's biodiversity. However, to date most of this information has not been digitized. Thus, in most cases, the only way a potential user can find data is to travel physically

to the place where the specimen is housed or to contact the repository where a relevant specimen may be housed. The sustainable use and management of biodiversity will require data about it be available when and where that information is needed by decision-makers and scientists alike. Because biodiversity information is not immediately at hand, it is often not applied in policy or management decisions that affect the organisms involved, nor is that information readily accessible by the users.

References:

- [01] F. A. Bisby, *Biodiversity informatics and Internet. Science*, 283:2309 (2000) [PMID: 11009408]
- [02] R. K. Colwell & J. A. Coddington, *Phil. Trans. R. Soc.*, 335:101 (1994) [PMID: 7972351]
- [03] W. Jetz, & C. Rahbek, *Science*, 297:1548 (2002) [PMID: 12202829]
- [04] J. Soberson & A. T. Peterson, *Phil. Trans. R. Soc. Lond.*, 359:689 (2004) [PMID: 15253354]
- [05] C. Rahbek, & G. R. Graves, *Proc. R. Soc. Lond.*, B267:2259 (2000) [PMID: 11413641]

Edited by P. Kanguane

Citation: Shanmughavel, *Bioinformatics* 1(9): 367-369 (2007)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.