



Review Article

<https://doi.org/10.20546/ijcmas.2018.712.345>

Significance of Bioinformatics in the Conservation of Biodiversity and Databases

Rakesh Ranjan¹, Saket Vinayak¹ and Shanker Kumar Pandey^{2*}

¹University Centre of Bioinformatics (Sub-DIC), T.M. Bhagalpur University,
Bhagalpur- 812007, India

²Department of Botany, Jamtara College, Jamtara, Jharkhand- 815351, India

**Corresponding author*

ABSTRACT

Biodiversity is commonly used for the totality and variability of species, genes and the ecosystems. They occupy at three levels i.e generic, species and ecology. Recently the advancement in bioinformatics is mainly accelerated by the study of molecular events using data obtained by exhaustive experiments such as whole genome sequencing, functional annotation, expression analysis and some others. However, the complete 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 databases. The massive development of biodiversity related information systems over the WWW (World Wide Web) has created much excitement in recent years. Biodiversity information can be considered the basic data on the occurrence and diversity of species (or indeed, any recognizable taxa), commonly in association with information regarding their distribution in space, time, or both. Such information may be in the form of retained specimens and associated information, for example as assembled in the natural history collections of museums and herbaria, or as observational records, for example either from formal faunal or floristic surveys undertaken by professional biologists and students, or as amateur and other planned or unplanned observations including those increasingly coming under the scope of citizen science. Providing online, coherent digital access to this vast collection of disparate primary data is a core. Biodiversity Informatics function that is at the heart of regional and global biodiversity data networks. There is a resonance between the needs of biodiversity science and the opportunities for globalization and interoperability provided by the internet. The main objective of the bioinformatics projects is to conserve the biodiversity and permit data interoperability and knowledge synthesis across wide arrays of local systems, and to embed them in global knowledge. This paper emphasizes the importance of bioinformatics in the conservation of biodiversity databases of scattered data for biotechnological and agricultural research.

Keywords

Bioinformatics,
Biodiversity,
Conservation,
Database, Research

Article Info

Accepted:

20 November 2018

Available Online:

10 December 2018

Introduction

Biodiversity stands for all living things on earth. It refers to the range of variations among a set of entities and is commonly used to describe variety and variability of living organisms in terms of genetic diversity, species diversity and ecological diversity. In simple terms, Biodiversity is the vast variety of natural plant and animal life existing in any region. The biodiversity is usually studied now by taxonomist who takes great pains in collecting, identifying, documenting and describing the elements of diversity. There is a resonance between the needs of biodiversity science and the opportunities for globalization and interoperability provided by the internet. Global biodiversity depends on several parameters such as biomass, ecosystems, phyla, floras and faunas, hot-spots, genetic erosion, the impact of aliens and others. 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. In the last three decades, 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. 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 more than 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 conservation and Bioinformatics

Biodiversity informatics is the application of information technology methods to the problems of organizing, accessing, visualizing and analyzing primary biodiversity data. Primary biodiversity data is composed of names, observations and records of specimens, and genetic and morphological data associated to a specimen. Biodiversity informatics may also have to cope with managing information from unnamed taxa such as that produced by environmental sampling and sequencing of mixed-field samples. The term biodiversity informatics is also used to cover the computational problems specific to the names of biological entities, such as the development of

algorithms to cope with variant representations of identifiers such as species names and authorities, and the multiple classification schemes within which these entities may reside according to the preferences of different workers in the field, as well as the syntax and semantics by which the content in taxonomic databases can be made machine query able and interoperable for biodiversity informatics purposes. 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. 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.

Biodiversity database and Bioinformatics

Taxonomists create the nomenclature and classification databases which can also contain all kinds of information about organisms, including their characteristics, economic importance, conservation and management. 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 for all these purposes. Biodiversity informatics includes the application of information technologies to the management, algorithmic exploration, analysis and interpretation of primary data regarding life particularly the species level organization. One major issue for biodiversity informatics at a global scale is

the current absence of a complete master list of currently recognized species of the world, although this is an aim of the Catalogue of Life project which has ca. 1.65 million species of an estimated 1.9 million described species in its 2016 Annual Checklist. "Primary" biodiversity information can be considered the basic data on the occurrence and diversity of species, commonly in association with information regarding their distribution in either space, time, or both. Such information may be in the form of retained specimens and associated information, for example as assembled in the natural history collections of museums and herbaria, or as observational records, for example either from formal faunal or floristic surveys undertaken by professional biologists and students, or as amateur and other planned or unplanned observations including those increasingly coming under the scope of citizen science. Providing online, coherent digital access to this vast collection of disparate primary data is a core Biodiversity Informatics function that is at the heart of regional and global biodiversity data networks, examples of the latter including OBIS and GBIF.

As a secondary source of biodiversity data, relevant scientific literature can be parsed either by humans or (potentially) by specialized information retrieval algorithms to extract the relevant primary biodiversity information that is reported therein, sometimes in aggregated / summary form but frequently as primary observations in narrative or tabular form. Elements of such activity (such as extracting key taxonomic identifiers, key wording / index terms, etc.) have been practiced for many years at a higher level by selected academic databases and search engines. However, for the maximum Biodiversity Informatics value, the actual primary occurrence data should ideally be retrieved and then made available in a

standardized form or forms; for example both the Plazi and INOTAXA projects are transforming taxonomic literature into XML formats that can then be read by client applications, the former using Taxon X-XML and the latter using the taXMLit format. The Biodiversity Heritage Library is also making significant progress in its aim to digitize substantial portions of the out-of-copyright taxonomic literature, which is then subjected to OCR (Optical Character Recognition) so as to be amenable to further processing using Biodiversity Informatics tools.

Challenges in biodiversity information

Biodiversity scientific information is highly varied and includes quantitative and coded tables, time-series as well as narrative and descriptive text. A common characteristic is that it is often spatially referenced i.e. in map form or with reference to point locations. The needed bioinformatics technology includes database management systems, GIS, image analysis, statistical analysis and modeling. Further there is a need to be able to locate an extract descriptive text, often involving large quantities in disparate locations. This leads to requirements for key wording, indexing, hypertext linking, distributed networking, and for meta-database technology to assist in locating appropriate data sources. 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. However, advancements in informatics capabilities for biological sciences in data management, in network connections, and in data content are still needed.

In conclusion, biodiversity Informatics benefits from the adoption of appropriate standards and protocols in order to support machine-machine transmission and interoperability of information within its particular domain.

Examples of relevant standards include the Darwin Core XML schema for specimen- and observation-based biodiversity data developed from 1998 onwards, plus extensions of the same, Taxonomic Concept Transfer Schema, plus standards for Structured Descriptive Data and Access to Biological Collection Data (ABCD); while data retrieval and transfer protocols include DiGIR (now mostly superseded) and TAPIR (TDWG Access Protocol for Information Retrieval). Many of these standards and protocols are currently maintained, and their development overseen, by the Taxonomic Databases Working Group (TDWG). 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 bioinformatics is very much significant in the conservation of biodiversity and database. Biodiversity information is neither 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

- Bioinformatics for Biodiversity? *Science*. 289: 2229–2440. 2000.
- Bisby FA. Biodiversity informatics and Internet Science. 2000;283:2309.
- Colwell RK, Coddington JA. *Phil Trans R Soc*. 1994;335:101.
- Jetz W, Rahbek C. *Science*. 2002; 297: 1548.
- Peterson AT and Vieglais D (2001). "Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem" (PDF). *BioScience*. 51: 363–371.
- Rahbek C, Graves GR. *Proc R Soc Lond*. 2000; B267: 2259.
- Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T., Brusca, R.C., *et al.*, (2015). "A higher level classification of all living organisms". *PLoS ONE*. 10 (4): e0119248.

How to cite this article:

Rakesh Ranjan, Saket Vinayak and Shanker Kumar Pandey. 2018. Significance of Bioinformatics in the Conservation of Biodiversity and Databases. *Int.J.Curr.Microbiol.App.Sci*. 7(12): 3008-3012. doi: <https://doi.org/10.20546/ijcmas.2018.712.345>