Biodiversity information is critical to a wide range of scientific, educational and governmental uses, and is essential for decision making. Most of the biodiversity information is neither readily available nor accessible. Recent developments in information and communication technology are allowing new experiences in the integration, analysis and visualization of biodiversity information [1]. A formidable challenge that lies ahead is to integrate these initiatives into an organized, well-resourced global approach to build and manage biodiversity data and information through collaborative efforts. The strong demand to integrate, synthesize and visualize this information for different purposes and by different end users is leading to the development of a new field of research, biodiversity informatics. This emerging field has great potential in diverse realms including basic and applied biology (biogeography, ecology and invasive species assessment and monitoring), agriculture (agricultural pests and bio-control measures) and public health (monitoring and control of infectious diseases). This potential, nonetheless, remains yet to be explored, as this field is only now becoming a vibrant area of inquiry and study [2]. Efforts to integrate data into viable resources for innovation in science and technology and decision-making are being developed as local, regional and global initiatives. Examples include the Brazilian Biota FAPESP Virtual Institute of Biodiversity [3], the World Information Network on Biodiversity [4], Australia’s Virtual Herbarium - AVH [5], the European Natural History Specimen Information Network - ENHSIN [6], the Inter American Biodiversity Information Network – IABIN [7], and the Global Biodiversity Information Facility – GBIF [8].

Most initiatives are focusing on biological species and specimen data as the first necessary information component of a global comprehensive data network on biodiversity. As non-biotic environmental and ecological data are increasingly being used in biodiversity informatics for modeling the distribution patterns of species and populations, they also need to be integrated in a transparent manner.

Taxonomic checklists that are being developed worldwide and the 2.5 billion specimen records available in scientific collections are the primary research archives documenting biological diversity on earth. This data infrastructure is the foundation of biology and document identities, habitats, history, and spatial distributions of the roughly 1.75 million described species of life on earth. The specimen vouchers and associated information provide a fundamental resource for biological systematics. Return on investment made over 250 years of global biological inventories, can be dramatically realized through the digitization and integration of species and specimen information. Most of the taxonomic checklists are not digitally integrated and most of the specimen records worldwide remain unavailable in the electronic domain.

In order to speed up the delivery of a comprehensive index of all known plants, animals, fungi, and micro-organisms, the Species 2000 [9] and the Integrated Taxonomic Information System [10] have joined forces to deliver an integrated synonymic species list prepared by worldwide taxonomic experts in the various groups of organisms. One of the principal uses for such a catalog is to index other biodiversity information by species, most particularly including synonymic indexing, where a species may be known by more than one name [11]. This Catalogue is being prepared in
Networking biological diversity information: linking local to global efforts

two editions, an *Annual Checklist* created as a new edition each year, and a *Dynamic Checklist* created as an online, distributed resource working with the contributing databases in real time. The *Annual Checklist* has been available since 2001 on the Web and on CD-ROM, and is now widely used on biodiversity data portals around the world. The fourth release of the Annual Checklist on CD-ROM contains a searchable database with more than 800 thousand scientific and common names of viruses, microorganisms, protists, fungi, animals and plants [12]. The *Dynamic Checklist* is still in prototype form, but will become available during 2004.

Biodiversity data are generated and compiled by thousands of researchers around the world. The result is a global array of distributed and heterogeneous biodiversity information resources. Integrating these data has been a goal of biodiversity informatics community for more than 10 years. Following the information and communication technology, the integration of distributed databases can now be dynamically processed according to users’ needs.

In order to integrate data from distributed sources in real time the adoption of standards and protocols is essential. The International Union of Biological Sciences Taxonomic Database Working Group – TDWG [13], in a joint initiative with the Committee on Data for Science and Technology – CODATA [14], has been working for several years on a standard called Access to Biological Collection Data – ABCD [15]. The main purpose of ABCD is to define global formats to exchange data and retrieve them from all kinds of biological collections. In practical terms, a highly structured XML schema has been defined in agreement with representatives from different projects and communities. The current version contains reusable data types and specific elements related to each different biological discipline, ranging from palaeontological collections to collections of living organisms. A more recent standard being used by many federated database networks of biological collections, known as Darwin Core, is a simple XML schema with a small number of elements covering mainly current taxon identification, some details about the collecting event (e.g. collector, location and date) and other basic information [16]. Both ABCD and Darwin Core are federated schemas that need a protocol to be used in networked systems. The Distributed Generic Information Retrieval (DiGIR) is an XML based protocol capable of working with a configurable federated schema [17]. The use of generic query elements and customizable record structures enabled not only decoupling from semantics, but also the possibility to develop wrapper software to encapsulate the heterogeneity of data providers. DiGIR is being adopted by several distributed networks of scientific collections worldwide, including GBIF.

The Brazilian Biota/Fapesp Virtual Institute of Biodiversity program officially launched in March 1999, aims at collecting, organizing and disseminating biodiversity information of the State of São Paulo, defining mechanisms for its conservation and sustainable use. The program is currently funding nearly 50 projects, involving more than 400 scientists and an equal number of students working at local research institutions and abroad. The information systems that are being developed for the program are SinBiota and speciesLink.

*SinBiota* is an information system developed to support the geo-spatial analysis and visualization of observational data. It is a centralized information system developed to receive and integrate observational data from associated projects and external data dynamically through the Internet. The use of standard field records developed by the scientific community and an associated list of species and geocoding of all observations is mandatory. A digital map base with a number of environmental layers integrated with the observational database is one of the key outlets for disseminating information of the program, the Atlas Biota [18]. PostgreSQL [19] is used as relational database and the Atlas is based on MapServer [20] both open source developments.
The speciesLink [21] network is a distributed information system integrating primary specimen data from biological collections at real time. The project makes use of the most current advances in communication protocols such as DiGIR - Distributed Generic Information Retrieval, and SOAP - Simple Object Access Protocol [22] and database management systems. The information system is being developed using free and open source software, mirroring techniques, and Internet 2 connectivity. The project is developing tools to help in the cleaning, utilization and visualization of data from 36 associated collections. The tools that are being developed include the openModeller [23], a species distribution modeling framework, spOutlier [24], an on-line tool to detect outliers in latitude, longitude and altitude, and geoLoc [25], a geo-referencing tool.

Another major product from the Biota FAPESP program is the Biota Neotropica [26], a peer reviewed on-line journal focused on the dissemination of neo-tropical biodiversity information.

Science is growing in size and complexity and is becoming more cooperative and cumulative. As processed information is more precise and systematically structured, new developments in information and communication technologies (ICT) are playing a significant role in science and innovation. The use of ICT is broadening the scope and scale of science and bringing new opportunities for international collaboration [27].

Good scientific information is fundamental for sound environmental decision making and the design of mechanisms to link scientific research to the decision making process is no easy matter [28]. Biodiversity informatics developments will directly benefit environmental education programs, resource management and conservation and biomedical and agricultural research.

REFERENCES AND URLS
[3] Biota/Fapesp Virtual Institute of Biodiversity program http://www.biota.org.br
Networking biological diversity information: linking local to global efforts

[21] speciesLink http://splink.cria.org.br
[23] openModeller http://openmodeller.cria.org.br
[26] Biota Neotropica http://www.biotaneotropica.org.br