

Global Biodiversity Information Facility Strategic Plan
October 2003

FOREWORD

The Global Biodiversity Information Facility (GBIF) is an open, independent organisation dedicated to making the world's biodiversity data freely and universally available via the Internet. Importantly, GBIF is open to participation by *all* countries, economic entities, and organisations that can benefit by the open sharing of biodiversity information on a global scale.

The concept of GBIF was developed by the Subgroup on Biodiversity Informatics of the OECD Megascience Forum Working Group on Biological Informatics. The Subgroup had met seven times between June 1996 and January 1999 to formulate their recommendations, which can be obtained at http://www.gbif.org/GBIF_org/facility/Blrepfin.pdf. (See also Box 2 in this document.) OECD Science Ministers endorsed the Working Group's report in June 1999, and explicitly recommended that GBIF be an independent organisation, open to any country or relevant organisation.

With the OECD Ministers' endorsement in hand, an Interim Steering Committee for GBIF met four times between July 1999 and December 2000 in order to negotiate the GBIF Memorandum of Organisation (MOU). GBIF came into being on 1 March 2001, when the conditions of the MOU were met. Full text of the MOU can be found at http://www.gbif.org/GBIF_org/facility/mou.html.

Although the work of GBIF is long-term and should be continued far into the future (the OECD Subgroup report planned for a duration of at least ten years), the initial MOU is in effect for five years. At the end of the third year (2004), there will be an external evaluation of GBIF's structure and progress toward its goals. As a result of this review, "course corrections" may be undertaken.

While reading this document, it is important to remember that GBIF only became a reality in March of 2001, and that the months that followed necessarily were devoted to selecting a site for the Secretariat, putting in place the Secretariat facilities and staffing the Secretariat with personnel selected after worldwide searches. The first Work Programme for GBIF was approved by the Governing Board in October, 2002, and went into effect in January, 2003. Thus, we regard this 5-year Strategic Plan as beginning in 2003.

This document is about the plans for building a megascience facility that will

- ☐ enable scientific research that has never before been possible,
- ☐ facilitate the use of scientific data in biodiversity policy- and decision-making, and
- ☐ make a whole world of biodiversity information--data that are currently exceedingly difficult to access-- freely and universally available via the Internet.

Global Biodiversity Information Facility Strategic Plan

EXECUTIVE SUMMARY

Biodiversity and biodiversity information are essential to a sustainable future for society. Moving scientific information into a wide array of uses (policy, management, conservation and education in addition to research) is a megascience activity. GBIF is a megascience facility that will

- enable scientific research that has never before been possible,
- facilitate the use of scientific data in biodiversity policy- and decision-making, and
- make a whole world of biodiversity information--data that are currently exceedingly difficult to access-- freely and universally available via the Internet.

GBIF will do for biodiversity information what the printing press did for the sharing of recorded information during the Renaissance — it will make recorded knowledge the common property of everyone, not just the elect few. Even better, GBIF will do it electronically, so the resource will be dynamic, interactive, and ever-evolving. Within five years, we expect that GBIF's portal will be the most-used gateway to biodiversity and other biological data on the Internet.

Sharing of economic and scientific benefits provided by biodiversity will be facilitated by GBIF's provision of access to scientific data and information about biodiversity that will be as easy and complete in Mongolia or Madagascar as it is in the U.K. or the U.S.

The Global Biodiversity Information Facility was established to carry out specific tasks that are essential to a worldwide infrastructure that can overcome current barriers to the universal availability of biodiversity information. It will, among other goals, improve the accessibility, completeness and interoperability of biodiversity databases; facilitate development of an electronic catalogue of the names of known organisms; develop curricula for biodiversity informatics training; implement programmes to enhance biodiversity informatics capacity in developing countries; and help to co-ordinate the biodiversity informatics programmes of its Participants.

To accomplish these goals, GBIF's activities are organized around six integrated thematic areas: Data Access and Database Interoperability, Digitisation of Natural History Collections, Electronic Catalogue of the Names of Known Organisms, Outreach and Capacity Building, *SpeciesBank*, and Digital Biodiversity Literature Resources. The purposes of these programmes are 1) To facilitate the full use of biodiversity and other databases by establishing an information architecture that enables interoperability and facilitates data-mining; 2) To facilitate the expansion of biodiversity knowledge by having legacy and newly acquired primary species occurrence data digitised and dynamically accessible; 3) To make integrated searching possible, as well as to facilitate the exploration and rapid expansion of biodiversity knowledge, by providing a complete, digital listing of the names of all known organisms; 4) To bridge biodiversity information technology "digital divides" through training and capacity building to ensure that people in every country have access to and can easily and freely use the world's biodiversity information; 5) To provide, in real time, a complete compendium of knowledge about particular species, including name and synonyms, distribution, natural history, physiology, etc., drawn from online information sources; and 6) To enable Web access to digitised versions of the published literature extending back in time at least to Linnaeus' publications of the 1750s, which are the basis of the system for scientific naming of organisms that is now in use.

The scope of GBIF's activities includes working toward making databases of information about all levels of biodiversity (genes, species, ecosystems) interoperable. For the first few years, GBIF is highly focused on those tasks for which it was specifically designed, which will bring species- and specimen-level data to the Web. The Participant Nodes are the key to GBIF's success; the Secretariat assists the Nodes to carry out the challenging tasks ahead.

GBIF's combined strategy of 1) focus on mission and specific goals, 2) inclusiveness in the manner in which it seeks advice, 3) openness in data sharing and software developments, 4) cost-effective partnerships, and 5) adherence to founding principles bodes extremely well for its

success and continued evolution toward the vision of its Participants. GBIF has the potential to advance by orders of magnitude our ability to exploit the Web's power, giving society true, worldwide, manipulable biodiversity information-at-our-fingertips, which will contribute to scientific innovation and progress *and* towards a sustainable society.

TABLE of CONTENTS

FOREWORD	ii
EXECUTIVE SUMMARY	iii
INTRODUCTION	1
Science.....	1
Society	1
Sustainability.....	1
THE GBIF VISION	2
Box 1: <i>GBIF, 2011</i>	2
THE GBIF MISSION	3
Box 2: <i>GBIF, 2001</i>	3
The Scope of GBIF's Activities.....	3
Figure 1: <i>Piecing Together a Complete Information Domain</i>	4
Box 3: <i>No Time to Compete</i>	5
Structure of the Facility	5
The Goals of GBIF	6
Sponsorship of GBIF Activities.....	8
THE FUTURE OF GBIF	9
APPENDIX 1: <i>List of Acronyms Used</i>	10
APPENDIX 2: <i>Strategic Outcomes of the Thematic Programmes</i>	11

INTRODUCTION

The Global Biodiversity Information Facility was established and is being maintained by an ever-growing consortium of forward-thinking countries, economies and international organisations. Participants in GBIF¹ recognise the importance of biodiversity and biodiversity information to a sustainable society. Thus, they have signed an MOU, agreeing that

“...a co-ordinated international scientific effort is needed to enable users throughout the world to discover and put to use vast quantities of global biodiversity data, thereby advancing scientific research in many disciplines, promoting technological and sustainable development, facilitating the equitable sharing of the benefits of biodiversity, and enhancing the quality of life of members of society. The importance of making biodiversity data openly available to all countries and individuals is underscored by various international agreements, especially the Convention on Biological Diversity.”²

The rationale for a megascience organization such as GBIF derives both from science and from society's need to build a sustainable future.

Science

Significant pieces of the tripartite (gene, species, ecosystem) biodiversity information resource are already online, such as the DNA data served by GenBank, EMBL and DDBJ, and other sequence (RNA, protein, etc.) data served by various sources such as ANGIS, RNABase, SCOR, SwissProt, and ExPASy. These persistent data stores, which are community research resources, have contributed substantially to medical, pharmaceutical and agricultural industries, and through these to society. Ecological, ecosystem and planet-wide data are provided online by, for example, MABnet, EIONet and (International) LTERnet, and other international and bi- or trilateral consortia. These online ecological resources are contributing to predictive modelling of global climate change and other large-scale ecological phenomena.

However, in almost all cases, the keystone of biodiversity information (that about individuals, populations and species of whole organisms) is not digitally available. These data are on the labels of natural history specimens, in libraries, and in handwritten notebooks or typewritten card files. If digitised, three hundred years' records of the life of planet Earth, a resource of inestimable value, would become available for many sorts of users and many kinds of uses; they would make a “GenBank equivalent” for biodiversity scientists—a persistent but dynamically accessible data store.

Society

The beauty of digital data is that they can be used and re-used so easily; such data can also be employed at more than one point on the globe at any one time to answer very different types of questions. Investments made in digitising scientific data and bringing them online in a well-planned information architecture are paid back many times over. Not least of these repayments is that digital data are usable by people other than scientists.

Sustainability

The ultimate source of global and national wealth is natural resources. Biodiversity, or “living natural resources”, provides an ever-increasing portion of that wealth. Biodiversity, as defined by the Convention on Biological Diversity³, has provided the basis of human survival (clean water and air, food, fuel, and fibre), not to mention prosperity, since *Homo sapiens* first set foot on Earth. If we humans are to continue to prosper and to leave future generations a healthy place in which to live, we must learn to use living resources in a sustainable way, and make it possible for all peoples to share in the benefits of the

¹ For a current list of GBIF Participants, see http://www.gbif.org/GBIF_org/participation

² Global Biodiversity Information Facility Memorandum of Understanding: Paragraph 1.

³ “Biodiversity” or “Biological diversity” means the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems (Convention on Biological Diversity, Art. 2, para 1).

sustainable use of biodiversity. The data made available by GBIF will help to form a foundation for measuring achievement of the goal of reducing the rate of biodiversity loss by 2010, as called for by the World Summit on Sustainable Development.

THE GBIF VISION

GBIF will do for biodiversity information what the printing press did for the sharing of recorded information during the Renaissance – it will make recorded knowledge the common property of everyone, not just the elect few. Even better, GBIF will do it electronically, so the resource will be dynamic, interactive, and ever-evolving. Within five years, we expect that GBIF's portal will be the most-used gateway to biodiversity and other biological data on the Internet.

GBIF will make possible simultaneous searching of databases at multiple levels of biological organisation, which in turn will make possible true 'data-mining' of biological information, and thus open new avenues of scientific investigation.

Sharing of economic and scientific benefits provided by biodiversity will be facilitated by GBIF's provision of access to scientific data and information about biodiversity that will be as easy and complete in Mongolia or Madagascar as it is in the U.K. or the U.S.

BOX 1: *GBIF, 2011*

Our vision is that ten years after its inception, GBIF will have become the premier biodiversity information source for the widest possible array of users. In 2011, the GBIF portal¹ is serving, at his, her or their desktop(s)...

- ... *a bureaucrat in the Ministry of Health* who needs to know what species of insects occur in her country that might become vectors for an emergent disease.
- ... *a graduate student in southeast Asia* who needs to know all the names (scientific and vernacular) that have ever been applied to a plant species from tropical Africa that he is studying for his thesis on its physiology and potential for cultivation.
- ... *a conservation biologist* who needs to understand the habitat requirements and naturally co-occurring species of an endangered species of primate (or whale, otter, orchid, parrot ...)
- ... *a molecular genetics researcher* who is looking for a gene in a member of the agriculturally important grasses that is analogous in function to one she has found in goosefoot that allows the plant to tolerate dry, hot, saline conditions.
- ... *a PhD taxonomist in Europe*, who is beginning the daunting task of monographing a genus of primarily tropical beetles that comprises at least 1,000 species.
- ... *a pharmaceutical chemist* who has found a promising drug compound in a fungus species, and would like to know if there are related species that produce similar compounds or the same compound in greater quantities.
- ... *a lawyer for a group of indigenous peoples* who needs to establish the exact identity of a plant on which they claim rights of intellectual property.
- ... *an elementary-school class* that is studying praying mantises, and wants to know about their prey and their predators.
- ... *a robotics researcher* who needs inspiration from nature about how to solve a particular engineering problem.
- ... *a curator at a storage site for genetic resources* who needs to know whether the tissue samples from an organism of interest in her collection are a thorough sampling from throughout the range of the species.
- ... *a government agency* that must interrogate multiple large datasets in order to set aside a biodiversity reservation that will preserve the largest possible number of species within the smallest possible area, while still providing opportunities for ecotourism and sustainable harvest of wild products.
- ... *a natural resource manager* who needs to prevent the advance or combat the depredations of invasive species.
- ... *a parataxonomist in Latin America*, who needs to identify the specimen he has just collected, and determine if it is known or is new to science.
- ... *and so on, and on, and on*. The applications and utility are endless, and of inestimable value. Because many different kinds of interfaces that each serve a different audience can be developed to access the same data resource, this one focused effort to provide primary data about biodiversity is an investment that by 2011 has paid off in multiple ways—and the payoffs will continue far into the future beyond 2011.

¹ The portal is a stable computing gateway that is founded on 1) common standards for data and metadata, 2) an information architecture that includes a registry of shared information, 3) a 90% complete catalogue of all the scientific names that have ever been published, and 4) hundreds of millions of digital specimen and observation records held by hundreds of globally distributed institutions.

THE GBIF MISSION

Make the world's biodiversity data freely and universally available via the Internet.

The Global Biodiversity Information Facility was established (see Box 2: **GBIF, 2001**) to carry out specific tasks that are essential to a worldwide infrastructure that can overcome current barriers to the universal availability of biodiversity information.

BOX 2: **GBIF, 2001**

GBIF was conceived by an international group¹ of scientists and administrators convened in January, 1996, by the OECD Megascience Forum as a Subgroup for Biodiversity Informatics of the Working Group on Biological Informatics. After nearly three years and seven meetings, as well as frequent consultations with existing biological informatics efforts and potential partners, the Biodiversity Informatics Subgroup concluded that:

- ☐ The biodiversity information domain is vast and complex, and critically important to society.
- ☐ At present, existing biodiversity information is neither readily accessible nor fully useful.
- ☐ Recent technological and political developments present leadership opportunities for OECD countries.

In its final report² to the OECD, the Subgroup proposed that OECD member countries should provide leadership to establish a global mechanism that would make biodiversity data and information openly accessible worldwide. The OECD Ministers accepted the Subgroup's report and endorsed its recommendation. They encouraged establishment of an international coordinating body that would implement a Global Biodiversity Information Facility (GBIF). As a result, representatives from 17 interested countries participated in an *ad hoc* meeting on Implementation of GBIF in March 1999. Outcomes from that meeting guided the efforts of the GBIF Interim Steering Committee, which met three times over the next year (delegates from 15 countries attended each of these meetings). The work of the Interim Steering Committee resulted in the Global Biodiversity Information Facility Memorandum of Understanding (MOU), which was opened for signature in December, 2000. The initialising conditions of the MOU were met by March, 2001, and GBIF was born.

1 Representatives from 21 countries (Australia, Belgium, Canada, Denmark, Finland, France, Germany, Israel, Italy, Japan, Korea, Mexico, Netherlands, Norway, Poland, Portugal, Russia, Sweden, UK, USA) and the European Commission.

2 Final report of the OECD Megascience Forum Working Group on Biological Informatics. January, 1999. 74 pp. See http://www.gbif.org/GBIF_org/facility/Birepfin.pdf

The MOU lays out the understanding of the Participants that GBIF is an open-ended coordinating body, as well as the principles of GBIF, which are that it will:

- ☐ be shared and distributed, while encouraging co-operation and coherence;
- ☐ be global in scale, though implemented nationally and regionally;
- ☐ be accessible by individuals anywhere in the world, offering potential benefits to all, while being funded primarily by those that have the greatest financial capabilities;
- ☐ promote standards and software tools designed to facilitate their adaptation into multiple languages, character sets and computer encodings;
- ☐ serve to disseminate technological capacity by drawing on and making widely available scientific and technical information; and
- ☐ make biodiversity data universally available, while fully acknowledging the contribution made by those gathering and furnishing these data.

Adherence to these principles is specifically intended to achieve benefits both for science and for society.

The Scope of GBIF's Activities

All of GBIF's tasks are aimed at making data about individuals, populations and species digitally available. These are the data that tie the whole biodiversity information pattern together (see Figure 1). Without the services that will be provided by GBIF, there is no way to obtain a transparent and seamless response to a query that requires a call upon both ecological and genetic data as well as species data, such as *"What other organisms live in the same kinds of habitats as this one from which I have extracted a gene that enables it to tolerate high levels of lead acetate, and are able to do so because of either homologous or analogous gene function?"*

At present, biodiversity databases are for the most part not interoperable, either with each other or with databases in the genetic or ecological domains. Interoperability is a desirable goal because it will enable the apparently simple (but actually complex) domain-wide querying described above. Attainment of interoperability will require certain key elements. These are

- 1) a complete electronic catalogue of scientific names,
- 2) digitisation of the label data from ever-increasing percentages of the world's natural history specimens and cultured microorganisms (and eventually digitised versions of biodiversity literature resources), and
- 3) an informatics infrastructure to link these together. This infrastructure will comprise
 - a. standards for data and metadata,
 - b. software developments, and
 - c. increases in distribution, on a worldwide basis, of Internet connectivity and means of access to the Internet.

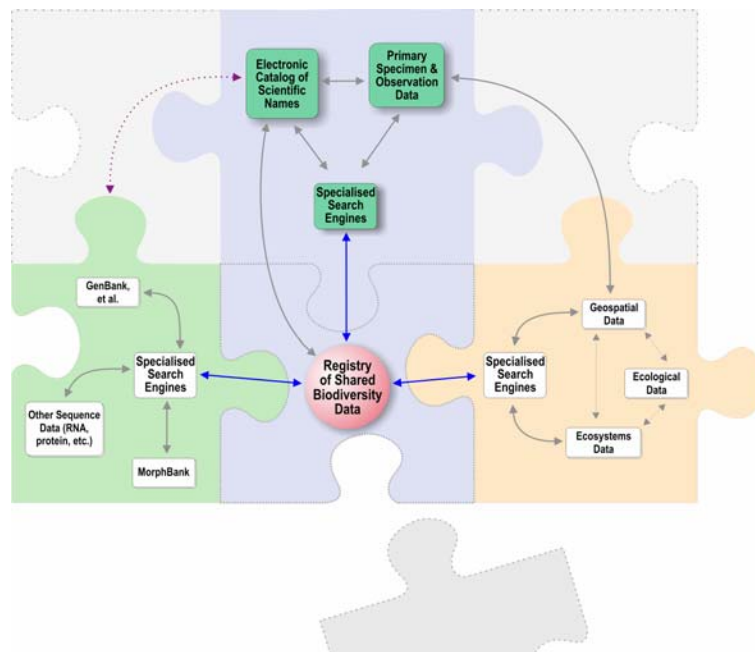


Figure 1. PIECING TOGETHER A WHOLE INFORMATION DOMAIN: The tasks that GBIF will carry out provide the centre pieces that tie other, existing parts of the digital biodiversity information domain together. Biodiversity is typically regarded as existing at three levels of organisation: genes (including molecular data—the green piece at the left of the diagram), species (including information about individuals and populations of organisms—the central, blue pieces), and ecosystems (the yellow piece at the right). Each of the three subdomains of the biological information domain will want to generate its own specialised search engines, but the taxonomic name service that centres around the Electronic Catalogue of Scientific Names and the registry of biodiversity data are unique GBIF tasks that will facilitate simultaneous searching across all levels of biodiversity. This is because scientific names are often the only data field(s) common to databases at all three levels.

GBIF has clearly identified the specific areas in which actions at a global level are needed to further biodiversity informatics, and has forged partnerships with other, existing organisations whose goals are similar. For instance, GBIF has formed a partnership with the Catalogue of Life consortium (Species 2000 and ITIS) that is aimed at increasing the rate of progress in producing a complete, concept-based listing of scientific names, and another partnership with the Taxonomic Databases Working Group (TDWG) to develop consensus on standards for data and metadata.

GBIF focuses only on the niche it was designed to fill. Avoidance of competition (see Box 3: *No Time to Compete*) with other organisations is cost-effective and time-efficient.

Cooperation and partnerships enhance the goals of all concerned. GBIF also solicits advice from as wide a constituency as may wish to comment, through the bringing together of Scientific and Technical Advisory Groups (STAGs), the membership of the Governing Board's Science Committee and the Participant Node Managers Committee (NODES).

BOX 3: NO TIME TO COMPETE

Biodiversity is disappearing from our planet at a frightening rate no matter how that rate is measured. The 187 Parties to the Convention on Biological Diversity recognise that humankind must use what it already knows, and learn more, about how humans should interact with the natural world that sustains us. There are many organisations and entities that are contributing to the breadth of knowledge we need. Each of these, just like species in their habitats, have a niche to which they are particularly suited. These pieces of the puzzle, when put together digitally using the Internet, create a whole that is not only greater than any of its parts, but also greater than the sum of its parts.

GBIF's niche is making scientific, species-level biodiversity data (primarily from collected specimens and authoritative observations) openly and globally accessible via the Internet. In doing this, GBIF

- ☐ provides information that enables the development of sustainable environmental policy;
- ☐ supports decision-making about biodiversity resource management;
- ☐ facilitates scientific research that contributes to these areas as well as to the accumulation of new knowledge about biodiversity for its own sake; and
- ☐ forms the essential bridge between the genetic and ecosystem levels of biodiversity information.

The informatics challenge of making the primary scientific data about species that has accumulated over the past 250 years in museums and libraries is, if not overwhelming, at least daunting even were GBIF to be funded at a level ten times that of its current budget. GBIF is, therefore, tightly focused on its own mandate. It cannot perform the tasks of others. There are not enough resources to allow that kind of competition. Even were the resources available, there is no time within which to allow competition to take place. Biodiversity is disappearing too rapidly. Ours is no time for competition; rather, it is a time for concerted, coordinated cooperation.

Seeking ways to leverage its efforts and limited funds, GBIF has formed partnerships, not least with the Convention on Biological Diversity. By coordinating and cooperating, GBIF and its partners are making more rapid progress toward common goals than might have been imagined even a few years ago.

Structure of the Facility

Unlike other megascience facilities that are built of bricks and mortar, GBIF operates as a virtual facility. The “bricks” of this facility are the databases, other information resources, and informatics tools made available by GBIF Participants. The “mortar” that holds the bricks together is the shared informatics infrastructure (software tools, operational protocols, and the Internet).

Key to GBIF's operations are the Participant Nodes and their associated databases (the “bricks”). In signing the MOU, Participants agree to establish and maintain at least one GBIF Node, which is defined as “a stable computing gateway that allows real-time inter-operational search of multiple institutional, national, regional and/or subregional databases containing primary or meta-level biodiversity data”. Participants agree to openly share biodiversity data held by their country or organisation. Implicit in this is the promise to fund within-country activities that will help to achieve the digitisation of biodiversity information from museums and libraries.

Represented through the NODES Committee, consisting of the Node Managers of all the Participants, the Nodes are “at the coal face” of information provision. They interact outwardly, with each other and the Secretariat, and inwardly, with the data providers in their countries and organisations. The Participant Nodes are the channels through which GBIF data and information will flow; their success is fundamental to fulfilling the promise of GBIF.

The GBIF Secretariat helps to provide the “mortar” by developing a Work Programme that provides standards for linking the data contained in the Nodes, helps to develop the content provided by the Nodes, and works with GBIF’s partner organisations to produce needed central resources, such as the Electronic Catalogue of Names of Known Organisms.

Goals of GBIF

By signing of the MOU, GBIF Participants have endorsed the following overall goals for GBIF:

- ☐ Improve the accessibility, completeness and interoperability of biodiversity databases, by:
 - ☐ Providing access to new and existing databases;
 - ☐ Contributing data and technical resources, within an intellectual property rights framework (such as that described in MOU Paragraph 8);
 - ☐ Developing novel user interface designs that incorporate features to support their functionality in a multi-lingual global context; and
 - ☐ Developing suitable tools and standards for accessing, linking and analysing new and existing databases, including standards and protocols for indexing, validation, documentation and quality control in multiple human languages, character sets and computer encodings;
- ☐ Facilitate development of an electronic catalogue of the names of known organisms;
- ☐ Design and implement SpeciesBank;
- ☐ Develop a digital library of biodiversity data;
- ☐ Build partnerships with other relevant organisations and projects;
- ☐ Improve high-speed networking and computation infrastructures;
- ☐ Share computational facilities, including high-volume data storage;
- ☐ Develop model curricula for biodiversity informatics training;
- ☐ Provide training for data managers and other relevant staff;
- ☐ Implement specific programs to enhance the biodiversity informatics capacity and technical skills base of developing countries; and
- ☐ Help to co-ordinate and harmonise the biodiversity informatics programs of the Participants.

To achieve these goals, the GBIF Work Programme is organised among six thematic programmes:

- ☐ Data Access and Database Interoperability (DADI).
- ☐ Electronic Catalogue of Names of Known Organisms (ECAT).
- ☐ Digitisation of Natural History Collections Data (DIGIT).
- ☐ Outreach and Capacity Building (OCB).
- ☐ SpeciesBank
- ☐ Digital Biodiversity Literature Resources

In the first three to five years of GBIF operations, the first four programme areas listed have priority. The Work Programme for each year of GBIF⁴ contains goals and mechanisms for each programme for that year. During 2004 (the third year of GBIF), initial discussions about planning

⁴ See http://www.gbif.org/GBIF_org/wp

for “SpeciesBank” will take place, and in 2004 or 2005 discussions will be had with the digital library community and the nascent Biodiversity Commons about how developments leading to truly “Digital Biodiversity Literature Resources” might best be undertaken.

The long-term visions and short to mid-term goals for each of the current four focal programmes are dictated by the mission and goals of the organisation, and the specific tasks that GBIF has identified as needing to be undertaken on a global basis in order to achieve those goals (see Appendix 2).

Programme	Vision	Goals / Tasks
DADI (with GBIF Secretariat ICT staff)	To facilitate digital biodiversity science and thus the full use of biodiversity and other databases by establishing an information architecture that enables interoperability and facilitates data-mining.	☐ Establish standards for data and metadata
		☐ Develop algorithms to search multiple databases simultaneously
		☐ Link species databases with genetic and ecological databases
		☐ Link biological databases with non-biological ones.
DIGIT	To promote digital biodiversity science and to facilitate the expansion of biodiversity knowledge by having legacy and newly acquired primary species occurrence data digitised and dynamically accessible.	☐ Discover extent of already-digitised collections
		☐ Make seed-money awards to speed progress of digitisation
		☐ Develop “best-practices” manual
		☐ Encourage Participants to fund museums and libraries in-country to speed digitisation
ECAT	To make integrated searching possible, promote digital biodiversity science, and facilitate biodiversity exploration and rapid expansion of knowledge by providing a complete, digital listing of the names of all known organisms.	☐ Provide content infrastructure to enable searches across multiple information domains
		☐ Make seed-money awards to speed progress of Catalogue development
		☐ Develop the Taxonomic Name Service function of GBIF information architecture
OCB	To reach out to all countries, and to bridge biodiversity information technology “digital divides” through training and capacity building to ensure that people in every country have open access to and can efficiently use biodiversity information.	☐ Develop biodiversity informatics training courses
		☐ Work with other organisations to overcome “digital divide” that exists around the world
		☐ Increase GBIF Participation by 10% each year
		☐ Increase awareness of GBIF among potential partner organisations
		☐ Encourage and promote synergies with biodiversity-related conventions
		☐ Support the further financial expansion of the Supplementary Fund

Long-term visions for the fifth and sixth programmatic areas are:

SpeciesBank - To provide, in real time, a complete compendium of knowledge about particular species, including name and synonyms, distribution, natural history, physiology, etc., drawn from online information sources.

Digital Biodiversity Literature Resources - To enable Web access to digitised versions of the

published literature extending back in time at least to Linnaeus' publications of the 1750s, which are the basis of the system for scientific naming of organisms that is now in use.

The programmes, the ICT staff of the GBIF Secretariat, and NODES work together in an integrated fashion to achieve GBIF goals. For instance, the work of DADI and ICT is key to the success of both ECAT and DIGIT; likewise the work of ECAT is fundamental to the successful use of digitised natural history collection data (DIGIT). OCB works with ICT, DADI, DIGIT and ECAT to provide targeted workshops and training sessions on topics identified and prioritized by Participants. Ultimately, SpeciesBank will depend heavily on the work of DADI, ECAT, DIGIT and doubtless the digital biodiversity literature resources; again, OCB will assist with workshops and training, while carrying out its other function of encouraging increased Participation in GBIF.

Details of deliverables and timelines for the four current focal programmes are provided in GBIF Work Programme documents, and therefore are not repeated here. However, Appendix 2 gives best-estimate timeframes for certain important activities of each programme.

Sponsorship of GBIF Activities

Most of the funds expended by GBIF Participants on GBIF activities are in-country or intra-organisational investments that capitalise on existing investments. Countries that participate in GBIF experience benefits from the establishment of a coherent national informatics structure that links to and supports the international efforts of GBIF.

On the other hand, GBIF Participants also should be aware of how their individual actions fit within the global GBIF architecture, and therefore understand how their investments affect the GBIF goals of digitisation of specimen data, expansion of network capabilities, and the construction of a scalable digital information facility that can accommodate future additions and changing needs. A GBIF document entitled *GBIF Node Requirements and Recommendations* has been produced to assist Participants in this regard.

Together, the financial contribution to the coordinating function of the GBIF Secretariat and the in-country investments will accomplish a truly megascience endeavour that no one country or organisation could ever accomplish alone. All will benefit, and both science and society will be well served. However, GBIF coordination activities and the Work Programmes could be carried out faster and on a larger scale if the funding was greater.

For this reason, the GBIF Governing Board established a Supplementary Fund. Contributions to this fund, in contrast to the general fund, may be made by not only by countries, but also by foundations, institutions, organisations or individuals. Examples of the purposes of the fund include:

- ☐ Assisting representatives of developing countries to attend important GBIF meetings
- ☐ Building informatics infrastructure (e.g., in developing countries, natural history institutions, the GBIF Secretariat)
- ☐ Furthering one or more of the Programme areas, such as organising training courses, and enabling developing country representatives to attend (OCB); providing support for the digitisation of collections of particular interest to the donor (DIGIT); enabling completion of a Global Species Database for a group of organisms of interest to the donor (ECAT); encouraging developments in information technology or information science that improve access to biodiversity data (DADI).
- ☐ Supporting interns or other visitors to work for a period of time with GBIF Secretariat staff
- ☐ Endowing the Ebbe Nielsen Prize for innovation in combining biodiversity informatics and biosystematics research

Donors to the GBIF Supplementary Fund may be assured that their funds will be carefully husbanded and properly used, because the Fund is maintained in accounts separate from those of the GBIF core budget, overseen by the Executive Committee of the GBIF Governing Board, and audited annually by an independent, international auditing firm that reports to the GBIF Governing Board and to donors to the Fund.

THE FUTURE OF GBIF

Transforming a 500 year tradition of (slow) information transfer by lines of type on paper into a digital (rapid) interchange among thousands of distributed databases, while at the same time dealing with the complexities of the information itself as well as the means of handling it, is no simple task, nor can this task possibly be accomplished in only five years.

This is why GBIF is a *megascience* activity, and why the Work Programmes of these early GBIF years promise “phase one” and “evolving” implementations of mechanisms to achieve the goals set out by the Governing Board.

Just as the Internet itself is evolving into a more and more useful tool, and becoming more and more integrated into the daily life of global society, so GBIF will evolve into a more and more useful gateway into the world of biodiversity information.

The burgeoning of information technology developments has already moved the mechanism of development of some GBIF activities away from the manner in which they were envisioned when the MOU was being designed. Doubtless, this will continue to happen, and GBIF staff and Participants, especially the Node Managers, will need to be flexible and adaptable in taking advantage of better and better information technology tools as these become available.

This will be particularly important as the immediate goals of the first years are accomplished, and possible new inclusions in the Work Programme are adopted. Some of these possibilities are:

- ☐ Creating the software and protocols that will enable interoperability among genetic, species-level, and ecosystem databases;
- ☐ Fleshing out the two remaining work programme areas: designing and implementing SpeciesBank and developing a digital library of biodiversity data and literature resources;
- ☐ Working with authors, publishers and other data providers to set up on-line data-entry services (not just concentrating on the legacy data already present in natural history collections);
- ☐ Exploring the possibility for GBIF to host on-line community monograph projects and other similar resources;
- ☐ Working with scientific societies to get their support for on-line publication of new taxonomic descriptions;
- ☐ Discussing with the various nomenclatural committees the possibility for GBIF to work with other organisations to act as a registrar of new taxonomic names and to make available on-line all new species descriptions;

These are challenging and exciting goals that are not without potential pitfalls. However, GBIF has been structured precisely to be able to take on these challenges, and is already making significant progress in interoperability developments, standards adoptions, the growth of the Electronic Catalogue of Names, the digitisation of specimen data, and outreach to potential new Participants as well as information providers at the gene and ecosystem levels of biological organisation.

GBIF's combined strategy of 1) focus on mission and specific goals, 2) inclusiveness in the manner in which it seeks advice, 3) openness in data sharing and software developments, 4) cost-effective partnerships, and 5) adherence to founding principles bodes extremely well for its success and continued evolution toward the vision of its Participants. GBIF has the potential to advance by orders of magnitude our ability to exploit the Web's power, giving society true, worldwide, manipulable biodiversity information-at-our-fingertips, which will contribute both to scientific innovation and progress, and towards achieving a sustainable society.

Appendix 1: List of Acronyms Used

ANGIS	Australian National Genetic Information Service	www.angis.org.au
DADI	Data Access and Database Interoperability (a GBIF programme)	www.gbif.org/prog/dadi
DDBJ	DNA Data Bank of Japan	www.ddbj.nig.ac.jp
DIGIT	Digitisation of Natural History Collections (a GBIF programme)	www.gbif.org/prog/digit
DNA	Deoxyribonucleic acid	
ECAT	Electronic Catalogue of the Names of Known Organisms (a GBIF programme)	www.gbif.org/prog/ecat
EIONET	European Environment Information and Observation Network	eionet.eu.int
EMBL	European Molecular Biology Laboratory	www.ebi.ac.uk/embl
ExPASy	Expert Protein Analysis System	us.expasy.org
GBIF	Global Biodiversity Information Facility	www.gbif.org
GenBank	Nucleic acid sequence database managed by the US National Center for Biotechnology Information	www.ncbi.nlm.nih.gov/Genbank
ICT	Information and Communications Technology	
ITIS	Integrated Taxonomic Information Service	www.itis.usda.gov
LTERnet	Long Term Ecological Research Network	lternet.edu
MABnet	Man and the Biosphere Network	www.unesco.org/mab
MOU	Memorandum of Understanding	
NODES	Participant Nodes Managers Committee (GBIF)	
OCB	Outreach and Capacity Building (a GBIF programme)	www.gbif.org/prog/ocb
OECD	Organization for Economic Cooperation and Development	
RNA	Ribonucleic acid	
RNABase	The RNA Structure Database	www.rnabase.org
SCOR	Structural Classification of RNA	scor.lbl.gov/scor.html
Species 2000	Species 2000	www.sp2000.org/
STAG	Scientific and Technical Advisory Group (GBIF)	
SwissProt	Protein knowledgebase	us.expasy.org/sprot
TDWG	Taxonomic Databases Working Group	www.tdwg.org/

Appendix 2: Strategic Outcomes of the Thematic Programmes.

GBIF Thematic Programme Area	Value Added to Pre-existing Databases	Utility to Users	Avoidance of Duplication	Contribution to Developing World	Performance Measures	Time frame	Linkages and/or Partnerships	GBIF Work Programme Integration
Data Access and Database Interoperability (including Information and Communications Technology staff of GBIF Secretariat) – DADI & ICT	- Combining information across biological information domains and between these and other information domains will open whole new fields of inquiry as well as new insights for science, the economy, and resource use sustainability, - The information-management infrastructural nature of DADI means that it will contribute significantly at all levels of the GBIF network	- Correlations of information from various sources of both biotic and abiotic data will facilitate new scientific, technological and economic advances - Standards adopted for data and metadata will enable cross-domain correlation and combination of data that has never before been possible - In the long term, will make possible transparent querying, such that users will be able to query the system without first knowing all the kinds of information available	- No other existing entity is specifically dedicated to the enabling of such linkages between sectors of the biology information domain; - Global coordination among national and institutional levels will help keep individual workers aware of others' efforts and thus avoid duplication of effort.	- Scientific correlations and insights enabled by GBIF-sponsored interconnections will facilitate biodiversity discovery and research in developing countries - All nations will gain access to key data necessary to understand their own national biodiversity - The work of DADI is essential infrastructure for data-sharing with countries of origin	- Capability to search several databases simultaneously and report the combined results - Integration of databases is 'seamless'—that is, 'user-transparent': many databases are used by the system but the user view is as though it is a single database	- Year 1 of GBIF: Workshops held to develop standards for database interoperability. - Year 2 of GBIF: Further development of GBIF standards and implementation among biodiversity database providers - Year 3 of GBIF: Planning for inter-communication of biodiversity databases with molecular and ecological databases - Year 4 of GBIF: Implementation of cross-domain communication capabilities - Year 5 of GBIF: Achieving database interoperability with robust search capabilities across multiple information domains	- Linkages between databases of biological nomenclature, taxonomy and collections, other databases of relevance to biodiversity research (e.g. geospatial, chemical, biochemical, physiological) will enable queries never before imagined - Interaction of GBIF with GenBank, LTERnet and other providers of biological and other data outside of the biodiversity information domain will capitalise on previous investments	- DADI plus ICT within GBIF Secretariat to develop overall information architecture - DADI with ECAT to develop Names Service architecture - DADI with DIGIT to ensure completeness of coverage in developing data and metadata standards for specimen information - DADI with DIGIT, ECAT and OCB to establish and build capacity in the NODES network - DADI with Taxonomic Database Working Group (TDWG) to develop standards and protocols for a searchable, distributed information system - ICT with International Grid Forum (IGF) and CODATA - GBIF interacts with the library community to develop ways to link primary data with scientific literature resources - GBIF works with other holders of primary biological data to integrate information domains

GBIF Thematic Programme Area	Value Added to Pre-existing Databases	Utility to Users	Avoidance of Duplication	Contribution to Developing World	Performance Measures	Time frame	Linkages and/or Partnerships	GBIF Work Programme Integration
Digitization of Natural History Collection Data -- DIGIT	- Will liberate (bring into digital accessibility) data that have been "imprisoned" for as much as 300 years by ink and paper - Will enable sharing of biodiversity data between developed-world museums and developing world - Will facilitate improvements in ecological, systematic and other biological and anthropological research - Will make possible for the first time the proper vouchering of molecular sequence and tissue samples	- Natural resource management will be greatly enhanced through use of digitised specimen data combined with ecological information - Economic uses of individual species can be better developed using the information stored in specimen data - Individuals, interest groups or civic organizations can develop checklists of organisms that occur in their area - Biodiversity policy development will have a better foundation based on specimen data - Ecological restoration and land reclamation will benefit from historical records of species occurrence - Invasive species can be tracked through time and routes of invasion predicted - Potential extent of species believed to be endangered can be predicted	- Co-ordination among natural history museums can reduce total workload and increase coverage of distribution data - Thorough digitization of specimen data will enable prioritization of future collections research, biodiversity discovery collecting, and sharing of results of collection trips - Digital availability of specimen records will enable multiple uses of datasets, increasing the return on investment in digitisation with each use	- Digitization of the ca. 75% of biodiversity specimen data held by institutions in the developed world allows these data to be shared with the countries of origin of the specimens - Availability of specimen data on the species that occur in a country will assist that country to develop sustainable use of the benefits from those species - Taxonomists in the developing world will have information that they need at their fingertips rather than having to travel to the developed world to gather it	- Pace of digitization of natural history collection data has increased markedly - The digitised data are used for cutting-edge research and to solve real-world problems - Best practices for specimen digitisation have been developed within the collections community - Collections community has come together to seek funding to complete distributed digitisation projects	- Year 2: Discovery of extent of already-digitised collections - Year 2(et seq.): Seed money awards made to stimulate digitisation projects - Year 5: Most GBIF Participant Nodes have worked with the institutions in their countries to fund the digitisation of data held by these institutions	- Links will be established between natural history museums to provide more thorough coverage of the globe for known species distributions - Digitised specimen data can be correlated with ecological and geospatial data to document and predict species occurrences - Digitised specimen data correlated with molecular information can enable unprecedented data-mining	- DIGIT works with DADI to develop standards for specimen data and metadata - DIGIT interacts with ECAT and DADI in developing the taxonomic names service that will enhance searchability of specimen databases - DIGIT and OCB interact to share best practices in digitisation with collections in GBIF Participant countries - DIGIT works with DADI, ECAT and OCB to establish and build capacity in the NODES network - DIGIT works with numerous inter-institutional organizations to enhance community-building among collections - GBIF works with other major networks that provide primary biological data to integrate information subdomains into a whole for biological information

GBIF Thematic Programme Area	Value Added to Pre-existing Databases	Utility to Users	Avoidance of Duplication	Contribution to Developing World	Performance Measures	Time frame	Linkages and/or Partnerships	GBIF Work Programme Integration
Electronic Catalogue of Names of Known Organisms -- ECAT	- Essential infrastructure to enable full use of collections, species-observation, molecular, ecological and abiotic data - Will be the first-ever complete listing of all the species currently known to science, which will greatly improve the rate of discovery and description of as-yet unknown biodiversity	- Is the unifying linkage among biological information sources (that in databases, physical libraries, and in natural history collections) - Will provide access to correlated biodiversity information for non-scientist users - Planned developments will provide a "taxonomic names service" to users who may know only vernacular names "Taxonomic names service" will indicate possible synonyms for the name used in a query, and eventually will include full taxon-concept data	- GBIF will partner with and enable the co-ordination of efforts already begun (e.g., Catalogue of Life) to concatenate nomenclatural databases produced independently by various experts - The Catalogue, once completed, will obviate the need for individual researchers or organizations to create <i>ad hoc</i> taxonomies in order to make their databases searchable--the ECAT Names Service will become the global "authority file" for scientific names of organisms	- Is the unifying linkage among digital biological information sources (that in databases, physical libraries, and in natural history collections)---it will serve both the developing and the developed worlds - Will enable digital access to collections data, library information, other databases and the like that has never before been possible - Will help reduce costs of biodiversity information retrieval by scientists from developing countries (the Internet search capabilities enabled by ECAT will reduce the need to travel to large information-centres in the developed world)	- Biodiversity data providers are using ECAT as the "authority file" for scientific names in their database systems - Biodiversity scientists are both using and contributing to ECAT during their regular work - ECAT will facilitate queries by automatically providing synonyms of the name entered by the user, and incorporating those into the queries that are presented to GBIF through any specialized search engine	- Year 2: Memoranda of understanding with Catalogue of Life project completed - Year 2: Initial developments for handling name services hammered out - Year 2 (et seq.): Seed money awards to stimulate garnering of name lists of all sorts Year 3 (end): 40% of all scientific names (including synonyms) included in ECAT Years 4--10: steady increase in name availability; also vernacular names and bibliographic information added - By Year 10: 90% of scientific names included in ECAT in taxonomic-concept-based form	- Enables links among existing sequence databases because names are the unique identifiers included in those sources - Will enable links between taxonomic and ecological databases not at present possible - Will in approximately year 4 or 5 begin to provide access to "SpeciesBank", the digital online information collation facility	- ECAT, DADI and ICT work to develop and implement the Taxonomic Name Service architecture - ECAT interacts closely with Species2000, ITIS, IPNI and other initiatives that provide taxonomic names to the web - ECAT with DIGIT, DADI and OCB to establish and build capacity in the NODES network - ECAT, DADI and ICT interact with counterparts in other information domains to enable cross-domain searching using unified or at least harmonized taxonomies - GBIF interacts with international nomenclatural bodies to work toward modernization of publication and validation of new scientific names

GBIF Thematic Programme Area	Value Added to Pre-existing Databases	Utility to Users	Avoidance of Duplication	Contribution to Developing World	Performance Measures	Time frame	Linkages and/or Partnerships	GBIF Work Programme Integration
Outreach & Capacity Building - OCB	- Will open new avenues of research both in computer science and information technology, and in biological sciences - Will expand user-base for GBIF information by helping to inform potential users of the utility of GBIF - Will advance science, technology and economy in both developed and developing world - Will encourage increased Participation in GBIF - Will solicit funding for GBIF from other sources in cooperation with the Executive Secretary	- Will engender good will among GBIF Participants and others - Will encourage scientists to provide data usable by non-scientist audience - Will facilitate workshops, training courses and other avenues to assist developing world to gain expertise in the use of primary biodiversity data for natural resource management and sustainability	- Every connection made will avoid duplication of effort because the Internet allows sharing and reorganization of the same information for different purposes - Primary scientific work will be shared globally, which will significantly reduce the need for the developing world and others to re-do what has already been done and therefore allow scientists and others to add to the body of human knowledge	- Research techniques and capacities will be transmitted to the developed and developing world - Developing countries will receive GBIF benefits, for example: access via the Internet to information they previously had to travel to the developed world to get, assistance with the use of repatriated data to benefit their populations, etc. - GBIF, in partnership with other organizations encourages the acquisition of Internet connections and promotes the GBIF concept	- Internet linkages are developed worldwide - GBIF concepts and principles of data-sharing are accepted and adopted globally - GBIF Participation increases by 10% or more each year - Training courses are provided and promoted by GBIF around the world	- By year 2 (et seq.): Organize a pilot programme of in-country focused support for developing countries with IT volunteers - By year 2 (et seq.): Develop with UNESCO a programme for Biodiversity Informatics Chairs - By year 2 (et seq.): Develop programme for School of Biodiversity Informatics - By year 3 (et seq.): Biodiversity informatics training of trainers courses are developed - By year 3: Modular programme for capacity building is available online - By years 4 & 5: Achieve the establishment of national and regional training programmes and activities	- Interactions between the biodiversity science and information technology communities will be encouraged in GBIF training courses - OCB works to bring representatives from developing countries to GBIF events such as Governing Board meetings in order to demonstrate the value to their countries of becoming GBIF participants	- OCB works with ICT, DADI, ECAT and DIGIT to establish and build capacity in the NODES network - OCB assists DIGIT and ECAT to identify supporting funds for projects in developing countries elsewhere - OCB works to identify individuals from Participating countries who can contribute to the Science subcommittees that advise DADI, DIGIT and ECAT - GBIF interacts with international biodiversity related organizations and conventions (e.g. CBD) to forge alliances and partnerships to achieve GBIF goals, which will also serve the goals of those organizations