Support to Building the Inter-American Biodiversity Information Network

Trust Fund #TF-030388

Taxonomic Authority Archives, Networks and Collections

(Report 7)

June 2004

Support to Building IABIN (Inter-American Biodiversity Information Network) Project

Taxonomic Authority Archives, Networks and Collections

Project Background

The World Bank has financed this work under a trust fund from the Government of Japan. The objective is to assist the World Bank in the completion of project preparation for the project Building IABIN (Inter-American Biodiversity Information Network) and for assistance in supervision of the project. The work undertaken covers three areas: background studies on key aspects of biodiversity informatics; direct assistance to the World Bank in project preparation; and assistance to the World Bank in project supervision. The current document is one of the background studies.

The work has been carried out by Nippon Koei UK, in association with the UNEP World Conservation Monitoring Centre.

Table of Contents

Annexes

Report Summary

This report provides a review of a number of taxonomic authority archives. The scope and types of authority networks are described, along with their role in biodiversity information management.

Four comprehensive reference sources, GBIF, ITIS, Species 2000, and All Species are reviewed and information on their structure, partnerships, and service are provided. The relationship between these four global players is described. ITIS and Species 2000 have formed the Catalogue of Life partnership. Together they have signed a Memorandum of Understanding with GBIF and will supply their species checklists to GBIF. In addition, a number of specialised reference sources are reviewed, along with their relationship to the comprehensive reference sources. These specialised reference sources are key sources of information to the comprehensive reference sources.

Many problems relating to taxonomy are encountered by both comprehensive and specialised reference sources. Expert review is a key tool employed to deal with such problems. A number of technologies are also being developed.

Recommendations are provided to assist with the further development of IABIN.

CHAPTER 1 INTRODUCTION

1.1 Scope

The purpose of this document is to provide a review of selected taxonomic authority archives (TAA). It describes selected taxonomic databases and electronic reference sources, the scope and structure of each, and examines the relationships between these reference sources. Problems encountered by TAAs are described, and tools and mechanisms that have been used to overcome these problems are presented. It is therefore a document that will provide guidance to IABIN on how best to compile and present taxonomic and other related information.

1.2 Taxonomy

Taxonomy involves the theory and practice of describing, naming and classifying organisms. Organisms are placed in a taxonomic hierarchy, which generally ranges from Kingdom down to species level, although the structure within each Kingdom varies. The species concept is central to the discipline of taxonomy, but there has been much discussion as to what constitutes a species, and there is no agreed definition across all groups. Taxonomic hierarchies are continually being revised as a result of the debate over how species should be defined, and because of newly described species and the unearthing of further information that may redefine already described species. Organisms previously considered as one species may subsequently be split into two or more species (e.g. on a geographic or morphological basis) and conversely, two species may be reassessed and 'lumped' into one. As a result, huge numbers of synonyms exist both in the taxonomic literature and in taxonomic databases and reference sources, and represent a sizeable challenge to their management. In sum, therefore, because taxonomy is a dynamic science and, because there is no such thing as a universally applicable species concept, the adoption of an accepted checklist of all species will effectively never be possible.

1.3 Role of taxonomy in biodiversity information management

The species name is key to the management and dissemination of biodiversity information. Before a species can be linked to any kind of information such as distribution, population size, trends, etc., that species must first be described and classified by taxonomists. Hence, despite all the debate and uncertainty, taxonomic information is essential to manage information on species, an important component of biodiversity.

The taxonomic status of any organism has implications on factors such as threats, endemism, etc., for that species. If a widespread and common species is split into

two species due, for example, to new scientific evidence, one of the newly described species may be quite restricted in distribution or rare, and may even be considered threatened, and *vice versa*. Hence it is important that we assess species populations and undertake conservation activities with the most up-to-date and accurate taxonomic information. Conservation efforts are often species-orientated but, without a sound taxonomic basis, appropriate species focused conservation actions are difficult to determine. In addition, since the majority of the world's species have not yet been described, it is impossible to assess overall species extinction rates, etc.

Users of taxonomic information and TAAs include scientists and governments, which require such information to, *inter alia*, assist in the formulation of policy, to use as biodiversity indicators, and to ensure that commitments to Multi-lateral Environmental Agreements (MEAs) are fulfilled. There are many agreements (e.g. Convention on Trade in Endangered Species of Wild Fauna and Flora, Berne Convention on the Conservation of European Wildlife and Natural Habitats) and many pieces of legislation (e.g. the EU Habitats Directive and the EU Birds Directive) that use lists of species, and so rely heavily on having sound taxonomic information. Taxonomic information is also required by conservation organisations and agencies (governmental, inter-governmental and non-governmental) to help formulate activities and determine conservation priorities.

1.4 Types of authority and reference networks

Taxonomic authority archives contain information on the taxonomic hierarchy in which a species lies and usually include information on author and date, and synonymy. Many TAAs also contain some additional information, most often geographical information and on location of specimens.

Most TAAs have been established by governments, inter-governmental organisations and agreements, institutions, organisations or, as is often the case, by partnerships between these groups. Because there are so many other types of information that could be included, there are important differences between a database that is fundamentally intended as a taxonomic resource and one whose main aim is to provide other information. The purpose of establishing a database has implications for the scope and type of the taxonomic information held in that database. For example, a database may be established with the intent to contain information on species listed in MEAs (e.g. CITES database) or on particular legislation. Financial and participant support for taxonomic databases has important implications for its quality and scope and hence its usefulness.

TAAs can obtain their information from a range of reference types and in many different ways. Some databases use only primary reference material, often in conjunction with expert comments, and all information is compiled by and often

in the institute 'housing' the database (e.g. ITIS). However, other taxonomic authorities are essentially networks that combine information provided by a range of other databases and institutions or by secondary reference material (e.g. GBIF), and may be centralised or decentralised. Taxonomic networks operate at three main levels – national (e.g. UK National Biodiversity Network), regional (e.g. EUNIS), and international (e.g. GBIF).

Many TAAs deal only with specific taxonomic groups, which may be on a global, regional or national scale (e.g. the global FishBase and Index Kewensis). Some (e.g. GBIF, Species 2000) aim to eventually include all taxonomic groups on the scale at which they operate. Species 2000 estimate that the existing global species databases may presently account for some 40% of the total known species.

1.5 Review of selected initiatives and identification of selected specimen collections in Japan and Europe that are relevant to the IABIN region

There are some specimen collections outside the IABIN region in Europe and Japan that have important collections of animals and plants relevant to IABIN. Selected examples are described in Annex 3. Some digitisation of the data relating to these specimen collections has been carried out, but this process needs to be accelerated. In addition, the current status of initiatives by organisations such as GBIF, GTI, BioNET, ITIS, and BioCASE, which are working on biological information sharing activities related to natural history collections are briefly described.

1.6 Interoperability

Interoperability can be understood as the communication between distributed systems. It enables the user to query a database and retrieve all data and information from a number of different sources at once, in a completely transparent way. As such, separate systems do not need to know the details of how they each work, but they do need to have enough common ground to reliably exchange information or data without error or misunderstanding.

Interoperability can be achieved in a number of ways. The first is when many data providers provide data to a central database, which is the only database accessed by the user. This approach is of limited value for broader taxonomic databases. The second is when identical structures and semantics are used by all databases and so can eventually be queried through a single web query interface. However, this limits the type of information that can be held, as different fields are often needed for different taxonomic groups, etc. The third and most versatile type uses databases that are different in structure map to a standard, and a common machine interface to the web. GBIF, for example, uses the Darwin Core, which is a profile describing the minimum set of standards for search and retrieval within natural

history collections and observation databases. The DiGIR protocol (the Distributed Generic Information Retrieval) used by GBIF, is intended to support the retrieval of structured data from multiple, heterogeneous databases, and it "wraps" the data in an XML envelope. Hence, when a user searches for information on a particular species, records from many different data providers and databases can be accessed.

Although such systems can be complex to develop, their functionality and usefulness for taxonomic databases are enormous. Hence, interoperability is one aspect of TAAs (particularly databases with broad taxonomic or geographic coverage) that is key to their overall usefulness and that will be an important consideration of IABIN.

1.7 Taxonomic and Nomenclature Codes and Standards

There are a number of procedural rules such as the International Code for Zoological Nomenclature and the International Code of Botanical Nomenclature, which generally aim to provide universality and continuity in the scientific names of species. Standards such as these are usually followed by TAAs and a five-kingdom system (Monera, Protista, Plantae, Fungi, and Animalia) has generally been adopted as a standard. However, these nomenclature codes deal with procedural issues and do not arbitrate between competing taxonomies.

Botanical nomenclature is independent of zoological and bacteriological nomenclature, and different conventions apply in each of the Codes. In the past, protists, have been treated according to either the zoological code (i.e. protozoans) or the botanical code (i.e. algae) and in some cases according to both. By initally assigning every taxon to one of the Kingdoms (Plants, Animals, Bacteria, Viruses) and applying the relevant Code within that Kingdom, some of the problems of having four different Codes can be overcome. However, the technical implications need to be considered.

The Taxonomic Database Working Group (TDWG), under the auspices of the International Union of Biological Sciences (IUBS) develops standards and protocols for taxonomic databases, which it issues in the form of recommendations on database design and information exchange. CODATA (Committee on Data for Science and Technology), an interdisciplinary Scientific Committee of the International Council for Science, has convened a Task Group on Global Species Data Networks who’s goal is to provide a uniform and validated quality index of names of all known species for use as a practical tool. The members of this group include particpants of several of the TAAs reviewed in this document including Species 2000, GBIF, ITIS, FishBase. It is important that procedural rules are followed by TAAs to ensure that taxa names are meaningful and consistent, and to ensure compatability with other systems. IABIN will need

to assess the standards followed by contributing organisations to ensure that they are compatible with its needs.

1.8 Problems

Although approximately 1.7 million species have already been scientifically described, an estimated further 10 to 100 million others remain undescribed. Taxonomy has not been a high priority discipline for the last 100 years and there are many gaps in our knowledge, particularly for certain taxonomic groups and geographic areas. Not enough trained taxonomists are available to do much of the necessary work and this presents a huge barrier in the retrieval of taxonomic information. In addition, much of the taxonomic information already collected has not yet been digitised, and so it is not readily accessible electronically. This further increases the gaps in the information available to TAAs. Accessibility to taxonomic information by conservationists has been considered problematic (e.g. plants (Lowry and Smith, 2003)). Additionally, official decision-making often tends to be slow, and hence TAAs may have difficulty in keeping up to date with all taxonomic changes.

The Convention on Biological Diversity (CBD) has acknowledged the existence of a "taxonomic impediment" to the sound management of biodiversity. In response, it has launched a Global Taxonomy Initiative, the purpose of which is to remove or reduce: the knowledge gaps in our taxonomic system (including those associated with genetic systems); the shortage of trained taxonomists and curators; and the impact these deficiencies have on our ability to conserve, use and share the benefits of our biological diversity.

A taxonomic network relying on multiple sources of information may retrieve conflicting information from its source databases with regards species names and an accepted taxonomy. These conflicts are encountered across the entire taxonomic spectrum and are often difficult to resolve. Hence, TAAs must develop effective ways to both highlight such conflicts and where possible or necessary, resolve them. In many cases highlighting conflicts is not sufficient if the information is to be used in a practical capacity e.g. to enforce legislation, and a decision on an accepted taxonomy may be required.

The problem of differing taxonomies has been encountered by a number of international agreements and Conventions. The Appendices of both the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and the Convention on Migratory Species (CMS) comprise lists of species that are subject to various degrees of protection or conservation action. The taxonomy followed by CITES is determined by the CITES Nomenclature Committee, and standard taxonomies have been adopted for many, but not all,

taxonomic groups. The taxonomy followed by CMS can differ in some instances from that used by CITES.

Parties to international conventions, such as those described above, are required to submit reports on their implementation of the conventions. Because different taxonomies are used by each convention, information compiled in the national reports submitted to each is difficult to combine and assess. In addition, the reporting requirements of international agreements can be demanding and the taxonomic discrepancies between conventions can further increase this burden. It has been suggested that the use of a common taxonomy between the species-based international and regional agreements would lead to improved access to and usefulness of the information that is compiled through national reports *etc*.

CHAPTER 2 COMPREHENSIVE REFERENCE SOURCES AND THE RELATIONSHIP BETWEEN THEM

2.1 Introduction

This section reviews four major international initiatives that provide broad taxonomic reference information: GBIF, ITIS, Species 2000 and All Species. A fifth TAA, Discover Life, is also briefly reviewed.

The potential overlap between these initiatives has been acknowledged. Formal agreements between several of the databases have been established to ensure cooperation, interoperability and information exchange. In addition, many board or committee members of one network are also participants or associate participants of the others.

2.2 GBIF

2.2.1 Background

The Global Biodiversity Information Facility (GBIF) was established in 2001. Its mission is to make the world's primary data on biodiversity freely and universally available via the Internet. It provides digital access to information on taxonomic hierarchies with links to further information where available.

A number of areas of emphasis have been identified by GBIF, namely: data access and interoperability; digitisation of natural history collections; electronic cataloguing of names of known organisms; and outreach and capacity building. According to their website *"In the near term, GBIF will provide a global metadata registry of the available biodiversity data with open interfaces. Anyone can then use it to construct thematic portals and specialised search facilities. Building on the contents of this registry, GBIF will provide its own central portal that enables simultaneous queries against biodiversity databases held by distributed, worldwide sources."*

Of particular note is the GBIF programme ECAT (Electronic Catalogue of Names of Known Organisms) which is working towards an electronic catalogue of the names of known organisms. It aims to provide content infrastructure to enable searches across multiple information domains, to make seed-money awards to speed progress of Catalogue development, and to develop the Taxonomic Name Service function of GBIF information architecture.

2.2.2 Partners and governance

GBIF originated through the work of the OECD Megascience Forum Working Group on Biological Informatics that was established in January 1996. A

Memorandum of Understanding (MOU) was opened for signature in December 2000, which provided that GBIF would be considered established once ten countries had signed and pledged their monetary contributions. This was achieved before March 2001. GBIF's Secretariat co-ordinates its activities, and GBIF is governed by a board comprising representatives from each of its participating nodes.

GBIF comprises participants and associate participants, all of whom provide it with data. Participants are those governments, economies or international organisations that are prepared to share biodiversity data and to develop a computing gateway or node for accessing such data. Participants in GBIF have signed a Memorandum of Understanding, and support network nodes through which they provide data. According to the most recent status reports delivered to the GBIF Secretariat, there are currently 43 nodes. Participants guarantee that they will openly and freely share biodiversity data, while at the same time, the data providers are assured that they will retain control of their own data. GBIF operates a data use agreement that requires users to publicly acknowledge the relevant data providers. Data providers include a range of organisations, from national initiatives such as Australian Biological Resources Study and Canadian Biodiversity Information Network, to comprehensive taxonomic reference sources such as ITIS and Species 2000, and more specialised databases such as AlgaeBase and FishBase.

2.2.3 Taxonomic information and options available through the GBIF website

Functionally, GBIF offers the users a number of search options: by scientific name, or by common name in any language. The search can be further narrowed by searching for a species in a particular country. The GBIF website is currently in English, although the search terms are also available in French and Danish.

Information on taxonomic hierarchy, including synonymy in some cases, common names, maps, locations of specimens or observations, and links are provided. Distribution lists either by country or broad geographic region are only provided when linked to a database that holds such information, e.g. FishBase, and so distribution information is not provided for all species. References (usually information provided by other databases and by partners) are sometimes (e.g. *Physeter catodon*) but not always provided (e.g. *Nemenzophyllia turbida*, where the database is sourced but no further information is available).

2.3 ITIS

2.3.1 Background

The Integrated Taxonomic Information System (ITIS) aims to provide authoritative taxonomic information on plants, animals, fungi, and microbes of

North America and the world. Although originally established with a North American focus, ITIS has since expanded and includes information on species from around the globe. The goal is to create an easily accessible database with reliable information on species names and their hierarchical classification. The ITIS includes documented taxonomic information on flora and fauna from both aquatic and terrestrial habitats. Currently coverage for some taxa is global, and for others it is as yet confined to North America, and there are many gaps in ITIS in the coverage of South American taxa. A summary of the status of data in ITIS can be found in the ITIS Data Status Summary Table at http://www.itis.usda.gov/status.html. The database is reviewed periodically to ensure high quality data with valid classifications, revisions, and additions of newly described species.

Each of the ITIS countries: US, Canada and Mexico, has a separate portal in which data can be queried. In addition, the ITIS data can be queried through the GBIF portal. Although the same data are used for all of these, different tools are available through each. For example, for some taxonomic groups, the Mexican portal provides different taxonomies that have been used for a given species, a tool that might be useful to IABIN.

2.3.2 Partners and governance

ITIS was originally formed as a partnership of a number of US agencies who signed a Memorandum of Understanding and formed a Steering Committee that directs two technical work groups - the Database Work Group (DWG) and the Taxonomy Work Group (TWG). ITIS has since expanded to form a partnership of US, Canadian, and Mexican agencies (ITIS-North America), other organisations, and taxonomic specialists.

Data sources used by and contributors to ITIS include a wide range of experts and information sourced from numerous published checklists and other databases. Data is reviewed by the TWG prior to incorporation in to ITIS.

2.3.3 Taxonomic information and options available through the ITIS website

Currently, portals for Canada, the US and Mexico query the same data, but each has developed tools targeted to their particular end users. Different language options are available through each portal: English (United States http://www.itis.usda.gov/), English and French (Canada http://www.agr.gc.ca/itis), Spanish (Mexico http://siit.conabio.gob.mx/), and Portuguese (Brazil http://www.itis.cria.org.br).

ITIS states that for each scientific name, it will include the authority (author and date), taxonomic rank, associated synonyms and common names where available, a unique taxonomic serial number, data source information (publications, experts,

etc.). Data quality indicators are also provided. Expert reviews and changes to taxonomic information in the database will be tracked. Information on broad geographic distribution is also provided, but country lists are not. IUCN Red List status and CITES listing for some, but not all, species are provided in the comments field of the database. In addition, ITIS has established a Taxonomic Resources Expertise Directory (TRED).

2.4 Species 2000

2.4.1 Background

Species 2000 is a "Federation" of database organisations working closely with users, taxonomists and sponsoring agencies. Species 2000 aims to provide an index of all known species in the world through an array of participant global species databases covering each of the major groups of organisms. Each such database will cover all known species in the group, using a consistent taxonomic system.

2.4.2 Partners and governance

Species 2000 arose from a workshop held in 1996 at which participants of 18 taxonomic databases met. The headquarters of the present databases are widely distributed throughout the world, and this trend will continue as new databases are added. Specialists also contribute information to Species 2000. Activities of Species 2000 are coordinated by a Secretariat based in the UK.

2.4.3 Taxonomic information and options available through Species 2000

The website of Species 2000 is in English. The Catalogue of Life 2003 checklist of the species of the world, produced in partnership with ITIS, currently holds information on 304,000 species, 486,000 synonyms, 217,000 common names, and 87,000 references, from 20 contributing databases. Through the Species 2000 website, access is provided to a Dynamic checklist which is continually updated, and an Annual checklist which is updated once a year. The annual checklist can be queried using scientific or common name. In addition, users can search for references by particular author or in a particular year. The Dynamic checklist can be queried using the genus name only.

Species searches provide information in a number of fields, namely: author and date; partial or complete taxonomic hierarchy (the standard is to Family level); distribution (though incomplete in some instances); specialist; date checked; remarks; further information; database; and search web for the species. Synonyms are sometimes, but not always, listed. In addition, information on CITES listing and IUCN Red list status is provided for some but not all species (this is due to the fact that information on some species is sourced from ITIS, which provides such information for some but not all species).

In addition, a Names Service is provided, which will be of particular use to professional taxonomists. The spelling and original publication details for a much wider range of names can be checked. It includes many groups of organisms not yet incorporated in the Species Locator service. Many of the names available here are no longer the accepted names of species recognised today, and some are long out of use, with the exception of the Bacteria, Archaea, Algae and some of the Fungi, which do give accepted names and a responsible synonymy.

2.5 All Species

2.5.1 Background

The All Species Foundation was established to catalogue every living species on earth. This inventory would need to enlist the support and cooperation of scientific organizations around the world. All Species is intended to be a temporary endeavour which will cease to exist in 25 years when its mission to compile a list of all species is completed. Information in All Species is sourced from both comprehensive and specialist TAAs around the world.

At the moment the activities of All Species as a TAA appear to have been surpassed by the activities of the other major TAAs, and it may be of limited use to IABIN as a TAA. Funding constraints have led to a scaling down of its activities. Due to economic circumstances the All Species officially closed its San Francisco office in December 2002, although work continues at the California Academy of Sciences by one staff member. In an effort to avoid duplication of the work being done by GBIF, All Species has shifted its emphasis towards encouraging the development of new technologies and tools, and publicity for taxonomic endeavours, which may be more relevant to IABIN.

2.5.2 Partners and governance

The All Species Foundation was established following a series of workshops and meetings between 2000 and 2002 attended by scientists and other professionals from around the world. It brings together participant global species databases. All Species has a number of strategic partners that are involved in funding, co-sponsorship, data exchange, etc. In addition, a new partnership agreement with Conservation International will fund one half-time All Species representative at their Washington DC office.

2.5.3 Taxonomic information available through All Species

The All Species website is in English. All Species is essentially a search engine through which a number of databases can be queried for particular species and links to the appropriate page. Hence the information provided for a given species will depend entirely on the source databases. The search function is available through the All tool kit.

2.6 Discover Life

The mission of Discover Life mission is to assemble and share knowledge about nature in order to improve education, health, agriculture, economic development, and conservation throughout the world. Its focus is to develop pages for the estimated 100,000 species that inhabit Great Smoky Mountains National Park, but will also include information on any species living anywhere. Discover Life is served from the University of Georgia, Athens, under the auspices of The Polistes Foundation, a non-profit think tank with a team of expert advisors.

Currently the species coverage in Discover Life is much more limited than that of ITIS, the other North American TAA. One of its long-term goals is to provide an easy and freely available gateway to query an up-to-date electronic encyclopaedia of life that includes everything from a network of virtual museums and herbaria to recommendations on how to control pests and disease vectors, diagnose and treat infections, and grow crops, trees, and flowers. However, it does not purport to be a TAA and does not currently address many of the taxonomic problems outlined previously.

2.7 Relationship between comprehensive reference sources

The previous sections have outlined selected comprehensive reference sources of taxonomic information, all of which have similar end goals i.e. to provide taxonomic and other information on the species of the world (or in the case of ITIS, for North America and then the world) across all taxonomic groups. However, in an effort to ensure cooperation and coordination and to avoid duplication of effort, two important agreements have been signed by the three main players, GBIF, ITIS and Species 2000. All Species appears to lie outside of the mainstream of these other taxonomic initiatives but it does work with and obtain data from them.

The relationship between ITIS and Species 2000 is still evolving. Together they produce the Catalogue of Life, a uniform and validated index to- or checklist of-the world's known species collated by taxonomists throughout the world. It is available on a CD and can also be downloaded from the web. Three editions of the CD have been produced so far. ITIS and Species 2000 signed a Memorandum of Understanding in November 2003 to further enhance collaboration.

GBIF has recently signed a Memorandum of Cooperation with the Catalogue of Life Partnership of ITIS and Species 2000. The Memorandum provides a basis for mutual co-operation and a framework for GBIF to access the Catalogue of Life and to use it in its services. The Memorandum covers an initial three-year period, but can then be extended on a rolling basis. The synonymic species checklists provided by the Catalogue of Life partnership will be made available to GBIF,

and it is anticipated that they will play a key role in the name-service and indexing functions of the GBIF portal. So far, GBIF has provided some funds to the Catalogue of Life Partnership to assist with their work.

The role of GBIF differs from that of the Catalogue of Life partners. While ITIS and Species 2000 will provide a checklist of species of the world, GBIF will be a portal both to that information as well as to large amounts of other information based on species collections from museums and herbaria throughout the world. GBIF will contain not just the Catalogue of Life species names but also names that have long since gone out of use, that have been misused, misspelled etc and location and other information linked to each such specimen. The initial focus of GBIF appears to be on museum specimens, which should provide new sources of information and will complement work already undertaken by ITIS and Species 2000.

All Species aims to "*co-operate with 'anyone and everyone' including 'competition'. It prefers to use and promote tools made by others*." Species 2000 is the largest data provider to All Species although no formal relationship exists between the two. All Species and GBIF are strategic partners, and All Species has signed an MoU with GBIF.

Of the comprehensive reference TAAs, those that will be most useful to IABIN will depend on the needs of IABIN: whether it requires a checklist of species with one accepted name per taxa only, or a wider list that may have all accepted names as considered by various taxonomies.

The comprehensive TAAs outlined above offer the future possibility of providing species names for all the species of the world. IABIN, by definition, will focus on the Americas. ITIS currently has a North American focus, with global coverage for some taxonomic groups. Expansion to full South American coverage would require substantial work and so would likely take some time. The checklists provided by 'Catalogue of Life' partnership could be a useful starting point for IABIN to compile a list of species of the Americas. However, GBIF also represents a strong international TAA, and it includes both the information described above and much more, and it is in the process of developing a number of technologies to deal with taxonomic problems.

CHAPTER 3 SPECIALISED REFERENCE SOURCES AND THE RELATIONSHIP BETWEEN THEM

3.1 Introduction

There is a wide range of specialised reference sources which aim to catalogue all of the species in particular taxonomic groups or geographic regions, or particular subsets of species, e.g. those listed in international conventions or as globally threatened. Many of the specialised reference sources have an explicit mechanism for deciding upon at least one responsible or consensus taxonomy, and for applying that taxonomy consistently. In addition, synonyms and alternative taxonomies are often cross-referenced. Many specialised reference sources represent the most comprehensive electronic sources of information available for their particular taxonomic expertise.

3.2 FishBase

3.2.1 Background

FishBase is an information system with key data on the biology of all fishes. Fishbase on the web probably contains practically all fish species known to science. The FishBase 2000 CD-ROM contains taxonomic, geographic and biological information on 25,000 species.

3.2.2 Partners

FishBase was developed at the International Center for Living Aquatic Resources Management (ICLARM), in collaboration with the Food and Agriculture Organization (FAO) of the United Nations and many other partners. It currently has 1,040 collaborators, comprising experts from around the world who provide fish data of all kinds to FishBase.

3.2.3 Information that the database contains

The FishBase website can be accessed in English, Spanish, Portuguese, Italian, German, Dutch, Swedish and Chinese. It currently contains information on 28,400 species, 79,100 synonyms, 183,300 common names, and 32,500 references. FishBase provides updates on the general changes that have been made to the database enabling users to quickly assess these changes, but these do not include specific taxonomic changes and additions.

FishBase contains information on taxonomic hierarchy, author and date, detailed information on synonyms, distribution and a wide range of other information. Although information on taxonomic hierarchy is provided in FishBase, the position of taxa within each level of the taxonomic hierarchy is unclear, and instead taxa are listed alphabetically.

3.3 MammalBase

Although MammalBase was explicitly referred to in the TOR for this project, no information on this resource could be found on the Internet, or from other sources.

However, the Mammal Species of the World (MSW) is a database of mammalian taxonomy, which provides a taxonomic hierarcy for all of the world's known mammal species. It is housed by the Smithsonian Museum of Natural History. The authors of this database have attempted to provide taxonomic lists consistent with recent literature. A comprehensive list of synonyms that have been used in the scientific literature for each taxon has been compiled. The entire Mammalia name hierarchy is also available as a checklist.

In addition, another mammalian network of relevance to IABIN is MaNIS (Mammal Networked Information System). This network of distributed databases of mammal specimen data was developed by 17 North American institutions (Natural History Museums) and their collaborators. Queries can be made through the portals of each institution and taxonomic and geographic information relating to specimens can be obtained. The ANSI/NISO Z39.50 standard for information retrieval is used to enable data sharing and exchange of knowledge among natural history collections.

3.4 GloBIS (The Global Butterfly Information System)

The Global Butterfly Information System (GloBIS) is an initiative of the Major Systematic Entomology Facility (MSEF) group. It aims to make available in electronic form taxonomic, distributional, bibliographic and ecological information about butterflies that is currently unavailable to users because it is 'hidden' in museum collections or in scholarly articles, and to train students in butterfly systematics where expertise is lacking. Although global coverage is not currently available from Globis over the internet, it is intended that such information will be available in the future.

3.5 Index Kewensis

3.5.1 Background

Index Kewensis (IKew) is a world list of names of seed-bearing plants with bibliographic references to their first publication. It should be strictly considered as an index and it is not a list of accepted names. Additionally, not all of the names listed are validly published according to the rules of the International Code of Botanical Nomenclature (ICBN). It should not be considered an authority regarding synonyms, in spite of the fact that synonyms are listed for many genera and species.

3.5.2 Partners

Compilation of IKew began in January 1882. It is currently compiled at the Royal Botanic Gardens, Kew, in the UK and draws upon support from libraries around the world.

The Royal Botanic Gardens, Kew, the Harvard University Herbaria, and the Australian National Herbarium all collaborate to produce the International Plant Names Index (IPNI), which is a database of the names and associated basic bibliographical details of all seed plants. For IABIN's purposes, IPNI is likely to be more useful than Index Kewensis.

3.5.3 Information that the database contains

IKew is available in English, on CD Rom or on the web via IPNI, and can be used to search for the published name of any seed-bearing plant (genus or species level) anywhere in the world, its author and bibliographic reference; records for a family, genera in a family, or species in a genus; synonym(s) to aid in a search of past publications or descriptions of a plant; a list of the genera and/or species described in a specific publication (e.g. those in a particular flora); the region(s) or habitat where the species occurs (or at least where the type was found);

On the Index Kewensis website a list of Authors of Plant Names is provided, giving the recommended standard forms of those authors' names, including abbreviations. Through this list the plant groups an author has studied/described can be determined.

The website also provides a list of inconsistencies and pitfalls in the database. Additionally, there are a number of errors that were introduced when the data were converted into an electronic form.

3.6 ILDIS

3.6.1 Background

The International Legume Database & Information Service (ILDIS) is an international project that aims to document and catalogue the diversity of the world's legume species in a readily accessible form. Information from participating research groups around the world is compiled in the ILDIS World Database of Legumes, which is updated every ten weeks. Information has been collected from local herbaria, national botanists, and from literature written in many different languages.

3.6.2 Partners

ILDIS is a collaborative organisation and it involves legume experts and institutions from all over the world. The activities of ILDIS are managed from the

ILDIS Co-ordinating Centre based in the UK. However, much of the information included in the ILDIS World Database of Legumes has been compiled at the ILDIS Regional Centres, which are located throughout the world.

3.6.3 Information that the database contains

The ILDIS website is in English. The ILDIS World Database of Legumes currently contains records for over 19,000 taxa. The data contain fully referenced, information, and comprise the World Species Checklist Data, the World Geographical Distribution Data and the World Botanical Information Data. The ILDIS World Database of Legumes can be searched by scientific name, and information is retrieved for accepted name, synonyms, taxonomic tribe, descriptors, geographical records (including whether native, introduced, or of uncertain status), sources, references and links to other databases.

A draft World Species Checklist is available on the website which provides a list of species, subspecies and botanical varieties, constructed and updated by a world-wide network of taxonomic experts. Information included in the Checklist includes accepted name, authority and references, legume tribe membership, position in the classification and vernacular names with references. Associated geographical and botanical information will also be available.

ILDIS provides updates on changes that have been made for each version of the database, enabling users to quickly assess such changes.

3.7 CABI

3.7.1 Background

CABI Bioscience is one of the two divisions of CAB International. CABI Bioscience houses a number of global taxonomic databases, chiefly of fungi, the IndexFungorum. The remit of the database is fungal nomenclature and systematics (i.e. fungi in a strict sense, fungal analogues, and (not comprehensively) fossil forms.

3.7.2 Partners

The Index Fungorum database is now jointly managed by CABI Bioscience and Centraalbureau voor Schimmelcultures (CBS) (http://www.cbs.knaw.nl/); the other resources on the website are wholly managed by CABI Bioscience.

3.7.3 Information that the database contains

IndexFungorum is a world database of fungal names and contains over 345,000 names of fungi (including yeast, lichens, chromistan fungi, protozoan fungi and fossil forms) at species level and below. It has been derived from a number of published lists. By the end of 2004 the database will account for 40% of the

known fungi (32,000 species of an estimated 80,000 total) by assigning taxonomic assessments to a subset of the 360,000 names.

The CABI website and database are in English. The database may be searched by either the species name or the specific epithet. There is almost complete coverage of all names at species, genus and family level; infraspecific names are incomplete for lichen forming species. A name record will usually have a reference to an entry in one of the bibliographic catalogues cited above and, in addition, more recent records from the Index of Fungi will have the full citation from the source publication (excluding those from the last 5 years). Many records include information on taxonomic synonymy and publication details derived from numerous acknowledged sources.

The Dictionary of the Fungi (currently 9th edition) contains the current consensus on the fungal taxonomic hierarchy to the rank of genus and through which the database may be searched for the status of generic names, or by following the hierarchy from the rank of Kingdom. The entries for each genus generally include authors and place of publication, together with the type species, and other data. A recent addition is the database of family names, which includes authors, place of publication and type genus. This database will eventually be expanded to include all supra-familial ranks.

3.8 CGIAR SINGER

3.8.1 Background

The Future Harvest Centres, a network of 16 food and environmental research centres supported by the Consultative Group on International Agricultural Research (CGIAR), hold more than half a million samples of crop, forage and agroforestry plants in trust for the world community under agreements signed in 1994 with the United Nations Food and Agriculture Organization (FAO). The CGIAR System-wide Information Network for Genetic Resources (SINGER) was established under the auspices of the CGIAR System-wide Genetic Resources Programme (SGRP).

SINGER links the genetic resources information systems of the individual Centres of the CGIAR around the world, allowing them to be accessed and searched collectively. SINGER contains information on the half a million individual samples of germplasm held in the CGIAR Centre gene banks.

SINGER is an initiative of the CGIAR System-wide Genetic Resources Programme (SGRP). The International Plant Genetic Resources Institute (IPGRI) provides the Secretariat for the SGRP. SINGER collaborates with numerous specialists to make available high-quality information and expertise on the plant collections.

SINGER encourages networking both on a crop and regional basis. Pilot networks linking holders of wheat, barley and sweet potato collections worldwide are currently being established under the lead of the Future Harvest Centres.

3.8.2 Information that the database contains

The SINGER website is in English. Users can search all holdings by crop, taxonomy, geography, acquisition or transfer, or can focus on the collections in one Centre or of one crop, and query data on the characteristics, performance and source of the samples.

SINGER offers a number of specialised and innovative data searching and retrieval features. The user interface provides multiple query functions that integrate multiple querying with mapping (global, regional, country), statistical (mean, variance and standard deviation) and graphical (scatter and distribution plots) functions. Data can be downloaded from SINGER for further analysis.

3.9 UNEP-WCMC Species database

3.9.1 Background

The UNEP World Conservation Monitoring Centre is the biodiversity assessment and policy implementation arm of the United Nations Environment Programme (UNEP).

In its Species database, UNEP-WCMC maintains information on species of conservation concern including those protected by multilateral environmental agreements (MEAs), such as the Convention on Trade in Endangered Species of Wild Fauna and Flora (CITES), the Convention on Migratory Species and a number of regional agreements. The database currently holds information on almost all described vertebrate species, and many species of plants and invertebrates - 69,672 animals and 85,535 plants as well as tens of thousands of subspecies, populations and synonyms. Wherever possible, data managed by UNEP-WCMC are placed in the public domain. Subsets of the data are available on request, however due to data provider agreements, entire datasets for some taxa cannot be downloaded.

3.9.2 Partners

UNEP-WCMC has a wide range of partners that supply information and data to its databases. Species lists in the database follow the taxonomy of CITES, where a standard taxonomy has been adopted. Other species lists are supplied by other organisations, although no formal agreements are in place, except with the conventions.

3.9.3 Information that the database contains

The UNEP-WCMC website is available in English, French or Spanish. Other language options, including Arabic, Russian, and Chinese are also being developed. Sub-sets of data from the Species Database have been tailored to different user needs and can be interrogated on the web, providing differing views of the central database.

Searches can be conducted at any taxonomic level from phylum to species, by scientific or common name and also by country. A complete taxonomic hierarchy is provided for all species, as well as species author and date, common names in a variety of languages where available, distribution, and information on whether it is listed on a number of MEAs. Additional information on literature references and relevant links to other databases and resources are also provided.

3.10 IUCN Red List

3.10.1 Background

The IUCN (World Conservation Union) Red List of Threatened Species provides information on taxonomy, conservation status and distribution for taxa that have been evaluated using the IUCN Red List Categories and Criteria. The IUCN Red List is not a Taxonomic Authority Archive; however, it does contain information on the global threat status of species, which will likely be useful for IABIN's purposes, and so is included in this section.

The main purpose of the IUCN Red List is to catalogue and highlight those taxa that are facing a higher risk of global extinction (i.e. those listed as Critically Endangered, Endangered and Vulnerable). Information on taxa that are Extinct, Extinct in the Wild; Near Threatened and on taxa that cannot be evaluated because of insufficient information (i.e. Data Deficient) are also given.

3.10.2 Partners

The IUCN Red Lists are produced under the direction and management of the Red List Programme and involve a network of scientists and partner organisations working in almost every country in the world.

3.10.3 Information that the database contains

The IUCN Red List can be searched in English. The list of threatened taxa is maintained in a searchable database. Information on species provided through the Red List website includes: higher taxonomy; scientific name (including authority details wherever possible); common names in English, French, and Spanish; Red List category and criteria; date of assessment; an indication if a petition about the status of the species has been lodged; distribution information; a rationale for the

listing (including any numerical data used or inferences made, that relate to the thresholds in the criteria); current population trends; major habitat preferences; major threats (past, present and future); conservation actions in place and needed; and other notes.

3.11 Zoological Record

The Zoological Record is a database summarising the citations to scientific literature on animals which, in the process, gathers information on taxon names. Zoological Record has been published continuously since 1865, but currently only the volumes from 1978 onwards are available electronically. The Index to Organism Names is a system created by BIOSIS UK to make animal nomenclatural data available on the web, using Zoological Record. The database currently holds more than 1.4 million names at all taxonomic ranks. These are available via a search function but not in a list form. Users are made aware that the system does not deliver a list of valid organism names; instead it offers a view of how names have been utilised in the scientific literature and what changes in their application have been formally proposed.

3.12 European Species of Conservation Concern (EUNIS)

3.12.1 Background

The European Nature Information System (EUNIS) is a set of public-access databases that store information collected within the framework of the Natura 2000 programme. The EUNIS web application provides access to data on species, habitats and sites compiled in the framework of NATURA 2000 (EU Habitats and Birds Directives), and data sources or material published by ETC/NPB (formerly the European Topic Centre for Nature Conservation). Additionally, information is provided on species, habitats and sites taken into account in relevant international conventions or from International Red Lists, as well as specific data collected in the framework of the European Environment Agency's reporting activities. Though developed separately, at the request of the European Commission, the database on Special Protection Areas (Sites designated under the EC-Birds Directive, which are also part of the NATURA 2000 Network) provides relevant data.

3.12.2 Partners

EUNIS was developed and is managed by the European Topic Centre for Nature Protection and Biodiversity (ETC/NPB in Paris) for the EEA and the European Environmental Information Observation Network (EIONET).

3.12.3 Information that the database contains

The EUNIS website is in English. EUNIS contains information on more than 23,000 species and subspecies in Europe, and on habitat types and sites. However, the amount of information collected on each species is variable.

EUNIS offers a search facility for species, habitat types and sites modules. There is a set of predefined search functions that perform the most often used types of queries. Information on species is provided in some or all of the following categories: names, taxonomic classification, groups, synonyms, country/region, international threat status, national threat status, legal instruments, habitat types, sites, and references. Although data on some species are limited, the synonyms database is useful.

3.13 Relationship of specialised reference sources with comprehensive reference sources

All of the specialised datasets considered in this report can be queried via the web. However, very often, their data can also be accessed through the web portals of comprehensive reference sources such as GBIF and Species 2000, which compile information from a range of specialised data sources. In many cases, official agreements and partnerships have been developed. A list of data providers to GBIF, ITIS, Species 2000 and All Species is given in Annex 4.

Data are usually provided by specialised TAAs to the comprehensive TAAs in one of two ways: either the data are actually sent, in one form or another, to the comprehensive TAAs, or the comprehensive TAAs query the database of the specialised TAA and retrieve the information required. Batch processing is usually possible in order to process multiple information types at once. Having established the type of data that it requires, IABIN can then determine the best method of data retrieval from the relevant TAAs. IABIN will need to adopt metadata exchange standards and communication protocols (selected examples are briefly presented in Chapter 2 of Annex 1).

Through its activities cataloguing fish species of the world, **FishBase** is an important partner to many comprehensive reference sources. It contributes data to the Catalogue of Life Partnership, and indirectly to GBIF via Catalogue of Life Partnership. Similarly **ILDIS**, which catalogues legume species of the world, is one of the founder members of Species 2000 and an important partner to many comprehensive reference sources. It provides data to the Catalogue of Life dynamic checklist of all known organisms, and indirectly to GBIF via Catalogue of Life Partnership.

CABI Bioscience catalogues fungal species of the world and co-ordinates the fungal component of the Species 2000 project. It is an important contributor to the Catalogue of Life Partnership, an associate participant in GBIF, and is also queried by All Species. Funding from GBIF (July 2003-December 2004) has been provided to augment IndexFungorum and expand SpeciesFungorum.

As Index Kewensis does not provide an accepted list of plant names or synonyms, and does not include an accepted taxonomy, it is of limited use to comprehensive reference sources. However, its **IPNI** partner, the International Organisation for Plant Information contributes to Species 2000. Copies, including electronic copies, may be made of the data held in IPNI for personal use or for use within an organisation, and larger datasets require the permission of the editors.

CGIAR SINGER is not currently a data provider to the Catalogue of Life Partnership or All Species, although its data can be accessed through FishBase. However, it is in the early stages of collaboration with GBIF and hopes to sign an MoU with GBIF in the near future. SINGER is committed to global information exchange by developing links with national, regional, crop and other genetic resources-related information systems.

UNEP has signed an MoU with GBIF and is a GBIF associate participant. **UNEP-WCMC** has been designated as the UNEP node to GBIF, although no exchange of data has yet occurred. UNEP-WCMC contains information on species listed under various multi-lateral environmental agreements which could be exchanged with GBIF. Discussions have taken place to this end but as yet nothing concrete has been implemented.

The **IUCN Red List** has signed an MoU with GBIF and is an associate participant. Although Species 2000 do not directly use the Red List, many of their partner organisations (e.g. FishBase) do. Red Lists for amphibians available through the All Species website via AmphibiaWeb.

EUNIS compiles information collected within the framework of NATURA 2000 (EU Habitats and Birds Directives) including EU and other publications, information from Member States, EU sponsored projects, relevant international conventions, international Red Lists and national threat status.

Data ownership is an issue of concern to many specialised reference sources. Considerable time, effort and money are required to compile taxonomic information. Hence, institutions and individuals can be reluctant to make data freely available and accessible so that others can ‘reap the rewards’ of their labour or in case the data provider is not fully acknowledged. Where data collection is specifically funded through governments or foundations, public access to the resulting data is often a condition of funding, so these issues may be of lesser

concern. In addition, many scientific journals now insist as a condition of publication that the data used in the paper should be made freely available (usually over the Internet). However, where data are gathered in a voluntary capacity, as part of the core activities of an institution, or in any commercial context, these problems with data ownership are likely to persist.

To address this concern, many of the comprehensive reference sources have data user agreements whereby the user must fully acknowledge the source database. Formal contracts can be useful to protect both partners to ensure proper acknowledgement and recognition of the data source, and in the case of disagreement. However, not all databases have formal contracts. For example, All Species sources information directly from specialist websites without necessarily consulting them, though it does fully acknowledge the data source.

CHAPTER 4 TAXONOMIC DIFFERENCES IN DATABASES

4.1 Introduction

As discussed in Chapter 1, large gaps and inconsistencies exist in current taxonomies and competing taxonomies lead to conflicts in information available. As a result, there are many taxonomic conflicts both between and within specialised and comprehensive reference sources. Although in many cases such differences and conflicts cannot easily be resolved, it is important that they are recognised and dealt with as best possible. This chapter outlines the main problems encountered by taxonomic reference sources and some tools and mechanisms that have been used to overcome them.

4.2 Taxonomic Differences in Databases

Following a standard taxonomy means that, for a given taxon, one scientific name should be considered as an accepted name and all other names which refer to that taxon should be considered synonyms. Consistently following a standard taxonomy enables the compilation of a reliable and useful dataset. Differing taxonomic hierarchies and nomenclature for many taxa are used by different organisations, which can be problematic. However, specialised taxonomic databases usually try to follow a standard taxonomy, even if that taxonomic hierarchy is not always complete. How that standard taxonomy is adopted is generally decided by specialist review or by taxonomic working groups established specifically to deal with taxonomic issues. The taxonomy followed is often based upon the most recent and authoritative references available. However, taxonomic decisions may be subjective, as particular experts have a ‘preferred’ taxonomy that they decide to follow. Taxonomic decisions may also be based on funding priorities.

Many comprehensive reference sources pool information from a number of specialised reference sources. Datasets for a particular taxon may be received from more than one reference source, and different taxonomic classifications may have been used by each reference source. Hence more than one accepted name for the particular taxon may be received. The comprehensive reference source may choose to follow one particular taxonomic classification or alternatively to display several taxonomic classifications for a particular species whereby, more than one genus or species name may be listed for a given taxon and the laborious and time-consuming issue of synonymy for many species may be overlooked in favour of including all species names in the databases. In many cases, comprehensive reference sources follow the taxonomies that are provided by their source databases.

As well as adopting a standard taxonomy, the issue of synonymy is a problem that taxonomic databases and reference sources must resolve. It is important that widely used synonyms are included in taxonomic databases and reference sources so that users following alternative taxonomies can find the taxa that they are looking for and so that information from data providers using synonyms can be accessed and understood. However, the huge number of synonyms in the scientific literature makes compiling a complete list of synonyms an almost impossible task.

Links between accepted names and their synonyms should be provided. For example, two species names that refer to the same taxon may both be considered accepted names by two separate data providers. In this case, both names will be retrieved from the database as accepted names, and there should ideally be a facility to indicate that the two are linked. However, this is not an easy task, particularly when data sources (e.g. specialised reference databases) do not supply such information, and particularly considering the number of synonyms that exist.

The occurrence in the literature of homonyms, that is a taxonomic name identical to one previously applied by a different author or to a different taxon, can also be problematic. Misspellings are also a problem.

One further point to note is that not all reference sources include information on the higher taxonomy of the given taxa. Furthermore, if such information is included, many higher taxonomies do not provide a taxonomic sequence, e.g. do not indicate where in a particular class that an order should be placed, but instead list the names of orders alphabetically.

Funding, staff and time constraints are very often the main difficulties encountered when addressing taxonomic problems. In the face of limited funds, most organisations adopt the best system they can manage within their budgets, and fully acknowledge the shortcomings of their systems and information.

4.3 Taxonomic Differences in Databases – Current Practice

An outline of the current practice and some of the tools and mechanisms used by selected taxonomic reference sources to combat taxonomic differences are provided below.

4.3.1 Comprehensive reference sources

GBIF and Species 2000 source all their information from specialised reference sources and/or collections. ITIS, on the other hand, compiles much of its information 'in house' using checklists and other reference material, working in collaboration with experts and specialist organisations. Hence, ITIS must determine the higher taxonomy that it follows, whereas the other two can effectively 'delegate' much or all of this work to the specialised reference sources.

The higher taxonomy currently followed in GBIF is generally determined by the partners that provide it with data. For example, all the higher taxonomy for fungi has been provided by the specialist database CABI (CABI receives contributions from many systematists who assist with such classifications). Additionally, GBIF and the Catalogue of Life Partnership have signed an MoU, and it is likely that the synonymic species checklists provided through the MoU will play a key role in the name-service and indexing functions of the GBIF portal. The forthcoming higher classification (kingdoms and phyla) for the 2005 Catalogue of Life will be useful to GBIF, but GBIF will most likely also remain open to other taxonomies from other data providers. GBIF are currently in the process of dealing with the problem presented when differing views on taxonomy are displayed, to ensure that users are also presented with a clear view of taxonomy. Filtering mechanisms and tools to display multiple taxonomies are being developed, and could be very useful to IABIN.

How GBIF will deal with the numerous museum records and specimens that it catalogues remains to be seen. Currently, higher taxonomy is not always given for museum specimens (e.g. *Loxogramme mexicana*), but instead detailed geographic and specimen information is provided (in this case through Biodiversidad de Costa Rica) but not taxonomic information.

Synonyms for many species are provided in GBIF. The data provider decides whether a name is considered to be a valid name. Although this is a reasonable approach in that it reduces the workload for GBIF, it does mean that the data providers must contain up-to-date and accurate information, particularly recent taxonomic reclassifications and newly described species. For many species, ITIS is used as the source data, and, although ITIS (in conjunction with its Catalogue of Life Partners) is undoubtedly one of the most comprehensive sources available, the huge task of ensuring that their information is up-to-date has been slow to be realised (see below). This problem is not unique to ITIS.

All information in **ITIS** follows the five Kingdom taxonomic code. The higher taxonomy followed and the level of synonymy reported by ITIS varies between taxonomic groups, and depends on the most recent references available and the capacity of the stewards reviewing that particular group. For a given taxonomic

group, several competing taxonomies may be entered into the database. ITIS Canada has developed a tool through which competing taxonomies can be compared. Unfortunately, adequate data sets that would enable proper use of this tool have not been received and so it is not currently in use. Much work would be required to input sufficient data, to use this tool across all taxonomic groups.

In general, the main tool used by ITIS to determine an agreed higher taxonomy is that of expert input and review. It is intended that all taxonomic groups will be reviewed and updated as soon as possible although the enormous workload has slowed this process somewhat. Emphasis is placed on species level information, so subspecies information is not complete for many groups. At present there is no formal system in place to enter information on newly described species. Instead, such information is entered into the database each time the taxonomic group is reviewed. However, to improve this situation ITIS is considering funding a position that would deal adequately with the addition of newly described species.

The current aim of ITIS is to enter widely used synonyms, and species level synonyms are a priority. For many species, ITIS lists synonyms and indicates whether a species is considered valid or not. However, for some species where there is taxonomic uncertainty (e.g. *Loxodonta cyclotis* and *Loxodonta africana*) both species are listed as valid. One shortcoming of this system is that a species that is considered by some authorities to be a synonym of another species, is not linked visibly to that species, so the user may not realise that both names may actually be synonymous.

Species 2000 has established a Taxonomic Working Group, which deals with taxonomic issues. This group will work this year to implement a revised higher (kingdoms and phyla) classification for the 2005 Catalogue of Life. Currently, the information that can be accessed through Species 2000 does not necessarily follow the five Kingdom taxonomic code. Hence, data providers are required to only offer information up to Family level. For many groups the higher taxonomy is supplied but for others e.g. many species of plant, it is not. Species 2000 lists synonyms where they are provided by their source databases. As much of the information in Species 2000 is provided by ITIS, it follows that the taxonomic conflicts present in the ITIS data will be mirrored. Owever, Species 200 are also considering the possibility of allowing users to view different taxonomic hierarchies for a given taxon.

The website of **All Species** states that "*although we recognize the importance of the species debate, we also recognize that ALL Species cannot solve it but hope that our efforts will contribute to an acceptable resolution"*. Advisors to All Species include taxonomists and systematists. However, because All Species links

directly to other databases, it does not need to deal with the taxonomic difficulties faced by the other databases.

4.3.2 Specialised datasets

FishBase follows a standard and consistent taxonomy, including synonyms, based on a number of reference sources that are cross-checked with other sources. The original goal was to provide an overall framework of orders, families and subfamilies (with occasional use of suborders) but difficulties with determining an agreed classification of fishes have hindered this to some extent, and emphasis was placed on inputting correct genera and species information. Fishbase does contain information for the 62 orders of fish down to species or subspecies level, but only about 120 families have been checked, and not all recent revisions have been employed. All family names and higher taxa, as well as the genera assigned to a family, have been matched electronically against *Eschmeyer's Catalog of Fishes* database.

To detect errors in the data, a number of approaches are taken. Names, authors, and distributional ranges are double-checked against available literature, using more than one source wherever possible. All assigned original combinations to all valid names are checked against Eschmeyer's (1998) *Catalog of Fishes* database. In FishBase 98, this was achieved for all valid names and for most junior synonyms, and it is routinely done for every new name that is added to FishBase. Additionally, FishBase matches its names against other available databases. To this end, a routine has been developed that examines lists of scientific names of fishes, identifies synonymous and misspelled names, and makes suggestions for the most probable correct name or spelling. FishBase now contains more than 70,000 synonyms, including junior synonyms, new combinations, misspellings, and misidentifications, along with over 25,000 valid names, all of which are drawn from references and checklists.

The **Index Kewensis** is an index of plant names, and its compilers emphasise that it should not be considered as a list of accepted names, or as an authority regarding synonyms. However, Index Kewensis and data from the Gray Card Index (GCI) and the Australian Plant Names Index (APNI) are compiled to produce the **International Plant Names Index** (IPNI). However, records from each of the three sources in IPNI contained different levels of information, which were of varying quality, and many records contained errors. In addition, IPNI did not want to automatically exclude duplicates as useful information may be tagged to the duplicate record. When merging the records, careful checking of the data by expert staff was undertaken, and many data have been checked back to their original place of publication. A number of editorial processes are planned to improve the quality of the data over the coming years, which include "de-duplication, standardisation and verification".

One tool used by IPNI with regards to the addition of new names into the database is the ability to flag new names, or amendments to existing name records so that the user can view the new name but is aware that it does not yet carry the stamp of editorial approval. Once the editor has reviewed the contribution it is flagged as 'approved' and the update is made to the database. If a contribution is rejected, it is stored and can still be viewed. In such a way the history of the names and the changing taxonomy can be viewed.

ILDIS uses a consistent classification, edited and updated by a network of experts, known as Taxonomic Co-ordinators. ILDIS provides information on the correct tribal positioning of all extant legume genera as determined by the participating experts. In addition, recent research and reference material is used to update the database.

Funding has been found to be one of the main obstacles to taxonomic work of **CABI Biosciences**. To compile taxonomic hierarchies, a number of primary references are used, as well as information provided by taxonomic experts and contributors to the Dictionary of Fungi. Taxonomic updates above the genus level are done every five years, but at the species level updating is less complete, due to lack of funding. Newly described species are sourced from the Index of Fungi.

The main mechanisms used by **CABI** to deal with taxonomic problems are simply expert input and the ability of the CABI staff to manage and organise data appropriately. Errors in the data are identified through informal messages and are subsequently corrected. Avoiding duplication of data is dealt with manually and no specific tools are available.

CGIAR SINGER uses internationally agreed conventions for describing germplasm accessions, as well as authoritative references for taxonomy and country names. Information is provided by a collaboration of experts responsible for managing the individual databases. SINGER is using its leading position to promote common standards, including for key taxonomic descriptors worldwide, to allow system-wide access and searches across multiple databases, whilst retaining the autonomous structure and management of the individual databases.

The taxonomy followed in the **UNEP-WCMC** for many taxonomic groups is that adopted by a number of MEAs. Where no such standard is used, the most recent and authoritative reference materials are used to determine higher taxonomy and synonymy.

The **IUCN Red List** does not purport to be a taxonomic authority. The website notes *"the lack of sufficiently clear taxonomic standards"*. However, taxonomic standards are being adopted and all species on the Red List should conform to these by the year 2004. All new species' listings, and any revisions to listings,

must be in accordance with these taxonomic standards, but deviations are permitted provided they are fully documented and substantiated. The documentation requirements and taxonomic standards will be reviewed at regular intervals. The taxonomic standards used by the IUCN Red List are determined by the SSC Specialist Groups and other experts and organisations, where appropriate, using recent authoritative reference sources.

4.3.3 Interoperability

GBIF and Species 2000 receive data from diverse sources, and those data are often in correspondingly diverse formats. A number of data standards have been developed by different organisations to share their particular data through one portal (e.g. BioCase, Spice). However, the development of an overall architectural framework for exchange of data is required where the protocol and data format(s) on offer are clearly identified, and where these data can be retrieved and compiled from many different sources. GBIF have been working on exchange models, which can successfully retrieve information from multiple sources, and which can recognise and filter duplicates. Moreover, they have done so thus far on a remarkably modest budget by using widely used modern software, protocols and standards. GBIF intends to eventually extend the interoperability of their database from specimen and name data, which has been the focus to date, to molecular data and ecological data. It has been suggested[1] that GBIF can provide a model for other federated database projects, which can in turn contribute new ideas to GBIF and possibly even share development costs for some kinds of basic federated database infrastructure. GBIF is also looking at ways to deal with synonymy and other such problems, as it compiles data from diverse sources.

Specialised reference sources house their data in a variety of ways in systems of varying complexity. It is important that the architecture developed by any taxonomic network such as IABIN will facilitate data retrieval from institutions that may not have significant technical capacity or expertise.

4.4 Case Study - National Biodiversity Network (NBN) of the UK

The National Biodiversity Network (NBN) is a project to build the UK's first network of biodiversity information. It is building tools to make wildlife information accessible in a digitised and exchangeable form. Many organisations provide biodiversity data and many data include taxonomic components or are entirely taxonomic in content. The NBN places copies of all datasets provided into a single database. This is considered to be a useful approach, particularly in order to identify errors. In addition, technical complexities are avoided, as many of their

[1] McCarthy, J.L. 2004. Technical review of network architecture for the Global Biodiversity Facility (GBIF).

data providers do not have the means to operate within more technically complicated systems.

In compiling lists of species and merging data from many sources, a number of problems have been encountered, and a number of mechanisms have been used to deal with them. The main problems include:

- The prevalence of homonyms in the data provided;
- The occurrence of synonyms in the data provided;
- Species names that are incomplete or that have incorrect information attached, most often missing authorities;
- Different recorders and organisations sometimes use different higher taxonomies presenting difficulties when querying data provider databases;
- In the NBN database species and genera names are stored in the same field, allowing complex names to be stored. However this presents problems when querying and linking to other databases;
- Of the multiple databases sourced, very few have fields in common.

One of the main tools that has been developed is the Species Dictionary. This provides a standard reference for names of organisms found in the United Kingdom. It is a project of the National Biodiversity Network and is hosted and managed by The Natural History Museum, London.

The Dictionary database is a collection of lists of names. Individual lists can be searched or organisms can be located by their common or scientific name. A basic classification of life is provided as a navigational tool. The dictionary can handle multiple classifications and comparisons can be run. Although initially designed to hold static published lists, it is intended to move towards a dynamic list with input from experts and partner organisations. In addition, through the dictionary a consensus higher classification will be built that should help organise and standardise the data. When determining the classifications that have been used, many sources and experts are referenced. When dealing with homonyms, assistance from organisations outside the UK was sought.

In order to deal with species names that are incomplete or that have incorrect information attached, a names server is being developed which will flag up misspellings etc and will convert it to the correct citation. However, none of the original lists provided is edited directly ensuring that it can be reproduced if necessary. This names server should be functioning for approximately six groups, including reptiles, amphibians, mammals, and lepidoptera, by May 2004.

So far, NBN has not achieved good coverage of synonyms. It contains accepted names for two thirds of all described taxonomic groups in the UK, partially accepted names for one quarter of groups and about one tenth with no approved

names. However, approximately 85% of these groups do not have good coverage of synonyms. This will be addressed according to user priority and may use information provided by Species 2000.

Because species and genera names are stored in the same field in the NBN database, difficulties have been encountered when querying and linking to other databases. To overcome these difficulties, an additional field has been added to the table that allow a 'word in name' query to be performed. NBN is developing tools to extract data from other databases where few of the databases have fields in common.

As many database searches are often through common name, a tool called 'Nature Navigator' has been developed. Nature Navigator provides a single access point to information on more than 8,000 of the best-known species that occur in Britain. It provides the scientific name and classification for the common names searched and is aimed at the general public. In addition, a 'Working Navigational Hierarchy' is being developed to do a super search of every preferred name across all of the providers' data, but this is not yet an effective tool.

The NBN has developed partnerships with many organisations in the UK. Many organisations came forward of their own accord, indicating their wish to provide data to the NBN. The NBN is currently analysing the gaps in the data obtained so far and is approaching organisations that might be able to fill in those gaps.

The NBN has developed data licence templates and uses formal contracts between them and their data supplier. This guarantees that NBN is authorised to use the data, and that the data provider will be fully acknowledged.

Note: Charles Hussey at the NHM, London has offered that the NHM could provide IABIN with further assistance both advisory and scientific, should it be required. Contact Vanessa Pike, email: v.pike@nhm.ac.uk

4.5 Case study – EUNIS

EUNIS - European Nature Information System collects and maintains data so that it can be used as a reference tool or dataset for assistance to the NATURA 2000 process (EU Birds and Habitats Directives), and coordinated with the related EMERALD Network of the Bern Convention; for the development of indicators (EEA Core Set, Bio-IMPS, IRENA); and for environmental reporting connected to European Environment Agency reporting activities. As a regional information system, it is likely to be of interest to IABIN. However, EUNIS collects information on only subsets of European species, and does not aim to collect information on all species in the region. Nor does it involve collaboration between numerous national organisations in the same way that IABIN will likely operate.

Instead, it includes data compiled in the framework of NATURA2000 (EU Habitats and Birds Directives), data collected from frameworks, data sources or material published by ETC/NPB (formerly the European Topic Centre for Nature Conservation), information taken into account in relevant international conventions or from International Red Lists and specific data collected in the framework of the EEA's reporting activities.

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

GBIF is emerging as a leading global player in the field of comprehensive reference material with strong support from the ITIS and Species 2000 partnership, the Catalogue of Life Partnership. The emphasis of All Species appears to have shifted away from the niches occupied by the other three.

Several of the specialised reference sources reviewed in this report, are the leading electronic sources of global information for particular taxonomic groups, The quality of information provided by such specialised reference sources often determines the quality of the information held by the comprehensive reference sources. Many of the specialised TAAs have already established partnerships with the comprehensive reference TAAs outlined above. These partnerships and the resulting provision of data are important factors in the taxonomic coverage and success of the comprehensive TAAs. The IUCN Red List and UNEP-WCMC species database provide important information on subsets of species that may be useful to IABIN.

A consistent taxonomic approach is essential for the reliable interpretation and use of data. The main mechanisms used by specialist databases to deal with taxonomic problems are their collaborations with experts and organisations from around the world. Access to authoritative reference sources, expert review and expert input to decision making appear to be the most important factors in determining the taxonomies used, the application of synonyms in the databases, and the elimination of problems and inconsistencies. Many specialised reference sources can provide an acceptable taxonomy that can be used as a standard taxonomy by the comprehensive TAAs.

Concerning the comprehensive reference sources, many of the taxonomic issues and problems encountered have not yet been fully addressed. Although for many taxa the specialist reference sources will have decided upon an appropriate taxonomy and synonymy to follow, some of the comprehensive reference sources include competing or conflicting taxonomies without necessarily displaying them as such. Although a tool attempting to deal with this has been developed by ITIS Canada, lack of proper data has prevented the use of this system. ITIS Mexico displays competing taxonomies in some instances. GBIF is currently developing technologies and tools to deal with synonyms and competing taxonomies in a more comprehensive way. Technologies that deal with data exchange formats and communication protocols are also being developed.

Although the incorporation of taxonomic information into IABIN could be achieved by forming partnerships and bringing together information from the specialist TAAs, this would seem to be needless duplication of the work of the comprehensive TAAs which already have established partnerships and compiled the relevant information. For many taxa, the coverage of ITIS has a North American focus, and it will probably take some time before coverage of taxa across the Americas will be more complete. However, clearly the American focus is of particular relevance to IABIN. In addition, the checklist of species produced by the Catalogue of Life could prove to be a very useful tool for IABIN depending on its requirements. However, GBIF includes information from ITIS and many other sources in its archive, and also includes a range of other information including that from museum and herbarium collections.

Upon inclusion of all species names in a comprehensive TAA, many types of information can then be attached to the species names. It is suggested that a complete distribution, legislation and global threat status may be useful types of information to include on species of the Americas due to the species specific nature and range of end users of such information. However, it must be noted that the distribution information in many of the comprehensive databases is not complete and at times is misleading. Regional and national organisations in the Americas may be able to assist in improving the quality of such information.

Where national biodiversity archives have been or are currently being developed e.g. that of Colombia, the information compiled can fill geographic gaps for particular taxa in the Americas, and can be used to include information on national status. However, there are also likely to be many taxonomic groups for which little information has been collected across the entire region.

5.2 Recommendations

The overall recommendation is that IABIN maximises the benefits available to the Network through the utilisation of all existing and available data and tools, rather than to duplicate work that has already been undertaken by other agencies. Financial and other resources will be most efficiently and usefully employed by identifying gaps in existing data and technologies, and focusing on these gaps. Full use of available expertise from national, regional and international sources will be key.

- Species lists with taxonomic hierarchy, accepted names, synonyms and common names will be the cornerstone of the taxonomic information held by IABIN. Such lists have already been developed for some taxa of the Americas. It is recommended that IABIN make maximum use of those lists that are already available, both through international reference sources such as the Catalogue of Life Partnership and GBIF (and its Electronic Catalogue

of Names of Known Organisms Programme (ECAT)), and through existing regional and national bodies. To this end, it is recommended that IABIN should liaise closely with the Catalogue of Life Partnership through which comprehensive species lists for some taxa are already available. In particular, the regional and scientific expertise of the ITIS partner CONABIO will be important.

- It is suggested that IABIN include all species names that are considered to be accepted names by different authorities so that relevant information can be retrieved from other databases or networks. However, it is suggested that IABIN uses technologies that are being developed by GBIF and others to display different taxonomic classifications and nomenclature, or adopts one standard taxonomic classification, so that relevant information can be consistently attached to the correct taxon name. Links between accepted names and their synonyms should be provided, again so that relevant information can be accessed. Discussions with ITIS, Species 2000, GBIF and regional data providers should provide some advice as to the various taxonomic classifications widely used.
- It is recommended that IABIN liaise closely with GBIF and its ECAT Programme on the most appropriate technologies to use. GBIF are developing technical standards and technologies to assist in the retrieval of information from various sources, and developing an overall architectural framework for data exchange. IABIN should avoid investing significant resources in developing their own tools that would duplicate GBIF's work. IABIN might find it useful to consult with other networks (such as the UK NBN) that receive data from sources that do not have technical capacity or expertise.
- It is important that IABIN should receive support from within the region. To this end, national and regional organisations should perceive IABIN as a partnership that is assisted by the aforementioned TAAs, but that is not directed by them. It is recommended that IABIN should identify appropriate taxonomic organisations at a national and regional level, and develop partnerships with them from the outset.
- Regional gaps in datasets are inevitable. It is suggested that IABIN should prioritise areas for new research along particular themes or in areas that will have a maximum number of uses and end users e.g. invasive species, species of particular ecosystems, bioprospecting, etc. Such themes should also provide practical examples of the political, economic and social benefits of taxonomic and related information, such as that provided by IABIN. Such thematic projects are more likely to attract external funding than stand-alone taxonomic projects. Seed money available through the GBIF Programme ECAT may be of particular interest to IABIN.

- The input and review of relevant experts and organisations is a fundamental aspect of the management of taxonomic databases. Many TAAs have established specific groups that liaise with the wider scientific community to deal with taxonomic problems. It is recommended that a taxonomic advisory group within IABIN should be established to liaise with the TAAs and to deal with issues such as taxonomic and other standards, synonyms, etc.
- It is also recommended that IABIN should develop formal agreements with its TAA partners to ensure that full acknowledgement is given to the source. This will help address the discomfort that many organisations experience when supplying their data in its entirety to be used by others.

ANNEX 1 –Key Specimen Collections Relevant to the Region

CHAPTER 1 INTRODUCTION

1.1 Summary

Some major collections of natural history specimens from the Americas are kept in museums/herbaria in European countries. In addition, some other countries such as Japan have smaller collections from the Americas. Although the digitisation of many of these collections has begun, the proportion of digitised records remains small. Some organisations are already working on biological information sharing activities, such as GBIF, ITIS, BioCISE, and could feed back into IABIN activities.

Seventeen natural history museums or institutions are identified as potential partners for IABIN regarding specimen collections. Material held by these institutions can be broadly categorised as (a) basic collections for faunal and floral research, (b) historical collections before the mid 20th century, and (c) local collections from the IABIN region held outside the IABIN region.

To ensure safe data exchange and information sharing through IABIN, metadata standards and protocols must be adopted.

1.2 Objectives

The aim of this annex is to review selected biological information sharing initiatives and specimen collections outside the IABIN region, the type of information that is held and the extent to which it is already digitised. The annex also assesses the current status of selected data standards to facilitate the exchange and sharing of biological information.

1.3 Research Method

Document reviews and web searches were adopted as the main research methods. Target specimen collections were as follows:

- Basic Collections for faunal and floral research;
- Historical Collections from before the 20th century;
- Local collections specific to the IABIN countries.
- Local collections in Japan specific to the IABIN countries.

With respect to the data standards and protocols for communication and exchange, several biological information-sharing initiatives were reviewed.

CHAPTER 2 REVIEW OF ONGOING INITIATIVES CONCERNING NATURAL HISTORY COLLECTIONS

2.1 Overview

This chapter reviews biological information-sharing activities, with particular emphasis on data standards, undertaken by the following initiatives:

- GBIF (Global Biodiversity Information Facility);
- GTI (Global Taxonomic Initiative);
- CHM (Clearing House Mechanism);
- BioNET;
- ITIS (Integrated Taxonomic Information System);
- BioCASE (Biological Collection Access Service for Europe);
- BioCISE (Biological Collection Information Service in Europe);
- ENHSIN (European Natural History Specimen Information Network).

2.1.1 GBIF (Global Biodiversity Information Facility)

GBIF emerged from the OECD Megascience Forum Working Group, and was established in 2001. The mission of the Global Biodiversity Information Facility (GBIF) is to make the world's primary data on biodiversity freely and universally available via the Internet. GBIF is now working on encouraging, coordinating and supporting the development of worldwide capacity to access the vast amount of biodiversity data held in natural history museum collections, libraries and databanks. Near-term GBIF developments will focus on species and specimen-level data.

After three years of activity the GBIF portal site has begun to provide a service for searching specimen information from providers in certain countries. As of February 2004, about 12 million specimen records, from 34 institutes of 14 countries, are available from the GBIF portal (http://www.gbif.org/).

2.1.2 GTI (Global Taxonomy Initiative)

The aim of the Global Taxonomy Initiative (GTI) is to enable the provision of appropriate taxonomic information and capacity to underpin decision-making in conservation of biological diversity (CBD), sustainable use of its components and equitable sharing of the benefits derived from the utilization of genetic resources. This will be achieved by addressing (a) the lack of taxonomic information for identifying the species components in biological diversity, and (b) the need to build capacity for taxonomic activity in all regions, but especially developing

countries. This includes reference materials, databases, and taxonomic expertise relevant to the objectives of the Convention on Biological Diversity.

2.1.3 CHM (Clearing-House Mechanism)

Recognising the importance of biodiversity to our daily lives and the pressure that human activities are placing on our living world, governments adopted the Convention on Biological Diversity in 1992 as a framework for action. From the start it was understood that scientific knowledge and technological know-how would have a vital role to play. However, expertise in managing information and technology varies enormously from country to country. For this reason, the Convention established a Clearing-House Mechanism to ensure that all governments have access to the information, technologies and experience they need for their work on biodiversity.

2.1.4 BioNET-International

BioNET- International, the Global Network for Taxonomy, is dedicated to creating sustainable mechanisms to assist developing countries to overcome the Taxonomic Impediment by becoming self-reliant in taxonomy (i.e. in the skills, infrastructure (collections etc.) and technologies needed to discover, identify, name and classify, and understand the relationships of all organisms on this planet).

BioNET aims to establish subregional LOOPs (Locally Organised and Operated Partnerships, of which there are two operating in the Caribbean (CARINET) and Andean Countries (ANDINONET), and others under development or consideration in Meso America, North America and the South of Latin America.

2.1.5 BioCASE (Biological Collection Access Service for Europe)

BioCISE and the subsequent 'BioCASE' are regional collaboration programmes among EU member countries, originating from the ENHSIN, a national programme for natural history museums to share biological information among themselves. All of these activities aim ultimately to share information through the Internet.

BioCISE (Biological Collection Information Service in Europe), the predecessor of BioCASE, was a means of identifying biological information resources (collections and databases), cataloguing interdisciplinary biodiversity database expertise, and providing guidelines for the incorporation of collection information in databases. BioCISE operated at the level of the collection (coverage, characteristics and other information) and worked with a particular data model to develop a consistent metadata system for collections.

BioCASE is based on the results of its predecessor project, BioCISE and also on ENHSIN. The BioCASE project is a project to build a networked biological collection information system for Europe and Israel. The network consists of 31 national nodes providing metadata on collections and collection databases in their country through a daily updating mechanism. These metadata are used to populate a database serving as a registry for clients accessing unit level databases.

2.1.6 ENHSIN (European Natural History Specimen Information Network)

The aim of the European Natural History Specimen Information Network (ENHSIN) is to enable the development of a shared, interoperable infrastructure of natural history specimen databases in European institutions. Although new developments have been made in producing frameworks for connecting global species databases and for providing access to the wider content of European natural history collections, there is, at present, no corresponding approach to facilitate access to specimen data. ENHSIN is intended to fill this gap. Seven European organisations are involved in the initial phase of the project. Their aim is to create an operational system for what is hoped to evolve into a pan-European network. ENHSIN is proposing to develop access at the level of the specimen, and it is essential that this development is compatible with outputs of the BioCISE and subsequent BioCise initiative for the collections. BioCISE dealt with harmonised approaches to information on a collections scale, whereas ENHSIN addresses the need for common access to information on the actual specimen. Both kinds of information are important to users and part of the ENHSIN task will be to address complementarity and possible harmonised development with initiatives such as BioCISE. A further important difference is in coverage: BioCISE dealt with biological collections, while the collections within EHNSIN have wider coverage – not only biological, but mineralogical.

2.2 Standards for Information Sharing

Because biodiversity information such as specimen data is managed by the organisation at which they are stored, it is necessary to establish a meta-database system for integrating all available information. This approach has been employed by many biodiversity databases.

Decisions must be made on a data exchange format and a communication protocol between meta-database and data providers. The TADWG (Taxonomy Database Working Group) and CODATA (Committee on Data for Science and Technology) have had extensive discussions regarding the production of a world standard for taxonomic databases. Recently, DADI (Data Access and Database Interoperability) of GBIF, decided to use DarwinCore ver.2 as a data exchange for specimen databases. Through the GBIF portal one can search many databases

of data providers all at once, and it is evident that the DiGIR protocol is used for querying and for data exchange between a portal and data provider.

Standards can be broadly classified into: (a) Data exchange standard (i.e. metadata), and (b) communication protocol for data exchange.

2.2.1 Metadata

(1) History

The discussion on the database schema standards in the field of biological information is rooted in the specimen exchange system. Originally this was developed for the purpose of interchanging specimen collections among natural history museums. These include: (a) Floristic regions of the world, (b) Herbarium Information standard and protocol for interchange data, (c) Index of Herbarium, (d) Plant name in botanical database, (e) Plant distribution and status scheme. The TDWG (Taxonomic Database Working Group) was established for discussion of these matters.

(2) ENHSIN

During a three-year pilot project phase, ENHSIN proposed metadata standards for exchanging information. This metadata consisted of the four elements of information:

- On species: Attribute (i.e. taxonomic information, such as family and kingdom);
- On species: Gathering Area (i.e. longitude, latitude, elevation, date);
- On Holding Institution; and
- For Data Management (i.e. revision, quality control).

The proposed metadata in ENHSIN defines the types of data for each item, (i.e. float, string, integer), together with a decoding rule in XML format.

(3) Darwin Core

Darwin Core is a simple XML-based exchange format for specimen and observation data. Originally it was developed in 2001 as version 1. Currently the Darwin Core is updated with additional data elements and exists as version 2. The current version of Darwin Core comprises field information such as taxonomic information on scientific name, kingdom, class, order, family, genus, species, and subspecies. It also contains information on location and data, together with the name of the collector. In Darwin Core 2.0, many data fields have been defined and added, and the number of required fields (the field that must be contained) is only

five in total. This is because this schema had been created with the priority of being easily adaptable with existing databases.

(4) ABCD (Access to Biological Collection Data)

ABCD is a metadata standard proposed by a joint initiative between CODATA and TDWG. It consists of five parts: (a) Conventions, (b) General hierarchy, (c) Full returned dataset, (d) Component dataset, and (e) Unit dataset. With respect to Unit dataset, the following information is supposed to be stored in an XML format.

- Gathering event data (Project name, gathering agent, Date/time);
- Gathering site data (Country, Coordinates, Altitude, Habitat etc.);
- Identification data.

This metadata standard is a possible future option in the field of biological information sharing, since GBIF is now evaluating this data schema as one of the potential standard Data Transmission Protocols.

Presently, there are some transmission protocols that are applied in TCP/IP on the Internet. These are: Z39.50, SOAP, XML and UDDI. These protocols act under the decentralised, distributional database environment on a server-client model on the Internet.

(5) DiGER (Distributed Generic Information Retrieval)

This is a protocol under the HTTP, XML environment. It was developed in 2001, as an output of the meeting at Santa Barbara in the United States. DiGER is now using PHP as a programming language and thus is an OS-independent protocol and offers a retrieval system. This system is available on WINDOWS and Linux, and can be accessed via the GBIF portal site.

This protocol is gradually being accepted as a standard communication protocol in application layer. For example, GBIF decided to employ this communication protocol between portal and data providers, and now 34 institutes have adopted this protocol for data exchange. Also, there are many potential users such as OBIS, MapSTeDI,SALVIAS.

(6) Z39.50

Z39.50 is a protocol based on the ANSI/ISO standard. It offers a query and retrieval between remote computers through the Internet. This standard includes: (a) query languages, (b) record syntax option for transferring a results record, and (c) query types information. This protocol is widely used in a Library retrieval system, such as OPAC.

(7) SOAP (Simple Object Access Protocol)

This XML-based protocol comprises four parts: (a) envelope, (b) encoding rule, (c) remote procedure call, and (d) response. This protocol acts on the HTTP protocol commonly used in the Web service.

2.2.2 Future of the Standard for Sharing Biological Information

Much discussion on metadata exchange formats and communication protocols is still required. IABIN may wish to adopt the same protocols as the GBIF portal providing specimens and observation data. GBIF use DarwinCore 2.0 as a data exchange standard, and DiGIR as a communication protocol. As DarwinCore 2.0 has a minimum requirement of essential fields, many variants of database schema could be created, because there are many optional fields. This flexibility is one reason for adopting this format as standard, because it would minimise the need to change the data structure in existing databases. One disadvantage of the DiGIR is its difficulty of implementation into each system.

Alternatively, particularly in the context of building a new database or network, IABIN may wish to adopt the ABCD data schema, which is comprehensive and compatible to the DarwinCore ver.2.0. However, there are many data providers that have adopted the DiGIR protocol, because it is used in the GBIF portal. The number of DiGIR providers will increase due to the implementation of national nodes of GBIF. A list of available DiGIR providers can be found at http://www.gbif.org

CHAPTER 3 NATURAL HISTORY COLLECTIONS RELEVANT TO IABIN

3.1 Introduction

This chapter gives an overview of the subsequent Chapters 4, 5, and 6 that deal with selected natural history collections relevant to the IABIN region. These specimen collections can roughly be classified into three broad categories, as follows:

- Basic collections for faunal and floral research;
- Historical collections (before mid 20th century);
- Local collections in Europe and Japan.

The collections reviewed in the following chapters are those held in institutions in Europe and Japan.

3.2 Basic collections for faunal and floral research

These collections were collected mainly for faunal and floral research, and thus include many type specimens. As these collections have usually been accurately identified, they should be a priority for digitising.

3.3 Historical collections (before mid 20th century)

Older collections are important both because they include many type specimens, and also because they help establish changes in distribution of species. Many historical collections from the Americas are preserved in Europe (particularly in the UK and Spain).

3.4 Local collections in each countries or regions

Large-scale collections, collected by zoological and botanical expeditions to the IABIN region, can be found in museums and herbaria in Europe and Japan. The coverage of these collections is not, however, sufficient to determine the detailed distribution of organisms, or their recent status.

Chapters 4, 5 and 6 below identify the locations of selected collections of the principal animal, plant, and fungal taxa, and provide summary descriptions of the collections, and sources of further information.

CHAPTER 4 ANIMAL COLLECTIONS

4.1 Introduction

4.1.1 Selected collections outside of the Americas

The Natural History Museum, London, U.K. (http://www.nhm.ac.uk)

The collections of the UK's Natural History Museum comprise all recent animal groups. With the exception of insects and arachnids, there are an estimated 27 million specimens, with a worldwide coverage of many taxa. The collections are particularly rich in material from former colonial countries and in type, rare and historic material. The bulk of the collections were assembled in the late nineteenth and early twentieth centuries, but they include some sixteenth century and modern material. Many specimens originate from the work of famous zoologists, including Linnaeus, Darwin, Wallace, Rothschild, Hooker, Sloane, Lyell and Sowerby, and from voyages of exploration, including those of Challenger, Alert, Investigator and Discovery.

National Museum of Natural History Leiden, Netherlands
(http://www.naturalis.nl/english/index2.html)

This collection includes 570,000 vertebrates, 2,290,000 invertebrates, excluding insects, and 5,250,000 insects.

The National Science Museum, Japan

This collection includes large numbers of mammal, fish, and invertebrate specimens.

4.2 Mammals

4.2.1 Selected collections outside of the Americas

The Natural History Museum, London, U.K. (http://www.nhm.ac.uk)

The mammal collection of the Natural History Museum contains about 359,000 specimens, including more than 8,000 types.

The National Science Museum, Japan

This collection holds the largest specimen collections in Japan. There are about 30,000 specimens of terrestrial mammals, and 3,600 specimens of marine mammals.

The Primates Research Institute, Kyoto University, Japan

The collection holds nearly 6,000 specimens, of which 1,383 are New World monkeys. Details of all the specimens have been entered into a database.

4.3 Birds

4.3.1 Selected collections outside of the Americas

The Natural History Museum, London, U.K

The London collection contains about 2.5 million specimens including more than 9,000 types.

Senckenberg NaturMuseum, Germany
(http://senckenberg.uni-frankfurt.de/sm.htm)

The museum holds 90,000 specimens, including about 400 types, covering about 75% of the known bird species. The geographical emphasis of the collection lies in Central Europe, North Africa and South America. The most important collection parts come from: E. Rueppell, H. Berlepsch (main South America), C. Erlanger (main North Africa), H. Bregulla (tropical Pacific), and F. Behn (Chile).

Yamashina Institute for Ornithology, Japan
The collection includes about 9,000 specimens collected from all over the world, but mainly from East Asia and the Northwestern Pacific.

4.4 Reptiles and Amphibians

4.4.1 Selected collections outside of the Americas

The Natural History Museum, London, U.K.

The collection contains about 200,000 specimens of reptiles and amphibians, including more than 8,500 types.

Senckenberg Museum,Germany

The collection in the Senckenberg Museum contains about 100,000 specimens of amphibians and reptiles including numerous types from all continents. At present the faunal and zoogeographical studies concentrate particularly on the following countries of Central America: Mexico, Nicaragua, Honduras, El Salvador.

4.5 Fish

Taxonomic information for fish is well organised and recorded in Fishbase (http://www.fishbase.org/home.htm), and most information can be obtained from the website.

4.5.1 Selected collections outside of the Americas

The Natural History Museum, London, U.K.
The collection of the London Natural History Museum contains about 2.5 million specimens, including more than 15,000 types.

Senckenberg Natural History Museum, Germany
The collection contains many offshore fish from South America

The National Science Museum, Japan
The collection contains marine fish collected offshore of Suriname. The specimens are well organised and details of some from the Americas have been digitised. The database of the collections is available at http://fishdb.kahaku.go.jp/, which has 6,231 specimens registered from American countries.

Hokkaido University, Japan
The collection contains marine fish collected from Patagonia and Peru. Details of some specimens have been digitised and are available at
http://fishdb.kahaku.go.jp/

Kyoto University (Musum, Maizuru Fisheries Research Station), Japan
The collection contains marine fish collected from Patagonia.

4.6 Invertebrates

4.6.1 Selected collections outside of the Americas

Natural History Museum, London, U.K.
The Natural History Museum collection of insects and other terrestrial arthropods, including spiders, mites and myriapods, comprises up to 28 million specimens. It is the most comprehensive in the world and includes named representatives of about half of the more than one million described species. The collection of Crustacea contains about 3.54 million specimens including more than 25,000 types. The mollusc collection contains about 8 million specimens including more than 80,000 types. The collection contains about 1,000,000 specimens of lower invertebrates, including more than 25,000 types.

Senckenberg Natural History Museum, Germany
There are many collections of insect specimens in the Senckenberg Museum; for example, an estimated 2.4 million specimens of Coleoptera, with about 2 million from the tropics.

The National Science Museum, Japan

The insect collection includes 170,000 Lepidoptera, featuring the Nakahara collection of world butterflies (about 5,500 specimens), which includes many specimens from the Americas. There are also 130,000 Coleoptera (including 1,800 type specimens) from all over the world, again including the Americas. There are up to 50,000 Hymenoptera, with more than 1,200 type specimens; 129,000 Diptera (>1,000 type specimens); and 110,000 Dragonflies (>1,300 type specimens). The Museum also holds the the largest collections of crustaceans in Japan, with about 15,000 specimens, mostly from Japan and adjacent regions, and including about 2,900 type specimens. Many mollusc specimens are also held, mostly from Japan and adjacent areas.

The University Museum, Kyushu University, Japan

This houses the largest collection of insects in Japan, containing 4 million specimens, including more than 10,000 type specimens. The main collections are: Lepidoptera collection from Asia (30,000); Hemiptera collection (140,000); World Lucanid collection; Curculionoid collection (250,000); Diptera collection (200,000); Hymenoptera collection; Sugitani collection of butterflies and moths (36000); Miyagawa collection of weevils (35000). The collections of Kyushu University are on their way to becoming digitised, and their bee collection database is available on-line (http://konchudb.agr.agr.kyushu-u.ac.jp/hanabachi/).

Hokkaido University, Japan

About two million entomological specimens are preserved in the Collection of Systematic Entomology of Hokkaido University, including 10,000 type specimens, mostly from Japan and the Asian region.

CHAPTER 5 PLANT COLLECTIONS

5.1 Introduction

Some of the largest collections of plant specimens (including vascular plants, bryophytes, and algae) from the Americas are preserved in the herbaria of the following institutes:

5.1.1 Selected collections outside of the Americas

Royal Botanic Gardens, Kew, U.K.
There are currently over 7 million specimens at Kew, representing nearly 98% of all of the genera in the world. It has the largest collection of historical plant specimens (including types), and plant specimens are found from all regions of the world. Kew also has many important resources for plant biodiversity, including Index Kewensis now available on-line. Digitisation is being carried out through project work and during routine specimen curation. Activity is currently concentrated on material from northeastern Brazil, and staff are also digitising the specimens of important plant collectors, such as Richard Spruce.

The Natural History Museum, U.K.
The Department of Botany of the Natural History Museum houses a major international collection of over six million preserved specimens of algae, diatoms, lichens, mosses, ferns, conifers and flowering plants. The collection is of great historical importance, dating back to the sixteenth century, and contains many collections from North, Central and South America.

The National Herbarium of the Netherlands

The National Herbarium of the Netherlands was established in 1999 by the merger of three major university herbaria in Leiden, Utrecht and Wageningen. With a collection of over 5.5 million specimens, the NHN is one of the largest herbaria in the world. All ca. 50,000 type specimens have been listed on the Internet. The Utrecht University branch has extensive collections of tropical American specimens of vascular plants and, during the past two decades, various collectors have brought in around 25,000 specimens from Guyana and French Guiana.

The University of Tokyo Herbarium, Japan
The collection contains 2,000,000 vascular plant specimens, some of which are from the Americas, including specimens collected in the Andes from the 1960s to the 1980s by Profs Fumio Maekawa, Mokoto Nishida and Mikio Ono. Some type specimens in the collection have been digitised and are available through JTYPES (http://foj.c.u-tokyo.ac.jp/jtypes/).

Kyoto University Herbarium, Japan
The collection contains 1,500,000 vascular plant specimens, some of which are from the Americas. Some type specimens in the collection have been digitised and are available through JTYPES (http://foj.c.u-tokyo.ac.jp/jtypes/).

The National Science Museum Herbarium, Japan
The collection contains 700,000 vascular plant specimens, some of which are from the Americas.

Makino Herbarium, Tokyo Metropolitan University, Japan
The collection contains 350,000 vascular plant specimens, some of which are from the Americas. All the type specimens in Makino Herbarium have been digitised and are available on-line, with photographs (http://taxa.soken.ac.jp/MakinoDB/makino/html_e/index.html).

Hokkaido University, Graduate School of Science Herbarium, Japan
The collection holds about 120,000 specimens of seaweeds, some of which are from Central and South America.

Hiroshima University, Japan
At Hiroshima about 400,000 bryophyte specimens are preserved in total, including the Horikawa Collection (70,000); the Suzuki Collection (30,000); the Ando Collection; and the Seki Collection. Two bryophyte collections from the IABIN region are mainly held in this university: one contains about 10,000 specimens collected in Chile and Peru in the 1980s, and the other contains about 1,000 specimens collected in Chile in the 1960s. The type specimens have been digitised and are available on-line.

Hattori Botanical Laboratory
The collection contains 462,948 bryophyte specimens in total, comprising 214,606 from Japan and 230,276 from elsewhere.

National Institute of Polar Research, Japan
The collection includes 2,000 bryophyte specimens collected in Chile by Dr Kanda.

5.1.2 Important historical collections in European herbaria

(1) Richard Spruce Collection

This is a collection by Richard Spruce from the Amazon and Andes in 1849 to 1864, which forms an important botanical, historical and ethnological resource. The Richard Spruce project has been proceeding with collaboration between the Natural History Museum of London and the Royal Botanical Gardens, Kew. This project is creating a database of Spruce collection holdings in the respective herbaria, to make information about them available to a wider audience of

botanists, historians and other parties interested in the exploration of the Amazon and Andes.

(2) *Historical collections of the Real Jardin Botanico de Madrid, Spain* (http://www.rjb.csic.es/)

This herbarium has several historically important collections from Central and South America as described below.

1) Herbario of Ruiz and Pavón
This collection is from the Botanical Expedition to the Virreinato of Peru and Chile (1777-1788). There are 2.264 drawings and 3.000 species.

2) Herbario of Mutis
This is the collection from the Real Botanical Expedition of the Virreinato de Nueva Granada, which visited Columbia, Ecuador and Venezuela. Material comprises 6,717 drawings and 6,000 specimens of 306 taxa.

3) Herbario of Sessé and Mociño
This is from the Real Botanical Expedition to the New Spain (1787-1803), which visited Mexico, Guatemala, Cuba, Puerto Rico, California, and Nootka Bay. It includes 1,335 drawings and 3,500 species, of which 2,500 were new to science).

4) Herbario of Neé
This collection is from the "Malaspina Expedition" of 1789-1794, by Commander Alexander Malaespina, which visited Uruguay, Argentina, Chile, Peru, Ecuador, Panama, Mexico, Gulf of Alaska, Archipelago of Nootka, Guam, the Philippines, Soundings, Molucas, New Guinea, New Zealand, and New Holland. There are 370 drawings and 30,000 specimens.

5) Herbario of Boldo
This collection comes from the Real Commission of Guantánamo, also referred to as 'Expedition to Cuba of Count de Mopox' (1796-1802). It covers Cuba and the USA and comprises 66 drawings and 2,000 specimens.

6) Herbario de Isern
This collection is from the expedition to the central Pacific of (1862-1866), which visited Brazil, Uruguay, Argentina, Chile, Bolivia, Peru and Ecuador. There are 8,176 species.

CHAPTER 6 FUNGI AND LICHENS

6.1 Fungi

6.1.1 Selected collections outside of the Americas

The National Science Museum Herbarium, Japan

The collection contains about 210,000 specimens, some of which are from Alaska, Chile and Peru.

6.2 Lichens

6.2.1 Selected collections outside of the Americas

The National Science Museum Herbarium, Japan

The collection contains about 150,000 specimens, some of which are from Canada, Chile, Mexico, Peru and the USA.

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Specimen Collections held in Natural History Museums and other Institutions

The above analysis has identified 17 natural history museums or institutions that could be potential partners for the IABIN with regards specimen collections. The collections held can be broadly categorised as (a) basic collections for faunal and floral research, (b) historical collections obtained before the middle of the 20th century, and (c) local collections in the IABIN region.

Digitisation of specimens has been undertaken at some institutions in the region, however, digitisation needs to be accelerated, particularly in the case of collections obtained through faunal and floral research. Local collections from the IABIN region held outside the region are mainly in small-scale museums and herbaria. With some exceptions, the budget in these museums is small, so it is difficult to commence digitisation without external assistance. This material will be useful to IABIN, so it is suggested that assistance could be provided to these institutions to enable them to digitise their collections.

7.2 Meeting the Standard for Biological Information Collection

Data exchange and information sharing through IABIN will require using metadata and data transfer protocol standards. As described above, GBIF currently collaborate with the CBD, and the CHM and the GTI under the CBD. Also TADWG, ITIS and Species2000 are partners of GBIF. On the basis of these successful collaborations, it may be appropriate to adopt the GBIF standard as a IABIN standard for metadata exchange and data transfer protocol, since these standards are emerging as a future *de-facto* standard.

7.3 Relevance of specimen collections in Japanese institutions to IABIN

In general, the main specimen collections held in Japan do not appear to hold a substantial number of specimens from the Americas. However, the significance of some species in some collections is inevitably greater than in others.

ANNEX 2 - Key Contacts

Donal Hobern
Programme Officer for Data Access and Database Inetroperability,
Global Biodiversity Information Facility Secretariat,
Universitetsparken 15,
DK-2100 Copenhagen, Denmark.
Tel: + 45 35321483
Email: dhobern@gbif.org
Web Site: www.gbif.org

James Edwards, Executive Secretary
Global Biodiversity Information Facility Secretariat
Universitetsparken 15
DK-2100 Copenhagen, Denmark
Tel: +45 35 32 14 70
Fax: +45 35 32 14 80
Email: jedwards@gbif.org
Web Site: www.gbif.org

Dr. Michael Ruggiero
Director, Integrated Taxonomic Information System (ITIS)
Smithsonian Institution
P.O. Box 37012
NMNH, Room CE-120, MRC – 0180
Washington, DC 20013-7012

Dr. Guy Baillargeon
Agriculture and Agri-Food Canada
c/o Maurice Lamontagne Institute, DFO
850 route de la Mer
P.O. Box 1000 Mont-Joli (Qc) G5H 3Z4
Canada
Telephone: +1-418-775-0741
Fax : +1-418-775-0679
E-mail: baillarg@agr.gc.ca

Prof Frank Bisby
Species 2000 Secretariat,
Centre for Plant Diversity and Systematics,
School of Plant Sciences,
The University of Reading,
Reading RG6 6AS, UK.
Tel: +44 118 378 6466
Fax: +44 118 378 8160
Email: sp2000@sp2000.org

Kevin Kelly
Governing Board Chair
All Species Foundation
P.O. Box 29462
San Francisco, CA 94129 USA
Tel: 415-750-7243
Email: info@all-species.org
Web: www.all-species.org

Boris Fabres,
WorldFish Centre – FishBase Project
3rd Floor, Khush Hall
IRRI College
Los Baños
Laguna 4031, Philippines
Tel: +44 118 378 6466
Fax: +44 63 2 845 0606
Email: b.fabres@cgiar.org
Web: www.fishbase.org

Rosemary Davies,
Index Kewensis IK/IPNI,
Herbarium, Royal Botanic Gardens,
Kew, Richmond,
Surrey, TW9 3AE.
Email: ik@kew.org or ipnieditors@ipni.org or ipnifeedback@ipni.org

Judith Heald
ILDIS Co-ordinating Centre
Centre for Plant Diversity & Systematics
School of Plant Sciences
The University of Reading
Reading RG6 6AS
United Kingdom
Tel: +44 (0)118 931 6466
Fax: +44 (0)118 975 3676
Email: ildis@ildis.org
Web: www.ildis.org

Dr Paul Kirk
CABI Bioscience UK Centre
Bakeham Lane
Egham, Surrey, TW20 9TY UK
Tel: +44 (0) 1491 829023
Fax: +44 (0) 1491 829100
Email. p.kirk@cabi.org
Web www.cabi-bioscience.org

CGIAR SINGER
System-wide Information Network for Genetic Resources (SINGER)
SGRP Secretariat
c/o IPGRI
Via dei Tre Denari, 472/a
00057 Maccarese (Fiumicino)
Rome, Italy
Tel: (+39)0661181
Fax: (+39)0661979661
Email: singer@cgiar.org
Internet: http://singer.cgiar.org

Dr Gerardo Fragoso,
Head of CITES, CMS & Species Programme
UNEP-WCMC
219 Huntingdon Road
Cambridge CB3 0DL
Telephone: ++44 (0) 1223 277314
Fax: ++ 44 (0) 1223 277136
Email: gerardo.fragoso@unep-wcmc.org
Web: www.unep-wcmc.org

Craig Hilton-Taylor
IUCN/SSC Red List Programme Officer
219 Huntingdon Road
Cambridge CB3 0DL
Tel: +44 (0) 1223 277966
Fax +44 (0) 1223 277845
Email: craig.hilton-taylor@ssc-uk.org
Web: www.redlist.org

Professor Ivan Vrana
EUNIS Secretariat
Dept of Information Engineering
University of Agriculture in Prague
Kamýcká, 165 21 Prague 6 – Suchdol
Czech Republic
Fax : +42 02 2438 2274
Email: vrana@pef.czu.cz
Web: www.eunis.org

ANNEX 3 - Acronyms and Abbreviations

APNI	Australian Plant Names Index
AVRDC	Asian Vegetable Research and Development Center
B	Herbarium, Botanischer Garten und Botanisches Museum Berlin-Dahlem, Zentraleinrichtung der Freien Universit 舩 Berlin
BC	Herbarium, Institut Botanic de Barcelona
BioCASE	Biological Collection Access Service for Europe
BioCISE	Biological Collection Information Service in Europe
BiOSCG	Biological Observations, Specimens and Collections Gateway
BM	Herbarium, The Natural History Museum, London
BR	Herbarium, Jardin Botanique National de Belgique
CABI	Commonwealth Agricultural Bureaux International
CANM	Herbarium, Canadian Museum of Nature
CBD	Conservation of Biological Diversity
CBS	Centraalbureau voor Schimmelcultures
CD	Compact Disc
CD-ROM	Compact Disc – Read Only Memory
Cge	Herbarium, University of Cambridge
CGIAR SINGER	The Consultative Group on International Agricultural Research - System-wide Information Network for Genetic Resources
CIMMYT	Centro Internacional de Mejoramiento de Maíz y Trigo
CIP	Centro Internacional de la Papa
CITES	Convention on International Trade of Endangered Species
CMS	The Convention on Migratory Species
COL	Herbarium, Universidad Nacional de Colombia
CONABIO	Comisión Nacional para el Conocimiento y Uso de la Biodiversidad
DEFRA	The Department for Environment, Food and Rural Affairs (UK)

DiGIR	Distributed Generic Information Retrieval
DUKE	Herbarium, Duke University, Durham
DWG	Database Work Group
ECP-GR	European Crop Genetic Resources Programme
EEA	European Environment Agency
EIONET	The European Environmental Information Observation Network
EMBL	European Molecular Biology Laboratory Reptile database
ETC/NPB	European Topic Centre for Nature Protection and Biodiversity
EU	European Union
EUNIS	The European Nature Information System
EURISCO	The European Catalogue Of Crop Genetic Resources Collections
F	Herbarium, Field Museum of Natural History
FAO	The Food and Agriculture Organization
FH	Herbarium, Harvard University
FI	Herbarium, Museo di Storia Naturale del'Universit, Firenze
FW	Herbarium, Texas Christian University
G	Herbarium, Conservatoire et Jardin botaniques de la Ville de Geneve
GBIF	Global Biodiversity Information Facility
GCI	Gray Card Index
GTI	The Global Taxonomy Initiative
GTI	Global Taxonomy Initiative
H	Herbarium, University of Helsinki
HAL	Herbarium, Martin-Luther-Universitat, Halle
HerpNET	Herpetological Database Network
IABIN	Inter-American Biodiversity Information Network
ICARDA	International Center for Agricultural Research in the Dry Areas
ICBN	International Code of Botanical Nomenclature

ICLARM	International Center for Living Aquatic Resources Management
IIBC	The International Institute of Biological Control
IIE	The International Institute of Entomology
IIP	The International Institute of Parasitology
IKew	Index Kewensis
ILDIS	The International Legume Database & Information Service
IMI	International Mycological Institute
IPNI	The International Plant Names Index
ITIS	The Integrated Taxonomic Information System
IUBS	International Union of Biological Sciences
IUCN	The World Conservation Union (formerly International Union for the Conservation of Nature)
JTypes	Type Specimen Database in Japan
MEA	Multilateral Environmental Agreements
MoU	Memorandum of Understanding
NBII	The National Biological Information Infrastructure
NBN	National Biodiversity Network (UK)
NCBI	The National Center for Biotechnology Information (US)
NOAA	National Oceanic & Atmospheric Administration (US)
OBIS	Ocean Biogeographic Information System
OECD	Organisation for Economic Co-operation and Development
SBML	Systems Biology Mark up Language
SGRP	System-wide Genetic Resources Programme
SIS	Species Information Service
SPP	The Species Plantorum Project

SSC	Species Survival Commission
TDWG	Taxonomic Database Working Group
TRED	Taxonomic Resources Expertise Directory
TWG	Taxonomy Work Group
UNEP-WCMC	United Nations Environment Programme World Conservation Monitoring Centre
USDA	United States Department for Agriculture
USDA-GRIN	United States Department for Agriculture Germplasm Resources Information Network
USGS	United States Geological Survey
WWF	The World Wide Fund for Nature

ANNEX 4 – Table of Data providers to Comprehensive Reference sources

Note: Through their partnerships with ITIS and Species 2000, many organisations also provide information indirectly to GBIF

	GBIF	**ITIS**	**Species 2000***	**AllSpecies**
Agriculture and Agri-food Canada		✓		
AmphibiaWeb				✓
Australian Antarctic Data Centre	✓			
Australian Biological Resources Study - Mollusca Pulmonata			✓	
Belgian Biodiversity Information Facility	✓			
Bernice Pauahi Bishop Museum	✓			
Biologiezentrum der Oberoesterreichischen Landesmuseen	✓			
Bird Studies Canada	✓			
Bishop Museum, Hawaii				✓
Burke Museum (UWBM)	✓			
CABI Bioscience UK Centre	✓			✓
California Academy of Sciences (CAS)	✓			
Canadian Biodiversity Information Facility	✓			
Carabidae of the Western Hemisphere				✓
CephBase			✓	
Conabio		✓		
Danish Biodiversity Information Facility	✓			
The Diptera Site, NMNH				✓
Ecological Monitoring and Assessment Network (EMAN)	✓			

	GBIF	**ITIS**	**Species 2000***	**AllSpecies**
The EMBL Reptile Database				✓
FishBase	✓		✓	
Global Biodiversity Information Facility (GBIF)		✓		
GBIF Spain	✓			
GTI Japan	✓			
Hexacorallians Database, Kansas Geological Survey			✓	
Hymenoptera Name Server				✓
ICTVdB (viruses) American Type Culture Collection			✓	
Instituto Nacional de Biodiversidad (Costa Rica)	✓			
ITIS	✓		✓	
James R. Slater Museum of Natural History (PSM)	✓			
Los Angeles County Museum of Natural History (LACM)	✓			
Missouri Botanical Garden	✓			
Museum of Texas Tech University (TTU)	✓			
Museum of Vertebrate Zoology (MVZ)	✓			
NatureServe		✓		
NBII		✓		
NCBI Taxonomy Page				✓
NOAA		✓		
Netherlands Biodiversity Information Facility	✓			
The Orthoptera File				✓

	GBIF	**ITIS**	**Species 2000***	**AllSpecies**
Royal Botanical Gardens, Kew	✓			
Smithsonian Institute		✓		
Species 2000	✓	✓		✓
The Swedish Museum of Natural History (NRM)	✓			
Taiwan Biodiversity Information Facility	✓			
The Tiara Biodiversity Project				
Trichomycetes Database			✓	
United Sates Environmental Protection Agency		✓		
United States National Park service		✓		
Univeristy of Alaska Museum (UAM)	✓			
University of Alberta	✓			
University of Helsinki, Department of Applied Biology	✓			
University of Kansas Museum of Natural History (KU)	✓			
University of New Mexico Museum of Southwestern Biology (MSB)	✓			
University of Turku	✓			
Utah Museum of Natural History (UMNH)	✓			
USDA		✓		
USGS		✓		
WFCC-MIRCEN World Data Centre for Micro-organisms (WDCM)	✓			
The World Spider Catalog				✓
Xylariaceae Database			✓	