## Consejo Superior de Investigaciones Científicas


# System of Information on Biodiversity - SIBIO

# (Spanish node of GBIF)

## Proposal and Action Plan

**CONTENTS**

## 1. Introduction

The GBIF is conceived as an interoperable network of databases on biodiversity and information technology tools that will enable different users to access the information and put it at the service of the economic, environmental and social interests of each country.

Although the GBIF is a worldwide project, it must be implemented at the national a regional levels. Thus the proposal to create national nodes to accomplish the aims of GBIF in each country. For Spain in particular, we propose that it be a kind of "coordination unit," that lends assistance to a "database federation" on biodiversity, which would consist of the centers holding the information, and which would enable access in a unified manner.

## 2. Global Biodiversity Information Facility (GBIF): Objectives

The objective of GBIF is to make available the huge amounts of information on biodiversity that are presently scattered among collections on natural history, herbariums, libraries and databases of diverse types.

The GPIF does not aim at increasing present knowledge of biodiversity, but rather boosting a new kind of synergy between the organizations that possess and generate this information and thereby create conditions that would allow this information to be accessible to all.

One of the main challenges faced by GBIF is the ensure interoperability between databases on species and specimens and those of other types on biological diversity at different levels, from information on genome sequences to data on ecosystems.

To accomplish these objectives, each participating country will create a national node that will coordinate the start-up of a national database network on biodiversity. This node will gather information from the institutions that have collections of interest, develop the computer systems necessary for establishing the network, and will contribute towards training personnel for promoting the use and access to information on biodiversity both inside and outside the country.

To be successful, a necessary priority will be the creation of an electronic catalog of the names of all known institutions, which will enable correlating information from quite diverse sources, which will coalesce into new knowledge on biodiversity.

**Box 1: General aims of GBIF**

Make available, free of charge, all data on worldwide biodiversity by means of:

- Creating a worldwide federation of databases.
- Promoting a national node in each participating country.
- Ensuring interoperability between species and specimen databases and other types of databases on biodiversity.
- Contributing to the training of personnel facilitate access and use of information on biodiversity on the part of all countries regardless of the financial capacities.

## 3. Possible strategies for creation of national node

In our view, the installation of a national GBIF node can be undertaken in one of two ways:

### 3.1. Independent center

The idea is to establish the node as an centralized system in an independent center that would lead, gather information, process it, and make it available to society. To use an example that is familiar, it would be something like the CINDOC system of the CSIC (Information and documentation center).

Advantages of this approach

- Greater capacity of representation.
- For the future, greater likelihood of a consolidated institution with its own personnel and budget.
- Greater autonomy in decision-making and activities.
- Greater operative capacity for participating in international organizations and handling any tasks assigned.

Drawbacks of this approach

- Slow start-up.
- Lack of flexibility in projects that require greater vigor at the beginning than in the future when the biodiversity network is already in place and most work will consist of maintenance.
- Difficult to scale at the beginning, as at present many aspects of Spanish collections on biodiversity are unknown, as are institutions' reactions to the project.

- Necessity to earmark funding and other resources to tasks (administration, training facilities, infrastructure) that would otherwise be covered without extra spending.
- The wariness that projects of this sort might arouse in centers containing collections of information, who may feel themselves relegated to secondary roles.
- The problem of keeping information updated owing to the lack of effective cooperation from information centers for the above-cited reasons.

### 3.2. Coordinating unit

The idea is to create a node that would be a sort of decentralized system with a coordinating unit that would organize the start-up of a federation of databases comprised of the centers that are holders of collections, and would provide the tools, consulting and training needed so that each collection can computerize its own collection and make that data available to the common system. To provide an familiar example as the previous case, it would be similar to the library coordination unit of the CSIC.

Advantages of this approach

- Quick start-up, owing to fact that a great deal of the infrastructure (areas, communications, experience, knowledge) is already in the CSIC and could be channeled towards this project.
- Cost-effectiveness of resources invested. This option enables using new resources where needed without duplication of already-existing infrastructures or services.
- Flexibility for adapting dimensions and resources of unit to different needs over time and the momentum taken on by the project.
- More active participation on the part of information centers as most of the resources invested in the project will directly result in the improvement of their infrastructures owing to computerization and standardization.
- More reliable updating of information as each member center will be responsible for updating as its collection grows.

Drawbacks of this approach

- Unit's capacity to act and margin for maneuver could be undermined by the limitations of the institution housing it or the latter's lack of commitment to the goals of the project. Easy start-up also means easy dismantling.

**Box 2. Possible strategies for accessing information**

Centralized database

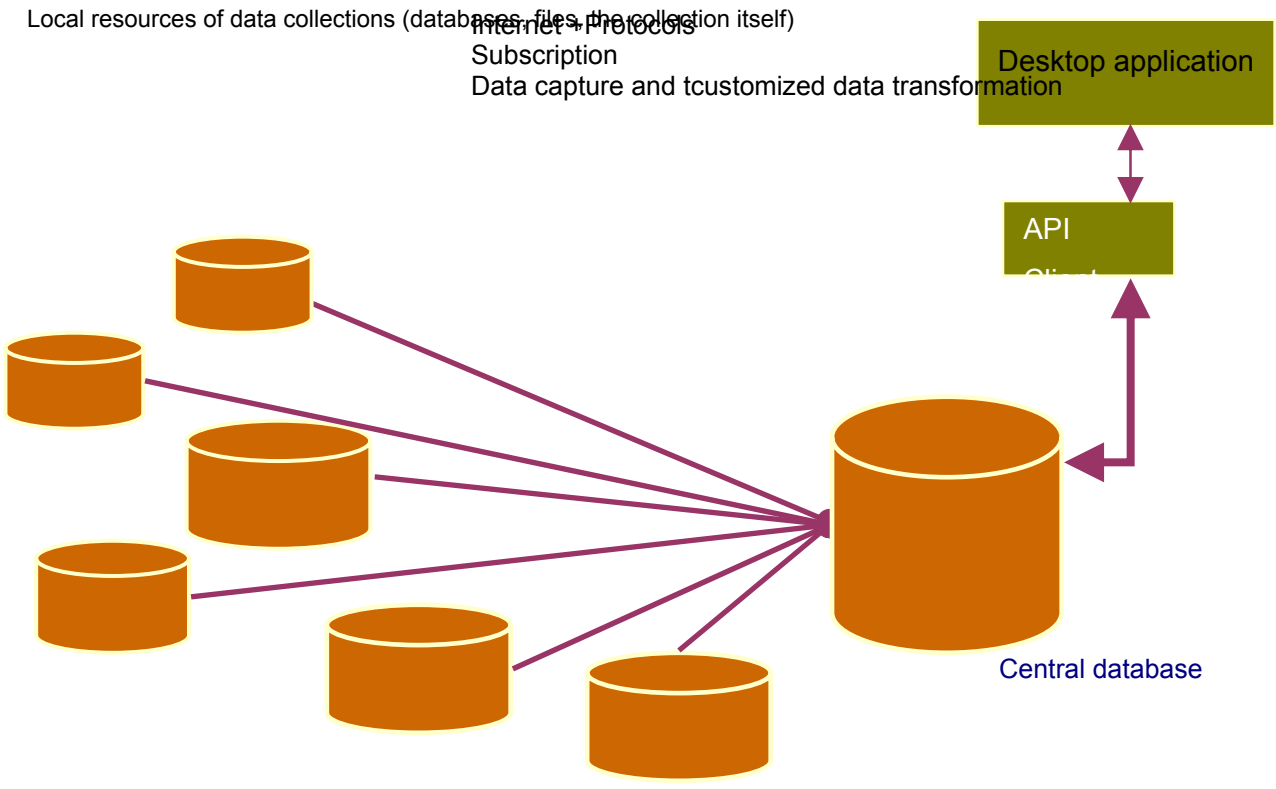
- Easy to implement.
- Requires a major effort to capture data.
- Problems of obsolescence.
- Copyright problems that may hinder cooperation with depositories of information.

Federation of databases

- More difficult to implement
- Data capture requires less work as data is entered and/or maintained by the depositories.
- With a cooperative approach, control over data remains in hands of provider, and therefore, problems of copyrights are minimized.

## Centralized information system

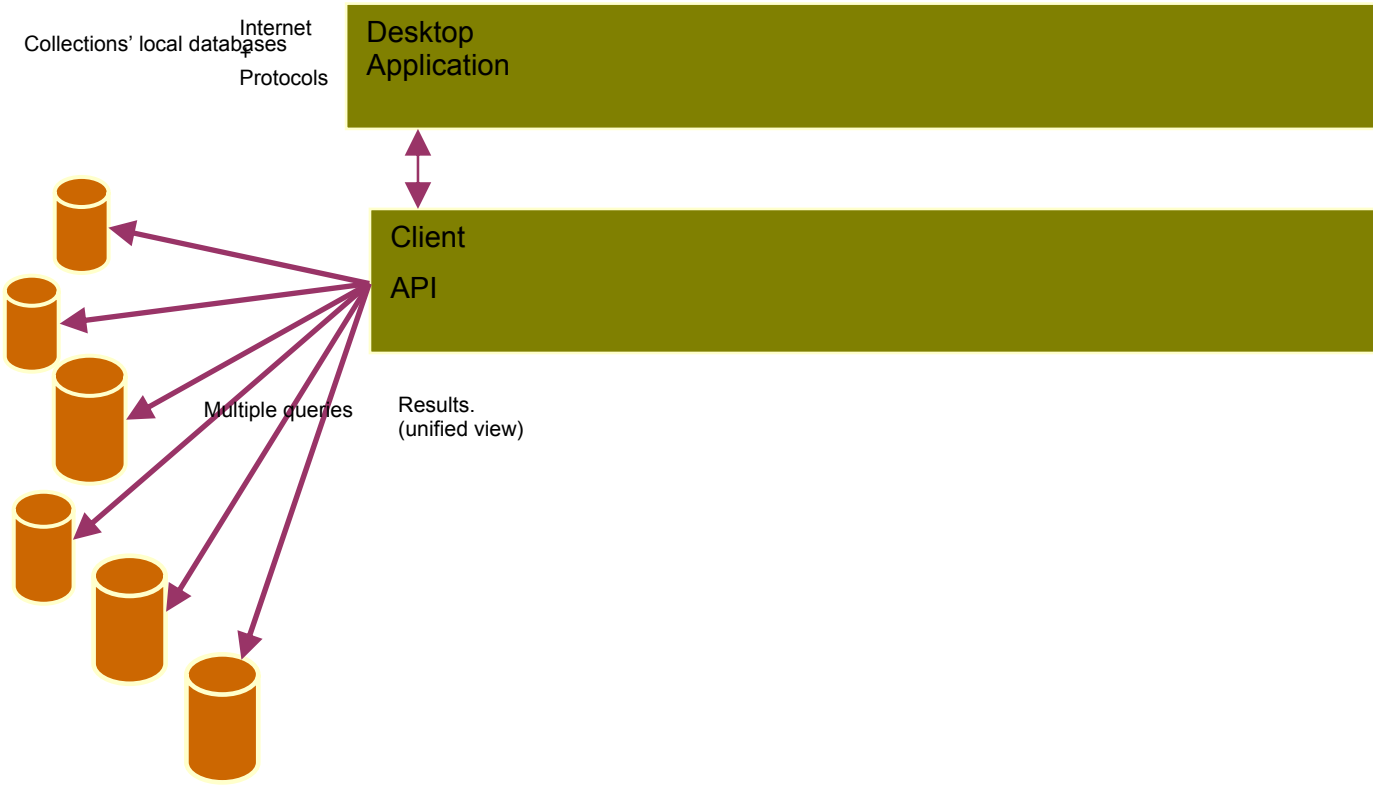
Involves collecting data from multiple sources into a single database to which all users will have access. These usually have a rapid response and are easy to implement, technically. On the other hand, data providers do not have direct access or control over the data, thereby creating hesitation among potential data providers. It is also necessary to implement a central database updating system, otherwise it will quickly become obsolete and thus of less usefulness.



**Distributed information system**

In these systems, the different databases that contribute are under the control and management of the providers, who keep them updated in their own interest. As the data remains under their control, they can restrict access to information they consider to be sensitive (locations of endangered species, new scientific discoveries, etc) Each query is propagated from the access point to the system and all the affiliated databases ("federation of databases") and the results gathered at this point and presented to the users.

Until recently, one of the drawbacks of distributed systems was that the information from different providers was not integrated into a single display screen. Moreover, they required specific software modules for each system database. An example is the current interface of Species 2000 (www.sp2000.org). The implementation of protocols like Z39.50 will integrate data from multiple sources into a single list, in addition to the improved speed, reliability and simplicity of installation.



## 4. Standards and emerging technologies

Implementation of the SIBIO with a distributed information model is more technically complex than with a centralized model, owing to the need to add to a centralized system an entire layer communications between databases and programs through the Internet. This problem could have meant failure, until recently, for a proposal like the one we are making herein; nevertheless, a set of standards and technologies of recent creation – in many cases still under development – presently add up to a different outlook, as they minimize the problems that had been inherent to distributed systems. Below we will briefly describe the most relevant innovations to a “federation of databases” like the one chosen.

ASP (Active Server Pages). This is a technology developed by Microsoft to prepare dynamic web pages by use of programming languages like Vbscript or Jscript. The appeal of this technology is its ease of use in creating Internet resources with access to databases.

ODBC (Open Database Connectivity).This is an applications programming interface (API) for access to databases. It enables creation of programs that interact with diverse types of databases in a general language without having to know the peculiarities of each database manager. It is based on X/Open and ISO/IEC specifications for call interfaces (CLI) for databases, and on SQL as an access language for databases. ODBC is Microsoft technology supported by most database manufacturers (Oracle, Informix, IBM, etc.)

XML (eXtensible Markup Language). This technology allows for putting structured data (spreadsheets, databases, etc.) in a text file. It is a set of rules that allows for creating other rules to enable displaying and manipulating the structured information. One easy use of XML is the creation of dynamic web pages, but its most powerful potential use lies in its ability to display or manage (combined with ASP) one or several databases that may reside in different computers and with different software, so that the complexity of the operation is hidden to the user. XML is, of course, expandable, independent of platform, and supports multiple languages; its technology is recent yet already well consolidated, both by the large software manufacturers and the projects already underway, and it is in the public domain.

Z39.50 (ANSI/NISO Z39.50 = ISO 23950: "Information Retrieval [Z39.50]: Application Service Definition and Protocol Specification). This standard specifies a client/server protocol for searching for and extracting information from remote databases. The initial implementations and its present scope of use is in libraries, where it was enhanced in order to access information in multiple catalogs in a single query.

ZBIG "Dublin Core" (Z39.50 Biology Implementers Group). A work group under the auspices of CIMI (a consortium of museums and other institutions that are responsible for the natural and cultural heritage of the entire world) whose first aim was the definition of a Z39.50 standard profile for specimens in natural history collections. There is an initial version of this profile – known as the Dublin Core – which is already being used in some projects.

## 5. Situation of Spanish Natural History Collections

The natural history collections can be considered scientific infrastructures maintained mainly by museums and botanical gardens, research centers, universities, regional authorities, etc. The main collections include specimens of flora and fauna, but it may also include tissues, micro-organisms, banks of genetic resources, etc. These collections are generally financed by local regional or national institutions.

Owing to their long history, European natural history collections are the most extensive worldwide databases on biodiversity, for many of the specimens are from scientific expeditions to areas of political influence, and there are even specimens of extinct species. In the case of Spain, the collections reflect the immense wealth in biodiversity the country possesses, the highest of any European Union country. Moreover, the significant collections from Latin American countries, the north of Africa and Equatorial Guinea are an excellent source of information on biodiversity in these areas.

The resources and work of each collection also varies widely, as they are more dependent on their research teams and means than on their size of history. Thus it is still unfortunately true that in spite of the country's major investment in creating and maintaining these collections, the information they contain is at times not very accessible, as it is kept in scattered files, archives, notebooks or other storage systems of little use beyond the local level. It is thus important to participate in the project of organizing the information with the ultimate aim of creating an interoperable network of databases, with access through a single portal that enables a simultaneous search for information in all databases.

Table 1. Total number of specimens in Spanish natural history collections.

| | Botanical | Zoological | Totals |
|---|---|---|---|
| CSIC | 1.426.620 | 5.209.365 | 6.635.985 |
| Other | 2.876.757 | 4.094.385 | 6.971.142 |
| **Totals** | **4.303.377** | **9.303.750** | **13.607.127** |

### 5.1. Botanical collections

We are aware of the existence in Spain of 34 botanical collections (herbariums) scattered throughout the country, containing somewhat more than 4 million specimens. In 1993 the Spanish and Portuguese herbariums decided to create an association called “Asociación de Herbarios Ibero-Macaronésicos” [Association of Iber-Macaronesic Herbariums]. The need was soon recognized of computerizing the collections, and thus of making an aim of the association promotion and assistance in these tasks. A standard herbaria program was adopted: a computer application developed in the Royal Botanical Garden (CSIC). The situation as regards computerization of the collections is summarized in table 2 of this report; in addition, some herbariums, like ABH and SANT, have declared their intention to undertake computerization in the near future or when they have the resources to do so. Among the successes of this initiative we would like to point to the Cryptogramia herbarium of the Royal Botanical Garden (CSIC), the only Spanish collection that can be accessed via Internet.

We can therefore say that the Real Jardín Botánico (CSIC) for the moment leads in initiatives for computerizing botanical collections, in the development of specific tools to do so, in the number of specimens computerized, in advising of other herbariums, and placement of collection databases on the Internet.

| INSTITUTION | Code | No. of specimens | Computerized |
|---|---|---|---|
| Universidad de Alicante | ABH | 19500 | |
| Universidad de Alcalá de Henares | AH | 62400 | |
| Universidad de Castilla, La Mancha | ALBA | ------- | |
| Estación Experimental de Zonas Áridas | ALME | 25220 | |
| Sociedad Aranzadi | ARAN | 55900 | |
| Institut Botànic de Barcelona | BC | 1133600 | |
| Universitat Autònoma de Barcelona | BCB | 35100 | |
| Universitat de Barcelona | BCC-BCF* | 260000 | 52000 |
| Universidad del País Vasco/EHU | BIO* | 19500 | 3900 |
| Universidad de Córdoba | COA-COFC | 37700 | |
| Universidad Politécnica de Madrid | EMMA-MAA- | 104000 | 1040 |
| Universidad de Oviedo | FCO | 28600 | |
| Universidad de Granada | GDA-GDAC* | 156000 | 124800 |
| Institut d'Estudis Ilerdencs | HBIL* | 11700 | 2340 |
| Universitat de Girona | HGI | 15600 | |
| Instituto Pirenaico de Ecología | JACA* | 296400 | 296400 |
| Colegio Universitario Santo Reino | JAEN | 32500 | |
| Jardín Botánico Canario Viera y Clavijo | JVC-LPA | 32500 | |
| Universidad de León | LEB | 57200 | |
| C.I.T.A.-Xunta de Galicia | LOU | 39000 | |
| Real Jardín Botánico | MA* | 1105000 | 331500 |
| Universidad Complutense | MACB | 39650 | |
| Universidad Complutense | MAF* | 104650 | 5232 |
| Universidad de Málaga | MGC | 32500 | |
| Universidad de Murcia | MUB* | 52000 | |
| Instituto Canario de Investigaciones | ORT | 42900 | |
| Universidad de Navarra | PAMP | 23127 | |
| Universidad de Salamanca | SALA-SALAF* | 78000 | 15600 |
| Universidad de Santiago de Compostela | SANT | 25350 | |
| Universidad de Sevilla | SEV-SEVF | 191100 | |
| Universidad de La Laguna | TFMC | 8450 | |
| Universidad de Extremadura | UNEX | 13000 | |
| Universitat de València | VAL* | 169130 | 118391 |
| Museo de Ciencias Naturales de Alava | VIT | 78000 | |
| **Totals** | | **4303377** | **951203** |

Table 2. Of the 34 Spanish herbariums, 11 have active programs for computerization (marked with an asterisk), 10 with the Herbar program, of these one is accessible over the Internet: Criptogamia herbarium of the Real jardín Botánico-CSIC. Data: Holmgren, P.K., Holmgren, N.H. & Barnett, L.C. (1990). Index herbariorum. Part I: The herbaria of the world. Eighth edition. *Regnum. Veg.* 120. And personal communication of collection custodians.

All this data is reflected graphically in figure 1.

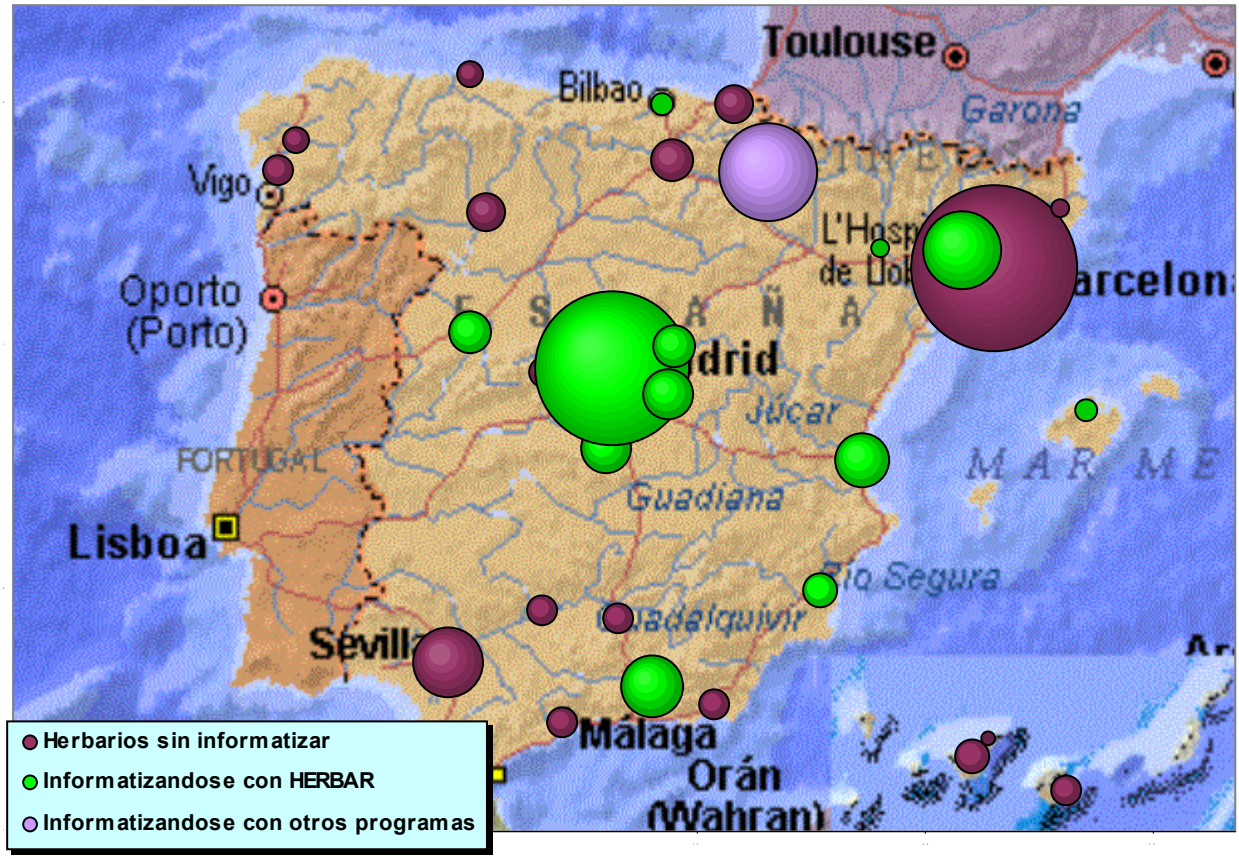



Fig. 1.

### 5.2. Zoological collections

At first sight, the zoological collections in Spain can be divided into three categories: Large collections, like that of the Museo Nacional de Historia Natural [National Museum of Natural History] (CSIC), with nearly 5 million specimens, the Museo de Zoologia of Barcelona and the Sociedad de Ciencias Naturales Aranzadi of San Sebastián. Specialized collections in university collections on animal biology, of which there are a large number all over the country.

Collections of small museums, other institutions and private collections, some of which contain valuable specimens and a long history, like that of the Museo de Ciencias of Málaga or the Instituto de San Isidro of Madrid.

Table 3 summarized the number of collections on which we have data, grouped by region. Unfortunately, complete data is not available for Asturias, La Rioja, Valencia, Castilla-La Mancha, Extremadura, Murcia or Canarias, though they undoubtedly have collections.

Table 3. Zoological collections by region

| | Collections | Specimens [1] | With inventory | Computerized[2] |
|---|---|---|---|---|
| Andalusia | 25 | 300.000 | 20 | ? |
| Aragón | 5 | 10.450 | 4 | 0 |
| Baleares | 11 | 420.000 | ? | ? |
| Castilla-León | 2 | 58.000 | 0 | 0 |
| Catalonia | 3 | 14.500 | 2 | 2 |
| Galicia | 1 | 800 | 1 | 0 |
| Madrid | 60 | 6.500.000 | 29 | 31 |
| Basque Country and Navarra | 274 | 2.000.000 | ? | ? |
| Totals | **381** | **9.303.750** | **56** | **33** |

## 6. Spanish node of GBIF

Our proposal for the Spanish node of the GBIF is the creation of the SIBIO (Spanish Information System on Biodiversity, Sistema Español de Información sobre Biodiversidad) which is conceived as an information network linked by a coordination unit (UC) and a federation of databases made up of the member centers of the network (CAR).

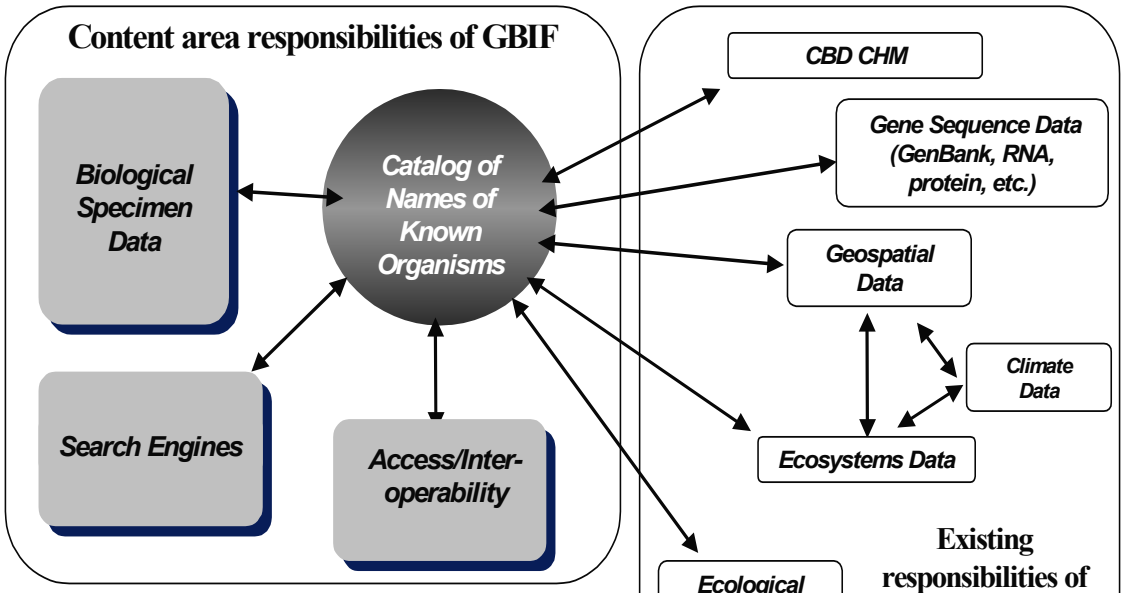
### 6.1. Objectives

The ultimate aim of the SIBIO is to make available, universally and free of charge, all Spanish information on biodiversity, and through GBIF be a part of a worldwide federation of databases on biodiversity.

An aim of the SIBIO, through its coordination unit (**UC**) is to manage and provide support to a federation of databases on biodiversity while enabling access to information in a unified manner. It must therefore

[1]This figure is much higher, if we take into account the fact that some collections do not have an inventory, and in others the figures given are of sets, and the number of specimens in each is unknown.

[2] At least in part.




GBIF would enable synergism among existing investments that is not possible at present.

**Box 3: Catalog of names of known organisms**

There are two main problems in organizing available information on biodiversity: first, the fact that we do not know how many species live on the planet, and second, the fact that knowledge on different groups of organisms is unequal.

Presently 1.8 million species have been described, and the estimations of the number truly existing are so imprecise that they range from 10 million to 100 million. This lack of precision is compounded by a heterogeneity of knowledge, for compared to fairly wee-known species like birds, mammals and vascular plants (of which 10,000, 4,000 and 250,000 are known, respectively, which amounts to 90% of the totals) there are others, which are the most numbers, like fungi or insects, of which we know barely 15% of the species actually in existence.

In the process of describing species, quite frequently the same species has been described several times, or the concept of them has changed, which has lead to the same species being given different names: the accepted one and the synonyms.

It is therefore vital, in the context of the GBIF to create a catalog of the names of known organisms, which would become a indispensable reference point for the organization of all information on biodiversity. This would thus minimize, at least in part, the fictitious growth of the number of species, given that each name represents a species, without taking into account that many are synonyms.

To prepare such a list, expert researchers in different taxonomical groups have to be consulted. A certain advance has been achieved in Europe in this regard, owing to the participation of the MNCN in the European Fauna project (financed by the V Framework Program) whose objective is to create a database with the list of European land fauna and its distribution either to the project EURO+MED (also financed by the V Framework Program) on Mediterranean flora and which has the participation of RJB.

create the coordination unit to reinforce adaptation and creation of databases in different depository centers of biodiversity collections (Network Associated Centers **CAR**) and facilitate common access to the information.

The aim of SIBIO is not to accumulate huge quantities of information in a server but rather to create and facilitate the infrastructure necessary, through the coordination unit

(**UC**) to make the information accessible, from the Network Associated Centers (**CAR**), to society

Thus, the SIBIO, through the coordination unit (UC) will:

1. Design and implement a national biodiversity information system.
2. Establish and support a coherent national strategy for computerizing information on biodiversity. This strategy would, first of all, contribute towards the national aims of knowledge, conservation and rational use of the information, and second, enable the integration of the information in international networks, as intended by GBIF.
3. Provide advising on technical and applied research aspects related to biodiversity to governmental institutions, social and private organizations.
4. To meet the general objectives of the GBIF, the national node will contribute to the training of personnel with a view towards promoting access and use of information on biodiversity on the part of all countries, which in the case of Spain can focus on Latin America.

**Box 4: SIBIO as the Spanish node of GBIF**

Conceived as a federation of databases made up of the centers associated to the network (CAR) and organized by a coordination unit (UC).

It will:

- Design a Spanish system of information on biodiversity..
- Establish and support a coherent national strategy for computerizing data on biodiversity.
- Provide technical support in this regard.

### 6.2 Strategy

The following shall be crucial aspects of strategy:

- Defining a standard data profile for the search and handling of computer records of natural history collections and other databases with information on biodiversity.

- Establishing directives and recommendations for the “proper computerization of biodiversity data and collections” that would be useful as a set of guidelines for initiatives in the field.

- Adopting standards, create tools and provide training to enable people and institutions with collections and information on biodiversity to computerize this data in a manner that is useful for their own purposes and those of potential users, both in Spain and internationally.

- Developing or adopting software for integrating in common interface the data held in different databases affiliated to the unit, databases that constitute a federation of databases.

- Developing an interface for browsing and analyzing the data accessible to the federation. This interface or access point need not be a single central one, assigned to a “central node”: all the nodes that make up the federation of databases can function as access points; thus, the federation is comprised of symmetrical nodes.

- Presenting data from the databases associated to the Unit to potential users (the government, researchers, environmental managers, non-governmental organizations, society at large), be they national or international, in a unified, useful manner..

#### **6.2.1.** *Recommendations linked to strategy*

The above actions must be consistent with project funding policy that must providesthe economic means to initiatives which goals coincide with those of the Spanish node of the GBIF.

Any design in this regard must be scalable. First, it must enable homogeneous access to information compiled on the basis of quite diverse realities, and second, it must be ready to incorporate new data in the future, whether from the addition of new taxonomical groups, or from the need to include new types of information.

In accordance with the general philosophy and strategy of the GBIF, the unit will promote the free use and circulation of data affiliated to the federation of databases. Nevertheless, the data will continue to be property of the providers (CAR) and this ownership will be recognized in the credits of use.

There is no point in undertaking the computerization of collections unless they are well-conserved and accessible to study. Thus, prior to computerization, preparation of those that require it is recommended.

**6.2.2.** *Projects that work under this proposal, and from which lessons can be drawn*

ABIF. Australian Biodiversity Information Facility. An initiative of the Australian Biological Resources Study. It seeks to place reliable and complete information on Australian biodiversity at the disposal of users. ABIF expressly recognizes the participation of all contributions. The Facility is made up of checklists and censuses of different groups of organisms, all electronically accessible. The basis of each module is a list of valid names for that group of organisms in Australia. In the next phase, ABIF intends to add complementary information, such as: bibliographical data, dates and places of publication, nomenclature, chronological and ecological information, descriptions, keys to identification, illustrations, biological data and links to other sites.

CONABIO. La Comisión Nacional para el Conocimiento y Uso de la Biodiversidad [National Commissionn for the Knowledge and Use of Biodiversity] (Mexico). A state agency, which is interministerial, and that funds projects and publications, develops computer tools brings together databases, all aimed at the sustainable knowledge and use of biodiversity. One of its projects, REMIB (Mexican Network of Information on Biodiversity) is very similar in aims and approach to what a Spanish GBIF node would be.

INBIO. Instituto Nacional de Biodiversidad [National Biodiversity Institute] (Costa Rica). A scientific association with a social orientation, public interest and non-profit, whose mission is to promote greater consciousness regarding the value of biodiversity, to achieve its conservation and improve the quality of life of human beings, to introduce and deepen knowledge of national biodiversity and discover intelligent, non-harmful forms of using it in a sustainable manner. These aims, which surpass the aims of our proposal, are to be achieved by means of combining the following processes: making an inventory of Costa Rican biodiversity, conduct research into the new species and their potential uses, promote conservation for development and social extension.

IPNI. International Plant Name Index. A collaborative project between the Royal Botanical Gardens, Kew, the Harvard University Herbaria, and the Australian National Herbarium. It is a database of the names and secondary names of all plant species. Its aim is to provide basic bibliographical information on the names of plants. It will be a dynamic resource, which will depend on the direct contributions of all the members of the botanical community, and of free access.

OBIS Ocean Biogeographic Information System. An initiative of the Institute of Marine and Coastal Sciences, of Rutgers, State University of New Jersey. It is a database of the global marine distributions both of animals and plants, enabling the correlation of geographic distribution with environmental factors. The ultimate aim is to design a tool

that would allow for describing regional and global models of distribution of biodiversity in oceans, and thereby lay the true bases for research into global biodiversity.

The Species Analyst. A project coordinated by the Natural History Museum of the University of Kansas. It develops standards and computer tools in order to access in a unified manner databases of collections and observations on biodiversity.

### 6.3. Work to be undertaken by the Coordination Unit (UC)

The Spanish node of the GBIF will work to accomplish its objectives through the "coordination unit," which will engage in the following activities:

1. In light of the unequal status of collections (see tables 2 and 3), and the frequent lack of information on them, the UC will gather from the institutions that house them all necessary data to ascertain their starting point concerning preparation and computerization. Of use here will be the experience gained in projects already underway in Europe such as the European Natural History Specimen Information Network (ENHSIN) or of professional associations like that of the Iber-Macaronesian herbaria (AHIM).
2. In an initiative of this sort, with participation by institutions of diverse types and where the data to be included is of varied origin and use, planning must allow certain margin for handling heterogeneous information. Yet it is likewise indispensable to ensure the unified functioning of the system, thus the need for a set of standards. Owing to the efforts of international working groups such as CODATA (Committee on Data for Science and Technology) or TDWG (IUBS Commission for Taxonomic Databases) a large part of the standards already exist; it is necessary only to create those standards that may be needed to accommodate the special conditions of Spanish collections. Similarly, to get the project going, decisions must be made concerning the technology to be used. These decisions – priorities, criteria, standards and technology – are prior to any other work.
3. The software needed for achieving these aims can divided into the following:

   Reference application for computerizing the collections. The selection or design of this application need not be single, exclusive, as the casuistry of all the collections cannot be accommodated in a single application. Nevertheless, the UC will design a documented reference application that will be free of charge, and that will enable computerization of collections in a manner that is reasonable and useful for the interests of the collections themselves and the node.

Generic software for computers of the coordination unit and affiliated collections: operating system, database manager, etc. (generic commercial software: Windows 2000, IIIS, ODBC drivers, suites)
Specific development (ASP development, ODBC connections, XML wrappers, etc.) for placing the databases on the Internet.
Specific development for creating a single access point to databases.
Development tools of IT applications (ASP pages, ZX client)

4. To be suitably used, any computer development requires appropriate documentation and a training program. Attending to both aspects is vital to ensuring that the application can be expanded and enhanced even when the team that created it is no longer available. It also enables potential users to employ it free of traumas and with the utmost usefulness possible from the start. Moreover, good documentation and a training program are always an investment, for they free the development team from a considerable amount of queries and requests.
5. To ensure the coherence of the project funding policy that must provide the economic means for the initiatives of the collections linked to the node, it seems vital for the UC to provide technical advisory to the Experts Committee to judge the proposals; advisory on the suitability, viability and budget of the proposals.
6. It is obvious that the UC will have responsibility for follow-up and quality control in those projects under its control.
7. In order to accomplish its ultimate goal of making available the information of node-associated databases to all users, a web server will have to be created and maintained. Such a server would not only be an access point to the information but also a source of resources where the participating collections can find programs, documentation, help files, links, etc., of aid in their work. It is also possible for the servers of the participating centers function as access points to the information. This possibility, which is more viable technically, blurs the traditional scheme of a central node plus satellite nodes in favor of a “network of equals” model.
8. It should not be forgotten that the Spanish node will be just another node in the national GBIF network and that the information it provides will have to be integrated into other supra-governmental and global networks. That is why the UC will seek to achieve interoperability with initiatives of a broader geographic scope (UE, GBIF, Euro+Med, European Fauna, Species 2000, etc.) and will coordinate with the national focal points that channel initiatives of the Biological Diversity Convention in every country.

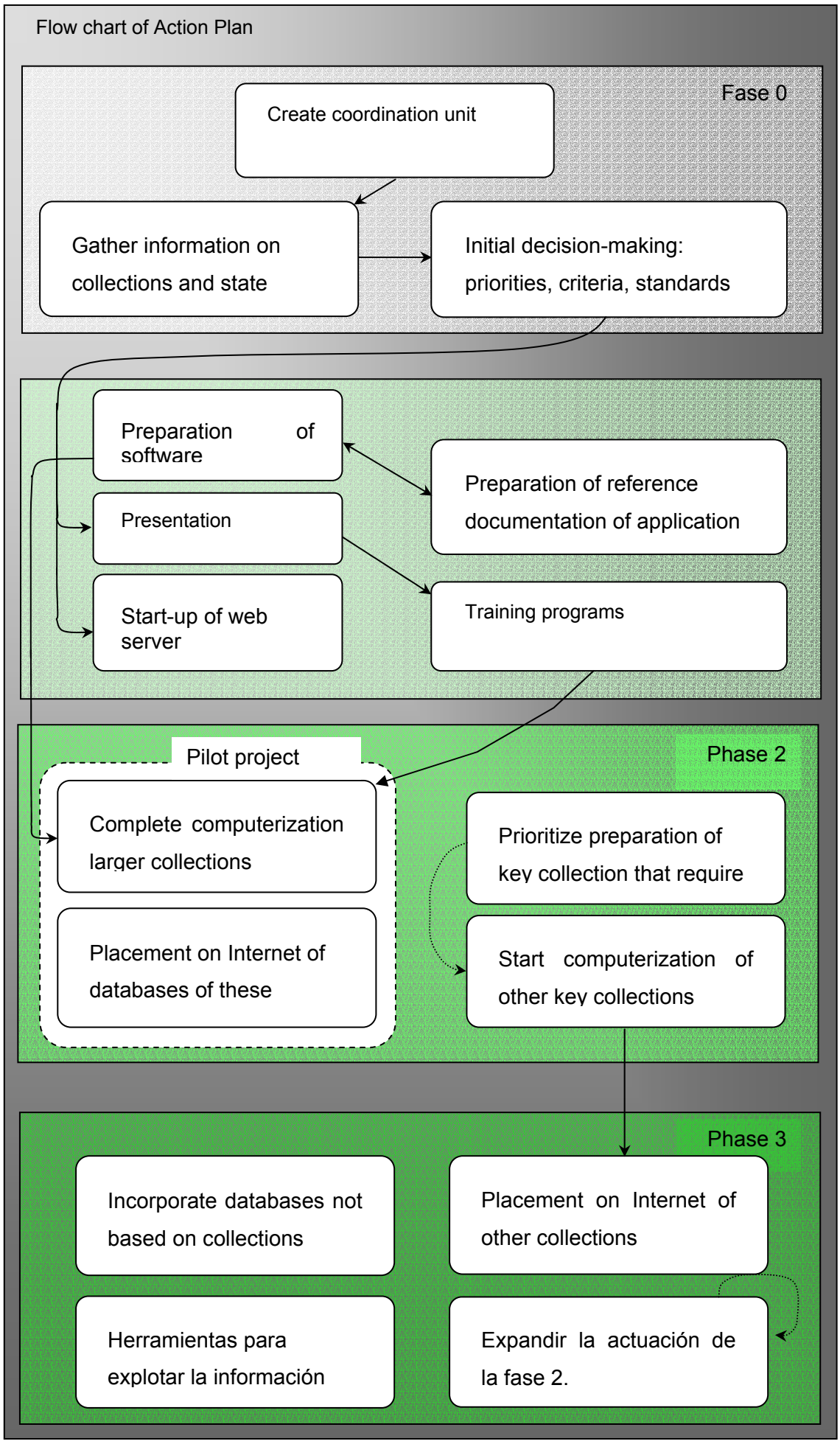
9. The resources and work of each center vary greatly, and are depend more on its research teams and means that its size or history. This why the UC must provide them with technical support and specific information technology advisory
10. Establishing access to all the information on biodiversity in the Spanish collections is only the first step. The ultimate goal of this entire project must be to ascertain and comprehend the biodiversity of the Earth as a whole. To move towards this goal, there must be development of tools of analysis and links to other databases to enable understanding of the distribution of biodiversity, creation of predictive models to help minimize the negative impact of climate change or of invading species; in short, the make possible more harmonious, sustainable development.
11. It cannot be forgotten that the richest countries in biodiversity are the tropical ones, and that by and large, they are also those with the least economic resources to implement projects such as this. GBIF aims at assisting them in developing their own capacity to study and catalog their biodiversity. This can be achieved through the national nodes, by placing at their disposal the necessary tools for developing such projects and contributing to the training of specialized personnel. As for Spain, it would seem logical for the national node to focus primarily on assisting and training nodes in Latin America.

### 6.4. Work to be undertaken by centers associated to the network (CAR)

1. Analyze the state of collections and provide this information to the coordination unit (UC) so that it may prepare a plan of action.
2. Prepare computerization projects in accordance with the directives of the SIBIO, in order to apply for funding from the relevant sources.
3. Prepare collections to be consulted and used.
4. Computerize collections.
5. Update information as the collection incorporates new specimens

**Box 5: Work of Coordination Unit (UC)**

1. Compilation of information and evaluation of the state of natural history collections with a view towards possible measures.
2. Establishment of criteria for computerization (standard and technology, types of data)
3. Design and/or adoption of software both for providing unified access to data and computerizing each collection.
4. Development of manuals that will enable use of tools to computerize collections and provision of necessary training.
5. Advisory and evaluation of proposals for computerization projects.
6. Follow-up and quality control of projects in its domain.
7. Creation and maintenance of website that will coordinate work and present information from the associated databases to users.
8. Achievement of interoperability of node with broader geographic or supra-governmental initiatives (UE, GBIF, Euro+Med, European Fauna, Species 2000, etc.) and will coordinate with national focal points that channel the initiatives of the Biological Diversity Convention in each country.
9. Provision of technical support and advisory to collections participating in the project.
10. Research into the forms of giving value-added to the data through development of tools of analysis and links to other databases with a view towards creating integrated information systems: ecology, climate, geography, customs, etc..
11. Collaboration with Latin American countries in the implementation of national nodes.


Flow chart of Action Plan
Fase 0
Create coordination unit
Gather information on collections and state
Initial decision-making: priorities, criteria, standards
Preparation of software
Presentation
Start-up of web server
Preparation of reference documentation of application
Training programs
Pilot project
Phase 2
Complete computerization larger collections
Placement on Internet of databases of these
Prioritize preparation of key collection that require
Start computerization of other key collections
Phase 3
Incorporate databases not based on collections
Herramientas para explotar la información
Placement on Internet of other collections
Expandir la actuación de la fase 2.

### 6.5. Action plan

We have divided the Action Plan into four phases:

*Phase 0*

1. Creation of coordination unit (UC)
2. Analysis of the state of collections
3. Drafting of general report on collections and their state
4. Initial decision-making (priorities, criteria, standards and technology).

*Phase 1*

1. Preparation of necessary software: algorithms for searching in multiple databases simultaneously, database servers and clients, links to external databases, user interfaces, etc.
2. Preparation of reference documentation for application (for the application itself and for the users' reference).
3. Presentation (informational sessions, workshops, Internet, congresses, forums, etc.)
4. Training programs (workshops, seminars, users' meetings).
5. Start-up of web server.
6. Preparation and application for projects with a view towards securing funding on the part of the centers associated to the network (CAR)

*Phase 2*

1. Acceleration and completion of computerization of the large collections with consolidated programs of computerization.
2. Access via Internet of databases of a pilot group of collections according to the database federation model.
3. Adjustments to the operating model in parallel with the start-up of the pilot group.
4. Involvement in computerization tasks of collections that are key either for their size or the nature of their contents.
5. Training programs (workshops, seminars, users' meetings).
6. For those collections that require it, preparation for access and study. In some cases, there will be specific improvement plans, while in others incentives for transfer of collections to be deposited at consolidated collections

*Phase 3*

1. Access via Internet of databases of collections that are key either for their size or the nature of their contents (phase 2. 3).
2. Expansion of scope, preparation and computerization of collection, placement on Internet of more collections, and more training courses.
3. Incorporation to node of databases on biodiversity not based on collections. (observations, literature).
4. Elaboration of tools of analysis and integration between databases with different types of information (on specimens, climates, genomes, GIS systems, etc. to create models that would enable us to comprehend biodiversity in a global perspective.

### 6.6. Resources needed for implementation

#### 6.6.1. *Coordination Unit*

Personnel

Director.
Webmaster.
Specialist in communication and databases (information technology).
Project coordinator
Coordinator of publications, courses, workshops and literature
Secretary

Hardware

6 workstations
2 servers (one for the unit and another for the website)
Local network (cabling, concentrators, router, etc.)
Peripherals (printers, scanner, CD-ROM stampers, etc.)

- Generic software: For operating system, database, communications, development, etc., plus use licenses.
- Communication system: High-speed Internet connection, fax, phones, etc.
- Organization of courses, seminars and presentations: For this item, the computer room at the CSIC could be used.
- Publications: Editing and publication costs for manuals developed by unit and of publications of interest to the project.
- Physical space and furniture: Office furniture
- Literature: Acquisition of books, subscriptions to magazines, journals.
- Non-inventory items: Fungibles
- Travel and per diems: Congresses, meetings, visits to collections, travel for the purposes of technical support and supervision of projects.

**6.6.2.** *Adjustment of software and manuals*

Acquisition, contracting and preparation of specific software (computerization applications, software for simultaneous access to multiple databases)
Design and layout of pages in website.

**6.6.3.** *Centers associated to network (CAR)*

In computerization of collections, there are three factors at work: the size of the collection, time and the state of the collection.

Personnel

A computerization team is made up of two types of professionals: the data recorder and the supervisor.
The data recorder enters the information in the base, corrections, makes backup copies and produces listings, and the supervisor sets priorities, controls the quality of the process, and settles doubts.
The proportion between these two types of professional per team will vary according to the state and type of collection; the number of teams will be conditioned by the time factor.

- Hardware: A workstation for a data recorder
- Generic and specific software: It is recommended that the generic software and its licenses be the responsibility of the depository center of collections.
- Non-inventory material: Fungibles
- Travel and per diems: To attend courses, seminars and workshops.

As for the preparation of the collections, it is quite difficult to calculate the resources needed, as the collections that require it are also those for which the least information is available, thus requiring data gathering (see point 1 of the activities section); nonetheless, we can identify three types:

- Personnel: Needed for setup, labeling and ordering of collections.
- Inventory material: Furniture needed for the proper placement of the already ordered collections.
- Fungibles: Boxes, labels, paper, needles, etc.

**6.6.4.** *International Cooperation*

Mainly training courses, plus travel and per diem expenses.

## 6.7 Budget

### **6.7.1.** *Coordination Unit (UC)*

Personnel

| Position | Annual total |
|---|---|
| Director. | 6.500.000 |
| Web server administrator | 6.200.000 |
| Specialist in communications and databases (IT specialist). | 6.200.000 |
| Project coordinator | 5.000.000 |
| Coordinator of publications, courses, workshops and literature | 5.000.000 |
| Secretary | 2.750.000 |
| **Annual total** | **31.650.000** |

Hardware

| Hardware | Cost |
|---|---|
| 6 workstations (350.000 x 6) | 2.100.000 |
| 2 servers: one for the Unit and another for the website (500.000 +1.200.000) | 1.700.000 |
| Local network (cabling, concentrators, router, etc.) | 1.500.000 |
| Peripherals (printers, scanner, CD ROM stampers, SAI, etc.) | 1.000.000 |
| **Total** | **6.300.000** |

Generic software

| | | |
|---|---|---|
| Operating system software, databases, development communications, and use licenses. | **Total** | **25.000.000** |

Communications system

| | |
|---|---|
| High speed Internet connection, fax, phones, etc. | |
| **Total** | **1.500.000** |

Organization of courses, seminars and presentations

| | |
|---|---|
| Organization, teaching materials, scholarships, guest professors, etc. | |
| **Annual total** | **1.500.000** |

Publishing

| | |
|---|---|
| Costs of editing and publishing of manuals developed by the unit and other publications of interest for the project. **Annual total** | **1.500.000** |

Physical space and furniture

| | | |
|---|---|---|
| Office furniture | **Total** | **3.000.000** |

Literature

| | | |
|---|---|---|
| Acquisition of books, subscriptions to magazines, journals, etc. | | |
| | **Annual total** | **800.000** |

Non-inventory material

| | | |
|---|---|---|
| Fungibles | **Annual total** | **1.000.000** |

Travel and expenses

| | |
|---|---|
| Congresses, meetings, visits to collections, trips for technical support and supervision of projects. **Annual total** | **4. 000.000** |

**6.7.2.** *Preparation of software and manuals*

| | | |
|---|---|---|
| Acquisition, contracting and adjustment of specific software (computerization applications, software for simultaneous access to multiple databases) | | |
| Design and layout of pages in website | **Total** | **30.000.000** |

**6.7.3.** *Centers associated to the network (CAR)*

The most logical manner of funding this part of the plan is through projects on the basis of an application by the institutions holding collections that participate in this proposal. These projects could enter under a specific program of the Spanish Ministry of Science and Technology, and following a positive assessment, gain funding. It is estimated that the average amount per specimen, for preparation and computerization, is 160 pesetas.

Many of the institutions holding collections can indirectly co-fund the work as they can contribute part of the personnel needed (curators, specialists, etc.) paid by the institutions to which they belong.

The approximate total, the total, maximum investment is calculated as follows:

Total specimens in Spanish collections: 14.000.000

Computerized: 30% = 4.200.000

Not computerized: 70% = 9.800.000

Average cost of computerization and preparation of each specimen: 160 pesetas.

Total cost: 9.800.000 x 160 = 1.568.000.000 pesetas.

***6.7.4.*** International cooperation

| | | |
|---|---|---|
| Scholarships for participation of Latin Americans in training courses (9 technicians year x 300.000 pesetas scholarship) | **Annual total** | **2.700.000** |

**Box 6. Summary of required budget**

| | **Total / year** | **One-time investment** |
|---|---|---|
| *Coordination unit* | 38.456.000 | 38.800.000 |
| *Specific software* | | 30.000.000 |
| *Computerization of collections*[3] | | |
| *International cooperation* | 2.700.000 | |
| **Total** | **43.150.000** | **68.800.000** |

[3] The part of the budget for computerization and preparation of the collections must be funded, as already noted, through application-based projects presented by the institutions holding collections that are participating in the proposal, and after the evaluation they receive the funds from the Spanish Ministry of Science and Technology through one of its programs or through calls of regional, national or European funding programs. It is estimated that the average amount per specimen for preparation and computerization is 160 pesetas.

**Box 7: Copyrights**

The establishment of GBIF may generate sensitive issues concerning copyrights, including legal protection of copyrights in the databases or the commercial use of information accessible through the GBIF. Since the intention of GBIF is to promote greater comprehension of the benefits of international cooperation in the field of information on biodiversity, it would seem to be appropriate to adopt a an international code of good conduct in accordance with the standards contained in the Convention on Biological Diversity. GBIF should also address the issue of sharing the benefits related to copyrights in biodiversity, though it must be emphasized that GBIF plays no role in the management of biological resources, but only in the management of information on biodiversity.

A set of basic principles has been approved on the issue of copyrights, and which are contained in the MOU (Memorandum of Understanding):

- GBIF should promote free access to information on biodiversity and should not claim any copyright over the databases developed by organizations that affiliate themselves to the GBIF, and should place in the public domain the data commissioned, created or developed directly by GBIF, in complete respect for the conditions set by the information providers who affiliate their databases to the GBIF.
- GBIF must ensure that the source of the information is expressly recognized.
- GBIF must not assume responsibility for the maintenance and updating of the affiliated databases, ensuring that control over databases remains in the hands of the provider.
- Owners of the databases have the right to block access to the data.
- GBIF shall have no responsibility whatsoever regarding the quality or reliability of data, or its suitability for particular purposes.
- When the obtaining of new data has involved access to biological resources, GBIF should ask for proof that said access was lawful and in accordance with prevailing regulations.
- GBIF may apply for copyrights on tools like search systems or software that are developed by GBIF.
- GBIF should promote the non-exclusive transfer of information technologies developed in this context to institutions in developing countries, together with training and skill-building programs.

# Appendix 1: Analysis of operational risks

**Potential difficulties in establishing coordination unit**

Lack of political will
Delays in decision-making
Personnel selection problems
Insufficient funding

**Potential difficulties for startup**

Delays in releasing funds
Delays in installation of unit

**Potential difficulties of strategy**

Insufficient basis for initial decision-making

Contingency plan

Contact experts of other national nodes, in collections, distributed systems, in standards.

**Potential difficulties in scaling work**

Insufficient base of knowledge on state of collections for initial decision-making.

Contingency plan

Prepare a good initial report.
Create a sufficiently agile system that enables rectification of errors without high costs or much complication

**Potential difficulties resulting from the technology chosen**

Insufficient basis for the correct selection of software
Lack of foresight of new casuistry that may arise regarding new types of collections.

Contingency plan

Contact experts from other nodes and follow the example or previous experiences.

Design a system that is scalable

**Potential difficulties in training programs**

Lack of economic resources
Lack of interest in centers associated to network

Contingency plan

Scholarship program
Attractive proposals based on work saving that could result from knowing software.

**Potential difficulties related to funding of computerization programs**

Lack of funding
Inconsistent funding
Interruption of work when already underway

Contingency plan

Set up systems of projects that, under control, ensure funding for the entire proposal.

**Potential difficulties in maintaining project.**

That after computerization of the collections, the centers holding them give up updating them.

Contingency plan

Courses, workshops or meetings to inform on new technologies and the benefits they bring.

**Potential difficulties in the implementation of computerization programs**

Lack of interest on the part of collection-holding centers in participating in proposal

Contingency plan

Attractive proposals- that offer a direct benefit to the collection – and sufficient funding.
Establish a flexible system of project funding.

**Potential difficulties resulting from the state of conservation of collections**

The state of collections- though the collections may be valuable – is so deficient, that they are not worth computerization.

Contingency plan

Establish the means for preparing collections.

**Potential difficulties resulting from lack of cost-effectiveness**

Work is not cost-effective. Either because the result is incomplete or not very reliable; access is slow or inconvenient; the information unsuitable or not the best available.

Contingency plan

Quality control

Quick implementation

Follow example of prior experiences.

# Appendix 2: Expected benefits

### A good scientific infrastructure

It will enable Spain to have one of the best scientific infrastructures in the world for the study of biodiversity (after the US, Mexico and Great Britain) which would be appropriate given the qualitative and quantitative biological diversity of Spain.

### Unified, global access

In biodiversity studies, where information on each species is scattered among thousands – of not millions – of specimens, the best collection cannot compete with the sum of all, so the quality and reliability of information brought by such access will have an immediate impact on:

- Accuracy of knowledge and characterization of each of the facets of biodiversity.
- The quality of decisions on biodiversity (environmental impact, conservation priorities, penetration of invading species, etc.)
- Boosting quality and efficiency of research in biodiversity and related fields.
- Global knowledge of previous work will allow for reducing costs of research due to existence of reliable database. It should be born in mind that at present it is often preferable to carry out field work starting from zero than to gather information scattered in a large number of collections

### Easy access to updated knowledge

This will allow for setting goals and priorities in scientific projects with a greater basis of knowledge and greater clarity, making research more cost-effective.

It will open up new fields of study that have so far been unexplored: models that explain the distribution of species, predictive models of climate change, biome-climate interactions, etc.

It will “democratize” and strengthen the study of biodiversity at a worldwide level by facilitating access to information to undertake work, by means of a database network. It is too often the case that certain fields of study are the private domain of research teams that have access to greater information.