IABIN Architecture and Interoperability
Version 1.1
By Boris Ramirez, IABIN Secretariat

1. Introduction

The Inter-American Biodiversity Information Network (IABIN) was conceived as a forum to foster technical collaboration and coordination among the countries of the Americas, to collect, share and use biodiversity information relevant to the decision making processes for biodiversity conservation and management, as well as for education, in the Region.

Under this objective the following goals were established:

- Build an infrastructure for the exchange of digital biodiversity information
- Strengthen the technical capacity to exchange information among the countries in the Americas without regard to political, linguistic and institutional frontiers
- Provide access to biodiversity information that would be helpful to decision makers in order to improve biodiversity conservation efforts
- Improve the capacity to store, use and distribute timely and scientifically sound biodiversity information
- Produce or adapt environmental decision making tools that would foster sustainable development in the Region.

Since 2005 the IABIN Secretariat, jointly with the Thematic Networks Coordinating Institutions, has led the process to achieve these goals and the objectives established.

This document presents the current state of implementation of each of the IABIN Thematic Networks, how these networks will be initially integrated, as well as the integration of the information of the data providers connected to IABIN and to other initiatives at the national, regional and international levels.

Several organizations which have been working in this area have collaborated in the whole process and IABIN has established agreements with them in order to adapt their technologies to the needs of IABIN and promote their standards and protocols

2. General Concepts

This chapter defines some of the general concepts and the terminology used in the implementation of the Network as well as the terminology used throughout this document.

It is important to clarify that IABIN is a decentralized data network in which the data providers are the key component for the network's existence. Without data providers IABIN could not exist.

a. **Basic concepts for the exchange of data**

In order to achieve IABIN's general objectives it is necessary to make available the biodiversity data that each organization possesses. These organizations will become data providers for IABIN the moment they share their data with the other Network members following the principles established by IABIN to this end (see annex 1: IABIN guiding principles).

To a great extent, the success of IABIN depends on its capacity and that of its partners (Focal Points and Coordinating Institutions) to connect data providers with high quality data, and that these data could be consulted in an efficient manner in real time with each data provider maintaining control over his data as well as the responsibility for maintaining them.

The data provider must have the freedom to decide which data he wants to share, in what format, and with which network wants to share that data.

In order to give this freedom to the data providers, however, IABIN had to resolve the following issues:

- How to integrate databases that contain biodiversity data which are in different organizations and in different formats?
- How to make available in real time the information of a data provider?
- How to know where the data are found?
- How to make available information which is not in digital form?

- *How to integrate databases that contain biodiversity data which are in different organizations and in different formats?* In order to answer this question IABIN identified the need to create/adopt standards for the exchange of information. In other words, the need to find a way for all the data providers to speak the same language.

To this end IABIN proceeded to identify or create standards that could be used, or which were already being used, to exchange data in each one of IABIN Thematic Networks, and the need to develop a strategy to promote their use and implementation. The result of this phase is found in Chapter 4 "Standards and Protocols" of this document.

- *How to make available in real time the information of a data provider?* In order to answer this question, an analysis of the existing technologies and how other networks were using them was conducted. Each one of the options found was evaluated. At the end of the process it was concluded that the final product must have the following characteristics:

- It should be easy to install and maintain. This means that it should allow the connection of the data providers' data in a transparent way and without the provider having to incur in any additional expenses.
- It should ensure the security of the data, and should not put in jeopardy the integrity of the data provider's internal information networks.

- It should provide the necessary information to be able to identify the provider and his intellectual property rights on the data.
- It should provide a unique solution to the whole Network. This means that the solution should be capable of serving data from Species, Specimens, Invasive Species, Pollinators, Protected Areas and Ecosystems.
- It should use the Internet as the primary channel of communication and data exchange. This Internet solution presents the following challenges:
    - Different speeds in the connections between data providers. In theory, under this scheme it is contemplated that the Network will function to the speed of the data provider with the slowest connection.
    - Connection availability. This means that the data provider should be available every time there is a query, since if the data provider is not on line it would not be possible to consult and retrieve his data when the user requests them.
    - Not all data providers would have Internet connection.
- It should allow other networks (if the data provider so desires) at the national, regional or global level to have access to his data without having to implement parallel solutions to this end and in order to avoid duplicate efforts.

As a result of the above, the need to create/adapt a "connector" was identified. This connector would have to be installed between the data provider and the Network, and it would have all the characteristics discussed above.

It was also identified that there are two kinds of data that need to be integrated and connected: The first kind are the primary biodiversity data (descriptions of specimens, species and observations) and the second kind are the cartographic data (country maps, species distribution maps, ecosystem maps, etc.).

For the first kind, the IABIN Technical Working Group decided that the best option would be to take the work led by the Biodiversity Information Standards Group (formerly known as TDWG, Taxonomic Database Working Group) and work jointly with them and GBIF in the implementation of an integral solution to access and present data called TAPIR (TDWG Access Protocol for Information Retrieval) which can be found in http://www.tdwg.org/activities/tapir/, for all the data providers of the different networks. TAPIR is the next generation of a project known as DiGIR, which was developed with the objective of sharing biological data under the Darwin Core Standard.

For the second kind, the IABIN Technical Working Group decided to continue with the strategy started by the IABIN Geospatial Network, which allows for the exchange of geographic information using standards such as WMS and WFS. It is expected that eventually, TAPIR implementations will be able to server WMS and WFS protocols.

- *How to know where the data is found?* Once all the previous challenges had been solved, a mechanism had to be created that would allow searching the Network and retrieving data of interest. Similarly, the question came up of what should be the characteristics of the Network's search engine and how it would differ from the popular search engines such as Google in order to justify the investment. As a result, a search tool called IABIN BioBot will be created (see Annex 2 *Differences between IABIN BioBot and Google*), along with a series of tools and strategies framed within the IABIN Catalog. These tools and strategies will allow searching the Network. It is worthy to note that the creation of metadata is paramount to achieving this.

This solution should also allow for the creation of an index to facilitate the retrieval of the data made available by each data provider. In other words, it should allow determining exactly what data there are in the Network, which providers own them and how to reach and retrieve them.

- *How to make available data which are not in digital form?* The Network also found a great amount of data which are not in digital format or which are in electronic documents that do not allow for an automatic data exchange. For this reason the Network should provide those data providers who want to digitize their data with the tools to do so. Under this scheme each IABIN Thematic Network should adopt or create an open source data digitizing tool.

b. IABIN basic architecture

As a result of the questions originally posed, the basic scheme to be implemented by each one of the IABIN Thematic Network (as would be allowed by the available technology, the idiosyncrasies of the thematic network and the characteristics of its data) was developed. This scheme is shown in Figure 1.

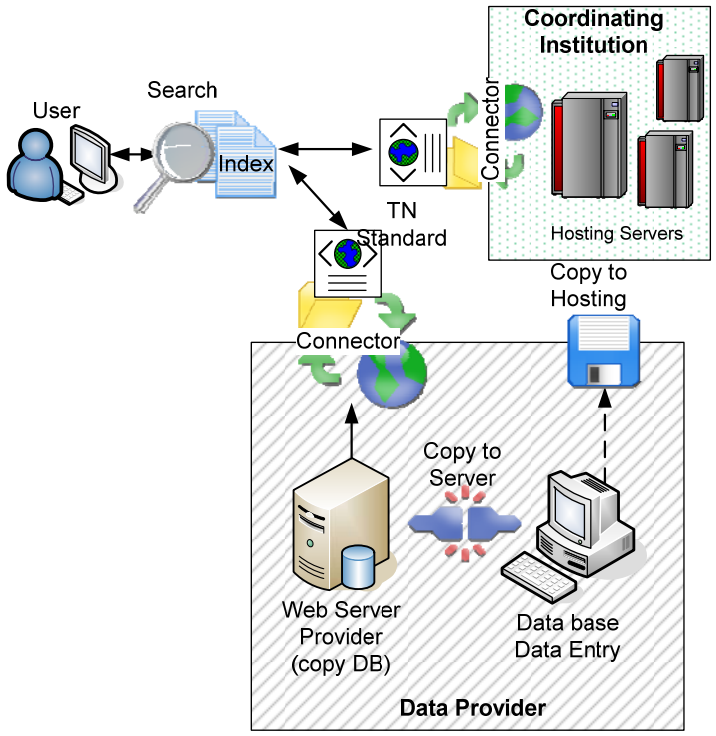


Figure - 1

The following elements can be appreciated in this scheme:

- ✓ **Data Providers:** Organizations willing to share their data through the Internet, following the policies established by IABIN. The data providers will continue to own their data. The data provider will be able to share his data through one of the following methods:
  - o Connecting directly to the Internet by installing a "CONNECTOR". For this, the data provider will need as a minimum:
    - An Internet server
    - Internet connection with 128 kbs bandwidth
  - o Or by taking advantage of the Hosting Service provided by the Coordinating Institution. For this, the data provider should give a copy of his data to the Coordinating Institution (or to any other organization, sometimes knows as "custodian", selected by the provider that belongs to the Network and could have a "CONNECTOR" installed). The data provider will be responsible for updating the data.

The following are the main commitments from the data providers:

- Permit access to his data
- Create and maintain metadata for the data he owns
- Ensure the quality of his data
- Update periodically his data
- Maintain communications with the Focal Point in his country and with the Coordinating Institution through which his data is being shared

✓ **Coordinating Institution:** Is the institution responsible for the implementation of one of the IABIN Thematic Networks. Among its main functions are:

- Create a Working Group
- Define information use policies for its Thematic Network (TN)
- Define, maintain and provide standards for its TN
- Provide Hosting Services to those organizations that would require them and which do not have the capacity to share their data through the Internet.
- Provide training (develop the course and the materials) on the use of the data digitizing tool
- Maintain the Thematic Network Internet portal
- Develop the data digitizing tool for its TN

✓ **Search engine:** This will search for, locate and index the data provided through the Network. This function will be carried out by the IABIN Catalog. This search engine will allow the user to search information about specific TN (Advanced search mode)

✓ **Connector:** This will be the element that would physically connect a data provider to the Network.

- Will receive and interpret the users requests sent by the Thematic Network or the IABIN Portal.
- Will search for the information requested in the provider's database
- Will return the requested information in the Standard established by the Thematic Network

This basic architecture will be the same to search and consult for biological or geographical data which can be accessed through the Network.

3. Implementation of the basic architecture of each of the Thematic Networks

The following chapter briefly describes how the basic IABIN architecture will be implemented by each one of IABIN Thematic Networks.

a. Species and Specimens Thematic Network

Objective	To make available existing information on Species (species description, observations and distribution) and Specimens (collections and observations).
Internet portal	http://specimens.iabin.net http://especies.iabin.net
Standards	✓ Specimens: Darwin Core and ABCD Schema ✓ Species: Plinian Core
Connector	TAPIR *Note: Data providers who are already using DiGIR to proved species data will be able to continue using this connector if they so desire.*
Digitalization	This TN is developing its own data digitizing tool which will integrate information on Species and Specimens. *Note: The data provider is free to choose the data digitizing tool of his preference as long as it fulfills the IABIN Standards.*
Integration with other IABIN Thematic Networks	Through: ✓ Taxonomic names ✓ Geo-referencing of specimens and observations ✓ Species distribution maps
Challenges	✓ Data quality ✓ Quality of the geo-referencing of existing data ✓ High number of data providers ✓ Possible duplication of data (the same data served through different providers or networks)
Observations	✓ This network is the result of merging the IABIN Species and Specimens Thematic Networks. It was a decision approved by the IABIN executive Committee in June, 2006.

b. Invasive Species Thematic Network

Objective	To make available existing information on Invasive Species by promoting the creation of a national database. This network is also known as I3N – IABIN Invasive Information Network. In addition to the biological information on the invasive species, this network collects information on the economic impacts and identified control measures.
Internet Portal	http://i3n.iabin.net
Standards	✓ I3N Standard *Note: The basis of the I3N Standard is compatible with the Darwin Core to which an extension for the management of data of interest for invasive species has been added (control, economic impacts, etc)*
Connector	TAPIR (future implementations) *Note: I3N is proposing to initially only deploy one TAPIR client in order to minimize impact in existing I3N data providers. The national databases presently operational will be connected in the near future.*
Digitalization	The network has developed its own data digitizing tool. *Note: The data providers must use the data digitizing tool developed by I3N since this tool captures the information of interest for the network (control, economic impacts, etc.)*
Integration with other IABIN Thematic Networks	Through: ✓ Taxonomic names. ✓ Geo-referencing of specimens and observations. ✓ This network will act as a data provider for the IABIN Species and Specimens Thematic Network, using the standards established for this network.
Challenges	✓ Obtaining the data ✓ Quality of the geo-referencing of existing data ✓ Adapting the existing software to use TAPIR ✓ Little knowledge in the countries about invasive species
Observations	✓ This is the most advanced of the IABIN Thematic Networks. ✓ The strategy for I3N calls for the identification and establishment of a National I3N Leader and the development of one National database for the management and control of invasive species in each country.

c. Pollinators Thematic Network

Objective	To make available the existing Pollinators data. In addition to the biological pollinator information, this network has plans to incorporate to the system the information about plants and their pollinating species. This information could have the associated economic information, in near future.
Internet Portal	http://pollinators.iabin.net http://polinizadores.iabin.net
Standards	✓ Specimens: Darwin Core and ABCD Schema ✓ Species: Plinian Core. ✓ *Note: The relationship between plants and their pollinators will be managed as an extension to the Darwin Core. This extension has to be developed and validated.*
Connector	TAPIR
Digitalization	To be determined. *At present there exist some tools that allow for the digitizing of pollinator collections. The option of using the same tool that is being developed for the Species and Specimens Thematic Network is being contemplated, just adding the extension for the plant-pollinator relationship.*
Integration with other IABIN Thematic Networks	Through: ✓ Taxonomic names ✓ Geo-referencing of specimens and observations, ✓ This network will act as a data provider for the IABIN Species and Specimens Thematic Network, using the standards developed for that network
Challenges	✓ Obtaining the data ✓ Quality of geo-referencing of the existing data ✓ Few pollinator databases in digital format ✓ The existing data digitizing tools for pollinators do not have information about the pollinating activity.
Observations	✓ The primary strategy of this network is to try to digitize the greatest amount of pollinator collections. ✓ The complex relationship between plants and pollinators needs to be included.

d. Ecosystems Thematic Network

<table>
<tr><td>Objective</td><td>To integrate the existing information on ecosystems (terrestrial, marine and continental waters) at the regional level. One of the main goals of this network is to create a cross-reference system that would allow carrying out crosswalks between the different ecosystems classifications used in the continent. In order to achieve this, a Standard Format was developed (GEOSS methodology) with five (5) levels.
<table>
<tr><th>Level</th><th>Theme</th><th>Terrestrial</th><th>Continental Waters</th><th>Marine</th></tr>
<tr><td>1</td><td>Global Macro-Bioclimatic & Biogeography</td><td>Polar, Boreal, Temperate, Med., Tropical (optional: add vegetation structural response [tree, shrub/herb, semi-desert, tundra, aquatic, sparse])</td><td>Polar, Temperate, Tropical</td><td>Polar, Temperate, Tropical</td></tr>
<tr><td>2</td><td>Meso-Bioclimate/ Biogeography</td><td>Biomes/phytogeography; relevant leaf phenology combinations + phytogeography</td><td>global biomes + zoogeographic basins/ecoregions (WWF/TNC)</td><td>marine regions defined by SST, sea surface elevation, mixed layer depth; biogeography</td></tr>
<tr><td>3</td><td>Geophysical</td><td>Plant-available soil moisture; hydrogeomorphology, specialized substrates</td><td>Ecological Drainage Units: climate/physiography/substrate porosity</td><td>Nearshore FW, Coastal Neritic, Oceanic</td></tr>
<tr><td>4</td><td>Geophysical Structure</td><td>Vegetation Structure Landscape mosaics</td><td>Surface water character determining biotic structure (e.g., lakes, rivers, streams)</td><td>e.g., Estuary, FW plume;
Reef;
Biomass estimate by water column layer</td></tr>
<tr><td>5</td><td>Biotic Composition</td><td>Gradient from secondary vegetation to undisturbed land cover; Characteristic Genera</td><td>Macrohabitats; Fish/invert communities</td><td>Coastal Macrohabitats; Plankton communities</td></tr>
</table>
Note: The countries will continue to use their own existing ecosystem classifications. The Standard Format is only a common way to describe each class.</td></tr>
<tr><td>Internet Portal</td><td>http://ecosystems.iabin.net
http://ecosistemas.iabin.net</td></tr>
<tr><td>Standards</td><td>✓ Standard Format for the description of an ecosystem.</td></tr>
<tr><td>Connector</td><td>✓ WS (Web Services) for the Standard Format and the Cross-reference.
✓ WFS for access to geographical data</td></tr>
<tr><td>Digitalization</td><td>This network developed its own data digitizing tool. This tool assists in filing in the Standard Format.</td></tr>
<tr><td>Integration with other IABIN Thematic Networks</td><td>Through:
✓ Geographical coordinates
✓ Lists of the dominant species in the ecosystem
✓ Geospatial integration.</td></tr>
<tr><td>Challenges</td><td>✓ Recollecting the data</td></tr>
</table>

	✓ The creation of cross-references. ✓ Several ecosystem types are used in the continent, which make it impossible to have 100% equivalency between one system and another. ✓ Difficult to fill in the information to Level 5 of the Format ✓ A large number of ecosystem information is found in maps. ✓ There is information on terrestrial ecosystems, but the information for marine and continental water ecosystems is scarce.
Observations	✓ It is expected that the Species, Specimens, Invasive Species and Pollinators Thematic Networks provide and digitize the information necessary (occurrences with coordinates or boundaries) to be able to determine the species existing within an ecosystem. ✓ This network will not digitize data on the species and specimens existing in an ecosystem.

e. Protected Areas Thematic Network

Objective	To make available the information regarding protected areas, having as the main priority the information about their management.
Internet Portal	http://protectedareas.iabin.net http://areasprotegidas.iabin.net
Standards	*To be approved:* ✓ WDPA Core Ver. 1.2 (World Data Base on Protected Areas)
Connector	*To be approved:* ✓ TAPIR. ✓ WFS for access to geographical data.
Digitalization	To be determined.
Integration with other IABIN Thematic Networks	Through: ✓ Geographical coordinates ✓ Lists of the dominant species in the protected area ✓ Geospatial integration.
Challenges	✓ Recollecting the data and updating the new version of the WDPA Core. (Version 1.2) ✓ A great deal of information about protected areas in found in maps.
Observations	✓ It is expected that the Species, Specimens, Invasive Species and Pollinators Thematic Networks provide and digitize the information necessary (occurrences with coordinates or boundaries) to be able to determine the species existing within a protected area. ✓ This network will not digitize data on the species and specimens existing in a protected area.

f. Geospatial Network

Objective	To make available the existing cartographic information. *Note: This network was not in the original plans for IABIN. It was created responding to the need to have access to the existing geographical information such as: country boundaries, cities, rivers, lakes, etc.*
Internet Portal	http://geoespatial.iabin.net http://geoespacial.iabin.net
Standards	✓ FDGC (Standard for spatial data)
Connector	✓ WFS for access to geographical data.
Digitalization	N/A.
Integration with other IABIN Thematic Networks	Through: ✓ Geographical coordinates ✓ Geospatial integration.
Challenges	✓ Standardization of the presentation of the different maps. ✓ In the integration of maps from different countries it will be necessary to reach agreement regarding boundaries. ✓ There is no coordinating institution for this network. It is expected that the Ecosystems and Protected Areas TNs will lead it. ✓ It is difficult to have access to the official cartographic information in each country. ✓ High Internet speed is needed.
Observations	✓ Its implementation is carried through the installation and integration of national map servers. ✓ It is possible that this network will disappear in the future and that the Ecosystems and Protected Areas TNs will assume a joint role in maintaining it.

g. Catalog

Objective	To integrate and facilitate the search for data and information provided by each Thematic Network. The IABIN Catalog will provide the following services: ✓ IABIN BioBot: Search engine to retrieve biological data in three languages (English, Spanish and Portuguese). ✓ UDDI: Registry of IABIN providers ✓ Geographical Index (Gazetteer) ✓ Organizational Index ✓ Common phrases ✓ Thesaurus ✓ Registry of Metadata ✓ Spatial Data Providers Registry *The Catalog will have the capacity to read and integrate databases of the data existing in the countries.* *The search engine will search for a word in English, Spanish and Portuguese, thanks to the Thesaurus and other multilingual control lists, but the content will be shown in its original language (it will not be translated).*
Internet Portal	N/A
Standards	✓ FDGC (Standard for spatial data) ✓ Dublin Core (Standard for documents, images)
Connector	✓ Web Services
Digitalization	✓ For digitization of metadata
Integration with other IABIN Thematic Networks	The IABIN Catalog will search through: ✓ Taxonomic names ✓ Common names ✓ Phrases ✓ Geographical coordinates and Geographical names
Challenges	✓ The Catalog requires high Internet connections. ✓ Quality of the metadata. ✓ Little development in metadata creation.
Observations	✓ IABIN will create a centralized thesaurus which will be fed from regional thesauri. Each term will be translated into three languages. ✓ The Thesaurus will also be available via web-services and for use/adoption by all TNs and the broader community within their applications and systems.

4. Integration of the Thematic Networks

Up to now we have described how IABIN will integrate information of the same kind (e.g. Invasive Species) owned by the different data providers.

However, in order for the information provided through each one of the IABIN Thematic Networks to be useful for decision making it will be necessary for the different kinds of data to be integrated as well.

The capacity to develop tools that would integrate in a homogeneous way this variety of data will allow for better decision making. Thus it is one of IABIN's greatest challenges to develop these tools.

It is also important that all IABIN partners understand the importance of standards for data digitizing, since it is these standards that will permit the integration.

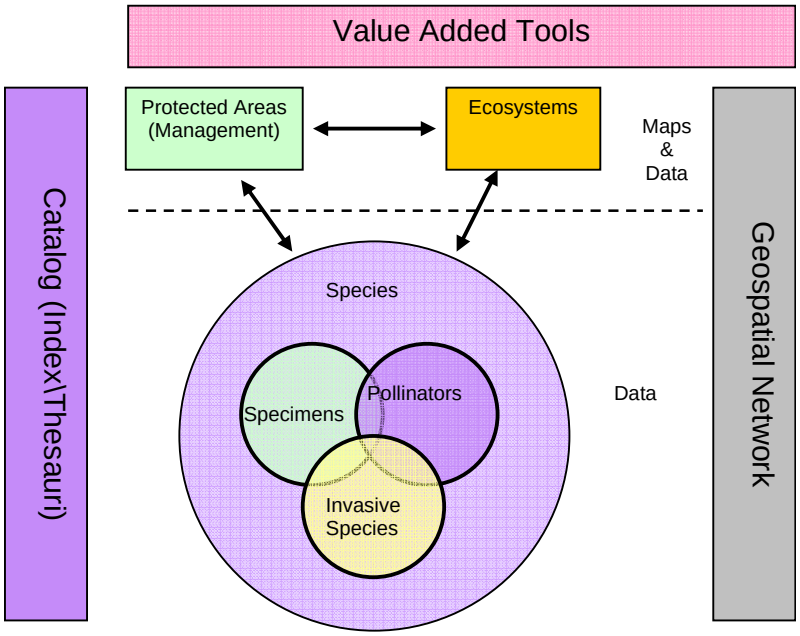


Figure - 2

a. Basic Integration Concept

The data that will be provided by the different IABIN data providers are basically of two types: primary biological data and spatial data. These kinds of data can be integrated by the following combinations:

- ✓ Biological data with biological data. This scheme will allow for the integration of the information existing about the different species and specimens. In order for this scheme to be applied it will be necessary to use common descriptors for the data.

- ✓ Biological data with spatial data. This scheme will allow for the integration of species and specimens information in the context of a map. For this scheme to work it will be necessary for the biological information to have coordinates for a geographical representation. The process of locating data in space is called geo-referencing.

- ✓ Spatial data with spatial data. This scheme will allow for the integration of the information contained in two maps, and will generate a new map.

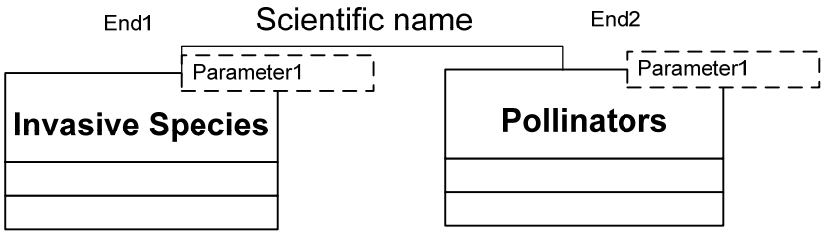
The process described above gives us IABIN's Architecture, which can be viewed in Figure 3. This figure shows how each one of the Thematic Networks and data providers will serve their data so all the other networks will be able to access them.

Due to the large variety of possible combinations that could be generated when combining different kinds of data, the IABIN Technical Working Group suggested that in a first phase the Network focuses on solving the basic integrations for each one of the integration schemes. These integrations will be:

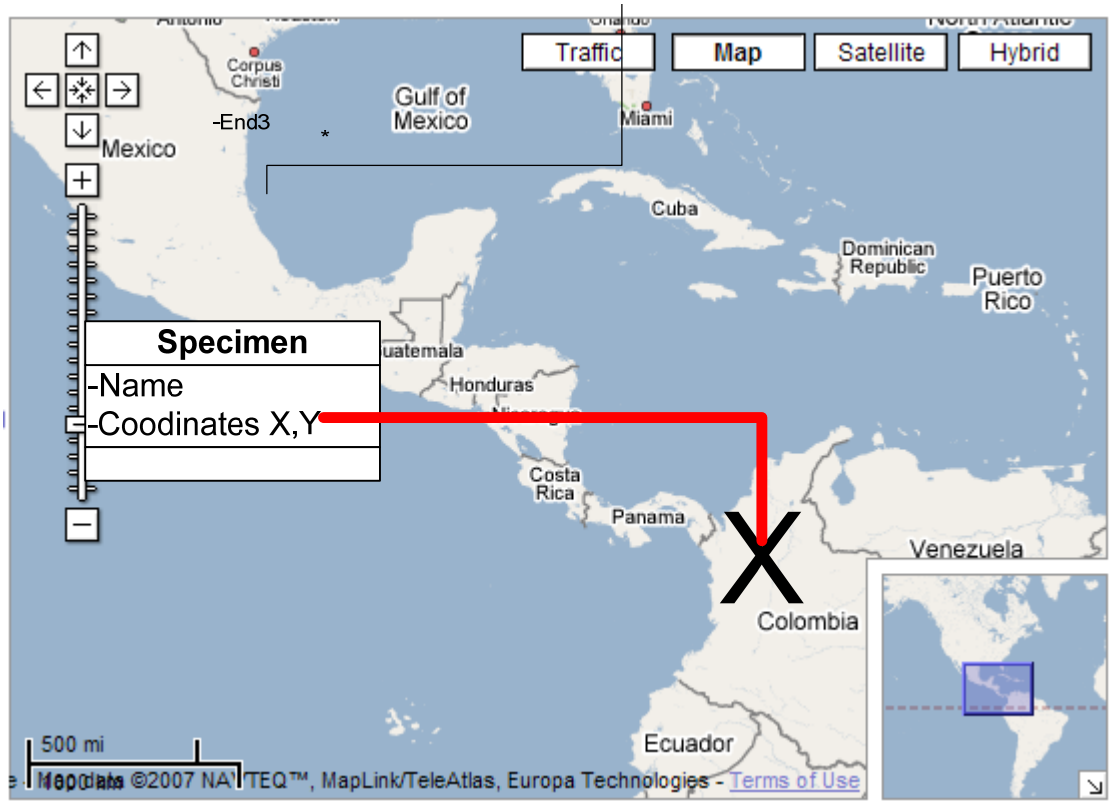
1 – The integration of biological data with biological data. It will be carried using only the scientific names of the species and specimens that are in each database. The reason for using only the scientific name is that this is the only element in all the record that in some way has been standardized internationally and for which databases from taxonomic authorities are being created, which will allow validating existing information. In addition, the scientific name is not affected by the language of the data provider.

It is important to clarify that the IABIN Catalog will be able to search the whole registry, for example using the common name, but the integration will only take into consideration the scientific name in order to determine if the registry makes reference to the same species or not.

The reason not to use other elements such as common name is that these may vary from one region to the next and from one country to another, in addition to making the search more complex in idiomatic terms.



2 – The integration of biological data with spatial data. This will only be carried out with those biological records that are geo-referenced; that is, those that will have geographic coordinates which will allow identifying the place where the species was collected or observed.



3 – *The integration of spatial data with spatial data.* This will be carried out through the superposition of maps.

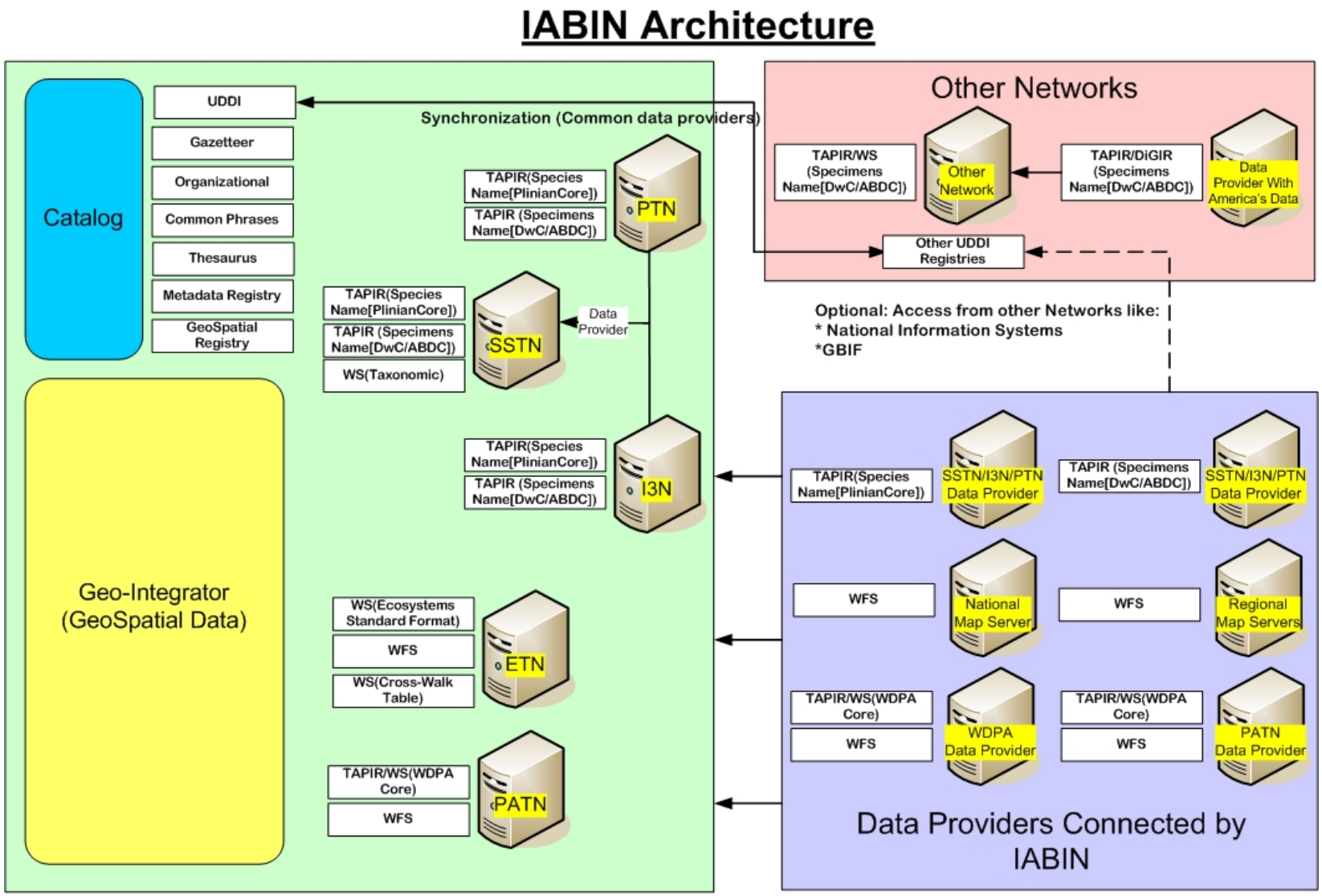

IABIN Architecture
Catalog
UDDI
Gazetteer
Organizational
Common Phrases
Thesaurus
Metadata Registry
GeoSpatial Registry
Synchronization (Common data providers)
Other Networks
TAPIR/WS (Specimens Name[DwC/ABDC])
Other Network
TAPIR/DiGIR (Specimens Name[DwC/ABDC])
Data Provider With America's Data
Other UDDI Registries
TAPIR(Species Name[PlinianCore])
TAPIR (Specimens Name[DwC/ABDC])
PTN
TAPIR(Species Name[PlinianCore])
TAPIR (Specimens Name[DwC/ABDC])
WS(Taxonomic)
SSTN
Data Provider
Optional: Access from other Networks like:
* National Information Systems
*GBIF
TAPIR(Species Name[PlinianCore])
TAPIR (Specimens Name[DwC/ABDC])
I3N
TAPIR(Species Name[PlinianCore])
SSTN/I3N/PTN Data Provider
TAPIR (Specimens Name[DwC/ABDC])
SSTN/I3N/PTN Data Provider
Geo-Integrator (GeoSpatial Data)
WS(Ecosystems Standard Format)
WFS
WS(Cross-Walk Table)
ETN
WFS
National Map Server
WFS
Regional Map Servers
TAPIR/WS(WDPA Core)
WFS
PATN
TAPIR/WS(WDPA Core)
WFS
WDPA Data Provider
TAPIR/WS(WDPA Core)
WFS
PATN Data Provider
Data Providers Connected by IABIN

Figure - 3

b. Use Cases in the First Phase (IABIN-GEF Project)

In the first phase of the implementation of IABIN it will be possible to carry out four kinds of integrated queries. These queries are called Use Cases, which will be:

1 – Search for a word or phrase using IABIN BioBot. It is expected that this will be the most common kind of search. In this search the user will enter a word such as "invasive species" and the IABIN BioBot will retrieve all existing information in the Network related to this phrase in English, Spanish and Portuguese. The retrieved information could be of any type: databases, tabular data, documents, images, etc.

2 – Search for species and specimens using their scientific name. The user will enter the scientific name of a species and the search engine developed by the Species and Specimens Thematic Network will retrieve all records available in the Network about that species.

3 – Search for the information existing within a geographical area selected by the user.

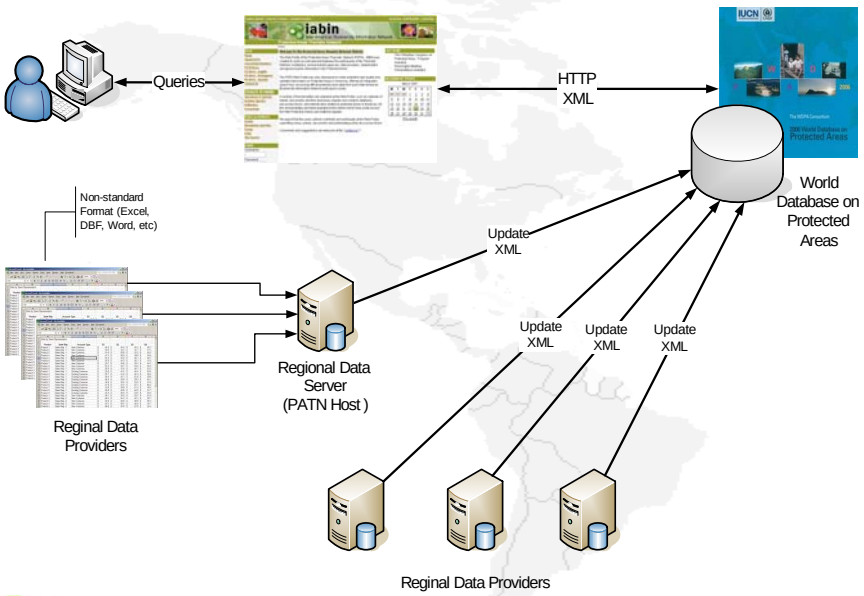
4 – Search for related information, around a radius of "X" kilometers, from a point determined by the user.

c. Integration with other Initiatives

It is foreseen that the information served by IABIN could be integrated with other initiatives, as long as these use the same standards and protocols for data exchange as IABIN.

When a data provider is connected to the Network, any other network could search his data, as long as the data provider agrees to this.

An example of this is the PATN proposed Architecture which has integration with World Database on Protected Areas (WDPA).



5. Standards and Protocols

The initial standards and protocols for IABIN were approved at the 4th IABIN Council Meeting, in April 2005.

It is an ongoing task of the IABIN Technical Working Group to review and update the approved standards. The following table shows the IABIN Standards and Protocols, updated at the last meeting of the IABIN ITWG in March, 2007.

Part of IABIN Architecture	Standard or Protocol Adopted
Architecture	Web Services
Registry Services	UDDI
Interface description	WSDL
Access protocols	TAPIR DiGIR (if the provider has it integrated)
Data coding	XML
Data transport	HTTP over TCP/IP
Metadata	
o For bibliographical data	Dublin Core
o For specimen collections and observations	Darwin Core ABCD Schema
o For Species	Plinian Core
o For Protected Areas	WDPA Core Version 2
o For Invasive Species	I3N Standard
o For Spatial Data	FGDC
o For general biological resources	CSDGM with Bio Profile
o For geographical data processing	Open GIS Consortium (OGC) WFS WMS (only if WFS is not available)
For document format	HTML, PDF, and ASCII
Graphic format	PNG, JPEG, GIF, WebCGM

6. General Considerations for the IABIN Architecture.

The architecture presented in this document, and at the 5th IABIN Council Meeting, is an architecture that is in continuous evolution and will have to be adjusted according to the technical requirements of each Thematic Network and to new technologies.

However, there are some general considerations:

a. Data quality

Ensuring the quality of the data served by the Network should be one of the main objectives of IABIN. Only thus it will be possible to ensure that the data can be used for decision making.

It is impossible for IABIN to check the quality of each data that is shared through the Network, but each Thematic Network will establish feedback mechanisms

through which the data providers will be notified of the errors encountered in their data so they can correct them.

This mechanisms will allow the search engine, the IABIN BioBot, to give greater priority (greater weight) to those data providers whose data are determined as highly trustworthy and, in the same manner, give lower priority to those data providers whose data are reported as faulty in successive searches.

With this mechanism the responsibility to ensure data quality will be that of the data providers who own the data.

IABIN, through its training component, will focus on training data providers on ways to correct their data. The most common problems found regarding data quality are of two types:

i. Biological data do not comply with the established standards. This is because most of the data were digitized when these standards did not exist yet, so the data will have to be modified to comply with the standards.

ii. Geo-referencing of the existing data. The geo-referencing methodologies using digital maps and global positioning systems (GPS) are relatively new. Thus, a large number of data collected before them have this type of problem and this has to be corrected in order for the data to be used in decision making.

7. Annexes

Annex 1: IABIN guiding principles

IABIN has adopted 11 guiding principles for interoperability formats, standards and protocols:

1. Seamless access to all types of IABIN data and information regardless of where it resides and interoperable with both CBD-CHM and GBIF;
2. Open, widely supported, non-proprietary standards;
3. Compatibility with emerging standards of key regional, global and national biological information networks;
4. Minimization of technology restrictions imposed by the network architecture;
5. Phased, incremental development;
6. Scalability, so that standards will be usable and applicable at different network scales;
7. Inclusion (e.g. facilitate local-language queries) in the design of applications;
8. Expertise and capabilities are shared throughout the network;
9. Respect for Intellectual Property Rights, Traditional Knowledge rights and rules for access and benefit sharing of Genetic Resources in accordance with the CBD principles and guidelines, and national legislations;
10. Future extensibility and backward compatibility;
11. Minimization of cost while ensuring reliable user services

Annex 2. Difference between IABIN BioBot and Google
(By Mike Frame, USGS)

All search engines are not the same due to a number of characteristics including search engine contents/scope, ranking algorithms deployed, or the use of controlled vocabulary or terminologies to support retrieval of data or information. Below is a basic comparison of the IABIN BioBot and a typical web-based search engine such as Google.

Google:

- Google's mission is to make the entire Web universe findable and visible. This does not include non-web-based content held in databases or restricted scientific domain servers.
- Google cannot interpret context of a user's search. Google executes a search and does not know the domain for the search. For instance, a query term of "hawks" in Google would return results for sports teams, species, and/or a TV/Movie star.
- Google does not understand what a document contains or what is its overall intent or context.
- Google's ranking algorithm is primarily based upon inbound links; resources more often linked to each other are ranked higher. This method for retrieval of scientific data may not be the best approach to insure duplication of research or innovative research occurs.

IABIN BioBot:

- IABINs mission is to make the biological or ecological Web more findable.
- IABIN BioBot interprets context within metadata (subject, taxonomic, geographic) to help present accurate and distributed access to resources.
- What a document is about is understood by IABIN BioBot due to the fact that its domain and/or primary data holdings relates to the biological sciences.
- IABIN BioBot's ranking algorithm is based upon weightings of metadata fields, and weighting of "Reliable/High Quality/Trusted" sources higher than none "partner" content. This helps to insure higher quality and specific relevant results to user queries.
- IABIN BioBot results are presented in a Tabbed interface to allow users to view "All" results by default, or select on specific resource types (i.e. Maps, Images, Journal articles, etc.). This helps to categorize information to users instead of simply presenting a long narrative list of results.
- IABIN BioBot deploys an underlining thesaurus to aid users in the retrieval of information that may not have been catalogued or described in the same context or with the same terminology as the user query. For instance, if a user queries for "invasive species", typical web-based search engines can only retrieve data/information with that exact phrase. Through the use of a thesaurus, IABIN BioBot users are presented with results from related terms to "invasive species" such as "alien species", "non-native species", etc.