

Technical and Financial Progress Report

Web-services of datasets and data providers using metadata standards Dublin Core and Federal Geographic Data Committee – Biological Data Profile (FGDC-BDP)



Prepared by:

Prof. Dr. Pedro Luiz Pizzigatti Corrêa (Coordinator)
Prof. Dr. Antônio Mauro Saraiva (Vice-Coordinator)
Leandro Halle Najm (Computer Scientist – Master Student) – Team manager
Marcelo Queiroz Leite (System Analyst – Master Student) – Software Developer Leader
{pedro.correa, saraiva, hallenajm, mql}@usp.br
Nelson Mimura Gonzalez (Computer Engineer student) – Software Developer
Rafael Pelegrini Domingues (Computer Engineer student) – Software Developer
{ nelson.gonzalez, rafael.domingues}@poli.usp.br

19/04/2010

This report is the IABIN Catalog Content progress report that is made to describe the status of the research and development at this point. The topics of this report are Introduction, Background and Outros

1. Introduction

The proposal sent to Request for Proposals (RFP) were to Specify and develop web-services of data providers and datasets in the Dublin Core and FGDC BDP standards, respectively, using architecture-based on services developed by the GBIF Portal. This proposal involves the extension of current GBIF services (see <http://data.gbif.org/tutorial/services>) to support Dublin Core and FGDC BDP metadata standards. Other objective is that the services will be integrated with the IABIN Pollinator Thematic Network (IABIN/PTN) Portal. This source code will be available to be used by IABIN Species and Specimen Thematic Network SSTN/IABIN and GBIF.

The data providers will be in Dublin Core metadata standard and data sets metadata should be in FGDC BDP standard.

According to the proposal, the Initiation plan was:

- I. *Communication and interviews with the IABIN Catalog Focal Point for best practices, more project details, and information contact exchange.*
- II. *Analysis of metadata standards of IABIN Catalog, Dublin Core and FGDC BDP specifically;*
- III. *How the data is currently stored;*
- IV. *What data can be collected;*
- V. *Definition and separation of activities to the project team;*
- VI. *Documentation.*

In this progress report this plan was explored and finished, and the Planning plan, is in progress. The description of the features is in topic 2, materials and methods.

2. Background

GBIF Data Portal of Biodiversity

The portals used by the IABIN of Species and Specimens Thematic Network (SSTN) and the Pollinators Thematic Network (PTN) are both based on the GBIF Data Portal (GBIF, 2010).

Web services to be developed in this project will be implemented integrated with the source code of the portal. So the following text presents the main features of the portal, in addition to software components and frameworks used in this implementation.

GBIF is an international initiative that has as main goal to make biodiversity data accessible everywhere in the world. To enable this data sharing, the portal has made use of various computer technologies, so as to enable sample collection directly from their providers. This portal was developed using Java and

has as main characteristic the use of frameworks like Spring and Hibernate, which provides modularity in your code aiding in future developments, as well as providing a good programming practice.

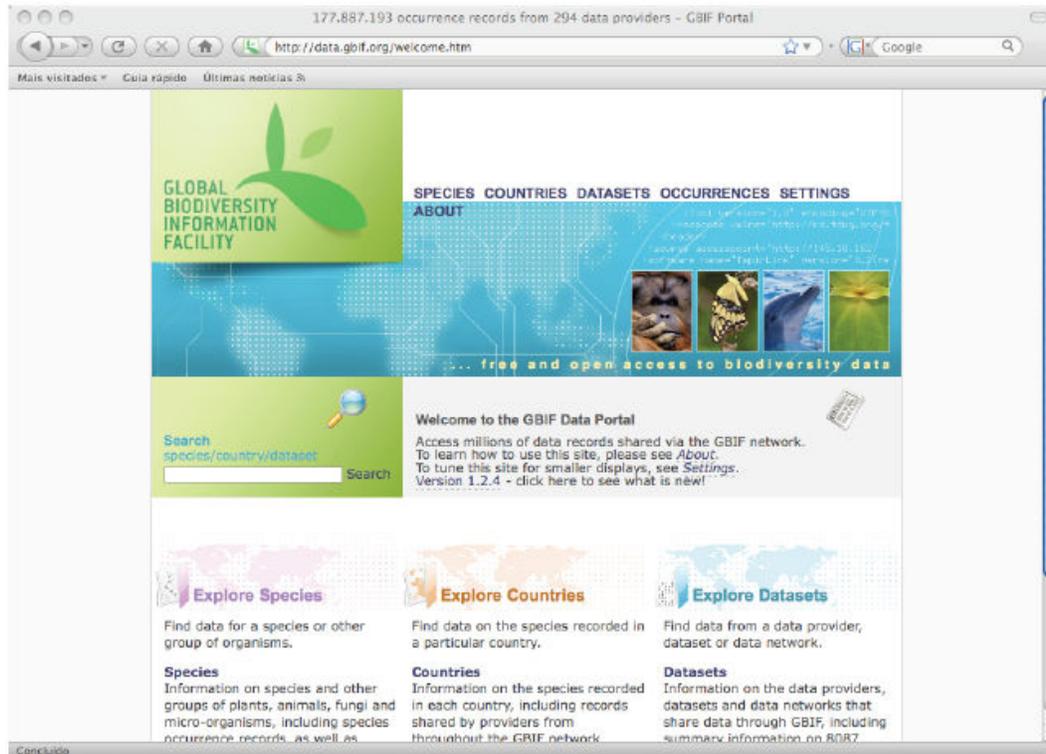


Figure 1 - Home page of the portal of GBIF (GBIF, 2010)

The GBIF Data Portal provides the main resources available to end users:

- Centralized storage of homogeneous data of specimens: based on specific fields of biodiversity (defined by the standard DWC) are stored records from providers, which are available for access on the Data Portal;
- Tool for occurrences, datasets and countries: with specific filters, the End User can choose to find records of biodiversity, scientific name, family, parents or even access to the original provider;
- Navigation in the taxonomic tree: This feature allows searching based on the taxonomic tree of the species;
- Display geo-referenced: Some records can provide information about the location of your collection, and in these cases, the portal offers a map where the location is displayed to the User;
- Support for multiple languages: Due to access to the portal is carried out by the world, this feature allows the researchers to choose the available language of your choice.

The portal was developed in a modular way, allowing their reuse for expansion of functionalities. The portal was developed in 5 modules, which have specific functions, as brief description below:

- portal-core: provides the components of utilities and processes used by other modules;

- portal-index: this module makes all the stage of indexing and communication with providers in order to make the data available for access by the tools of the portal;
- portal-service: provides the DAO (Data Access Objects) / DTO (Data Transfer Objects) and direct integration with Hibernate for data access by the tools of the portal;
- web portal: contains all the graphics and visual portal. It is this module that generates the web environment, necessary for access. It requires other modules for operation;
- portal web service: This module is responsible for the resources to extend the portal, making possible integration of other specific tools and turn available new web services.

All modules are integrated and managed through the Java framework Spring.

"... The Spring provides a lightweight solution for building applications business and at the same time, continue to support the possibility of use of transaction management of remote access to its logic services using RMI (Remote Method Invocation) or Web services, resources for sending e-mail and various options for persisting your data a database. The Spring provides an MVC framework (Model View Controller), transparent ways of integrating AOP (Aspect Oriented Programming) to your software and a well-structured hierarchy of exceptions, including automatic mapping from the hierarchies own exceptions.

The Spring, however, may be the ideal place for all your applications corporate, it is modular, allowing you to use parts of it, without add the remaining parts ... " (HEMRAJANI, 2006)

According to Figure 2, the Data Portal organization MVC between the different components of the portal:

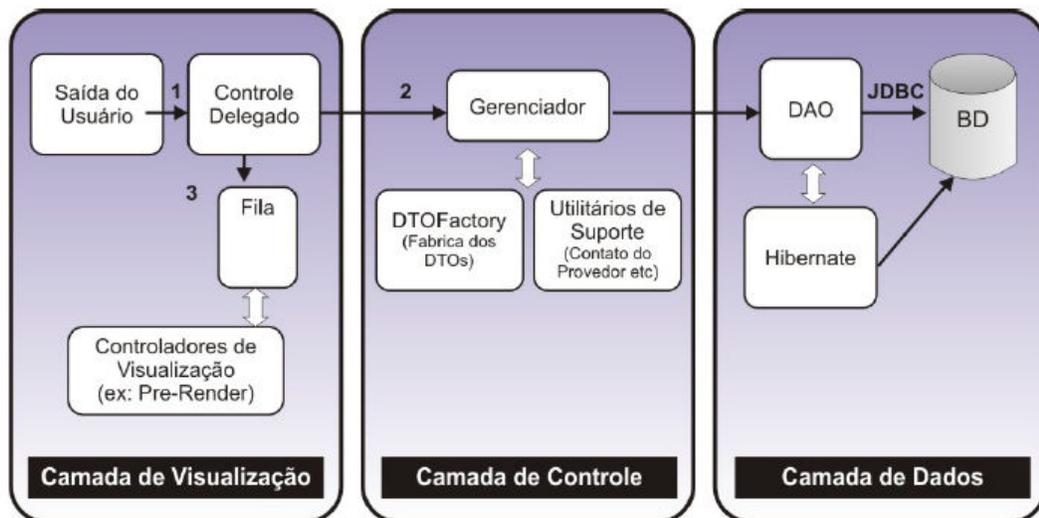


Figure 2 - Layer MVC components used by the GBIF portal.

2.2 IABIN Pollinator Thematic Network Data Portal

Inter-American Biodiversity Information Network (IABIN)

The Inter-American Biodiversity Information Network (IABIN) is a forum to foster technical collaboration and coordination among countries of the Americas in collection, sharing, and use of biodiversity information relevant to decision-making on natural resources management and conservation, and education to promote sustainable development in the region.

Goals

1. Build up an infrastructure for biodiversity information exchange.
2. Strengthen technical capacity to exchange biodiversity information and expertise across political, linguistic, and institutional boundaries.
3. Provide access to biodiversity information useful to decision-makers to improve biodiversity conservation
4. Enhance capacity to store, use, and distribute scientifically sound and update biodiversity information.
5. Produce or adapt information products for decision makers (tools for decision-making) so they formulate effective environmental management policies and promote sustainable development in the region.

IABIN Pollinators Thematic Network (IABIN-PTN)

The IABIN Pollinators Thematic Network (IABIN PTN) is one of six thematic networks of the IABIN. It is developing a Western Hemisphere network for access to and exchange of pollinator information. Bees, bats, and hummingbirds are some of the most important pollinators of wild and cultivated plants in the Western Hemisphere. Pollinators provide an essential ecological service to wild and cultivated plants. However we do not yet have the basic knowledge required to design a comprehensive action plan to monitor, manage, and conserve wild pollinators.

We need to know how many pollinating species there are, their names, where they live, and what plant species they pollinate. Without this information we will not be able to link the scattered biological, ecological and agricultural data in an efficient retrieval system. By facilitating access to and exchange of these data, the IABIN PTN will support successful monitoring and management of pollinator populations throughout the Western Hemisphere.

The Pollinator Data Portal (PDP) <<http://pollinators.iabin.net/portal/>> allows data owners to publish and standardize digital pollinator observation, specimen, and interaction data online for scientific research and public access. These data are generously made available through the IABIN by a wide range of institutions and organizations from around the world.

The types of data being shared through the IABIN PTN data portal are:

1. **Species occurrence records (based on specimens and observations)** - information about the occurrence of species at particular times and places (pollinators and plants).

2. **Names and classifications of organisms** - information on the names (both scientific and common) used for species and on the classification of those organisms into taxonomic hierarchies.
3. **Species Interaction records** – information about interactions among pollinators and plants.

The PDP was developed based on the project and source code of the Global Biodiversity Information Facility (GBIF) Data Portal <<http://www.gbif.org>>. It uses the JAVA programming language, and various OpenSource technologies including: SpringFramework, Hibernate, DisplayTag, JFreeChart, Tiles, Struts, Quatz, DWR, XmlBeans, Maven Server and MySql Databases.

The Pollinator Data Portal differs from the GBIF Data Portal in that the layout of the user interface and the harvesting process have been modified to accommodate species interactions data and additional resources for researching and displaying interaction data.

2.3 Data Providers and Datasets

Although it is the intent of the IABIN PTN to provide free and unrestricted assistance to data owners who wish to share their pollinator data, the consortium recognizes the need to ensure a level of quality for both data owners and users of the PDP.

The IABIN PTN will accept datasets contributed by recognized organizations such as museums, universities, and research organizations. Datasets will also be accepted from individual data owners if the data owner can demonstrate the validity of the dataset and their expertise in collecting/curating the dataset to the satisfaction of the IABIN PTN Consortium. Each data owner will be evaluated on a case by case basis.

Accepted evidence for an individual data owner's expertise includes a degree or other certification in the biological, computer, natural or library sciences conferred by a recognized tertiary or other educational institution; and/or evidence of experience including at least three professional references who can attest to the applicant's expertise/experience in relation to the dataset.

Dataset Mapping

Data owners will need to map the fields of their pollinator dataset to the fields of the PDD. Detailed instructions on how to do this are available in the PDD Manual.

In addition to mapping the fields of their dataset to those of the PDP, data owners must prepare their dataset for publishing by defining which fields will be made available through the PDP and configuring the system to index them correctly.

Dataset Harvesting

It is the process of publishing information from their owners to the PDP in order to be available on the Internet. For this process to concretize must register the provider in this site and from time to time get information from this provider.

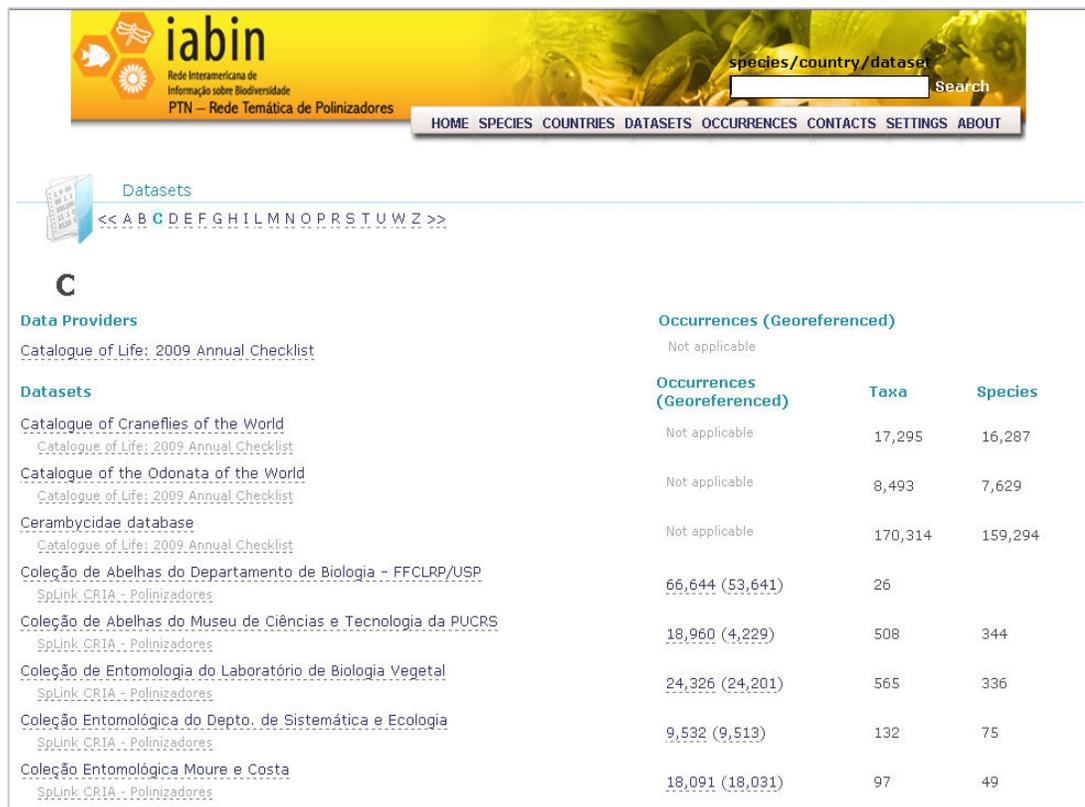
Over time only the new or changed data will be sought, considering that initial harvest will bring all records from this server.

Data Sorted By Data Provider and Dataset

The **Datasets** area of the PDP can be accessed by clicking on the **Datasets** link in the navigation bar at the top of the **Home** page. An alphabetical list of data providers and datasets is displayed in the **Datasets** area along with the number of Georeferenced Occurrence records, Taxa, and Species records published by each data provider or in each dataset.

Click on a letter in the alphabetical list at the top of the page to display a list of data providers and datasets whose names start with that letter.

Click on a data provider or dataset name to view more information about the selected data provider (Data Provider overview page, Section 0) or dataset (Dataset overview Page, Section **Error! Reference source not found.**).



The screenshot shows the iabin website interface. At the top, there is a navigation bar with the following links: HOME, SPECIES, COUNTRIES, DATASETS, OCCURRENCES, CONTACTS, SETTINGS, and ABOUT. Below the navigation bar, there is a search bar with the text "species/country/dataset" and a "Search" button. The main content area is titled "Datasets" and features an alphabetical navigation bar with letters from A to Z. The letter "C" is selected, and the page displays a list of datasets under the heading "Data Providers".

Data Providers	Occurrences (Georeferenced)	Taxa	Species
Catalogue of Life: 2009 Annual Checklist	Not applicable		
Datasets	Occurrences (Georeferenced)	Taxa	Species
Catalogue of Craneflies of the World Catalogue of Life: 2009 Annual Checklist	Not applicable	17,295	16,287
Catalogue of the Odonata of the World Catalogue of Life: 2009 Annual Checklist	Not applicable	8,493	7,629
Cerambycidae database Catalogue of Life: 2009 Annual Checklist	Not applicable	170,314	159,294
Coleção de Abelhas do Departamento de Biologia - FFCLRP/USP SpLink CRIA - Polinizadores	66,644 (53,641)	26	
Coleção de Abelhas do Museu de Ciências e Tecnologia da PUCRS SpLink CRIA - Polinizadores	18,960 (4,229)	508	344
Coleção de Entomologia do Laboratório de Biologia Vegetal SpLink CRIA - Polinizadores	24,326 (24,201)	565	336
Coleção Entomológica do Depto. de Sistemática e Ecologia SpLink CRIA - Polinizadores	9,532 (9,513)	132	75
Coleção Entomológica Moure e Costa SpLink CRIA - Polinizadores	18,091 (18,031)	97	49

Figure 3 - List of Data Providers and Datasets.

Data Provider Overview Page

For data providers that publish datasets that contain georeferenced records, the **Data Provider** overview page includes an **Actions** box, **Occurrence Overview** map, list of **Indexed Data** - including Event Logs (Section **Error! Reference source not found.**) and Indexing History, descriptive information about the data provider, and a list of other datasets published by the data provider. Click on the name of a dataset in the list of datasets to view a **Dataset** overview page.

For data providers that publish datasets that do not contain georeferenced records, the **Data Provider** overview page only includes descriptive information about the data provider and a list of other datasets published by the data provider (Figure 10). Click on the name of a dataset in the list of datasets to view a **Dataset** overview page (Section **Error! Reference source not found.**).

2.4 Software Tools

Java

Java is an interpreted programming language and interpreted are in *bytecode* format. The *bytecode* may be interpreted by a JVM (Java Virtual Machine) or compiled JIT (Just in Time) by it. This handicap, that may jeopardize efficiency, is compensated by the fact that it is easy to develop and deploy on Java.

Java can be downloaded at Sun's website:

<http://java.sun.com/javase/downloads/index.jsp>

It is important to download the latest version.

Eclipse

Eclipse is a multi-language IDE (Integrated Development Environment) developed in Java. One feature that had great weight on the choice for the IDE is the extensible plug-in system, that makes the environment flexible for development. Such extensions allow the easy development of web-applications, GUI (Graphical User Interface) and many other kind of applications.

The necessity of a remote repository is suppressed by the Subclipse plug-in, that offers an integrated Subversion client. This client uses a GUI for handling requests, as well as context menus.

Since Eclipse is under the Eclipse Public License, it is free and open source software, which reduces the cost for development and gives us freedom to adapt the software to our necessities.

Eclipse can be obtained at:

<http://www.eclipse.org/downloads/>

It is required the version for Java EE (Enterprise Edition).

Web services

Web services are application components that can be used over a network by other applications. Since Web services communicate using open protocols, are self-contained and self-describing, they are ideal to proffer a remote API. Its basic platform consists of XML (Extensible Markup Language) and HTTP (Hypertext Transfer Protocol), and its elements are: SOAP (Simple Object Access Protocol), UDDI (Universal Description, Discovery and Integration) and WSDL (Web Services Description Language).

Axis2

A Web services framework developed by the Apache Software Foundation, available in C and Java. It provides the capability to add web services interfaces to Web applications, as well as a standalone

server application. The possibility to work with WS-* interfaces of Web services and REST simultaneously was decisive on the choice of the framework.

Axis2 usually comes embedded to Eclipse as a plugin, so its installation may be not necessary. During project creation some additional steps will be provided in case any manual installation is needed.

Tomcat

Tomcat can be downloaded here:

<http://tomcat.apache.org/download-60.cgi>

For Windows the best solution is installing Tomcat as a Windows Service. For other operating systems, the binary distribution is enough.

Eclipse allows the creation of well-structured web services using Java as programming language, Axis2 as web services framework and Tomcat for service publishing. This tutorial aims at creating a simple and intuitive web service using these tools.

3. Partial Results

The partial report approaches to equalize the IABIN Catalog and USP Project Team knowledge and main goals, discover the exactly number of digitalized data and how to stored it and define technologies and strategies to achieve the main objective.

3.1 Focal Point communication

The Focal Point of IABIN in Brazil is Braulio Dias, from Environmental Ministry of Brazil (Ministério do Meio ambiente). The contact was made by e-mail and phone call (55 61 20282028), the conversation was made with his secretary, Cibelli Castro. The e-mail is copied below.

From: **Pedro Corrêa** <pedro.correa@usp.br>
Date: 2010/3/30
Subject: IABIN Catalog Approved Project at USP/BRAZIL
To: Braulio.dias@mma.gov.br
Cc: "Huber, Richard" <RHuber@oas.org>, "Neira, Luisa Fernanda" <LNeira@oas.org>, cibelly.castro@mma.gov.br

Prezado Braulio Dias,

Sou docente da Escola Politécnica da USP e recentemente submeti e foi aceito um projeto a uma Chamada da IABIN (Inter-American Biodiversity Information Network).

Encaminho o Documento Anexo da proposta do projeto denominado "Web-services of datasets and data providers using metadata standards Dublin Core and Federal Geographic Data Committee – Biological Data Profile (FGDC/BDP)", juntamente com a carta de aceite.

Por favor, gostaria de solicitar que possa responder a esse e-mail, dando ciência do projeto. Essa solicitação deve-se ao fato de que uma das atividades iniciais do projeto é comunicar ao "Focal Point" da IABIN no Brasil sobre o desenvolvimento do Projeto.

Estou a disposição para maiores esclarecimentos,

Muito obrigado.

Atenciosamente,
Pedro Luiz Pizzigatti Corrêa.

English version:

Dear Braulio Dias,

I am a professor at the Engineering School of the USP (University of São Paulo - Brazil) and recently submitted and was accepted a project to a call of IABIN (Inter-American Biodiversity Information Network).

Please, find attached the document of the project proposal entitled "Web-services of data providers and datasets using Dublin Core metadata standards and Federal Geographic Data Committee - Biological Data Profile (FGDC / BDP), together with a letter of acceptance.

Please, I would ask you that can reply this e-mail, informing the project. This request is due to the fact that one of the activities of the project is to communicate to the IABIN "Focal Point" in Brazil on the development of the project.

I am available for further clarification

Thank you very much.

Prof. Dr. Pedro Luiz Pizzigatti Corrêa	University of São Paulo
Lecturer	School of Engineering
pedro.correa@poli.usp.br	Computer Engineering Department
www.pcs.poli.usp.br	PB: 61548 São Paulo-SP, Brazil - Zip: 05508-900
Phone: 55+11+3091-5366	Fax: 55+11+3091-5294

 **DataSet-DataProviderProposal-review.doc** 107K

 **IABIN Acceptance Notification OAS Catalog- USP Web-services.pdf**
627K

3.2 Data-set and Data-provider Metadata

The Standards required to this work were researched and studied, as well as the IABIN Portal actual metadata used for records, DarwinCore.

The access to the IABIN Portal is complete, and the search returns data-providers and data-sets in a GBIF metadata, different from the records, that is Darwin Core standard:

The IABIN Portal data-sets and data-providers data is stored in the GBIF metadata as can be seen below:

```
<gbif:dataResource> Where is the data-provider;  
<gbif:name>Name of the data-provider  
<gbif:description>Description of the data-provider  
<gbif:defaultBasisOfRecord>Type of the records  
<gbif:created>When the data-provider was created  
<gbif:modified> Last data-provider was modified  
<gbif:accessPoints> Access point URL
```

One example of returned by GBIF data-provider metadata record is shown below:

```
<gbif:dataResource rdf:about="http://200.144.189.73/portal/ws/rest/resource/get/28" gbifKey="28">
```

```
  <gbif:name>Catalogue of the Odonata of the World</gbif:name>
```

```
    <gbif:description>The Global Species Database of Odonata is a list of all names of extant species of dragonflies and damselflies of the world. The present list is based on a database built during a period of twenty years. The original information was derived from a published catalogue by Davies & Tobin (1984, 1985). However, original publications were consulted to check the spelling and original combination of (virtually) all names. The present taxonomic status of every name in this list is based on published authoritative sources. Addition of the names of fossil taxa, and setting up a management committee for this database, are to be expected in the near future. The original information of the GSD Odonata is managed using a FoxPro application.</gbif:description>
```

```
    <gbif:defaultBasisOfRecord>taxonomy</gbif:defaultBasisOfRecord>
```

```
      <gbif:created>2009-08-04 03:34:53.0</gbif:created>
```

```
      <gbif:modified>2009-08-04 03:34:53.0</gbif:modified>
```

```
      <gbif:accessPoints></gbif:accessPoints>
```

```
</gbif:dataResource>
```

One example of returned GBIF data-set metadata record is showed below:

```
<gbif:dataResource rdf:about="http://200.144.189.73/portal/ws/rest/resource/get/3" gbifKey="3">
```

```
  <gbif:name>AlgaeBase</gbif:name>
```

```
    <gbif:description>Algaebase aspires to include all organisms regarded as algae, including the Cyanophyta (or Cyanobacteria). Primarily a taxonomic database, it has diverse information on freshwater, marine (seaweeds) and terrestrial algae, including uses and many pictures. Seagrasses are also included for the present. About 55,000 names are currently documented. AlgaeBase was developed at the Martin Ryan Institute, National University of Ireland, Galway, with support from the Higher Education Authority (Irish Department of Education and Science) and the European Commission.</gbif:description>
```

```
    <gbif:defaultBasisOfRecord>taxonomy</gbif:defaultBasisOfRecord>
```

```
      <gbif:created>2009-08-04 03:34:53.0</gbif:created>
```

```
      <gbif:modified>2009-08-04 03:34:53.0</gbif:modified>
```

```
      <gbif:accessPoints> </gbif:accessPoints>
```

```
</gbif:dataResource>
```

The further step is to relate this metadata standard into the FGDC.

3.3 Collected data

The total of records catalogued in Iabin Portal is 7 data-providers, listed below:

1. Catalogue of Life: 2009 Annual Checklist
2. Instituto de Ciencias Naturales
3. Instituto de Ciencias Naturales – Interaction
4. Portal
5. SpLink CRIA – Polinizadores
6. Webbee pollinator interaction provider (CEPANN)
7. Webbee pollinator specimen provider (CEPANN)

The total of records catalogued in labin Portal is 80 data-sets, listed below:

1. Coleção de Abelhas do Departamento de Biologia - FFCLRP/USP
2. Coleção de Abelhas do Museu de Ciências e Tecnologia da PUCRS
3. Coleção de Entomologia do Laboratório de Biologia Vegetal
4. Coleção Entomológica do Depto. de Sistemática e Ecologia
5. Coleção Entomológica Moure e Costa
6. Coleção Entomológica Paulo Nogueira-Neto - IB/USP
7. Coleção Entomológica Pe. Jesus Santiago Moure (Hymenoptera)
8. Laboratório de Ecologia e Biogeografia de Insetos da Caatinga
9. A World Catalogue of Centipedes (Chilopoda) for the Web
10. AlgaeBase
11. Aphid Species File
12. Australian Faunal Directory (Pulmonata)
13. Bacteriology Insight Orienting System
14. BioSystematic Database of World Diptera
15. Blattodea Species File Online
16. Brachiopoda & Phoronida World Databases
17. Catalogue of Craneflies of the World
18. Catalogue of the Odonata of the World
19. Cerambycidae database
20. Computer Aided Identification of Phlebotomine sandflies of Americas
21. Conifer Database
22. COOL: Cercopoidea Organised On Line
23. Cumacea World Database
24. Droseraceae Database
25. Electronic Catalogue of Weevil names (Curculionoidea)
26. FishBase
27. Fulgoromorpha Lists On the WEB
28. Geranium Taxonomic Information System
29. Global Butterfly Information System
30. Global Taxonomic Database of Gracillariidae
31. Hexacorallians of the World
32. Home of the Xylariaceae
33. ILDIS World Database of Legumes
34. International Committee on Taxonomy of Viruses / Master Species List
35. IOPI Global Plant Checklist
36. ITIS World Bee Checklist

37. Lacistemataceae Holistic Database
38. Mantodea Species File Online
39. Mites Global Species Databases
40. Moss TROPICOS Database
41. Neuropterida Species of the World
42. Nomenclatural Database of Eumycetozoa (Myxomycota)
43. Orthoptera Species File Online
44. Parhost World Database of Fleas
45. Phasmida Species File
46. Phyllachorales
47. Phylogeny and taxonomy of Glomeromycota (arbuscular mycorrhizal (AM) and related fungi)
48. Plecoptera Species File Online
49. Psylloidea database
50. Rhytismatales
51. Rotifera Database
52. Solanaceae Source
53. Species 2000 additional Common Names for the Catalogue of Life
54. Species Fungorum
55. Systematic Database of the Scale Insects of the World
56. Systematic Myriapod Database
57. The Global Lepidoptera Names Index
58. The Integrated Taxonomic Information System
59. The Lecythidaceae Pages
60. The World Spider Catalog
61. TicksBase
62. TIGR Reptile Database
63. Tineidae NHM: Global taxonomic database of Tineidae (Lepidoptera)
64. Trichomycetes & Fungi Associated with Arthropods
65. UNESCO-IOC Register of Marine Organisms
66. Universal Chalcidoidea Database
67. World Biodiversity Database (Euphausiacea)
68. World Checklist of Selected Plant Families
69. World database of Proseriata & Kalyptorhynchia
70. World Ophiuroidea Database
71. World Porifera Database
72. World Scarabaeidae Database
73. World Species List of Annonaceae
74. Zoological-Botanical Database (Vespoidea)
75. Zygomycetes
76. Portal Index
77. Coleção Entomológica Paulo Nogueira-Neto (CEPANN) - Interações
78. Coleção Entomológica Paulo Nogueira-Neto (CEPANN)
79. Instituto de Ciencias Naturales
80. Instituto de Ciencias Naturales

3.4 Software Tools and methodologies

The tools and structure needed to the project is already been coded, and the source code is located and will be managed in Google Code:

<http://code.google.com/p/iabincatalog/>

In the actual status, there are two tools in finalization progress:

- iabincatalog-direct-access

Allows direct access to data portals of IABIN and GBIF. The XML generated by these portals is downloaded directly (if we were to get the information datasets and data providers using the services already provided by these portals).

- iabincatalog-xml-parser

First steps for translating XML. Upon completing this project he will be able to get one and translate XML tags and any necessary information via a mapping file in CSV format.

All source code will be available in Google Code®, based in GBIF and IABIN PTN source code, that are full available at respective links below:

The code to the GBIF Portal can be accessed in <http://code.google.com/p/gbif-dataportal/>

The code to the IABIN PTN Portal can be accessed in <http://jabuticaba.pcs.usp.br/svn/laa/iabin/versao-20-03-2010/>.

4. Timeline

Below is the actual timeline of the project:

Iabin Catalog Timeline																								
Activity	Weeks																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Internal Meeting - Planning	■																							
Iabin Focal Point Contact	■			■				■				■					■							
Research		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Code										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Internal Meeting - Technologies										■														
Progress Report														■										
Test Client															■	■	■	■	■	■	■	■	■	■
Internal Meeting - Finalization																							■	
Final Report																								■

Current and future activities is being concentrate in the execution itself of extraction of data, map the FBDC BDP standard in the actual metadata and the execution itself of encapsulation of data with standards metadata.

There was a meeting with Prof. Mike Frame (USGS/NBII and University of Tennessee) for clarifications about the mapping of metadata which occurred on April 20, 2010.

It is also planned another meeting between the USP group with Iabin Catalog Coordinator Prof. Mike Frame for adjustments in of the services.

5. References

Corrêa, P. L. P.; Lopes, P. V.; Sellers. E.; Saraiva, A. M.; Harrison, C.; Adams, L. D. Pollinator Data Portal. User Manual and Training Guide Version 1.0 Beta. February, 2010.

DWC.DarwinCore Schema. Available in <<http://www.tdwg.org/activities/darwincore/>>. Accessed in: 20 feb. 2010.

GBIF.Global Biodiversity Information Facility. Available in: <<http://www.gbif.org/>>. Accessed in: 20 feb. 2010.

HEMRAJANI A. Desenvolvimento ágil em JAVA com Spring, Hibernate e Eclipse. São Paulo: Pearson Prentice Hall, 2006. 290p.

IABIN-PTN. IABIN – Polinators Thematic Network. Available in: <<http://pequi.pcs.usp.br/portal/>>. Accessed in 28 mar 2010.

<http://java.sun.com/> - Accessed in 28 mar 2010.

<http://www.eclipse.org/> - Accessed in 28 mar 2010.

<http://tomcat.apache.org/> - Accessed in 28 mar 2010.

<http://www.findjar.com/> - Accessed in 28 mar 2010.

<http://www.jarfinder.com/> - Accessed in 28 mar 2010.

Annex A

Create the project

Firstly it is necessary to create an Eclipse project:

1. "File" > "New" > "Other"
2. "Web" > "Dynamic Web Project" > "Next"
3. "Project name" > type "TestWS" (without quotes)
4. "Target runtime" > "New" > "Apache Tomcat v.6.0" > "Next"
5. "Tomcat installation directory" > "Browse" > choose Tomcat directory > "Finish"
6. "Dynamic web module version" > "2.5"
7. "Configuration" > "Modify" > mark "Axis2 Web Services" > "OK"
8. "Next" > "Next" > "Finish"

If the error message "Failed while installing Axis2 Web Services Core 1.1" shows up, it is required a manual installation of Axis2. Proceed as following:

1. Download the latest version of Axis2 in binary format
2. Copy Axis2 folder from the compressed file to any location
3. At Eclipse go to "Window" > "Preferences" > "Web Services" > "Axis2 Preferences"
4. At the tab "Axis2 Runtime" > "Browse" > choose Axis2 directory previously copied
5. (re)Create the project

Now the web service project is set. It is time to create the service.

Create the class

1. Right-click on "Java Resources: src" > "New" > "Class"
2. "Name" > type "ServiceClass" (without quotes) > "Finish"

Our service will print and return a simple "Hello World!" message. To do so the code should be like this:

```
public class ServiceClass
{
    public String serviceMethod()
    {
        System.out.println("Hello World! [print]");
        return "Hello World! [return]";
    }
}
```

3. Save

Create the web service

1. Right-click on the project name ("TestWS") > New > Other
2. Select "Web Services" > "Web Service"
3. "Service implementation" > "ServiceClass" (class previously created)
4. "Web service runtime" > "Apache Axis2" > "Finish"

Some warnings may appear, just click on "OK".

Use!

To use the web service it will probably necessary to install one or more JARs, like httpcore JAR. All additional JARs can be obtained at:

<http://www.findjar.com/index.x>

or

<http://www.jarfinder.com/>

To add the JARs to the project, use the "Build Path" options. Remember to export the JARs (both at the "Build Path" and "Java EE Module Dependencies" options).

To easily test the recently created web service it is possible to use Eclipse functionalities:

1. Right-click on "TestWS" (project name) > Run > Run As > Run on Server
2. Change the address to <http://localhost:8080/TestWS/axis2-web/>
3. Click on "Services"
4. Click on "ServiceClass"; the WSDL description will be shown
5. Copy the WSDL address (<http://localhost:8080/TestWS/services/ServiceClass?wsdl>)
6. Eclipse: Run > Launch the Web Services Explorer
7. Click on the icon "WSDL Page" (top right)
8. Click on "WSDL Main"
9. WSDL URL: paste the address previously copied > "Go"
10. At the "Navigator" select
"WSDL Main\<address>\ServiceClass\ServiceClassSoap11Binding\serviceMethod"
11. Click on "Go"; the message "Hello World! [print]" will be printed to Eclipse console, and the message "Hello World! [return]" can be seen at "Status".

Using the browser, open the following address:

<http://localhost:8080/TestWS/services/ServiceClass.ServiceClassHttpSoap11Endpoint/serviceMethod>

A XML document will be returned and the message "Hello World! [return]" will be encapsulated inside the <ns:return> tag.

Annex B

Specification of the fields of FGDC BDP metadata standard.

