University of Abomey-Calavi

=*=*=*=*=

Faculty of Agricultural Sciences

=*=*=*=*=

CAPACITY BULDING WORKSHOP ON BIODIVERSITY DATA MOBILIZATION

Abomey-Calavi, Maison d'Accueil Sainte Anouarite, October 12th to 13th 2016

WORKSHOP REPORT

Abomey-Calavi, October 12-13, 2016

Contents

Introduction

The Global Biodiversity Information Facility (GBIF) aims at mobilizing and publishing biodiversity data, in order to make them available to users (students, researchers, developers, etc.), and support decision making. Benin, member of the GBIF since 2004, is actively involved in the mobilization and publication of biodiversity data, and their uses in decision making. In order to increase the capacity of the various partners involved in data mobilization, GBIF Benin organized a capacity building workshop in the framework of "Biodiversity Information for Development" National and Regional projects. This workshop took place on Wednesday 12th and Thursday13th of October, 2016 at Maison d'Accueil Ste Anouarite of Abomey-Calavi.

A total of forty participants attended this workshop. The workshop agenda and list of participants are presented in Appendices 1 and 2 respectively.

The workshop was organized into six sessions:

- the opening ceremony;
- importance of biodiversity data digitization and related benefits;
- basic concepts of data cleaning tools;
- basic concepts of data publication;
- presentation of individual BID projects;
- closing ceremony.

This report summarized the key outcomes of the activities conducted across these sessions.

1 Opening Ceremony

1.1 Welcome speech of BID Project National Coordinator

The workshop started at 9 AM. In his address, Professor Jean Ganglo, Coordinator of BID National and Regional Projects, welcomed the participants and thanked them for their cooperation. Then he made a brief history of the creation of the Global Biodiversity Information Facility (GBIF) and presented its objectives. He expressed his profound gratitude to the Dean of the Faculty of Agricultural Sciences of the University of Abomey-Calavi, Professor Joseph Hounhouigan for his constant support to GBIF's activities in Benin. He also thanked Dr. Gaston Akouehou, Head of Delegation of GBIF Benin. Finally, he thanked the participants and wished a fruitful workshop.

1.2 Opening address of the representative of the Dean of the Faculty of Agricultural Sciences of the University of Abomey-Calavi

Speaking on behalf of the Dean of the Faculty of Agricultural Sciences, the Secretary General, Mrs. Aimée Hounga, welcomed participants. Then, he presented a brief history of the "Global Biodiversity Information Facility" (GBIF). She stated that GBIF is a global network of biodiversity information created in March 2001 to support open access to biodiversity data via the internet. GBIF counts currently 57 countries including 38 voting members and 39 international organizations. Benin is GBIF participant since December 2004 and has acquired voting status in October 2011. She presented the purpose of the workshop: increase the quality of data available for the needs of users and decision makers. Finally, she assured that the FSA will keep supporting GBIF Benin in its activities, and wished success to the workshop.

2 Overview of national and Regional BID projects

The coordinator of the national and regional BID projects, Professor Jean Ganglo, presented both projects. Funded by the European Union via the GBIF, the BID Regional project (Capacity building and biodiversity data mobilization for conservation, sustainable use, and decision making in Africa and Madagascar) gathers eight countries (Senegal, Ivory Coast, Guinea, Mali, Democratic Republic of Congo, Madagascar, Niger, and Benin), and aims at establishing and strengthening international cooperation for the mobilization of biodiversity data. In the framework of the regional project, a training workshop for all partners will take place in November 2016. The second objective is to increase the availability of biodiversity data, by strengthening the capacity of partners at national, regional and sub-regional levels, and by giving them financial and technical assistance. The third objective is to use biodiversity data to meet conservation priorities, by establishing distribution maps of threatened and invasive species. Furthermore, the effectiveness of protected areas in the conservation of biodiversity will be assessed. Results will be made available to decision makers through various means including the mass media and communication workshops.

The National BID project concerns strengthening biodiversity data mobilization capacities to address national priorities in health and food security in Benin. The project will increase the availability of biodiversity data on medicinal and agroforestry species, by providing technical

and financial assistance to partners. Finally, these data will be used to meet national priorities by developing distribution maps of target species, and conservation strategies as well.

This dual presentation by Professor Ganglo was followed by discussions with the participants. Discussions focused on the challenges faced by the project management team in mobilizing data, especially at regional level where contexts are not the same, with various levels of political stability. To this question, the Coordinator explained that he supports the national coordinators of consortium countries with patience and determination to gradually upgrade the capability of these countries in the GBIF network.

Participants were interested in terms and conditions of projects funding by GBIF. The coordinator explained that GBIF is financially supported by some donors for activities across different continents according to the priorities of donors. The selection follows a highly competitive process.

3 Importance of Biodiversity Information

Professor Jean Ganglo presented the importance of biodiversity information, and the benefits of publishing data. After giving a definition of biodiversity, the presenter listed the various services provided by biodiversity (environmental regulation, cultural and religious services, supporting and provisioning services, etc.). GBIF's mission is to facilitate free access via the Internet to biodiversity primary data to support scientific research, conservation of natural resources and sustainable development. GBIF has a database of more than 625 million of occurrence data from nearly 2 million of plant, animal and fungal species. Unfortunately, Africa has less than 4% of this data, and has a crucial lack of scientific expertise in that field. GBIF Benin aims at supporting sustainable development of Benin through the mobilization and use of biodiversity data.

Biodiversity informatics is a relatively new field of science that deals with large-scale mobilization of biodiversity data, data processing, analysis… in order to inform decision-making on the conservation and sustainable use of biodiversity services at all levels.

Then the presenter clarifies the concept of primary data which include: They are data with at least three attributes:

- The name of the organism observed whatever kingdom it belongs to
- Detailed description including geographic coordinates of where it is observed

- Full date in terms of day, month, and year of observation

These data must be encoded in GBIF spreadsheet, which is the publishing format data used by GBIF community.

When occurrence biodiversity data are displayed in geographic space using GIS, one can understand the environmental factors that control their distribution at present. Furthermore, under the different concentration pathways of climate change one can model the distribution of targeted species and derive possible strategies for biodiversity conservation and therefore, guide decision makers in natural resource management.

Prof. Ganglo also presented the state of biodiversity data on the portal of GBIF Benin. This portal contains 207 occurrence datasets with over than 240,000 occurrence data published by 22 countries including Benin. GBIF Benin has published more than 75% of those data. The data published so far are still insufficient and even more concentrated in the south of the country.

4 The types of publishable data on GBIF website

First, Prof. Ganglo explained again the concept of primary biodiversity data, reminding the essential attributes of primary data. Those data can be found in museum and herbaria on the labels of specimens or in laboratories where species are sometimes preserved in formalin. Fossils, skulls etc. can also be described and linked to occurrence data to be published on the website of GBIF. The insectariums also contain numerous and rich information that can be published by scientists in the field. There are also primary data from the literature (scientific publications, graduation dissertations, research protocol, etc.). Observation, radar or satellite, and other supports are also sources of primary data. Professor Ganglo explained the concept of sample data. Those are data collected on the basis of a defined scientific protocol such as forest, floristic, animal inventory protocols. He also clarified the notion of abundance data.

He explained that secondary data are syntheses obtained on the basis of primary data. These include e.g. species distribution areas, checklists, descriptions of natural history of the species.

Metadata are structured datasets descriptions. They provide essential details such as geographic and taxonomic scope of dataset to be published, the methods of collection and observation, contacts, and requirements of the quote. They help give context to data and allow users to be informed of the dataset.

Finally, some taxonomic references were presented to participants. Issues such as security of data not yet published by the holders were addressed. To this end, participants were reassured that publications by GBIF do not affect the integrity of scientific data at the level of suppliers. Moreover, when it comes to sensitive data, for example data on species threatened, the geographical coordinates will include a level of uncertainty, so that the species cannot easily be located.

5 Principles of data management in the data publication process

In this communication, Mr. Jaures Gbetoho stated that data quality is ensured by the suppliers; and it is better to prevent errors than trying to correct them later. Also, using the standards, provides as much detail as possible on data. Data quality depends on the use pursued by the users. However, to ensure that the information provided is of quality (the minimum required), contributors are expected to provide the key attributes without errors. Mr. Gbetoho described potential sources of errors in data publishing process, and invited participants to avoid mistakes. Further clarifications were given during exchanges with participants.

6 Practical demonstration on digitization: tools and feedback

Professor Jean Ganglo started this presentation by defining the concept of "Darwin core archive". The "Darwin core archive" is a data presentation standard of biodiversity informatics that uses the specific terms of biodiversity to produce a single set of data on the site of GBIF. Once data are published, the "Darwing core archive" is automatically generated. It is a set of text file with a simple descriptor on how files are organized. Then GBIF spreadsheet was presented, with practical examples and details instructions to participant to help them successfully filling the data. Data licensing issues were also explained by the presenter.

7 Cleaning and formatting data

7.1 Data management tools

In this communication delivered by Mr. Jaures Gbetoho, the various data cleaning tools were presented to participants. Among other tools, there are GBIF's ECAT name Parser (www.tools.gbif.org/nameparser), Iplantcollaborative (www.iplantcollaborative.org), ITIS

(www.itis.org), the coordinates converter (www.splink.cria.org.br/conversor) etc. Practical applications were made using different tools.

Iplantcollaborative can be used to determine an error in scientific name of a plant species. It is also possible to go to www.eol.org that makes suggestions of scientific names.

In the ITIS database, it is possible to make a comparison of scientific names regardless the kingdom. ITIS also offers a list of the scientific names: valid, outdated names, unqualified names. From there, EOL can search the rejected or outdated names one by one. The procedures of data importation from the various websites were shown.

7.2 Practical session on data cleaning

Prof. Jean Ganglo explained that data cleaning is a process to determine inaccurate, incomplete, or undesirable data and to improve their quality through suspicious data corrections. Data cleaning is essential, because it allows improving the quality to make them suitable for use by users by reducing errors in data and improving their documentation and presentation.

Mistakes can have various origins: for example, data without geographic coordinates. To clean data, tools are developed to assist institutions to add geographical information (e.g.: Geolocate can help determine the geographical coordinates of the place of collection provide that the collectors fully describe the locations of data collection).

To prevent errors in the accurate determination of the species observed or recorded, it is recommended to consult specialists or taxonomic documents. One should also be careful when encoding data to prevent input errors. Scientific names are the major key to ensure data quality.

This introduction helped to address some practical steps of data verification and cleaning by GBIF Benin:

- Once the dataset is received, at first a conversion in csv format is done; then QGIS is used to make a quick projection on the map of the country (Benin). The data displayed out of Benin are automatically eliminated. The screening procedure was shown to participants.

- In a next step, a dynamic cross-tab in Excel software is used to verify the existence of duplicated data. Participants were urged to make these kinds of cleaning before submitting their datasets for publiscation.
- Demonstrations were done on the use of Google refine and QGIS. For example, how to use them to identify and correct errors or to generate taxonomic ranks of the species, or to rename the different columns of the spreadsheet based on data that are presented. At this point, several interactions enabled participants to have clarification on the use of different tools.

8 Checklists and use of Geolocate

In this presentation, Mr. Gbetoho defined the concept of checklist and how to publish them. The mandatory fields in checklists are: Taxon ID, Scientific Name, Taxon rank. The other fields can be generated automatically using specific tools. A checklist format (templates) was presented with instructions on filling.

In the second part of this communication, Mr. Gbetoho trained participants on georeferencing data using Geolocate. This application enables to reference the geographical coordinates of data collection sites. The input is the name of the locality and the software presents the geographical coordinates as output. If the data are numerous, it is possible to use a CSV file to upload them on Geolocate. Nevertheless, it was agreed that researchers should make the effort to take the coordinates in the field, and that the use of Geolocate should be a last resort, to limit potential bias.

9 Presentations of BID Individual Project

Three individual BID projects were presented to participants. The first presentation was that of Charles M Gangnibo of the Faculty of Agricultural Sciences whose project deals with birds. He presented the context of this project. He then presented the importance of birds in human life. The expected results of the project are among others: (i) sensitizing data holders and decision makers on the importance of biodiversity, (ii) strengthening the capacity partners on data mobilization; (iii) Development of ecological niches of two species most threatened birds...

The second presentation was done by Mr. Isidore AMAHOWE from the Forest Service, whose project deals with *Afzelia africana*. This project will increase the quantity of available

data, predict the evolution of this species and utilize the outputs to enlighten decision makers. Project activities also encompass the strengthening of the capacity of actors.

The third project, on conservation of gregarious forest plant species, was presented by Mr. Romaric LOKOSSOU from the Centre for Studies, Research and Training in Forestry (CERF). The project involves the setting up of a seed bank of native species used for reforestation; these species are of interest because there are resilient and adapted to the current context of climate change. Activities on native species seed bank are already conducted in southern Benin, and the project will extend them to the center of Benin. Besides data publication on GBIF site, the project will build the capacity of partners in mobilizing data.

Participants unanimously appreciated the individual BID projects. Clarification questions were asked, and suggestions for improvement were made. These suggested improvements are related to the targeted species by the projects as well as the methodology. Other participants expressed their expectation to benefit from support of GBIF Benin in order to apply to this kind of project in the future. The Node Manager of GBIF Benin urged the individual BID projects holders, to invite him to their various activities.

10 Closing ceremony

Speeches were delivered by the head of delegation and the Node Manager of GBIF Benin in the closing ceremony.

In his address, Dr Gaston Akouehou, Head of Delegation, presents his sincere thanks to the participants for their availability. He noted that participants were actively involved in the various sessions of the training, hence their enthusiasm for formatting, and publishing data in the GBIF system. He also thanked the trainers, especially Prof. Jean GANGLO.

In the closing remarks, Prof. Jean Ganglo, Node Manager, thanked the audience for its enthusiasm and active participation. Likewise, he thanked the individual project leaders, before reassuring the Head of Delegation of GBIF Benin that data will be published according to the commitments made by the various participants. Finally, he urged participants to publish data for biodiversity conservation, and asserted the readiness of his team to support the stakeholders in need.

Conclusion

The objectives of the capacity building workshop on biodiversity data mobilization, formatting, cleaning, and publishing within the framework of national and regional BID projects were achieved. Participants were informed about the importance of Biodiversity Informatics in species conservation. They were also trained on the use of different formatting, cleaning and publication of biodiversity data tools. More than 240,000 occurrence data have been published on Benin and more than 75% of this by GBIF Benin. Several participants expressed their desire to publish data. Individual BID projects in Benin were also presented with an overview of their activities.

Appendices

Appendix 1: Programs of the workshop

Time	Activities	Responsible
	Day 1: Wednesday, October 12, 2016	
Session 1	**Opening**	
08:30–9:00	Welcome and installation of participants	M. Smith Dossou
09:00–09:30	- Welcome by the BID National Project Coordinator - Opening address by the Dean of the FSA-UAC	Prof. Jean C. Ganglo Prof. Joseph Hounhouigan
09:30–09:45	Introduction of participants, including the areas of work, and interests in biodiversity.	Facilitator
09:45–10:00	Presentation of the objectives of the workshop	Facilitator
10:00–10:30	Presentation of the project Prof. National BID	Prof. Jean C. Ganglo
10:30–11:00	**Family photo + Coffee Break**	**Organizing Committee**
Session 2	Digitization of biodiversity data	
11:00–12:15	Importance of information biodiversity: Why to publish biodiversity data? Advantages of data publication	Prof. Jean C. Ganglo
12:15–13:00	Types of biodiversity data considered in the context of GBIF activities	Prof. Jean C. Ganglo
13:00–14:00	**Lunch Break**	**Organizing Committee**
14:00–14:30	Data management principles in the digitization process	M. Jaures Gbetoho
14:30–15:30	Practical demonstration on digitization: tools and feedback	Prof. Jean C. Ganglo
Session 3	**Cleaning, formatting, and transformation of data**	
15:30–16:30	Basics of data cleaning	Prof. Jean C. Ganglo
	Day 2: Thursday, October 13, 2016	
08:30–09:30	Data management tools	M. Jaures Gbetoho
09:30–10:30	Using OpenRefine in data cleaning	Prof. Jean C. Ganglo
Session 4	**Publication data**	**Organizing Committee**
10:30–11:00	**Coffee Break**	**Organizing Committee**
11:00–12:30	Basic Concepts of data publishing	M. Jaures Gbetoho
12:30–13:30	**Lunch Break**	**Organizing Committee**
13:30–14:15	Constraints to data publication and possible solutions	Prof. Jean C. Ganglo
Session 5	**Specificities of BID individual Project**	**Organizing Committee**

Time	Activities	Responsible
14:15-15:00	BID Individual projects: objectives and expected outcomes	M. Charles Gangnibo Isidore Amahowe Romaric Lokossou
15:00-16:15	Publication of data on bird species: importance/uses, types of information, data formatting and some helpful tips	M. Charles Gangnibo
Session 6	**Closing**	
16:15–16:30	Evaluation of the training	M. Smith Dossou
16:30–17:00	Speech of the Coordinator BID national and regional Projects Speech by the Dean of the FSA-UAC	Prof. Jean C. Ganglo Prof. Joseph Hounhouigan

Appendix 2: List of participants

N°	Name and surnames	Organization	Origin	Contact	
				Phone number	E-mail
1	ADOHOU Sylvana	Herbier national	Abomey-Calavi	67760700	sadohou@yahoo.fr
2	AGBANI O. Pierre	JBZ-EA/UAC	Abomey-Calavi	97608987 94321655	pierreagbani@gmail.com
3	AGODJI Michel	FSA	Abomey-Calavi	96303615	m.agodji1@gmail.com
4	AGOKOLI Mélodie	EPAC/UAC	Cotonou	61321228	neagokoli@gmail.com
5	AHOTON Ganse Yvette	FSA	Abomey-Calavi		
6	AKOUEHOU S. Gaston	CERF	Cotonou	97367678	gastonakouehou@gmail.com
7	AMAHOWE O. Isidore	DGEFC	Cotonou	97326013	ogoudje.amahowe@gmail.com ogoudje_amahowe@yahoo.fr
8	AOUDJI Augustin	FSA	Abomey-Calavi	97481280	augustin.aoudji@gmail.com
9	ASSONGBA Faustin	FAST/Dassa/UPA	Dassa-Zoumè	97602632 95673397	yedjanlognon@yahoo.fr yedjanlognon@gmail.com
10	ATCHADE Nicolas	DGEFC	Abomey	97576703	atchade_s@yahoo.fr
11	BIO Anselme	FSA/UAC	Abomey-Calavi	95231499	bioanselme@gmail.com
12	CODJIA A. Sylvie F.	LSF/FSA	Cotonou	96144320	sf.codjia@gmail.com
13	DJOTAN A. Kévin Géoffroy	LSF/UAC	Cotonou	95475392	geoffroydjotan@yahoo.fr
14	DOSSA GBO C. Florent	LSF/UAC	Abomey-Calavi	97161059	florentcof@yahoo.fr
15	DOSSOU Smith	FSA/UAC	FSA	96486194	smithdossou@yahoo.fr
16	GANGLO C. Jean	LSF/FSA	Abomey-Calavi	66363770	ganglocj@gmail.com

N°	Name and surnames	Organization	Origin	Contact	
				Phone number	E-mail
17	GANGNIBO N. Charles	LSF/UAC	Abomey-Calavi	97398973	charlaco2000@yahoo.fr
18	GBAGUIDI Roméo	CERF	Cotonou	95242460	gbaguidiromeo@gmail.com
19	GBETOHO Alain Jaures	LSF/FSA	Cotonou	96833454	gljaures@gmail.com
20	HOUENON G. Hurgues Aristide	Herbier National du Bénin	Abomey-Calavi	66096163	houenonhurgues@gmail.com
21	HOUNGA Aimée	FSA	Abomey-Calavi	66844921	
22	HOUNHOUIGAN D. Joseph	FSA	Abomey-Calavi		
23	HOUNKPONOU Fawaz Deen	Nature Tropicale ONG	Cotonou	96342046	fawazdeen.hounkponou@yahoo.fr
24	HOUNSA Michaël Archange F.	LSF/FSA	Cotonou	97847789	houmichey@gmail.com
25	HOUSSOU Doris	PAPDFGC	Cotonou	97122232	risdodo@gmail.com
26	KIKI Paterne	LSF/UAC	Cotonou	97869453	kikipaternep@gmail.com
27	KINGBO Armand	LSF/FSA	Abomey-Calavi	96269768	armandkingbo@gmail.com
28	KOURA Kourouma	LSF/FSA	ADJAHONME (KLOUEKANME)	95403969 96716130	kourakourouma@yahoo.fr kourakourouma@gmail.com
29	LEGBA S. Ingride	CERF/ SEFTBPPFNL	Abomey-Calavi	96084672	ingride.legba@gmail.com
30	LOKOSSOU Romaric	CERF	Cotonou	95795173	lokossouromaric@yahoo.fr
31	MAHOUSSI Tranquillin	FSA	Abomey-Calavi	96570737	
32	MASSOEBI Epiphane Frederic	FSA	Abomey-Calavi	96992949	

N°	Name and surnames	Organization	Origin	Contact	
				Phone number	E-mail
33	OROU MATILO Augustin	CERF	Cotonou	97842182	matilorou@yahoo.fr
34	SOGBOSSI Erick	LaCarto/UAC	Abomey-Calavi	96114887/95538562	ericksogbossi@gmail.com
35	SOSSOU Danielle	Nature Tropicale ONG	Cotonou	96122135	danielle.sossou@yahoo.fr
36	TOGNI Yolande	LSF/FSA	Abomey-Calavi	66248666	ytogni@yahoo.fr
37	TOKO Ismaïla	Lacarto/DGAT/ UAC	Abomey-Calavi	97772872	ismael_toko@yahoo.fr
38	TRAORE M. Ibrahim	LSF/UAC	Cotonou	96280598	traoremoussilimii@gmail.com
39	VINYOR B. Hermann	CERF	Cotonou	97176581	boris731@live.com
40	YEHOUENOU TESSI Donal R.	LSF/UAC	Cotonou	97696308	tdonald3@yahoo.fr