

BID Capacity Enhancement Workshop: Biodiversity Data Mobilization

Information Brochure

Version 1, July 2017



1. Introduction

In 2015, the GBIF Secretariat officially started **Biodiversity Information for Development (BID)**, a multi-year, €3.9 million programme funded by the European Union with the aim of increasing the amount of biodiversity information available for use in scientific research and policymaking in the 'ACP' nations of sub-Saharan Africa, the Caribbean and the Pacific. One of the main implementation mechanisms of the programme is a set of public calls to support projects directly contributing to the objectives of the programme. More information about the programme is available at <http://www.gbif.org/programme/bid>.

This **workshop** aims to contribute to enhancing the capacity of the project teams to plan and implement biodiversity data digitization efforts effectively and according to GBIF standards. It will have a strong focus on the technical aspects of the data mobilization—in particular everything related to the data lifecycle: digitization, management and online publishing. The social aspects of the process will also be considered.

The event will have online and onsite components and both will have a strong practical approach (including a significant component of group work). Participants will have the possibility to evaluate their learning and receive an official certification in the form of a Mozilla Open Badge. The official language of the workshop will be English.

Only individuals nominated by a BID project representative and mentors selected by the GBIF Secretariat can subscribe to this workshop. Please address any questions you may have to PID@gbif.org.

This brochure includes a **draft agenda** for the event. Time is expressed in UTC/GMT +2 (CAT).



The BID programme is funded by the [European Union](#) and co-funded by [GBIF](#).

2. Prerequisites for the Workshop

To make best use of the activities around this workshop, the participants should possess the following skills and knowledge:

- Basic skills in computer and internet use, and in particular in the use of spreadsheets, databases, and tools for geographical data representation (e.g. google maps, GIS software).
- Basic knowledge about geography and biodiversity informatics: geography and mapping concepts, basic taxonomy and nomenclature rules, data(set) types, basic knowledge about GBIF and other relevant initiatives working in biodiversity informatics and how they operate.
- Basic understanding of the data publishing model of GBIF, the tools available and how to use them for publishing.
- A good command of English. Course activities will be conducted in English. Translated materials will be available in French and Spanish.
- Experience as a trainer, for example, leading seminars, workshops or other training activities in topics related to biodiversity informatics. Participants need to be able to adapt generic training materials to specific contexts and languages maintaining their instructional value.

3. Pre-Workshop Activities: 10-21 July 2017

Participants should reserve around **10 hours** to complete online activities in the two weeks leading to the onsite workshop.

Module 0: Introduction to the course and platform

The very first module of the online preparatory activities will help participants to familiarize themselves with the online platforms that will be used during the workshop and provide means to self-assess the participants' current knowledge.

Module I: GBIF and other BDI initiatives

The landscape of biodiversity informatics initiatives in the 21st century is complex, with many organizations working on different aspects of it. This module reviews the most relevant ones for the global and Pacific context.

Module II: Basics of Biodiversity Informatics

This module aims at helping the student to consolidate some basic concepts used in biodiversity informatics, and to get familiar with the basic characteristics of biodiversity datasets.

Module III: Workshop sessions preliminary work

The trainers will present the students with some additional preparatory materials and activities on the topics that will be studied during the main onsite workshop.

4. Onsite Workshop: 24-27 July 2017

Draft schedule

This draft schedule is provided to allow participants a sense of the contents planned for the event. It is likely that there will be changes before the final contents of the workshop are fixed.

Monday, 24 July 2017
<p>Module IV: INTRO TO THE ONSITE EVENT <i>This module will enable the participants to get acquainted with each other, with the platforms that we will use during the workshop and with the basic theoretical concepts that will be used across modules.</i></p>
08:30 – Registration
<p>09:00 - Session 1: Welcome and introduction to the workshop <i>This session will include a combination of an explanation of all the practical information that we will need during the course; a review of the pre-course activities; and participant introductions.</i></p>
<p>09:45 - Session 2: Foundations for the course <i>During this session we will have the opportunity to review the contents planned for the course and discuss key concepts that we will be using in all modules, in particular about biodiversity digital data management. It will be a theoretical interactive session where students will have the opportunity to contribute with their answers and experiences.</i></p>
10:45 - Coffee/tea break
<p>Module V: PLANNING A DIGITIZATION PROJECT <i>This module will help up to understand the different stages that should be taken into account when planning a digitization project, and how to adapt them to our specific situation.</i></p>
<p>11:00 - Session 3: Digitization project planning: how to decide what to do and when to do it? <i>This session will include a theoretical presentation and discussion around the key project planning stages for successfully implementing a digitisation project and how to create a viable workflow. We will cover topics such as: What things should be formally structured at the beginning?; What can be left wait and see?; Who needs to know what and when?</i></p>
12:30 - Lunch break
<p>13:30 - Session 4: Digitization planning practical exercise: Will it work? <i>Practical role-play group session. Using the methods discussed in Session 3 groups will be tasked to create an actual executable project plan/work-flow. Each member of the group will be assigned a stakeholder role and members will work together to create a workflow for a BID-like small digitisation project. These will then be presented to the whole group for discussion of commonalities and differences.</i></p>
15:00 - Coffee/tea break
15:30 - Session 5 : Digitization planning practical exercise: Suggested solution and discussion
<p>Module VI: BIODIVERSITY DATA DIGITIZATION <i>This theoretical/practical module will help identifying the type of data each participant is working with, and how to best digitize relevant information using best practices and existing tools and techniques.</i></p>
<p>16:00 - Session 6: Introduction to Darwin Core <i>During this session, participants will be introduced to the Darwin Core standard and it's components which be used throughout the remainder of the course.</i></p>
17:00 - End of the day

Tuesday, 25 July 2017

Module VI: BIODIVERSITY DATA DIGITIZATION CONTINUED

This theoretical/practical module will help identifying the type of data each participant is working with, and how to best digitize relevant information using best practices and existing tools and techniques.

09:00 - [Session 7](#): Biodiversity Data Types: With which kind of data are we working?

Practical session to identify which kind of data people are dealing with and useful information to digitize in priority. Includes group discussions and roleplays to familiarize with different kinds of data (collections, observations, sample-based, taxonomic, possibly pictures) and how to best share the related information.

10:30 - Coffee/tea break

11:00 - [Session 8](#): Principles of data management in digitization

Theoretical and practical session focused on best practices and data quality principles in the context of data digitization. Includes a short presentation on data quality and coherence (especially on subjects like georeferencing, dates, names and taxa cross-checking), followed by a group discussion on good practices and a practical exercise using simple spreadsheets.

12:30 - Lunch break

Module VII: DATA CURATION, FORMATTING & TRANSFORMATION

This mainly practical module will help with the basic tools and concepts used for data validation and cleaning, and how data can be converted into Darwin Core.

13:30 - [Session 9](#): Basic concepts of data cleaning

The first part of this session will enable the participants to get familiarized with the main concepts of data cleaning, related tools, and best practices in the curation process. The second part will consist in a theoretical/practical exercise with examples of technical and consistency validation checks.

15:00 - Coffee/tea break

15:30 - [Session 10](#): Use of OpenRefine

Practical/theoretical session focused on OpenRefine as an easy tool to improve the quality of biodiversity datasets by using the default features, existing web services and regular expressions.

17:00 - End of the day

Wednesday, 26 July 2017

Module VII: DATA CURATION, FORMATTING & TRANSFORMATION CONTINUED

This mainly practical module will help with the basic tools and concepts used for data validation and cleaning, and how data can be converted into Darwin Core.

09:00 - [Session 11](#): Other data management tools

Practical/theoretical session focused on easy to use tools to validate and clean datasets in three main categories: nomenclatural, format, and geographical.

10:30 - Coffee/tea break

Module VIII: DATA PUBLISHING USING IPT

This module focuses on the process of making biodiversity data freely available online, also known as publishing the data online. We will use the main tool that GBIF makes available for this purpose: the Integrated Publishing Toolkit (IPT).

11:00 - [Session 12a](#): Data publishing concepts and introduction to the IPT

Presentation and discussion on subjects such as licenses, metadata, mandatory fields, hosting of data sets of different institutions on the same IPT installation, etc. Presentation and demonstration covering the basics of publishing using the IPT tool (principles, user interface, workflow, metadata, dataset visibility, etc). Additional demonstration and discussion covering IPT features and publication of a complex, sample-based dataset where emphasis will be put on the use of extensions and the core/extension relationship.

12:30 - Lunch break

13:30 - [Session 12b](#): Data publishing concepts and introduction to the IPT continued

15:00 - Coffee/tea break

15:30 - [Session 12c](#): Data publishing concepts and introduction to the IPT continued

17:00 - End of the day

Thursday, 27 July 2017

Module IX: FREE PRACTICE

This modules will feature free practice sessions where the students will have the opportunity to practice using examples provided by the organizers, or using their own examples and datasets. A focus will be suggested for each session but the participants will have the option to choose where to concentrate their efforts based on their skills and needs.

09:00 - [Session 13](#): Practical session

The students will perform hands-on exercises comprised of techniques learned during the course on data digitization, data management, and data publishing.

10:30 - Coffee/tea break

11:00 - [Session 14](#): Practical session continued

Module X: EVALUATION

Last module of the onsite workshop where we will review the contents covered, evaluate it, and discuss follow-ups.

12:20 - [Session 15](#): Summary of the course and evaluation of the onsite event.

During this session we will review the contents covered by the onsite workshop, and the participants will have the opportunity to provide feedback via an evaluation session. We will also discuss the follow-up activities after the onsite workshop, in particular everything connected with the assessment of the participants and the certification process.

12:30 - Lunch break

13:30 - End of the training event/beginning of regional meeting

5. Post-Workshop Activities: 07-25 August 2017

Participants should reserve around **10 additional hours** to complete online activities in the three weeks following the onsite workshop.

Module XI: Data access via GBIF.org

As a complement to the contents seen to date, the last content module will focus on the use of GBIF.org to access data published through the network. An important component of this module will concentrate in the feedback mechanisms that GBIF.org offers to data publishers

Certification-related activities

This final component will include the submission of final assignments by the participants in the workshop, their assessment by the trainers and the delivery of the certification and badges to those who comply with the requirements defined for the workshop.