

Capacity Enhancement Training: Biodiversity Data Mobilization

Training program - Minsk 18-22 February 2019

Last updated 2019-02-15



1. Introduction

This [training course](#) aims to enhance the capacity to plan and implement biodiversity data digitization and mobilization efforts effectively and according to GBIF standards. It will have a strong focus on the technical aspects of data mobilization – in particular everything related to the lifecycle: planning, capture, quality and publishing in order to increase the amount, richness and quality of the data published through the GBIF network and made available for use in scientific research and policymaking. The social aspects of the process will also be considered. The training event will have online and onsite components and both will have a strong practical approach including a significant component of group work.

The course is organised by the National Academy of Sciences of Belarus, Natural History museum at the University of Oslo, GBIF-Norway, ForBio, and Global Biodiversity Information Facility (GBIF) Secretariat in Minsk, Belarus, February 18-22 2019. The course is part of the [BioDATA](#) project (2018-2021) funded by the *Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education (DIKU)*.

Course topics

Foundations of biodiversity informatics

- Learn key concepts of biodiversity informatics, particular to biodiversity digital data management.
- Introduction to the Darwin Core Standard and its components.

Planning digitization and data mobilization

- Learn to understand the different stages for planning a digitization project and how to adapt them to a specific project.

Data capture

- Learn to identify the type of data and how to best digitize relevant information using best practices and existing tools and techniques.

Data management

- Learn the basic tools and concepts used for data validation and cleaning.

Data publishing

- Learn the process of making biodiversity data freely available online, also known as data publishing, utilizing GBIF's Integrated Publishing Toolkit (IPT).

Learning outcomes

- Develop a data mobilization strategy customized to a given institutional framework.
- Apply a model to build a data mobilization strategy and the associated digitization and data capture protocols.
- Evaluate a data mobilization strategy to identify potential gaps, inefficiencies and pitfalls.
- Apply a digitization protocol to produce digital biodiversity data from analogue sources.
- Use software tools designed to facilitate biodiversity data digitization.
- Apply data cleaning protocols to evaluate and increase the fitness for use of a biodiversity dataset.
- Use software tools to evaluate the fitness-for-use of a biodiversity dataset.
- Use software tools designed for (biodiversity) data cleaning.
- Apply a biodiversity data publishing mechanism.
- Define the publishable data types and subtypes (if any) for a biodiversity dataset.
- Use the GBIF IPT to publish biodiversity datasets using the appropriate extensions.
- Capacitate others in the digitization, management and publishing of biodiversity data.

Course trainers

Teachers

- Dag Endresen - GBIF-Norway, University of Oslo
- Laura Russell - GBIF Secretariat
- Dmitry Schigel - GBIF Secretariat

Mentors

- Rukaya Johaadien - GBIF-Norway, University of Oslo
- Piotr Tykarski - GBIF-Poland, University of Warsaw
- Maxim Shashkov - GBIF.ru team, Institute of Physicochemical and Biological Problems in Soil Science
- Natalya Ivanova - GBIF.ru team, Institute of Mathematical Problems of Biology RAS

Software tools

- e-Learning platform at: <http://elearning.gbif.es/courses/GBIFDMBIODATA2019/>
- IPT - Integrated data Publishing Toolkit at: <https://data.gbif.no/ipt-test/>

Language

English, with limited support in Russian.

Assignment and credits

Upon successful evaluation and completion of the course, participants have the opportunity to receive an official certification in the form of a Mozilla Open Badge.

2. Prerequisites

To make best use of the activities around this course, 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, and basic knowledge about GBIF and other relevant initiatives working in biodiversity informatics.
- A good command of English. Course activities will be conducted in English. Russian-speaking trainers and mentors will be available and translated materials will be available.
- Willingness to disseminate the knowledge learned in the course with partners and collaborators in your projects by adapting the biodiversity data mobilization training materials to specific contexts and languages maintaining their instructional value.

3. Pre-course activities: 5 - 15 February 2019

Participants should reserve around 10 hours to complete online activities in the two weeks leading to the onsite course. Access information and login to the [e-learning platform](#) and the [course IPT](#) will be provided to all participants. [Preparatory materials are available here.](#)

Module I: Introduction to e-learning

The first module of the online preparatory activities will help participants to familiarize themselves with the online platforms that will be used during the course. ([e-Learning platform](#))

[Session 01a](#): Introduction to the course, overview, and practical information (*Preparatory reading*)

[Session 01b](#): How to use the e-learning platform (*Preparatory activity*)

[Session 01c](#): Introduction to learning online (*Preparatory activity*)

[Session 01d](#): GBIF and other initiatives (*Preparatory reading*)

Module II: Foundation activities

This module introduces students to biodiversity informatics and the Darwin Core standard.

[Session 02a](#): Biodiversity informatics (*Preparatory reading*)

[Session 02b](#): Data quality (*Preparatory reading*)

[Session 02c](#): Digitisation workflows (*Preparatory reading*)

[Session 02d](#): Software tools (*Preparatory activity*)

Module III: Preliminary work

This module provides additional preparatory materials, references, and activities on the topics that will be covered during the main onsite workshop.

[Session 03a](#): Software installation.

Please try to come prepared to the course with some softwares preinstalled ([OpenRefine](#), [Java JRE](#), [IPT](#)).

We will of course assist you with software installation during the workshop.

*Please remember to **bring your laptop** and ensure that you have **administrator rights** to install the necessary software.*

[Session 03b](#): Pre-course questionnaire.

Please fill in the [pre-course questionnaire](#) before 15 February 2019.

4. Onsite course: 18 - 22 February 2019

Day 1 - Monday 18 February 2019

Module IV: INTRO TO THE ONSITE EVENT

This module enables the participants to get acquainted with each other, with GBIF, and with the platforms that we will use during the course. Basic theoretical concepts that will be used across modules.

08:30 - Registration (30 minutes)

09:00 - [Session 04a](#): Welcome and introduction to the course (60 minutes)

This session will include a welcome from our hosts that will lead into an explanation of all the practical information that we will need during the course; a review of the pre-course activities; and participant introductions. (Presentation and introductions)

10:00 - [Session 04b](#): Why share data? (60 minutes)

This session will introduce participants as to why we should share data. (Presentation and discussion)

11:00 - Coffee/tea break (30 minutes)

Module V: BIODIVERSITY INFORMATICS FOUNDATIONS

This module provides introductions to key concepts and terminology that will be used across the workshop modules.

11:30 - [Session 05a](#): Foundations: Terminology and standards (30 minutes)

During this session we will review discuss key concepts that we will be using in all modules, in particular about biodiversity digital data management and the concepts of standards. We will review the importance of documentation and will focus on metadata and data mapping in preparation for data publishing. (Presentation)

12:00 - [Session 05b](#): Foundations: Darwin Core (20 minutes)

During this session, participants will be introduced to the Darwin Core standard and its components, which will be used throughout the remainder of the course. (Presentation)

12:20 - [Session 05c](#): Foundations: Data quality (20 minutes)

During this session, participants will be introduced to the Darwin Core standard and its components, which will be used throughout the remainder of the course. (Presentation)

12:40 - [Session 05d](#): Foundations: Documentation (20 minutes)

During this session, participants will be introduced to the Darwin Core standard and its components, which will be used throughout the remainder of the course. (Presentation)

13:00 - Lunch break (60 minutes)

Module VI: PLANNING

This module introduces participants to methods for planning a biodiversity data mobilization project.

14:00 - [Session 06a](#): Planning: workflows and documentation (45 minutes)

This session describes the key project planning stages for successfully implementing a mobilization project and how to create a viable workflow. Topics include: What things should be formally structured at the beginning? What can be left to wait and see? Who needs to know what and when? Use case 1 will be introduced in this session. (Presentation)

15:00 - [Session 06b](#): Planning: Identifying key stakeholders and roles (75 minutes)

Practical group session. Using the methods discussed in the previous session, groups will be tasked to create an idealised project plan/workflow based on Use Case 1. Mentors may, if it is appropriate, visualize the groups own project. (Exercise)

15:30 - Coffee/tea break (30 minutes)

16:00 - [Session 06c](#): Planning: Who does what and when - tasks and stages (30 minutes)

Practical group session. Using the methods discussed in the previous session, groups will be tasked to create an actual executable project plan/workflow. These will then be presented to the whole group in the next session for the discussion of commonalities and differences. (Exercise)

16:30 - [Session 06d](#): Planning: suggested solution (60 minutes)

Groups will report back to the classroom a summary of their exercise outcomes. Presenter will go through the suggested solution. (Discussion)

17:30 - End of the day

Day 2 - Tuesday 19 February 2019

Module VII: BIODIVERSITY DATA CAPTURE

This module introduces participants to biodiversity data types and provide best practices for data capture.

09:00 - [Session 07a](#): Biodiversity data origins and types (60 minutes)

Practical session to identify what kinds of data people are dealing with and useful information to prioritize data capture. Includes group discussions to familiarize with different kinds of data (collections, observations, sample-based, taxonomic, and possibly media) and how to best share the related information. (Presentation and discussion)

10:00 - [Session 07b](#): Data capture, processing and quality (60 minutes)

Practical session focused data capture utilizing the previously introduced use cases and Biodiversity Challenge. (Presentation and discussion)

11:00 - Coffee/tea break (30 minutes)

11:30 - [Session 07c](#): Data capture, processing and quality (30 + 60 minutes)

Practical session focused data capture utilizing the previously introduced use cases. (Presentation and exercise)

13:00 - Lunch break (60 minutes)

Module VIII: DATA CLEANING AND STANDARDIZATION

This module will introduce participants to the basic tools and concepts used for data validation, cleaning, and how data can be standardized for publishing as Darwin Core.

14:00 - [Session 08a](#): Basic concepts of data cleaning (30 + 60 minutes)

The first part of this session will familiarize participants with the main concepts, related tools, and best practices for data cleaning and standardization. Followed by a practical exercise with examples of technical and consistency validation checks. (Presentation and exercise)

15:30 - Coffee/tea break (30 minutes)

16:00 - [Session 08b](#): Data cleaning using other tools (30 + 60 minutes)

This session focuses on tools used to validate and clean datasets in three main categories: nomenclatural, format, and geographical. Followed by a practical exercise. (Presentation and exercise)

17:30 - End of the day

Day 3 - Wednesday 20 February 2019

Module VIII: DATA CLEANING AND STANDARDIZATION (continued)

This module will introduce participants to the basic tools and concepts used for data validation, cleaning, and how data can be standardized for publishing as Darwin Core.

09:00 - [Session 08c](#): Data cleaning using OpenRefine (120 minutes)

A presentation of OpenRefine - an easy tool to standardize and improve the quality of datasets. Followed by a practical exercise using the default features, existing web services and regular expressions. (Presentation and exercise)

11:00 - Coffee/tea break (30 minutes)

Module IX: DATA PUBLISHING (USING IPT)

This module focuses on the process of making biodiversity data freely available online. We will use the GBIF Integrated Publishing Toolkit (IPT).

11:30 - [Session 09a](#): Data publishing using IPT (90 minutes)

Presentation 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). 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. (Presentation and demonstration)

13:00 - Lunch break (60 minutes)

14:00 - [Session 09b](#): Data publishing using IPT - try for yourself (90 minutes)

Exercises in groups where you try for yourself to use the IPT to publish demo datasets or your own datasets in a sandbox environment linked to a demo version of the GBIF portal in Copenhagen. (Exercise)

15:30 - Coffee/tea break (30 minutes)

16:00 - [Session 09c](#): Data publishing using IPT - discussion (60 minutes)

Discussion on the IPT exercise and looking at the datasets that was published by the student groups. (Presentation and Discussion)

17:00 - [Session 09d](#): Data publishing with “data papers” (30 minutes)

Data papers provide a scholarly peer review publication for describing a dataset to increase the fitness for reuse of data by others. (Presentation)

17:30 - End of the day

Day 4 - Thursday 21 February 2019

Module IX: DATA PUBLISHING (continued)

This module focuses on the process of making biodiversity data freely available online. We will use the GBIF Integrated Publishing Toolkit (IPT).

09:00 - [Session 09e](#): FAIR open data (30 minutes)

FAIR data is findable, accessible, interoperable, and reusable. What is the value of the FAIR data principles for your datasets. (Presentation and discussion)

Module X: HOW TO TEACH THE COURSE

This first BioDATA event is for training the trainers/mentors that will teach the course in later training events. The group assignment address how to teach the course in your respective home countries.

10:00 - [Session 10a](#): Introducing group assignment: How to teach the course (30 + 60 minutes)

Introduction to the group assignment. This first course has the goal to train-the-mentors who will contribute to teach the course during the upcoming BioDATA student-courses. (Presentation and discussion)

11:00 - Coffee/tea break (30 minutes)

11:30 - [Session 10b](#): Discussion on how to teach the course in your home countries (90 minutes)

A brief discussion and summary from the planning of how to solve the groups assignment. (Discussion and exercise)

13:00 - Lunch break (60 minutes)

14:00 - [Session 10c](#): Work on the groups assignment - How to teach the course (90 minutes)

Continue working on the group assignment. (Exercise)

15:30 - Coffee/tea break (30 minutes)

16:00 - [Session 10d](#): Work on the groups assignment - How to teach the course (90 minutes)

Continue working on the group assignment. (Exercise)

17:30 - End of day

Day 5 - Friday 22 February 2019

Module X: HOW TO TEACH THE COURSE - GROUP PRESENTATIONS

Each presentation should be about 20 minutes + 10 minutes reserved for questions from the audience.

09:00 - [Session 10e](#): Group 1 Presentation (30 minutes)

Group presentations with proposed solution. (Group presentation)

09:30 - [Session 10f](#): Group 2 Presentation (30 minutes)

Group presentations with proposed solution. (Group presentation)

10:00 - Discussion (60 minutes)

Discussion on the proposed solution on how to teach the course from group 1 and 2.

11:00 - Coffee/tea break (30 minutes)

11:30 - [Session 10g](#): Group 3 Presentation (30 minutes)

Group presentations with proposed solution. (Group presentation)

12:00 - [Session 10h](#): Group 4 Presentation (30 minutes)

Group presentations with proposed solution. (Group presentation)

12:30 - Discussion (30 minutes)

Discussion on the proposed solution on how to teach the course from group 3 and 4.

13:00 - Lunch break (60 minutes)

14:00 - [Session 10i](#): Discussion and planning for the next BioDATA events (90 minutes)

Lessons learnt from this train-the-mentors course to bring forward in the planning of the next BioDATA training events for students in your home country. (Discussion)

15:30 - Coffee/tea break

Module XI: ONSITE CONCLUSION

During this last module of the onsite course, we will review the contents covered and discuss follow-up activities. Participants will have the opportunity to submit their evaluation of the course. Certificates presented.

16:00 - [Session 11](#): Onsite conclusion (60 minutes)

(Presentation, discussion, and course evaluation)

17:00 - End of the training event



5. Follow-up activities: 4 - 15 March 2019

Participants should reserve around 10 hours to complete online activities in the two weeks after the onsite course.

Module XII: Prepare dataset from use case 3 and publish on IPT

The follow-up activity provides an opportunity to demonstrate the skills you have acquired while following this course. These exercises will form the basis for the student evaluation for the ECTS points and the GBIF certification.

[Session 12a](#): Use case 3 (sessions and exercises will be added here when the onsite course is completed)

Day 1 Monday 18 Feb	Day 2 Tuesday 19 Feb	Day 3 Wednesday 20 Feb	Day 4 Thursday 21 Feb	Day 5 Friday 22 Feb
09:00: Presentation 04a. Introduction & GBIF participation 10:00: Presentation / discussion 04b. Why share data?	09:00: Presentation / discussion 07a. Biodiversity data origins and types 10:00: Presentation / discussion 07b. Data capture, processing & quality	09:00: Presentation / exercise 08c. Data cleaning using OpenRefine	09:00: Presentation 09e. FAIR open data 10:00: Presentation / discussion 10a. How to teach the course in your own country train-the-mentors	09:00: Group presentations 10e. Group 1 presentation 09:00: Group presentations 10f. Group 2 presentation 10:00 Discussion
11:00 Coffee break	11:00 Coffee break	11:00 Coffee break	11:00 Coffee break	11:00 Coffee break
11:30: Presentation 05a. Terminology & Standards 12:00: Presentation 05b. Darwin Core 12:20: Presentation 05c. Data quality 12:40: Presentation 05d. Documentation	11:30: Presentation / exercise 07c. Data capture, processing & quality	11:30: Presentation / demo 09a. Data publishing with IPT	11:30: Discussion / exercise 10b. Introducing the group assignment - How to teach the course in your own country	11:30: Group presentations 10g. Group 3 presentation 12:00: Group presentations 10h. Group 4 presentation 12:30 Discussion
13:00 Lunch break	13:00 Lunch break	13:00 Lunch break	13:00 Lunch break	13:00 Lunch break
14:00: Presentation 06a. Planning: workflows and documentation 15:00: Exercise 06b. Planning: stakeholders & roles	14:00: Presentation / exercise 08a. Basic concepts of data cleaning	14:00: Exercise 09b. Data publishing with IPT - try for yourself	14:00: Group exercise continued 10c. Preparation of group assignment	14:00: Discussion 10i. Discussion and planning for the next BioDATA events
15:30 Coffee break	15:30 Coffee break	15:30 Coffee break	15:30 Coffee break	15:30 Coffee break
16:00: Exercise 06c. Planning: tasks and stages 16:30: Discussion 06d. Planning: Suggested solution	16:00: Presentation / exercise 08b. Data cleaning using other tools	16:00: Presentation / discussion 09c. Data publishing with IPT 17:00: Presentation / discussion 09d. Data papers	16:00: Group exercise continued 10d. Preparation of group assignment (continued)	16:00: Conclusion 11. Conclusion and wrap-up
17:30 End of day	17:30 End of day	17:30 End of day	17:30 End of day	17:00 End of day