

Data skills in biodiversity education and research in Belarus: Biodiversity Data Mobilization

11 to 15 November 2019, Bialowieza Forest, Belarus

Last updated 2019-11-07



Image: [European red deer \(*Cervus elaphus*\) Bialowieza Forest March 2017, CC-BY Aleksander Bolbot \(iNaturalist\)](#).

Introduction

[This course](#) aims to enhance the capacity for the attending students 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 in Belarus 11 to 15 November 2019 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. The course is part of the [BioDATA](#) project (2018 to 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 digitization, management and publishing of biodiversity data.

Course teachers

Local organizers

- Oleg Borodin, GBIF Belarus, National Academy of Sciences of Belarus (*local organizer*)
- Nina Voronova, National Academy of Sciences of Belarus (*local organizer*)

Trainers

- Dag Endresen - 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, RAS, Moscow
- Natalya Ivanova - GBIF.ru team, Institute of Mathematical Problems of Biology, RAS
- Dmitry Schigel - GBIF Secretariat, Copenhagen
- Laura Russell - GBIF Secretariat, Copenhagen (*offsite support*)

Mentors

- Ana Sheiko
- Hleb Kryshyn
- Stanislaw Krasouski
- Yauheni Sysoliatin

Software tools

- e-Learning platform: <http://elearning.gbif.es/courses/GBIFDMBIODATA2019/>
- IPT - Integrated data Publishing Toolkit: <https://ipt-biodata.gbif.no/belarus/>
- iNaturalist: <https://www.inaturalist.org/projects/biodata-belarus-2019>

Language

English and 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 GBIF Mozilla Open Badge.

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.
- Bring your own laptop to the onsite course for the best learning experience.

Pre-course activities: 28 October to 8 November 2019

Participants should reserve around 10 hours to complete online activities in the two weeks before 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 01](#), [module 02](#).

[Module 01](#): Preparatory 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*) [[Russian](#)]

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

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

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

[Module 02](#): Preparatory foundation activities

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

Session 02a : Biodiversity informatics (Preparatory reading) [Russian slides]
Session 02b : Data quality (Preparatory reading) [Russian slides]
Session 02c : Digitisation workflows (Preparatory reading) [Russian slides]
Session 02d : Software tools and installation (Preparatory activity) [Russian slides] (Software database) Please try to come prepared to the course with some softwares preinstalled (OpenRefine , Java JRE). We will of course assist you with software installation during the workshop, if needed. Please remember to bring your laptop and ensure that you have administrator rights to install the necessary software.
Session 02e : Preparatory literature [Russian slides] (reading list)
Session 02f : Pre-course questionnaire [Russian slides] Please fill in the pre-course questionnaire [in Russian] before 9 November 2019.

Onsite course: 11 to 15 November 2019

Material for the onsite course are available from the e-Learning platform and for download [here](#). Short [agenda](#) and program [in [English](#) and [Russian](#)] (this current document).

Day 1 - Monday 11 November 2019
Module 03: INTRODUCTION TO THE ONSITE EVENT <i>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.</i>
08:30 - Registration (30 minutes)
09:00 - Session 03a : Welcome and introduction to the course (60 minutes) [Russian slides] <i>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).</i>
Module 04: BIODIVERSITY INFORMATICS FOUNDATIONS <i>This module provides introductions to key concepts and terminology that will be used across the workshop modules.</i>
10:00 - Session 04a : Foundations: Terminology and standards (45+15 minutes) [Russian slides] <i>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 and discussion).</i>
11:00 - Coffee/tea break (30 minutes)
11:30 - Session 04b : Foundations: Darwin Core (30+15 minutes) [Russian slides] <i>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 and discussion).</i>
12:15 - Session 04c : Foundations: Data quality (30+15 minutes) [Russian slides] <i>During this session, participants will be introduced to generic data quality principles, which will be used throughout the remainder of the course (presentation and discussion).</i>

13:00 - Lunch break (60 minutes)

Module 05: PLANNING

This module introduces participants to methods for planning a biodiversity data mobilization project. A document with description of [roles](#) and [stages](#) (in both English and Russian) is made available for the card game.

14:00 - [Session 05a](#): Planning: workflows and documentation (45 minutes) [[Russian slides](#)]

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).

14:45 - [Session 05b](#): Planning: Identifying key stakeholders and roles (45 minutes) [[Russian slides](#)]

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 05c](#): 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 05d](#): Planning: suggested solution (60 minutes) [[Russian slides](#)]

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 12 November 2019

09:00 - [Session 03b](#): Why share data? (60 minutes) [[Russian slides](#)]

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

Module 06: BIODIVERSITY DATA CAPTURE

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

10:00 - [Session 06a](#): Biodiversity data origins and types (45+15 minutes) [[Russian slides](#)]

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).

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

11:30 - [Session 06b](#): Data capture, processing and quality (45 minutes) [[Russian slides](#)]

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

12:15 - [Session 06c](#): Data capture, processing and quality (45 minutes) [[Russian exercise](#)]

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

13:00 - Lunch break (60 minutes)

Module 07: 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 07a](#): Basic concepts of data cleaning (45 minutes) [[Russian slides](#)]
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).

14:45 - [Session 07b](#): Data cleaning using other tools (45 minutes) [[Russian slides](#)]
This session focuses on the tools used to validate and clean datasets in three main categories: nomenclatural, format, and geographical. Followed by a practical exercise (presentation and exercise).

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

16:00 - [Session 07c](#): Data cleaning using OpenRefine (30+60 minutes) [[Russian slides](#)]
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).

17:30 - End of the day

Day 3 - Wednesday 13 November 2019

Excursion: Field trip - [BioBlitz](#) with [iNaturalist](#)

Each BioDATA training course includes a bioblitz with an introduction to the [iNaturalist platform](#) for reporting species observation data, <https://www.inaturalist.org/projects/biodata-belarus-2019>

Before the field trip and bioblitz you should consult the self-learning [Session 6c](#).

09:00 - Visit to Bialowieza Forest by walking from the course venue. All students are asked to register their own user account at the [iNaturalist platform](#) and join [the course bioblitz project](#).

14:00 - Lunch break (60 minutes)

15:00 - [Session 04d](#): Foundations: Documentation (30 minutes) [[Russian slides](#)]
During this session, participants will be introduced to data documentation best practices, which will be used throughout the remainder of the course (presentation).

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

[Module 08](#): DATA PUBLISHING

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

16:00 - [Session 08d](#): Data publishing with “data papers” (45 minutes) [[Russian slides](#)]
Data papers provide a scholarly peer review publication for describing a dataset to increase the fitness for reuse of data by others. (Presentation)

16:45 - [Session 08e](#): FAIR open data (45 minutes) [[Russian slides](#)]
FAIR data is findable, accessible, interoperable, and reusable. What is the value of the FAIR data principles for your datasets (presentation and discussion).

17:30 - End of the day

Day 4 - Thursday 14 November 2019

[Module 08](#): 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).

<p>09:00 - Session 08a: Data publishing using IPT (90 minutes) [Russian slides] <i>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).</i></p>
<p>11:00 - Coffee/tea break (30 minutes)</p>
<p>11:30 - Session 08b: Data publishing using IPT - try for yourself (90 minutes) [Russian slides] <i>Exercises in groups were you try for yourself to use the course 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).</i></p>
<p>13:00 - Lunch break (60 minutes)</p>
<p>14:00 - Session 08c: Data publishing using IPT - discussion (90 minutes) [Russian slides] <i>Discussion on the IPT exercise and looking at the datasets that was published by the student groups (presentation and Discussion).</i></p>
<p>15:30 - Coffee/tea break (30 minutes)</p>
<p>16:00 - Data publishing using IPT - continued (90 minutes) [Russian slides] <i>Exercises in groups were 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).</i></p>
<p>17:30 - End of day</p>

Day 5 - Friday 15 November 2019

Group assignment: Exercise with [Use Case 2](#)

Group assignment

09:00 - Work on the group assignment - [Use Case 2](#) (90 minutes)

Work on the group assignment (exercise).

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

11:30 - Group assignments continued (90 minutes)

Continue work on the group assignment (exercise).

13:00 - Lunch break (60 minutes)

14:00 - Presentations and discussion (90 minutes)

Discussion on the group assignment results.

15:30 - Coffee/tea break

[Module 09](#): 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 \(RU\)](#) of the course. Course certificates presented.

16:00 - [Session 09](#): Onsite conclusion and wrap-up (30 minutes) [[Russian slides](#)]

Presentation, discussion, and [course evaluation](#) [[evaluation form in Russian](#)].

16:30 - End of the training event



Follow-up activities: 25 November to 6 December 2019

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

Module 10: 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 10a:](#) GBIF data portal. [[Russian slides](#)]

[Session 10b:](#) GBIF data use. [[Russian slides](#)]

[Session 10c:](#) Final assessment and [Use Case 3](#). [[Russian slides](#)]

Grading of the course will be based on your submitted solutions for the exercises with [Use Case 2](#) and [Use Case 3](#).

SUMMARY AGENDA

Belarus 11 to 15 November 2019

Day 1 Monday 11 Nov	Day 2 Tuesday 12 Nov	Day 3 Wednesday 13 Nov	Day 4 Thursday 14 Nov	Day 5 Friday 15 Nov
09:00: Presentation 03a. Introduction & GBIF participation	09:00: Presentation, discussion 03b. Foundations: Why share data?	09:00: Field trip / bioblitz Field trip with bioblitz (and iNaturalist)	09:00: Presentation / demo 08a. Data publishing with IPT	09:00: Exercise Exercise with Use Case 2
10:00: Presentation / discussion 04a. Foundations: Terminology & Standards	10:00: Presentation / discussion 06a. Biodiversity data origins and types	Self-learning session 06c before the field trip and bioblitz 06c. Data capture with iNaturalist		
11:00 Coffee break	11:00 Coffee break		11:00 Coffee break	11:00 Coffee break
11:30: Presentation 04b. Foundations: Darwin Core	11:30: Presentation & exercise 06b. Data capture, processing & quality	Field trip / bioblitz (continued)	11:30: Exercise 08b. Data publishing with IPT - try for yourself	11:30: Group presentations Exercise with Use Case 2 continued
12:15: Presentation 04c. Foundations: Data quality				
13:00 Lunch break	13:00 Lunch break	14:00 Lunch break	13:00 Lunch break	13:00 Lunch break
14:00: Presentation 05a. Planning: workflows and documentation	14:00: Presentation / exercise 07a. Basic concepts of data cleaning	15:00: Presentation 04d. Foundations: Documentation	14:00: Presentation / discussion 08c. Data publishing with IPT	14:00: Discussion Results from Use Case 2
15:00: Exercise 05b. Planning: stakeholders & roles	14:45: Presentation / exercise 07b. Data cleaning using other tools			
15:30 Coffee break	15:30 Coffee break	15:30 Coffee break	15:30 Coffee break	15:00 Coffee break
16:00: Exercise 05c. Planning: tasks and stages	16:00: Presentation / exercise 07c. Data cleaning using OpenRefine	16:00: Presentation / discussion 08d. Data papers	16:00: Exercise 08c. Data publishing with IPT (continued)	15:30: Conclusion 09. Conclusion and wrap-up
16:30: Discussion 05d. Planning: Suggested solution		16:45: Presentation / discussion 08e. FAIR data		Final follow-up assignment Follow-up task Use Case 3
17:30 End of day	17:30 End of day	17:30 End of day	17:30 End of day	16:00 End of day

[Course home page](#) -- Preparatory materials: [module 01](#), [module 02](#). IPT <https://ipt-biodata.gbif.no/belarus/>
Venue: [Białowieża forest](#), [Kamenyuki](#), Belarus 11 to 15 November 2019. Agenda last updated 2019-11-07. [EN](#) & [RU](#)