

BioDATA

Accelerating biodiversity research through DNA
barcodes, collection and observation data

Online course (Georgia & Armenia) - June 2021



[Common bumblebee](#) (*Apoidea*) pollinating vetch (*Vicia villosa*) CC BY-NC-ND 2.0 [Jimmy Smith](#)

Summary

Course Description	3
Aim	3
Prerequisites	3
Intended Learning Outcomes	3
Duration	3
Instructors	4
Preparatory Activities	4
Registrations and accounts	4
Day 1, 15 June 2021 (GET/AMT UTC+4 Time Zone)	5
Day 2, 16 June 2021 (GET/AMT UTC+4 Time Zone)	7
Day 3, 17 June (GET/AMT UTC+4 Time Zone)	8
Day 4, 18 June 2021 (GET/AMT UTC+4 Time Zone)	9

Course Homepage

<https://www.nhm.uio.no/english/research/projects/biodata/activities/210615-digital-georgia.html>

Course Description

Aim

This course will teach you how to use DNA barcodes, collection and observation data to resolve research questions in biodiversity. The four-day intensive program will use a combination of lectures, tutorials, and hands-on exercises. You will learn to handle biodiversity data including DNA barcoding. You will gain practical experience in using open and digitally documented biodiversity data through GBIF and BOLD to answer biodiversity research questions. You will understand and practice capturing observation, collection, and genetic data from analogue and digital sources. Finally, the course provides basic skills in data publishing through GBIF and BOLD.

Prerequisites

The course is suitable for MSc and PhD students in biology and other professionals in relevant fields. Participants should have an affinity or professional interest in biodiversity. Participants need to have the motivation and interest to handle DNA barcodes, museum collection data, and observation data. A good understanding of English is necessary to follow the course, carry out the exercises, and get good support during the teaching.

Intended Learning Outcomes

- Understand and be able to explain the concept of species delimitation.
- Learn to use genetic sequence data as a DNA barcode to identify a species.
- Learn to publish and retrieve data from GBIF and BOLD.
- Learn the basics of data capture, cleaning, storage, geo-referencing, and citation.
- Critically assess the quality of own and external data and their fitness for purpose.
- Practice the key tools and approaches to maximize data quality, data linking & data reuse.
- Explore the benefits of FAIR and open data principles in biodiversity research and collaboration.
- Understand the value of data management as a research enabling tool.
- Broadly understand the importance of international biodiversity infrastructures, and how these can contribute to biodiversity assessments, monitoring, conservation, and red-listing.

Duration

Four days, plus eight hours in preparatory time.

Instructors

Dag Endresen (DE); Dmitry Schigel (DS); Helena Wirta (HW); Hugo de Boer (HdB); Laura Russell (LR); Stefaniya Kamenova (SK)

Preparatory Activities

To prepare for the course lectures, participants are strongly encouraged to go through the following preparatory materials. These foundations materials are terms-heavy and likely new, and these resources will prepare for the course lectures; please capture questions as you go through and ask them during the course.

- Foundations videos (67 minutes) playlist at https://youtube.com/playlist?list=PLY6tlKN_kHB8N5_1Li19e2sXpNHjsbziG
 - Terminology
 - [Definitions](#) (12:03)
 - [Software](#) (5:58)
 - [Structures](#) (13:10)
 - [Darwin Core](#) (12:26)
 - [Data Quality](#) (9:48)
 - [Documentation](#) (15:37)
- Excel tips
 - <https://docs.gbif.org/course-data-mobilization/course-docs/Excel-tips-EN.pdf> (can link here or download and attach to email)

Optional: Introduction to GBIF <https://docs.gbif.org/course-introduction-to-gbif/en/>

The course will also include a practical part where you will design a research project and its data elements. This project and its “data journey” will be on a plant-pollinator system. To prepare for this part, please go through the following. Light reading, feel free to skip if you are working on pollinators yourself already.

- Honey detective work raises fears for bees
<https://www.bbc.com/news/science-environment-55662985>
- Loss of bees causes shortage of key food crops, study finds
<https://www.theguardian.com/environment/2020/jul/29/bees-food-crops-shortage-study>

Registrations and accounts

- Check if you have / [create](#) ORCID.
- Create a personal (user) account at GBIF.org, top right corner. You can login to GBIF using ORCID.
- [Create](#) a personal (user) account at BOLD. You can login to BOLD using ORCID.
- [Create](#) an account at iNaturalist. (You can link to your ORCID in settings)

Day 1 - 15 June 2021 ([GET/AMT UTC+4 Time Zone](#))

09:45 - 10:00 Zoom session open

10:00 - 12:15 INTRO SYMPOSIUM - MORNING SESSION

10:00 - 10:30 Talk 1 (HdB)

- Introduction linking barcoding, data and biodiversity

10:30 - 11:00 Talk 2 (DS)

- GBIF for science and policy. Introduction lecture, but diplomatically presented. Showing more the advanced state of progress that Georgia has already reached, and how it crucially contributes to the international community already.

11:00 - 12:15 [Initiatives in Georgia and the Caucasus ecoregion]

- Biodiversity data projects in Armenia - Armine Abrahamyan (15 mins, confirmed)
- Biodiversity data projects in Georgia, incl Georgian Biodiversity Database and CaBOL - David Tarkhnishvili (15 min)
- Data for policy and practice in Southern Caucasus - Salome Nozadze, Ministry of Environment and Natural Resources Protection of Georgia (15 mins, confirmed)

12:15 - 12:30 COFFEE BREAK

12:30 - 13:15 TEACHING BLOCK 1 - INTRO (SK)

- Welcome & practical information
- Introduce the course structure, ILOs, DJC & resources.
- Roundtable in the coffee break for participants to introduce themselves.
- Respond to any questions.

13:15 - 13:45 Lecture 1 "Data as a first-class research citizen" (DE)

- Open data, FAIR [requires machine-readable metadata]
- Local and online databases [pathway from data generation to data publication to data aggregation/portal]

13:45 - 15:00 LUNCH BREAK (75 minutes)

15:00 - 16:00 TEACHING BLOCK 2

15:00 - 15:15 Introduce the workshop Data Journey Case [DJC]

- The overall DJC is an exercise that combines BOLD and GBIF into a single practical example that is followed through the entire workshop, that forms a research journey from the field to paper.

15:15 - 16:00 DJC Step #1 (HW, DS)

- Practical exercise around biodiversity data.

Task 1	5 mins	Think and prepare	→ Draw a schematic presentation of your research plan/questions
Task 2	25 mins	Explore and discuss	→ Search for different species based on their names. Are there any records? → How much overlap in available records do you observe between different databases? → What possible research questions and objects can you formulate? → Do you have the necessary data?
Task 3	15 mins	Think and review	Report on experiences and develop a vision of project progress

16:00 - 16:30 COFFEE BREAK

16:30 - 18:15 TEACHING BLOCK 3 (HW)

16:30 - 17:30 Lecture 2 “Why do we need to identify species?” (HW)

- How to identify species?
- The concept of DNA barcoding: Theory and empirical examples.
- International initiatives (add maybe NorBOL, CABOL, etc.).

17:30 - 18:15 DJC Step #2 (DS, HW)

- Explore GBIF and the Georgian Biodiversity Database to find all records (collections and observations), and visualize this on a map.
- Recap discussion in groups.

Task 4	15 mins	An overview of <i>Bombus</i> & <i>Vicia</i> records in Georgian Biodiversity Database (GBD) & GBIF	Based on GBIF & GBD: → What species of <i>Bombus</i> and <i>Vicia</i> are known to occur in Georgia? → Combine lists into a final list in Excel of species names occurring in Georgia. → Manual editing / cleaning of the list from Georgian Biodiversity Database to make it work at Task 5
Task 5	15 mins	A clean list of species with validated taxonomic names	→ Handle synonyms: https://www.gbif.org/tools/species-lookup
Task 6	15 mins	Recap discussion in groups Ask your questions	Recap what has been learned during the day & provide overture for Day 2

Day 2 - 16 June 2021 ([GET/AMT UTC+4 Time Zone](#))

10:00 - 12:00 TEACHING BLOCK 4

10:00 - 12:00 DJC Step #3 (HW, DS)

- Explore BOLD to find sequence records about the *Bombus* & *Vicia* species occurring in Georgia.
- Explore different ways of accessing DNA barcoding records.

Task 7 (HW)	45 mins	A <i>Bombus-Vicia</i> records in BOLD	→ Are there DNA barcodes available for the species of <i>Bombus</i> and <i>Vicia</i> you listed from Georgia? → If yes, download the data. How many specimens per species? → Which and how many of these species lack a DNA barcode? Mark in your list.
Task 8 (DS)	15 mins	Explore more automatized ways for accessing DNA barcoding data in BOLD	→ Use the GBIF shortcut (at least for <i>Bombus hortorum</i> https://www.gbif.org/species/1340542) → Click around for different species → Do you manage to access all and the same type of data?
Task 9 (HW)	45 mins	Check how much information is needed for referencing a DNA barcode	→ Find out what data and metadata is available for each DNA barcode for the specimen you just located → Examine what information is given in BOLD for a specimen and download the available data for one <i>Bombus</i> specimen

12:00 - 12:30 COFFEE BREAK

12:30 - 13:30 TEACHING BLOCK 5

12:30 - 13:30 Lecture 3 “Species identification and delimitation with DNA” (HdB)

- Methods and principles.
- Barcode Index Numbers, genetic divergence thresholds, marker influence, etc.
- Quality concerns and wet lab basics needed for adequate sequence handling.
- Taxonomy and nomenclature (Light repetition with examples).
- Linnaean and OTU names (Light repetition with examples).

13:30 - 14:45 LUNCH BREAK

14:45 - 16:15 TEACHING BLOCK 6

14:45 - 15:45 Lecture 4 “Barcode reference depositories (BOLD, GenBank, etc.)” (SK lecture + BOLD demo by HW)

15:45 - 16:15 DJC Step #4 “Data and tools in BOLD” (HW, DS, HdB)

- Practical exercise around nomenclature and identification.
- Provide them with several sequences from museum collections with given names, then search sequences in BOLD, and see that all resolve to the same species.
- Show how synonyms are linked to a single accepted name [GBIF example].

Task 10	30 mins	How to get an identification for a sequence & download all sequences for all available <i>Bombus</i> species known from Georgia in BOLD?	→ Copy-paste an “unknown” DNA sequence to BOLD → Which species is it? 10 animal & 10 plant sequences (some with a perfect match, some not at all)
---------	---------	--	---

16:15 - 16:45 COFFEE BREAK

16:45 - 17:15 DJC Step#5 Molecular identification and tree making in BOLD (HW)

Task 11	30 mins	Use BOLD tools	→ Use the Sequence Analysis tools within Dataset 1) Make a tree of them to check how well the sequences group by species 2) Look at the BINs vs species 3) Look at barcode gaps 4) Among the XX barcodes for <i>Bombus</i> , are all of them useful for your research? 5) Are there cryptic species or misidentifications? → Option: give a glimpse of GBIF tool https://www.gbif.org/tools/sequence-id
---------	---------	----------------	--

17:15 - 18:00 TEACHING BLOCK 7

17:15 - 18:00 Lecture 5 “Data types, processing” (LAR)

- Data types
- Data types quiz?
- Data processing

Day 3 - 17 June ([GET/AMT UTC+4 Time Zone](#))

10:30 - 11:00 Online meetup to check that everyone understands the task and has iNaturalist figured out (HdB, DS)

iNaturalist project for the course:

<https://www.inaturalist.org/projects/gbif-bold-course-bombus-and-vicia-in-s-caucasus>

11:00 - 13:00 Excursion: mini-BioBlitz in the botanical gardens or similar iNaturalist event

13:00 - 14:00 Lunch and back to the course

14:00 - 17:45 TEACHING BLOCK 8

14:00 Show the iNaturalist charts, leaders, but winners & prizes at the end of the course (DS)

14:30 - 15:15 DJC Step #6 (LAR)

- Typing and scanning for capturing label information
- Transform written data into computer-read formats

Task 12	45 mins	Data capture Analog to digital (Specimen label digitization)	1-2 images to process by each student
---------	---------	---	---------------------------------------

15:15 - 15:45 Lecture 7 “Data management” (LR, DE)

- Principles of data organization and personal data management (LAR)
- Your data management plan (DE)

15:45 - 16:00 DJC Step #7 (LAR)

- data cleaning practice

Task 13	15 mins	Data quality check (Work with one family and 2 records)	
---------	---------	---	--

16:00 - 16:30 COFFEE

16:30 - 17:45 DJC Step #7 (LAR)

- cleaning practice continued

Task 14	15 mins	Other tools (convert coordinates, dates, etc.)	Play around with the different tools
Task 15	60 mins	Data cleaning	Open refine

Day 4 - 18 June 2021 ([GET/AMT UTC+4 Time Zone](#))

10:00 - 12:15 TEACHING BLOCK 9

10:00 - 11:00 Lecture 8 “Metabarcoding, eDNA applications and biodiversity assessments” (SK)

- Explain how metabarcoding works and differs from regular DNA barcoding.
- Explain how metabarcoding enables working with complex samples.
- Speed presentations of different (meta)barcoding studies presenting a series of examples across the board.
- Reference database, and especially custom database for specific applications (bioinformatically or manually).

- Back to sequence clustering in OTUs, and explain that these are similar to DNA barcode sequences.

11:00 - 11:15 Lecture 9 “Publishing concepts” (LAR)

- “Publishing concepts” (15 min).

11:15 - 11:45 DEMO 10 Publishing to GBIF (LAR)

11:45 - 12:15 COFFEE

12:15 - 12:45 DJC Step #8 (LAR)

- Publishing to GBIF

Task 16	45 min	Check the outcome of already published datasets	→ Look at the occurrence records in GBIF → Look at the maps & download the occurrences
---------	--------	---	---

13:00 - 14:15 LUNCH

14:15 - 15:15 DJC Step #9 (Levan Mumladze)

- Publishing to BOLD

Find a video to watch together and then discuss practicalities of BOLD publishing process

Task 17	60 mins	DEMO BOLD data publishing	Create a new project in BOLD Upload specimen data Upload images Upload sequence & trace files
---------	---------	---------------------------	--

15:15 - 15:45 COFFEE

15:45 - 16:15 Lecture 11 “Data publishing benefits” (DE, DS)

- Data exposure: why and when
- Data citation and DOI
- Data papers

16:15 - 17:15 Final hour

- Course evaluation (fill in the evaluation form).