

Data Management Plan

GBIF Norway 2021-2025

Contact people	Dag Endresen dag.endresen@nhm.uio.no , ORCID: 0000-0002-2352-5497 Rukaya Johaadien rukayasj@uio.no , ORCID: 0000-0002-2857-2276 Michal Torma michal.torma@nhm.uio.no , ORCID: 0000-0002-2690-9509
Based on	Life Sciences DSW Knowledge Model - ELIXIR Norway localization, 0.3.0 (elixir.no:lifesciences-elixir-norway:0.3.0)
Project phase	Before Submitting the DMP
Created by	Dag Endresen (dag.endresen@gmail.com) UiO
Generated on	05 Feb 2024

Data Management Plan created in Data Stewardship Wizard «ds-wizard.org»

HISTORY OF CHANGES		
Version	Publication date	Changes
<i>There are no named versions</i>		

Contributors

The following contributors are related to the project of this DMP:

- Dag Endresen
dag.endresen@nhm.uio.no, ORCID: [0000-0002-2352-5497](https://orcid.org/0000-0002-2352-5497)
Roles: Contact Person, Project Leader
- Rukaya Johaadien
rukayasj@uio.no, ORCID: [0000-0002-2857-2276](https://orcid.org/0000-0002-2857-2276)
Roles: Contact Person, Data Manager, Work Package Leader
- Michal Torma
michal.torma@nhm.uio.no, ORCID: [0000-0002-2690-9509](https://orcid.org/0000-0002-2690-9509)
Roles: Contact Person, Data Manager, Data Steward

Projects

We will be working on the following projects and for those are the data and work described in this DMP.

GBIF Norway 2021-2025

Start date	2021-01-01
End date	2025-12-31
Funding	Norges Forskningsråd : 322444 (granted)

2.4 Project description and project summaryThe project description for project no. 322444is provided in the document dated 18.11.2020.Changes in the project description must be approved by the Research Council. Requests for changes must be submitted to the Research Council via“My RCN Web” > “Projects/Reports” > “View / Change project”.The following project summary may be published by the Research Council:GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.GBIF operates through a network of Nodes which coordinates the biodiversity information facilitation of participant countries and organizations, collaborating with each other and the Secretariat to share skills, experiences, and technical capacity. The data are published by many institutions from around the world and include scientific names data, primary distribution and monitoring data on plants, animals, fungi, and microbes. By encouraging and helping institutions to publish data according to common standards, GBIF enables research that could previously not be performed and informs better decisions to conserve and sustainably use the biological resources of the planet.Strategic priorities1. Empower global networkEnsure that governments, researchers, and users are equipped and supported to share, improve and use data through the GBIF network, regardless of geography, language, or institutional affiliation.2. Enhance biodiversity information infrastructureProvide leadership, expertise, and tools to support the integration of all biodiversity information as aninterconnected digital knowledgebase.3. Fill data gapsPrioritize and promote mobilization of new data resources which combine with existing resources to maximize the coverage, completeness, and resolution of GBIF data, particularly with respect to taxonomy, geography, and time.4. Improve data qualityEnsure that all data within the GBIF network are of the highest possible quality and associated with clear indicators enabling users to assess their origin, relevance, and usefulness for any application.5. Deliver relevant dataEnsure that GBIF delivers data in the form and completeness required to meet the highest-priority needs of science and, through science, society.

Section A: Data Collection

1. What data will you collect or create?

Data formats and types

We will be using the following data formats and types:

- Darwin Core

It is a standardized format. This is a suitable format for long-term archiving. We will have only a small amount of data stored in this format.

2. How will the data be collected or created?

There will be no instrument dataset in this project.

Section B: Documentation and Meta-data

3. What documentation and meta-data will accompany the data?

List of data to be published is given in Section E, Question 9. This also includes information about catalogs where the data can be found. Information about data types used is given in Section A, Question 1.

We will use other solution than (electronic) lab notebooks to make sure that there is good provenance of the data analysis: Planning further extension of tools for tracking content differences between updated versions of datasets, <https://archive.gbif.no>.

We will be documenting the data with Dublin Core and DataCite metadata standards.

Metadata will be openly available.

Section C: Ethics and Legal Compliance

4. How will you manage any ethical issues?

Data we produce

For the data we produce, the ethical aspects are as follows:

- **We have 1613 datasets as of 25 April 2022 - no information on personal nor sensitive data.**

5. How will you manage copyright and Intellectual Property Rights (IPR) issues?

We will be working with the philosophy *as open as possible* for our data.

All of our data can become completely open over time.

Limited embargo will not be used as all data will be opened.

We have made the following arrangements regarding the data ownership: Data will remain owned by the individual data publishers, and data reuse will be tracked back to data owners.

Section D: Storage and Backup

6. How will the data be stored and backed up during the research?

Storage needs are small at the beginning and will grow later.

All essential data is also stored elsewhere to prevent a total loss of data. Backup of project data that is not in the work space is left to the project participants.

7. How will you manage access and security?

Project members will not store data or software on computers in the lab or external hard drives connected to those computers. They will not carry data with them (e.g. on laptops, USB sticks, or other external media). All data centers where project data is stored carry sufficient certifications. All project web services are addressed via secure HTTP (<https://...>). Project members have been instructed about both generic and specific risks to the project.

The possible impact to the project or organization if information is lost is small. The possible impact to the project or organization if information is leaked is small. The possible impact to the project or organization if information is vandalised is small.

We are not using any personal information.

Only project members will have read access; only selected project members will be able to write data.

Section E: Selection and Preservation

8. Which data are of long-term value and should be retained, shared, and/or preserved?

We plan to produce the following datasets:

- **We have 1613 datasets as of 25 April 2022** (published) – This data set will be kept available as long as technically possible. – The metadata will be available even when the data no longer exists.

9. What is the longterm preservation plan for the dataset?

- **We have 1613 datasets as of 25 April 2022** (published)
The distributions will be stored in:
 - Domain-specific repository: Global Biodiversity.

None of the used repositories charge for their services.

We have a reserved budget for the time and effort it will take to prepare the data for publication.

Section F: Data Sharing

10. How will you share the data?

- **We have 1613 datasets as of 25 April 2022**
The dataset has the following identifiers:
 - DOI: [All of our datasets have DOIs](#)

The distributions will be available as follows:

Information about used repositories (i.e. where will potential users find out about the data) is provided in Section E, Question 9.

Embargo on the data is described in Section C, Question 5, and Section F, Question 11.

11. Are any restrictions on data sharing required?

Ethical and legal restrictions are documented under Section C. We have used the Data Stewardship Wizard, which made us aware of options to minimize the restrictions.

No data sharing agreement will be required.

Section G: Responsibilities and Resources

12. Who will be responsible for data management?

Dag Endresen is responsible for implementing the DMP, and ensuring it is reviewed and revised.

Rukaya Johaadien and Michal Torma are responsible for maintaining the finished resource.

Michal Torma is responsible for the management and proficiency of data including data processing, data policies, data guidelines, and data availability.

13. What resources will you require to deliver your plan?

To execute the DMP, additional specialist expertise is required. We will be training existing staff.

We require the following hardware or software in addition to what is usually available in the institute: GBIF integrated data publishing toolkit (IPT); for standardized data publishing in compliance with domain routines and standards.

Charges applied by data repositories (if any) are mentioned already in Section E, Question 9.