

Major Types of Biological Data



Five major types of biological data have been defined in [GFBio](#) and are recognised by the [GFBio Data Centers](#) in NFDI4Biodiversity. They are used for the "Service Description" of the individual Data Centers as well as in the context of the Technical Documentations of processing tools.

Types of biological data:

- [Type 1: Biodiversity and Occurrence data](#)
- [Type 2: Taxon Data](#)
- [Type 3: Environmental Biological and Ecological Data](#)
- [Type 4: Non-Molecular Analysis Data](#)
- [Type 5: Molecular Sequence Data](#)
- [Additional Information](#)

Type 1: Biodiversity and Occurrence data

These are the data from the classical collection and alpha-diversity research domain, i.e. digital objects with taxon name(s), georeferences, e.g. locality, date and often referenced resources as multimedia objects. We distinguish between:

- Type 1a: Collection Data (with reference to physical object)
- Type 1b: Observation Data (without reference to physical object)

Used standards:

- [ABCD](#) (Access to Biological Collection Data) and extensions
- [DwC](#) (Darwin Core) and extensions
- [DC](#) (Dublin Core) as included in ABCD and DwC for basic bibliographic information

Used identifiers:

- primary identifier: biological (digital) object (digital specimen or observation)
- main secondary information: geo-information and time, related (multimedia) resources

Example packages:

- [Curators Herbarium B \(2020\). Digital specimen images at the Herbarium Berlinense. \[Dataset\]. Version: <2020-10-07>. Data Publisher: Botanic Garden and Botanical Museum Berlin. <https://data.bgbm.org/dataset/gfbio/0001/>. \[Please cite individual specimens with their stable ID, for images add the image ID.\] \(digital specimens\), also accessible via GFBio VAT](#)
- [Schott, H. \(2018\). IBF Monitoring of Orthoptera, University of Regensburg. \[Dataset\]. Version: 20181205. Data Publisher: Staatliche Naturwissenschaftliche Sammlungen Bayerns – SNSB IT Center, München. \[http://www.diversitymobile.net/wiki/IBForthopteracoll_About\]\(http://www.diversitymobile.net/wiki/IBForthopteracoll_About\). \(digital observations\), also accessible via GFBio VAT](#)
- [Rakotoarison, A.; Scherz, M. D., Bletz, M. C.; Razafindraibe, J. H.; Glaw, F. & Vences, M. \(2019\). Media and additional measurements belonging to the description of *Cophyla fortuna* \(Microhylidae, Cophylinae\). \[Dataset\]. Version: 1.0. Data Publisher: Zoological Research Museum Koenig - Leibniz Institute for Animal Biodiversity. <https://doi.org/10.20363/media-cophyla-fortuna-1.0>. \(digital specimens\), also accessible via GFBio VAT](#)

Notes

The time investment for individual scientific data curation to be done by data providers and GFBio data managers before and during data transformation is varying.

[Back to top](#)

Type 2: Taxon Data

These are taxon-related data (e.g. in a catalogue, checklist or so-called red list).

Used standards:

- [ABCD](#) (Access to Biological Collection Data) and extensions
- [DwC](#) (Darwin Core) and extensions
- [DC](#) (Dublin Core) as included in ABCD and DwC for basic bibliographic information

Used identifiers:

- primary identifier: class name (taxon), e.g., as defined by the nomenclatural rules of the three International Codes of Biological Nomenclature
 - [International Code of Nomenclature for algae, fungi, and plants](#)
 - [International Code of Zoological Nomenclature](#)
 - [International Code of Nomenclature of Prokaryotes](#)

- main secondary information: taxonomic classifications and concepts, synonymy, vernacular names, geo- and conservation status information etc.

Example packages:

- [Taxon list of vascular plants from Bavaria, Germany compiled in the context of the BFL project](#), also accessible via GFBio terminology service and as taxon backbone in GFBio portal
- [Taxon list of animals with German names \(worldwide\) compiled at the SMNS](#), also accessible via GFBio terminology service and as taxon backbone in GFBio portal

Notes

The time investment for individual scientific data curation to be done by data providers and GFBio data managers before and during data transformation is varying.

[Back to top](#)

Type 3: Environmental Biological and Ecological Data

These are environmental biological and ecological study data including functional and phylogenetic trait data and other kind of analysis data.

Used standards:

- [EML](#) (Ecological Metadata Language)
- [DELTA](#) (Description Language for Taxonomy, for trait data)
- [SDD](#) (Structured Descriptive Data, for trait data)
- [GML](#) (Geography Markup Language) and ISO 19139 metadata

Used identifiers:

1.
 - a. primary identifier: biological class concept (e.g., OTU or OFU)
 - b. main secondary information: trait and environmental (analysis, measurement, transformation, translocation) information
2.
 - a. primary identifier: environmental and ecological study item and event
 - b. main secondary information: biological and ecological information, measurements and description of the environment

Example packages:

- SDD example with EML for basic bibliographic information
 - see [DOI: 10.25897/5/nhc7-0d72](#) and [DOI: 10.25897/5/tyc9-k378](#) (SNSB data publication pipeline under construction)
- EML example with CSV table data structured according the EAV data model
 - [Ferber, Stefan; Schleuning, Matthias; Hemp, Andreas; Howell, Kim; Böhning-Gaese, Katrin \(2018\): Various investigations to analyze the effects on species richness of birds during the KiLi \(Kilimanjaro\) Project. PANGAEA, <https://doi.org/10.1594/PANGAEA.896128>](#)
 - see [DOI: 10.25897/5/j2cs-q186](#) and [DOI: 10.25897/5/kk8s-7a12](#) (SNSB data publication pipeline under construction)

Notes

The time investment for individual scientific data curation before and during data transformation of (matrix) data into a highly structured and standard schema-compliant format at data item level might be high. Thus, the data management process has to be agreed between data provider and GFBio data curator before starting (see DMPs).

[Back to top](#)

Type 4: Non-Molecular Analysis Data

These are non-molecular analysis data (data sets and/or data packages) in its original data file format (often RAW format).

Used standards

- [EML](#) (Ecological Metadata Language) for basic bibliographic information
- [DC](#) (with [Pansimple XSD](#)) for basic bibliographic information

Used identifiers:

- primary identifier: as provided by data producer
- main secondary information: as provided by data producer

Example packages:

- *coming soon*

Notes

This type of data is accepted, as far as well documented and with a core set of standard-compliant metadata and appropriate for long-term archiving.

The time investment for individual scientific data curation to be done by data providers and GFBio data managers before and during data transformation might be limited.

[Back to top](#)

Type 5: Molecular Sequence Data

These are molecular sequence data including MixS-compliant metadata.

Used standards:

- [MixS](#) (Minimum Information about any (X) Sequence)

Used identifiers:

- primary identifier: molecular sample accession
- main secondary information: geo-information and time

Example package:

- [PRJEB26997](#)

Notes

The time investment for individual scientific data curation to be done by data providers and GFBio data managers before and during data transformation might be limited.

[Back to top](#)

Additional Information

For more details see also

- [Data exchange standards, protocols and formats relevant for the collection data domain within the GFBio network](#)
- [Technical documentation of GFBio publication of **type 1** data](#)
- [Technical documentations](#)

[Back to top](#)



Do you have questions, feedback or need help?

[Contact our Helpdesk](#) for direct support.

