



## Eligibility criteria for GFBio portal data providers

During the 6th GFBio Steering Committee meeting (December 2016), it was decided to define eligibility criteria for external data providers from major joint initiatives wishing to disseminate their metadata and data via the GFBio portal, with or without using the archiving services of GFBio. The following document defines some basic rules which data providers, such as external data archives or large scale research projects, need to comply with before metadata on their data assets or their assets themselves can be added to the GFBio data portal.

Eligibility of data providers requires approval from the GFBio Steering Committee.

Data providers must cooperate with GFBio and sign a written agreement (Memorandum of Understanding) defining the tasks of both partners which complies with the following criteria:

### Data archive criteria

- Data providers must offer data in the thematic scope of GFBio.
- Data providers must have appropriate quality assurance mechanisms in place. These need to be documented.
- Data providers must guarantee midterm perspective for a stable data management for 10 years.
- A dedicated technical and organizational contact person must be named.
- Appropriate metadata (and data) formats have to be provided (see below).
- Appropriate metadata (and data) exchange interfaces have to be provided (see below).
- Data providers have to keep submitted metadata records up to date and synchronized.
- An open access policy has to be agreed to (see below).
- Data providers must deliver a concept for long-term data archiving, e.g. in cooperation with an institutional data repository which must either include:
  - A strategy to transfer data to a certified or mandated long-term data archive
  - A strategy to achieve a certification as a long-term archive by e.g. DINI

### Metadata access criteria

- Metadata need to be provided in at least one of the following formats:
  - Ecological Metadata Language EML (v. 2.1)



- ABCD (v.2.06) in the case of collection and occurrence monitoring data
- Directory Interchange Format (DIF)
- INSPIRE compliant ISO19139
- GFBio Indexing Format: Extended Dublin Core (pansimple: <http://ws.pangaea.de/schemas/pansimple/pansimple.xsd>)
- ABCD data needs to be provided as ABCD xml archive (zip) with BioCASE provider endpoint which must be registered at the GFBio BioCASE Monitor Service.
- All other metadata files need to be provided either via:
  - OAI-PMH (preferred)
  - FTP directory

## Data access criteria

- In the case of collection and occurrence monitoring data, data has to be provided as ABCD (v.2.06) (which have to be registered at the GFBio BioCASE Monitor Service (BMS)).
- In the case of taxonomic or ecological data DwC, EML or a similar TDWG-agreed content standard for data is preferred.
- At least one data set or file must be provided for each metadata set.
- Metadata in one of the above defined formats has to be provided for each data set.
- Data must be accessible via the internet (download option).
- Data must be accessible in an appropriate format (Desktop formats (e.g. xls) ASCII, XML).

## Data policy criteria

- Metadata must be offered as open access without restrictions.
- ABCD structured data must be offered as open access under an appropriate licence such as the Creative Common License.
- Necessary rights to harvest and publish metadata (in the case of ABCD also data) via the GFBio website must be granted.
- Data for which metadata (and data) is provided via GFBio shall be offered as open access.
- Data providers must comply with Directive 2003/4/EC or the Umweltinformationsgesetz (UiG) respectively.
- Data offerings in relation to genetic material must comply with the Nagoya Protocol.