

iDiv Data & Code Sharing Policy and Guidelines

Executive Summary

iDiv represents a large-scale collaborative biodiversity science research effort, whose central mission is to promote theory-driven synthesis and data-driven theory. It is widely agreed upon that data preservation and sharing is essential to further biodiversity research in general and to reach these aims in particular. To support this, many funding agencies – including the German Research Foundation (DFG) – have strong expectations with respect to data management and reuse.

iDiv aims to be a world leader in this respect. In committing to make data and code high quality, open and FAIR (i.e. findable, accessible, interoperable and reusable, by both humans and machines), we wish to lead by example. We will make our science reproducible, widely accessible, and preserve it in a way that enables future generations to build upon it. Openness and FAIRness are key building blocks to support capacity building (globally) among young scientists and society. With this, iDiv meets its responsibility as a world-renowned research centre. These principles of openness and FAIRness are also reflected in iDiv's Code of Conduct & Scientific Integrity¹.

Our Data & Code Sharing Policy reflects these principles. It provides guidelines on how we will incentivize and support the provision of open and FAIR data and code, and outlines consequences of not following these rules.

In a nutshell, this policy says: As a rule, data and code produced within the iDiv context (e.g. in the core and member groups, in sDiv workshops, by external users of iDiv platforms) shall be made openly available in an appropriate repository (the iDiv Data Portal being one example). Care shall be taken that this data and code are well curated and FAIR.

Formalities

Within iDiv, the **Data & Code Unit** (iBID) provides support and advice on all aspects of open data and code, and administers the iDiv Data Portal, a repository for research data management including data storage, sharing, and reuse. The iDiv Data Portal Usage Policy and the FAQ on the Portal (<https://idata.idiv.de>) describe the rules governing usage of this infrastructure and of the data provided through it.

Definitions of key terms used in the text are provided at the end of this document.

¹ <https://www.idiv.de/code-of-conduct>

iDiv Data & Code Sharing Policy

Aim:

As a leading research centre in biodiversity research, iDiv strives to become a world leader with respect to following the principles of open science and FAIR data. In committing to make data and code high quality, open and FAIR (i.e. findable, accessible, interoperable and reusable, by both humans and machines), we wish to lead by example. We will make our science reproducible, widely accessible, and preserve it in a way that enables future generations to build upon it.

This means:

At iDiv we commit to making *all* data, code and potentially additional materials (e.g. research software, electronic lab notebooks) that have been created within the iDiv context and are of potential future use, openly available and FAIR.

What does “created within the iDiv context” mean?

- Work that has received iDiv funding. iDiv funding is interpreted as resources provided to iDiv professors and their core staff, sDiv postdocs, or scientists financed via the Flexpool or other iDiv funds. For members, the conditions apply to those components of their work which were funded by iDiv and/or have relied upon iDiv equipment, platforms, facilities, or personnel for their generation.
- In the spirit of iDiv’s Code of Conduct & Scientific Integrity, we strongly encourage members to interpret this generously, and to apply this policy to all their biodiversity-related data and code.
- iDiv platforms are encouraged to include this policy into their Usage Rules and thereby request external users to adhere to it.
- For collaborative work with people or organisations outside of iDiv, it should be the aim of the iDiv scientists involved to strive for adherence to the iDiv rules.

What does “potential future use” mean?

- “Potential future use” in particular is given if data or code has been used in a publication or other research output, or if the measurements are unrepeatable or cannot easily be repeated, e.g. field observations taken on a particular day or under specific conditions.
- If in doubt, err towards FAIR and open preservation.

What does “open” mean?

- The short version of the definition provided by the Open Knowledge Foundation² is “data and content can be freely used, modified and shared by anyone for any purpose”.
- “Open” does **not** mean that the rules of good scientific practice, in particular the rules with respect to attribution, do not apply. They are still valid.

What do I need to do to meet iDiv’s minimal requirements for making my data and code open?

- Data providers are encouraged to select the least restrictive license possible. For example, at the time of writing this Policy, in the iDiv Data Portal most datasets use CC0 (public domain) or CC BY 4.0³. However, we recognize that making datasets open is not an all-or-nothing process, and particularly in the case of data collected or aggregated at sDiv workshops, or third-party funded research, the interests of a diverse group of internal and external data owners may need to be balanced. In cases where external partners are reluctant to make their data open, iDiv researchers might suggest the adoption of a staged process, for example opening up selected parts of their joint data initially, and progressing to a greater proportion over time.
- We follow the maxim “as open as possible, as closed as necessary”. For example, privacy or conservation concerns can be addressed by obfuscating certain components of a dataset if necessary (e.g. via anonymization, or blurring of geographic coordinates); such concerns need not prevent the remainder of a dataset being made open.
- Similarly, for code, the least restrictive license possible should be used. During code development, care should be taken whenever possible to avoid the use of existing libraries that would prevent publication of the code under an open license or create conflicting licenses.
- Data/code should be made publicly available as early as possible. In iDiv-associated repositories, the default position is to make datasets open within 2 years, starting from the end of data acquisition. It is possible to extend this embargo period by additional 1-year increments, in rare well-motivated cases, upon request. Data custodians undertake to comply when the Data Manager of an iDiv repository follows up at the end of an embargo period, or make an adequate case for continued restriction of access. iBID staff will provide advice on these issues.
- Upon request, iBID can provide the means to make datasets stored in the iDiv Data Portal available to subsets of individuals *within* iDiv, or to journal editors and reviewers, prior to the end of any embargo period. Data custodians can indicate the conditions under which data is to be released.
- Individuals must adhere to the authorship rules outlined in iDiv's Code of Conduct & Scientific Integrity.

² <https://opendefinition.org/>

³ <https://creativecommons.org/licenses/by/4.0/>

What does “FAIR” mean?

- The FAIR guiding principles⁴ require data (and other artefacts) to be:
 - Findable
 - Accessible
 - Interoperable
 - Reusable
- Important aspects to achieve these goals are, among others:
 - Data and metadata should have persistent identifiers, e.g. DOIs
 - Open, non-proprietary formats should be used, e.g. CSV rather than .xlsx
 - Data should be described by rich metadata
 - Data and metadata should be deposited in sustainable, searchable repositories
 - Data usage conditions should be clear, e.g. a license should be specified
- The same principles can also be applied to code.

What do I need to do to meet the minimum requirements of iDiv towards making my data and code FAIR?

- iDiv itself offers or contributes to multiple repositories (e.g. the iDiv Data Portal, JEXIS) or iDiv-associated solutions (e.g. AquaDiva) within which data (and code) can be archived and published with a DOI. For data produced within the context of iDiv but without iDiv funding (e.g. in third-party funded projects of iDiv members) and which is to be made publicly available, the iDiv Data Portal may offer to store the data or information about it. The conditions of usage (for example any associated cost, in the case of large datasets) will be negotiated on a case by case basis.
- If a specialized repository exists and is widely used in a given subdiscipline, scientists are encouraged to deposit their data there (for example GenBank or MetaboLights for "-omics" data, or the TRY database for plant trait data). For more general datasets, use of the iDiv Data Portal is encouraged, but users are free to use other general (ecological) repositories (e.g. PANGAEA). However, we strongly advise scientists to choose repositories that meet quality criteria and in particular provide *curation* of (meta)data, as these offer the best chance that data will be appropriately archived and ultimately reusable without future users resorting to contacting the data provider or custodian(s), and are often certified. re3data.org is a useful resource for finding such repositories. iBID staff are happy to advise on the choice of repository.
- Provision of rich metadata (i.e. information describing the data) is an essential component. It is the metadata content which makes data/code findable and accessible by both humans and machines (e.g. search engines and other external services), hence increasing FAIRness. Rich contextual information is also essential if the data are to be effectively re-usable by others.

⁴ Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, *et al.* (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3: 160018. <https://doi.org/10.1038/sdata.2016.18>

- Similar to data, code (e.g. statistical scripts, research workflows and software) shall be made publicly available, in addition to the metadata and data *per se*. iBID staff can advise researchers in understanding best practice requirements, for example the choice of a license, and the need to provide a README file and adequate documentation of the codebase for i) developers, and ii) end users.
- In the future, relevant services and platforms will be offered by NFDI4Biodiversity and other NFDI consortia. These developments will be closely monitored and publicized within iDiv.

When should I start thinking about making my data and code FAIR and open?

- The short answer to this is: as early as possible, preferably *before* you start producing data or code.
- To encourage this, all collaborative efforts within iDiv or involving iDiv need to **prepare a data management plan** (DMP) prior to their start. For example, researchers should consider their data storage needs and back-up strategy well in advance, in order to acquire funding and/or negotiate appropriate storage space from their institution (the Data & Code Unit handles datasets that are intended to be made publicly available, rather than storage of data/code intended primarily for researchers' internal use). Collaborative efforts include externally funded projects (like Collaborative Research Centres), but also informal agreements to collaborate on a certain topic and projects funded by sDiv or the Flexpool. Usage of the DMP tools provided by gfbio.org or (in the future) NFDI4Biodiversity is strongly encouraged.

How will iDiv support making data and code FAIR and open?

- iBID, iDiv's Data & Code Unit, has been established with the aim of facilitating openness and FAIRness.
 - iBID staff can support you in applying for data management resources in funding proposals.
 - iBID staff can provide guidance in writing Data Management Plans.
 - The iBID helpdesk provides support with all aspects of data and code curation, and can also advise on suitable repositories.
 - iBID runs the iDiv Data Portal and other iDiv data solutions and can support you in uploading your data and obtaining a DOI after a curation process.
 - iBID supports networking and training of research data management experts in iDiv groups.
 - iBID contributes to training the next generation of iDiv scientists on open and FAIR science.
 - iBID keeps track of national and international developments (e.g. requirements of funding agencies, emerging standards, licenses) and will inform users where relevant.
 - iBID will prioritize support requests that will result in open and FAIR data and code and/or are of strategic importance to iDiv.

- Authors are requested to **include a brief statement acknowledging the contribution of iBID personnel** in their manuscripts, if an associated dataset has been through the curation process or the researcher has otherwise received substantial consultation and assistance from iBID staff.
- To highlight both individuals' and iDiv's efforts to support open science, the DOI (or URL, where no DOI has been obtained) for data, code and/or other material that has been made publicly available will be displayed on the iDiv website (Publications section), alongside the DOI for the associated article.
- In evaluating applications for funding (e.g. Flexpool or sDiv workshops) or other resources, iDiv reserves the right to use i) information provided in the proposal about plans to make the eventual data/code publicly available, and ii) the proportion of an individual's published work over the preceding 5 years for which associated data and/or code were made publicly available. For applications to iDiv's Open Access fund (for publishing articles), it is a condition that the specific data/code associated with the publication for which support is requested is made publicly available.
- Lack of adherence to the Data & Code Sharing Policy and Data Portal Usage Policy with respect to data provision and/or usage may result in loss of access rights to the iDiv Data Portal, the withholding of funds, or refusal to allow future applications for resources from iDiv including, but not limited to, funding or the use of infrastructure, equipment, platforms, facilities or personnel provided by iDiv.

If a provision of this agreement is or becomes illegal, invalid or unenforceable in any jurisdiction, that shall not affect:

1. The validity or enforceability in that jurisdiction of any other provision of this agreement; or
2. The validity or enforceability in other jurisdictions of that or any other provision of this agreement.

I accept the terms and conditions of the iDiv Data & Code Sharing Policy and Guidelines.

Name: _____ Affiliation: _____

Date: _____ Signature: _____

Definitions

In this policy the following definitions are used:

(Primary) Data: includes measurements, observations, model output, output of synthesis analyses, etc.

Metadata: descriptive information about individual or aggregate datasets, necessary for the interpretation, reuse, and disclosure of data. In case of externally stored data, the links to these repositories are provided.

iDiv Data Portal User: any person using the iDiv Data Portal, for the discovery, download or upload of datasets.

iDiv Member: as defined in the Bye-laws of iDiv⁵.

iDiv Scientist: any person (PI, Postdoc, PhD student, technical staff, Master or Bachelor student working on their thesis, research assistants, interns) working within iDiv.

Data Collector: the person or persons responsible for collecting/creating the data.

Data Custodian: person responsible for the provision of a specific set of data and metadata to a data repository or similar. From the point of view of iDiv, this person represents all involved data owners and is responsible for ensuring that the intellectual property rights of data owners and collectors are not violated.

Data Owner: the data owner is the university or research institution for which the data are collected. It is represented by the principal investigator (PI) of the project or part of the subproject for which they are responsible. The data owner has to ensure that other people associated with the data (in particular data collectors) are appropriately considered and that their intellectual property rights are not violated. The PI can delegate the tasks associated with data ownership to other PIs or group members.

Data Provider: person uploading data or metadata, for example to the iDiv Data Portal.

iBID: the iDiv Data & Code Unit, which administers the iDiv Data Portal.

Data Manager: the head of the iDiv Data & Code Unit (iBID) or the manager of an iDiv repository.

⁵ https://www.idiv.de/fileadmin/content/Files_CentralManagement/Ordnung_iDiv_210820.pdf