

Title	Version	Last consented (Due for review)
IBL data release and sharing policy	1.2.0	8 March 2022
Authors		
Gaelle Chapuis, Niccolò Bonacchi, Karel Svoboda, Sonja Hofer, Hannah Bayer, Julia Huntenburg		
Editors		Working Group
		DAWG, Governance
Abstract		
A key mission for the International Brain Laboratory is to make our data open and accessible to the community. This policy describes the mechanisms for data release and sharing.		

Table of contents

Goals and Scope	3
Definitions	3
Data	3
Registered project	3
Data acquirer	3
Timeline terms	4
Data ownership	4
Data public release	4
Data curation effort	4
Data Documentation	5
Data format and platform	6
Authorization of release	6
Release timeline	6
Versioning	7
Data access	7
Access rights	7
Usage rights	8
Citation and publication rights	8
Critical Checkpoints for Data Release	8

Goals and Scope

- Ownership
- Release
- Access

Definitions

Data

Data is defined here as encompassing raw experimental data, processed datasets created by standard post-processing pipelines, **and the associated metadata**.

This policy concerns only **data acquired by IBL members using IBL infrastructure as part of IBL-registered projects**. This includes both personal and platform projects.

This policy does not concern, and thus does not support, data that is **not** part of IBL-registered projects.

Registered project

All projects are registered according to the [IBL Charter](#). A registered project can either be a platform or a personal project. Oversight for registered personal projects is in accordance with the [Project Board Policy](#); oversight for registered platform projects is provided by the relevant Task Force and the Executive Board.

Data acquirer

The data acquirer is the person who acquired the data, i.e. physically ran the experiment.

The name of the acquirer (“user name”) is retrievable at both the subject and the session level from the database. If the subject and session “user names” differ (i.e. the subject “responsible user” has delegated the running of the experimental procedure to someone else), both users are considered data acquirers.

Timeline terms

- **Acquisition Date.** The acquisition date is the date at which data acquisition is completed, which is when the experimental session is completed and the associated raw data is registered and stored onto the centralized IBL database. This typically happens within days of the experimental procedure having taken place.
- **Collection Date.** The collection date is the date at which **the complete** dataset (raw and processed) gathered as part of a project has been formally controlled for quality and approved for public release.

Data ownership

Ownership of data remains with the laboratory in which it was acquired. Data registration, storage and access on the centralized IBL database, as described in this policy, are mandatory.

Data public release

Data curation effort

The IBL will make an effort to curate and make publicly accessible all aspects of a dataset.

All data will be publicly released with the exception of:

- Data known to be incorrect.
- Data that did not pass a quality control.

Data with known quality errors or which has not been checked for quality control may be released upon request, but only with an explicit statement of the errors or absence of quality control.

Curation thus entails:

- Establishing necessary quality checks on the data (inclusive of metadata)
- Validating the data passes all necessary quality checks (and documenting any errors)

For **platform projects**, the relevant working groups will develop and implement quality checks across the entire dataset. Automated quality checks are preferred, but manual checks are acceptable if no practical automated alternative is available. The task force will be responsible

for establishing which quality checks and passing criteria are appropriate for the project. The task force is also responsible for validating and documenting that all data passes these quality checks, and may ask for staff assistance with this.

For **personal projects**, the data acquirer is responsible for establishing which quality checks are appropriate for the data and validating that the data passes all necessary quality checks. If an appropriate quality check has already been developed for a platform project, this check should be applied to personal project data. However, if no appropriate quality check currently exists, one must be created and approved as described below. While individuals can create quality checks for personal projects, it is recommended that they be created by a set of stakeholders, e.g. a Working Group, a Task Force, or a group of individuals working on related personal projects. Whenever possible, these quality checks should be created before experiments are begun.

Personal project data is expected to be held to the same standards as platform project data. We strive to reach a uniform IBL standard for data quality, ensure that only high quality data is released and enable data to be widely reused. Therefore all quality checks must be developed and documented sufficiently for reuse by other researchers within the IBL. All personal project data must be on the central server prior to data release.

The rigor and documentation level of the quality checks must be approved prior to final curation. New quality checks for platform projects will be evaluated by a relevant Task Force. New quality checks for personal projects will be evaluated by the Project Board. The Executive Board (EB) will make a final decision to approve or reject new quality checks based on the combined recommendation of the quality check creator and evaluator.

Data Documentation

To help the scientific community use the released data, the IBL will share documentation with all released data.

For a data release associated with a paper, the documentation will consist of the code used to make the figures in the paper. For a data release not associated with a paper, the documentation should consist of code for accessing all the fields of the data and metadata and creating a minimal visualization. For more information about sharing code, see the [IBL Code Reproducibility Policy](#).

The data documentation will also contain versioning information and any disclaimer or information about known QC errors, as required by the outcome of the curation process.

For a **personal project**, the data acquirer is responsible for supplying the data documentation. For a **platform project or a release not associated with a project**, the Data Architecture WG

(DAWG) is responsible for supplying the data documentation.

Data format and platform

For final public release, a copy of the data and the documentation will be made publicly accessible on a server that can support large-scale download traffic. Copies of the data can also be shared via other platforms and in other formats. If the IBL is no longer able to provide support for a released dataset, we will attempt to permanently archived at least the processed/compressed in a publicly accessible location that is independent of IBL resources.

For **platform projects**, the DAWG is responsible for (1) determining the data format(s) and platform(s) used for release; (2) formatting the data and (3) releasing it onto the chosen platform.

For **personal projects**, we strive for data storage and release using the same mechanisms as for platform projects. However, DAWG must be consulted on the most appropriate data format and platform to be used, and has the final say on these matters. The data acquirer is responsible for the data formatting and release onto such a platform (the DAWG can provide support if required).

Authorization of release

All curation, formatting, documentation and upload to the approved platform must be completed prior to data release. In addition, the requestor must get permission from the data acquirer for release of any data that the requestor themselves did not acquire. Requestors should use the [Data Release Request Form](#) to document that all steps have been completed.

As part of the publication process for **both platform and personal projects**, the documentation that these steps have been completed should be provided to the [paper review board](#). The review board will examine the documents to ensure that the data is fit for release. Documentation of the request and the approval should be archived along with the other paper review board materials.

For releases not associated with a publication, documentation of the completed steps should be completed by the requestor and provided to the Executive Board for final approval. The EB can inspect the documentation itself or appoint an ad hoc Data Review Board to make a recommendation about whether the data is ready for release. Documentation of the request and the approval should be archived by the data requestor in the EB share drive.

Once the release request has been approved, the requestor should provide documentation of the approval to the Data Architecture Working group to initiate the release. The DAWG will have up to 6 weeks to prepare the release, and requestors should include this preparation time in

their expectations about the timeline for posting a preprint.

Release timeline

Public Commitment:

The International Brain Laboratory will release all data sets within 12 months of collection, or upon acceptance for publication of an associated manuscript, whichever comes first.

For a **platform project**, the Executive Board is responsible for ensuring this deadline is met.

For a **personal project**, the associated [Project Board](#) is responsible for ensuring this deadline is met.

Most data will be released at the time a preprint is posted. However, in order to meet our public commitment, any data that has not yet been released 12 months after approval of the quality control process will be automatically released as a stand alone dataset as per the mechanisms described above.

As per the [Publications Policy](#), If a personal project manuscript uses data from platform projects (see [Alyx User Guide](#)), the platform project data must be released before the manuscript is deposited publicly or submitted for peer review. Any exceptions to this policy will be decided by consent of the Executive Board.

Subsets of data may be publicly released prior to the completion of collection (i.e. not all datasets acquired as part of a project are ready for release), as long as the same release mechanisms are applied.

Data may be publicly released based upon external request, at the discretion of the EB, and with the same release mechanisms applied. In this case, DAWG will be responsible for formatting the data, providing documentation of the included data and metadata fields, and indicating that QCs may not have been applied. Code for accessing data fields and the results of any QCs that have been applied should also be provided if they are available.

Versioning and removing published data

For personal projects, the data acquirer is responsible for any secondary release of similar datasets. The DAWG will provide assistance.

For platform projects, the DAWG is responsible for any secondary release of similar datasets.

Two cases should be distinguished:

1. If previously released data is deemed entirely unusable, the data will be removed from the public database (example: videos with incorrect timestamps, impossible to fix).
2. If an improved version of previously released data becomes available, a revised version of the data will be published, in addition to the original data (example: new version of spike sorter).

In both cases, the updates will be announced in the same channels that were used to announce the original release in the first place.

Data access

Access rights

IBL members can access the data acquired as part of both personal and platform projects, as soon as the data is registered on the centralized database.

Usage rights

Usage encompasses:

- Conducting analysis/processing on the data leading to further scientific discoveries
- Presenting data or derived results at conferences or other form of publication
- Using the data as material for teaching or outreach purposes

Once the data is published, any person can use the data without asking for the consent of the data collector, consistent with the CC-BY license.

Prior to the data being published:

- IBL members can use the data acquired as part of a **platform project** without asking for the consent of the data collector. Such usage must be within the scope of another registered project (either platform or personal) that has met the requirements of the [Project Board](#) and [Publications](#) policies.
- IBL members may use the data acquired as part of another registered **personal project** of which they are not part of **only with the consent of the data acquirer**. Authorship should be discussed with the data acquirer in case of a publication being envisioned.

Citation and publication rights

All data will be shared externally using the [Creative Commons CC BY license](#). This license requires that reusers give credit to the creator and allows reusers to distribute, remix, adapt and build upon the material in any medium or format, even for commercial purposes. No warranties are given to reusers.