



EBRAINS

EBRAINS Data Provision Protocol ("DPP")

version 1.1

This document describes the ethical, legal and technical requirements for EBRAINS Data curation.

Note: This DPP reflects the current EBRAINS curation workflow. This workflow is constantly under development and the DPP will be updated as technical or other changes are made. Please contact the Curation Support curation-support@ebrains.eu for further questions regarding technical considerations.

Table of Contents

Introduction	4
Definitions	4
General Requirements for Data Providers	4
Part 1: Ethical and Legal Requirements	5
1. Provision of Human Datasets	5
1.1 “Identifiable” Human Datasets	5
1.2 Consent Requirement	5
1.3 Anonymised Human Datasets	6
1.4 Data Controllorship	6
1.5 Data Subject Rights	6
2. Providing Non-human Animal Data	7
3. Provision of Model and Software	7
3.1 Models	7
3.2 Software	7
4. Provision of Externally-Hosted Datasets	7
4.1 Human Datasets and Non-Human Animal Data/Datasets	7
4.2 Models and Software	7
5. Data/Datasets Generated Outside the EU	8
5.1 Human Datasets Generated Outside the EU and Third Countries with EC Data Protection Adequacy	8
5.2 Non-Human Animal Datasets Generated Outside the EU	8
5.3 Model Data and Software Generated Outside the EU	8
6. Secondary Data Providers	8
6.1 Human Datasets	8
6.2 Non-Human Animal Datasets	9
6.3 Models and software	9
7. Authorship	9
7.1 Sharing Personal Data of the Author	9
Part 2: Technical Requirements / Recommendations	10
8. Data Custodian (requirement)	10
9. EBRAINS Account (requirement)	10
10. Metadata Provision (requirement)	10
11. Curation Request (requirement)	10
12. Data Descriptor (requirement)	11
13. Data Provision (requirement)	11
14. Data Organisation (recommendation)	11
15. Licences (requirement)	11
16. Embargo Policy (recommendation)	12
16.1 Publishing a Dataset that Accompanies a Peer-Reviewed Journal Publication	12

Version

Version	Description	Date
1.0	Initial version	2021-11-11
1.1	Yearly DPO review 2021	2021-11-28

Introduction

Capitalised terms and definitions used in this Policy have the same meaning as in the Terms of Use governing the EBRAINS Services, unless another definition is explicitly provided for specifically for the context of this Policy.

Definitions

Data Providers: The individuals who provide Data to the EBRAINS research Platform or Services.

Pseudonymization: The processing of Personal Data in such a manner that the Personal Data can no longer be attributed to a specific Data Subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the Personal Data are not attributed to an identified or identifiable natural person.

Pseudonymized Data: A Human Dataset which has been the subject of Pseudonymisation provided by the Data Provider for making it available via the EBRAINS Platform and/or Services.

Research Participant: An individual whose Personal Data form part of Human Datasets provided by Data Providers to the EBRAINS Platform or Services.

General Requirements for Data Providers

All EBRAINS Data Providers must:

- agree to the [EBRAINS General Terms of Use](#)
- agree to the [Knowledge Graph Terms and Conditions](#)
- choose [a license for their dataset](#)
- confirm that the Data are a product of their own work, and cite any related publications, or provide information on the source(s) if the data are not generated by them as per the applicable citation rules
- agree to the [requirements for authorship](#)
- have an active EBRAINS account as per the Access Policy
- ensure that any provided metadata are complete and of high quality
- **consult and agree with the sections below for specific ethical and legal requirements relating to the Data submitted for curation.**

Part 1: Ethical and Legal Requirements

The following section details the ethical and legal requirements for the following types of Data.

1. Provision of Human Datasets

Data Providers submitting Human Datasets for use on the EBRAINS Platform and/or via the Services must:

- a. Provide **copies of ethics approval for the collection of the Data from a local or national research ethics authority**,
- b. Identify and demonstrate an **adequate lawful basis for the processing of the Data**,
- c. Be aware that **all Pseudonymised Data will be shared via the EBRAINS Human Data Gateway** which provides additional security for the data.

Any supporting documentation which is not in the English language is to be provided with a short but comprehensive summary in English. All documents submitted will be assessed by the EBRAINS Curation Request Review Committee.

1.1 “Identifiable” Human Datasets

We consider all Data originating from Data Subjects, thus individual living human beings, as Personal Data which fall under the provisions of GDPR.

Post-mortem human data (data relating to deceased human beings) do not fall under the protection of the GDPR.

For Personal Data, we distinguish between pseudonymised Personal Data (potentially re-identifiable) and anonymous statistics (irreversibly anonymised data which are proven not to be eligible for any re-identification) derived from Personal Data in accordance with Recitals 26 and 28 of the [EU General Data Protection Regulation \(GDPR\)](#).

For all pseudonymised Personal Data, EBRAINS allows data sharing on an access-controlled basis on the EBRAINS Platform. Access to pseudonymised Personal Data is provided only to Users with an active EBRAINS account and Registered Access as per the Access Policy.

Publishing Human Datasets to make them available to Users with Registered Access is subject to the following conditions:

- The Data Provider must include thorough documentation of consent and approvals from a competent research ethics authority for the data collection.
- The Data Provider must ensure that Data provision to EBRAINS complies with the Law and the GDPR.
- The Data Provider must implement the pseudonymisation of that Data while respecting the need not to impair the scientific value of the Human Dataset.

1.2 Consent Requirement

Data Providers of Human Datasets must ensure that prior written informed consent has been obtained from all Data Subjects for the collection of the Human Dataset as well as for the sharing of such Human Dataset on international research infrastructures such as the EBRAINS Platform.

This consent must also include consent to participate in the study from which the Human Dataset is obtained, as required by the principles of the [Nuremberg Code](#), [the Declaration of Helsinki](#) and [the Belmont Report](#).

The information notice provided to the Data Subject must include comprehensive information about the handling of incidental findings (findings that may occur during analysis but are unrelated to the goals of the research). Where consent is not used as the legal basis for processing, the Data Provider will be asked to identify relevant basis for processing.

1.3 Anonymised Human Datasets

For the use of anonymous statistics derived from Human Datasets, the GDPR does not apply.

Publishing Anonymised Human Datasets to make them available to Users with Registered Access is subject to the following conditions:

- The Data Provider must ensure that Data provision to EBRAINS complies with the Law and the GDPR regarding anonymisation requirements.
- The Data Provider must implement and ensure total, irrevocable and definitive irreversible anonymisation of the Personal Data in the Human Dataset.

1.4 Data Controllership

Data Providers must be aware of their legal [responsibility](#) to establish a legal basis for processing, and the ethical demands to maintain data protection, privacy and confidentiality of the Personal Data of Data Subjects.

As persons or entities which define and control the purposes and essential means of processing of the Data, the Data Providers remain *joint data controllers* with EBRAINS and as such are jointly liable in cases of non-compliance with data protection regulations such as the GDPR. This means that procedures to address data protection, privacy (including pseudonymisation of Human Datasets), information security, security and provisions for Data Subject access requests must be put in place by the Data Providers prior to the submission of any Human Dataset.

Further explanations of the Data Provider's responsibilities and liabilities are contained in the [EDPB guideline on data controllership](#).

1.5 Data Subject Rights

The responsibility of meeting any applicable requirements around Data Subjects' rights and requests such as the rights to access, rectification, erasure/withdrawal of consent, restriction of processing, data portability and object, in light of research exemptions remains with the Data Provider.

When such requests are made directly to EBRAINS, Data Subjects will be directed to the specific Data Provider(s) responsible for the Human Datasets in question.

Further information on Data Subject rights can be found on the [European Commission website](#).

2. Providing Non-human Animal Data

All animal data provided to EBRAINS must be generated through research that meets the applicable ethical and legal, international, national or regional standards such as EU Directive [2010/63/EU](#) and the [Three Rs principles](#).

Evidence of all necessary ethics approvals must be provided to EBRAINS by the Data Provider.

3. Provision of Model and Software

3.1 Models

Submitted models should be a product of the Data Provider's own work. If the Data Provider has built on existing code, they must confirm that they are respecting any associated software licences.

3.2 Software

Data Providers must confirm that the software for which they submit metadata is their own work, or clarify on whose behalf they submit the metadata. If they or the person(s) on whose behalf they submit metadata have built on existing code, they must respect any associated software licenses.

4. Provision of Externally-Hosted Datasets

EBRAINS provides an option to share only the metadata of certain Data through the EBRAINS Knowledge Graph. In those instances, EBRAINS retains joint responsibility with the Data Provider for the metadata hosted on the ERAINS RI, but not for the Data itself which is hosted externally and remains the responsibility of the Data Provider only.

4.1 Human Datasets and Non-Human Animal Data/Datasets

Data Providers can choose to host the Data externally, and only store the metadata on the EBRAINS RI which goes through the Data curation request mechanism. Data controllership and legal responsibility for externally hosted Human and Non-Human Animal Data remain solely with the Data Provider.

Access to these Datasets can only be requested through the external system where the Datasets are stored, not via the EBRAINS RI.

4.2 Models and Software

EBRAINS RI can share metadata of externally hosted models and software. There are no specific ethical requirements in those cases.

5. Data/Datasets Generated Outside the EU

Any Human or Non-Human Animal Data or Datasets generated outside the EU must adhere to all relevant EU guidelines and the above relevant chapters of this DPP.

In particular, Data Providers must:

- Provide copies of ethics approvals and other authorisations or notifications to EBRAINS,
- Confirm that the activity could have been legally carried out in an EU country (for instance, an opinion from an appropriate ethics structure in an EU country) ([as per the EU guidance on ethics self-assessment](#)),
- Respect the specific requirements as detailed in articles 5.1-5.3 below.

5.1 Human Datasets Generated Outside the EU and Third Countries with EC Data Protection Adequacy¹

Any Human Dataset generated outside the EU or non-EU countries with EU data protection adequacy must adhere to all relevant EU guidelines. In particular, Data Providers must:

- provide copies of ethics approvals and other authorisations or notifications, including copies of consent forms and other supporting documentation.

5.2 Non-Human Animal Datasets Generated Outside the EU

Any Non-Human Animal Datasets generated outside the EU must adhere to all relevant EU guidelines. In particular, Data Providers must:

- provide copies of ethics approvals and other authorisations or notifications to EBRAINS,
- where genetically modified non-human animals are involved, copies of GMO authorisations detailing the harm-benefit analysis must be provided.

5.3 Model Data and Software Generated Outside the EU

There are no particular additional requirements for models or software generated outside the EU.

6. Secondary Data Providers

6.1 Human Datasets

Data Providers submitting pseudonymised Data which they did not collect directly from the Data Subjects must provide information on the source of the Data and identify and demonstrate adequate lawful basis for processing in accordance with article 6 of the GDPR (consent, contract, legal obligations, vital interests, public task or legitimate interests).

When the Data are identified as special category Data or sensitive Data, a further basis from article 9 of the GDPR is also required. This may be: explicit consent; employment; vital interests; made public by the Data Subject; carried out for not-for-profit organization; legal claims, preventive or occupational medicine; public interests in the area of public health; archiving purposes in public interests, scientific or historical research purposes or statistical purposes; substantial public interest on the basis of EU or State law.

¹ https://ec.europa.eu/info/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

6.2 Non-Human Animal Datasets

Data Providers submitting Non-Human Animal Datasets that were not generated by their own research should ensure that the Data they are submitting have been collected in accordance with EU Directive 2010/63/EU on the protection of animals used for scientific purposes or the [Guide for the Care and Use of Laboratory Animals published by the Institute for Laboratory Animal Research](#).

6.3 Models and software

Data Providers submitting models or software which they did not generate themselves should ensure they appropriately cite the original creator of the models or software. Otherwise there are no particular requirements associated with secondary data of this type.

7. Authorship

7.1 Sharing Personal Data of the Author

The Data Provider agrees to publish her/his name as an author/custodian of the Data, as well as the names of the contributors to the Data on the public EBRAINS Knowledge Graph (KG) Search landing pages. The custodian of the Dataset is responsible for acquiring informed consent from the contributors of the Data to publish their names along with the associated dataset on the EBRAINS Platform.

For the Data Descriptor (a structure containing information that describes data) connected to each Dataset, affiliation and email addresses are also requested. Although not required, Data Providers are urged to add this information as it will improve the quality of the Data publication.

The EBRAINS Curation Team will request confirmation before publishing the Dataset on EBRAINS RI, allowing the Data Provider to have full control over which Dataset her/his name is connected to.

Part 2: Technical Requirements / Recommendations

8. Data Custodian (requirement)

The Data Provider must name a Data Custodian for the provided Data and metadata. The Data Custodian is responsible for the content and quality of the Data and metadata, and is the person to be contacted by EBRAINS CS in case of any misconduct related to the Data. It is the obligation of a Data Custodian to keep EBRAINS informed about changes in the contact information of the authors of the Datasets provided by them.

9. EBRAINS Account (requirement)

Data Providers are asked to upload their Data to the EBRAINS RI via respective dedicated data buckets issued by the Curation Services through the EBRAINS [Collaboratory](#) Bucket service which currently stores the Data at one of the [Fenix RI sites](#).

Before curated Data are published (made publicly available) through the EBRAINS KG, the Data Provider has to review and approve the correspondingly integrated metadata in the Database. This review is facilitated by the EBRAINS KG Preview Service that allows Data Providers to view the KG landing pages of their Data before they are publicly released. Please note that this service only simplifies the review of the registered metadata by the Data Provider, but it is not needed to finalise the curation process and publish the Data.

The usage of the EBRAINS Collaboratory Bucket service and KG Preview service require the Data Provider to have an EBRAINS user account. Details on how to apply for an EBRAINS account can be found in the Access Policy and the register form on the EBRAINS website: <https://ebrains.eu/register>. Details on the usage conditions and restrictions of EBRAINS Services can be found in the Terms of Use and Access Policy.

10. Metadata Provision (requirement)

Data Providers who want to publish their Data via the EBRAINS Knowledge Graph (KG) have to provide a corresponding set of Metadata compliant with the running version of the openMINDS metadata schemas. All versions of the openMINDS metadata schemas are defined and documented on GitHub: <https://github.com/HumanBrainProject/openMINDS>.

The different ways of providing openMINDS-conformant metadata to the Curation Services will be listed and described on the [EBRAINS Curation collab](#).

11. Curation Request (requirement)

The Data Provider is required to make an official Curation Request via the EBRAINS Curation Services (CS). This request can be made directly to the EBRAINS CS by filling out the [Curation Request web-form](#). Requests for curation will be reviewed by the EBRAINS Curation Request Review Committee.

Information concerning the Curation Request process can be obtained by contacting the Curation Services: curation-support@ebrains.eu.

12. Data Descriptor (requirement)

The Data Provider is required to deliver a human-readable Data Descriptor that accompanies the provided Data. If the Data Provider did not already publish a Data Descriptor in a peer-reviewed journal, the EBRAINS Curation Services (CS) provide an EBRAINS Data Descriptor template on the Drive of the [EBRAINS Curation collab](#). Please note that the Data Provider does not have to deliver a finalised version of the Data Descriptor together with the initial request for Data Curation.

13. Data Provision (requirement)

Data that is assigned with an EBRAINS DOI has to be hosted on the EBRAINS Data Storages (DS).

In addition, software source code and model code/Data should be hosted on a public version-controlled repository (e.g., GitHub, GitLab).

For sensitive/special category Human Data, the EBRAINS CS also accepts that the Data Provider stores these Data on a locally managed and secure long-term repository with strict access restrictions.

The Data Provider should be aware that all externally hosted sensitive/special category of Personal Data that are referenced by the EBRAINS KG will not receive an EBRAINS DOI.

14. Data Organisation (recommendation)

EBRAINS has no requirements on how Data should be organised on delivery, but the Data Provider is encouraged to deliver Data in a consistently organized manner. Best practice guidelines and potential standards for Data organization can be found on the [EBRAINS Curation collab](#).

The Data Provider should be aware that the EBRAINS Curation Services (CS) have the right to refuse poorly organized Data or can request changes to improve the organization of delivered Data on a case-by-case basis.

15. Licences (requirement)

The Data Provider is required to define an appropriate licence for the provided Data. The EBRAINS Curation Services (CS) accept the following licenses:

- For Animal Data or Human Data or metadata, the EBRAINS CS accept any Creative Commons (CC) licenses (except no-derivative features). The appropriate license can be determined with the help of the CC license chooser:
 - <https://creativecommons.org/choose/>
- For Pseudonymised Data requiring additional protection and shared via the EBRAINS Human Data Gateway, CC licences will not apply. The KG datacard presenting the Data will have the following statement: *"The use of this Dataset requires that the user cites the associated DOI and adheres to the conditions of use that are contained in the Data Use Agreement."*
- For models, the EBRAINS CS accept any open source licenses or CC licenses (except no-derivative features). Information and a list of appropriate open source licenses can be found here:

- <https://opensource.org/>
- <https://choosealicense.com/>
- For software, the EBRAINS CS accept any type of license, although the deployment of open source licenses is strongly recommended.

16. Embargo Policy (recommendation)

There is no embargo policy at the moment for Data Providers requiring an embargo on their Datasets. They should seek information from EBRAINS Curation Services on what options are available.

16.1 Publishing a Dataset that Accompanies a Peer-Reviewed Journal Publication

If you are in the process of submitting your Dataset for publication along with a manuscript for a peer-reviewed journal, you have the different options outlined [here](#).