# Brain MRI data sharing guide

**Start: I want to share identifiable brain MRI data**

Deidentify the MRI data and possible accompanying data as much as possible, e.g., deface the MRI data, use random identifiers, remove names, dates and contact information, etc.

I cannot do this

Are you using informed consent as a legal basis to share personal data?

No, I can't (anymore)

No, I don't need to

Yes

1. my institution considers deidentified MRI data anonymous
2. at my institution, we use a different legal basis for personal data sharing

Children (<16)

To what purposes of data sharing did your participants consent?

No MRI data sharing

I'm not sure

I'm certain

Public data sharing: all purposes, also outside EU

Data sharing under conditions

My sample consists of …

My sample consists of …

Vulnerable participants

Vulnerable participants

Healthy adults (> 16)

Healthy , non-vulnerable participants

Talk with a data steward, privacy officer, information manager, or trial officer at your institute

**Deidentified brain MRI data - publicly**
Provided you have a legal basis (e.g., explicit informed consent) and you don't have unique data from vulnerable populations, you can share your deidentified raw MRI data publicly. Please note that sharing data publicly is irreversible, so when in doubt, please contact a professional at your institute.

Possible platforms:
- Openneuro
- EBRAINS
- Open Science Framework
- Zenodo
- 4TU data
- DANS EASY
- TraIT
- Institute-specific repositories, e.g., DataverseNL

**Deidentified brain MRI data - access restrictions**
Provided you have a legal basis (e.g., explicit informed consent), you can share your deidentified raw MRI data with access restrictions, e.g., by granting access after:
- signing a data use agreement
- accepting terms or using a restrictive license

The restrictions may depend on the content of the informed consent form, your study population or your supporter's advice. Possible platforms:
- EBRAINS
- 4TU data
- DANS EASY
- TraIT
- Institute-specific repositories, e.g., DataverseNL
- During research (not FAIR): SURFdrive / SURF filesender / SURF Research Drive

**Derived brain MRI data**
If your data cannot be deidentified or you have no legal basis to share personal data, you cannot share raw MRI images.

However, you can still make your data useful in the following ways:
1. NeuroVault Share spatially normalized or otherwise preprocessed individual MRI data (excluding transformation matrices)
2. NeuroVault: Share statistical values from (individual) regions of interest
3. NeuroVault: Share group maps, e.g., t-maps of contrasts
4. Share a synthetic dataset
5. Use federated/distributed learning, e.g., with COINSTAC, ENIGMA, PySyft, , OHDSI, Personal Health Train

AND

When in doubt, talk with a data steward, trial officer or information manager | Preferably use BIDS | Always aim to include a persistent identifier (e.g., DOI) to the (meta)data in your manuscript.

# About this guide

With the increasing amount of data sharing initiatives, researchers collecting **brain MRI data** cannot stay behind. However, many EU researchers do not know:

- where to start
- what they can share and where
- where they can find information or support

For these reasons, we developed this MRI data sharing guide. The guide is meant for researchers who want to share their brain MRI data, and who are new to data sharing. The current document is a supporting document to be used in combination with the Brain MRI data sharing guide.

**Scope**

This flowchart and documentation are currently most applicable to researchers in the Netherlands. It might also be helpful for researchers from other EU countries, but local and national regulations could differ. We encourage reusing and adapting this flow chart to local circumstances.

Back to top

# What is identifiable data? [1/2]

The General Data Protection Regulation (GDPR) distinguishes 2 types of data:

1. **Anonymous data**: data that is not in any way traceable to an individual. Strictly speaking, re-identification should not be possible with anonymous data. Importantly, anonymous data do not fall under the GDPR. Therefore, privacy-wise, **sharing anonymous data publicly is possible without requiring a legal basis**.

2. **Personal data**: data that can identify individuals **with reasonable means**:

- Directly identifying information: for example, name, picture, or contact information
- Indirectly identifying information:
    - If your data are so **unique**, they can only refer to one person. For example, if your data includes a man 105 years of age from a small Dutch village, those living in that village will likely be able to say: 'Hey, that's Bob!'
    - If combining (public) datasets can lead to identification through a "**linkage attack**", i.e., linking de-identified data with data that contain identifying information. For example, the combination of zip code, date of birth and gender can already identify 87% of the United States population (source). An attacker could use public data (e.g., a voter list) that contains names, gender, zip code and date of birth to cross-reference and identify the individuals (see link).
    - If **re-identification** is still possible. Data that look anonymous but can still re-identify individuals are called pseudonymous. One example is when the name-number key still exists. A strict interpretation of the GDPR states that the data can **never** be anonymous as long as this key file exists. A more liberal interpretation however is that the data can be considered anonymous if the key file is not shared and is stored securely and the data are anonymous in all other aspects as well.

Back to top

Next page

# What is identifiable data? [2/2]

**All personal data fall under the GDPR**. In order to be able to process and share these data, you need a legal basis. In research, this is most often **informed consent** (but there are other legal grounds to consider as well).

## Are my data personal?

It is sometimes difficult to estimate whether your data are anonymous (see next section). Generally, you can follow the following steps:

1. **Deidentify** your data, e.g., remove directly identifying information (see also)
2. After deidentification, estimate whether it is possible to re-identify someone using **reasonable means**
3. Think about whether these means would be reasonable **within a few years' time** as well, e.g., considering the speedy development of Artificial Intelligence and Big Data.
4. Think about how **unique** your study sample is and evaluate the risk of identification for this population
5. Before making a decision on anonymity, consider all of the above together with a **privacy officer** or data steward at your institute.

See also this presentation by Enrico Glerean from Aalto University (2020-10-27).

Back to top

How is brain MRI data identifiable?

# How is brain MRI data identifiable?

Raw brain MRI data can be identifiable in the following ways:

- **DICOM headers and .PAR files** (Philips scanners) can contain personally identifying information such as name, date of birth, date of data collection, etc.
- Structural MRI scans often contain **facial features**
- Each brain is unique. Studies have found that **both anatomical and functional patterns can act as a fingerprint** (Abramian & Eklund, 2018, *preprint*; Ravindra & Grama, 2019, *preprint*; Peng et al., 2019, *preprint*; Finn et al., 2015). If a database linking MRI data with a name is published publicly elsewhere, releasing deidentified MRI data of the same participant can lead to them being identified through combination with the named data.
- If shared in combination with **other data** that can lead to identification (see link), e.g.:
  - Identifiers, such as name, ID number, location data, phone number, IP address, etc.
  - Demographic information, e.g., zip code, date of birth, gender
  - Sensitive personal data: genetic, biometric and health data, racial and ethnic origin, political opinions, religious or ideological convictions, trade union membership
  - Open text fields in questionnaire data
  - Highly unique phenotypic and clinical data

Because brain MRI data is unique biomedical data and a unique profile can be generated from them, we consider both raw **and** defaced MRI data **personal data** (see this pseudonymization guide for more information on different levels of pseudonymization). **Please note** however that policies and opinions of data protection officers may differ on this matter: some do consider deidentified MRI data *an sich* anonymous. Contact your data protection officer or privacy manager when in doubt.

^ Back to top

> What are reasons not to share brain MRI data?

# What are reasons not to share brain MRI data?

Valid reasons not to share raw MR images of the brain are:

- When data are not considered anonymous and you have no legal basis to share them, e.g., you have not obtained informed consent
- There is no way to share the data securely (e.g., if the data receiver cannot guarantee safe storage and processing)
- Your institution does not allow sharing personal data with others at all (this does not happen often however)
- There are ethical concerns due to the nature of the research (e.g., conducted among vulnerable or high-risk groups)

If your reason(s) is/are not in this list, you can likely share the raw and/or derivative MRI data with the ways described in this guide. If they are, please see this section.

Back to top

What precautions do I need to take for brain MRI data?

# What precautions do I need to take for brain MRI data?

Because brain MRI data is unique biomedical data, we consider brain MRI data **personal data**, although policies and opinions of data protection officers may differ on this matter.

To share brain MRI data safely, you need to take the following precautions:

- Obtain participants' informed consent to share their data (see section). Alternatively, you could consider using a different legal basis, ask your data protection officer or privacy officer for help.
- Deidentify your brain MRI data (see section).
- Preferably use a repository that stores data in the EU. This is because the EU is bound to the GDPR and other countries may not be.
    - Because data will be hosted on an external server (in- or outside of the EU), the institution responsible for the data needs to have a processing agreement with the repository.
    - If the data are hosted outside of the EU, additional agreements need to be set up with the repository on top of the processing agreement. This is necessary to ensure that the repository will handle the data in line with the GDPR.
- For personal data, it is recommended to use a Data Use Agreement before sharing (see section). This is because the institution that is considered to be the data controller remains responsible for the data and may want to control what recipients are permitted to do with the (personal) data.

Back to top

How do I deidentify my brain MRI data?

# How do I deidentify my brain MRI data?

1. Remove identifying header information in DICOM or PAR files, such as name, date of birth, date of data collection, etc.
2. Skull strip or deface anatomical images to get rid of facial features
3. Anonymize or pseudonymize **accompanying data** as much as possible (see also)
    - Aggregate data, for example:
        - Date of birth → Age → Age category
        - Clinical score → Clinical category
        - Postal code → Place of residence → Province / Living area
    - Remove directly identifying information, for example:
        - Name
        - Date of measurement
        - Pictures
        - Extreme or unique values that can be identifying (for example extremely old, extreme clinical scores, etc.)
4. Use **random identifiers** (e.g., subject codes) and/or randomize these subject codes before data sharing
5. Do **not include keyfiles** such as participant-identifier links, identifier mapping and transformation matrices in the shared dataset.

Back to top

What needs to go into my informed consent form?

# What needs to go into my informed consent form?

If you are planning to share personal data with others, include at least the following information in the data sharing section of the information letter/privacy notice:

- What personal data will be shared?
- For what purpose will the personal data be shared? E.g., what type of research(er) can reuse the data (e.g., “future research in the field of functional MRI”)?
- With whom will the personal data be shared? E.g., publicly, only with researchers from a university, within the EU, outside of the EU, upon request, etc.
- Pay special attention if you collect and want to share sensitive personal data

Please refer to a GDPR-compliant consent form by the Open Brain Consent initiative here (includes several translations).

**Please note**:

- contact your privacy officer or information manager to make sure you are complying to all necessary regulations and policies.
- if you study children under 16, informed consent is the only legal basis allowed for sharing personal data

Back to top

When does my data require extra protection?

# When does my data require extra protection?

We recommend not sharing data publicly and introducing additional access restrictions in the following situations:

- When you have collected data from high-risk or vulnerable individuals, such as:
  - children (<16)
  - clinical populations (especially when individuals show unique abnormalities)
- When your data are easily identifiable, e.g., because they are highly unique or contain identifying additional information (see section) and pseudonymizing them would lead to data loss.

Back to top

When do I need a data use agreement?

# When do I need a data use agreement?

## What is a DUA?

A Data Use Agreement (DUA; sometimes: data sharing agreement) is a legal document between two parties (often 2 institutions) in which an agreement is made about which data can be used by whom and how. In a way, a DUA can function as a substitute for a license if you just share the data case by case.

Use a DUA when:

- you want to take additional measures to better protect your participants' privacy.
- you want to take care of ownership issues.

**Please note**

- A DUA is **not** a legal basis to share personal data: either obtain informed consent or use a different basis.
- Make sure an **authorized signatory** of your institution signs the agreement. In most cases, the institution claims ownership over all collected research data. Therefore, in most cases, someone who is authorized to make decisions on behalf of the institution should sign the agreement, **not** individual researchers.

## Content

A DUA often contains at least the components mentioned in this template. It is important to have your legal department take a look at the agreement:

- formulations may differ per institution
- your institution may have model agreements available to use as a template
- all legal agreements need to be approved by them.

Back to top

What license should I use for MRI data?

# What license should I use for MRI data?

Not all commonly used licenses are appropriate for data(bases). Some recommended licenses are:

- **Sharing data publicly: CC0:** your data will become public domain. This allows people to do anything they want with the data, technically even without acknowledging the source.
- **Sharing data publicly or on request: CC-BY 4.0:** allows people to do anything they want with the data, while acknowledging the source.
- **Sharing data case by case: DUA**: a Data Use Agreement (see section) can be seen as a specific license that is designed for data and can take into account that, besides the rights of the licensor (e.g., the researcher or university), there may also be rights of the participants whose data is included, for example relating to re-identification. You could call an "agreement" or a "contract" between data provider and data downloader a "license", but calling it a "data use agreement" makes it more explicit that it is not about the (re)use of a creative work (like written text or code by an author) but reuse of measured/observed data (source: Open Brain Consent)

Back to top

Where can I share MRI data?

# Where can I share MRI data? [1/3]

If there is no suitable solution in this overview, try searching for one here.
Please note that storage of personal data within the EU is needed to comply with the GDPR.

### International services

| Name | Description | Storage location | Access restrictions | Storage limit | Costs |
|---|---|---|---|---|---|
| Openneuro | share raw (defaced) MRI datasets publicly using the BIDS format | USA* | - Public<br>- Embargo (6 months), after which data become public (CC0) | No limit | Free |
| EBRAINS | share raw (defaced) MRI data, requires institution subscription, reviewed before publication | Switzerland | - Public<br>- Embargo<br>- Restricted access (registered users) | Unknown | Unknown |
| Open Science Framework | Project management | USA, Germany, Canada, Australia | - Public project<br>- Private project<br>- Private link | - Public project: 50GB<br>- Private project: 5GB | Free |
| NeuroVault | share unthresholded statistical maps, parcellations, and atlases | USA* | - Private collection<br>- Private link<br>- Public collection | No limit | Free |
| Zenodo | General purpose repository for all research output | Switzerland (CERN) | - Public<br>- Embargo<br>- Closed (can specify license in description field) | 50 GB | Free |

Back to top

To national services

# Where can I share MRI data? [2/3]

**National services**

| Name | Description | Storage location | Access restrictions | Storage limit | Costs |
|---|---|---|---|---|---|
| 4TU data | international data repository for science, engineering and design | The Netherlands | - Open<br>- Embargo<br>- Restricted<br>- Metadata-only (incl. link to original location of data)<br>- Private URL | Up to 1 TB | Free up to 10 GB,<br>1 TB free only for affiliated researchers |
| DANS EASY | Dutch generic archiving service | The Netherlands | - Public<br>- Open for registered users (DANS license)<br>- Restricted: request permission, possible to add additional conditions<br>- Other access: accessible via another repository | No limit | Free up to 100 GB for individual researchers |
| TraIT by Health-RI | Dutch imaging platform for clinical data management, processing, collaboration and archiving. | The Netherlands | - Public<br>- Protected: project description public, request access to data<br>- Private: only visible to logged in project members | No limit | See pricing |

Back to top

To institute-specific services

# Where can I share MRI data? [3/3]

### Institute-specific services

| Name | Description | Storage location | Access restrictions | Storage limit | Costs |
|---|---|---|---|---|---|
| Yoda (SURF) | Dutch service for data management, archiving and publishing | The Netherlands | Public, Metadata only | Institution-dependent | Free for affiliated researchers |
| DataverseNL | belongs to publication, guidelines may differ between institutions | The Netherlands | Public, Available on request, Allows formulating custom terms and conditions, Private URL | max. 9.3 GB per file | Free for affiliated researchers |
| EUR data repository | sharing research data for EUR researchers | Figshare: AWS Ireland | Public, Private, Private URL, Allows defining specific terms of use | 5 GB per file, 20 GB private space, unlimited public | Free for affiliated researchers |
| Donders repository | Repository for the Donders Institute to deposit data | Nijmegen, the Netherlands | Open, Restricted (login required), Custom license | No limit | Free for affiliated researchers |
| AmsterdamUMC db | Medical database for sharing clinical ICU data, not usable for everyone | Amsterdam, the Netherlands | Restricted | Unknown | Unknown |
| OMERO | Manage and share imaging data (originally for microscopy data) | Leiden, the Netherlands | Private, Read-only, Read-annotate, Read-write | | Unknown |
| SURFdrive | Dutch service for collaborating in the cloud. Suitable for sharing case by case, not FAIR. | The Netherlands | Create links, Share folder(s) | 500 GB | Free for affiliated researchers |
| SURF research drive | Dutch service for collaborating on in the cloud (for teams). Suitable for sharing case by case, not FAIR. | The Netherlands | Create links, Share folder(s) (read, write (delete, edit, move), reshare) | institution-dependent | Institution-dependent |
| SURF filesender | Dutch service for securely sending files (replacing email servers), not FAIR. | The Netherlands (saved 14-21 days) | Download via email or URL, Encrypted (password-protected) | 1 TB (2 GB encrypted) | Free for Dutch researchers |

Back to top

How to make my brain MRI data useful without sharing raw data?

# How to make my brain MRI data useful without sharing raw data? [1/2]

If you do not want to share any raw brain MRI data at all, you can still make them valuable in the following ways:

### 1. Share derived MRI data

Instead of sharing raw MRI data which can contain sensitive information, you can always share processed/derived MRI data, such as:

- spatially normalized or otherwise preprocessed MRI data, provided the transformation matrices (containing the warping parameters for normalization) are excluded from the shared dataset. This is because such matrices can be seen as a re-identification key to get back to identifiable MRI data.
- statistical (or other) values from (individual) regions of interest
- group maps such as unthresholded statistical maps (via NeuroVault). This is highly recommended, because sharing group maps this way allows for more accurate meta-analyses and makes peer reviewing your manuscript much easier.

### 2. Synthetic data (functional MRI data)

Creating a synthetic dataset can be useful to capture the statistical idiosyncrasies of your real dataset. This synthetic dataset can be used to reproduce the results of your analysis, without violating any privacy or intellectual property regulations. Synthesizing MRI data relies on having a very clear and precise “forward model” of what the BOLD response of a given voxel will be, depending on some response parameter of that voxel. Therefore, synthetic MRI data can also be used for power analysis. Read more here and here.

Back to top

More ways to do this

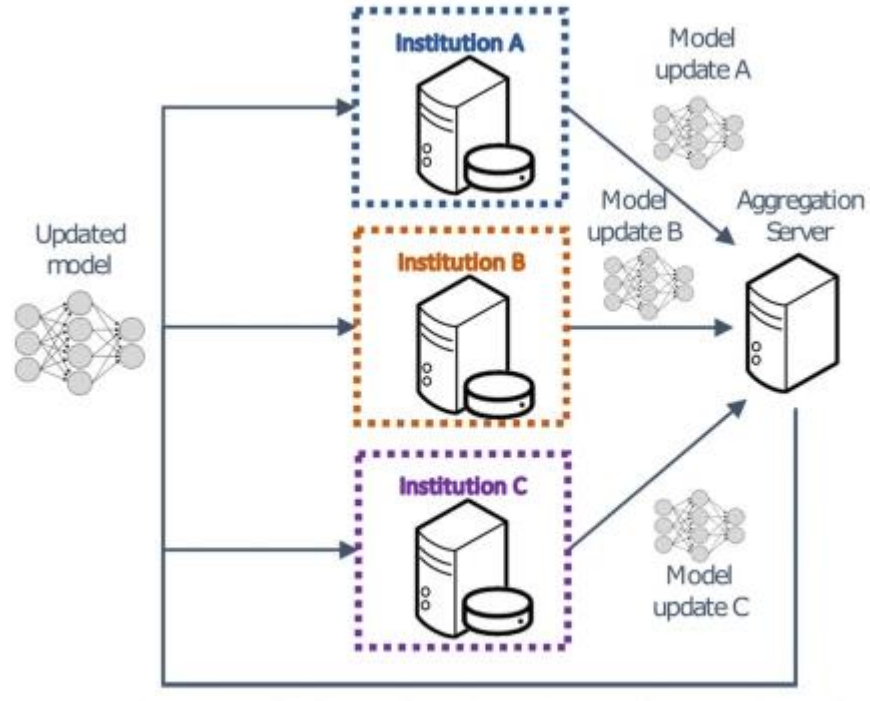
# How to make my brain MRI data useful without sharing raw data? [2/2]

### 3. Federated/distributed learning

Federated or distributed learning arises from the field of Artificial Intelligence and relies "on the principle of remote execution—that is, distributing copies of a machine learning algorithm to the sites or devices where the data is kept (nodes), performing training iterations locally, and returning the results of the computation (for example, updated neural network weights) to a central repository to update the main algorithm." (Kaissis et al., 2020). This means that you do not move your data, while still providing valuable information about it.

Some federated learning tools and projects:

- COINSTAC
- PySyft
- ENIGMA consortium: Consortium with several working groups. Share pre- and post-processing analysis scripts, the leading site will conduct the meta-analysis
- OHDSI (Observational Health Data Sciences and Informatics): collaborative to bring out the value of health data through large-scale analytics.
- Personal Health Train, part of Health-RI (official website here)




(b) Data-private Collaborative Learning using Federated Learning

*Source: Sheller et al., 2020*

Back to top