university of groningen

Digital Competence Centre

Status: revised (August 2023)

Terms of Use for Datasets

The FAIR data-principles (Findability, Accessibility, Interoperability, and Reusability) help researchers make their data 'as open as possible and as closed as necessary'. This guide focuses on the "A" of FAIR—accessible under well-defined conditions. You will learn more about the Terms of Use for both open data and data with restricted access, as well as how others can access your restricted data.[1]

Table of contents

Why do I need to specify Terms of Use?

It is always important to define how your research data can be reused by others. When there are no limitations, the data can become openly available to everyone. However, there may be reasons to restrict access to the data, for instance to protect personal data or to respect third party licences. No matter whether your research data is open or restricted, the Terms of Use of the dataset should always be defined. The Terms of Use are part of the metadata. As such, the information about the intended use of the data is always publicly available, even if access to the dataset is restricted.

[1] Note that the UMCG has its own policies, rules and regulations (and legal department, Loket Contract Research) and these overrule any recommendations made in this document. Researchers of the UMCG are thus encouraged to go to the UMCG Research Toolbox (login with UMCG account).

When do I need to think about Terms of Use?

At the start of your research project, you should already consider how you want to make the data available for reuse during and after your project. Determine at an early stage if there are legitimate or practical reasons to restrict access to the data, inform (potential) collaborators and/or other third parties involved in the research about the intended future use of the data, and inform your data subjects in case your research involves the collection of personal data (e.g. by means of the consent form and information sheet). All these considerations should be specified in your Research Data Management Plan (RDMP).

Near the end of your research project, you will need to prepare the dataset for archiving and/or publishing. For example, if the dataset contains personal data, you may need to de-identify the data before it can be made available for reuse. The degree of de-identification may have implications for your Terms of Use.

How do I set Terms of Use for open and restricted data?

In this section you learn more about terms of use for open data and terms of use for data with restricted access. It is possible to make part of the dataset publicly available, whilst restricting access to other parts. DCC data stewards are available for tailored advice.

Open data

The most straightforward way of publishing your data is to make them openly available via a public trusted repository. By depositing your data and software in a data repository you enhance their findability, accessibility and impact. You not only allow others to validate your research findings, but also enable fellow researchers to 'reuse' your research results.

By depositing your data and software in a data repository you enhance their findability, accessibility and impact. Data repositories such as the UG's default data repository DataverseNL generally allow you to add a machine readable Creative Commons licence to the metadata of your dataset, which specifies under what conditions your data are available and can be reused by others.

Terms of Use for open data

Open data published without limitations and restrictions can be used by any interested party. To maximise the reuse of your research data, our advice is to publish the data using the Creative Commons licence CC0 1.0, which places the data in the public domain. This is also the default licence for research data made available via the UG default data repository DataverseNL. CC0 is recommended in order to optimise the reusability of research data. Alternatively, you could use the CC BY 4.0 licence, which means that attribution should be given. The CC BY licences complicate data reuse because attribution should then also be given to the sources used to create your dataset.

Keep in mind that according to the standards for good research practices (e.g., see Netherlands Code of Conduct for Research Integrity), researchers should always cite their sources accurately. That means that researchers using the data should always give you credit, even if you make the data available under a CC0 1.0 licence.

Data with restricted access

Although data can often be shared openly, sometimes access needs to be restricted. This means that your data is only shared with interested parties under well-defined conditions. There can be various reasons for restricting access to research data which are not mutually exclusive. That is, a single dataset can have multiple reasons for restricted access. There can be both practical reasons (e.g., related to size and costs) and legal reasons (e.g., respecting third party licences) for restricted access. You are encouraged to think carefully about the reasons for restrictions.

Size and costs

Practical reasons for restricted access to research data can be related to the size of the dataset and the costs for data publication. Most data repositories have a maximum size for data publication. Currently, datasets larger than 50 GB usually come with additional conditions or are not supported at all. Furthermore, there are generally costs associated with depositing data in public repositories. If costs cannot be covered or if the dataset is too large to be deposited in a public repository, you are encouraged to archive the data at UG facilities instead (e.g., Y: drive or RDMS). Consider publishing a sample of the dataset in DataverseNL.

Terms of Use

Assuming there are no further restrictions, datasets archived at UG facilities can still be shared with interested parties under a CC0 1.0 (or CC BY 4.0) licence.

Embargo

It is possible that you can make the data publicly available, but that you need additional time to analyse or publish your findings. In the field of astronomy, for example, it is common practice to receive a time window to analyse data from a telescope, after which the data automatically becomes publicly available. The embargo ensures that researchers who created the dataset have time to publish their research without worrying about competition. Depending on the platform used for data publication, you can specify a date or a period of time after which the embargo should be lifted or whether the embargo should be lifted manually. This period generally ranges from six months to three years after depositing the data.

Terms of Use

When the embargo period ends and no other reasons for restricted access apply, the data become publicly available with a CC0 1.0 (or CC BY 4.0) licence.

Personal data

When a dataset contains data collected from human subjects, you will need to comply with the General Data Protection Regulation (GDPR), which prohibits you from sharing personal data openly with the general public. Before publishing your data, it is important to think carefully about whether 1) there is a possible risk of re-identification of the participant, for instance by combining different datasets, due to demographics, or outliers. 2) participants provided consent to the reuse of their data.

Datasets containing personal data can for instance be deposited in the UG default data repository DataverseNL, but access to the datafiles usually has to be restricted, which can be done on the platform. Take into account that DataverseNL is not suitable for highly sensitive personal data. In that case you are encouraged to archive the research data at the Y: drive or RDMS.

Terms of use

Creative Commons licences are not suitable for data containing personal data with access restrictions. Instead, custom Terms of Use have to be set which will largely depend on the consent given by the participants and the degree of de-identification. As such, the custom Terms of Use have to reflect what is allowed according to the informed consent and, more generally, the GDPR. Terms of Use for these data are set in a Data Transfer Agreement (DTA) - see below for more information on the DTA.

Sensitive and/or Confidential data

Data does not necessarily need to be personal data (anymore) in order to be sensitive. There may be circumstances where the public release of data might put research participants, vulnerable groups or the public at risk. Sensitive data can, for instance, include information on domestic energy usage that could possibly be used to determine occupancy patterns in participants' homes. In addition, confidential data could refer to data which reveals private information about a company or could threaten national security (e.g., nuclear research). You are encouraged to archive the research data at the Y: drive or RDMS. DataverseNL is not suitable for highly sensitive or confidential data.

Terms of Use

Creative Commons licences are not suitable for sensitive and/or confidential data with access restrictions. Terms of Use for these data are set in a Data Transfer Agreement (DTA) - see below for more information on the DTA.

Third party licences

If you are making use of data from other sources (e.g., existing datasets or databases), you are encouraged to carefully read the Terms of Use or licence set at the source. One example is the use of X, formerly Twitter, data. Researchers can scrape data from this platform via the API. When you use the API, you agree with the terms in the licence which states that X IDs or user IDs can only be shared with other researchers, and not with the general public.

You should always comply with third party conditions. These apply to the primary dataset and may be different for your processed data. If the licence allows you to make (parts) of the data publicly available, your dataset could be deposited in the UG default data repository DataverseNL. Datafiles usually have to be restricted, which can be done on the platform. Instead, you are encouraged to archive the research data at the Y: drive or RDMS.

Terms of Use

Reuse of primary data cannot extend to what was specified at the source. This may be different for your processed data. You can contact the DCC for tailored advice.

Data Transfer Agreement (DTA)

A DTA—sometimes referred to as a Data Sharing Agreement or a Data Use Agreement—is a legal contract that defines the specific purposes for which the data may be used by the requesting party. As such it is the most comprehensive specification of Terms of Use. The DTA also describes the rights and obligations of both parties involved and sets out the measures for data protection. The UG has its own model DTA that can be tailored for the dataset of your research project. The Privacy and Security coordinator of your Faculty or the DCC can help you with this. Moreover, the DTA always has to be signed by an authorised person. At the UG, this is the Managing Director of your Faculty or the Dean of the Board of the University, depending on the degree of sensitivity of the data.

How to provide access to restricted data?

An access procedure ('terms of access') explains the way in which to request access to research data and how this request will be handled.[2] Depending on where the data is stored (e.g., Y: drive, RDMS, or DataverseNL) and the Terms of Use, the access procedure differs.

[2] This guide does not include information about granting access to committees involved in audits and/or inquiries concerning scientific integrity, as these follow strict procedures.

In any case, you should inform others about the data through, for example, a Data Availability Statement in your publication. Make sure it is clear how interested parties can get in touch with the persons or services facilitating the transfer of the data. This is not necessarily the corresponding author. You should also consider what information is required from the requesting party to facilitate the transfer of the data (e.g., email or address).

A simple procedure can merely consist of a check of the credentials of the requesting party. For instance, if the Terms of Use specify that the data may only be shared with researchers affiliated with a university or not-for-profit research institute. If this is the case for a dataset published in DataverseNL, support staff of DataverseNL at the UG can validate the request on behalf of the authors and grant access to the requesting party via the platform.

If the data is archived on the Y: drive or RDMS, the responsible faculty members (e.g., researchers, ethics board, or other relevant staff members) can validate the request and grant access to the data by sending the dataset to the requesting party through SURFfilesender or Unishare.

Example: facilitating data reuse with a DTA

An example of facilitating the reuse of research data that includes the signing of a DTA is the procedure for the dataset of UG Faculty of Arts researcher dr. Tommaso Caselli: "DALC - Dutch Abusive Language Corpus"[3], which is available under restricted access via DataverseNL. This dataset contains personal data. The Terms of Use for this dataset limit the use of the data for the purpose of conducting not-for-profit scientific research only.

Researchers interested in the dataset can request access via DataverseNL. The administrators and curators of DataverseNL at the UG are notified of the request and will contact the requesting party to ask for additional information such as their affiliation, line of research, and reason for the request. The curators will subsequently check the credentials and consult with the UG researcher for scientific relevance. When the request is deemed valid, the curators will send the tailored DTA to the requesting party, who will need to fill in the following details in the DTA:

1. A detailed description of research/intended use of the data
2. The first and last name of the principal investigator of the requesting party, including their job title
3. A description of security measures
4. The name and job title of the contact person

[3] Caselli, Tommaso; Weultjes, Marieke; Schelhaas, Arjan; Leistra, Folkert; van der Veen, Hylke; Robben, Menno; Timmerman, Gerben; Ruitenbeek, Waard; Zwart, Victor; van der Noord, Robin; Gnezdilov, Zhenja; Theodoridis, Dionysios, 2023, "DALC - Dutch Abusive Language Corpus", https://doi.org/10.34894/HOINL3, DataverseNL, V1

An authorised person at the requesting institute has to sign the DTA. The requesting party can return the signed document to the UG curators, who will subsequently send the DTA to General & Legal Affairs (ABJZ) for a legal check. Once ABJZ gives the go-ahead, the UG curators will ask the Privacy & Security coordinator of the Faculty to have the DTA signed by the Managing Director of the Faculty. The Privacy & Security coordinator will return the signed DTA to the UG curators, who will subsequently send a copy to the requesting party. Finally, the curators will grant access to the dataset via DataverseNL.