

## Improving the Transparency of Data Access Conditions in the SSH Domain

### Recommendations based on a small-scale analysis of the conditions applied to restricted access datasets

Deborah Thorpe  
DANS

Ricarda Braukmann  
DANS

Angelica Maineri  
ODISSEI

#### Abstract

The benefits of making data available for reuse are recognised by many. While some datasets can be openly available, others contain sensitive or personal data that needs to be protected. To allow the sharing of the latter datasets, many trustworthy digital repositories provide options to publish data restricted access. However, detailed standardised information about the access conditions for these restricted access datasets is often lacking from the metadata. Researchers interested in reusing these datasets can thus not judge whether they are eligible to reuse data and under what conditions.

To get a better understanding of how we can increase the transparency of access conditions, this paper aimed to investigate the access conditions and procedures that are commonly applied by depositors within the (Dutch) Social Sciences and Humanities (SSH) research community. The results of a survey that was conducted (n=45) indicated that various conditions are applied, and while some datasets can have few restrictions, for others reuse is highly restricted. Most respondents limit reuse to research purposes and prohibit commercial use. Some datasets are available for students or teaching, but often with additional requirements. A large majority of respondents required a motivation letter to evaluate before allowing reuse. Notably, respondents often chose 'It depends' when asked whether a specific condition was applied, showing a lot of nuances in the conditions and the evaluation of access requests. An important result from our survey was that clear procedures and decision-making guidelines seem to be lacking for many respondents. Requests are often evaluated ad-hoc and through email. Decisions are said

*Submitted 15 January 2025 ~ Accepted 20 February 2025*

Correspondence should be addressed to Deborah Thorpe, DANS, Anna van Saksenlaan 51, 2593 HW, The Hague, The Netherlands. Email: [Deborah.Thorpe@dans.knaw.nl](mailto:Deborah.Thorpe@dans.knaw.nl)

This paper was presented at the International Digital Curation Conference IDCC25, 17-19 February 2025

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution License, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



to require an evaluation of the quality of the application, yet the evaluation criteria seem to be rarely specified and explicitly communicated at the time of data deposit.

Based on the results of the small-scale survey, we conclude with a set of six recommendations directed at data owners and repositories outlining how information about access conditions and procedures can be made more transparent in the future. This work should be seen as a starting point to improve the accessibility and reusability of restricted access datasets in the SSH domain.

## Introduction

The benefits of making data available for reuse are recognised by many. While some datasets can be openly available, others contain sensitive or personal data that needs to be protected. In the Social Sciences and Humanities (SSH) domain, many researchers collect personal data. SSH data may be collected, gathered, or curated using surveys, interviews, oral histories, analyses of social media content, and many more methodologies. The resulting datasets often contain personal data and/or sensitive information, and, although anonymisation is possible in some cases, the value of the data often lies in the personal information that is being studied.

Depending on the type and sensitivity of data, reuse can be impossible or highly restricted (e.g., requiring that data are analysed within a controlled secure environment). For datasets that do not require such a high level of protection but cannot be shared fully open, many repositories provide options to publish datasets with restricted access: the downloading of the data for reuse is only granted after a request is evaluated by the data owner. This paper focuses on the access conditions that depositors apply to such datasets.

The Accessibility principle of FAIR (Wilkinson et al., 2016) states that detailed and standardised information about access procedures and requirements needs to be available. Thus, repositories that aim to promote and implement FAIR should be supporting data depositors in making their restrictions transparent. Currently, however, information about the restrictions applied to the data is often incomplete or lacking from the metadata. This makes it difficult for users to assess whether they are eligible to reuse data and under what conditions. Requirements that data owners apply can vary and may include: limitations on the individuals that are allowed to reuse data; the need for a motivation letter; and/or limitations on the reuse purpose. Importantly, even if conditions are specified by the data owner, there is no commonly accepted, standardised, machine-actionable way to express these conditions in the metadata.

Recognising that such a lack of transparency is a barrier to FAIR, DANS, the Dutch national centre of expertise and repository for research data, has taken steps to support data owners to provide adequate information about data reuse conditions for datasets that are published under restricted access. To provide guidance for depositors, a guidebook on restricted access was published in 2024 (Braukmann, Verburg, & Mahabier 2024a). It provides guidance for depositors explaining what to consider and how to create a data access protocol.

Providing standardised ways to describe data access requirements is also one of the goals of ODISSEI,<sup>1</sup> the Dutch Open Data Infrastructure for Social Science and Economic Innovations, and in particular within the development of the ODISSEI Portal.<sup>2</sup> As outlined in the user guide (Huis in 't Veld et al., 2024), the Portal was created so that metadata from various Dutch social science data providers can be found in one interface. The Portal aims to also provide standardised and machine-actionable information about data access conditions and broker access to datasets through the Data Access Broker (DAB).

While we thus recognise the importance of providing guidance and alignment of the descriptions of data access procedures, more information is needed on the data access conditions currently applied within the SSH research community. Datasets from this community in the Netherlands are primarily published through the DANS Data Station SSH<sup>3</sup> or through local university repositories, such as DataverseNL.<sup>4</sup> These datasets typically do not contain detailed information outlining data access conditions, and no standardised description is currently available. Hence knowledge on the applied restrictions remains anecdotal. To gather input from the community on standardising access conditions, DANS and ODISSEI initiated discussions in two workshops (Hugo, van Kemenade, & Braukmann, 2022; Braukmann, Hugo, & van Kemenade, 2023). To build on these workshops, we launched a survey in April 2024. In this paper, we present the results of this survey and distil a set of recommendations for data owners

---

<sup>1</sup> ODISSEI: <https://odissei-data.nl/nl/>

<sup>2</sup> ODISSEI Portal: <https://odissei-data.nl/nl/odissei-portal-nl/>

<sup>3</sup> DANS Data Station SSH: <http://ssh.datasetations.nl>

<sup>4</sup> DataverseNL: <http://dataverse.nl>

and repositories to move towards a further standardisation of access conditions. Such standardisation efforts would significantly facilitate the uptake of the FAIR principles and in particular of the Accessibility principle.

## Survey on Access Conditions

### Aim

The aim of the survey was to gather information about access conditions applied in the (Dutch) SSH community. We intended that the survey results would aid the identification of common conditions that should be included in standardisation efforts for the DANS repository services. We wanted to use the survey results to create a set of recommendations which could then be shared for community feedback and further alignment. Finally, the results should also contribute to improving our existing guidance for depositors. It should be noted that while our focus was on the SSH domain and the Dutch community and was promoted through DANS and ODISSEI with a large audience in this domain, we did not explicitly exclude respondents based on these factors.

### Design

The survey contained three sections with a total of 19 questions (see Appendix A). The survey opened with a section on background information, followed by information about the data the respondent has archived with restricted access. Then, we asked detailed questions about the data access conditions that are applied. We asked about different possible conditions which were based on the DANS Data Access Protocol template (Braukmann, Verburg, & Mahabier, 2024b), which included, for instance, reuse only for research, teaching, and reuse for commercial purposes. Respondents were encouraged to list any additional conditions. For all questions, respondents were able and encouraged to elaborate in open fields, and we included a final question where any additional information could be entered. In the closing statement of the survey, we invited respondents to email us if they were interested in taking part in a future user group on the same topic. Three individuals expressed their interest in the user group.

The survey was executed using EUSurvey (v1.5.3.1), which is open source (GitHub<sup>5</sup>), built by DG DIGIT, and funded under the ISA and Digital Europe Programme (DIGITAL).

### Recruitment

The survey was open from 10 April 2024 until 1 July 2024. The survey was distributed primarily via DANS and ODISSEI communications channels to reach the Dutch SSH community. In addition, we were also able to contact data depositors who have deposited restricted access data in the DANS Data Station SSH. Participation in the survey was voluntary, and, by default, responses were anonymous.

We also promoted the survey through our networks aimed at both researchers and research supporters, such as the Dutch National Coordination Point Research Data Management (LCRDM) and the Data Stewardship Interest Group (DSIG), as well as our personal social media channels and newsletters with a SSH focus (e.g., from the Digital Repository of Ireland<sup>6</sup> and DARIAH-EU<sup>7</sup>). Lastly, we promoted the survey at the national SURF Research Day, where a short 'pecha kucha' presentation was given about its aims and delivery (Thorpe & Braukmann, 2024).

---

<sup>5</sup> EU Survey: <https://github.com/EUSurvey>

<sup>6</sup> Digital Repository of Ireland: <https://dri.ie/>

<sup>7</sup> DARIAH-EU: <https://www.dariah.eu/>

## Analysis

Of the 19 questions, eight were closed-ended and included the option to elaborate on the answer. Eleven questions were open-ended. We summarised the results to the closed-ended questions in bar graphs using R Studio (Version 2024.04.2+764). The scripts used for the analysis are available on Github.<sup>8</sup> We performed a qualitative analysis of the answers to the open text questions. It is in these open-ended questions that we expected to receive nuanced information about how data depositors establish and manage the access restrictions they apply.

## Results

We received 46 responses. One response was excluded as they indicated that they did not work with restricted access data, and they did not respond to any of the other questions. The results discussed below are thus from 45 respondents. The raw results of the survey were exported from EUSurvey in .ods format and are published on Zenodo under a CC-BY licence (Braukmann & Thorpe, 2024).

### Respondents (Q1–Q5)

The majority of respondents were from the Netherlands (see Table 1) and mostly working at a university. Nine respondents classified their organisation type as ‘other’: four were working in a University Medical Centre or hospital; two indicated they were part of an independent research organisation; two worked for an NGO or social organisation; and two were retired or currently unemployed. Most respondents were either in a research position (14) or working as data managers (8) and data stewards (7). In addition, there were two librarians, two project managers, and individual responses from, amongst others, a privacy officer, policy manager, an IT consultant, an information specialist, and a lecturer. Twenty-six of the 45 respondents reported that they had archived data at one of the repository services hosted by DANS.<sup>9</sup>

### Types of Data (Q6)

Most respondents indicated that they work with personal data. Sixteen respondents explicitly mentioned working with interview data, but cohort studies and surveys were also represented amongst the responses. Five respondents indicated working with health data, and one respondent indicated they work with sensitive corporate data. Respondents often indicated that the data they work with has a certain additional sensitivity, including researching sensitive topics or working with vulnerable populations.

### Reasons for Restriction (Q7)

Most often respondents mentioned **privacy and GDPR** as reasons to restrict access. The **informed consent (IC)** that is gathered from participants was mentioned to be an important guideline for the restriction of access. Some respondents indicated that the IC promises limitations on reuse to **scientific purposes** only or prohibits reuse for **commercial** purposes. Three respondents indicated that data is restricted as it cannot be anonymised or **anonymisation** is too difficult. Three respondents indicated that they have legal agreements or contracts with third parties which create a need for restrictions on reuse. Two respondents mentioned copyright as the reason for restricting reuse.

Some respondents mentioned that restrictions are derived from concern about the **security** of the research participants. In some cases, this was to mitigate risk for participants from vulnerable

---

<sup>8</sup> Analysis Scripts available on Github: <https://github.com/odissei-data/SurveyAccessConditions>

<sup>9</sup> DANS services: <https://dans.knaw.nl/en/products-services/>

groups (mentioned by three respondents). One respondent was working with educational datasets, which contain student submissions for assignments. In this case, the depositor needed to restrict access because answers for assignments that are still in use should not be openly available. In some cases, respondents indicated that **embargos** are placed on the data, restricting reuse before a certain time has passed.

There were a few individual responses that deviated from the majority of the answers. One respondent indicated that they restricted access to the data because they were afraid of **misinterpretation**. Two respondents stated that their reasons for restriction related to **recognition and reward**, wanting to use the data themselves first, or wanting to co-author publications from secondary analyses.

**Table 1.** Background of the participants. DS=DANS Data Stations, DV-NL=DataverseNL, EASY=DANS EASY archive (predecessor of the Data Stations).

Country		Organisation		Archived at DANS	
Netherlands	37	University	24	No	19
Norway	2	Other	9	EASY & DS	9
Austria	1	Research institute	5	DV-NL	8
Belgium	1	Infrastructure provider	4	EASY	7
Canada	1	Research institute; other	1	DS	1
France	1	University of applied sciences	1	EASY & DV-NL	1
Ireland	1	University; infrastructure provider	1		
UK	1				

## Evaluation of Access Requests (Q8)

In addition to understanding the reasons for restrictions and the conditions applied, we asked about the process by which data depositors receive, manage, and evaluate requests for access to data. The themes that came out of the responses were: 1) information about the ‘mechanisms for access’ to the data; 2) information about forms that are required as part of the access request; and 3) information about who the decision makers were in the access request process.

Many of the responses also expressed a **lack of protocol** for organising access requests, or explicit statements that the process is badly organised and/or difficult.

### Mechanisms and forms for access

A large number of respondents indicated that requests are initiated through a point of contact (usually themselves: a data steward/researcher/other), who often forwards this request to other(s), for example, researchers. Within these responses, there was often no clear indication that there is a structured evaluation process involved. In some cases, the respondent explicitly stated that the process is handled by email, whereas others did not explicitly state what means of communication are involved—but email at some or all parts of the process may be inferred.

Several respondents indicated that a structured order form is involved, but there was no indication as to whether this is managed by a repository, or through a separate form. Five respondents stated that the mechanism for access is through a repository such as DataverseNL, or another repository or platform.

## Decision making

Many respondents indicated or stated explicitly that there is a single decision maker, who may work in consultation with the researcher/research group (e.g., '[the request is] assessed by the dataset owner (Principal Investigator) if the reason behind request is valid'). Nine respondents mentioned an organised committee (e.g., 'proposal needs approval from our study scientific committee'), and two specified the involvement of an ethics board (e.g., 'we also have cases that, due to the sensitivity of the dataset, the deliberation goes to a faculty ethical committee instead of the PI').

## Types of Restrictions (Q9–Q18)

Q9–Q17 of our survey focused on the various restrictions respondents may or may not apply to their datasets. These restrictions are of different natures: some refer to the purpose of reuse (e.g., commercial use, use for teaching, other purposes than research); some refer to the characteristics of the requestor (e.g., allowing access to students, restricting access from certain regions or organisations).

Figures 1 and 2 present an overview of the responses to the closed questions for each of the possible restrictions.

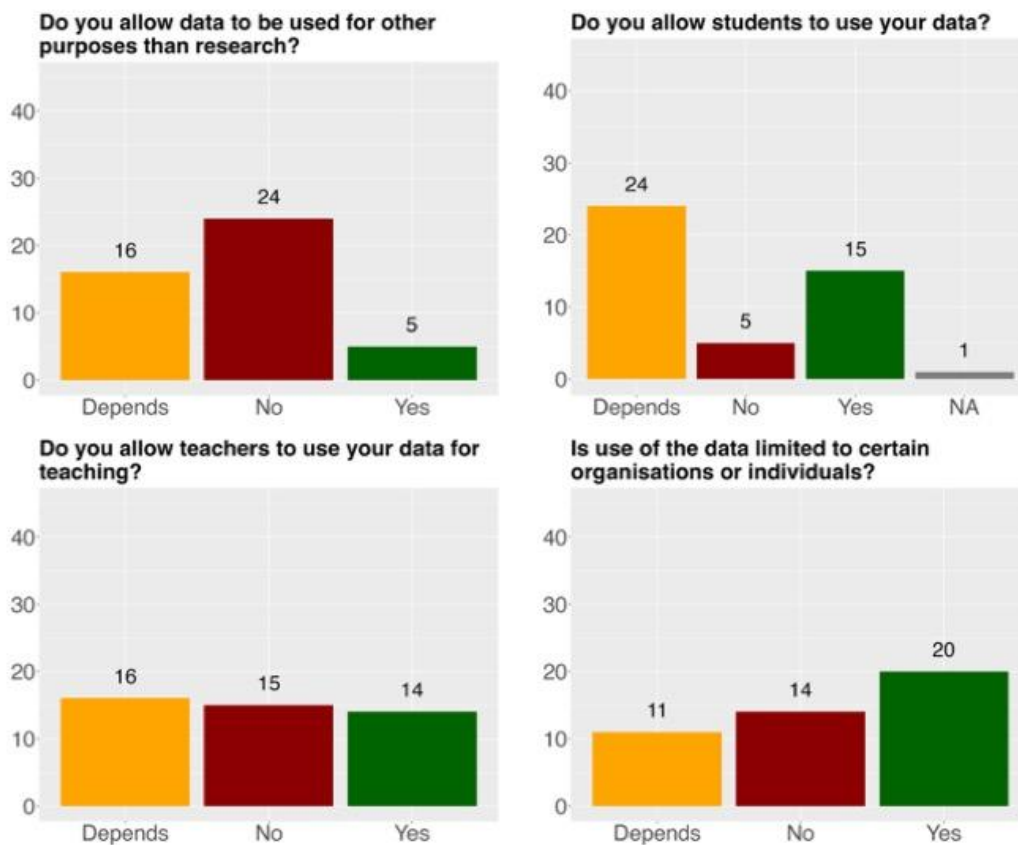
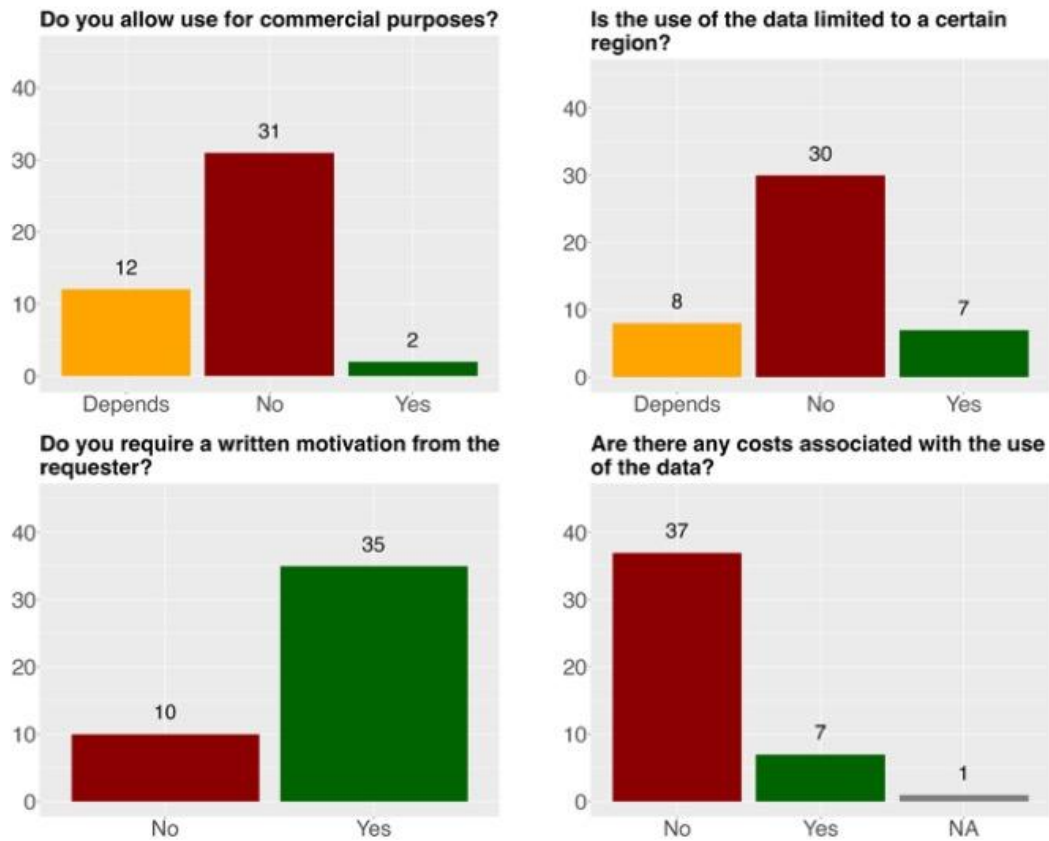


Figure 1. Overview of responses to Q9–12.

From the plots, it can be seen that some conditions are applied by most respondents, like prohibiting use for commercial purposes and requiring a motivation letter. On the other hand, only a few respondents indicate that reuse is associated with costs. What is interesting is that many respondents selected the 'It depends' option and chose to clarify their answers in the open response fields. Note that the 'It depends' option was not available for the questions regarding costs and the motivation letter.

We did not assess further how the different conditions relate to each other (e.g., whether some conditions are always applied together). However, we did see that there was only one respondent who indicated that reuse was very widely allowable: they allowed reuse by students, for teaching, for commercial purposes, and reuse was not limited to particular individuals, organisations, or regions. This concerned data from an NGO stored at DANS. Notably, they did require a motivation letter from the requester. On the other hand, there was also only one respondent who indicated that reuse was very heavily restricted: they did not allow reuse by students or for teachers, they prohibited commercial purposes, and reuse was limited to certain individuals, organisations, and regions. This concerned interview data stored at a university, for which reuse was only permitted within a particular project. All the other respondents applied a mixture of conditions.



**Figure 2.** Overview of responses to Q13–Q15 and Q17.

Below, the different conditions are described in more detail, taking into consideration the answers to the open question, where respondents could elaborate.

### Use of data outside of research (Q9)

Five respondents stated that they do allow data to be used for purposes other than research, 24 did not allow it, and 16 stated that it depends. This is interesting because teaching, for instance, may be defined as a ‘non-research’ usage. However, several respondents stated that they do not allow use of data outside of research, yet selected ‘It depends’ when asked if they allow use for teaching. This may indicate that the exact implications of the restrictions that they are applying are not as clear as they should be, and/or perhaps there is a level of ambiguity around what comprises ‘research’.

Several respondents outlined specific types of use that would be granted, namely, policy making, educational use, and to inform patients/clinical use. There was an indication from two respondents that a judgement would be made on the intention behind the reuse, especially that it needed to be in the public interest. Informed consent (IC) was an issue that was brought up by some respondents indicating that the IC defined certain terms of access and use.



Four respondents indicated a level of ambiguity: they could not think of a potential type of reuse outside of research that would apply to their data, there were no regulations on this yet, or the request would need to be discussed with the department or researcher.

### **Use of data by students (Q10)**

When asked whether they allowed students to access their data, 15 respondents selected 'Yes', whilst only five selected 'No', and 24 respondents answered that 'It depends'. In the open field, some respondents indicated that they required more information and/or involvement of the supervisor. For example, they would allow access to the students of researchers who were already working with the data or were part of the project team. Several mentioned that supervisors needed to request the data for the students. Others mentioned that the quality of the research and reuse purpose would need to be checked. Others indicated that the background of the student was important, for example, they must be from specific Masters programmes, or their own departments, or sometimes a broader restriction (e.g., in the SSH domain). One respondent stated that the answer would depend on the 'status and affiliation of the applicant'.

Some respondents stated that reuse by students requires additional agreements. For instance, good procedures at the student's host institution (e.g., guarantees of anonymity). Or the institute (and in some cases, the students) must sign additional agreements (i.e., data processing, confidentiality).

In several cases, it seemed to be a case-by-case decision that is not formally documented (and, thus, also not directly included in the study preparations and informed consent discussions with participants).

### **Use of data in teaching (Q11)**

Our respondents were nearly equally divided between allowing reuse for teaching (14), not allowing it (15), and indicating that 'It depends' (16). Some elaborated that access would only be allowed for anonymised data and with appropriate safeguards. One respondent answered that it would depend on the request or the course—so a motivation letter would be evaluated (see Q15). Some respondents mention that affiliation is important, but this was mentioned less than it was in relation to the use of the data by students.

With regards to documentation and processes for making decisions regarding access for teaching, one respondent did describe a clear policy in place for their dataset. However, in several cases, it seems, decisions are not formally documented. Some respondents also explicitly said that they are unsure whether reuse would be possible, indicating that such a request had not happened yet.

### **Organisational or individual limitations (Q12)**

Twenty respondents indicated that they limit reuse to certain organisations or individuals, while 14 indicate they do not. Eleven responded 'It depends'. Most respondents elaborating on the restriction indicated that reuse is limited to research-performing organisations. In some cases, reuse by users from educational or clinical institutes is allowed, and one respondent stated that users from non-profit organisations would be allowed to access the data. Several stress the importance of the user's organisation being non-commercial, and in some cases, individuals would need to belong to an approved organisation, or specific agreements need to be signed.

A few respondents indicated that reuse is only possible if the requester works in collaboration with the data owner, and one said they generally require co-authorship with the requester. As in the case of the questions above, it was often specified that permission to access the data depends on the proposal submitted and can be mediated by the agreements in the IC forms. For some respondents, again, the status of the applicant is relevant.

As with the answers to previous questions, it appears that for several respondents no formal agreements are in place in relation to this type of restriction.

### **Commercial use (Q13)**

Most respondents (31) do not allow reuse for commercial purposes. Only two said they do allow commercial use, and 12 respondents responded 'It depends'. In the open field, it was often

explained that this would be decided on a case-by-case basis and in agreement with researchers—or in one case, a data access committee. Access may be limited to organisations that are collaborators, and the purpose of the research was given as an important evaluation criterion.

Several respondents explained that they have not yet received a request for use for commercial purposes, and several respondents said that legal agreements would need to be developed. One respondent remarked that the definition of ‘commercial’ can be interpreted very broadly, making it difficult to provide a definite answer.

### **Regional limitations (Q14)**

The majority of respondents (30) stated that access would not be limited by region, seven said it would, and nine indicated ‘It depends’. The responses in the open field often stated that data would need to remain in the EU or EEA due to GDPR. Some said it can be shared outside of the EEA with extra agreements or with countries for which the EU has adopted an adequacy decision.

In some cases, parts of the data could be shared, and IC was mentioned as a determining factor. Certain respondents noted they would only share data with collaborators (as in the case of the answers to Q12 above). One respondent mentioned that organisations may prohibit data reuse for certain countries, and another indicated that data reuse was restricted to the Netherlands.

### **Motivation letter (Q15)**

Thirty-five respondents indicated that they require a written motivation letter. We asked them to outline what information they require in the letter. The responses were varied, but can be categorised into the following types of information:

- **A research plan:** Respondents wanted information about the research proposal plan (22 respondents), descriptions of the overall purpose of accessing the data (eight respondents), or an outline of what exactly the requester plans to do to/with the data (six respondents).
- **Some indication of ‘good’ research or researcher quality:** Respondents wanted to know whether there is a ‘good’ research plan in place so as to ensure the quality of professional and/or ethical standards (two respondents), evaluate whether the requester had ‘good intentions’ (one respondent), or ensure that the research will be likely to make a contribution to the understanding of the topic (one respondent).
- **Meeting ethical, legal requirements and consent:** Respondents mentioned maintaining anonymity/confidentiality (four respondents), ensuring that reuse proposed fits with the consent gained (one respondent), and potential agreements not to share the data with others (one respondent).

Interestingly, there was often no specified format or length for the letter of motivation (19 respondents). Six respondents indicated that a form is used, and three that a template is in place. One person answered that they require a form, plus an analysis plan. Where respondents did specify a word length, there was considerable variety in the specified length, ranging from ‘more than 100 characters’ to ‘more than 100 words’, to a ‘half A4 page’.

When asked about the criteria for evaluating written letters of motivation, these can be divided into three main themes: 1) Is the research ‘good quality’/interesting/ethical?; 2) Is the researcher themselves a ‘bona fide’ researcher, working at a recognised institute?; 3) Does the data fit the research proposal, and are they requesting the right (amount of) data?

Other criteria included whether the reuse is compatible with the IC gained from research participants, and whether the requester can guarantee anonymity/protect the privacy of research subjects.

A cluster of six respondents indicated that there either were no criteria, or that the checks were at a very basic level (i.e., establishing that the purpose of the request is ‘scientific’ or ‘serious’). For example: ‘No actual criteria, just to verify it’s not an automatically generated request.’

### Costs for reuse (Q17)

Only seven respondents indicated that they would charge for reuse. Several respondents specified that a fee may be requested to cover data management, handling and protection services, as well as the work needed to prepare the data for reuse. In some cases, costs are a standard fee, while others determine costs based on the size of the dataset. Costs for making data available in a specialised data-processing environment are also mentioned by one participant.

### Additional requirements for reuse (Q16, Q18)

A variety of other written agreements to be signed were listed by respondents in response to Q16, including data access agreements; data transfer agreements; data sharing agreements; and confidentiality agreements.

Requirements that were not covered in our questions but mentioned by respondents in response to Q18 were that requesters were required to analyse data in a **specific environment**, the need to **cite the dataset** in a publication, or a limitation for reuse within a **specific research field**. An interesting note was that one respondent indicated that reuse may be limited if there is a **conflict with other approved data requests**.

Some respondents have specific agreements with reusers (such as those mentioned in response to Q16), where conditions are specified, and these may differ from case to case. One respondent said that the restrictions are detailed in their data management plan (DMP), though since this was an anonymous survey, we do not have access to the DMP for more information.

An interesting comment from one respondent was that ‘European law does not allow discrimination of EU citizens. Any user that satisfies the criteria must be given access under the same conditions’.

One respondent commented that it can be difficult to evaluate reuse for data mining and AI use, as it can be hard to assess the purpose and decide whether it is in line with the original purpose specified in the informed consent.

### Additional Remarks (Q19)

The final question was an open field for additional information or comments. As might be expected, the responses were wide-ranging. Some respondents indicated that the process of filling in the questionnaire itself was a reminder of the **need to set up protocols** for data access: for example, ‘this stresses the need for a good though-through protocol which we still need to set up’ and ‘I wish there was a faster and easier way to share data. As data sharing becomes more common, I hope researchers will have a better idea of the conditions under which they would be prepared to share data, and this process can be more streamlined’.

Others stated that the **lack of a standardised procedure** was due to the fact that the numbers of access requests were still small: ‘Due to the limited number of requests, the procedure is mostly ad-hoc at the moment’.

On the topic of reasons for access restrictions, one respondent wanted to point out that they might not be so clearly defined, and when datasets are reused and combined, they may stack up: ‘**Multiple data sensitivities** may intertwine or even collide in newly assembled sets of data for reuse’.

## Discussion and Recommendations

The answers to our survey demonstrate that data owners apply a wide variety of access restrictions and conditions. Some restricted access datasets may be available with very few conditions, while multiple conditions may apply for others. Importantly, many respondents chose the option ‘It depends’ and elaborated in the open text fields, indicating that conditions may be more nuanced and may not be easily specified in binary categories.

Responses to the open-ended questions suggest that requests for access are often sent via email. Since in these cases the data repository is not involved in the communication, no formal

documentation of the communication is saved, risking a lack of transparency and equity. The unstructured email format increases the likelihood of inadequate information being requested or provided, and/or misunderstandings between the requester and the person managing access.

While some respondents indicate that a request form or template is already in use, a form that is included in the access process at a repository could standardise this process further, making it more straightforward for both the data requester and the data depositor. Forms can increase transparency around decisions, when both the questions being asked of the requester and the requester's answers are recorded within the repository system—alongside the final decision that is made about whether to grant access.

One significant finding of this survey is that the majority of respondents (35 out of 45) require a motivation letter for granting access. This is important to know as it results in additional work for the data user, in particular when it is not clearly specified what information such a letter needs to contain and how it will be evaluated. It will be important to further investigate whether the information that those managing access require in such a letter could be adequately gathered using a comprehensive request form within the repository.

The impression is that though data depositors require the written letter to check the background and motivations of the requester (or even just as a sanity check to ascertain that it is a real person making the request), they are less sure about what the letter should contain and how they will evaluate it. Expectations that the letter will indicate the quality of the researcher and their institution and/or their 'good intentions' are at best vague, and at worst have the potential to introduce inequalities in data access. Thus, any request form should be used carefully in combination with documentation in which the data owner considers and outlines all of the information that is needed to evaluate the request. An example for such a form is the DANS Data Access Protocol template (Braukmann, Verburg, & Mahabier, 2024b).

In general, the variety and nuance in the answers to each of the open questions suggests that access conditions are not always easily categorised and have not always been decided upon prior to depositing the data. This underlines the importance of describing both access conditions and the request procedure, so that users can quickly and easily ascertain if they will be eligible for reuse, and how their request will be processed.

Ambiguity about access, coupled with an unnecessarily time-consuming or confusing request process, may result in researchers deciding not to select restricted access data for reuse in their projects, in an environment where they are still required to publish quickly and often. Conversely, a more transparent and streamlined process may encourage the reuse of restricted access datasets, providing of course that the access requirements are met by the requester.

Of course, we also need to note that our survey is not representative. While the 45 responses do give us a general idea of the conditions applied by researchers in the SSH field, in particular in the Netherlands, the results should be further validated. We see the recommendations outlined below as a starting point which should be discussed with the wider community to assess if they reflect broader perspectives. This will be an important next step in our ambitions to further harmonise the description of access conditions.

## Recommendations

Based on the results of the survey and the discussion points outlined above, we put forward the following recommendations:

### **Data owners: Make data access planning part of project preparation**

Processes and restrictions are often not well defined, and several respondents indicated that requests are handled on a case-by-case basis. This hampers transparency and can lead to arbitrary decisions.

It is vital that conditions for access are discussed at an earlier point in time. We recommend making this an active part of project preparations, with access conditions considered and articulated in the data management plan. Since many respondents indicated the importance of the IC forms in guiding data sharing, data sharing needs to be included in the ethical approval process and the design of the IC procedures.

### **Data owners: Commit to ‘as closed as necessary’**

According to the H2020 Program Guidelines on FAIR Data (European Commission, 2016), data should be ‘as open as possible and as closed as necessary’. The phrase ‘as necessary’ makes it clear that restrictions should be imposed to respect privacy, data sensitivity, or intellectual property rights, but it also means that *only* the restrictions that are needed should be applied.

We recommend critically considering whether each potential restriction is necessary and carefully articulating the reasons for applying it. Consider whether justifications for restricting access align with the FAIR principles and Open Science values: limiting access to research collaborators, for instance, is likely to contravene these principles. The process of making each restriction explicit and justified will help to align the sharing of restricted access data with the FAIR principles and the values of Open Science. As decisions can in some cases be difficult, we stress the importance of available support from someone who can provide advice, such as an institutional data steward.

We also want to stress that **better understanding of the GDPR** seems to be needed. Many respondents note the GDPR and other legal requirements as a reason to limit data sharing, stating, for instance, that data needs to remain within the EEA. Yet it is questionable whether this is always, by definition, the case. We advocate for a deeper understanding of the implications of the GDPR and welcome more education from legal experts.

### **Data owners: Publish a data access protocol alongside your data**

All conditions applied should be documented and published. The information needs to be linked to any (meta)data publication that describes the data so that reusers are informed before they start the application process. Our survey clearly shows that, while some depositors have clear conditions, many decide on a case-by-case basis. To avoid arbitrary decisions and to follow the Accessibility principles of FAIR, this information should be collected and published. Conditions can be outlined for example through a data access protocol.

In addition to the conditions that apply, the process should be communicated as well. Who will evaluate the response, based on what information, and what steps are made to reach a decision? As we have seen that many depositors ask for a motivation letter, we especially encourage that the requirements for such a letter and how they will be assessed are explicitly included in the documentation.

### **Repositories: Provide standard to record access conditions**

We acknowledge that, to our knowledge, currently no widely accepted standardised way of describing access conditions and processes is available. Repositories should put efforts into extending metadata and standardised descriptions that can capture common conditions applied to restricted access datasets. Work on this has been initiated, for instance, investigating the use of ODRL (Open Digital Rights Language) to describe access conditions in a standardised way (Bell, 2024).

### **Repositories: Improve the support for handling access requests**

Our survey indicates that many respondents rely on email conversation to discuss reuse. The potential limitations of and threats from this process are explained in the Discussion section above. We advocate that access requests should, where possible, be performed through existing repository infrastructure. Many repositories offer an access request functionality, and in the context of ODISSEI, work has been done on a Data Access Broker (Van Der Meer & Hesam, 2024) through which access to different restricted access datasets can be requested and handled in a uniform way. As outlined above, collecting the required information through a request form can aid the evaluation process. Further assessing the needs of the data owners to process requests through the infrastructure is an important step to improve transparency.

### **Repositories: Collect and share insights on data reuse**

While our survey provides a small insight into the practices around conditions of reuse for restricted access, a better understanding is still very much needed. We encourage repositories and infrastructure providers to collect, analyse, share, and discuss more comprehensive information

about the reuse of restricted access data. Capturing more information about access requests and decisions within the infrastructure, as opposed to outside the system through email, would contribute to this, since the ‘fuller picture’ could then be analysed by specialists working at the repositories.

Transparency about reuse for research purposes, but also for teaching or commercial purposes, can provide insights into current practices. In addition, this information will allow us to identify practices that contradict the Open Science values and contribute to existing inequities. Examples might include: restricting reuse of data to colleagues without a solid reason for doing so; or requiring co-authorship in all cases of reuse.

## Future Directions

This paper has contributed some information about the types of conditions that are applied to restricted access data within the SSH community. While our results provide some guidance on moving towards a more transparent process for accessing restricted data, our sample was very small, and the observations and recommendations should be seen as a first step.

To bring this effort further into alignment with the broader community, including CESSDA (Consortium of European Social Sciences Data Archives)<sup>10</sup> and the Research Data Alliance<sup>11</sup> (RDA, e.g. through the Sensitive Data Interest Group<sup>12</sup>), would be an important next step. Within the ODISSEI consortium, there have been plans to further assess the use of ODRL (Open Digital Rights Language)<sup>13</sup> for access interoperability based on work from the UK Data Service (UKDS) within the FAIR-Impact project (Bell, 2024). These are important developments that will need to be considered and discussed with the wider community.

Overall, with this work, we hope to have shed more light on current practices and inspire the adoption of a more structured description of data access conditions and procedures to enhance the ‘A’ of FAIR (Wilkinson et al., 2016): providing the exact conditions under which the data are accessible.

## Acknowledgements

The work on this survey has been conducted as part of the ODISSEI Roadmap project funded by the Dutch Researcher Council under NWO Grant Number 184.035.014 and the SSHOC-NL project funded by the Dutch Researcher Council under NWO Grant Number 184.036.020.

## References

- Bell, D. (2024). *Advancing access interoperability with ODRL* (FAIR-Impact Blog). Retrieved from <https://fair-impact.eu/use-cases/advancing-access-interoperability-odrl>
- Braukmann, R., Hugo, W., & van Kemenade, J. (2023). *Harmonising Access Procedures for Sensitive Data—Workshop at the Open Science Conference* (Zenodo). Retrieved from <https://doi.org/10.5281/zenodo.8108414>
- Braukmann, R., & Thorpe, D. (2024). Survey on data access conditions for sensitive data—Raw Data and Documentation [Data set] (Zenodo). <https://doi.org/10.5281/zenodo.12805137>

---

<sup>10</sup> CESSDA: <https://www.cessda.eu/>

<sup>11</sup> RDA: <https://www.rd-alliance.org/>

<sup>12</sup> Sensitive Data Interest Group: <https://www.rd-alliance.org/rationale/sensitive-data-interest-group/>

<sup>13</sup> ODRL: <https://www.w3.org/TR/odrl-model/>

- Braukmann, R., Verburg, M., & Mahabier, W. (2024a). *Guidebook—Depositing Restricted Access In The DANS Data Stations (1.0)* (Zenodo). Retrieved from <https://doi.org/10.5281/zenodo.10887484>
- Braukmann, R., Verburg, M., & Mahabier, W. (2024b). *Data Access Protocol Template (1.0)* (Zenodo). Retrieved from <https://doi.org/10.5281/zenodo.10887572>
- European Commission. (2016). *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020 Version 3.0* (Luxembourg, European Commission, Directorate-General for Research & Innovation). Retrieved from <http://dx.doi.org/10.25607/OBP-774>
- Hugo, W., van Kemenade, J., & Braukmann, R. (2022). *Harmonising Access Procedures for Sensitive Data—Workshop at the FAIR Data Day* (Zenodo). Retrieved from <https://doi.org/10.5281/zenodo.7382780>
- Huis in 't Veld, L., Braukmann, R., Thorpe, D. E., & Maineri, A. M. (2024). *ODISSEI Portal User Guide (1.0)* (Zenodo). Retrieved from <https://doi.org/10.5281/zenodo.14524428>
- Thorpe, D., & Braukmann, R. (2024, June 10). *Improving the Transparency of Data Access Procedures* (Zenodo). Retrieved from <https://doi.org/10.5281/zenodo.11546569>
- Van Der Meer, L., & Hesam, A. (2024). *The complete SSH workflow for sharing sensitive data. EGI Conference 2024* (Zenodo). Retrieved from <https://doi.org/10.5281/zenodo.14279126>
- Wilkinson, M., Dumontier, M., Aalbersberg, I., 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>

## Appendix 1: Survey Description and Questions

Title	Survey on data access conditions for sensitive data
Information about the survey	
Description	<p>Thank you for considering filling out this survey led by Dr Ricarda Braukmann and Dr Deborah Thorpe on behalf of <a href="#">ODISSEI</a> and <a href="#">DANS</a>, on data access conditions. The purpose is to survey the current practices in the field by asking data owners to provide some information about the processes involved in selecting data access restrictions and your underlying motivations.</p> <p>Our survey is aimed primarily at data owners who use services to archive and publish datasets that require access restrictions. Should you choose to participate, you will be asked a number of questions about how you manage datasets; about the range of potential access restrictions that you may apply to your data; and about how requests to access these restricted data sets are applied. We expect the survey to take approximately 10-15 minutes.</p> <p>Participation is entirely voluntary and by default your responses are anonymous. If you have any concerns or queries then please contact us by email at <a href="mailto:fairsupport@odissei-data.nl">fairsupport@odissei-data.nl</a></p> <p>The information you provide may contribute to research publications, conference presentations and policy briefs in an anonymous format. A subset of the anonymous data and supporting metadata will be made openly available for reuse and verification as FAIR and Open Data in a repository for indefinite long-term preservation.</p> <p>If you agree to take part in the survey please proceed to answer the questions. Clicking the final 'Submit' button of the survey at the end will constitute you providing consent to participate.</p> <p>This survey will close on Monday 1st July.</p>



About you	
Description	In this section we want to collect some anonymous information about your type of position, organisation and the types of data that you usually manage
Q1	What is your position / job title?
Q2	<p>What type of organisation do you work for?</p> <p><input type="checkbox"/> University</p> <p><input type="checkbox"/> University of Applied Sciences</p> <p><input type="checkbox"/> Research Institute</p> <p><input type="checkbox"/> Infrastructure provider</p> <p><input type="checkbox"/> Other</p> <p>If you selected 'Other', please elaborate (optional)</p>
Q3	What country?
Q4	<p>Have you used services from DANS to archive and publish data?</p> <p><input type="checkbox"/> Yes - DANS EASY</p> <p><input type="checkbox"/> Yes - DANS Data Station</p> <p><input type="checkbox"/> Yes - <u>DataverseNL</u></p> <p><input type="checkbox"/> No</p>
Q5	<p>Do you manage datasets that require access restrictions?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p>

About the data	
Description	<p>We would like to know a bit more about datasets that you manage that require access restrictions</p> <p>If you selected no to the question above and so do not manage datasets that require access restrictions, the rest of the survey will not be applicable. If you have any questions about depositing restricted access data in DANS, please contact the relevant <a href="#">Data Station Manager at DANS</a>.</p> <p>If you do manage datasets with access restrictions, we invite you to continue with the survey below:</p>
Q6	What kinds of datasets do the access restrictions concern?
Q7	What is the main reason for restricting access?
Q8	How is the evaluation of data access requests organised?

More about the data access restrictions	
<b>Description</b>	The following questions regard the restrictions for access you may impose on a requester
<b>Q9</b>	<p>Do you allow data to be used for other purposes than research?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> It depends (please elaborate below)</p> <p>If you selected 'It depends' to the question about reusing the data for other purposes than research, please elaborate...</p>
<b>Q10</b>	<p>Do you allow students to use your data?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> It depends (please elaborate below)</p> <p>If you selected 'It depends' to the question about reusing the data for other purposes than research, please elaborate...</p>
<b>Q11</b>	<p>Do you allow teachers to use your data for teaching?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> It depends (please elaborate below)</p> <p>If you selected 'It depends' to the question about reusing the data for other purposes than research, please elaborate...</p>
<b>Q12</b>	<p>Is use of the data limited to certain organisations or individuals?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> It depends (please elaborate below)</p> <p>If you selected 'It depends' to the question about reusing the data for other purposes than research, please elaborate...</p>
<b>Q13</b>	<p>Do you allow use for commercial purposes?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> It depends (please elaborate below)</p> <p>If you selected 'It depends' to the question about reusing the data for other purposes than research, please elaborate...</p>

Q14	<p>Is use of the data limited to a certain region (for instance data needs to remain in the EU or The Netherlands?)</p> <p><input type="checkbox"/> Yes (please elaborate below - which region?)</p> <p><input type="checkbox"/> No</p> <p><input type="checkbox"/> It depends (please elaborate below)</p> <p>If you selected 'Yes' or 'It depends' to the question about reusing the data for other purposes than research, please elaborate...</p>
Q15	<p>Do you require a written motivation from the requester?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p>
Q15a	<p>If you require a written motivation letter, what information do you expect?</p>
Q15b	<p>Do you expect a certain format or length for the letter of motivation?</p>
Q15c	<p>What are the criteria that you use to evaluate the motivation letter?</p>
Q16	<p>List any other agreements that the requester has to sign to gain access to the data</p>
Q17	<p>Are there any costs associated with the use of the data?</p> <p><input type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p>If there are any costs, please elaborate</p>
Q18	<p>Please list any other restrictions that you apply for data access</p>
Q19	<p>Any other information or comments?</p>
Final section and button to submit	
Description	<p>We are planning to work on this topic more, for instance by <b>establishing a user group</b> to discuss how DANS could improve application forms for restricted access and what information should be provided to the applicants.</p> <p>If you are interested in taking part in such a user group, <b>please email us</b> at: <a href="mailto:fairsupport@odisseei-data.nl">fairsupport@odisseei-data.nl</a>. We will then get in touch with you with more details.</p> <p>We would like to thank you for taking the time to fill out this survey. Your effort is appreciated, and will greatly contribute to our understanding of access restrictions.</p> <p>Reminder: by clicking submit below, you indicate your consent to participate.</p>