

# IMPLEMENTATION

# STORIES



**FAIRSFair**  
Fostering Fair Data Practices in Europe



THEME



**ENSURING TRUSTED  
CURATION OF FAIR DATA**

# IMPLEMENTATION STORIES

THEME 7 ENSURING TRUSTED  
CURATION OF FAIR DATA

## Making data FAIR but not open

Authors  
Laura Molloy (CODATA)

From interviews with  
Anja Perry and Oliver Watteler, GESIS–Leibniz Institute for the Social Sciences,  
Cologne, Germany, and staff of the Research Data Section (RDS) and the  
European Social Survey Data Archive (ESS), both at the Norwegian Centre for  
Research Data (NSD), Bergen, Norway.

gesis Leibniz-Institut  
für Sozialwissenschaften

NSD  
NORSK SENTER FOR  
FORSKNINGSDATA

### Introduction

This implementation story discusses approaches in the data archive world towards tackling the challenges of making data as FAIR as possible, when there are compelling reasons for the data to be restricted or unavailable.

This topic was included in FAIRsFAIR deliverable D3.4, Recommendations on practice to support FAIR data principles, under the theme “Ensuring trusted curation of data”. Within that set of recommendations, the FAIRsFAIR project committed to supporting change in good practice for researchers, repositories and ethics committees on selecting and preparing sensitive data to be FAIR. This implementation story aims to support that goal.

### *FAIRsFAIR recommendation*

*“Develop and implement guidance and support for making sensitive data FAIR for reuse.”*

*FAIRsFAIR Recommendations on practice to support FAIR principles*

## Approach taken

### Data archives

Data archives often hold datasets that are either embargoed for a specific period, or are unavailable unless certain access conditions are met. This may happen for a variety of reasons. For example, research may have been undertaken in collaboration with industry on the basis of legal agreements that restrict data sharing or data publishing for commercial reasons. Data from earth sciences are restricted when disclosure would provide locations of rare species of plants, animals or minerals. In the social sciences, some studies involve participation by individuals from vulnerable communities or may expose participants to legal or safety consequences, should they be identified. In some cases, the particular context of the study makes it difficult or impossible to completely anonymize responses and in such cases, access to some data may be restricted in order to protect the safety of participants.

A dataset, therefore, cannot always responsibly be made entirely open. None of the cases above are due to a lack of desire by the researcher to share their data, but the driving principle in all cases should be to make data as open as possible, as closed as necessary.

*"Policies should be aligned and consolidated to ensure that publicly-funded research data are made FAIR and Open, except for legitimate restrictions. The maxim 'as Open as possible, as closed as necessary' should be applied proportionately with genuine best efforts to share."*

Turning FAIR into Reality, recommendation 17, p. 70

This widely-quoted principle encourages responsible, intelligent openness and helps the researcher and research professional staff to think through the legal and ethical consequences of their data curation decisions. Such 'thinking through' informs – and should be captured in – the data management plan (DMP).

*"Workflows, business processes and safeguards need to be established to ensure that research groups benefit by exposing DMP content within acceptable levels of risk. Where a project will use sensitive personal data, for example, a DMP that describes processing details may need to have access limitations."*

FAIRsFAIR Recommendations on practice to support FAIR principles, p. 20

When finding the best stewardship location for such datasets, a domain-specific data archive can offer confidentiality protections and specialist help that is not offered by less formal solutions such as online repository services and some off-the-shelf institutional repository installations.

One of the many advantages of depositing data in a domain-specific data archive is the deep expertise of archive staff in making data findable, accessible, interoperable and reusable to the greatest extent, whilst accommodating the ethical and legal requirements of sensitive data curation for such data. It is important for legal, financial and scholarly reasons that, when a dataset exists, its location and access conditions are known and documented in a standardised and accepted way. The researcher must justify their use of time and resources to their funder; the data archive must also understand and plan for sufficient and appropriate preservation and access infrastructure over time; and scientific endeavour as a whole benefits from the knowledge produced by research activity in order to advance science. These reasons for FAIR-aligned practice exist even when data produced by a study is classified as sensitive.

Accordingly, it is important that even sensitive datasets are as FAIR as they possibly can be. This implementation story looks at how these challenges are met by several well-established European data archives in the social sciences: GESIS–Leibniz Institute for the Social Sciences, Cologne, Germany; and the Research Data Section (NSD-RDS) and the European Social Survey Data Archive (ESS), both at the Norwegian Centre for Research Data (NSD), Bergen, Norway.

## Identifying, archiving and providing access to sensitive data in practice

### Data archives

The data archives at GESIS-Leibniz Institute for the Social Sciences and the Norwegian Centre for Research Data (including both RDS and ESS) deal with large data holdings. Datasets that can be considered sensitive are the minority of data holdings, but their presence must nevertheless be identified by staff, and subsequently appropriately accommodated by policy and security measures.

At GESIS, personal data is defined according to the GDPR<sup>1</sup> as those that can be used to re-identify an individual person; and any dataset defined by Article 9 of the GDPR as sensitive data is classified accordingly.

GESIS uses a restricted class of access for sensitive datasets, but this is also used for some non-sensitive datasets in order for data providers to monitor access. At the time of interview, GESIS held 6,548 published studies, 1,029 (~ 16 %) of which are available under access class C, meaning 'restricted access'. Of these holdings, 21 are 'highly sensitive' datasets which are only available for on-site use. GESIS data access classes are currently under review and are in the process of being re-designed. They currently use data tags for the newly developed access classes. The ones currently in use have been developed over the years according to the needs of data providers and do not follow any controlled lists or standards.

At NSD-RDS, the GDPR is also used to define sensitive data. In addition, to decide whether the data are personal identifiable, NSD carries out a manual control of anonymity<sup>2</sup>. Sensitive data are treated with extra security and access to such data is limited. Holdings are mainly datasets that are available for research purposes. Sensitive data holdings are a small minority of the data held. NSD's focus is to archive data that can be shared and made available for reuse. Similarly to GESIS, they use their own classification, not a standard controlled vocabulary.

The European Social Survey Data Archive (ESS) holds, at a conservative estimate, raw data for about 70% of countries participating in the last five rounds of ESS, that is to say ESS5 to ESS9 inclusive, which amounts to 80 sensitive datasets. This is because even if all direct individual identifiers are removed from ESS data, the main dataset contains a lot of factual information about the respondents. Combining demographic variables such as ancestry, country of birth, language spoken at home, household composition (size, age), religion, occupation, level of education, age, and NUTS region<sup>3</sup> increase the risk of disclosure of the respondents. In particular, outliers are in the risk zone of being unique in their characteristics and thus identifiable.

Published data at ESS are in general not sensitive. However, researchers who contact the ESS Data Archive for access to sensitive data can apply for this access via a procedure that involves deletion of data after time-limited use. This procedure is not publicly available on the ESS website. Sensitive data are kept as back-up data for security purposes, rather than for research itself.

1. European Commission (2016). EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 119/1.

2. [https://o.nsd.no/arkivering/en/control\\_anonymity.html](https://o.nsd.no/arkivering/en/control_anonymity.html)

3. <https://ec.europa.eu/eurostat/web/nuts/background>

## Policies and procedures for data that is FAIR but not open

### Data archives

Data archives that have been in place for several decades have often already invested considerable time and focus on activities that could now be described as 'FAIR-aligned'. For example, at GESIS, archive staff have been working on making data findable, accessible, interoperable and reusable in a continuous process since 1960. All metadata are openly accessible and follow the DDI standard. All data made available also receive a persistent identifier in form of a DOI, and technically the data formats are widely usable (e.g. outputs from common statistical software).

The idea for offering highly sensitive data for on-site use was developed in the late 2000s. In 2014, they started training for researchers in research data management, to improve practice, although they note that demand has declined as training in research performing organisations has increased. The skills problem has not yet been resolved, however, as "there are still some researchers we haven't reached yet. Unfortunately, this becomes apparent when we cannot share their data due to mistakes in or missing informed consent."

There was never a moment when GESIS decided to make their data FAIR; rather, they adopted the term recently as it became popular to reflect their FAIR-aligned activities that have been in continuous development for several decades.

Most data archived at NSD-RDS are anonymous, and data made available by NSD-RDS are in general not sensitive. It is possible to archive personal data within NSD if there are permissions to archive and share data with direct or indirect personal identifiers.

Metadata are findable and searchable for all NSD-RDS datasets regardless of access conditions, and whether the data are sensitive or not. If researchers contact the archive for access to sensitive data, the data controller for projects that wish to use personal data from NSD-RDS must ensure and demonstrate that they have legal basis for processing personal data according to the GDPR.

In order to make data as FAIR as possible (meta) data are assigned a persistent identifier – a practice which has been in place since 2017. All datasets are documented following the DDI metadata standard<sup>4</sup> at least at study level. Documentation is based on the archiving forms, questionnaires and any reports or summaries received from the data provider. Metadata for all studies are searchable and findable through the NSD discovery portal.

To facilitate access and re-use, NSD staff have been working on a new discovery portal and have recently started using a new documentation tool to better document metadata. Information on access conditions, and how to get access, are available in study-level metadata.

*"TDRs can provide services for secure storage and managed access to all types of data, including sensitive data."*

FAIRsFAIR Recommendations on practice to support FAIR principles, p. 26

4. Strictly speaking there are three DDI standards: Codebook, Lifecycle, and Cross-domain integration (DDI-CDI): <https://ddialliance.org/learn/what-is-ddi>

# IMPLEMENTATION STORIES

ENSURING TRUSTED  
CURATION OF FAIR DATA

## Policies and procedures for data that is FAIR but not open

At the ESS, efforts for FAIR have been made since 2002 including open online access to data with no embargo period, an approach structured with the DDI suite of standards as the conceptual framework.

In 2019, DOIs were implemented, to be added at study level for data and documentation, and in 2020, CC licences for ESS data and documentations were added into the conditions of use<sup>5</sup> as [CC BY-NC-SA 4.0](#) for data and [CC BY-SA 4.0](#) for documentation. Machine-readable metadata are in the process of implementation now by use of Colectica-based repositories, based on the DDI Lifecycle specification<sup>6</sup>.

All of these interventions are applied across the ESS data holdings including sensitive and non-sensitive datasets. The benefits they bring include further integration into community-accepted and standardised models and specifications which improve accessibility and interoperability, both important aspects of infrastructure development for making and keeping data FAIR.

Present work at the ESS for non-sensitive data within the SSHOC project<sup>7</sup>, specifically SSHOC project Task 5.5 ESS as a service: a pilot making cross-national survey data FAIR, and further deliverables Recommendations for FAIR repository (January 2022), and Report on preparing ESS data for EOSC (April 2022), and cloud-adapted data processing programmes, could be applied to policy for FAIR-compliant dissemination of sensitive data in the future; this work would need to be prioritized and funded by ESS ERIC.

Ultimately, success in dealing with the complex challenges of making sensitive data FAIR is made more likely when researchers communicate clearly and well with data repository/archive staff members, and early in the project lifecycle.

As GESIS staff commented, *"We welcome people who ask us about RDM very early, preferably during the grant proposal stage. This saves a lot of trouble in the long run."*

5. [https://www.europeansocialsurvey.org/data/conditions\\_of\\_use.html](https://www.europeansocialsurvey.org/data/conditions_of_use.html)

6. <https://ddialliance.org/Specification/DDI-Lifecycle/>

7. <https://sshopencloud.eu/about-sshoc>

# IMPLEMENTATION STORIES

ENSURING TRUSTED  
CURATION OF FAIR DATA

## Further information

**GESIS–Leibniz Institute for the Social Sciences. (n.d.). Share your data!**  
<https://www.gesis.org/en/services/archiving-and-sharing/sharing-data>

### NSD-RDS:

NSD (n.d.). Archiving research data. <https://www.nsd.no/en/archiving-research-data>

NSD (n.d.). Personal data. (n.d.). [https://o.nsd.no/arkivering/en/personal\\_data.html](https://o.nsd.no/arkivering/en/personal_data.html)

Examples of sensitive data documented on study level:

<http://nsddata.nsd.uib.no/webview/index.jsp?node=0&submode=ddi&study=http%3A%2F%2F129.177.90.161%3A80%2Fobj%2FStudy%2FNSD2481&language=no&mode=documentation&top=yes>

<http://nsddata.nsd.uib.no/webview/index.jsp?node=0&submode=ddi&study=http%3A%2F%2F129.177.90.161%3A80%2Fobj%2FStudy%2FNSD2791&language=no&mode=documentation&top=yes>

### NSD-ESS:

European Social Survey, European Research Infrastructure Consortium (ESS ERIC) (n.d.). European Social Survey. <https://www.europeansocialsurvey.org>

European Social Survey, European Research Infrastructure Consortium (ESS ERIC) (2018). Round 9 Survey Specification for ESS ERIC Member, Observer and Guest Countries, version 3. [https://www.europeansocialsurvey.org/docs/round9/methods/ESS9\\_project\\_specification.pdf](https://www.europeansocialsurvey.org/docs/round9/methods/ESS9_project_specification.pdf)

European Social Survey, European Research Infrastructure Consortium (ESS ERIC) (2019). ESS9 2018 Data Protocol Edition 1.4. [https://www.europeansocialsurvey.org/docs/round9/survey/ESS9\\_data\\_protocol\\_e01\\_4.pdf](https://www.europeansocialsurvey.org/docs/round9/survey/ESS9_data_protocol_e01_4.pdf)

(Data protocol ESS10 (not yet published: June 2020, but revision foreseen for end of June 2021, to be published in Sept. 2022)

## ■ *About FAIRsFAIR Implementation Stories*

---

FAIRsFAIR Implementation stories illustrate good practices in research communities and organisations to support the implementation of the FAIR principles. These practices encompass 'FAIR-enabling' actions as recommended in the EC Expert Group on FAIR report [Turning FAIR into Reality](#) and the [FAIRsFAIR Recommendations on practice to support FAIR principles](#). FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

## ■ *FAIRsFAIR - Fostering FAIR Data Practices in Europe*

---

European Commission Grant Agreement No. 831558  
H2020-INFRAEOSC-2018-4  
[www.fairsfair.eu](http://www.fairsfair.eu) - [support@fairsfair.eu](mailto:support@fairsfair.eu)

## ■ *Acknowledgements*

---

This report has been produced by the FAIRsFAIR (GA No. 831558) project, which received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019. FAIRsFAIR gratefully acknowledges everyone who we could interview for this series.

## ■ *Disclaimer*

---

The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content. **January 2022**

## ■ *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/>

**JOIN OUR COMMUNITY!**



@FAIRsFAIR\_eu



/company/fairsfair

[www.fairsfair.eu](http://www.fairsfair.eu)



FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.