



**HAL**  
open science

## Data Management Plan of ANR-21-FRAL-0005 (DeNazDB)

Victor Gay, Jan Stuckatz, Selina Hofstetter, Mikkel Dack

► **To cite this version:**

Victor Gay, Jan Stuckatz, Selina Hofstetter, Mikkel Dack. Data Management Plan of ANR-21-FRAL-0005 (DeNazDB). ANR (Agence Nationale de la Recherche - France). 2023. hal-04043652v1

**HAL Id: hal-04043652**

**<https://hal.science/hal-04043652v1>**

Submitted on 23 Mar 2023 (v1), last revised 19 Jun 2024 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Data Management Plan

## ANR-21-FRAL-0005 (DeNazDB)

### Information on data management plan

Version	Initial version
Research subject	Political science, Social and economic geography, History and archaeology
Language	English
Date of creation	2023-03-23
Date of last modification	2023-03-23
Identifier	Hal-04043652
License	<a href="#">Creative Commons Attribution 4.0 International</a>

### Information on project

Project title	Who Became a Nazi? A Structured Database of the Denazification Questionnaires from French and American Occupation Zones in Germany, 1945–1949
Acronym	DeNazDB
Summary	What led Germans to support and join the Nazi party? Did they benefit materially from their membership? And how did they justify their choices?

Nazi ideology caused a world war and one of the deadliest genocides in history. About 8.5 million Germans were members of the Nazi party (NSDAP). However, what caused individual Germans to vote for and join the Nazi party in the 1930s is still not fully understood, in part because current research relies on either aggregate quantitative or small-scale qualitative data. Moreover, existing historical individual-level data does not allow researchers to follow individuals over time, is often unrepresentative, and overlooks important socio-demographic information.

In this project, we propose to collect individual-level longitudinal data from the denazification questionnaires (*Fragebogen*) distributed in Germany's US occupation zones between 1945 and 1949. The *Fragebogen* contain detailed information about individuals' past membership in Nazi organizations, retrospective voting records, as well as academic and occupational histories, income, assets, and other socio-demographics. With grant ANR-21-FRAL-0005, we plan to digitize a population-representative sample of 12,500 questionnaires for the former American occupation zone, from the approximately 500,000 denazification files located at the National Archives (NARA) in College Park, Maryland. The data will allow us to answer which demographic groups voted for or joined the Nazi movement, and which life events (e.g., unemployment) led to these decisions. Moreover, leveraging quantitative analysis techniques, we will draw representative inferences about the individual causes for vote choice and party membership.

The project will contribute to our understanding of individual motivations to vote for and participate in populist radical right-wing movements. It will further add to political science, economic, and historical accounts of how the Weimar Republic slid back from democracy to dictatorship. Additionally, the assembled data will be made available to scholars and citizens.

Funding source    Agence Nationale de la Recherche : [ANR-21-FRAL-0005](#)

Deutsche Forschungsgemeinschaft: GZ: STU 782/1-1, Project Number: 469232628

Start date            2022-06-01

End date 2025-05-31

Partners Fondation Jean-Jacques Laffont and Copenhagen Business School (CBS)

## Research outputs

- **The Fragebogen collection.** This research output consists in a digital collection of 12,500 Fragebogen questionnaires created during the denazification campaign (1945–1949) in the US Occupation Zone.
- **The Fragebogen database.** This research output consists in a relational database of the content of the Fragebogen collection.

## Contributors

Name and ORCID	Affiliation	Rôles
Gay, Victor <a href="https://orcid.org/0000-0001-9912-3841">https://orcid.org/0000-0001-9912-3841</a>	Fondation Jean-Jacques Laffont	Principal investigator, contact for the Fragebogen collection and DMP.
Stuckatz, Jan <a href="https://orcid.org/0000-0002-3666-2570">https://orcid.org/0000-0002-3666-2570</a>	Copenhagen Business School	Principal investigator, contact for the Fragebogen database.
Hofstetter, Selina <a href="https://orcid.org/0000-0002-2333-7517">https://orcid.org/0000-0002-2333-7517</a>	Tecnológico de Monterrey in Monterrey	Contact for the Fragebogen database.
Dack, Mikkel <a href="https://orcid.org/0000-0003-3459-9202">https://orcid.org/0000-0003-3459-9202</a>	Rowan University	Contact for the Fragebogen collection.

# 1. Description of data collected

## The Fragebogen collection

- a. How will new data be collected and produced and/or how will pre-existing data be reused?

### *Location and accessibility*

The archival material for the Fragebogen collection is located at the National Archives at College Park (NARA) at 8601 Adelphi Road, College Park, MD, 20740-6001. It is part of the Series "Fragebogen Files Relating to Denazification, 1945–1948" (NAID [7062935](#)) of the Record Group 260 "Records of U.S. Occupation Headquarters, World War II, 1923–1972." It is spread across 2,183 legal archives boxes.

Both access and use of the material in this series is unrestricted. It was declassified under Declassification Project Numbers NND 775051 and NND 775119.

### *Data collection procedures*

The archival material will be collected in digital form by a team of trained research assistants equipped with high quality digital cameras mounted to downward-facing tripods. More precisely, research assistants will be equipped with Olympus OM-D E-M10 Mark II Mirrorless cameras with 14-42mm EZ lens, mounted on NEEWER 72 inch camera tripods.

### *Type of data collected*

Three sets of digital copies images are to be collected:

- Contextual information such as photographs of the boxes containing the material, the cover of the folder containing the material, etc.,
- The Fragebogen questionnaires (usually 6 pages),
- Other documents appended to the back of the Fragebogen questionnaire, such as letters, court proceeding (Spruchkammern) records, personal biographies (Lebensläufe), testimonies by third persons (e.g., explaining support for jewish friends/neighbors), written statements by individuals explaining membership in Nazi organizations.

- b. Which type of data will be collected or produced?

### *Data collected*

Each digital image collected will be stored in JPG format. JPG ([Joint Photographic Expert Group](#)) is an open standard and stable image format. Each image has a size of 2.80 MB. We expect the entire archival material collected to have a size of about 200 GB.

### *Data produced*

To describe the digital images collected, we will produce metadata stored in JSON-LD format. JSON-LD ([JavaScript Object Notation for Linked Data](#)) is a lightweight open standard and stable Linked Data format extended from JSON. The volume of this file will be negligible.

The creation of this metadata will be done through [Tropy](#), a free open-source software that helps organize and describe photographs and research material. This software generates a TYPY file in SQLite 3 format. [SQLite 3](#) is an open standard and stable database file. The volume of this file will be negligible.

### **The Fragebogen database**

- a. How will new data be collected and produced and/or how will pre-existing data be reused?

### *Data reused*

The Fragebogen database will be constructed through the reuse of the Fragebogen collection.

### *Data produced*

The Fragebogen database will be produced in two steps:

- **Relational database.** First, a relational database will be produced through an [MS Access](#) based on the content of the questionnaires available in the Fragebogen collection.
- **Analysis dataset.** Then, the data from the relational database will be extracted through a [Stata](#) script to produce an analysis dataset, which will consist of an individual-level panel dataset.

- b. Which type of data will be collected or produced?

### *Data reused*

See the description of the type of data collected to construct the Fragebogen collection above.

### *Data produced*

The relational database will be stored in ACCDB format. ACCDB is an MS Access database file. The volume of this file will be about 20 MB.

The script to construct the analysis dataset will be stored in DO (or R) format, the script format used by the software Stata (or R). The volume of this file will be marginal.

The analysis dataset will be stored in two formats: in DTA format, a Stata dataset file, and in CSV format, an open-standard comma-separated values text file. The volume of the DTA file will be about 25 MB, and that of the CSV file, about 35 MB.

## 2. Documentation and data quality

### The Fragebogen collection

- a. Which metadata and documentation will accompany your data?

### Metadata attached to each element of the Fragebogen collection

#### *Metadata standard*

Each image collected will be documented with metadata following the [Dublin Core](#) metadata standard and using controlled vocabularies when appropriate. All elements related to a particular Fragebogen will have the same metadata (except for the title—see *Naming conventions* below). The metadata will document the following elements:

Label	DCMI Term	Definition	Data type
Title	<a href="#">title</a>	Name of the source	<a href="#">XSD String</a>
Creator	<a href="#">creator</a>	Organization responsible for creating the source	<a href="#">XSD String</a>
Period	<a href="#">period</a>	Date the source was collected (time interval)	<a href="#">DCMI Period</a>
Type	<a href="#">type</a>	Type of source	<a href="#">DCMI-TYPE</a>
Archive	<a href="#">source</a>	Name of the holding archive	<a href="#">XSD String</a>
Record Group / Series	collection*	Record group / Series in which the source can be found	<a href="#">XSD String</a>
Box	box*	Box in which the source can be found	numeric

Folder	<a href="#">folder*</a>	Folder in which the source can be found	numeric
Identifier	<a href="#">identifier</a>	Local identifier (NAID)	<a href="#">XSD String</a>
Language	<a href="#">language</a>	Language of the source	<a href="#">ISO 639-2</a>
Spatial Coverage	<a href="#">spatial</a>	Landkreis where the source was collected	<a href="#">GeoNames</a>
Temporal Coverage	<a href="#">temporal</a>	Year when the source was collected	<a href="#">ISO 8601-1</a>
Rights	<a href="#">rights</a>	Intellectual property rights information of the digital copy of the source	<a href="#">URI</a>
Date Created	<a href="#">created</a>	Date of creation of the digital copy of the source	<a href="#">ISO 8601-1</a>
Contributors	<a href="#">contributor</a>	Names of persons responsible for creating the resource	<a href="#">XSD String</a>
Format	<a href="#">format</a>	Format of the digital copy	<a href="#">MIME</a>
Subject	<a href="#">subject</a>	Topic of the resource	<a href="#">LCSH</a>
Collection	<a href="#">isPartOf</a>	Collection the item belongs to	<a href="#">XSD String</a>
License	<a href="#">license</a>	License of the item	<a href="#">Creative Commons</a>

\*custom Tropy properties

### *Metadata creation process*

The creation of metadata of the elements of the Fragebogen collection will be done through [Tropy](#), a free open-source software that helps organize and describe photographs and research material. In particular, we



will use the [Tropy Generic metadata template](#), which we will customize in order to better reflect the specific form of the archival documents digitized (see items marked with \* in the table above).

### *Naming conventions*

These naming conventions apply to fields without default values (see below).

- Title: keep the original digitized copy name.
- Contributors: LastName\_FirstName [role], where role is among the following: ProjectLeader, Photographer, MetadataCreator.

### *Default values*

- Creator: [Department of Defense. European Command. Office of Military Government for Germany \(U.S.\). Civil Administration Division. Public Safety Branch. 3/15/1947-9/21/1949.](#)
- Period: name=Denazification; start=1945; end=1948
- Type: Image
- Archive: [National Archives and Records Administration](#)
- Record Group / Series: [260/7062935](#)
- Identifier: <https://catalog.archives.gov/id/7062935>
- Language: German
- Rights: Public domain
- Contributors: Dack\_Mikkel [ProjectLeader]; Gay\_Victor [ProjectLeader]; Hofstetter\_Selina [ProjectLeader]; Stuckatz\_Jan [ProjectLeader]\*
- Format: JPEG
- Subject: Germany—History—1945-1955
- Collection: *The Fragebogen Collection*
- License: Attribution 4.0 International ([CC BY 4.0](#))

\* This default value can be modified

### *Tags*

In addition to metadata, we will associate [tags](#) to the digital material in order to enable us to filter out specific types of images when inputting data in the Fragebogen database or when exporting images to the Fragebogen collection to the repository (see below). We will use 3 tags: Context, Questionnaire, Other.

### *File organization*

To help us navigate the collection through Tropy, we will organize items into hierarchical [lists](#). This hierarchy will reproduce the geographic stratification of the data into Bundesland and Landkreis. The particular file structure of the collection will be as displayed in the Appendix.

### Metadata attached to the overall the Fragebogen collection

The overall Fragebogen collection will be documented following the [Dublin Core](#) metadata standard with the following elements:

Label	DCMI Term	Definition	Data type
Title	<a href="#">title</a>	Name of the collection	<a href="#">XSD String</a>
Creator	<a href="#">creator</a>	Organization responsible for creating the source	<a href="#">XSD String</a>
Period	<a href="#">period</a>	Date the source was collected (time interval)	<a href="#">DCMI Period</a>
Type	<a href="#">type</a>	Type of source	<a href="#">DCMI-TYPE</a>
Archive	<a href="#">source</a>	Name of the holding archive	<a href="#">XSD String</a>
Record Group / Series	collection*	Record group / Series in which the elements of the collection can be found	<a href="#">XSD String</a>
Identifier	<a href="#">identifier</a>	Local identifier (NAID)	<a href="#">XSD String</a>
Language	<a href="#">language</a>	Language of the collection	<a href="#">ISO 639-2</a>
Spatial Coverage	<a href="#">spatial</a>	Location where the elements of the collection were collected	<a href="#">GeoNames</a>
Rights	<a href="#">rights</a>	Intellectual property rights information of the digital copies contained in the collection	<a href="#">URI</a>

Date Created	<a href="#">created</a>	Date of creation of the collection	<a href="#">ISO 8601-1</a>
Date Available	<a href="#">available</a>	Date of public availability of the collection	<a href="#">ISO 8601-1</a>
Contributors	<a href="#">contributor</a>	Names of persons responsible for creating the collection	<a href="#">XSD String</a>
Format	<a href="#">format</a>	Format of the digital copies of the collection	<a href="#">MIME</a>
Subject	<a href="#">subject</a>	Topic of the collection	<a href="#">LCSH</a>
Citation	<a href="#">bibliographic Citation</a>	Citation to the collection	APA
License	<a href="#">license</a>	License of the collection	<a href="#">Creative Commons</a>

*Default values*

- Title: *The Fragebogen Collection*
- Creator: [Department of Defense. European Command. Office of Military Government for Germany \(U.S.\). Civil Administration Division. Public Safety Branch. 3/15/1947-9/21/1949.](#)
- Period: name=Denazification; start=1945; end=1948
- Type: Collection
- Record Group / Series: [260/7062935](#)
- Identifier: [doi of the repository where the collection is deposited]
- Language: English; German
- Spatial Coverage: [Bavaria](#); [Hessen](#); [Baden-Wuerttemberg](#); [West-Berlin](#)
- Rights: Public domain
- Date created: date when the collection was extracted in its entirety and deposited in the repository.
- Date available: date when the collection was made publicly available through the repository.
- Contributors: Dack\_Mikkell [ProjectLeader]; Gay\_Victor [ProjectLeader]; Hofstetter\_Selina [ProjectLeader]; Stuckatz\_Jan [ProjectLeader]; LastName\_FirstName [ImageCapture]
- Format: JPEG
- Subject: Germany—History—1945-1955

- Citation: Dack, Mikkell, Victor Gay, Selina Hofstetter, and Jan Stuckatz. 2025. *The Fragebogen Collection* [database]. RepositoryName.
- License: Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))

*Documentation*

Upon publication of the Fragebogen collection, each element of the collection will be characterized by the above metadata. Moreover, the Fragebogen collection itself will be documented through a peer-reviewed data paper.

- b. Which quality controls of the data will be put in place?

Quality controls will be implemented through a verification of the metadata of each element of the Fragebogen collection during the production stage of the Fragebogen database.

**The Fragebogen database**

- a. Which metadata and documentation will accompany your data?

*Metadata standard*

The Fragebogen database (analysis dataset) will be documented following the [Dublin Core](https://dublincore.org/) metadata standard with the following elements:

Label	DCMI Term	Definition	Data type
Title	<a href="#">title</a>	Name of the dataset	<a href="#">XSD String</a>
Creator	<a href="#">creator</a>	Individuals responsible for creating the dataset	<a href="#">XSD String</a>
Period	<a href="#">period</a>	Date the source was collected (time interval)	<a href="#">DCMI Period</a>
Type	<a href="#">type</a>	Type of dataset	<a href="#">DCMI-TYPE</a>
Identifier	<a href="#">identifier</a>	DOI of the database	<a href="#">URI</a>
Language	<a href="#">language</a>	Language of the dataset	<a href="#">ISO 639-2</a>

Spatial Coverage	<a href="#">spatial</a>	Location where the elements of the dataset were collected	<a href="#">GeoNames</a>
Temporal Coverage	<a href="#">temporal</a>	Year when the elements of the dataset were collected	<a href="#">ISO 8601-1</a>
Rights	<a href="#">rights</a>	Intellectual property rights information of the digital copies contained in the collection	<a href="#">URI</a>
Date Created	<a href="#">created</a>	Date of creation of the collection	<a href="#">ISO 8601-1</a>
Date Available	<a href="#">available</a>	Date of public availability of the collection	<a href="#">ISO 8601-1</a>
Contributors	<a href="#">contributor</a>	Names of persons responsible for creating the dataset	<a href="#">XSD String</a>
Format	<a href="#">format</a>	Format of the digital copies of the collection	<a href="#">MIME</a>
Subject	<a href="#">subject</a>	Topic of the collection	<a href="#">LCSH</a>
Citation	<a href="#">bibliographic Citation</a>	Citation to the collection	APA
License	<a href="#">license</a>	License of the collection	<a href="#">Creative Commons</a>

*Default values*

- Title: *The Fragebogen Database*
- Creator: [Department of Defense. European Command. Office of Military Government for Germany \(U.S.\). Civil Administration Division. Public Safety Branch. 3/15/1947-9/21/1949.](#)
- Period: start=2022-06-01 ; end=2025-05-31
- Type: Dataset
- Identifier: [doi of the repository where the dataset is deposited]

- Language: English
- Spatial Coverage: US occupation zone, Germany
- Temporal Coverage: start=1945 ; end=1948
- Rights: Public domain
- Date created: date when the dataset was deposited in the repository.
- Date available: date when the dataset was made publicly available through the repository.
- Contributors: Dack\_Mikkel [ProjectLeader]; Gay\_Victor [ProjectLeader]; Hofstetter\_Selina [ProjectLeader]; Stuckatz\_Jan [ProjectLeader]
- Format: dta ; csv
- Subject: Germany—History—1945-1955
- Citation: Dack, Mikkel, Victor Gay, Selina Hofstetter, and Jan Stuckatz. 2025. *The Fragebogen Dataset* [database]. RepositoryName.
- License: Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))

### *Documentation*

Beyond the above metadata, the Fragebogen database will be documented by information on a web platform (see *Dissemination*) and a peer-reviewed data paper.

- b. Which quality controls of the data will be put in place?

Quality controls will be implemented through a series of sanity checks against existing datasets during the production stage of the Fragebogen database. In particular, we will compare the content of the Fragebogen database against Falter and Brustein (2015), which is a database of 42,004 NSDAP membership cards created between 1925 and 1933. The cards contain information for several demographic characteristics of NSDAP members: dates of birth and death, civil status, location of residence and birth, occupation, and dates of entry and exit. This data provides a rich cross-section of individuals who joined the NSDAP in its early years. Sample cards were randomly drawn from the NSDAP membership registers (Ortskartei and Reichskartei), which contain 7.2 million original and official individual German Nazi membership cards and are located in the Berlin Document Center (Brustein 1996).

Brustein, William. 1996. *The Logic of Evil: The Social Origins of the Nazi Party, 1925-1933*. New Haven, CT: Yale University Press.

Falter, Jürgen W., and William Brustein. 2015. “Die Mitglieder der NSDAP, 1925– 1945”. Arbeitsbereich Vergleichende Faschismusforschung, FU Berlin; Department of Sociology, University of Minnesota; Forschungsprofessur des Instituts für Politikwissenschaft, JGU Mainz. Datenfile Version 2.0 (unpublished data file).

### 3. Storage and backup during the research process

#### The Fragebogen collection and database

- a. How will the data and metadata be stored and backed up during the research process?

During the collection and production processes of both research outputs, the data (digital copies of archival material, metadata, PTY file, relational database, analysis dataset and script) will be stored and backed up on the Dropbox Business cloud system at Copenhagen Business School. The DeNazDB project has been designated 10 TB of space by CBS IT services for the duration of the project. Dropbox Business offers incremental backup if changes are made to records. Dropbox Rewind allows users to restore entire backups from specific points in time. The storage capacity of the backup services is limited by the storage capacity under the specific DB plan. File metadata is stored in a MySQL-backed database service. Block Servers process files from the Dropbox applications by splitting each into blocks, encrypting each file block using a strong cipher, and synchronizing only blocks that have been modified between revisions. The Block Storage Servers act as a Content-Addressable Storage (CAS) system, with each individual encrypted file block retrieved based on its hash value.

- b. How will the security of data and protection of sensitive data be ensured throughout the research process?

#### *Information security*

Dropbox files and Dropbox Paper documents at rest are encrypted using 256-bit Advanced Encryption Standard (AES). To protect data in transit between Dropbox apps (currently desktop, mobile, API, or web) and our servers, Dropbox uses Secure Sockets Layer (SSL)/Transport Layer Security (TLS) for data transfer, creating a secure tunnel protected by 128-bit or higher Advanced Encryption Standard (AES) encryption. Similarly, data in transit between a Paper client (mobile, API, or web) and the hosted services is encrypted via SSL/TLS. Dropbox for Business is compliant with the following ISO norms: ISO 27001 (Information Security Management), ISO 27017 (Cloud Security), ISO 27018 (Cloud Privacy and Data Protection), and ISO 22301 (Business Continuity Management). Dropbox uses independent third-party auditors to test our systems and controls against some of the most widely accepted security standards, such as ISO 27001 and SOC 2. These reviews occur at least annually and are conducted by independent audit and security firms that are independent. For example, ISO 27001 certification is performed by Ernst & Young CertifyPoint, which maintains ISO accreditation from the Dutch Accreditation Council.

#### *Access*

Access to Dropbox for Business will be limited to the research team (PIs, collaborators, research assistants). Access is granted by setting up a guest researcher email account at CBS, which is secured by dual authentication. The access of research assistants will be limited to the specific folders or regions they will work

on. Both access to the DB for Business folders as well as the guest researcher account at CBS will be limited solely for the duration of their employment for the DeNazDB project. Research assistants will be limited to viewing digitized Fragebogen, but will not be able to download photos to their local machine. Apart from the research team, only CBS IT services responsible for the maintenance of the DB for Business infrastructure have access.

#### *GDPR-compliance*

CBS has entered into a centralized agreement with Dropbox which ensures that data is stored within the EU according to GDPR considerations. Dropbox is certified with ISO/IEC 27018 Code of Practice for Protecting Personal Data in the Cloud and ISO/IEC 27701 Extension to ISO/IEC 27001 and ISO/IEC 27002 for Privacy Information Management. Dropbox provides control and visibility features that can help you manage your data protection obligations, including GDPR compliance obligations. The Dropbox Business admin dashboard enables monitoring team activity, view connected devices, and audit sharing activity. Lost device protection, version history, and file recovery can help protect against accidental loss, damage, or destruction of personal data, and can help with the ability to restore availability and access to personal data in a timely manner in the event of an incident. Two-factor authentication is used for access. Audit logs and activity logs facilitate processing and record keeping.

## **4. Legal and ethical requirements, Storage and backup during the research process**

### **The Fragebogen collection and database**

- a. If personal data are processed, how will the provisions in the legislation on personal data and their security be ensured?

Several considerations are in order regarding the personal data nature of the data collected for the Fragebogen collection and database in order to determine whether it is subject to GDPR regulations.

#### *Personal data processing*

According to [Article 4 \(GDPR. Definitions\)](#), the data of the Fragebogen collection and database could *a priori* constitute personal data (Article 4.1: “personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”).

Moreover, the nature of the operations we employ during the data collection process would constitute “cross-border processing” under the GDPR definitions (Article 4.2: “processing’ means any operation or set of



operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction”; Article 4.23.a.: “‘cross-border processing’ means [...] processing of personal data which takes place in the context of the activities of establishments in more than one Member State of a controller or processor in the Union where the controller or processor is established in more than one Member State”). Note that even if the data were collected in the U.S., which is not a Member State, it will be processed in two Member States: France and Denmark.

Hence, According to Articles 2 and 3 ([GDPR. Material Scope](#) and [GDPR. Territorial scope](#)), the data of the Fragebogen collection and database would be subject to GDPR regulations (“This Regulation applies to the processing of personal data wholly or partly by automated means and to the processing other than by automated means of personal data which form part of a filing system or are intended to form part of a filing system” and “This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union”).

#### *Exemptions to GDPR regulations*

However, two elements imply that the data of the Fragebogen collection and dataset are not subject to GDPR regulations,

First, the data collected concern deceased persons. And according to [Recital 27](#): “This Regulation does not apply to the personal data of deceased persons.”

Moreover, the purpose of the Fragebogen collection and dataset is related to archiving, and historical and scientific research. And according to Article 89.2 ([GDPR. Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes](#)): “Where personal data are processed for scientific or historical research purposes or statistical purposes, Union or Member State law may provide for derogations from the rights referred to in Articles 15, 16, 18 and 21 subject to the conditions and safeguards referred to in paragraph 1 of this Article in so far as such rights are likely to render impossible or seriously impair the achievement of the specific purposes, and such derogations are necessary for the fulfillment of those purposes.”

France and Denmark have provided derogations for these purposes, which our research output falls under.

- France, [Loi Informatique et Liberté](#), Section 4 (Traitements à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques), article 78: “Lorsque les traitements de données à caractère personnel sont mis en œuvre par les services publics d'archives à des fins archivistiques dans l'intérêt public conformément à l'article L. 211-2 du code du patrimoine, les droits prévus aux articles 15,16 et 18 à 21 du règlement (UE) 2016/679 du 27 avril

2016 ne s'appliquent pas dans la mesure où ces droits rendent impossible ou entravent sérieusement la réalisation de ces finalités.”

- Denmark: [Danish Data Protection Act](#) (Act no. 502 of 23/05/2018): Chapter II, Part 3, 10.-(1): “Data as mentioned in Article 9(1) and Article 10 of the General Data Protection Regulation may be processed where the processing takes place for the sole purpose of carrying out statistical or scientific studies of significant importance to society and where such processing is necessary in order to carry out these studies”, and Chapter III, Part 6, 22.-(5): (5) “Articles 15, 16, 18 and 21 of the General Data Protection Regulation shall not apply if the processing of data takes place exclusively for scientific or statistical purposes”.

Finally, according to [Recital 158](#): “Member States should also be authorised to provide for the further processing of personal data for archiving purposes, for example with a view to providing specific information related to the political behaviour under former totalitarian state regimes, genocide, crimes against humanity, in particular the Holocaust, or war crimes.”

Hence, the data collected and processed for the Fragebogen collection and database do not fall under GDPR restrictions. It therefore does not require specific authorization relative to ethical questions.

- b. How will other legal aspects, such as the ownership and property rights of the data, be handled? What is the appropriate legislation?

#### *Reuse of work subject to copyright*

Copyright law to which the content of the Fragebogen collection and database are subject to emanate from the United States, since this is where the base material used in the production process originates. The archival material collected at NARA is in the public domain, so that no specific authorization is needed to reuse and disseminate this material (“Generally, copies of photographic records held by the National Archives may be published without special permission or additional fees. The National Archives does not grant exclusive or non-exclusive publication privileges. Copies of Federal records, as part of the public domain, are equally available to all.”) See <https://www.archives.gov/research/still-pictures/permissions>.

#### *Rights imparted to the producers and authors of the Fragebogen collection and database*

Intellectual property rights law to which the content of the Fragebogen collection and database are subject to emanate from the United States, France, Denmark, and Mexico, as the producers and authors of the Fragebogen collection and database are located across these four countries. We provide below the provisions relative to intellectual property rights law in France that apply to the research outputs.

The Fragebogen collection and dataset are databases that structure its content in an “original” manner per the French accordance of the term “intellectual copyright.” Moreover, it is an ensemble compiled thanks to a substantial investment of its producers. These databases are therefore protected by the *sui generis* regulation

of producers of databases. They are also protected by intellectual copyright in the sense that these databases are organized in an original manner, attesting an “intellectual effort” on the part of the authors.

Since articles L. 341-1 et L. 341-2 of the regulation regarding intellectual copyright relative to *sui generis* regulation of producers of databases do not mention the existence of a plurality of producers, we will refer to intellectual property rights regulations, which specify this case. The Fragebogen collection and database is the output of an intellectual effort of a plurality of authors who collectively created the work under the aegis of other parties—project leaders Mikkel Dack, Victor Gay, Selina Hofstetter, Jan Stuckatz—and the inputs of each contributor is undistinguishable in the ensemble. It is therefore a collective work, whose rights are attributed to the physical person of the project leaders (Mikkel Dack, Victor Gay, Selina Hofstetter, Jan Stuckatz) who disseminate it.

- c. How will potential ethical questions be handled, and deontology codes abided by

The research protocols used to process the digital images collected at NARA (which are in the public domain) will be evaluated by the Internal Review Boards of the respective organizations of project leaders. This item will be completed in future iterations of this DMP.

## 5. Data dissemination and long-term preservation

### The Fragebogen collection

#### *Repository*

Upon completion of the Fragebogen collection, we will deposit all of its content in [Nakala](#) for long-run preservation and dissemination, the repository of the Data Infrastructure [Huma-Num](#) (image files and associated metadata). It is specialized in hosting archival digital collections. This repository stores data on servers managed by Huma-Num and housed in France, at the [calculator center of IN2P3](#). The data are stored through a network storing system of type [NAS](#), which backs-up data daily through a robot of [CCIN2P3](#) using the software [TSM](#) edited by IBM. Metadata are stored as a SQL database (MariaDB), which is backed-up daily. They follow Dublin Core metadata standards. Metadata are also exposed in RDF format via a [Triple-Store](#) (GraphDH), which is also backed-up daily on the Human-Num infrastructure. More technical details can be found [here](#). Nakala is managed by Huma-Num, which ensures its development, evolution, and maintenance. It is free of charge to users and there is no limitation on storage volume for public data.

#### *Dissemination*

While all of the material of the Fragebogen collection will be deposited on Nakala for long-run preservation, only the items tagged as *questionnaires* will be made available to the public—Nakala offers the possibility of designating each item as private or public. The metadata management of Nakala makes it harvestable by external search engines such as [Google Dataset Search](#) or [ISIDORE](#).

The Fragebogen collection will be disseminated under the license Attribution 4.0 International ([CC BY 4.0](#)).

Moreover, Huma-Num offers the possibility of displaying the content of the repository through a web platform powered by [Omeka](#), which is a free, open-source content management system for online digital collections. It will enable us to publish and exhibit the digital items of the Fragebogen collection.

### *Embargo*

Nakala offers the possibility of applying a temporary embargo on public data. We will put the Fragebogen collection under embargo until the data paper describing this dataset is published (or at least when it is out as a working paper).

### *Persistent identifiers*

Each item of the Fragebogen collection deposited in Nakala and made public will be attributed a persistent identifier ([DOI](#)).

## **The Fragebogen database**

### *Repository*

Upon completion of the Fragebogen dataset, we will deposit the analysis datasets in [ADISP-PROGEDO](#) of the data infrastructure [PROGEDO](#), in the section “[Historical Data](#).” It is specialized in hosting numeric datasets. Like Nakala, this repository stores data on servers managed by PROGEDO and housed in France, using secured and daily back-up processes. ADISP-PROGEDO is managed by the PROGEDO infrastructure, which ensures its development, evolution, and maintenance. It is free of charge to users and there is no limitation on storage volume.

### *Dissemination*

The metadata management of ADISP-PROGEDO makes it harvestable by external search engines such as [Google Dataset Search](#) or [ISIDORE](#). Its management is powered by the [Dataverse](#) system. The deposited data will be available to the public upon registration with an institutional affiliation and a motivated request.

The Fragebogen analysis dataset will be disseminated under the license Attribution 4.0 International ([CC BY 4.0](#)).

Note that we will disseminate the content of the Fragebogen analysis dataset through a webmapping platform that will enable users to visualize its content. While users will be able to generate maps and export them, they will not be able to download the underlying analysis dataset without going through ADISP-PROGEDO.

### *Embargo*

ADISP-PROGEDO offers the possibility of applying a temporary embargo on deposited data. We will put the Fragebogen dataset under embargo until the data paper describing this dataset is published (or at least when it is out as a working paper).

### *Persistent identifiers*

The Fragebogen analysis dataset deposited in ADISP-PROGEDO will be attributed a persistent identifier ([DOI](#)).

## **6. Responsibilities and resources in terms of data management**

### **The Fragebogen collection and database**

Victor Gay is responsible for the data management aspects of the Fragebogen collection and database. He will ensure they follow [FAIR](#) data management principles. He will allocate a research time of 2 person.month to these aspects.

## Appendix

Folder structure of the Fragebogen collection (hierarchical lists in Tropy):

- Bavaria Oberand Mittel Franken
  - Eichstett
  - Eschenbach
  - Muenchberg
  - Rehau
  - Stadtsteinach
  - Wunsiedel
- Berlin Sector
- Hesse
  - Alsfeld
  - Bergestrasse
  - Biedenkopf
  - Buedingen
  - Darmstadt
  - Dieburg
  - Dillkreis
  - Erbach
  - Eschwege
  - Frankenberg
  - Frankfurt
  - Friedburg
  - Fritzlar-Homberg
  - Fulga
  - Gelnhausen
  - Giessen
  - Gross Gerau
  - Hanau
  - Hersfeld
  - Hofgeismer
  - Huenfeld
  - Kassel
  - Lauterbach
  - Limburg
  - Maintaunus Kries
  - Marburg

- Melsungen
- Obertanus-Kreis
- Offenbach
- Rhinegau
- Rotenburg
- Schluchtern
- Schwalbach
- Usingen
- Waldeck
- Weilburg
- Wetzlar
- Wiesbaden
- Witzenhausen
- Wolfhagen
- Ziegenhain
- Niederbayen-Oberpfalz
  - Deggendorf
  - Landshut
  - Mainberg
  - Oberviechtach
  - Passau
  - Riedenberg
  - Rottenburg
  - Viechtach
  - Vilsbiburg
- Oberbayern
  - Aichach
  - Altoetting
  - Donauworth
  - Freising
  - Friedberg
  - Munich
  - Pfaffenhofen
  - Tolz
  - Wasserburg
  - Wolfratshausen
- Schwaben
  - Augsburg
  - Dillingen

- Fuessen
- Guenzburg
- Illertissen
- Kaufbeuren
- Kempten
- Krumbach
- Markt Oberdorf
- Mindelheim
- Neu Ulm
- Schwabmunchen
- Wertingen
- Wuerttemberg
  - Boeblingen
  - Esslingen
  - Heidelberg
  - Karlsruhe
  - Sinsheim
  - Stuttgart
- Wuerttemberg-Baden
  - Aalen
  - Backnang
  - Bruchsal
  - Buchen
  - Crailsheim
  - Esslingen
  - Geoppingen
  - Heidenheim
  - Heilbronn
  - Karlsruhe
  - Kneuzelsau
  - Leonberg
  - Ludwigsburg
  - Mannheim
  - Mergentheim
  - Oehringen
  - Pforzheim
  - Schwabisch Gmund
  - Schwabisch Hall
  - Stuttgart



- Tauberbischofsheim
- Ulm
- Waiblingen