



European Holocaust Research Infrastructure
Implementation Phase
HORIZON-INFRA-2023-DEV-01-02
GA no. 101129732

Deliverable 1.2

Project Data Management Plan (DMP)

Mike Bryant
NIOD Institute for War, Holocaust and Genocide
Studies

Reto Speck
NIOD Institute for War, Holocaust and Genocide
Studies

Start: February 2024 [M1]
Due: July 2024 [M6]
Actual: July 2024 [M6]



[EHRI is funded by the European Union](#)

Document Information

Project URL	https://www.ehri-project.eu
Document URL	https://www.ehri-project.eu/deliverables-ehri-ip-2024-2026
Deliverable	D1.2 Project Data Management Plan (DMP)
Work Package	WP1
Lead Beneficiary	1 KNAW-NIOD
Relevant Milestones	MS2
Dissemination level	Public
Contact Person	Mike Bryant (m.bryant@niod.knaw.nl)
Abstract (for dissemination)	This Data Management Plan (DMP) reports on the types of data used and processed by the EHRI-IP project, the Findable, Accessible, Interoperable and Reusable (FAIR) characteristics of that data, and the project's approach to improving its FAIR compliance.
Management Summary	(required if the deliverable exceeds more than 25 pages) N/A

Document Control

Version	Date	Change	Author
0.1	03/07/2024	Initial version	Mike Bryant

Glossary

API	Application Programming Interface
CC	Creative Commons
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
EHRI-IP	European Holocaust Research Infrastructure Implementation Phase
EOSC	European Open Science Cloud
FAIR	Findability - Accessibility - Interoperability - Reusability
GDPR	General Data Protection Regulation
LOD	Linked Open Data
ORCID	Open Researcher and Contributor Identifier
ORDP	Open Research Data Pilot
PID	Persistent Identifier
PDF	Portable Document Format
RDF	Resource Description Format
RDM	Research Data Management
SSHOC	Social Sciences & Humanities Open Cloud
WP	Work Package

Table of Contents

1. Project Abstract.....	6
2. Data Summary.....	6
2.1. Procedural data.....	6
2.2. Dissemination.....	7
3. FAIR Data.....	8
3.1. Findability.....	8
3.2. Accessibility.....	8
3.3. Interoperability.....	8
3.4. Reusability.....	9
4. Other research outputs.....	9
4.1. Deliverables.....	9
4.2. Code & Software.....	9
4.3. Scientific publications.....	10
5. Allocation of resources.....	10
6. Data security.....	10
7. Ethical aspects.....	10
8. Other issues.....	11
8.1. General Data Protection Regulation (GDPR).....	11
8.2. Horizon Europe Open Access requirements.....	11

1. Project Abstract

The mission of the European Holocaust Research Infrastructure (EHRI) concerns trans-national Holocaust research, commemoration and education, and its main challenge is the wide dispersal of sources and expertise across many institutions. EHRI overcomes such fragmentation by connecting sources, institutions and people. The EHRI Portal enables online access to information about Holocaust sources, no matter where they are located. The Conny Kristel Fellowship gives researchers access to the resources of the world's twenty leading Holocaust archives. EHRI's extensive programme of networking and training brings people together. Last but not least, EHRI promotes innovative tools that advance the digital transformation of Holocaust research.

EHRI was added to the ESFRI roadmap in 2018 and is working towards transitioning from a succession of projects to a permanent organisation in the form of a European Research Infrastructure Consortium (ERIC). A successful Preparatory Phase project (EHRI-PP) ran from December 2019 to May 2023 and performed the necessary strategic, legal, financial and technical work necessary to prepare an ERIC application. This Implementation Phase (EHRI-IP) project will run for two years (2024-2025) and will guide the of the ERIC through initial implementation and early operation phases. This Data Management Plan (DMP) summarises the types of data used in the EHRI-IP project, the metadata formats used (where relevant), and security and ethical aspects related to the processing of data.

2. Data Summary

EHRI-IP is primarily a strategic and organisational project. In the course of its operation, non-public procedural data is generated, along with that related to dissemination and public outreach, with the main outputs being written documentation and reports, along with code that is produced in the process of developing and refining the ERIC's technical architecture (see section "Other research outputs".) Other types of data and metadata used and generated by the EHRI-3 project and in the operation of EHRI-ERIC are considered out-of-scope for this DMP. Readers seeking information about Data Management for the EHRI-3 project should consult EHRI-3 Deliverable D12.2 Updated Data Management Plan.¹ Information about EHRI-ERIC's future data management activities can consult EHRI-PP Deliverable D7.3 Data Management Plan EHRI-RI², due to be revised and updated in the course of EHRI-IP WP4 T4.3.

2.1. Procedural data

Procedural data is here defined as that relating to the internal organisation, communication, and operation of the EHRI-IP project. Additional non-public data relating to the EHRI-IP project is generated by consortium partners in the course of their activities and is also considered out-of-scope of this plan. Personal data may be generated by surveys, registration for public events, and other activities relating to project management and will be processed according to relevant statutes, in particular the GDPR.

1

<https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D12.2-%20-%20Updated%20Data%20Management%20Plan.pdf>

2

<https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D7.3-%20-%20Data%20managemen%20plan%20EHRI-RI.pdf>

EHRI's main project management tool is Basecamp, an online software-as-a-service (SaaS), in which groups are maintained for the overall project consortium, project management board (PMB) and individual work packages. Basecamp data is only accessible to project consortium members and participants in project advisory activities. Additional procedural data is generated by the use of email and online communication tools such as Zoom and Microsoft Teams, and is considered ephemeral and non-sensitive in nature. Data such as minutes for project meetings and attendance records are managed by consortium members using institution-specific procedures and will be uploaded onto Basecamp where required.

2.2. Dissemination

EHRI-IP uses a number of channels for dissemination, including a mailing list, the project website, and social media accounts:

Mailing List

The EHRI mailing list is managed using the MailChimp SaaS. Users can optionally sign up to the mailing list via links on the EHRI project, Document Blog, and Portal websites in order to receive the monthly EHRI Newsletter and other selected communications. The sign-up form³ is GDPR-compliant and Personally Identifiable Information (PII) related to recipients of the mailing list (name, email address, and professional field) is not shared outside the MailChimp platform but may be used on an aggregate basis for internal project analytics.

Website

Operation of the EHRI Project website is discussed in the EHRI-PP DMP and forthcoming EHRI-RI DMP. The website hosts newsletter articles, information about upcoming EHRI events, and other news relating to Holocaust research such as information about participants in EHRI's Conny Kristel Fellowship programme. The website is currently hosted on the Drupal platform and is deployed, managed, and backed up using EHRI's Ansible-based Infrastructure-as-Code (IaC) tools.

Analytics for the EHRI Project website and other EHRI sites — including the Portal, Document Blog, Online Editions and Geospatial Repository — are collected anonymously by an EU-based web analytics provider and exported periodically to a database accessible to EHRI's Dissemination team. Anonymous and GDPR-compliant web analytics may be used in aggregate to inform project members about the constitution and behaviour of EHRI's audience.

Social media channel

EHRI maintains social media accounts on Facebook⁴, LinkedIn⁵, and X (formerly Twitter)⁶. Access to post on these accounts, along with related analytics, is restricted to specific members of the Dissemination team. Social media analytics may be used in aggregate to inform project members about the make-up of EHRI's audience and their preferences with regard to EHRI's output and programmes.

³ <https://ehri-project.us2.list-manage.com/subscribe?u=26b57b75c8a6999778bc4495d&id=58024f80e2>

⁴ <https://www.facebook.com/EHRIproject/>

⁵ <https://nl.linkedin.com/company/ehri-european-holocaust-research-infrastructure>

⁶ <https://x.com/EHRIProject>

3. FAIR Data

EHRI projects to the date and the EHRI-RI build extensively on FAIR data, as detailed in the EHRI-RI DMP. EHRI is a Governing Board member of the Social Science & Humanities Open Cloud (SSHOC) Open Science Cluster and outputs such as the EHRI Portal part of the SSHOC Marketplace, and by extension the European Open Science Cloud (EOSC) ecosystem. EHRI-RI is committed to expanding the FAIR compliance of its data through a number of initiatives described below.

3.1. Findability

Under Work Package 4 (Technical Development) preparations for the introduction of persistent Identifiers (PIDs) throughout the EHRI-RI ecosystem is ongoing. EHRI is also participating in the FAIR-IMPACT support programme for creating EOSC-compliant PID policies and engaging with the development of the FAIRCORE4ESOC Compliance Assessment Toolkit (CAT)⁷, intended for assisting the creation of machine-actionable policies, supported by automated policy evaluation and compliance monitoring.

While EHRI's scholarly research outputs are assigned PIDs (specifically Digital Object Identifiers: DOIs) via the HAL research repository, other textual outputs, particularly "grey literature" (including posts on EHRI's Document Blog and Online Editions) currently are not. Establishment of the EHRI-ERIC will make possible the creation of a new, EHRI-specific DOI namespace through which we intend to mint DOIs for these resources. This namespace will be available for use by members of the EHRI-ERIC consortium. Preparations for the assignment of PIDs (specifically Archival Resource Keys (ARKs)) for descriptions of collection-holding institutions (CHIs) and archival descriptions is also in progress and planned for the first year of ERIC operation.

3.2. Accessibility

Also under WP4, EHRI-IP will continue to develop and expand the range of APIs and services that provide access to EHRI FAIR data. EHRI-IP is committed to facilitating the broadest possible access for material it produces and making this available in formats that suit researchers with a wide range of skills, specialties, and preferences. All EHRI-RI FAIR data is free and available without geographical restriction. A small subset of user generated data available on the EHRI Portal (specifically, public notes left by users) requires the use of an account that is free to all users who acknowledge the EHRI privacy policy⁸ and agree to abide by the EHRI terms of service.⁹

3.3. Interoperability

The EHRI-RI DMP details the data and metadata formats used by EHRI's FAIR data across its range of services, including the EHRI Portal, Online Editions, and Geospatial Repository. These include, but are not limited to, International Council of Archives (ICA) conceptual standards for the description of archival material, institutions and authorities, along with relevant technical standards for the encoding and transmission of this data in structured formats such as XML. EHRI's range of public APIs include established protocols such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), a GraphQL endpoint, and a RESTful JSON-based

⁷ <https://faircore4eosc.eu/eosc-core-components/compliance-assessment-toolkit-cat>

⁸ <https://portal.ehri-project.eu/data-policy>

⁹ <https://portal.ehri-project.eu/terms>

search interface mirroring that of the Portal website, allowing EHRI services to be interoperated in a range of contexts.

EHRI-IP WP4 and parallel initiatives will also continue to develop the EHRI-RI Linked Open Data (LOD) repository¹⁰ — currently available as a preview/test — which makes data from the EHRI Portal (and in future other sources) available in the Resource Description Format (RDF), with links to a number of popular LOD sources such as WikiData and GeoNames. The EHRI LOD repository will also offer a SPARQL endpoint, allowing users to query it in a flexible and interoperable manner.

3.4. Reusability

EHRI FAIR data is intended to be reusable and is, in cases such as archival descriptions provided by CHIs, already reused by EHRI. The default EHRI content licence is CC BY 4.0, which allows users to share and adapt content under the condition that they provide attribution to the author(s) of the work. EHRI-IP WP4 will seek to make content licences more explicit on EHRI's web properties.

4. Other research outputs

4.1. Deliverables

Of the 24 deliverables in EHRI-IP, 21 are public and 3 are deemed “sensitive” with internal-only access. Internal-only documents are made available to project participants via the project's Basecamp management software and uploaded to the EC Participant Portal. Public EHRI-IP deliverables are **not** currently assigned persistent identifiers (PIDs) such as DOIs. EHRI-ERIC will explore the possibility of assigning PIDs (potentially via a platform such as HAL) to historical EHRI project documentation and public outputs.

Public deliverables will be made available in unencrypted Archive PDF/A format on the EHRI project website¹¹ EHRI-IP deliverable page¹² within one month of being uploaded to the EC Participant Portal. All deliverables will be archived on internal systems at KNAW-NIOD. Other public outputs will be delivered in formats that have wide compatibility across devices and computing environments, where possible reducing the use of patent-encumbered and/or proprietary formats. EHRI-ERIC will investigate the use of sustainable repositories such as HAL for the depositing of historical EHRI deliverables and other outputs, including those relating to EHRI-IP.

4.2. Code & Software

In addition to documentation, EHRI-IP may produce new code in the process of maintaining and developing EHRI's technical services.

All code repositories will contain at minimum a README file in Markdown format detailing the purpose of the software and technical information required to build and/or run it.

¹⁰ <https://lod.ehri-project-test.eu/>

¹¹ <https://www.ehri-project.eu/>

¹² <https://www.ehri-project.eu/deliverables-ehri-ip-2024-2026>

Code will be made available on Github in public repositories under the EHRI organisation¹³, with the exception of cases where private repositories are required for operational and/or security purposes (e.g. Infrastructure-as-Code configuration files containing sensitive information.)

All non-derivative code repositories produced under EHRI-IP will be licensed using the European Union Public Licence (EUPL) version 1.2¹⁴, an OSI-certified¹⁵ GPL-v3 compatible F/OSS licence available in 22 European languages. EUPL-1.2 contains copyleft provisions intended to ensure that derivative works (e.g. software produced using the licenced code) use a compatible open-source licence. This makes it less permissive (flexible) than alternatives such as the BSD or MIT licences.

4.3. Scientific publications

Scientific publications produced under EHRI-IP will be uploaded to EHRI's organisational repository on HAL.¹⁶ HAL is an open archive repository currently undergoing CoreTrustSeal¹⁷ certification as part of the HALiance project.¹⁸ Deposited papers are available with digital object identifiers (DOIs) and are licensed under CC-BY 4.0.¹⁹

5. Allocation of resources

Costs incurred to ensure the long-term availability of EHRI-IP outputs will be borne by the overall EHRI technical budget spread across the EHRI-3 and EHRI-IP projects, and that of the EHRI-ERIC organisation after its launch. Responsibility for ensuring observance of relevant statutes lies with the Data Protection Officer at the EHRI-IP coordinating institution (KNAW-NIOD).

6. Data security

To ensure the safety and long-term availability of EHRI-IP outputs, regular backups will be made in a manner that places digital assets in a minimum of two distinct off-site storage locations. Best-practices in the management of infrastructure will be observed, including:

- Ensuring technical components can be updated rapidly to patch bugs and security vulnerabilities
- Using modern cryptographic techniques for password hashing
- Using multi-factor authentication (MFA) for infrastructure administration related tasks

7. Ethical aspects

Work packages that gather survey data will abide by the personal data restrictions of the GDPR and where possible anonymise individual responses, though due to the nature of the project it is unlikely that significant quantities of privacy-relevant data would be involved. Raw survey data will not be disseminated beyond those conducting the research and, on conclusion of the research, will be either deleted or archived in secure storage for a period of no more than five

¹³ <https://github.com/EHRI>

¹⁴ https://commission.europa.eu/content/european-union-public-licence_en

¹⁵ <https://opensource.org/license>

¹⁶ <https://hal.science/EHRI>

¹⁷ <https://www.coretrustseal.org/>

¹⁸ <https://www.ccsd.cnrs.fr/en/haliance-project/>

¹⁹ <http://creativecommons.org/licenses/by/4.0/>

years. Participants will be offered the opportunity to agree to these terms prior to taking part in research activities.

8. Other issues

EHRI-IP is a Horizon Europe project and as such abides by the following pieces of applicable legislation and guidance:

8.1. General Data Protection Regulation (GDPR)

In May 2018 the new European Regulation on data protection and privacy, the General Data Protection Regulation (GDPR), became applicable for all member states. The aim of the GDPR is to strengthen the rights of individuals, so that they can control their own data. Controllers and processors of personal data must have technical procedures and policies to implement the data protection principles (Accountability). The GDPR Principles are: Lawfulness, fairness and transparency; Purpose limitation; Data minimization; Accuracy; Storage limitation, and Integrity and confidentiality.

8.2. Horizon Europe Open Access requirements

Horizon Europe's documentation on Open Science²⁰ supports the "as open as possible, as closed as necessary" model of scientific publication. All publications arising from Horizon Europe funding must be made available as open access as soon as possible after publication, either by depositing in an open access public repository, or by publication in an open access journal.²¹

²⁰ <https://data.europa.eu/doi/10.2777/18252>

²¹ <https://www.openaire.eu/how-to-comply-with-horizon-europe-mandate-for-publications>