European Holocaust Research Infrastructure
H2020-INFRAIA-2019-1
GA no. 871111

Deliverable 12.2

Updated Data Management Plan

Mike Bryant
KCL / NIOD KNAW

Rebecca Dillmeier
USHMM

René van Horik
DANS KNAW

Rachel Pistol
KCL

Start: September 2020 [M1]
Due: August 2023 [M36]
Actual: July 2023 [M35]

Document Information

Project URL	https://www.ehri-project.eu
Document URL	https://www.ehri-project.eu/deliverables-ehri-3-2020-2024
Deliverable	D12.2 Updated Data Management Plan
Work Package	WP12
Lead Beneficiary	14 - USHMM
Relevant Milestones	MS3
Dissemination level	Public
Contact Person	Rebecca Dillmeier (rdillmeier@ushmm.org)
Abstract (for dissemination)	This report contains the final version of EHRI-3's Data Management Plan (DMP) building on Deliverable D12.1 developed using DMPOnline. The DMP builds on the FAIR data principles, which provide guidelines for making data Findable, Accessible, Interoperable, and Reusable. The plan outlines the data sources EHRI-3 has built and manages, as well as their technical specifications.
Management Summary	(required if the deliverable exceeds more than 25 pages) N/A

Table of Contents

Glossary

Research data: information (particularly facts or numbers) collected to be examined and considered, and to serve as a basis for reasoning, discussion or calculation.

API: Application Programming Interface

CDEC: Centro di Documentazione Ebraica Contemporanea

CPA: Content Provider Agreement

CHI: Collection Holding Institution

DMP: Data Management Plan

DOI: Digital Object Identifier

EAC: Encoded Archival Context

EAD: Encoded Archival Description

EAG: Encoded Archival Guide

EC: European Commission

FAIR: Findability - Accessibility - Interoperability - Reusability

GDPR: General Data Protection Regulation

GML: Geographic Markup Language

H2020: Horizon2020

ISAD(G): International Standard for Archival Description (General)

ISO: International Organization for Standardization

KG: Knowledge Graph

LOD: Linked Open Data

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting

OGC: Open Geospatial Consortium

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting 2.0

ORCID: Open Researcher and Contributor IDentifier

ORDP: Open Research Data Pilot

PID: Persistent IDentifier

RiC: Records-in-Context

RDF: Resource Description Framework

SCORM:Shareable Content Object Reference Model

SKOS: Simple Knowledge Organisation System

TEI: Text Encoding Initiative

URI: Uniform Resource Identifier

URL: Uniform Resource Locator

XML: eXtensible Markup Language

1. Introduction

The EHRI-3 project runs from 2020 to 2024. It deepens the integration of Holocaust archives and research that has been undertaken by EHRI since 2010 when EHRI-1 was first established. EHRI exists as a federated organisation organised into work packages responsible for producing different work products and assets.

The EHRI-3 project will lead to:

- a substantial increase in the coverage of the EHRI Portal, particularly with regards to the holdings of micro-archives;
- the development of new digital tools that connect dispersed Holocaust sources;
- the delivery of state-of-the-art training and education opportunities for researchers and archivists;
- the continuation and enhancement of the Conny Kristel Fellowship Programme.

The EHRI-3 project had a close relationship with both the EHRI Preparatory Phase Project (EHRI-PP), which ran from 2019 to 2023, and the subsequent EHRI Implementation Phase Project which is intended to run from 2024 until early 2025. The aim of the EHRI-PP and EHRI-IP projects were and are to transform EHRI from a fixed-duration project into a permanent European Research Infrastructure Consortium (EHRI-ERIC) that will help secure the future of trans-national Holocaust research, commemoration and education. Policies for data management will be an important component of the permanent EHRI Research Infrastructure (EHRI-RI). The data management and open science policies of EHRI-PP and EHRI-IP have a close relation with those of EHRI-3.

A main difference between the projects is that the EHRI-3 data management policies have a more descriptive character based on available (and — during the lifetime of the EHRI-3 project — enhanced) data collections and services whereas the EHRI-PP and EHRI-IP data management policies are more prescriptive concerning the management of data in the future EHRI Research Infrastructure (RI).

This deliverable is defined in the EHRI-3 project proposal as an Open Research Data Pilot (ORDP)[1], as initiated by the European Commission. Background information on the ORDP is given below. The main part of this report describes the process to define the EHRI-3 Data Management Plan. The appendix contains the EHRI-3 Data Management Plan, updated from the initial plan published in 2022.

2. Open Research Data Pilot

In the Horizon2020 Framework program, the European Commission launched a pilot for open access to research data and this DMP follows that model as an ORDP. Open access concerns the right to access and reuse of digital research data under the terms and conditions set out in the grant agreement. The ORDP aims to improve and maximise access to and re-use of research data.[2] The EU recognizes that some research data cannot be made open and applies the principle that data should be "*as open as possible, as closed as necessary*." It is therefore possible to opt out of research data sharing at any stage - before or after the signature of the grant agreement - but reasons have to be given e.g. for intellectual property rights (IPR) concerns, privacy/data protection concerns, national security

---

[1] https://www.openaire.eu/what-is-the-open-research-data-pilot
[2] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

concern, if it would run against the main objective of the project or for other legitimate reasons.[3]

There are two pillars to the ORDP: developing a Data Management Plan (DMP) and providing open access to research data, if possible. The conditions for the EHRI-3 project to adhere to, are:

- Develop (and keep up-to-date) a Data Management Plan (DMP).
- Deposit research data in a research data repository.
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate research data related to EHRI.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate research.

The ORDP applies to:

- The data (and metadata) needed to validate results in scientific publications.
- Other curated and/or raw data (and metadata) that you specify in the DMP.

3. Data Management Planning

DMPs are a key element of good data management and the creation of a DMP is encouraged by the EU.[4] According to the EU a DMP describes the data management life cycle for the data to be collected, processed and/or generated by a project. As part of making research data Findable, Accessible, Interoperable and Re-usable (FAIR), a DMP should include information on:

- what data will be collected, processed and/or generated
- the handling of data during and after the end of the project
- which methodology and standards will be applied
- whether data will be shared/made open access and
- how data will be curated and preserved (including after the end of the project)

The EHRI-3 DMP is based on the Horizon 2020 DMP template as published by the European Commission.[5] This deliverable is an update of the DMP that was published earlier and contains the final version of the EHRI-3 DMP.

FAIR principles

The FAIR principles are a key component of the H2020 DMP template. The principles intend to improve the findability, accessibility, interoperability and reusability of digital assets. For each of the four principles practical implementation guidelines are stated.[6] The figure below can act as a reference of the meaning of the FAIR principles. The compliance list contains action points that help to assess the "FAIRness" of data collections.

---

[3] See General Annex L of the 2017 Work Programme adopted on 25 July 2016).
[4] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm#A1-template
[5] Guidelines on FAIR Data Management in Horizon 2020, version 3 (July 2016)
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf#page=10
[6] https://www.go-fair.org/fair-principles/

Figure 1. FAIR·principles. Source: https://ccafs.cgiar.org/open-access-and-fair-principles
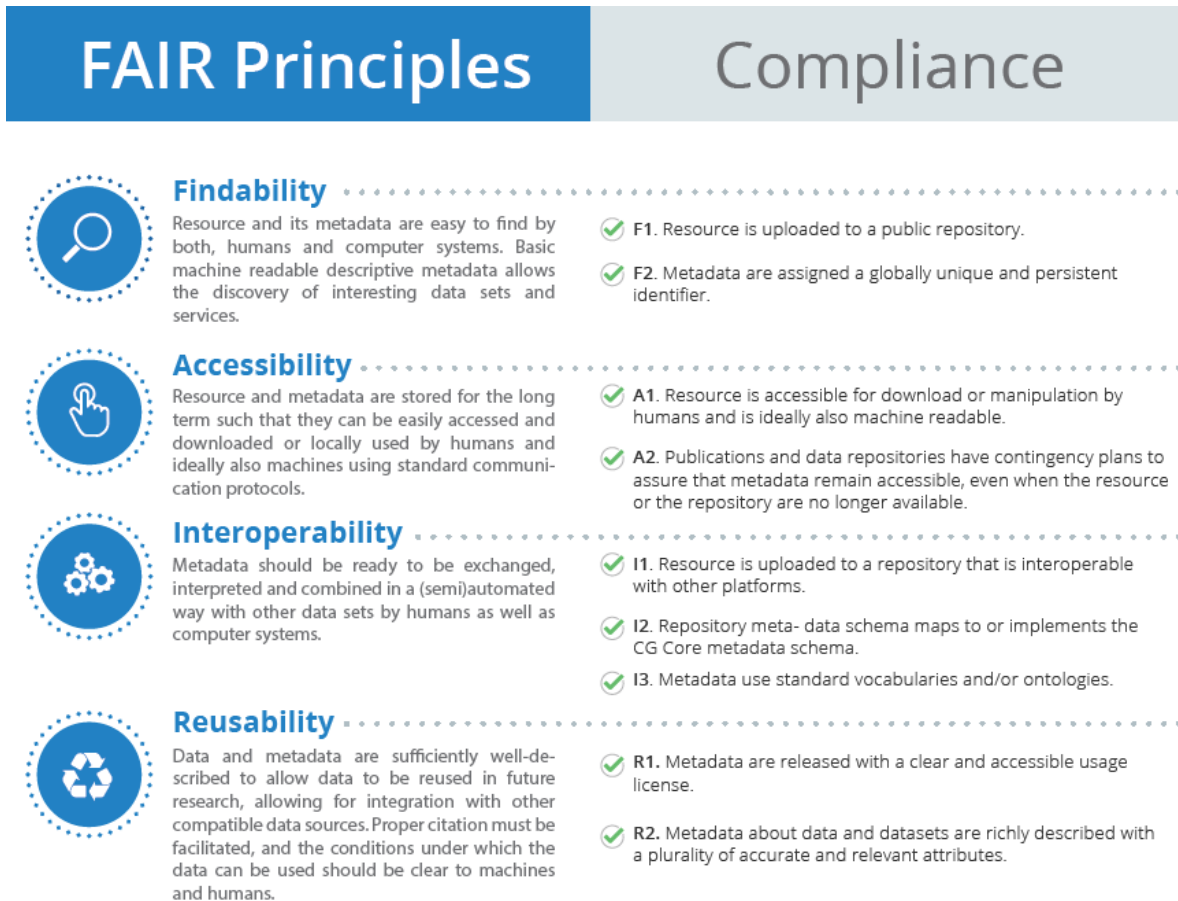
4. Compilation of the EHRI-3 Data Management Plan

The EHRI-3 Data Management Plan is based on the Horizon 2020 DMP Template.[7] The DMPonline Tool[8] was used to create the first version of the EHRI-3 Initial DMP along with this updated version. "DMPonline is a web-based tool that supports researchers to develop data management and sharing plans. It contains the latest funder templates and best practice guidelines to support users to create good quality DMPs."[9]

As stated in the description above, the DMPonline tool supports researchers. However, the EHRI-3 project has a broader target group. The data management issues are not only directed towards researchers but also other stakeholders such as archivists, the "general public", and the EHRI project partners that manage, create and share data collections.

The EHRI-3 project manages a wide range of data collections, ranging from websites, information systems to publications and this complicates the "one size fits all" character of the H2020 DMP template. The H2020 DMP is based on the assumption that a researcher or group of researchers creates or (re)uses a specific data collection in relation to a specific

---

[7] The H2020 DMP Template is available at:
https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.
[8] See: https://dmponline.dcc.ac.uk/
[9] https://www.dcc.ac.uk/dmponline

research project. The mission of the EHRI-3 project concerns "Trans-national Holocaust research, commemoration and education … and its main challenge is the wide dispersal of sources and expertise across many institutions".[10]
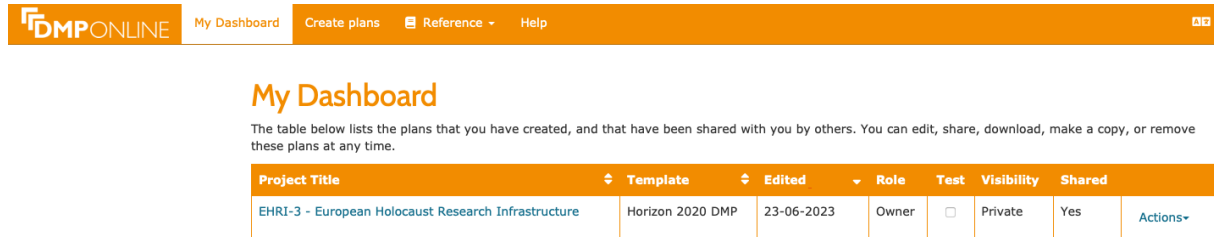

Figure 2. DMPonline Tool detail of screen: EHRI-3 DMP is using Horizon 2020 DMP template. Updated June 2023.

The DMPonline Tool contains a facility to download the DMP. The EHRI-3 DMP is included in this report in the appendix.

5. Next steps

Since the initial report D12.1, various EHRI-3 work packages have progressed in their planned activities, giving us a clearer view of the size and rate of growth of digital assets, and this plan has been updated accordingly. In some cases, however, uncertainty still remains. The EHRI Geospatial Repository, for example, is a platform where data management practices need to be further refined, and where improved FAIR compliance will come over time as usage increases and experience is gained managing fairly complex types of digital assets.

The development of clear and easy to follow policies and procedures for data management are essential for its success. Digital asset management and digital preservation require policy and people, in addition to technology, to define and manage systematic procedures. Even the most comprehensive digital asset management and digital preservation systems are limited without clear definitions of policy and procedure to guide them.

Key features of a successful data management plan that serves digital asset management and digital preservation functions include:

- Clearly defined goals for data management, including digital preservation (why)
- Clearly defined preservation worthy assets (what)
- Clearly defined procedures for EHRI participants to follow, as well as dissemination of those procedures (how)
- Implementation of a plan with an understanding of the associated risks and costs

While this is not a technology initiative alone, there is no doubt that technology plays an essential role in the management of and access to digital resources.

Based on the "EHRI-3 DMP" (see below) and in consultation with EHRI stakeholders, focus areas will be defined that will improve the "FAIRness" of the data collections created and (re)used in the EHRI project.

These focus areas are:

- Define metadata policies for all data collections in EHRI-3
- Implementation of persistent identifier (PID) services in EHRI

---

[10] https://www.ehri-project.eu/about-ehri

- (e.g. ORCID for people[11], DOI for publications[12], Handle for data[13])
- Increase durability of data repositories by applying for the CoreTrustSeal certification[14]
- Stimulate usage of certified repositories for data collections. For this, registries of research data repositories can be consulted.[15] Assess (if relevant) usage of repository services (e.g. Zenodo for publications, Dataverse for data[16])
- Increase durability and usability of online information (e.g. using "robustify link" services[17] (including storage of information in Internet Archive) for EHRI Blogs and other online resources)

---

[11] See: https://orcid.org/

[12] Zenodo is an example of a service that creates DOIs for publications (and links the DOI with the publication). See: https://zenodo.org/

[13] B2Handle provides services for managing "handles". See: https://eudat.eu/catalogue/b2handle

[14] See: https://www.coretrustseal.org/

[15] See for instance: https://www.re3data.org/

[16] See: https://dataverse.org/. An example of a Dataverse repository that manages Holocaust data can be found at: https://ssh.datastations.nl/dataverse/root?q=holocaust

[17] See: https://robustlinks.mementoweb.org/. The service was tested in 2020 with a Blogpost created by the EHRI-project. See: https://archive.ph/cje6q.

EHRI-3 - European Holocaust Research Infrastructure - DMP

A Data Management Plan created using DMPonline

Creators: Mike Bryant, Rebecca Dillmeier, René van Horik
Affiliation: EHRI-3 project
Funder: European Commission (Horizon 2020)
Template: Horizon 2020 DMP
ID: 72794
Start date: 01-09-2020
End date: 31-08-2024
Last modified: 23-06-2023

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. The EHRI-3 project (that runs from 2020 until 2024) deepens the integration of Holocaust archives and research that has been undertaken by EHRI since 2010.

The project will lead to:

- a substantial increase in the coverage of the EHRI Portal, particularly with regards to the holdings of micro-archives;
- the development of new digital tools that connect dispersed Holocaust sources;
- the delivery of state-of-the-art training and education opportunities for researchers and archivists;
- the continuation and enhancement of the Conny Kristel Fellowship Programme.

EHRI-3 - European Holocaust Research Infrastructure - Initial DMP

1. Data summary

Provide a summary of the data addressing the following issues:

- **State the purpose of the data collection/generation**
- **Explain the relation to the objectives of the project**
- **Specify the types and formats of data generated/collected**
- **Specify if existing data is being re-used (if any)**
- **Specify the origin of the data**
- **State the expected size of the data (if known)**
- **Outline the data utility: to whom will it be useful**

The purpose of the data collection and data generation in the EHRI-3 project is to increase the coverage of the EHRI Portal, to develop new tools that connect dispersed Holocaust sources, to deliver training and education opportunities for researchers and archivists, and to continue the Conny Kristel Fellowship Programme.

To meet the aims of the EHRI-3 project, data will be collected, processed and/or generated. The table below provides an overview of the data collections that play a role in the EHRI-3 project.

#	Data Collection	Current Web address
01	EHRI Website	https://www.ehri-project.eu/
02	EHRI Portal	https://portal.ehri-project.eu/
	Country Reports in EHRI Portal	https://portal.ehri-project.eu/countries
	Archival Institutions in EHRI Portal	https://portal.ehri-project.eu/institutions
	Archival Descriptions in EHRI Portal	https://portal.ehri-project.eu/units
	Vocabularies in EHRI Portal	https://portal.ehri-project.eu/vocabularies
	Authority Sets in EHRI Portal	https://portal.ehri-project.eu/sets
03	EHRI Document Blog	https://blog.ehri-project.eu/ Digital images and visualisations related to the blog are accessible via web address: https://visualisations.ehri-project.eu/

04	Online Editions	https://www.ehri-project.eu/new-ehri-online-editions Online edition 1: https://begrenzte-flucht.ehri-project.eu/ Online edition 2: https://early-testimony.ehri-project.eu/ Online edition 3: https://diplomatic-reports.ehri-project.eu/ Online edition 4: https://nisko-transports.ehri-project.eu/
05	International Research Portal for Records Related to Nazi-Era Cultural Property	https://irp2.ehri-project.eu/
06	EHRI Bibliography	https://www.ehri-project.eu/biblio
07	EHRI Collection of Digital Tools	https://www.ehri-project.eu/ehri-collection-digital-tools-guides https://github.com/EHRI/ehri-data-analysis-tools
08	EHRI Fellowship Program	https://www.ehri-project.eu/Conny-Kristel-Fellowships
09	EHRI Courses in Holocaust studies	https://training.ehri-project.eu/ + linked websites
10	EHRI News and Events	https://www.ehri-project.eu/ehri-getting-involved + linked websites
11	EHRI Work plan and Deliverables	https://www.ehri-project.eu/division-work + linked websites
12	EHRI Facebook	https://www.facebook.com/EHRIproject
13	EHRI Twitter Channel	https://twitter.com/EHRIproject
14	EHRI Basecamp	https://basecamp.com/2786019/
15	EHRI Youtube Channel	https://www.youtube.com/@europeanholocaustresearchi424
16	EHRI Geospatial Repository	https://geodata.ehri-project.eu

17	EHRI Publications Repository	https://hal.science/EHRI
18	EHRI podcast	https://www.ehri-project.eu/ehri-podcast-for-the-living-and-the-dead (with links to Podcast apps: Buzzsprout, Spotify, Apple Podcasts, Google Podcasts
19	EHRI Moodle "Aligning Holocaust data with Open Research and FAIR data principles"	Moodle in internal review mode and not open accessible yet. Current Moodle server is for content development and review.

Table 1: Inventory of EHRI-3 Data Collections and current Web addresses

All EHRI-3 data collections will be maintained and updated via the web-address given, at least during the lifetime of the EHRI-3 project.

The data collections are aimed at the following target groups:

- Researchers: #01, #02, #03, #04, #05, #06, #07, #08, #09, #10, #15, #16, #17
- General public: #01, #10, #12, #13, #15, #16, #17, #19
- EHRI project partners: #11, #14

The data formats and data types of the data collections created and (re) used by the EHRI-3 project can not be classified in a clear and unambiguous way. The websites that identify the data collections obviously are formatted in HTML, generated by an underlying content management system that supports a range of data types. In the table below for each data collection a short description is given of the way the data collections are created and maintained. Also information on the data formats used and approximate storage size is given.

#	**Data Collection - Data Format and Data Type**
01	*EHRI Website*
	The EHRI-website is formatted in HTML. Software: Drupal Web Content Management Framework (Open Source) Databases: MySQL Size: <1GB
02	*EHRI Portal*
	Software: Bespoke Play Framework-based Scala application Databases: Neo4j, PostgreSQL Size: <10GB

03	*EHRI Document Blog*
	Software: Wordpress Omeka Open Source Web Publishing software Databases: MySQL Digital images: PNG, JPEG format Size: <10GB
04	*Online Editions*
	Omeka Open Source Web Publishing software with custom plugins Databases: MySQL Digital images: PNG format Size: <10GB
05	*IRP2: International Research Portal for Records Related to Nazi-Era Cultural Property*
	Software: Bespoke Flask-based Python application Databases: Static Size: <1GB
06	*EHRI Collection of Digital Tools*
	Software: Github software coding platform Size: <10GB
07	*EHRI Courses in Holocaust studies*
	Software: Drupal Database: MySQL Size: <1GB
08	*EHRI Geospatial Repository*
	Software: Geoserver, GeoNetwork Database: PostgreSQL Vector data: GeoPackage Raster data: GeoTIFF Size: <100GB
09	*EHRI Moodle*
	Software: Moodle (open source) Moodle server: Moodlecloud.com (to be replaced by EHRI-production service) Size: < 1GB excluding the instruction videos Videos (interviews. Raw data: 35 GB) Stored in Cloud (SURFdrive)
10	*EHRI podcast*
	Software: (various platforms via Buzzsprout cloud based hosting) (related context information on the podcast on the EHRI website) Format: standard sound format

Size: <100GB

Table 2: Data Formats and Data Types of EHRI-3 Data Collections

Issues that obstruct the durability of the EHRI-3 data collections are:

1. Link rot and content drift (web links that do not resolve anymore, web-information that is changed over time)

2. Obsolesce of software, e.g. to compile the HTML pages

3. Lack of resources to maintain and update information services

4. Changing (or absent) EHRI policies concerning the application of information services (e.g. concerning the use of Twitter / Facebook / YouTube) and role of emerging technologies and information services.

In cooperation with the EHRI-PP project an assessment should be carried out on the data collections that will be used after the EHRI-3 project ends. This assessment consists of the following steps:

1. Selection of data collections that will be maintained and supported in the long-term

2. Assessment of standards, services and policies that are available to manage the data collections

2. FAIR data

2.1 Making data findable, including provisions for metadata

- **Outline the discoverability of data (metadata provision)**
- **Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?**
- **Outline naming conventions used**
- **Outline the approach towards search keyword**
- **Outline the approach for clear versioning**
- **Specify standards for metadata creation (if any). If there are no standards in your discipline describe what metadata will be created and how**

This section provides an overview of "findability" aspects of EHRI-3 data collections, such as metadata standards and identifier mechanisms.

Table 3 outlines the metadata standards used by the EHRI-3 data collections.

	Metadata standard	Applied at
01	EAD (2002) - Encoded Archival Description	EHRI Portal - Archival Descriptions EHRI publishes archival descriptions in a hierarchical form broadly corresponding to the ISAD(G) conceptual standard. Top-level, or collection-level, descriptions are grouped by holding institution and

		can contain an arbitrary number of child items, which likewise may be nested to arbitrary depths.
		Support for importing and exporting the newer revision of the EAD schema — known as EAD-3 — has been implemented and is targeted for public availability in Q3 2023.
02	EAG (2012) - Encoded Archival Guide	**EHRI Portal - Archival Institutions** EHRI publishes information about archival or other institutions that hold some form of Holocaust-related material. These descriptions correspond to the ISDIAH (International Standard for Describing Institutions with Archival Holdings) conceptual standard, incorporating names, street addresses and contact information, along with predominantly textual accounts of their historic context and holdings.
03	EAC (2010) - Encoded Archival Context of metadata about corporate bodies, families and persons (CPF)	**EHRI Portal - Authority Files** Authority records consist of historical information concerning corporate bodies, persons and families that have some function in the creation of or relate to archival material, described in accordance with the International Standard Archival Authority Record (ISAAR) and downloadable as Encoded Archival Context (EAC) XML.
04	RDF/XML representations of controlled vocabularies in the Simple Knowledge Organisation System (SKOS) format.	**EHRI Portal - Controlled Vocabularies** EHRI currently manages three public controlled vocabularies: EHRI Terms: a hierarchically organised, multilingual set of subject headings for describing Holocaust-related material EHRI Camps: a list of camps based on the USHMM Encyclopedia of Camps and Ghettos EHRI Ghettos: a list of ghettos based on the Encyclopedia of Ghettos of Yad Vashem and the Encyclopedia of Camps and Ghettos of the USHMM
05	TEI - Text Encoding Initiative (encoded as XML)	**EHRI Online Editions** Four online editions are published at the time of writing with a further two in development. Each edition is based around a specific topic explored via the use of primary sources for which scanned images have been encoded as TEI XML. The TEI documents typically embed references to corporate bodies, persons or places for which data can be retrieved for additional context via the EHRI APIs or Geonames.
06	BibTex / RFT / Tagged / XML format for bibliographic records	**EHRI Bibliography** A selection of publications relating to the EHRI project and Holocaust research.
07	ISO 19115/19139 encoding for the representation of geospatial information	**EHRI Geospatial Repository** The EHRI Geospatial Repository contains datasets tailored for use in geographic or spatial applications. The underlying information can consist of either vector data, such as points or geometry conforming to a particular coordinate reference system (CRS), or

		raster data such as digitised maps. Dataset metadata is encoded using the ISO19139 XML implementation of the ISO19115 metadata model, and can be downloaded from the Geospatial Repository.
08	SCORM (Shareable Content Object Reference Model) standard enables interoperability of courses between e-learning systems	EHRI Moodle
09	Standard sound file formats (MP3 / WAV)	EHRI Podcast

Table 3. Metadata standards used in EHRI-3 Data Collections

Uniform resource locators (URL) and persistent identifiers.

As an aggregator and custodian of web-based information resources, EHRI ensures that it follows best-practices with regard to the consistency, persistence and accessibility of Uniform Resource Locators (URLs). Information is provided below about how resource URLs for the EHRI Portal are currently formed and the methods for ensuring that URL changes, when they are necessary, can be dealt with without disruption to 3rd parties.

URL accessibility versus suitability for long-term use.

In some cases, the accessibility of URLs is in conflict with their persistence to a certain degree. For example: to make URLs more meaningful and readable (rather than consisting of, for example, a random alphanumeric string) the primary URL for archival units in the EHRI Portal incorporates in a hierarchical form the identifier assigned to it by its holding institution. A disadvantage of this approach is that in rare cases holding institutions can change their local identifiers, and these changes propagate down to EHRI's URLs when collection metadata is synchronised. Likewise, there are cases where archival descriptions that have been manually added to the EHRI portal require changes that affect their local identifiers, although these changes are likewise quite rare.

In those cases where local identifiers *do* change and we can obtain information from the originating institution that allows us to perform a cross-reference, EHRI's policy is to generate new hierarchical identifiers for the items in question and add entries to a database table that maps old URLs to their new locations. This information is then used to dynamically generate "Moved Permanently" (HTTP 301) status codes with the appropriate redirect header when necessary. Web browsers and automated tools such as search crawlers are therefore able to handle items that have "moved" without encountering missing (HTTP 404) pages. In cases where records are removed from the EHRI portal for privacy or other reasons, rather than being moved to a new address, the portal returns a "Gone" (HTTP 410) status code, informing the user or crawler that the record is no longer available.

While web redirects are an effective way of avoiding "link rot" for many web usage scenarios they introduce a layer of indirection that can be undesirable due to the added technical complexity involved. A future version of the EHRI portal could remedy this by introducing an alternative system of more opaque identifiers, at the expense of URL readability and transparency.

Identifiers

The EHRI portal does not currently provide any guarantees that URLs are persistent in the manner of Digital Object Identifiers (DOIs) or other handle systems, nor does the infrastructure provide such a service for project partners.

For content such as blog posts, online editions or geospatial datasets EHRI intends to begin providing DOIs when a permanent, EHRI-specific prefix can be used. At the current time this is likely to be after EHRI-ERIC obtains status as a legal entity.

Uniform Resource Identifiers (URIs)

Distinct from resource *locators* discussed above (URLs), EHRI also makes use of Uniform Resource *Identifiers* (URIs) for a subset of data types, principally those that maintain compatibility with the Simple Knowledge Organisation System (SKOS) and are available as RDF. Along with the work discussed below (see *Linked Open Data*) URIs for the "data.ehri-project.eu" domain will resolve to the LOD repository or a service such as the EHRI Portal based on HTTP content negotiation. At present there are inconsistencies between the implementation URI scheme and that discussed in EHRI-2 deliverable D11.4 (section 4.3)[18] that we intend to resolve prior to completion of the project in 2024.

Outgoing links within portal content

Due to the large quantities of aggregated metadata present in the EHRI Portal there is the potential for links to external content within the body of archival descriptions to go stale due to external websites undergoing changes or being taken offline. Currently no automated processes exist to detect and correct such content issues but these may be developed in future.

2.2 Making data openly accessible

- **Specify which data will be made openly available? If some data is kept closed provide rationale for doing so**
- **Specify how the data will be made available**
- **Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?**
- **Specify where the data and associated metadata, documentation and code are deposited**
- **Specify how access will be provided in case there are any restrictions**

EHRI-3 supports the Open Science paradigm. Legal and ethical regulations can lead to keeping data closed. EHRI-3 follows the principle "as open as possible, as closed as necessary". Currently EHRI-3 data collections (see table 1) are not stored in certified data repositories.

The EHRI-3 project provides several structured data collections, or application programming interfaces (APIs), tailored to different purposes and use-cases.

Search API

The Search API allows basic search and retrieval tasks, similar to those facilitated by the EHRI portal's web interface, to be performed with structured data, specifically JSON over

---

[18] D11.4 Report Standards https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D11%204%20Report%20standards.pdf

HTTP. The request and response formats conform to those of the JSON API specification. Example use-cases include:

- searching for items containing a specific phrase, among all supported data types
- retrieving JSON data describing country reports, institutions, documentary units and authority file using their EHRI identifiers
- searching documentary unit items held by a specific institution
- searching institutions by proximity to a given geographical point

The Search API facilitates straightforward integration with third-party federated search tools.

GraphQL API

The GraphQL API is intended primarily for more advanced and comprehensive retrieval tasks, compared to the Search API. Described as a "query language for your API" GraphQL provides a machine-readable schema and mechanism for exploring deeply-nested or hierarchical data, over HTTP. Use cases include:

- retrieving a wider range of data than supported by the Search API, including user-generated data and information relating to archival provenance
- streaming larger amounts of data, without pagination
- fetching very specific data attributes for, e.g., user-interface development

OAI-PMH 2.0

The Open Archives Initiative Protocol for Metadata Harvesting 2.0 is an XML-over-HTTP interface that allows supporting tools to harvest metadata about the collections in the EHRI portal in either Dublin Core or EAD format. The portal's implementation is fully-conformant with the version 2.0 specification and supports all protocol features, including deletion markers, country- and institution-level sets, and granular time-based harvesting.

XML Export

The EHRI Portal also provides the means to download XML representations of data types not supported by the OAI-PMH protocol, directly from item web pages. These include:

- Encoded Archival Guide (EAG) 2012 encodings of metadata about institutions
- Encoded Archival Context (EAC) encodings of metadata about corporate bodies, families and persons (CPF)
- RDF/XML representations of controlled vocabularies in the Simple Knowledge Organisation System (SKOS) format

EHRI Portal - Sub collection	Search API	GraphQL API	OAI-PMH	LOD RDF
Country reports	✔	✔		✔
Institution metadata	✔	✔		✔
Archival descriptions	✔	✔	✔	✔

Controlled vocabularies		✔		✔
Authority files	✔	✔		✔
User profiles				
Annotations		✔		
Tertiary Metadata		✔		

Table 4. Data access facilities to EHRI-3 metadata

Linked Open Data

In 2023 EHRI-3 began the process of making portal metadata available as Linked Open Data (LOD). With archival data primarily encoded using the Records-in-Contents (RiC) ontology, along with additional properties from schema.org, this LOD Knowledge Graph (KG) facilitates integration with other LOD sources such as DBPedia and Wikidata, as well as authority data available from the EHRI-3 partner institution Centro di Documentazione Ebraica Contemporanea (CDEC). At the time of writing the LOD repository is in a testing phase due to the *beta* (0.2) status of RiC ontology and data conversion infrastructure. We anticipate that the repository will be available at https://lod.ehri-project.eu/ in early 2024.

Open Geospatial Consortium APIs

The EHRI Geospatial Repository makes available a number of APIs standardised by the Open Geospatial Consortium (OGC) for interacting with spatial data over the internet. These include:

- Web Map Service (WMS): for retrieving map images from a spatial data source as (typically) raster information.
- Web Feature Service (WFS): for retrieving vector-based spatial data in a variety of formats (e.g. GML (Geographic Markup Language) XML or GeoJSON.
- Catalogue Service for the Web (CSW): for retrieving metadata about geospatial records

2.3 Making data interoperable

- **Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.**
- **Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?**

Interoperability means allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible

compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins).

The metadata standards used, as well as the APIs (both described in the preceding sections), support the interoperability of the data collections of EHRI-3. The application of the SKOS standard to manage three public controlled vocabularies (EHRI Terms, EHRI Camps, EHRI Ghettos) facilitates the interoperability of these vocabularies with other information systems.

From 2024, the EHRI-3 Knowledge Graph will greatly increase the interoperability of metadata from the EHRI portal by its use of RDF, standardised ontologies (RiC, schema.org), and connections to other LOD sources. Using the SPARQL query language users can incorporate the EHRI-3 KG into federated queries of multiple, distributed data sources in a seamless manner.

2.4 Increase data re-use (through clarifying licences)

- **Specify how the data will be licenced to permit the widest reuse possible**
- **Specify when the data will be made available for reuse. If applicable, specify why and for what period a data embargo is needed**
- **Specify whether the data produced and/or used in the project is usable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why**
- **Describe data quality assurance processes**
- **Specify the length of time for which the data will remain re-usable**

Concerning licences related to the EHRI-3 data collections the EHRI Grant Agreement contains the following statement: "Intellectual Property Rights in all reports and other documents produced by the project will be retained by the authors and host institutions but will be made freely available on a non-exclusive licence as required by the European Commission. At the same time, where necessary we will respect the licence model of all third-party software used during the project and collaborate closely with the licence holders. Great care will be applied to the protection of knowledge and the accurate identification of its owners. These measures ensure that exploitation of project results will not be unreasonably impeded.

Any software created during the project will be made available to the community on an open source basis and be published on GitHub or another equivalent code management platform. Inside EHRI, we will establish an open culture of sharing research and development results and will follow a liberal application of IPR within the project. This will allow for a smooth and seamless integration of partners into distributed project deliverables. At the same time, EHRI partners have legitimate interests that their potentially large investment into the software and archives technologies that is integrated into the infrastructure is protected. The Grant Agreement, to establish the relationship with the European Commission, will be complemented by a Consortium Agreement, which will ensure the formal agreement and commitment of the partners on the rules for confidentiality, liability, publicity and on the IPR regime to be applied during the project lifetime and after it."[19]

3. Allocation of resources

---

[19] EHRI-3 Grant Agreement page 36.

Explain the allocation of resources, addressing the following issues:

- **Estimate the costs for making your data FAIR. Describe how you intend to cover these costs**
- **Clearly identify responsibilities for data management in your project**
- **Describe costs and potential value of long term preservation**

ThisEHRI-3 Data Management Plan describes the current state of art. The costs for managing the data collections as well as to improve the "FAIRness" of the data collection during the remainder of the project are covered by the EHRI-3 project budget. Data management is part of several work packages in EHRI, but has its focal point in WP12.

The separate EHRI-PP DMP[20] further elaborates responsibilities and practices for data capture, metadata production, data quality, storage and backup, data archiving and data sharing, while the future EHRI-IP project will continue that work.

4. Data security

Address data recovery as well as secure storage and transfer of sensitive data

The EHRI-3 project is not responsible for the storage of (digital) Holocaust archives that are managed by the participating content providers. EHRI provides access to the archival metadata. Personal data is protected according to EU regulations including the General Data Protection Regulation (GDPR).

The EHRI-3 project endeavours to follow reasonable security best-practices in the maintenance of information technology infrastructure and data management. While this is always a moving target that requires constant vigilance we undertake to:

- Enforce 2-factor authentication for infrastructure administration tasks;
- Reduce vulnerability to brute-force attacks by enforcing the use of public/private key authentication — as opposed to password authentication — where possible;
- Ensure any stored passwords are hashed with modern cryptographic algorithms such as BCrypt;
- Frequently update servers to ensure that they run the latest security patches.

EHRI-3 infrastructure is deployed and managed using Infrastructure-as-Code (IaC) tools with configuration centralised in a private Github repository. Secret tokens are stored in an encrypted database within this repository, and where necessary tokens and keys renewed at regular intervals. Extensive automation of infrastructure and software deployment is intended to ensure that recovery time following security incidents and/or accidental data loss is kept to a minimum.

5. Ethical aspects

To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former

EHRI-3 confidential deliverable D13.1 POPD - Requirement no. 1 [3] specifies procedures in regard to ethical data collection in project activities.[21] Work packages that gather survey data

---

[20] https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D7.3%20-%20Data%20management%20plan%20EHRI-RI.pdf

[21] D13.1 POPD – Requirement No. 1 [3] (Confidential)

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 years.

6. Other

Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

EHRI Content Provider Agreement

EHRI-3 requests that CHIs whose collection descriptions will be presented in the EHRI Portal sign the Content Provider Agreement (CPA) with NIOD-KNAW, EHRI-3's coordinating institution and legal representative. The CPA formalises the relationship between EHRI-3 and the CHI and clarifies various definitions with respect to applicable laws and regulations such as the GDPR, including that of personal data (PD) and special categories of personal data. CHIs that sign the CPA acknowledge EHRI-3's right to publish their metadata during and beyond the duration of the project and agree, for the project's duration, not to transfer to NIOD-EHRI data that contravenes the aforementioned laws and regulations or to hold NIOD-KNAW legally responsible for any matters arising from such violation.

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.

Horizon 2020 Open Access requirements

All peer-reviewed scientific publications arising from Horizon 2020 funding have to be made available in open access as soon as possible. There are two ways that the beneficiary can provide open access to a publication. The beneficiary can deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications, or the beneficiary can publish the research in an open access journal. In this last case the publication should also be submitted to a repository.

In case of publication in an open access journal, open access should be provided immediately, or within six months of publication in any other case. Only for publications in the social sciences and humanities a period of 12 months applies.