



Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

D12.1 EHRI-3 Data Management Plan

Bryant, Mike; Dijk, Elly; Dillmeier, Rebecca; van Horik, René; Levy, Michael; Pistol, Rachel

2021

document version

Publisher's PDF, also known as Version of record

document license

CC0

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

Bryant, M., Dijk, E., Dillmeier, R., van Horik, R., Levy, M., & Pistol, R. (2021). *D12.1 EHRI-3 Data Management Plan*. <https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D12.1%20EHRI-3%20Data%20Management%20Plan.pdf>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl



**European Holocaust Research Infrastructure
H2020-INFRAIA-2019-1**

GA no. 871111

D12.1

EHRI-3 Data Management Plan

**Mike Bryant
KCL / NIOD KNAW**

**Elly Dijk
DANS KNAW**

**Rebecca Dillmeier
USHMM**

**René van Horik
DANS KNAW**

**Michael Levy
USHMM**

**Rachel Pistol
KCL**

Start: September 2020 [M1]

Due: February 2021 [M6]

Actual: April 2021 [M8]



EHRI is funded by the European Union

Document Information

Project URL	https://www.ehri-project.eu
Document URL	[www.....]
Deliverable	D12.1 EHRI-3 Data Management Plan
Work Package	WP12
Lead Beneficiary	14 - USHMM
Relevant Milestones	MS1
Dissemination level	Public
Contact Person	Michael Levy, mlevy@ushmm.org
Abstract (for dissemination)	<p>This report contains the first, initial version of EHRI-3's deliverable 12.1 - its Data Management Plan (DMP). The first part of this report lays out the design and framework for compiling the final DMP due in August 2024. Between now and the delivery of the final DMP, WP12 will continue to define its goals for data management, its preservation worthy assets, procedures for EHRI participants to follow and the dissemination of those procedures, and the implementation of a plan with an understanding of the associated risks and costs. There is a recognition that a successful DMP will require both human managed systems and technology. The goal of the DMP is to bring EHRI data into closer compliance with FAIR principles. The initial DMP outlines data sources as they are currently understood, including descriptions of their locations, data formats, size and "findability". Much of the work between this initial DMP and its final version will be to improve metadata standards and identifier mechanisms for data collections, which are necessary for FAIRness compliance. This will require consultation with EHRI partners and the dissemination of standards and practices established in the DMP.</p>
Management Summary	<p>(required if the deliverable exceeds more than 25 pages)</p> <p>[Max. 500 words]</p>

Table of Contents

1. Introduction	5
2. Open Research Data Pilot	5
3. Data Management Planning	6
FAIR data principles	6
4. Compilation of the EHRI-3 Initial Data Management Plan	7
5. Next steps	8
1. Data summary	10
Data Collection - Data Format and Data Type	12
2. FAIR data	14
2.1 Making data findable, including provisions for metadata	14
Uniform resource locators (URL) and persistent identifiers.	16
URL accessibility verses suitability for long-term use.	16
Identifiers	16
Uniform Resource Identifiers (URIs)	16
Outgoing links within portal content	17
2.2 Making data openly accessible	17
Search API	17
GraphQL API	17
OAI-PMH 2.0	18
XML Export	18
2.3 Making data interoperable	19
2.4 Increase data re-use (through clarifying licenses)	19
3. Allocation of resources	20
4. Data security	20
5. Ethical aspects	21
6. Other	21
EHRI Content Provider Agreement	21
General Data Protection Regulation (GDPR)	21
Horizon 2020 Open Access requirements	22

Glossary

API: Application Programming Interface

CPA: Content Provider Agreement

CHI: Collection Holding Institution

DMP: Data Management Plan

DOI: Digital Object Identifier

EAD: Encoded Archival Description

FAIR: Findability - Accessibility - Interoperability - Reusability

GDPR: General Data Protection Regulation

IAC: Infrastructure as Code

IRI: Internationalised Resource Identifier

IPR: Intellectual Property Rights

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

LOD: Linked Open Data

MPT: Metadata Publishing Tool

OAI-PMH: Open Archives Initiative Protocol for Metadata Harvesting

ORCID: Open Researcher and Contributor Identifier

ORDP: Open Research Data Pilot

PD: Personal Data

PID: Persistent Identifier

RDBMS: Relational Database Management System

RDF: Resource Description Framework

SCM: Source Code Management

SLA: Service Level Agreement

SKOS: Simple Knowledge Organisation System

TEI: Text Encoding Initiative

URI: Uniform Resource Identifier

URL: Uniform Resource Locator

XML: eXtensible Markup Language

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

1. Introduction

The EHRI-3 project will run 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 organization organized 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 EHRI Conny Kristel Fellowship Programme.

The EHRI-3 project has a close relationship with the EHRI Preparatory Phase Project (EHRI-PP) that will run from 2019 to 2022. The aim of the EHRI-PP project is to transform EHRI from a fixed-duration project into a permanent organisation 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 will have a close relation with those of EHRI-3.

A main difference between the two projects will be 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 data management policies will be more prescriptive concerning the management of data in the future EHRI-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 Initial Data Management Plan. The EHRI-3 Updated Data Management Plan (to be published as D12.2) is scheduled for month 36 of the EHRI-3 project. The compilation of the updated DMP consists of consultations with EHRI partners and service providers.

2. Open Research Data Pilot

This deliverable is classified as an Open Research Data Pilot (ORDP). In the Horizon2020 Framework program, the European Commission has launched a pilot for open access to research data. 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

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

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 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 pilot: 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 Open Research Data Pilot 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 will be based on the Horizon 2020 DMP template as published by the European Commission.⁵ This deliverable contains a first, initial version of the EHRI DMP that will be further developed during the project. An updated, final, EHRI-3 Data Management Plan will be published as Deliverable 12.2. It is foreseen that implementation of some aspects of the DMP requires advocacy, training and support for EHRI-3 partners. For this, consultation and training will be organised in cooperation with WP5.

FAIR data 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

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

each of the 4 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.



Figure 1. FAIR data principles. Source: <https://ccafs.cqiar.org/open-access-and-fair-principles>

4. Compilation of the EHRI-3 Initial Data Management Plan

The EHRI-3 Initial Data Management Plan is based on the Horizon 2020 DMP Template.⁷ The DMPonline Tool⁸ is used to create the first version of the EHRI-3 Initial DMP. “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

⁶ <https://www.go-fair.org/fair-principles/>

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>

As stated in the description above, the DMPonline tool supports researchers. 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” (including politicians), 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 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

Project Title	Template
EHRI-3 - European Holocaust Research Infrastructure	Horizon 2020 DMP

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

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

5. Next steps

The initial DMP (see appendix) contains a first inventory of the types of assets that are potentially preservation worthy and where they currently reside. With a better understanding of the size and scope of EHRI’s assets, as well as an understanding of the scale of growth, an updated DMP can be further developed and put in place. This is planned for month 36 in the EHRI-3 project when D12.2 EHRI-3 Updated Data Management Plan will be published.

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.

¹⁰ <https://www.ehri-project.eu/about-ehri>

Based on the “EHRI-3 initial DMP” 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.

Potential focus areas are:

- Define metadata policies for all data collections in EHRI-3
- Implementation of PID services in EHRI
 - (e.g. ORCID for people, DOI for publication, Handle for data)
- Increase durability of data repositories by applying for the CoreTrustSeal certification
- Stimulate usage of certified repositories for data collections
- Assess (if relevant) implement usage of repository services (e.g. Zenodo for publications, Dataverse for data)
- Increase durability and usability of online information (e.g. using “robustify link” services (including storage of information in Internet Archive) for EHRI Blogs and other online resources)
- Creation of policies to manage (including deletion) of project information (e.g. Basecamp information)
- Define EHRI policy for the usage of social media applications (Facebook, Twitter, YouTube)

The progress made on improving the FAIRness of EHRI data collections will be reported in the “EHRI-3 updated Data Management Plan” (D12.2). This deliverable will be created in close collaboration with the EHRI-PP project that aims for the establishment of a permanent EHRI research infrastructure. This moving from “project” to “program” requires a long-term vision on the management of data collections.

EHRI-3 - European Holocaust Research Infrastructure - Initial DMP

A Data Management Plan created using dmproadmap

Creators: Mike Bryant, Elly Dijk, Rebecca Dillmeier, René van Horik, Michael Levy

Affiliation: EHRI-3 project

Funder: European Commission (Horizon 2020)

Template: Horizon 2020 DMP

Project abstract

The mission of the European Holocaust Research Infrastructure (EHRI) concerns transnational 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 EHRI 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 EHRI Conny Kristel Fellowship Programme.

Last modified: 31-03-2021

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**

This is the Initial Data Management Plan of the EHRI-3 project (D12.1). It will be updated during the project. The Updated Data Management Plan (D12.2) will be published in August 2024.

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 EHRI 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/vocabulareis
	Authority Sets in EHRI Portal	https://portal.ehri-project.eu/sets
03	EHRI Document Blog	<p>https://blog.ehri-project.eu/</p> <p>Digital images and visualisations related to the blog are accessible via web address: https://visualisations.ehri-project.eu/</p>
04	Online Editions	<p>https://www.ehri-project.eu/new-ehri-online-editions</p> <p>Online edition 1: https://begrenzte-flucht.ehri-project.eu/</p> <p>Online edition 2: https://early-testimony.ehri-project.eu/</p>
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	<p>https://www.ehri-project.eu/ehri-collection-digital-tools-guides</p> <p>https://github.com/EHRI/ehri-data-analysis-tools</p>
08	EHRI Conny Kristel Fellowship Programme	<p>https://www.ehri-project.eu/ehri-fellows-2016</p> <p>https://www.ehri-project.eu/ehri-fellowship-call-2016-2018</p>
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 Workplan 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/channel/UCI-Y23CupqRs_Rm4ezo0pRw

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

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

The data collections are aimed at the following target groups:

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

The data formats and data types of the data collections created and (re) used by the EHRI-3 project cannot 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	<i>EHRI Website</i>
	<p>The EHRI-website is formatted in HTML.</p> <p>Software: Drupal Web Content Management Framework (Open Source)</p> <p>Databases: MySQL</p> <p>Size: <1GB</p>
02	<i>EHRI Portal</i>
	<p>Software: Bespoke Play Framework-based Scala application</p> <p>Databases: Neo4j, PostgreSQL</p>

	Size: <10GB
03	<i>EHRI Document Blog</i>
	<p>Software: Wordpress</p> <p>Omeka Open Source Web Publishing software</p> <p>Databases: MySQL</p> <p>Digital images: PNG, JPEG format</p> <p>Size: <10GB</p>
04	<i>Online Editions</i>
	<p>Omeka Open Source Web Publishing software with custom plugins</p> <p>Databases: MySQL</p> <p>Digital images: PNG format</p> <p>Size: <10GB</p>
05	<i>IRP2: International Research Portal for Records Related to Nazi-Era Cultural Property</i>
	<p>Software: Bespoke Flask-based Python application</p> <p>Databases: Static</p> <p>Size: <1GB</p>
06	<i>EHRI Collection of Digital Tools</i>
	<p>Software: Github software coding platform</p> <p>Size: <10GB</p>
07	<i>EHRI Courses in Holocaust studies</i>
	<p>Software: Drupal</p> <p>Database: MySQL</p> <p>Size: <1GB</p>

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

Issues that obstruct the durability of 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 will be carried out on the data collections that will be used after the EHRI-3 project will end. 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
3. Compilation of the Updated Data Management Plan (D12.2) that contains the plan to manage 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 provides 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

		<p>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.</p> <p>Support for importing and exporting the newer revision of the EAD schema — known as EAD-3 — has been partially implemented but is not publicly available as of Q1 2021.</p>
02	EAG (2012) - Encoded Archival Guide	<p>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.</p>
03	EAC (2010) - Encoded Archival Context of metadata about corporate bodies, families and persons (CPF)	<p>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.</p>
04	RDF/XML representations of controlled vocabularies in the Simple Knowledge Organisation System (SKOS) format.	<p>EHRI Portal - Controlled Vocabularies EHRI currently manages three public controlled vocabularies:</p> <p><u>EHRI Terms</u>: a hierarchically organised, multilingual set of subject headings for describing Holocaust-related material</p> <p><u>EHRI Camps</u>: a list of camps based on the USHMM Encyclopedia of Camps and Ghettos</p> <p><u>EHRI Ghettos</u>: a list of ghettos based on the Encyclopedia of Ghettos of Yad Vashem and the Encyclopedia of Camps and Ghettos of the USHMM</p>
05	TEI - Text Encoding Initiative (encoded as XML)	<p>EHRI Online Editions Two online editions are published at the time of writing with a further three 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.</p>
06	BibTex / RFT / Tagged / XML format for bibliographic records	<p>EHRI Bibliography A selection of publications relating to the EHRI project and Holocaust research.</p>

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 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.

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. While such services may be provided in future, they require a level of institutional sustainability and a guarantee of long-term funding that has hitherto been lacking in EHRI's time-limited phases to date.

A future version of this plan will discuss the provision and use of DOIs for second-order intellectual outputs such as publications and project deliverables.

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. For legacy reasons, URIs currently lack consistency with the URL scheme discussed above and are not resolvable (accessible over HTTP.) Work to address these deficiencies is

planned to be undertaken before the end of EHRI-3, using the EHRI-2 deliverable 11.4 (section 4.3) as a basis.¹¹

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 external websites undergoing changes or being taken offline. Currently no automated processes exist to detect and correct such content issues but 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 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 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 than the Search API, included user-generated data and information relating to archival provenance

¹¹ D11.4 Report Standards <https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D11%20%20Report%20standards.pdf>

- 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	Other
Country reports	✓	✓		
Institution metadata	✓	✓		
Archival descriptions	✓	✓	✓	
Controlled vocabularies		✓		TTL RDF
Authority files	✓	✓		
User profiles				
Annotations		✓		
Tertiary Metadata		✓		

Table 4. Data access facilities to EHRI-3 data collections

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.

An assessment will be carried out to improve the level of interoperability of the data collections. The results will be published in the Updated Data Management Plan (D12.2).

2.4 Increase data re-use (through clarifying licenses)

- **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 license 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.”¹²

3. Allocation of resources

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**

This report contains the EHRI-3 initial Data Management Plan and 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.

During the remainder of the project the roles and responsibilities for data management activities will be outlined, e.g., for data capture, metadata production, data quality, storage and backup, data archiving and data sharing. This will be done in cooperation with the EHRI-PP project.

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.

¹² EHRI-3 Grant Agreement page 36.

Over time we are undertaking to introduce more automation into the provisioning and configuration of IT resources which, among other benefits, will allow components to be rapidly and reliably redeployed in the event of a security breach or other compromise situation.

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.¹³ 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 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.

¹³ D13.1 POPD – Requirement No. 1 [3] (Confidential)

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.