European Holocaust Research Infrastructure
H2020-INFRAIA-2014-2015
GA no. 654164

Deliverable 9.4

Resource reports

Sigal Arie-Erez & Judith Levin
Yad Vashem

Giles Bennett & Anna Ullrich
Institut für Zeitgeschichte

Charlotte Hauwaert & Francesco Gelati
ARA/Cegesoma

Veerle Vanden Daelen
Kazerne Dossin (WPL)

Start: May 2015 [M1]
Due: April 2018 [M36]
Actual: April 2018 [M36]

Document Information

Project URL	www.ehri-project.eu
Document URL	[www……]
Deliverable	D9.4 Resource reports
Work Package	WP9
Lead Beneficiary	P2 ARA/Cegesoma (May 2015 - June 2016); P23 Kazerne Dossin (July 2016 – present)
Relevant Milestones	MS3
Dissemination level	Public
Contact Person	Veerle Vanden Daelen, veerle.vandendaelen@kazernedossin.eu, +32 15 28 86 23
Abstract (for dissemination)	The Resource Reports present background information on how resources were identified, selected and integrated into the EHRI infrastructure. They provide a concise overview about the scope of the EHRI data content, its origins as well as details about the methodologies employed to identify, select and integrate data.
Management Summary	n/a

Table of Contents

Glossary

CHI	Collection Holding Institution
CSV	Comma-Separated Value
EAD	Encoded Archival Description: an XML-based standard for describing archival materials
EAD2002	the 2nd edition of EAD, released in 2002
ECT tool	EAD Conversion Tool, developed by WP10, described in D10.1
MPT tool	Metadata Publishing Tool, developed by WP10, described in D10.2
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OAI-RS	Open Archives Initiative ResourceSync Framework Specification
XML	Extensible Mark-up Language

Collection Holding Institutions acronym references

BE-CEGESOMA	Cegesoma in Belgium
BE-KD	Kazerne Dossin in Belgium
CZ-JMP	Jewish Museum of Prague in Czech Republic
DE-BARCH	Bundesarchiv in Germany
DE-IfZ	Institut für Zeitgeschichte in Germany
DE-ITS	International Tracing Service in Germany
FR-MS	Mémorial de Shoah in France
GR-JMG	Jewish Museum of Greece
HU-HJA	Hungarian Jewish Archives
IL-YV	Yad Vashem in Israel
IT-CDEC	Foundation Jewish Contemporary Documentation Centre in Italy
IT-ICAR	Central Institute for Archives in Italy
LT-VGSJM	Vilna Gaon State Jewish Museum in Lithuania
NL-NIOD	Institute for War, Holocaust and Genocide Studies in the Netherlands
PL-ZIH	Jewish Historical Institute in Poland
RO-INSHR-EW	Elie Wiesel National Institute for Studying the Holocaust in Romania
SK-HDC	Holocaust Documentation Centre in Slovakia
UK-Wiener	Wiener Library in the UK
US-USHMM	United States Holocaust Memorial Museum in the USA
US-YIVO	Yiddish Scientific Institute in the USA

Introduction

The Resource Reports present background information on how resources were identified, selected and integrated into the EHRI infrastructure. They provide a concise overview about the scope of the EHRI data content, its origins as well as details about the methodologies employed to identify, select and integrate data. As such, the resource reports are directly related to Task "T9.4. Methodological analysis", which analyses and reflects upon the methods employed in the identification, investigation and selection work undertaken in tasks T9.1 and T9.2.

The reports are aimed at both research users of the EHRI infrastructure and archives that wish to integrate their holdings into EHRI. The former will be provided with concise information about the methodologies, which have been followed in developing the infrastructure's data content (part 1 of the Deliverable). This will enable researchers to make informed and controlled use of the EHRI infrastructure, as it enables them to understand the 'story behind the data'. For archives, the resource reports will provide valuable case studies about possible data integration pathways and models (part 2 of the Deliverable).

1   "The Story behind the Data" – Explaining to the users how the data are integrated into the EHRI portal

Information on the data content of the EHRI portal, or "the story behind the data", is available via the "Help" button under "Frequently Asked Questions" on the EHRI Portal (). The FAQs address questions concerning the structure, content (including content-selection and representation), and architecture of the portal, as well as the feedback options for users of the portal. Once the PMB approves the DL, the questions and answers below will replace the currently available text.

What type of information is available via the EHRI portal?

> The EHRI portal provides an overview of archival institutions that hold Holocaust-relevant sources as well as descriptions of Holocaust-relevant materials. It does not provide the archival material itself, but descriptions about it (metadata). To frame this information, the EHRI portal contains country reports providing concise per-country information on the respective country's Holocaust history, archival situation and the research EHRI has undertaken. A general introduction to the country reports can be found (here)[http://www.ehri-project.eu/country-reports], or via clicking on the link in the Countries section of the portal on the top right.

How does EHRI identify archival institutions that hold Holocaust-related materials?

> Holocaust-relevant archival institutions included in the portal have been identified by using the following major sources:
> - the Directory of Holocaust-Related Archives (Conference on Jewish Material Claims Against Germany)
> - the Guide des archives sur la Shoah (Mémorial de la Shoah)
> - an overview list of institutions Yad Vashem worked together with and which it has copied archival material from
> - an overview of restitution and compensation archives from the Conference on Jewish Material Claims Against Germany
>
> For some countries, additional information was gathered from:
> - the online catalogue of the United States Holocaust Memorial Museum

- specific additions from national archival guides, experts, and published studies on the subject

While the main tracing services relevant to the Holocaust have been covered by EHRI, further investigation may yield more insights.
EHRI continues updating the archival institutions.

How does EHRI select collections for the EHRI portal?

EHRI wishes to keep the portal as open-ended as possible; therefore, all collection descriptions on Holocaust-relevant materials are welcome in the portal. When authoring descriptions of archival materials itself, EHRI focuses on the period from 30 January 1933 to today. However, in case of so-called victim sources (sources from Jews or people considered as Jews under Nazi rule, or victim organizations) sources from the interwar period or earlier can be included. Our first geographical focus is on Germany and Eastern Europe, the main crime site of the Holocaust, our second one on other occupied countries and Axis countries, and our final one on other countries such as refugee countries. EHRI uses a top-down approach: from national, to regional, to local, to individual. A collection is considered a Holocaust-related collection if it contains one file on the Holocaust.

Who writes the Holocaust-related archival descriptions for the EHRI portal?

EHRI brings together both already existing descriptions and writes new descriptions. The sources can be listed as follows:
1. Descriptions written by EHRI
2. Descriptions written by the archival institution itself
3. Descriptions written by a third party, other than EHRI or the collection-holding institution itself (such as finding aids on Jewish sources or research guides on sources on the Second World War)

How can I distinguish between the 3 types of sources (EHRI-written, written by the archival institution or written by a third party)?

The fields "Process Info", "Archivist Note" and "Rules and Conventions" as well as the "online history" information box (available on the right-hand side of the screen) can provide additional information about the authoring, selection and integration of descriptions into the EHRI portal.

Is it possible that a collection in the portal has more than one description?

Yes, there are cases where a collection has more than one description. A collection may be described by various sources, and in various languages. All these descriptions can be helpful for a researcher as they have varied perspectives on the material being described and are often written in different languages. EHRI wishes to provide all these sources of information. Therefore, individual descriptions that relate to the same material are represented as parallel descriptions.

How can the various descriptions of one collection be accessed?

Information about the existence of parallel descriptions, and a mechanism to switch between them, is provided in the box labelled "Archival Descriptions" that is available on the right-hand side of the screen when viewing a description of archival materials.

What happens when a number of archival institutions have the same collection?

In cases where one institution holds an original collection and other institutions hold copies of the same collection, each institution describes the collection in its possession and our goal is to create connections between the descriptions. Links to other items are displayed at the end of the description, under "Other Connected Items". Please note that

this is work-in-progress: not all original-copy connections in the portal have been entered in this way and plans to do so are already underway.

I can find many descriptions of archival materials for some countries but only few for others. Why is this and is this still going to change?

The identification of sources is an ongoing effort. Differences in details of information should become less significant as more descriptions are added to the EHRI portal.

Which languages are used in the EHRI portal?

All country reports and descriptions of archival institutions are written in English. Descriptions of archival materials that are authored by EHRI are written in English as well. However, when existing descriptions of archival materials were available in languages other than English, they are integrated in their original language. On the difference between EHRI-authored descriptions and descriptions authored by others, see (here)[LINK to question "How can I distinguish between the 3 types of sources (EHRI-written, written by the archival institution or written by a third party)?"].

Does EHRI use a thesaurus/authority files?

EHRI developed a thesaurus which is available in ten languages as well as authority files for corporate bodies and persons. Descriptions of archival materials authored by EHRI are consistently indexed with subject terms and authorities. For the already existing collection descriptions which are integrated in the EHRI portal, the mapping of keywords with the EHRI thesaurus and authority files is in a test phase.

Which archival standards are used by EHRI to structure the data in the EHRI portal?

EHRI works with Guidelines for Descriptions that are in accordance with the standards of the International Council on Archives (ICA): the International Standard for Describing Institutions with Archival Holdings (ISDIAH, for archival institutions); the General International Standard Archival Description (ISAD(G), for archival descriptions); and the International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR(CPF), for authority records for corporate bodies, persons and families). For more information on these standards see (here)[http://www.ica.org/10206/standards/standards-list.html].

What are the "research guides"?

The aim of the EHRI research guides is to create comprehensive, innovative and easy to use tools that bring together dispersed and fragmented archival material from different archives and empower further research. In the guides, we test new approaches that combine detailed descriptions from the EHRI portal with narrative information and new ways to visualise the data. The research guides can be consulted (here)[https://ehri-project.eu/visit-ehri-research-guides].

Is the data presented in the research guides also available in the EHRI portal and vice versa?

The archival descriptions collected for the research guides are stored and available through the EHRI portal, as well as through the dedicated user interface. The guides typically contain detailed archival descriptions for selected archival material, going down to file or even document level in some cases.

I have found factually incorrect or incomplete information in the EHRI portal. How can I alert you about this?

> We do our utmost to provide high quality information in the portal. We therefore welcome your feedback if you detect any inaccuracies or omissions in any of the information we provide about countries, archival institutions and archival materials.

> You can alert us about such problems in two ways.
> - Public note: if you are a registered user of the portal, you can add a public note to the relevant description that explains the nature of the problem you have detected.
> - Via email: send us an email with relevant details to feedback@ehri-project.eu.

> Depending on the nature of your feedback and on whether the problematic information was written by EHRI or by a partner institution, we may either correct and/or enhance the description directly, add a note to the description, and/or alert the partner institution that has authored the description.

Does EHRI offer access to the archival materials that are described in the EHRI portal?

> No, the EHRI portal contains descriptions of Holocaust-related archival materials. It does not offer access to these materials (or to digital surrogates thereof). In general, if you want to access archival materials described on the EHRI portal, you need to check the access options and policies of the relevant holding institutions. When provided, the entry includes a link to the original cataloguing system website, where archival materials may be consulted.

2 Methodology for Data Identification and Selection

As written in EHRI-I deliverable 15.6, "the field of Holocaust studies relies on a huge variety of archives. One of EHRI's most important tasks was to create an inventory of Holocaust-related archival institutions and collections in order to share information about them and help provide access to the research community"[1]. Both phases of the EHRI project have gathered information on collection-holding institutions and their Holocaust-related collections in an ICA-standard compliant EHRI-customized database. Next to this the identification, selection and description of Holocaust-relevant collections have been the main focus of EHRI's data integration.

For its Data Identification work EHRI's WP9 has experts within the Work Package who work on data integration and hires local experts for those areas beyond the expertise gathered within WP9. Furthermore EHRI worked on input via community building and reaching out to experts via data integration workshops. Such workshops covered the following geographical areas: the former Soviet Union, Italy, Slovakia, Bulgaria and the war-time borderlands of Hungary and Romania[2]. For EHRI, the workshops resulted in a more precise outline on how the various archives in different countries are organised, what type of archival holdings they preserve, how they structure and describe Holocaust-relevant archives and what level of digitalisation and/or technical IT-solutions they possess to share descriptions of their holdings. Moreover, all the WP9 workshops promoted the EHRI project and Portal and were an opportunity to convey EHRI's missions to new (meta-)data providers and to encourage their readiness to cooperate with EHRI.

---

1 EHRI-1 deliverable 15.6, p. 8.
2 For more information please refer to EHRI D9.2.

3 Possible data integration pathways and models in EHRI-2

Since EHRI's second phase under Horizon2020 has started, there have been four major workshops to discuss methodologies on data integration. The first workshop, "Methodology and Workflow Workshop for WP9-10-11", took place in Athens on October 2015. One of the results of the first workshop was an outline of the possible ways for data integration in EHRI and the challenges ahead. The second methodological workshop was the workshop on linking at Yad Vashem on 15-17 March 2016. The workshop focused on linking collections, the identification and expression of copy-original relationships regarding collections held by CHIs in the portal and their practical implementation. The aim of the workshop was to see if and how linking could be implemented and its challenges as well as all the possibilities linking offers in the EHRI portal and the challenges this can represent. The third workshop was held at Kazerne Dossin in Mechelen on 26-27 October 2016. The workshop was aimed at discussing the different scenarios of data integration (manual, fully IT-based and semi-IT-based) and the technical aspects regarding data integration by the tools being developed by WP10. The latest WP9-10 workshop was held in Milan on 6 and 8 March 2017 where the WP10 tools were shown in their entirety via a simulation. These last two workshops delivered an introduction to the tools and made planning possible for testing the tools by Summer 2017, so that the tools were ready for use by CHIs outside the EHRI consortium in the autumn of 2017. Throughout the project WP9 and WP10 have invited members of each other's WPs to each others meetings where required and a communication meeting between the two WPs takes place every month. Additional meetings are scheduled as required.

In order to have a clear overview of the workflows for EHRI-2, WPs9-10-11 have developed use cases in two different formats, one from the perspective of the CHIs and a second from EHRI's perspective. The use cases from a CHI-perspective include 3 types of CHIs: "Hidden CHIs", "Digital Open CHIs" and "Advanced Digital CHIs" (for definitions, see below). Based on functionality, the following use cases have been developed:

- selection of relevant archival material
- IT-based import
- manual input
- update of existing descriptions

High-level categorization of CHIs that EHRI is likely to encounter: "Hidden CHIs", "Digital opened CHIs" and "Digital advanced CHIs".

- *Hidden CHI*: This type of CHI only has paper collections that cannot be found on the internet. The CHI may have finding aids, but these are also only in paper form. The CHI is not used to work with projects such as EHRI.
- *Digitally Open CHI*: This type of CHI has digital metadata and possibly digital content. Finding aids may be in paper and/or digital format. There is information available on the internet. The CHI is able to export from its content managements systems, sometimes in standardized form such as EAD. The CHI is not very well experienced in cooperating with projects such as EHRI.
- *Advanced Digital CHI*: This type of CHI has all its metadata available in a digital format and collection/archives can be found on the internet. The CHI is able to export in different standardized formats and has publication tools such as e.g. OAI-PMH. It is used to participate in project with third parties, such as EHRI.
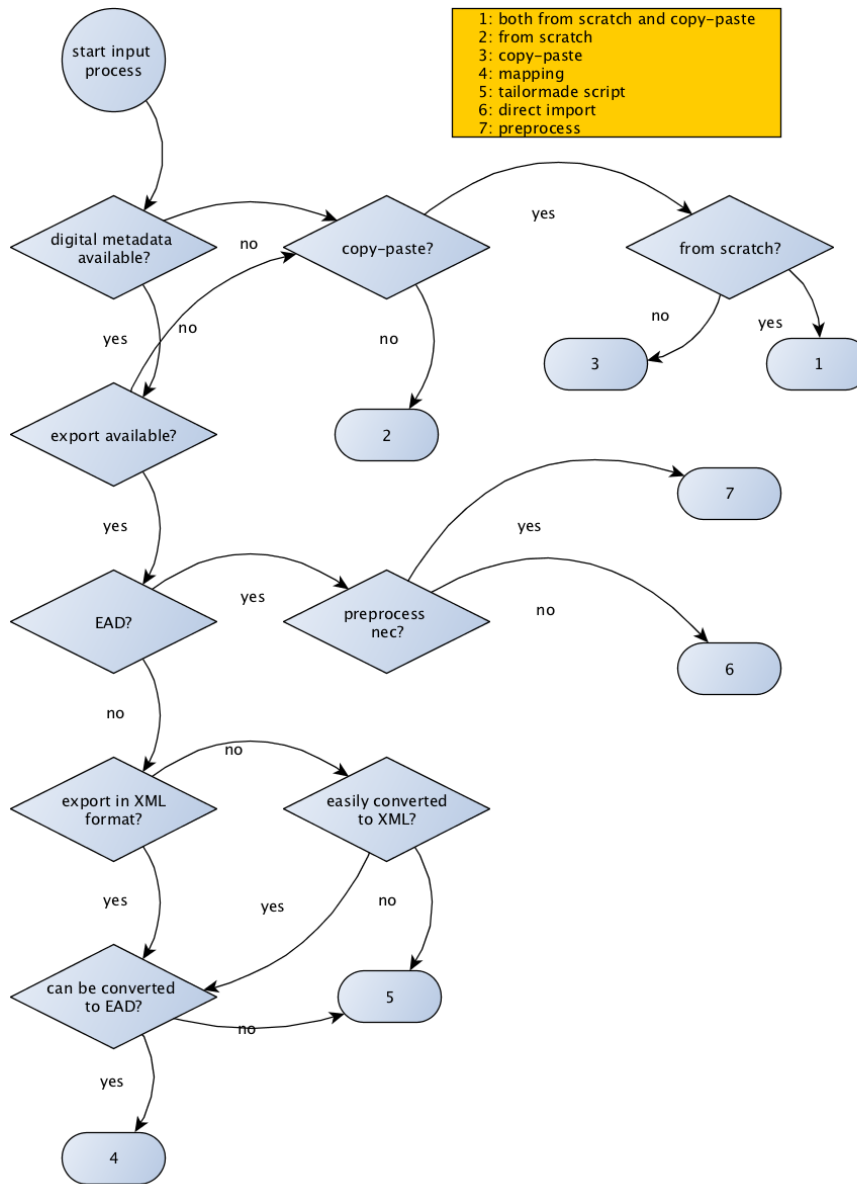
3.1 EHRI-1 scenarios for data integration

In EHRI-1, the following scenarios for data integration have been applied, and they map as follows to the 3 types of CHIs from the use cases:

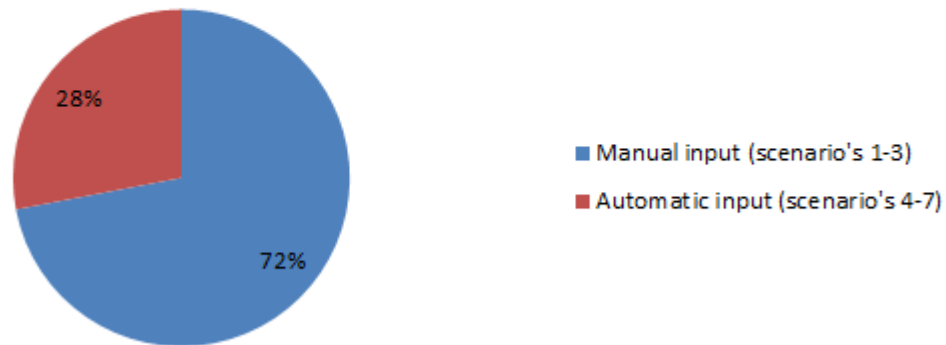| EHRI-1 scenario's for data integration	Types of CHIs
Manual input	
1. Both from scratch and copy-paste	Hidden CHIs (+ input from researchers and digital opened and advanced CHIs which were for various reasons not cooperating with EHRI)
2. From scratch	Hidden CHIs (+ input from researchers and digital opened and advanced CHIs which were for various reasons not cooperating with EHRI)
3. Copy-paste	Hidden CHIs (+ input from researchers and digital opened and advanced CHIs which were for various reasons not cooperating with EHRI)
IT-based input	
4. Mapping	Digital Open CHIs + Advanced Digital CHIs
5. Tailor-made script	Digital Open CHIs + Advanced Digital CHIs
6. Direct import	Advanced Digital CHIs
7. Pre process	Digital Open CHIs + Advanced Digital CHIs

While the first 3 scenarios are manual work only, it is very important to note that method 1-3 did not only include hidden CHIs but equally input from researchers and digital opened and advanced CHIs which were for various reasons not yet cooperating with EHRI:

- either because EHRI did not have the manpower in WP19 to do the import, cf. Archives of the American Jewish Joint Distribution Committee (JDC archives), for example;
- or because the CHIs were not able to work with EHRI's WP19, cf. Polish national archives, for example;
- or, because selecting the EHRI-relevant collections was more time-consuming by IT-driven means so that it made better sense to carry out the whole input manually.
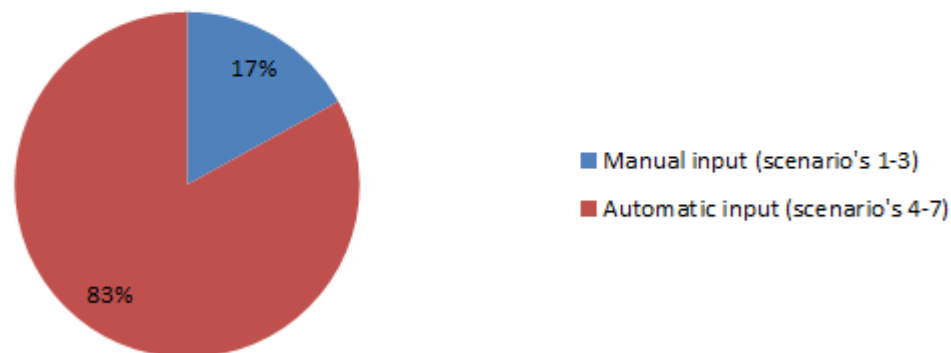
Graph 1 : Overview of data integration in EHRI-1 (Graph by Linda Reijnhoudt with input Veerle Vanden Daelen).

CHIs encountered by EHRI-1 for scenario's manual & automatic input

28%

72%

■ Manual input (scenario's 1-3)
■ Automatic input (scenario's 4-7)

Top level descriptions in the portal via scenario's manual & automatic input

17%

83%

■ Manual input (scenario's 1-3)
■ Automatic input (scenario's 4-7)

Graphs 2 and 3: Data integration in EHRI-1

Without USHMM, whose top-level descriptions included many items (shoes, coins, etc. described individually as top-levels), the proportion scenario 1-3 versus 4-7 becomes 35% versus 65%. A detailed overview of data integration in EHRI-1 is available on request.

From this analysis we can conclude that:
- EHRI-1 was very successful at bringing to the fore hidden CHIs: manual input was responsible for the collection descriptions of 72% of all the institutions listed in the portal for which collection descriptions are available. Thus manual input has put a high number of institutions on the radar of Holocaust researchers.

- Potential capacity of methods 4-7 (i.e. IT-based methods) leads to more descriptions in the portal: with 28% of the institutions with collection descriptions bringing in 83% of the collection descriptions, it is clear that these methods can bring in a much higher number of descriptions in a much shorter time-span than manual data integration

The following Deliverables from EHRI-1 form a firm basis to continue the developmental work, the data integration and standardization of the EHRI portal in EHRI's second phase:
– D17.2 (data model)
– The update of D17.3 via annex 3 in D15.6 (guidelines)
– D19.4 (workshop)
– D19.5
– Information on EHRI data integration from the EHRI country reports ().
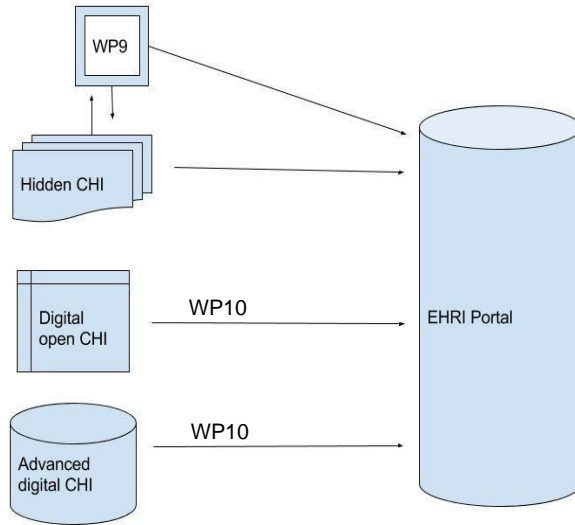
3.2 Workflow in EHRI-2

Metadata in EHRI-2 (since May 2015) is delivered by researchers (within the EHRI-project and local experts), collection-holding institutions (CHIs) and aggregators. Making the division for EHRI-2 by the three types of CHIs (Hidden, Digitally Open and Advanced Digital CHIs), the overview of scenarios for data integration from EHRI-1 (see 3.1) looks as follows:

Researchers / Hidden CHIs: Manual input
1. Combination of from scratch and copy paste - flows through WP9
2. From scratch only - flows through WP9
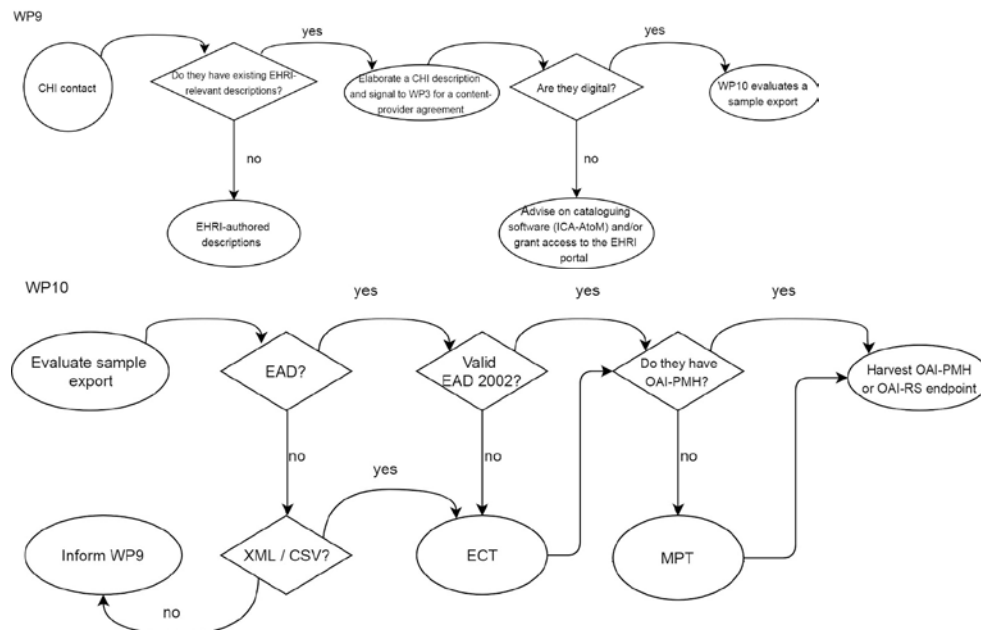3. Copy paste only - flows through WP9
Digital Open & Advanced Digital CHIs: IT-based input
4. Mapping - flows through the tools of T10.1 & T10.2
5. Tailor made script - flows through the tools of T10.1 & T10.2
6. Direct import - flows through T10.3
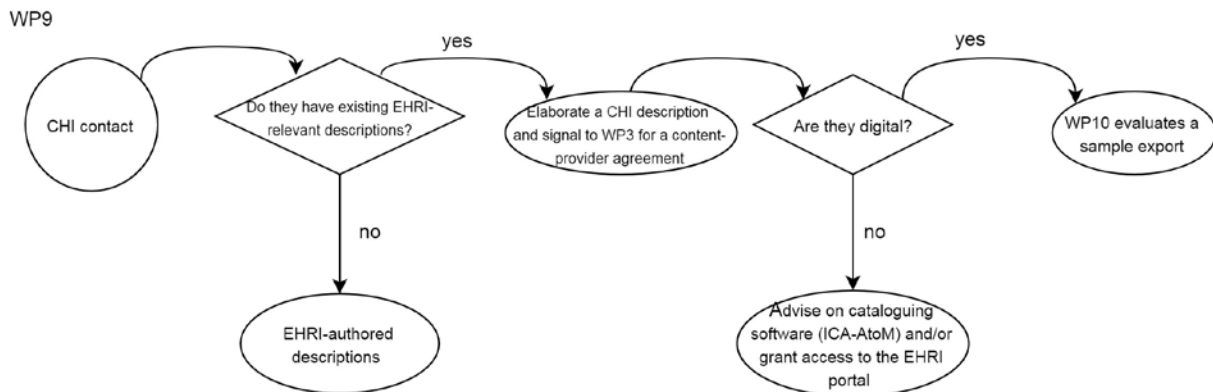7. Pre-process - flows through the tools of T10.1 & T10.2

Graph 4: Data integration in EHRI-2 by Linda Reijnhoudt

However, before we analyse the use case, it is imperative to look at the steps WP9 takes to assess when and how a CHI qualifies for which approach. In graph 5 an overall view is presented between the work WP9 is doing (selection of content, determine if selection is necessary, how to proceed, etc.) and at which point WP9 involves WP10 into the process. When descriptions are digitally available and can be exported, WP9 and WP10 will figure out how the export/import can be achieved. When the CHI has no XML or CSV files than WP10 will notify WP9 to work out a solution on how the import of the descriptions can take place.

Graph 5: Workflow between WP9 and WP10 regarding data imports into the EHRI infrastructure
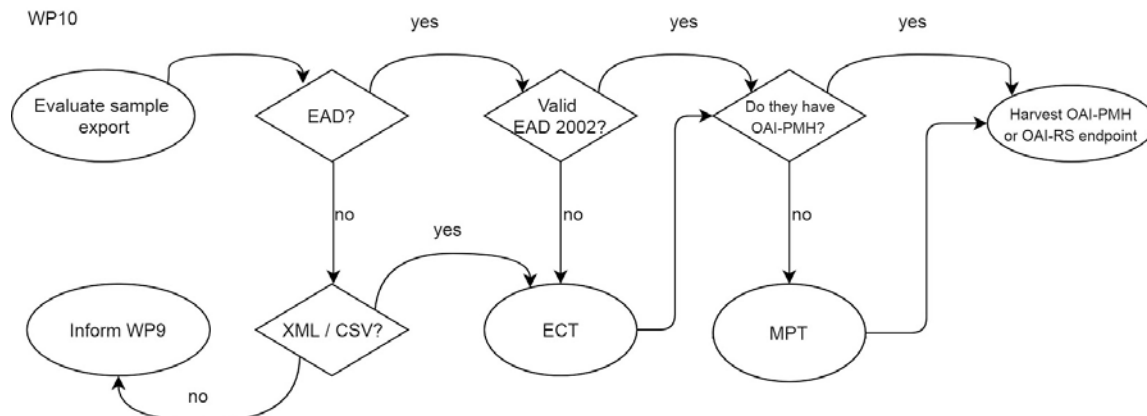
If we look into the details of the workflow of WP9, it is clear that WP9 initiates the first contact with the CHI or, for example after workshops, the CHI makes contact with a member of WP9 to discuss the possibilities regarding data integration into the EHRI portal. Once the first contact has been made, WP9 will evaluate the descriptions the CHI has and determine if they are relevant, digital and if they can be easily exported. In the ideal case, all those three premises are asserted and then WP9 involves WP10 to discuss the details of the future import on a technical level. However, if the CHI cannot provide the ideal case, WP9 will help the CHI to get their descriptions in line with the EHRI standards and to catalogue them manually. WP9 will also suggest a specific collection management software that can generate EAD files, as it is the case for the open-source software ICA-AtoM[3]. The CHI can also be given access to the EHRI Portal in order to add descriptions in there; in this scenario, the added descriptions can be exported in EAD files from the EHRI Portal to the CHI for their own use (publication or creation of website for example).

Graph 6: WP9 workflow for data integration

Once WP9 has established that the CHI has possibilities to do an IT-based import, WP9 contacts WP10 to resolve the details of the import. WP10 will thus further determine if the CHI's descriptions are in valid EAD, which open data sharing program they have (i.e. OAI-PMH) and suggest which steps the CHI has to undertake to complete the import of the data into the EHRI portal. Most CHIs who come in contact with WP10 have possibilities to integrate their data into EHRI on an IT-based import. However, it can occur that some CHIs do not have the right file-format (XML/CSV). When that happens, WP10 will involve WP9 to contact the CHI and to work together towards a solution so that the import of the data can be achieved.

---

3 See: http://www.ica-atom.org/

Graph 7: WP10 workflow for data integration

These graphs illustrate the perfect scenarios for data integration into the EHRI portal. Nonetheless, the selection for data integration can have its own challenges. Therefore, similar to EHRI-1, Digital Open and Advanced Digital CHIs may still fall under the use case "manual input" if those archives are for instance not yet ready to work with EHRI and if selection poses a challenge (see 2.3).

A key point and challenge is communication throughout the process with all involved stakeholders. With WP9 and WP10's aid, all EHRI consortium partners with EHRI-relevant collections have created a Google Docs file providing the content and technical partners with the necessary content, selection and IT-related information to prepare for the data integration and updates into the portal. Also, EHRI created a special webpage (http://ehri-project.eu/ehri-for-institutions) that explains data integration from the CHI's point of view (and not from EHRI point of view), and which provides both contact e-mail addresses and links to documentation on how to work within the EHRI Portal and with the EHRI tools.

3.2.1 Adding (meta-)data to the EHRI portal

Several options are available to publish (meta-)data in the EHRI portal: A first option to integrate data into the portal is to create a sustainable connection via OAI-PMH. The Open Archives Initiative Protocol for Metadata Harvesting[4] provides an application-independent interoperability framework based on metadata harvesting, i.e. regular IT-based import and updates of archival descriptions. This form of data integration is a permanent sustainable connection between the CHI and EHRI. The challenge with this system is that it requires the CHI to have an OAI-PMH facility, which is often not the case.

A second option is available for institutions with valid EAD but no OAI-endpoint. For such institutions, EHRI's WP10 has created a Metadata Publishing Tool (MPT). The MPT tool aims at publishing data from CHIs in a sustainable way. This means it can also publish updated, altered, and new data or even delete information. The tool publishes only in EAD format based on the Open Archives Initiative ResourceSync Framework Specification[5] (OAI-RS). The tool allows the CHI to remain in control of when and what to publish in the EHRI portal, since the CHI will operate the tool at its own institution. The role of EHRI consists in providing the tools for download and to generate an automatic search to see if institutions have updated their collections, to collect the change-lists and to implement them on the EHRI server in order to ultimately publish them in the EHRI portal as well.

---

4 See: http://www.openarchives.org/pmh/
5 See: http://www.openarchives.org/rs/toc

The third option remains manual input into the portal.

To address the needs of institutions who have digital (meta-)data, but not in valid EAD 2002, EHRI WP10 has created a mapping tool which can generate EAD-files semi-automatically, called the ECT tool (EAD Creation Tool). This tool transforms spreadsheets (CSV-files) into valid EAD files, the standard format used by EHRI. The CHI has to provide its metadata in spreadsheets, using one sheet for each collection, and one row for each item. Each item should contain all obligatory fields (such as identifier or title) as well as – where applicable - as many of desirable fields as possible (such as parallel form of title or scope and content). Thus, a mapping configuration is to be established by a cooperation between the CHI and WP10. Both the ECT tool and the MPT tool can be downloaded for free. The ECT tool can be used entirely separate from other tools, but also in combination with the MPT tool.
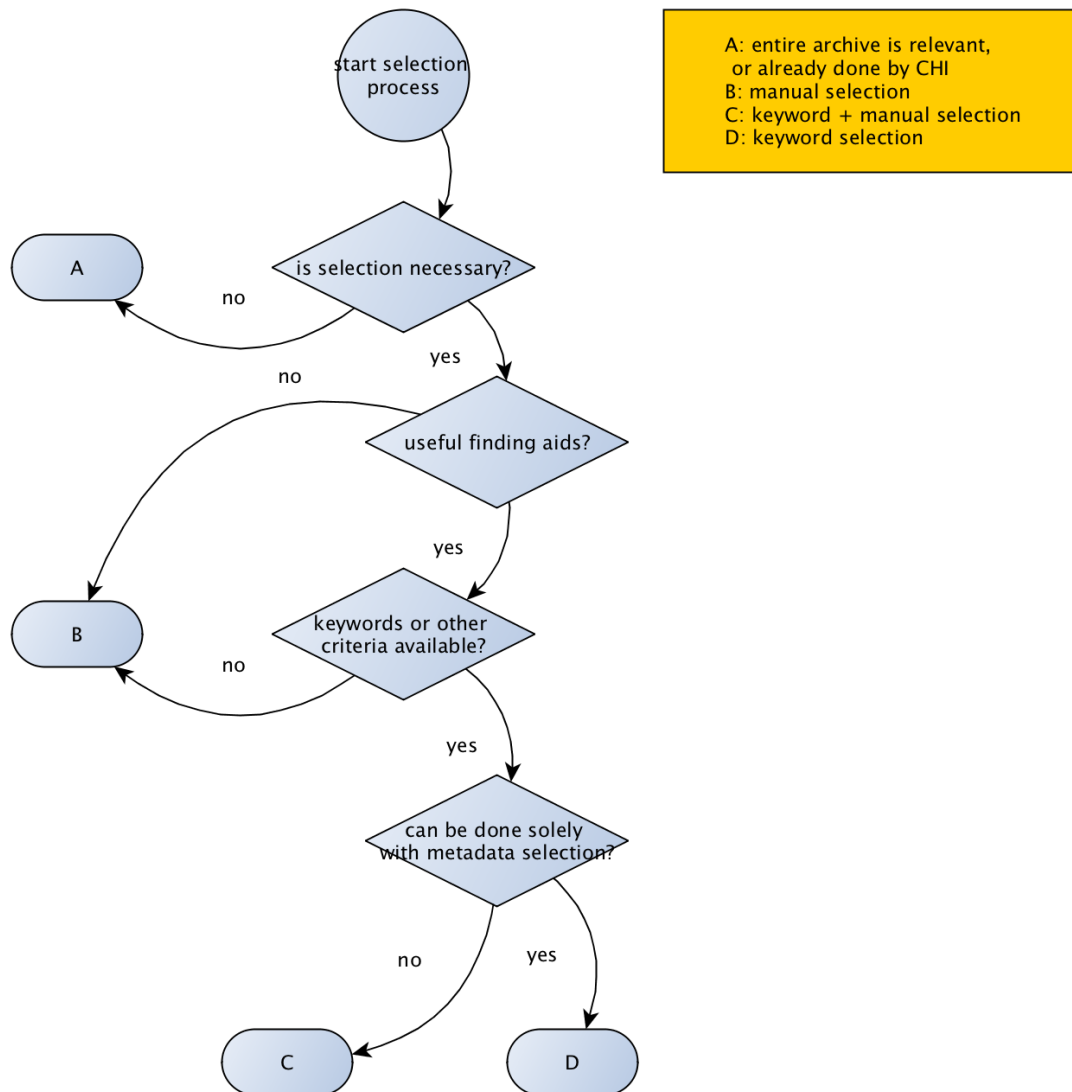
3.2.2 Selection as an additional challenge for data integration

EHRI deals with different types of archival institutions, some of which are focussing exclusively on the Holocaust, whereas others have a wider scope, such as Jewish history and culture, the history of the Second World War or much broader. Therefore, the latter categories will require selection of Holocaust-relevant archival descriptions for import while the first does not (since all its holdings are relevant to be imported into the EHRI portal). The following graph on "Use cases by functionality: selection" gives an overview of the decision process and workflow concerning selection.

CHIs under selection category A in the following graph (entire archive is Holocaust-relevant or the selection is done by the CHI), which meet the technical requirements of WP10, are ideal candidates for the IT-based scenario. If they fall within the parameters of WP10 (ECT/MPT/OAI-PMH), they can be dealt with by WP10 alone (and WP9 need not be involved).

CHIs that fall under category B (manual selection) are those who lack sufficiently detailed descriptions to make a selection based on keywords or other tools. They require a fully manual selection of relevant collections. This is mostly the case for Eastern and Central Europe, where keywords like Holocaust, Jews, etc. are often missing. This can mean that there is quite some manual work to be done, even if the CHI is theoretically a Digital Open or Advanced Digital CHI.

In between are those archives which fall under category C (combination of selection via keywords and manual selection) and D (selection via keywords). For category D, WP9 needs to study and indicate the relevant keywords, after which WP10 can work independently from WP9. Category C combines the working methods of B and D.

Graph 8: Use cases by functionality: selection (Graph by Linda Reijnhoudt)

3.2.3 New use case in EHRI-2: sustainable connection to the EHRI portal

Establishing a sustainable connection between the EHRI portal and CHIs concerning the metadata import has been a major priority of EHRI-2. In this way, the EHRI portal will show the most up-to-date information, and will automatically import newly created archival descriptions. Many CHIs, both inside and outside the consortium, constantly introduce new Holocaust-relevant information in their own collection management systems. The EHRI portal may automatically receive new and updated information thanks to a sustainable connection. This would also ensure that the information in the EHRI portal is always updated. In EHRI-1 all imports were non-sustainable, which also meant that no automated updates from the CHIs to the EHRI portal could take place. EHRI-2 Work Packages 9 and 10 are now working together with several CHIs, both inside the consortium (e.g. BE-KD, IT-CDEC, US-USHMM, SK-HDC, GR-JMG) and outside the consortium (e.g. IT-ICAR, US-YIVO) in order to establish a sustainable data connection, which would also allow for automated updates from CHIs to EHRI

Portal. The sustainable connection is indeed the ultimate goal of all automated import procedures. Automated updates are going to take place by means of an Open Archive Initiative Protocol, either the OAI-PMH, or the OAI-RS[6]. The CHI can autonomously decide how often updating metadata is updated. EHRI WP10 suggests a weekly update; the EHRI data harvester runs on a weekly basis. For a more exhaustive description of the status quo and of the foreseen steps which need to be taken with each institution in order to establish a sustainable connection, please see EHRI-2 deliverable 10.4 (Report on the description integration services), chapter 3 (Status data integration per Collection Holding Institution). EHRI hopes to be able to establish a sustainable connection with the biggest number of CHIs – with a special focus on the EHRI consortium partners - until April 2019.[7]

3.2.4   New use case in EHRI-2: linking of original and copies

A major improvement in the usability of the portal is given by the possibility to directly navigate between related sets of records. The transnational character of the Holocaust is responsible for the fact that multiple institutions have fully or partly copied archival materials and made them available for research purposes in their respective repositories. The link between originals and copies, however, is often partially or fully missing. WP9 has thus been very active in connecting originals and copies of the same documentary unit. It this way, the user can have access from the original set of records to the copies hold by other institutions, or backwards from copies to originals, thanks to the fields "Existence and Location of Copies" and "Existence and Location of Originals".

The departmental archives in France, for example, have been surveyed by USHMM and Mémorial de la Shoah. Copies were made and described at the latter institutions. The CHIs where copy-holding institutions have made copies are systematically being added to the EHRI portal. However, instead of searching and copying French-language collection descriptions from the various online finding aids, the Holocaust-relevant collections are traceable via links between the original-holding CHI and the copy-institutions, and as such provide collection descriptions.

It is important to know that a simple search query does not always permit to find all entries about the same desired item. In fact, originals and copies often do not bear the same title: sometimes the title is attributed by the archivist, so two archivists may give a different title to the very same set of records; sometimes language (or even script) barriers make originals and copies hard to be recognised. In this way, a direct link between originals and copies already represent a major step forward for the user's experience. The fields "Related Units of Descriptions" and "Separated Units of Descriptions" permit finally to navigate between records or records groups that in the former case are related (for instance by provenance), and in the latter case were part of the same record group.

3.3   Status and plans

WP9 is working manually in the EHRI portal. The tasks consist of adding and updating descriptions of collection-holding institutions and manually adding collections (EHRI-written or copied), institutions and country reports. If EHRI had continued working along the lines of EHRI-1, we could expect a similar proportion between scenarios for manual and automatic input (scenarios 1-3 & 4-7) as in EHRI-1. However, EHRI-2's WP10 is working differently than EHRI-1's WP19, namely thanks to the development of new tools. Both above-mentioned WP10 ECT and MPT tool are already fully operational, and many EHRI partner institutions, e.g. BE-

---

6 See: http://www.openarchives.org/rs/toc

7 For more details about sustainable connection, please refer to D10.4

Cegesoma, BE-KD, GR-JMG, SK-HDC, already successfully used one or both tools. In this way, WP9 and WP10 are fruitfully cooperating on automating data integration.

EHRI outcomes equally have to be in line with the vocabularies (WP11), EHRI Guidelines on Description and the General International Standard Archival Description.[8] Task 9.3 in WP9 on linking original-copy and related descriptions equally needs to be considered in the general EHRI data integration plans, as are maintenance and updating (and the need for persistent unique identifiers). EHRI-authored and manually created descriptions have to fully respect EHRI standards; for this reason, EHRI is double-checking the quality and the standard compliance of these entries. We have however to mention the fact that sometimes CHIs are not able to share information, which is marked obligatory by the EHRI Guidelines. Heritage data sometimes lacks information which is presently considered essential. In these cases, EHRI and partner institutions have to agree on which fields cannot be completed because of force majeure.

The amount of manual work for WP9 will remain considerable. There are CHIs where the survey is going to be mostly manual work and personal contact (such as in Slovakia or Greece). For CHIs that are not able to engage with EHRI at this time, the integration falls into WP9 field of work. This is most likely also going to be the scenario for those institutions that have a very limited number of relevant collections and for those where selection is not possible via text or keywords in the collection descriptions. The investment of trying to establish a sustainable connection indeed depends on the number of relevant collections this can provide to the EHRI Portal. This kind of scenario – a limited number of high-level collections with very general descriptions – is common in many countries. It may be more productive to bring in these descriptions manually and in a second step consider a more sustainable connection. This may also be a good scenario for data sources that do not cooperate with EHRI via import at this moment: continue with manual entry and offer a sustainable connection later. Equally, this is a scenario that could be considered in order to add the official descriptions from the CHIs in cases where EHRI authored English-language descriptions (for example in Croatia).

Also, even for those institutions which do have a significant number of relevant standard-compliant and digitally available collection descriptions to integrate, selection remains a major challenge; this work needs a lot of researcher-driven reading (relevant key words / full text search, as most often there are no structured access points).

The only instances where a fairly easy integration with limited input from WP9 is possible is for those archives which have standard-compliant, digitally available descriptions, where all preserved sources are EHRI-relevant (USHMM, YV, Mémorial de la Shoah) or for those which are fairly easy to select from. Also, there are countries where the digital finding aids are not ready to be used by the tools of WP10 as far as we can see at the moment. This mostly concerns PDF documents (which is the case in Italy and Romania, for example). Still, making the distinction between what will be manual and what will be IT-supported work remains difficult and a central task between WP9 and WP10.

---

8 See: https://www.ica.org/en/isadg-general-international-standard-archival-description-second-edition