



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

Deliverable 9.5

Final report on data identification and integration

**Veerle Vanden Daelen
Kazerne Dossin**

**Giles Bennett, Anna Ullrich & Dieter Pohl
Institut für Zeitgeschichte**

**Adina Babeş & Matthew Haultain-Gall
Cegesoma**

With input and feedback from all WP9 members

**Start: May 2015 [M1]
Due: July 2019 [M51]
Actual: July 2019 [M51]**



EHRI is funded by the European Union

Document Information

Project URL	www.ehri-project.eu
Document URL	Not applicable
Deliverable	Deliverable D9.5 Final Report on Data Identification and Integration
Work Package	WP9
Lead Beneficiary	P23 – KD
Relevant Milestones	MS4
Dissemination level	Public
Contact Person	Veerle Vanden Daelen, Veerle.vandendaelen@kazernedossin.eu , +32 15 28 86 23
Abstract (for dissemination)	The final report of WP9 provides an overview of the data that was identified and integrated into the EHRI infrastructure over the course of the project, including a statistical and qualitative analysis of the data as well as information concerning the employed methodology. It provides information about the achievements, challenges and plans for the future of data identification and integration in the EHRI Portal.
Management Summary	Not applicable

Table of Contents

Glossary	4
1 Introduction	5
2 Identification methodology	6
2.1 Countries.....	6
2.2 Archival Institutions	7
2.3 Archival descriptions	8
3 Integration methodology	10
3.1 Sustainable connections as the ultimate goal	12
3.2 Linking.....	12
4 Community and communication	14
5 Achievements, challenges and future plans.....	16
5.1 Achievements.....	16
5.2 Challenges for data identification and integration	17
5.3 Future plans	18

Glossary

CHI	Collection Holding Institution
CSV	Comma-Separated Value
EAD	Encoded Archival Description: an XML-based standard for describing archival materials
ECT	EAD Conversion Tool, developed by EHRI
MPT	Metadata Publishing Tool, developed by EHRI
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting

1 Introduction

The European Holocaust Research Infrastructure (EHRI) started its work in October 2010. EHRI's mission is to support the Holocaust research community by building a digital infrastructure and facilitating human networks. To achieve this, EHRI has developed an online Portal which provides online access to information about dispersed sources relating to the Holocaust, in combination with the development of tools and methods that enable researchers and archivists to work collaboratively with such sources. When the EHRI Portal was launched in March 2015, it provided information on 46 countries, listed 1,853 archival institutions (also referred to as Collection Holding Institutions, or CHIs, within the project to allow for various types of repositories, national and state archives and a wide variety of private and other archives) across 50 different countries and 152,691 archival unit descriptions (of which 18,231 were top-level units, e.g. fonds, record groups, collections) across twenty different countries. As such, EHRI's first phase provided the project with a solid foundation, but there were still many areas that had minimal or no coverage.

In order to attract a critical mass of researchers/users and to facilitate new transnational and innovative research, it is crucial that EHRI identifies and integrates a sufficiently large and new body of knowledge and information on Holocaust-related collections. However, integrating a critical mass of such information is a time-consuming and challenging endeavour as it needs to address the key challenges posed by the existing landscape of Holocaust archives. First, many important Holocaust archives currently remain 'hidden' from view either because they have not been described at all, or because they have been described at a generic level that masks their relevance to the Holocaust. Second, only a relatively small percentage of Holocaust-related archival materials and finding aids have been digitised, and the descriptions that are available online are frequently not standard-compliant or machine-readable. Therefore, the objectives of EHRI data identification and integration has been to:

- Identify Holocaust-related archival materials
- Write new descriptions and map existing descriptions to the EHRI standards and guidelines (based on ICA standards)
- Integrate descriptions of these resources into the EHRI Portal

This report describes the working methods, achievements, challenges of our data identification and integration work. It further outlines broad plans for further work beyond the confines of the current project.

2 Identification methodology

The field of Holocaust studies relies on a huge variety of archives spread throughout the world. One of EHRI's most important tasks has been to create an inventory of Holocaust-related CHIs and their holdings in order to share information about them and help provide access to the research community. The identification, selection and description of Holocaust-relevant archival holdings have been the main focus of EHRI's data integration work. For its data identification work, EHRI has worked with local experts from within and outside the EHRI consortium. Furthermore, EHRI has worked on input via community building and reaching out to archivists and researchers 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 Romania¹.

In its identification work, EHRI has proceeded in a top-down fashion by first dealing with the national archives and the largest institutions preserving materials concerning the Holocaust in a particular country. Special attention is given to those regions where a high number of victims lived or were deported to, and to places where knowledge about Holocaust-related archives is not easily available.

EHRI has had to develop and apply a variety of methods (e.g. selection by keyword matching, date ranges, and provenance information) in order to cover a diverse range of archives; to identify relevant material across dissimilar types of collections; and understand differently structured metadata and data management systems.

2.1 Countries

In order to structure its data identification work, EHRI organises its work based on present-day countries and provides a country report introducing the history, archival situation and EHRI research status for each country. These reports have been written according to a systematic and structured framework. The summaries give readers a concise overview of Holocaust history and sources in the countries described. The reports combine knowledge from important Holocaust encyclopaedias with new findings from recent publications as well as original EHRI research. Due to their concise format, the national report summaries only focus on broad outlines and do not include details or specific bibliographical references. The EHRI country reports are first and foremost a tool to frame the identification and investigation of sources on the Holocaust. An introduction, further explaining each of the sections and the working methods can be found on the EHRI project website (see <https://www.ehri-project.eu/country-reports>).

At this moment, EHRI provides country reports for 63 countries (<https://portal.ehri-project.eu/countries>). This is an increase of 17 since the beginning of EHRI's second phase, where the Portal included 56 countries and 46 country reports. The following map gives an idea of the EHRI Portal coverage. Countries in the European and African theatres of war, wholly or partially controlled, occupied or influenced by Nazi Germany and its allies, are marked in dark blue. All other countries represented in the EHRI Portal are indicated in light blue. The ultimate goal is to include a country report on every country that holds Holocaust-related archival collections and to keep the reports up to date in order to create a dynamic, growing resource for the identification of Holocaust-sources.

¹ For more information, please refer to EHRI D9.2 Integration workshops.

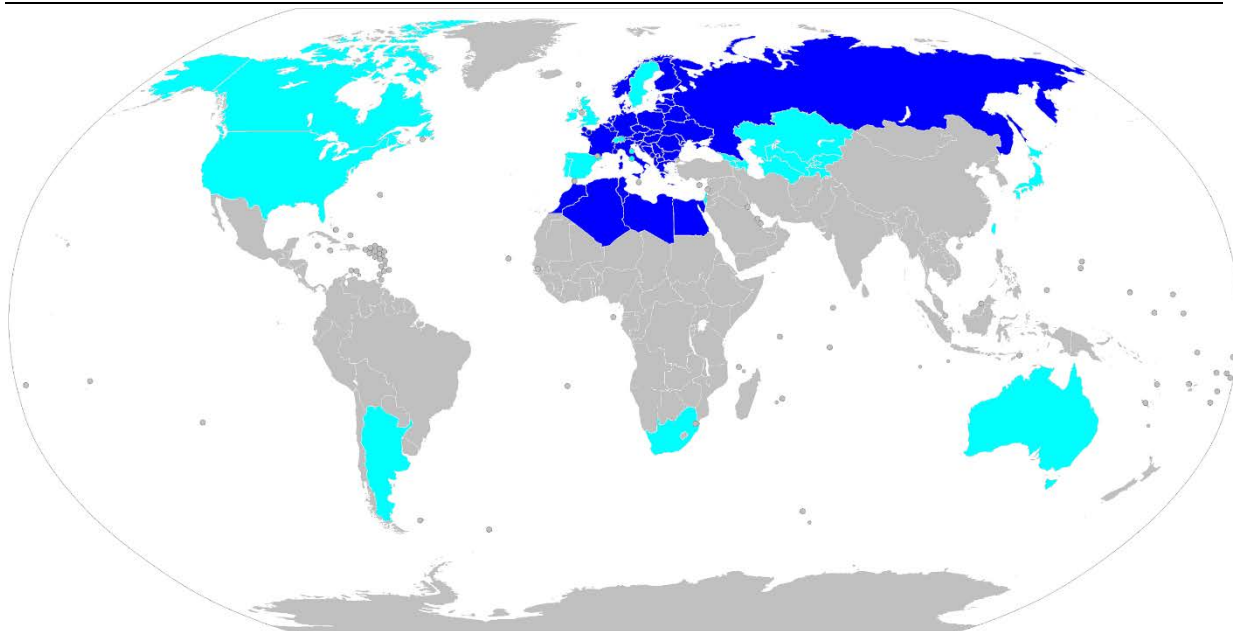


Illustration 1: Map with EHRI Portal coverage

2.2 Archival Institutions

In addition to the country reports, the EHRI Portal provides information about archival institutions (the term used on the Portal to signify CHIs) that hold Holocaust-relevant sources. This is a highly diversified group of institutions, including national archives, local archives, memorial sites, libraries, Holocaust museums, private archives, etc. During EHRI's first phase, 1,853 Holocaust-relevant archival institutions were identified. The project used 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 with which Yad Vashem worked together and has copied archival material from
- An overview of restitution and compensation archives from the Conference on Jewish Material Claims Against Germany
- And, for some countries, additional information was gathered from: the online catalogue of the United States Holocaust Memorial Museum and specific additions from national archival guides, experts, and published studies on the subject

As of 16 May 2019, the EHRI Portal included information about 2,137 archival institutions. There are only four countries that possess a country report that have no archival institutions described in the EHRI Portal (Algeria, Egypt, Libya and Tunisia). The EHRI project does describe sources on these countries preserved in other countries. About 20 of the 63 countries listed on the EHRI Portal have seen a significant increase in the number of identified Holocaust-relevant institutions, including eight countries which did not have any archival institutions included on the Portal by the start of EHRI's second phase. This includes countries in Western and Southern Europe, like France and Italy, as well as Central and Eastern Europe, Central Asia and the Caucasus. It is also worth noting that quality control was performed on the institution descriptions to ensure the data are up-to-date (reflecting mergers, new names and addresses, etc.) and to avoid possible duplication. Overall, the number of archival institutions on the Portal increased by more than 15 per cent over the course of EHRI's second phase.

2.3 Archival descriptions

EHRI wishes to keep the Portal as open-ended as possible; therefore, all collection descriptions on Holocaust-relevant materials are welcome. When writing descriptions of archival materials itself, EHRI focussed on the period from 30 January 1933 to today. An exception was made for “victim sources” (sources from Jews or people considered as Jews under Nazi rule, or victim organizations). Victim sources on Jewish life at the eve of the persecution, including the interwar period have been included. EHRI’s first geographical focus was on Germany and Eastern Europe, the main crime site of the Holocaust, its second focus was on other occupied countries and Axis countries, and our final focus was on other countries such as refugee countries. EHRI used a top-down approach: from national, to regional, to local, to individual. A collection is considered a Holocaust-related collection if it contains at least one file on the Holocaust.

EHRI brought together both already existing descriptions and wrote new descriptions. The sources can be listed as follows:

- Descriptions written by EHRI
- Descriptions written by the CHI
- Descriptions written by a third party, other than EHRI or the CHI (such as finding aids on Jewish sources or research guides on sources on the Second World War)

There are many levels of descriptions depending on the level of hierarchies in the collection management of the respective institutions.

There are cases where an archival unit has more than one description. An archival unit may be described by various sources, and in various languages. All these descriptions can be helpful for a researcher as they provide varying 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. 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.

The identification of sources is an on-going effort. Descriptions of archival materials that are authored by EHRI are written in English. However, when existing descriptions of archival materials are available in languages other than English, they are integrated in their original language. The integration of these descriptions is mostly in the original language, but an English translation is sometimes also possible.

At the start of EHRI’s second phase of funding, the EHRI Portal held 152,691 archival unit descriptions of which 18,231 were top-level archival unit descriptions (such as collection, fonds or record group descriptions, not including child items), held at CHIs in 22 different countries. Between 1 May 2015 and 16 May 2019, the number of top-level archival unit descriptions rose by about 45 per cent and the total number of archival unit descriptions nearly doubled to over 300,000 descriptions. Overall, the Portal currently provides archival unit descriptions from 751 CHIs located in 38 countries.

It is also useful to mention that over 24,500 archival units have parallel descriptions, meaning that these archival units either have descriptions in multiple languages, or descriptions provided by different authors. Given the multitude of languages within the sources, providing descriptions in more than one language can prove useful for researchers. Moreover, archival unit descriptions produced by different authors each working in their own specific context

(timeframe and context, such as a local archive, a Jewish sources survey, or a Holocaust survey, for example) provides complementary information.

Well over one third of the archival institutions in the EHRI Portal have descriptions of archival units and many of the others hold a general description of their Holocaust-related holdings in their institution's description. It is also important to note that the quantity of identified archival units does not necessarily reflect the amount of materials covered, as archives in Central and Eastern Europe, for example, tend to have very large collections comprising several linear meters of archive.

Of key importance is the fact that the EHRI Portal provides information about the archival units held at the largest and most important Holocaust-relevant CHIs that offer a supranational perspective on the Holocaust. These include institutions such as Yad Vashem, the United States Holocaust Memorial Museum (USHMM), the Arolsen Archives, and the Bundesarchiv in Germany. EHRI further focussed on those regions where the Holocaust took place and on the regions with a considerable number of victims. Thanks to existing inventarisation efforts at a national level, the Holocaust-relevant sources within the following countries – in alphabetical order – have been thoroughly covered: Austria, Belarus, Belgium, Czech Republic, the Netherlands, Poland and Ukraine. EHRI invested time and effort with local experts to equally provide overviews for the following countries: Bosnia and Herzegovina, Bulgaria, Croatia, Estonia, Greece, Hungary, Latvia, Lithuania, Luxembourg, Moldova, North Macedonia, Romania, Russia, Serbia and Slovakia. In countries like France and Italy, the historians and archivists within the EHRI team were able to achieve significant results in identifying and integrating Holocaust-relevant collections in a sustainable way. Coverage was also significantly increased for the following countries: Canada, Denmark, Finland, Ireland, Norway, Portugal, Sweden, Switzerland, the United Kingdom, and the Vatican.

3 Integration methodology

As noted above, the EHRI Portal works in a top-down approach, starting from present-day countries (as opposed to the borders from the time of the war). Each country in the Portal receives a so-called “country report” outlining the history and archival situation in the respective country and the status of EHRI’s research there. Attached to the countries are descriptions of archival institutions. Archival institutions in their turn hold archival descriptions and metadata describing the actual archival sources.



The EHRI portal offers access to information on Holocaust-related archival material held in institutions across Europe and beyond. For more information on the EHRI project visit <https://ehri-project.eu>.

Countries

EHRI national reports provide an overview of the Second World War and Holocaust history as well as of the archival situation in the covered countries.

Browse 63 country reports.

Archival Institutions

An inventory of archival institutions that hold Holocaust-related material.

Browse 2,146 archival institutions in 59 countries.

Archival Descriptions

Electronic descriptions and finding aids of Holocaust-related archival material.

Browse 353,096 archival descriptions in 746 institutions.

Illustration 2: Screen shot of the home page of the EHRI Portal, 20 May 2019

All data in the EHRI Portal are structured and organised according to the EHRI standards and guidelines which are in accordance with the standards of the International Council on Archives (ICA): the International Standard for Describing Institutions with Archival Holdings (ISDIAH, for CHIs); 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). Content provider agreements were established where needed, and privacy legislation was monitored by the project.²

Metadata is delivered by researchers (within the EHRI-project and local experts), CHIs and aggregators. A specific page on the EHRI website is dedicated to “EHRI for Collection Holding Institutions” (see <https://ehri-project.eu/ehri-for-institutions>) and explains which routes are possible to integrate data into the EHRI Portal.³

While the work on country reports and CHIs is manual work, there are - in a nutshell - the following options to publish (meta)data of archival descriptions in the EHRI Portal.

² See D.3.3, Report on Content Provider Agreements and Privacy Legislation.

³ This process is also explained in detail in D9.4, “Possible data integration pathways and models in EHRI-2”, p. 9-21, <https://www.ehri-project.eu/sites/default/files/downloads/Deliverables/D9%204%20Resource%20reports%20M36.pdf> retrieved on 20 May 2019.

- A first option to integrate data into the Portal is to create a sustainable connection via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).⁴ OAI-PMH provides an application-independent interoperability framework based on metadata harvesting, i.e. regular IT-based ingests and updates of archival descriptions. This form of data integration is a permanent sustainable connection between the CHI and EHRI. In other words, the CHI can publish updated, altered, and new data or even delete information. This allows the CHI to remain in control of when and what to publish in the EHRI Portal. The challenge with this system is that it requires the CHI to have an OAI-PMH facility, which is often not the case.⁵
- A second option is available for institutions with valid EADs but no OAI-endpoint. For such institutions, EHRI has created a Metadata Publishing Tool (MPT). The MPT 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 Specification⁶ (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 of 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.⁷
- A third option is direct manual input into the Portal. As not all archives have descriptions of their holdings readily available or they do not possess the necessary technical infrastructure to support a direct ingest of their descriptions via the EHRI tools, a significant amount of manual identification, investigation and description work has been undertaken and will remain necessary. A full tutorial, reworked and updated according the current standards and guidelines in EHRI's second phase, can be found under "Manual integration" on the EHRI for Collection Holding Institutions page.⁸

To address the needs of institutions that have digital (meta)data, but not in valid EAD 2002, EHRI has created a mapping tool which can generate EAD-files semi-automatically, called the ECT (EAD Conversion Tool).⁹ 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) and – where applicable - as many desirable fields as possible (such as parallel form of title or scope and content). Thus, a mapping configuration is established by cooperation between the CHI and EHRI. Both the ECT and the MPT can be downloaded for free. The ECT can be used entirely separate from other tools, but also in combination with the MPT.

Working in the EHRI Portal entails close cooperation between historians, archivists and IT specialists. The incentive to work with certain institutions was always motivated by the content of the archives, followed by the way in which they could provide the materials. The actual work required smooth communication between the various abovementioned specialisations, both within the institutions as well as within the EHRI consortium.¹⁰

⁴ See <http://www.openarchives.org/pmh/>

⁵ See <https://ehri-project.eu/ehri-for-institutions#Automated>

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

⁷ See <https://ehri-project.eu/ehri-for-institutions#MPT>

⁸ See <https://ehri-project.eu/ehri-for-institutions#Manual>

⁹ See <https://ehri-project.eu/ehri-for-institutions#EAD>

¹⁰ See D10.3 Training resources and workshop for the publication of archival descriptions, D10.4 Report on the description integration services and D11.5 Report Archivists and Collection Holding Institutions as Users.

3.1 Sustainable connections as the ultimate goal

Establishing a sustainable metadata import pathway from the different archival information systems of CHIs to the EHRI Portal has been a major priority and achievement during EHRI's second phase of funding. Whereas all other data integration pathways are crucial and a step in the right direction, only our sustainable pathway ensures that the EHRI Portal always shows the most up-to-date information by automatically importing newly created or updated archival descriptions from a given CHI's information system. Many CHIs, both inside and outside the consortium, constantly introduce new Holocaust-relevant information in their own collection management systems. In EHRI's first phase all imports were non-sustainable, which meant that no automated updates from the CHIs to the EHRI Portal could take place. During EHRI's second phase not only were the tools developed, but also implemented, resulting into sustainable connections for the following institutions: Kazerne Dossin: Memorial, Museum and Research Centre on Holocaust and Human Rights in Belgium; Yad Vashem in Israel; the Jacob Rader Marcus Center of the American Jewish Archives, the Fortunoff Video Archive for Holocaust Testimonies, and the USHMM in the United States of America. Altogether this resulted in 62,663 imported or updated entries in the EHRI Portal by 16 May 2019. The Centre for Historical Research and Documentation on War and Contemporary Society (Cegesoma) in Belgium; the Foundation Jewish Contemporary Documentation Centre in Italy; and the NIOD Institute for War, Holocaust and Genocide Studies in the Netherlands are as of 22 May 2019 on the verge of having a sustainable connection in place. This would allow for another 10,776 archival units, bringing us to a total of 73,439 sustainably imported archival unit descriptions in the EHRI Portal.

At the same time, the one-off ingests (bulk imports) of descriptions from the Belgian State Archives in Belgium; the University of Toronto Archives and Records Management Services and the Vancouver Holocaust Education Centre in Canada; the Jewish Museum of Prague in Czech Republic; the Dachau Concentration Camp Memorial Site, the Free University of Berlin, and the Arolsen Archives: International Center on Nazi Persecution in Germany; the National Archives and the Mémorial de Shoah in France; Jewish Museum of Greece; the Hungarian Jewish Archives; the Historical Archives of the Jewish Community of Rome in Italy; the National Archives and the Institute for War, Holocaust and Genocide Studies in the Netherlands, and the Archives Hub in the United Kingdom resulted in 109,179 imported or updated entries in the EHRI Portal by 16 May 2019.¹¹

However, the manual work carried out by EHRI is equally pioneering as, for the majority of these manually integrated Holocaust-relevant descriptions, it is the first time they have been made available in a standardised, digital format. As such, they open up Holocaust-relevant sources that until now were completely hidden. Also, this trajectory has proved to be a solution for those institutions where the process of creating a sustainable connection or doing a one-off ingest was, for many reasons, not possible.

3.2 Linking

A major improvement in the usability of the Portal is 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. Making clear connections between originals and copies re-contextualises archival materials. Different descriptions of the same archival units each have their own merits. Since the descriptions are made in a given context, namely a specific country, era and institution, they shape the metadata, its subjects and emphasis. As such, the metadata echo the history of each organization and its institutional and cultural agenda. Multiple descriptions provide additional perspectives for

¹¹ The ingests are described in detail in D10.4, Report on the description integration services.

research and thus it is not only interesting to connect the metadata of original and copy descriptions, but this re-contextualisation also enhances knowledge and facilitates research, including that concerned with the relationships between archival units and their metadata.

The departmental archives in France, for example, have been surveyed by both the USHMM and Mémorial de la Shoah. Copies were made and described at the latter institutions. The departmental archives in France where copy-holding institutions have made copies have systematically 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.

EHRI has been very active in connecting originals and copies of the same documentary unit. In this way, the user can gain access from the description of the original set of records to the description of the copies held by other institutions, or the inverse – from copies to originals – thanks to the fields “Existence and Location of Copies” and “Existence and Location of Originals”. Often even central documents are available in copied form in many locations without knowledge about the location and surrounding context of the original. Solid interlinking and integration of information can help to solve this problem. This not only provides extra information to researchers, but also allows them to decide on the most suitable location for consultation.

As a first step, a definition of methodology for linking and a theoretical framework has been developed. Practical, methodological and machine-based approaches were discussed at various workshops. In practice, EHRI has implemented the following options to connect copies to originals:

- Copy archival unit to original archival unit
- Copy archival institution to original archival institution
- Copy archival unit to original archival institution
- Copy archival institution to original archival unit

Standards and guidelines on linking original-copies were implemented in the EHRI tutorial.¹² In March 2019, there were 1,941 original-copy links implemented in the EHRI Portal. Most of the links - 1,865 of the 1,941 - are from a copy archival unit to an original-holding archival institution. Most of these, 1,767 in total, are made from USHMM copy collections. This makes the USHMM copy-links the core of the implemented links. They provide crucial information on Holocaust-relevant collections across many different countries and CHIs.

¹² See Sigal Arie-Erez, Yehudit Levin, Veerle Vanden Daelen, Laura Brazzo, Anna Ullrich, Giles Bennett, Adina Babeş, Francesco Gelati, Kepa Rodriguez, Linda Reijnhoudt, D9.3 Report on Linked Original-Copy Relationships in archival unit Descriptions, April 2019.

4 Community and communication

EHRI finds it extremely important to inform its users about how the information they are consulting was put together. Therefore, 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 (<https://portal.ehri-project.eu/help/faq>). 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.

The various EHRI data identification and integration 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 EHRI data identification and integration workshops promoted the EHRI project and Portal and were an opportunity to convey EHRI’s missions to new (meta)data providers and to encourage cooperation with EHRI.

At the same time, EHRI spent a considerable amount of effort in communicating to a wider audience about its successes of data integration into the EHRI Portal. This included articles in various EHRI communication tools such as the EHRI newsletters, EHRI Facebook posts and Twitter feeds, as well as contributions on the EHRI Document Blog.¹³ EHRI also invested time and effort in spreading news via external conferences and different types of publications.¹⁴ On Wednesday 6 November, EHRI will be engaging further with its key users

¹³ See, for example, “50,000 Archival Descriptions Added – Major Increase in EHRI Portal Collection Descriptions”, EHRI News 11 April 2019, <https://www.ehri-project.eu/50000-archival-descriptions-added> retrieved on 22 May 2019; European Holocaust Research Infrastructure Facebook page post on 16 April 2019; EHRI Tweets @EHRIproject on 3, 5, 8, 10, 12 and 15 April 2019; Dorien Styven, Marius Caragea, Veerle Vanden Daelen, “The Learning Curve in Sharing Data with the EHRI Project: The Example of a Memorial Site, Kazerne Dossin, Mechelen”, EHRI Document Blog, <https://blog.ehri-project.eu/2018/06/19/kazerne-dossin-mechelen/>, retrieved on 20 May 2019.

¹⁴ See, for example, Veerle Vanden Daelen, “Data Sharing, Holocaust Documentation and the Digital Humanities: Introducing the European Holocaust Research Infrastructure (EHRI)”, *Umanistica Digitale*, theme issue “Data Sharing, Holocaust Documentation and the Digital Humanities: Best Practices, Case Studies and Benefits”, 4 (2019), <https://umanisticadigitale.unibo.it/article/view/9036/8934> (<http://doi.org/10.6092/issn.2532-8816/9036>); Giles Bennett & Veerle Vanden Daelen, “The European Holocaust Research Infrastructure (EHRI): Taking Stock”, in: Frank Bajohr & Dieter Pohl (eds.), *Right-Wing Politics and the Rise of Antisemitism in Europe 1935-1941* Göttingen: Wallstein Verlag/München: Institut für Zeitgeschichte – Zentrum für Holocaust-Studien (2019) 249-259 (European Holocaust Studies, vol.1); Veerle Vanden Daelen, “Making sure the data fit the researchers – Data identification and investigation in European Holocaust Research Infrastructure (EHRI)”, in Agiatis Benardou, Erik Champion, Costis Dallas and Lorna M. Hughes (eds), *Cultural Heritage Infrastructures in Digital Humanities* (London/New York: Routledge, 2018) 97-111; Giles Bennett and Veerle Vanden Daelen, “Reaching out to Memorial Sites and Archivists: the European Holocaust Research Infrastructure in its Second Phase”, lecture at the Internationale Datenbanktagung der Gedenkstätten, Bergen-Belsen, 14-16 September 2016; Adina Babeş, László Csósz, Jan Hlavinka & Veerle Vanden Daelen, “The European Holocaust Research Infrastructure (EHRI)’s work on Hungary, Romania and Slovakia: A digital infrastructure and a human network connecting archives, archivists and researchers”, Presentation at the Academic Working Group of the International Holocaust Remembrance Alliance, Bucharest, 24 May 2016; Veerle Vanden Daelen, “The European Holocaust Research Infrastructure (EHRI)’s work on the former Soviet Union: A digital infrastructure and a human network connecting archives, archivists and researchers”, Presentation at the International Conference: Holocaust Documentation in the Territories of the FSU. Current Issues of Mapping, Accessibility and Usage, Yad Vashem, Jerusalem, 18-20 October 2015; Yael Gherman, Petra Links & Veerle Vanden Daelen, “Presenting the European Holocaust Research Infrastructure”, Presentation at ‘Europeana Research Invites – Using European Infrastructures for Humanities research: Scoping Content, Tools and Users, Digital Curation Unit (DCU), Athens, 12-13 October 2015; Veerle Vanden Daelen, “Introducing the European Holocaust Research Infrastructure”, Presentation at the 1st Workshop on Digital Humanities organised by the LINDAT/CLARIN Center, Faculty of Mathematics and Physics, Charles University in Prague (under the auspices of the Ministry of Education, Youth and Sport of the Czech Republic), Prague, 24 September 2015; Veerle Vanden Daelen, “The European Holocaust Research Infrastructure: from FP7 to H2020”, DARIAH-FED presentation digital humanities projects federal research institutions (Belspo), Brussels, 28 August 2015.

– academic researchers during the “Special Lessons and Legacies Conference: The Holocaust and Europe: Research Trends, Pedagogical Approaches, and Political Challenges” (Munich, 4-7 November 2019) with a workshop entitled “In dialogue with the Researcher: Exploring the Offerings of the European Holocaust Research Infrastructure and its Interactions with the Research Community”. All of this broadens knowledge and improves visibility, which is not only beneficial and appreciated within the EHRI consortium, but equally by all, including data providers, who contribute to its success.

5 Achievements, challenges and future plans

5.1 Achievements

During EHRI's second phase of funding, significant progress was made on the content provided in the EHRI Portal. With more countries covered and 17 new country reports, EHRI is now covering a much wider geographical area. Moreover, all information provided in the country reports during EHRI's first phase was revised and updated according to the work carried out since May 2015.

The overview of CHIs has been further completed with a considerable number of new institutions added to the Portal. Furthermore, many existing institutional descriptions were enhanced with more detailed information, and EHRI executed quality control of the data that were already available.

With a duplication of the number of archival unit descriptions provided in the EHRI Portal and, for the first time, the establishment of sustainable connections between various institutions and the EHRI Portal, EHRI has significantly enlarged the content it offers.

However, the numbers do not say it all. On a qualitative level, EHRI has continued working on key regions where the Holocaust happened and has focused on the victim sources. EHRI has both integrated existing information and actively surveyed and provided new descriptions. The work carried out does not necessarily result in equal numbers of identified institutions and archival units per country or in numbers that are relative to the size of the country. Centralised archival systems, for example, lead to less CHIs and very large archival units. These systems are typically found in former communist countries where it is difficult to identify other repositories which may be included in the EHRI Portal. Changing borders and regimes led to language issues and challenges for local experts. The crucial contribution of a project like EHRI is that it can bring together experts from such regions to advance knowledge through cooperation. This has been the case for Hungarian, Romanian and Slovak sources, for example. Moreover, thanks to EHRI's efforts, archival material which remained unprocessed until now (or processed without sufficient information for Holocaust research) has become available via the EHRI Portal (this is often the case for peripheral archives).

The work of Yad Vashem and USHMM as the two major aggregators remains unprecedented and crucial for Holocaust research. In order not to duplicate their efforts, the EHRI project does not follow the same method of copying sources; EHRI concentrates on bringing together information about the sources and, as such, serves as a kind of roadmap for Holocaust research. Altogether, the EHRI Portal now covers the most important types of Holocaust sources and, as a result, offers a comprehensive overview. Although archival cultures are very different one from another, we have a mid-level homogeneous overview of what is available and what is highly relevant for Holocaust research. So far, the EHRI project covers 63 countries in the EHRI Portal. In areas of central importance to the Holocaust, EHRI has unrivalled coverage in a balanced and even structure.

Last, but not least, the EHRI project contributes in a considerable way to the re-contextualisation of Holocaust-related archives by developing and implementing a methodology to link originals and copies to each other via the CHIs and archival unit descriptions.

Overall, the work carried out in the EHRI Portal enhances awareness to Holocaust-related sources and the institutions preserving them. The archival overviews are a clear call to keep on improving access to the sources and open them up for research.

5.2 Challenges for data identification and integration

An important challenge for surveying Holocaust-related materials is the specific archival culture of each country. For example, in former communist countries, like Belarus, Russia, and Ukraine, the archival systems are highly centralised and bureaucratised, with a tradition of limited access, both to finding aids and files. Even in countries with more liberal access traditions, certain archives may not be open for consultation, such as military and secret police files and documents relating to active administrations. Such access restrictions can also apply to particular church and private archives.

The state of archival infrastructures can also be a challenge. Certain countries already have Holocaust-related databases, archival guides and a variety of research publications, whereas others have only limited and fragmented information. EHRI's research team was confronted with very basic, non-digital infrastructures, often caused by lack of funding.

The sheer volume of materials in Germany and countries most heavily affected by the Holocaust, such as Poland, creates a major challenge in itself.

Another problem is the state of general knowledge on the Holocaust and Holocaust-related material (often combined with language challenges, see below), which sometimes leads to insufficient sensitivity to the subject in archival cataloguing and unspecific archival descriptions, especially of occupation files. This is often combined with limited local research and lack of attention to the subject of the Holocaust in local academia. Nevertheless, we are convinced that contact with EHRI will definitely have raised awareness of this topic. Furthermore, EHRI is working in a different environment now as compared to its start in 2010: the politics of memory and new tendencies in memorialisation do not always provide a political culture that facilitates EHRI's work.

Another key challenge is language-related, both for the researchers and the local staff (for example, the language of a finding aid and of the described material may thus often not be the same). For example, borderland regions with frontier and regime changes, shifting minorities and populations, pose language challenges, as we have, for example, seen in our workshop on Holocaust archives, research and education pertaining to the borderlands area of northern Transylvania (Romania), north-eastern Hungary, southern Slovakia, Ukraine and Moldova.¹⁵

While many institutions show a great deal of enthusiasm for our project (including CHIs with a limited percentage of Holocaust-related collections among their holdings), a lack of resources (both financial and staff) often hinders the integration of information on the Portal (e.g. several Israeli archives), unless local resources are available and the staff is very committed to the task.¹⁶ The latter, however, remains the exception and entails quite some work from the CHIs themselves (see for example Yad Vashem or Kazerne Dossin). In the

¹⁵ Adina Babeş and Marius Cazan, "EHRI Conference in Bucharest on Borderlands: 'Borders Are Imaginary'.", <https://www.ehri-project.eu/ehri-workshop-bucharest-borderlands-borders-are-imaginary>, retrieved on 24 May 2019.

¹⁶ This is also evidenced in the fact that the largest proportion of respondents – over 60 per cent - to an EHRI survey indicated that up to 25 per cent of their holdings were Holocaust-relevant (Annelies van Nispen, Reto Speck, Rachel Pistol, Marco Braghieri, *D11.5 Report Archivists and Collection Holding Institutions as Users*. November 2018, p. 9, 18, https://www.ehri-project.eu/sites/default/files/downloads/ehri_downloads/EHRI%20News/D11%205%20Report%20archivists%20and%20collection%20holding%20institutions%20as%20users.pdf, retrieved on 22 May 2019.

majority of the cases results were achieved only when EHRI invested time and resources, which was not possible in every desirable case. It was not always easy for institutions to install free software provided by an external project (for example in public institutions, installing a piece of software is a decision not to be made by the archivist or IT person alone, but by the larger structure and management).

Finally, the EHRI Portal with its data integration from various different sources and carried out by a wide range of people, poses a challenge for standardisation. While EHRI has established guidelines to help structure data integration, we acknowledge that there are still deviations from the standards set. By their very nature, free text fields lend themselves to varying interpretations when users fill them out. This is part of the heterogenous nature of the project. The guidelines have always been the guiding principle, yet when they were not followed 100% by data-providers from outside or inside the consortium, we decided to include rather than to exclude and keep this information from the researchers (who in general do not even notice more creative interpretations of the standards and guidelines). Thorough and systematic quality control has already addressed most issues.

5.3 Future plans

Even though EHRI has already investigated an impressive number of repositories and archival holdings, the challenges for identification and investigation remain significant, mainly due to the vast amount of materials to integrate and the fact that difficult access issues slow down the process. The EHRI Portal needs to continue the identification and integration of resources to further support the advancement of knowledge about the Holocaust.

To ensure that the EHRI Portal is as sustainable as possible, the project needs to continue gaining insight into how to better engage with professional users, notably researchers and CHIs. In a handful of cases, such as the Archives Hub in the United Kingdom, the Dutch national archives and the French national archives, EHRI was able to work very smoothly thanks to the fact that these have OAI-PMH endpoints open to the public, meaning EHRI could come and make a sustainable connection with their data without having to contact anyone. This, however, is still the exception as, for most CHIs, bringing their data into the EHRI Portal involved not only online contact, but also in-person meetings, and negotiations on changes to be made to the local systems to work with the EHRI tools. Challenges of adequate staffing, finances, and technology remain significant. As only half of the respondents to an EHRI survey indicated that they provided digital finding aids online (many institutions, especially memorial sites, only provide access on site to their digital catalogues due to the sensitive nature of their archival materials and privacy issues), this means that EHRI will have to take this into account and keep on offering scenarios and solutions for various degrees of digital openness. Especially smaller institutions (i.e. 1-10 employees) have less digital collections and therefore require a different kind of support than large institutions, which are much more likely to have digital catalogues.¹⁷ The more EHRI understands how to reach out in the most efficient way for both the project and the institutions, the better.

As a digital humanities project, it is key for EHRI to further engage with projects like DARIAH, (Digital Research Infrastructure for the Arts and Humanities) and Parthenos (Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies) to further strengthen interoperability and coordination between various initiatives. During both the first and second phases of funding, EHRI has been in touch with both projects, as well as with other digital humanities projects that have produced online databases on sources that at least partly overlap with EHRI's field of activity. These were

¹⁷ van Nispen, Speck, Pistol, Braghieri, *D11.5 Report Archivists and Collection Holding Institutions as Users*. p. 10-11, 13, 17, 23-24.

transnational projects, like CENDARI (Collaborative European Digital/Archival Infrastructure), APEX (Archives Portal Europe), Judaica Europeana, the Yerusha project (<http://yerusha.eu/>)¹⁸, as well as projects covering multiple countries in a certain region like Project Judaica (<http://www.jtsa.edu/project-judaica>), and national infrastructures and initiatives such as the Austrian, Belgian, British, Dutch, Israeli, Polish or Ukrainian ones EHRI engaged with.

To connect the sources to educators and a general audience, it is of key importance for EHRI to ensure its online teaching materials and tools, the Document Blog, and the EHRI Portal are well connected to each other. Moreover, the EHRI project should especially accommodate local interest and approaches with its wealth of information as this allows teachers and historians, as well as a general audience, to find information about their own geographical region, research area and points of interest. To accommodate to these “Heimatgeschichte”-tendencies, EHRI needs to further invest in proper geo-coding and geo-tagging so that sources about specific locations, which have been scattered worldwide, can easily be found by those interested. In this way, more and disparate sources can be included in local research, education and commemoration. In general, EHRI’s communications efforts should continue including news about the EHRI Portal as this is both very effective and motivating for the data providers and end users.

¹⁸ Gabor Kadar, “Unifying Our Scattered Heritage – The Yerusha Project”, EHRI News 2 April 2019, <https://www.ehri-project.eu/unifying-our-scattered-heritage-yerusha-project>, retrieved on 22 May 2019.