

## Data Sharing, Holocaust Documentation and the Digital Humanities: Introducing the European Holocaust Research Infrastructure (EHRI)

Veerle Vanden Daelen

Kazerne Dossin, Memorial, Museum and Documentation Centre on Holocaust and Human Rights,  
Mechelen, Belgium

[veerle.vandendaelen@kazernedossin.eu](mailto:veerle.vandendaelen@kazernedossin.eu)

**Abstract.** The European Holocaust Research Infrastructure (EHRI) started its work in October 2010 with financial support from the European Union. The project, which is currently under its second funding phase, continues developing according to its' mission to support the Holocaust research community by building a digital infrastructure and facilitating human networks and by helping networking of Holocaust researchers and archives. EHRI provides online access to information about dispersed sources relating to the Holocaust through its Online Portal. Tools and methods are developed that enable researchers and archivists to collaboratively work with such sources and explore new methodologies within digital humanities. This contribution seeks to present the resources and services EHRI has to offer to the research community, with a special emphasis on the EHRI Portal.

European Holocaust Research Infrastructure (EHRI) è un progetto lanciato nel 2010 grazie al sostegno dell'Unione Europea. Fa parte degli obiettivi del progetto la creazione di una infrastruttura digitale a supporto della comunità degli studiosi della Shoah e l'implementazione del networking fra ricercatori e istituti di conservazione. Nell'ambito di EHRI metodi e strumenti di lavoro sono sviluppati con l'obiettivo di favorire il lavoro collaborativo fra ricercatori e archivisti e, di studiare nuove metodologie per le Digital Humanities. Grazie al suo portale web (Online Portal) EHRI fornisce l'accesso alle informazioni sulle risorse d'archivio per la storia della Shoah disperse nei numerosi archivi europei e internazionali. Proprio i servizi e le risorse disponibili attraverso il portale web EHRI costituiscono il principale argomento di questo paper.

## Introduction

The European Holocaust Research Infrastructure (EHRI) started its work in October 2010 with initial financial support from the European Union Seventh Framework Programme for four years. Thanks to the continued EU support - EHRI is currently (2015-2019) a Horizon2020 EU-financed project with a total budget of almost eight million Euros - the project keeps on developing. The consortium in EHRI's second phase under H2020 consists of 24 partner institutions from 17 different countries and includes research institutions, libraries, archives, museums, memorial sites and e-science specialists.<sup>1</sup> Apart from this core working group, EHRI equally relies on the support of many other individuals and organisations in the broad fields of Holocaust studies and digital humanities. EHRI is devoted to building a Holocaust research infrastructure that is sustained by its network and will have a right of existence on its own accord.

The mission and main objective of the European Holocaust Research Infrastructure (EHRI) is to support the Holocaust research community by building a digital infrastructure and facilitating human networks and by helping networking of Holocaust researchers and archives. EHRI provides online access to information about dispersed sources relating to the Holocaust through its Online Portal. Tools and methods are developed that enable researchers and archivists to collaboratively work with such sources and explore new methodologies within digital humanities. Apart from providing an online platform, EHRI also facilitates an extensive network of researchers, archivists and others to increase cohesion and co-ordination among practitioners and to initiate new transnational and collaborative approaches to the study of the Holocaust. EHRI thereby seeks to overcome one of the hallmark challenges of Holocaust research: the wide dispersal of the archival source material across Europe and beyond (because of the geographical scope of the Holocaust, attempts to destroy the evidence, migration of Holocaust survivors etc.) and the concomitant fragmentation of Holocaust historiography with a multiplicity of documentation projects. By bringing together experts from different fields, and by building an innovative digital infrastructure supported by a large community, EHRI is a flagship project that showcases the opportunities for historical research in the digital age.

With this presentation at the EHRI workshop “Data Sharing, Holocaust Documentation, Digital Humanities: Best Practices, Case Studies, Benefits”, we would like to present the

---

1 NIOD, Institute for War, Holocaust and Genocide Studies (Amsterdam), Yad Vashem (Jerusalem), National Archives Belgium/CEGESOMA (Brussels), King's College (London), Institute for Contemporary History (Munich), Jewish Museum in Prague, DANS (Den Haag), Wiener Library (London), Vienna Wiesenthal Institute for Holocaust Studies, Jewish Historical Institute, ŻIH (Warsaw), Mémorial de la Shoah (Paris), International Tracing Service (Arolsen), United States Holocaust Memorial Museum, USHMM (Washington D.C.), Bundesarchiv (Berlin / Koblenz), Elie Wiesel National Institute for the Study of the Holocaust in Romania (Bucharest), Hungarian Jewish Archives (Budapest), Vilna Gaon State Jewish Museum (Vilnius), Dokumentačné stredisko holokaustu (Bratislava), Contemporary Jewish Documentation Center Foundation CDEC (Milan), The Jewish Museum of Greece (Athens), Ontotext (Sofia), INRIA (Le Chesnay), Stowarzyszenie Centrum Badań nad Zagładą Żydów (Warsaw), Kazerne Dossin: Memorial, Museum and Documentation Centre on Holocaust and Human Rights (Mechelen).

resources and services EHRI has to offer to the research community, with a special emphasis on the EHRI Portal. EHRI resources and training include: Online Portal with information on Holocaust-related archival material held in institutions across Europe and beyond, Online Training in Holocaust Studies; Seminars and Workshops; Fellowship Programme; Conferences; Online Document Blog; Online Research Guides; and Tools and Methods for Digital History.

### **EHRI Fellowships, Online Courses, Research Guides, Document Blog & Workshop**

The EHRI Fellowships support and stimulate Holocaust research by facilitating international access to key archives and collections as well as archival and digital humanities knowhow. The fellowships intend to support researchers, archivists, curators, digital humanists, and younger scholars (for information on past fellows and open calls, see <https://ehri-project.eu/ehri-fellowship-call-2016-2018>). The EHRI online courses also address the researchers, the general public and data managers/archivists (<http://training.ehri-project.eu>). There is on the one hand an unguided online course with 6 units from EHRI's first phase (<https://training.ehri-project.eu/>), as well as now the development of an interactive tutored online course with six lessons (<https://ehri-project.eu/interactive-ehri-online-course-holocaust-studies>) and a Bundesarchiv-written course on German Archivistics (Aktenkunde).

Whereas in EHRI's first phase two Research Guides were published online (<https://portal.ehri-project.eu/guides>; e.g. on Theresienstadt, <https://portal.ehri-project.eu/guides/terezin>), EHRI is now exploring the options of its relatively new EHRI Document Blog (<https://blog.ehri-project.eu/>). The EHRI document blog provides a space to share ideas about Holocaust-related archival documents; it provides an innovative platform for the presentation, visualization, contextualization and interpretation of the data and metadata, using digital tools. EHRI furthermore reaches out and explores new methodologies via workshops and methodological seminars. These include, for example, a seminar for conservationists working on Holocaust-related materials and workshops on specific topics.

### **The EHRI Portal**

The EHRI portal offers information on 57 countries, descriptions on 1,939 archival institutions across 51 countries, and 231,888 archival descriptions in 479 institutions (<https://portal.ehri-project.eu/> July 17, 2017). The data in the portal are structured in a top-down fashion: from countries with country reports on the history, archival situation and status of EHRI's research on the country, to an inventory of institutions which preserve Holocaust-relevant sources within these countries, to top-level collection descriptions (be it record groups, fonds, subfonds, collections or any other way the institution describing the sources structures them). EHRI's goal is to provide information on the archives, not to provide digital representations of all the archival materials. EHRI focuses on collection descriptions and is not

aiming to be a “scan depository”, nor does it aim to be a complete public database on the (often privacy sensitive) file or document level (although we will take those descriptions if we can have them).

So, instead of a digitization project giving direct access to the sources, it should be seen more as a “routeplanner”, assistance to researchers to identify the sources they need and to see information on the sources they are looking for across different institutions, languages and countries. If EHRI is aware of where the sources themselves are consultable in a digital format, it will include the link into the description. Storing all digital images of Holocaust-relevant archives is at this point not something the project can support or is aiming for. However, the contextualization, the merging of information on the sources across this many different countries and institutions is a tremendous help for researchers to identify their sources and it also allows the institutions preserving these sources to communicate to the international research community which sources they are holding, as quite often sources ended up in unexpected places.

As such, EHRI is making sources visible in a systematic fashion in order to counteract the fragmentation of the sources. The project reveals interconnections (e.g. through a multilingual thesaurus with approx. 5470 terms; collation of authority files; relationships between originals and copies). The goal is to keep expand and enriching the online inventory of institutions and collections pertaining to the Holocaust in Europe, Israel and beyond, and to connect archives and users. The contextualization of the sources they preserve is indeed useful for the archives, as well as receiving potential expert user feedback. As such, the project is mutually useful for both researchers and archives, and the positive news on this ever-growing portal is that it has attracted in a relatively short time after its launch a high number of unique users, who make frequent use of this resource.

## **Integration of Metadata into the EHRI Portal**

In EHRI’s first phase, the metadata in the portal were either bulk-imported by EHRI-IT or manually added by historians within the project. However, all data entry remained non-synchronized as even the bulk imports were one-time only harvests or imports. In EHRI’s second phase, the key factor of attention is on ensuring sustainable, meaning updatable, connections between the metadata providers and the project’s portal.

### ***Multiple Scenarios for Metadata Integration***

The integration of metadata into the EHRI portal can be done in various ways and can involve interaction with historians, archivists and IT/digital humanists within the project (as identified in the figures beneath). Typically, the historians and archivists (Work Package 9 - WP9 in the figure) will indicate which archives contain Holocaust-relevant collections. They will verify whether or not the institution already has descriptions of the sources or not. In the latter case, EHRI may decide to write collection descriptions itself or hire a local expert to do so. In the first case, with descriptions available, the first follow-up question is whether or not the

descriptions are digital and if so, in which format and in an exportable way or not. Samples of exports are provided to the EHRI IT of Work Package 10 (WP10) so that EHRI can assess which pathway would be possible to ingest information on the institution in case into the EHRI portal (hence the reference to 10 in the figure).

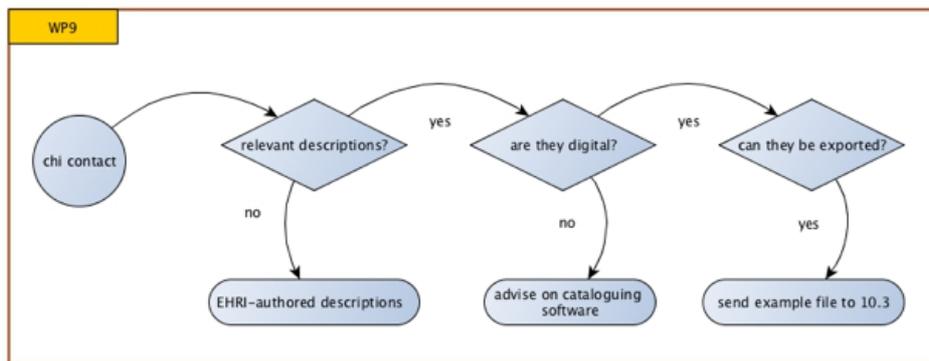


Figure 1: Workflow between WP9 and WP10 regarding data imports into the EHRI infrastructure (2) (EHRI, D9.4 Resource Reports, Update April 2017).

WP10 consequently verifies the sample export to evaluate whether or not the export is valid Encoded Archival Description or EAD and whether or not the institution has a Protocol for Metadata Harvesting (PMH) Endpoint. When both these questions get affirmative answers, establishing a connection between the institution and the EHRI project is a fairly straightforward endeavor, which only then entails the signing of a content-provider agreement (CPA) to ensure a sustainable connection to the EHRI project.

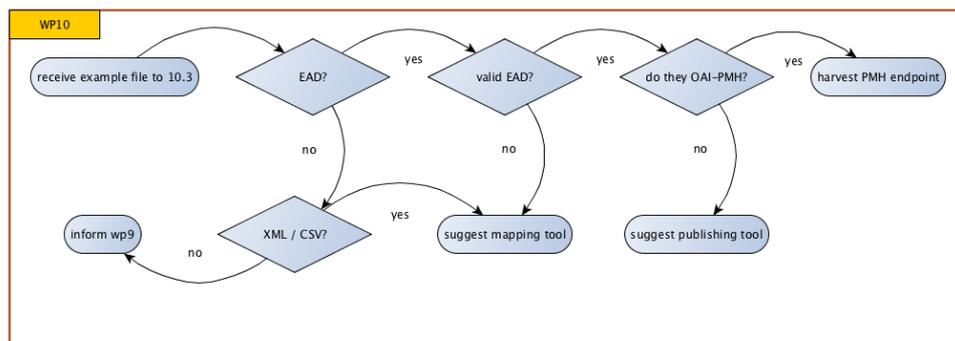


Figure 2: Workflow between WP9 and WP10 regarding data imports into the EHRI infrastructure (2) (EHRI, D9.4 Resource Reports, Update April 2017)

### *The EHRI EAD Mapping Tool and the EAD Publishing Tool*

For those who send in sample data that did not provide valid EAD, the EHRI project has developed an EAD Mapping Tool. The **EHRI Mapping Tool** allows for the mapping of local metadata fields to the international EAD standard. The tool can be installed by the institution itself or by the EHRI IT and will take the metadata of the institution, map them to valid EAD and convert consequently all metadata passing through this mapping tool. If they consequently have OAI-PMH (Open Archives Initiative – Protocol for Metadata Harvesting), they are then ready to share their data with the EHRI project, as seen in the figure below for Collection Holding Institution A. In case there is no OAI-PMH for EHRI to harvest the metadata, EHRI can assist further by the installation of an EAD publishing tool (Resource Sync). The **EHRI Metadata Publishing Tool** has been created to help archives to publish their metadata in a sustainable way (allowing for semi-automatic updates of descriptions in the EHRI Portal). Here as well, the institution can install the tool itself or with the help of EHRI IT. As soon as the program is installed and the metadata are being stored on a for EHRI accessible place on the institution's website, the sustainable connection to the project is a fact, as shown in the example of Collection Holding Institution B. Those able to provide valid EAD but without OAI-PMH can also install the EAD publishing tool in order to provide the EHRI project with their metadata in a sustainable way, as seen in the figure below for collection-holding institution C.

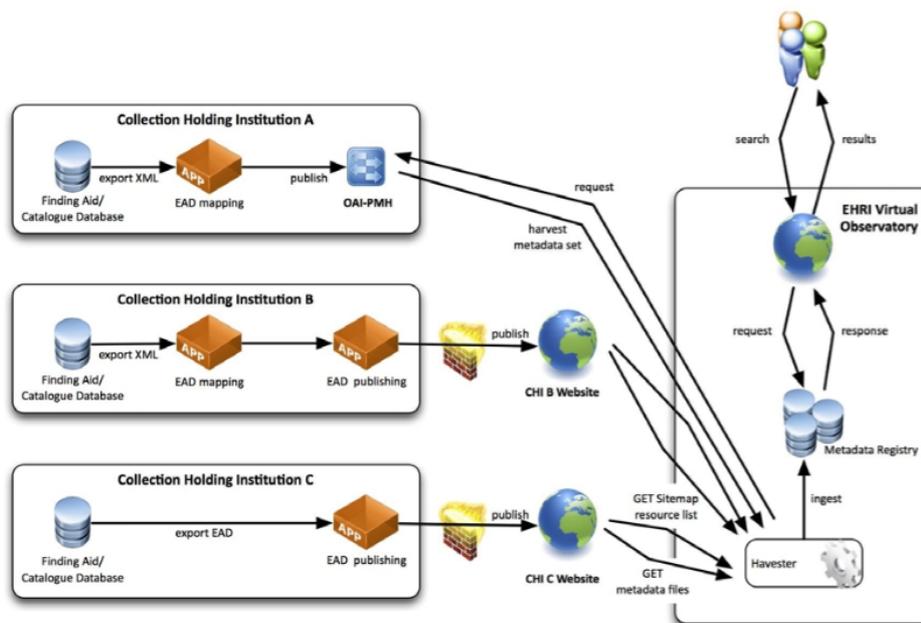


Figure 3: EHRI Data Infrastructure (EHRI D10.1 and D10.2 Collection description publishing services)

### ***EHRI Manual Data Entry and Follow-up***

In any case, the institutions preserving the documentation concerning the Holocaust are center stage in EHRI. It is not only those who have all the necessary knowhow and IT-tools available that are able to connect to the project. Also those institutions that are not yet having digital metadata or metadata in a format which would not be compatible with the use of the above-mentioned tools, are invited to share their metadata in the portal and make their metadata more openly available online. As already mentioned, where appropriate, archives can also be covered by **manual surveying and manual data entry**, either by EHRI staff or local experts, or by the collection-holding institution itself which in that case receives direct access to its own institution description in the EHRI portal and can – from there – add collection descriptions and child descriptions to its repository description. The screen shots below give an idea of how repositories, collections and child item descriptions can be created and updated within the EHRI portal. The extra tool behind the scenes is that every field from the ICA-standards, which form the basis for the forms, is explained when one clicks on the field itself. Moreover, all metadata entered in the portal can be exported in valid EAD to the respective institutions and as such, EHRI opens possibilities for further use of the data beyond the project itself.

The screenshot displays the 'Create Repository' form in the EHRI portal admin site. The form is organized into several sections:

- Identity Area:** Contains a text input for 'Identifier' (with a minimum length of 2), a dropdown for 'Language of Description', and a text input for 'Authorized Form Of Name'. Below these are radio buttons for 'Parallel Names' and 'Other Names'.
- Address Area:** Features a red 'Add Address' button.
- Description Area:** Includes three text areas: 'History', 'Geographical and Cultural Context', and 'Mandates/Incomes of Authority'.

On the right side, a vertical navigation menu lists the following areas: Identity Area, Address Area, Description Area, Access Area, Services Area, Control Area, and Administration. At the bottom of the form, there is a text box for 'Tell people what you're doing (optional, 400 characters max)' and two buttons: 'Create Repository' and 'Cancel'.

Figure 4: Illustration: Screen shots from the EHRI portal admin site

### *Open Source and Data Sharing Beyond the Project*

Because EHRI does not only wish for the metadata to be published in its own portal, but equally on the portal or website of the collection-holding institution or on other project's websites, EHRI has developed tools that allow for this, i.e. the mapping and publishing tool work equally for institutions to publish on their own website as well as share their data with other projects. The same is true for those institutions for which the data are manually added to the portal. EHRI can export the data back to the institutions. With some assistance from a web designer or basic explanation on how to create a website by yourself, the institution can further communicate about the data via its own ways of communication. Those without a website can go for a minimum scenario by integrating a link to their repository and its holdings in the EHRI portal in their email signature and spreading the news as such. EHRI furthermore provides tutorials and a helpdesk for each of the explained pathways to bring metadata into the EHRI portal. So, all together, the EHRI portal and the open source EHRI tools help archives not only to join the EHRI project, but equally to publish their own data themselves and to exchange data with other archives, memorials, projects and portals.

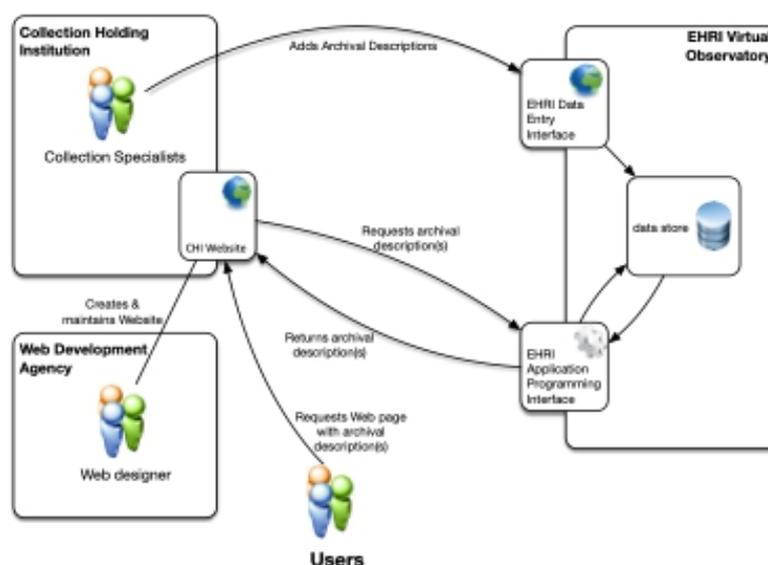


Figure 5: ADEMP (internal EHRI figure by Mike Priddy)

To stay informed about EHRI's activities and products, there are multiple options: the EHRI project website which includes links to all above named products (<https://www.ehri-project.eu>), the EHRI Facebook page (<https://www.facebook.com/EHRIproject/>), the EHRI newsletter (<https://ehri-project.eu/ehri-newsletter>) and the possibility to follow EHRI on Twitter ([@EHRIproject](https://twitter.com/EHRIproject)).

### **Acknowledgments**

Projects like EHRI are a group effort. The author would like to thank all her colleagues who contribute or have contributed to the EHRI project.