Reto Speck · Anna Ullrich

# IS DIGITAL BETTER? LESSONS LEARNED FROM BUILDING THE EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE (EHRI)

## 1. Introduction

The goal of the European Holocaust Research Infrastructure (EHRI) is to overcome the dispersal of archival material and knowledge on the Holocaust via dedicated offers to a European and international community of researchers, archivists, and – more generally speaking – people interested in learning about and preserving the history of the Holocaust. The EHRI was inaugurated in 2010 and has since been implemented via three EU-funded projects: EHRI-1 (2010-2015), EHRI-2 (2015-2019), and EHRI-3 (2020-2024). It brings together more than 20 of the most significant Holocaust archives, libraries, and research institutions in 17 countries. EHRI is an interesting case in the context of this volume because it prides itself on being both a digital infrastructure *and* a human network. How this works as well as the challenges and opportunities it precipitates form the focus of this article. First, we outline the activities and services of the EHRI, with special attention directed toward the EHRI Portal, one of the core (digital) offerings of the project. Second, we discuss two particular challenges, one regarding EHRI's data integration strategy, the other regarding more generally the question of digital and in-person activities. Finally, the paper details how the EHRI overcomes these challenges and which lessons we can learn from asking ›Is Digital Better?‹.

## 2. The European Holocaust Research Infrastructure (EHRI)

The need for the EHRI lies in the broad dispersion of Holocaust-related sources worldwide, a fact that has long challenged researchers of the topic. To date, the EHRI has identified more than 2,200 archives worldwide that hold Holocaust-relevant material. The EHRI's vision is to secure seamless access to all sources and expertise from across Europe and beyond needed to study the Holocaust and to set standards for excellence in transnational

Holocaust research, documentation, education, and remembrance. It continually develops and maintains a distributed, digital and human infrastructure providing a diverse community with access to archival resources and services. In doing so, it also acknowledges the relevance of Holocaust research for free and open societies with shared democratic values.

Before shifting the focus to the EHRI Portal, we would like to provide a brief overview of the tools and services – both digital and in-person – EHRI has so far developed to implement its vision. The EHRI's digital offerings are available via the Virtual Observatory, an expanding integrated online platform that provides free online access to (a) information about dispersed collection-holding institutions (CHIs) and their collections via the EHRI Online Portal and (b) scholarly digital editions of Holocaust sources via the EHRI Edition Platform.[1] Furthermore, it offers (c) visualization, enhancement, and contextualization of Holocaust sources via the EHRI Document Blog;[2] and (d) self-study and guided training on Holocaust sources and their interpretation via the EHRI Online Training Course.[3] The EHRI is continuously developing new services, with preparations for a Dashboard and a Massive Open Online Course (MOOC) currently underway.

As for in-person offerings, the EHRI's services include fellowships, seminars, and workshops. The Conny Kristel Fellowship Programme[4] is a competitive scheme that provides researchers, collection specialists, and other users transnational in-person access to the sources and expertise available at the most important institutions in Europe, Israel, and the United States holding Holocaust-relevant collections. Transnational training seminars and methodological workshops offer extensive training and networking opportunities to researchers, reference service specialists, educators, and other users at different career stages. The special emphasis in these offerings lies in exploring innovative digital methodologies.

Note that the services provided by the EHRI have generally been delivered either digitally or in person. Since the Covid-19 pandemic as well as the Russian war in Ukraine, however, it has become necessary to provide more flexibility, which resulted in transforming in-person meetings in some cases to online or hybrid events. While travel was restricted, some Kristel

---

1   EHRI Online Editions (https://www.ehri-project.eu/ehri-online-editions, accessed 8 May 2023).
2   EHRI Document Blog (https://blog.ehri-project.eu/, accessed 8 May 2023).
3   EHRI Online Course in Holocaust Studies (https://training.ehri-project.eu/, accessed 8 May 2023).
4   Conny Kristel Fellowship (https://www.ehri-project.eu/Conny-Kristel-Fellowships, accessed 8 May 2023).

Fellowships were offered remotely, depending on the virtual accessibilities of holdings at the respective institutions. Several workshops and seminars were also moved online since they otherwise could not have taken place at all. While the feedback for the specific events has been favorable, the EHRI is inclined to continue to offer some of its services as in-person events for various reasons – more about this below.

In addition to providing comprehensive research services, the EHRI also pursues a social mission. Through its dissemination and outreach activities, it plays a vital role in the fight against Holocaust denial and distortion, racism, and antisemitism, ensuring that Holocaust research keeps on informing a societal discourse on issues such as tolerance, cultural diversity, and human rights.

## 3. Virtual Integration of Dispersed Holocaust Sources: The Need for the EHRI Portal

The core service the EHRI developed to overcome the enormous fragmentation of Holocaust archives across Europe and beyond is the EHRI Portal, an online environment that provides an overview of archival institutions that hold Holocaust-relevant sources and descriptions of Holocaust-relevant materials. It is important to note that it does not offer the archival material itself but only descriptions thereof (metadata) and a range of tools to find, explore, organize, and share such information. Currently, the Portal includes about 400,000 archival descriptions and information on more than 2,200 Holocaust-relevant collection-holding institutions in 63 countries.

To find items in the Portal, one can apply a global search, which locates information across all the data content of the EHRI Portal. Alternatively, one can search within one of three predefined categories of information (countries, archival institutions, and archival descriptions) by selecting the appropriate category. A list of facets is available for both global and category searches. Facets allow narrowing the scope of the initial search or filtering the result lists. Available facets include, among others, timeframe, language, and the collection-holding institution.

To frame the archival information, the EHRI Portal contains country reports with concise per-country information on the respective country's Holocaust history, archival situation, and EHRI's research.[5] The country

---

5 Introduction to the EHRI Country Reports on Holocaust History and Archives (https://www.ehri-project.eu/country-reports, accessed 8 May 2023).

reports reveal a central challenge to a unified repository of archival information of the Holocaust: the large variety of archives and sources. Collections on the Holocaust are not just dispersed worldwide; they are also present in notably diverse archival institutions. Germany, for example, exhibits a highly differentiated archival system with little centralisation. Besides the Federal Archives and their departments, each federal state has its own state archive, and there are over 1,300 municipal archives. Further, there are archives of the various religious communities in Germany as well as universities, (larger) companies, noble houses, and associations. One could go so far as to say that it is almost impossible to find a contemporary history archive in Germany that does *not* contain material on the persecution of the Jews.

To some degree, this is true in most countries in which the Holocaust took place. Until 1989, the United States, Israel, and Western Europe were the main centers for Holocaust research. Auschwitz became the symbol for the Holocaust worldwide, because it was the largest death camp and the camp where Jews from Western and Central Europe were murdered. More specific collections have been set up in many regional centres on research and commemoration in the past decades. The opening up of archives in Eastern Europe, in particular in Eastern Germany, and the opening of formerly classified archives in Western Europe resulted in a substantial increase in the available source material. At the same time, the number of institutions in European countries that hold Holocaust-related collections and are active in research and commemoration has increased since 1989 (especially in Eastern Europe but also in Germany and most other European states). These institutions, old and new, have their own collections and their own (increasingly digital) archival infrastructures, which often do not support scholarly requirements. Different institutions use their own distinct systems and different metadata schemas. Many different languages are contained in the original documents and in catalogues, necessitating translation and hampering comparability.

The vast majority of Holocaust victims lived and were murdered in Eastern Europe. It is still far more difficult to conduct research and documentation on this part of Europe than in Germany or Western Europe. Finally, one of the major challenges for every scholar of the Holocaust is dealing appropriately with the prevalence of perpetrators' sources to avoid muting the voices of the persecuted Jews. The documents of Jewish organizations or relief organizations often followed the fate of their owners; in many cases, they were destroyed or dispersed. For instance, to gain insight into the activity of the Jewish refugee organisations in Prague in the 1930s, a researcher would have to study the fragments of reports saved in several archives, especially in the United States, Israel, the Czech Republic, and Germany. While there

are numerous testimonies given by Holocaust survivors after the liberation, original diaries, letters, or testimonies from the time of persecution are more difficult to find. Over the last years, a growing consensus has emerged in Holocaust historiography that Jewish sources and views must be more integrated into the narrative(s) of the Holocaust.

The fragmentation of archival sources results not only from the wide geographic scope of the Holocaust but also from the Nazi attempts to destroy the evidence. Moreover, and closely related to the geographical scope, Holocaust sources were written in many different languages, whereby the language of the documents is not necessarily the same as the cataloguing and description language, thus further complicating the picture. Furthermore, Holocaust survivors migrated to places across the globe, taking their documentation with them.

A plethora of documentation projects developed after the Second World War, which the example of Hans Günther Adler illustrates. Adler was a Czech Jew born in Prague and imprisoned in Theresienstadt, Auschwitz, and Buchenwald during the war. Following his liberation, he worked at the Jewish Museum in Prague before emigrating to the United Kingdom in 1947. A prolific writer throughout his life, Adler's works and letters –both the original manuscripts and copies – are distributed in many different institutions throughout Europe and beyond, including the EHRI partners Jewish Museum Prague, the NIOD Institute for War, Holocaust and Genocide Studies in Amsterdam, the Arolsen Archives in Bad Arolsen, King's College London, and Yad Vashem in Jerusalem. Because of this wide distribution, it is difficult for contemporary researchers to gain a coherent overview of the output of this important figure in Holocaust scholarship.[6]

While the identification of Holocaust-related sources continues apace and an increasing amount of sources is becoming available digitally, many ›hidden‹ collections still exist that are currently not widely employed in research. Collections may be ›hidden‹ because they are only insufficiently described or because they are not easily obtainable in a suitable form. In the context of the increasing application of digital methods to Holocaust sources, this situation may become exacerbated in the sense that (digital) research may increasingly focus on a minority of sources simply because they are already machine-readable – even if these sources are not the most suitable ones for addressing the research question being investigated. Avoiding this demands continued funding for source description and source digitisation.

---

6 Mike Bryant et.al.: The EHRI Project – Virtual Collections Revisited, in: Luca Aiello, David McFarland (eds.) Social Informatics. Basel 2014, p. 300.

Particularly important regarding Holocaust studies is making funding and expertise available to smaller collection-holding institutions and to institutions in regions with limited access to larger infrastructures (e.g., Eastern Europe or the Balkans), to enable them to become fully ›plugged into‹ the evolving digital landscape of Holocaust studies. Similarly, the willingness and ability of collection-holding institutions to make their (digitised) collections and related resources (metadata, thesauri, and other knowledge organisation frameworks) openly available in bulk to (digital) research are uneven. We need increased advocacy work that aims to demonstrate to collection holders the full benefits of openness and sharing and the availability of guidelines/templates to regulate such sharing. Consequently, Holocaust-relevant documentation can today be found in many notably diverse collection-holding institutions, spread across a vast geographic area – often in very surprising locations.

## 4. Challenges Encountered and Lessons Learned

While outlining the need for the EHRI Portal above, we became aware of one of the main challenges: How to develop standardized procedures to integrate the description of data (›metadata‹) from collection-holding institutions so diverse in their geographies, languages, levels of maturity and available collections?

We developed several methods to integrate collection metadata into the EHRI Portal: The *Manual Data Entry* method allows collection-holding institutions to directly enter collection descriptions into the EHRI Portal using an administrative account. The *Batch Import* method enables EHRI to import collection descriptions in bulk, preferably in the standardised Encoded Archival Description (EAD) format. The ideal scenario is integration via the *Repeatable Batch Imports* method, which allows EHRI to automatically download and import the collection metadata from a web-accessible location, regularly repeating the process to ensure that the descriptions are up-to-date.

Automatic data integration involves institutions to share their archival metadata in a structured format – typically XML with a standard schema exported from their internal IT systems – and to make the integration of larger, more complex collections possible with considerably less effort than manual integration would entail. EHRI-2, which ran from 2015 until 2019, had two main means of sustainable integration of metadata to the EHRI Portal: In the first scenario, EAD files were published through an OAI-PMH endpoint (provided the institution had an endpoint available that serves EAD files).

In this case, the connection between the institution and the EHRI Portal would be straightforward and, more importantly, sustainable, continuously integrating updates. The second scenario came into play when collection data could be exported as XML. The conversion to EAD was then performed by the data provider using the EAD Conversion Tool (ECT) and ResourceSync manifests generated using the Metadata Publishing Tool (MPT). The resulting manifests and EAD files were then hosted on a standard HTTP server from which they were harvested by EHRI's ResourceSync aggregator tool.[7]

While the two scenarios described above allowed the integration of many sources and the establishment of many sustainable connections with the EHRI Portal, it nevertheless became apparent that the second scenario was insufficient for some use cases and unemployable for others. Most notably, it expected data providers to use the EAD Conversion Tool – which depended on available time and expertise that was often not feasible at the collection-holding institutions. These issues resulted in the idea of creating an EHRI Mobile Data Integration Lab, which would be responsible for helping the institutions by teaching them the integration – or even performing it on their behalf. This plan, however, soon ran into issues of its own: The outbreak of the Covid-19 pandemic made it impossible to realize the ›Mobile‹ aspect of the Data Integration Lab as EHRI's data-integration experts had to stay put and exchange mapping rules and other necessary information and input mainly via email, making close guidance of the processes very tedious. In addition, software updates and patching were complicated, as an institution's staff had to download, set up, and run the latest version of the relevant tools to ensure everything worked as expected.

Subsequently, we decided that the data mapping tools would be delivered as a software service hosted by the EHRI itself. At the beginning of the EHRI-3 project, key parts of the data-integration process were integrated into the EHRI Portal so that, now, instead of operating offline as desktop or command-line applications, they could be accessed through the Portal's online admin interface. The main change concerning the workflow is that the data transformation is executed *after* the metadata harvesting. This results in a shared environment for all the data-integration stakeholders, while also relieving institutions from the need to implement the complex transformation process themselves, albeit using EHRI tools.

A brief look at the ETL (Extract, Transform, Load) processes shows how the new workflow made it easier for institutions to share their metadata –

---

7  Veerle Vanden Daelen et.al.: Interim Report on Data Integration (confidential deliverable), 2022, p. 8.

and for EHRI to integrate it into the Portal. In the ›Extraction Phase,‹ it is now possible to receive the data in different formats, whereas previously the institution had to provide their data already in the EAD format. This significantly lowers the threshold to be represented in the Portal, especially for smaller institutions that cannot convert their metadata into EAD (and who sometimes provide an overview of their holdings only in Excel).

Any necessary data transformations are performed in the ›Transform Phase.‹ In EHRI-2, each institution had to deal with this phase offline, eventually delivering EAD-conformant files. Now the data transformations are hosted in the EHRI Portal, allowing users to build EAD files from arbitrary XML formats. At the same time, it makes it easier for the data-integration specialists within EHRI to provide guidance and support immediately through the administrative account. This new approach allows for a very flexible and dynamic environment that can adapt input data to the EHRI EAD data model in a very fine-grained manner. In the ›Load phase,‹ previously transformed data are imported and stored in the EHRI Portal, still allowing for small(er) customisations.[8]

The new datalab workflow hosted within the EHRI Portal not only solved the problem of software updates (as now everybody is using the same version), the need to package and install desktop applications, and the exchange of mapping rules (which are hosted online); it generally makes it easier for the collection-holding institutions to share their data, independent of their digital maturity. For our data-integration specialists, it is a chance to closely guide the data standardization central and directly within the EHRI Portal. The workflow adaption can be considered a central ›lesson learned‹ for the digital advancement of the EHRI Portal – and the entire project.

In addition, the new data-integration pipeline has made it much easier for collection-holding institutions to integrate metadata into the Portal, lowering the technical hurdles they need to overcome. However, we know that, for very small holding institutions such as private collectors and community archives, integrating their metadata into EHRI can still be daunting and sometimes exceed their capabilities. Because such ›micro-archives‹ often hold very important Holocaust-relevant collections, and to ensure that these holdings are increasingly plugged into the global research landscape, we are currently investigating how to bridge the gap between microarchives with very low digital maturity and EHRI's digital infrastructure.

From the beginning of the EHRI project, the idea was to improve access to information on Holocaust-related archival material *and* connect people

---

8  Ibid., pp. 12-13.

in Holocaust research, preservation, education, and commemoration. In this way, even the EHRI Portal, one of EHRI's core digital offerings, should eventually be used to enhance the exchange between researchers and archivists and to encourage researchers to explore collections not (yet) digitised but crucial to advancing their research.

A recent survey conducted by the EHRI to identify user needs and understand the access challenges users face confirmed this.[9] The two main challenges the respondents identified were online access to archival descriptions (49 percent) and archival material (83 percent); 33 percent named fees associated with acquiring digital scans as a challenge. Additional challenges included contacting the archives, where 24 percent of respondents perceived finding the contact information for archives challenging, while 25 percent noted the difficulty in reaching out to reference and information services of archives. 34 percent of the respondents considered the usability of the archive's website a challenge. Lastly, 18 percent experienced language barriers as a challenge. While the demand for access to online archival material remains high, it is notable that, in a follow-up question on the importance of »the role of the archivist in your research,« 58 percent deemed the role as »extremely important« and 38 percent as »somewhat important.« The questions related to access demonstrate the critical need for access to archival descriptions among users – and this among all user groups. However, it is important to consider the question regarding the role of the archivist when evaluating the use of specific services. While access to descriptions and online material is important to the user, help from specialists within the field, such as archivists, is also vital.[10] EHRI solves these seemingly contradictive demands by making detailed descriptions and contact possibilities for archives

---

9  The survey was conducted as part of the Work Package »User, Access and Training Strategy« within the EHRI Preparatory Phase (EHRI-PP). It was shared widely on the EHRI website and other social media platforms. A total of 129 responses from 26 different countries were received, most of which came from Western Europe (31 percent), North America (24 percent), Israel (14 percent), and Eastern Europe (13 percent). The vast majority of respondents were researchers, 88 percent of respondents identified themselves within the academic research field (Ph.D., postdoc, lecturer, independent researcher), 11 percent were identified as archivists, 10 percent as educators, 4 percent librarians, 3 percent journalist/film industry, 3 percent M.A. level students, one respondent was at the B.A. level. See Emmanuel Moscovitz et al., User Needs Analysis (confidential deliverable), pp. 24-25.

10  See also Reto Speck and Petra Links: The Missing Voice: Archivists and Infrastructure for Humanities Research, in: International Journal of Arts and Humanities Computing, 7 (1-2), pp. 128-146.

digitally available via the Portal, while at the same time facilitating on-site access through the Conny Kristel Fellowships.

The answer to the question of whether participants would prefer an in-person or digital setting for participating in conferences, seminars, workshops, or fellowships further underscored the need for continued in-person events. Most respondents indicated a preference for a physical setting for all services listed: 58 percent for conferences and seminars, 62 percent for workshops, and 69 percent for fellowships. While the virtual setting has proved advantageous in certain contexts (easier to reach a wider audience, lower costs), most users would prefer to participate in the different activities in a physical setting. Asked about the main disadvantages of offering such services online, 83 percent of the respondents considered the lack of face-to-face communication and networking possibilities the most disadvantageous. 16 percent indicated fearing a lack of self-discipline needed for online programs.[11]

In this way, the survey confirmed the lesson learned through years of offering EHRI services – in-person and digitally: It is desirable to make as much information on archives, their metadata, and, where possible, the documents themselves available digitally. However, allowing a community to grow, network, engage, and gain expertise will be possible only through in-person events.

## 5. Conclusion – EHRI: A Digital Infrastructure *and* a Human Network

The digital transformation of archives is a powerful enabler for EHRI. On the one hand, the ability to virtually unite what is physically dispersed is a key affordance of the digital age. The EHRI Online Portal is built around this affordance by providing unified access. At the same time, digital approaches to research also have significant potential to generate new insights and disseminate these in novel ways.

However, going digital is no panacea to all challenges, and the shift from (mainly) analogue modes of access and research to predominately digital ones must be handled with care. For instance, there always lurks the danger of widening existing, or opening up new, ›digital divides‹ with the associated risk that researchers focus predominately on sources that are conveniently (digitally) available rather than the ones best suited to answering their research questions. To alleviate this, we need continued investment in the

---

11  Emmanuel Moscovitz et al., User Needs Analysis, p. 34

digitisation of archives in general, and in the identification, description, and integration of currently ›hidden‹ microarchives in particular. Furthermore, archivists and other collection specialists have traditionally been important interlocutors for researchers and purveyors of extensive, informal knowledge about the sources under their custodianship. It is vitally important not to lose this source of knowledge in the move towards remote instead than on-site access to archives. EHRI solves this by being both, a digital infrastructure *and* a human network, by carefully and continuously evaluating the field of Holocaust research, the tasks at hand, the needs of the community, and the resources available.