



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

GA no. 871111

Deliverable 11.3

Implementation of EHRI services for micro-archives (in iterations)

**Maria Dermentzi, Mike Bryant, Rachel Pistol
KCL**

**Sigal Arie Erez, Giora Zwilling
Arolsen Archives**

Start: March 2022 [M19]

Due: August 2023 [M36]

Actual: August 2023 [M36]



EHRI is funded by the European Union

Document Information

Project URL	www.ehri-project.eu
Document URL	https://www.ehri-project.eu/deliverables-ehri-3-2020-2024
Deliverable	D11.3 Implementation of EHRI services for micro-archives (in iterations)
Work Package	WP11
Lead Beneficiary	15 - ITS
Relevant Milestones	MS3
Dissemination level	Public
Contact Person	Maria Dermentzi, maria.1.dermentzi@kcl.ac.uk
Abstract (for dissemination)	<p>This DL provides an overview of the work undertaken by T11.2, which aims to develop new services that will allow micro-archival communities to open up, curate and preserve their collections. The report begins by detailing the activities that were carried out to specify the user requirements for the services that EHRI could develop to facilitate access to micro-archival records for the international Holocaust research community. Furthermore, this DL introduces the Micro Archive Publication Tool (MAPT), an experimental prototype that simplifies the process of publishing image-based micro-archives in a sustainable, user-friendly and standards-compliant manner. The DL concludes by summarising lessons learned and outlining future work directions.</p> <p>[Max. 150 words]</p>
Management Summary	[Max. 500 words]

Glossary

AI	Artificial Intelligence
API	Application Programming Interface
ASR	Automatic Speech Recognition
AWS	Amazon Web Services
DNS	Domain Name System
EAD	Encoded Archival Description
HTML	Hypertext Markup Language
ICA	International Council on Archives
IIIF	International Image Interoperability Framework
ISAD(G)	General International Standard for Archival Description
MAPT	Micro Archive Publication Tool
OCR	Optical Character Recognition
URL	Uniform Resource Locator
XML	Extensible Markup Language

Table of Contents

1	Introduction	5
2	Service Requirements as Specified by T11.1 in D11.2.....	6
3	Service Requirement Specification Work in T11.2.....	7
4	Micro-Archive Publication Tool (MAPT)	9
4.1	Overview	9
4.2	Limitations & assumptions	9
4.3	Using MAPT	10
4.4	Describing the collection	11
4.5	Publishing the data	16
4.6	The output: a published website	16
4.7	How the viewer works	17
4.8	Harvesting MAPT-generated EAD metadata	18
4.9	Limitations and future work	18
5	Lessons Learned.....	19
6	Conclusion	20
7	References.....	21

1 Introduction

The purpose of this deliverable is to report on the work undertaken by T11.2. T11.2 seeks to develop services that will allow micro-archival communities to open up their collections for use by the international Holocaust research community.

In the context of WP11, a micro-archive is broadly defined as a family archive, an association, a memorial, a grassroots initiative, a researcher's archive, or a very small archival institution, which is not run by local authorities or by the state. In early 2021, T11.1 prepared a call for micro-archives that was published on the EHRI website, newsletter and social media platforms and was widely shared via EHRI members' channels and networks. This outreach was further supported by extensive desk research which led to over 200 potential micro-archives being contacted about the call. The purpose of the call was to survey potential micro-archives on the contents and condition of their collections as well as on potential needs in terms of new EHRI services and technological tools that would frame the work of T11.2. The response to the call was much smaller than expected (only around 20 to 25 micro-archives got in touch). This was attributed to various factors, including the lack of in-person contact that would encourage representatives of smaller collections to engage with EHRI given the fact that these efforts took place mostly during the COVID-19 pandemic (Erdösi et al., 2022).

Another possible reason for the small number of responses was that the questionnaire on which the survey was based may have been perceived as quite technical from the perspective of family micro-archive representatives. People in possession of micro-archives are often not acquainted with archival practices and standards and their familiarity with digital technology might be limited. As a result, they may find it hard to articulate specific needs and requirements due to a lack of familiarity with professional archival practices. Instead, it was suggested that it would be more efficient to ask micro-archive owners for their feedback and reactions to tools that were already available (Erdösi et al., 2022).

Having taken into careful consideration the output of the survey carried out by T11.1, this task group agreed on a twofold strategy. First, T11.2 sought to further explore and refine potential user requirements for new EHRI tools and services that would support micro-archives in cataloguing, curating and/or preserving their collections. For this reason, several video-conferencing meetings with other WP members were held to discuss potential services targeted at micro-archives. Additionally, T11.2 organised a WP11 engagement workshop in partnership with the Wiener Holocaust Library in London which took place in May 2023, where a focus group was organised to further assess micro-archives' requirements in terms of services and expertise. Second, a prototype of the Micro Archive Publication Tool (MAPT) was developed so that we could demonstrate a concrete example of a potential service to be offered by EHRI while probing deeper into micro-archives' expectations and needs. MAPT was designed as a first prototype of a potential future EHRI service that would simplify the process of publishing image-based micro-archives in a user friendly yet standards-compliant way, whilst also helping us better understand how the needs of micro-archives in terms of metadata provision and description differed from—or aligned with—other archival collections.

In what follows, Section 2 will review the suggestions for prospective services for micro-archival communities as detailed in the *Report on micro-archival communities and user requirements (Deliverable 11.2)* (Erdösi et al., 2022). Section 3 will touch upon further user requirement specification work that took place in T11.2. Section 4 will describe the MAPT in detail. Section 5 will critically reflect on lessons learned from this task. Section 6 will serve as a conclusion and suggest future work.

2 Service Requirements as Specified by T11.1 in D11.2

In its efforts to frame the work of WP11, T11.1 designed a comprehensive questionnaire to survey micro-archive owners on the contents and condition of their collections and potential support they require in terms of expert advice and tools. Details on the challenges faced by T11.1 and the blank questionnaire itself can be found within D11.2 (Erdösi et al., 2022). The survey helped EHRI to refine its definition of what a micro-archive is and to better understand the ways in which it could support micro-archives of different sizes and resources.

According to D11.2, the levels of knowledge of archival practices and digital literacy varied greatly among the questionnaire respondents, as did their access to nearby larger institutions that could assist them in cataloguing, indexing, digitising, curating, or preserving their materials. This variation makes it challenging to formulate “one-size-fits-all” solutions for services that would support or facilitate the work of micro-archives in general, irrespective of their context and location (Erdösi et al., 2022). One part of the micro-archival community does not necessarily require access to theoretical advice and tools by EHRI because it is already receiving the support that it needs from larger local institutions. Others require financial support to outsource certain activities such as the digitisation of their collections due to lack of time and resources to do it in-house. There are also cases in which in-person support and guidance by representatives from institutions specialised in archival standards would be more beneficial. This is because most of the materials held by micro-archives have either not been catalogued at all or, even if they have, they have not been catalogued in a way that complies with archival standards, which is an issue when trying to make these micro-archives more accessible to a wider audience.

Among the proposals for new services that were listed in D11.2, some were more actionable for T11.2 than others. Starting with the more actionable proposals that were put forward, one of them had to do with offering a way for micro-archives that do not already have a web presence to acquire one through EHRI. It is worth noting that when a holding institution is added to the EHRI Portal, an institutional page is created about it with all the basic details about its location, history, sources etc, while the archival descriptions of its collection are also presented on the same page and are additionally searchable through it. To the extent possible, this information is also compliant with ICA standards although this task group understands that more can be done to ensure compliance for already added institutions (small and large) as well as future ones. Another actionable proposal was to offer digitisation services for micro-archives that are not yet ready to hand over their collections to other institutions, which is often the case with family archives. WP4 is already offering digitisation services in certain regions of focus, however this task could investigate ways to scale this. As described in the following sections, these two proposals were carefully considered and, to the extent possible, addressed. There was also the suggestion for EHRI to offer a service to digitally preserve the websites of micro-archives. However, similar services of this kind already exist, with the most notable one being the Internet Archive’s Wayback Machine (<https://web.archive.org/>) and therefore this was not deemed as a priority by this task group.

Other service proposals included the idea to enable micro-archives to publish posts about their material in the EHRI Document Blog, encouraging further visibility and recognition of their work. This is already possible and there is a page with guidance on how interested parties may contribute to the EHRI Document Blog: <https://blog.ehri-project.eu/about/how-to-contribute-to-the-ehri-document-blog/>. However, T11.1 and T11.2 decided that this service could be improved by creating a new category of posts specifically for micro-archives, allowing users to filter the Document Blog’s content based on whether a post is related to a micro-archive or not. For this reason, we suggest that from now on posts relating to micro-archives will carry a dedicated “micro-archive” category tag. Another idea was to include micro-archives’ activities in EHRI newsletters. The EHRI newsletter frequently features news and calls for papers that are not directly related to EHRI activities but might be related to one of our partners’ activities

and this could include micro-archives. That said, since more and more micro-archives are now getting indexed by the EHRI Portal, steps will be taken in cooperation with other work packages to discuss updating the newsletter template by adding a text that will explicitly invite partners to update us about any work/exhibitions that they want us to highlight in the next newsletter. Another idea could be to add a "Micro-Archive Spotlight" section that will highlight a different micro-archive collection each time a newsletter is sent. Overall, T11.2 maintains that both the EHRI Document Blog and the EHRI Newsletter as already existing EHRI services are ready to support these two requests, but a continued outreach is necessary to make sure that micro-archival communities know what services are available to them.

Some of the proposals in D11.2 were deemed to be outside the immediate scope of this task but could be considered as potential EHRI services to be offered by the future EHRI-ERIC National Nodes or the Central Hub. Such proposals included scholarship or internship programs that would allow students and researchers to visit micro-archives in situ and to offer practical support over extended periods of time. The proposal to bring micro-archives in contact with professional archives of their respective regions is also something that the National Nodes will be better suited to pursue since they will have a better overview of local micro-archives as well as of the capacities of local institutions to offer targeted support. The same is true for the proposal to find ways to include objects and documents from micro-archives in exhibitions of local memorials, museums, universities, and similar institutions. The proposal to consider offering legal advice and support regarding copyright and other legal matters also falls within this category of ideas to be explored in the scope of EHRI as a permanent research infrastructure or as a service offered by the National Nodes and it is already being discussed within the consortium and to a limited degree already offered by WP4 in collaboration with WP3.

The overall conclusion of D11.2 is that for micro-archives to be compliant with professional archival practices and standards, actual guidance is needed more than support in the form of services or tools. D11.2 also suggests that instead of creating new services after surveying micro-archives about their needs, local EHRI experts should be involved in the development of services in collaboration with other work packages which could then be tailored in WP11 to better suit micro-archives.

3 Service Requirement Specification Work in T11.2

During the course of T11.2, many steps were taken to refine the proposals for future EHRI services for micro-archival communities. In addition to the several WP11 meetings that were set up to discuss potential paths, task leader KCL organised a micro-archive engagement workshop in partnership with the Wiener Holocaust Library (WHL) in London, which has a rich experience of working with collection owners that would fall within EHRI's micro-archive definition. The workshop, which took place at the WHL in London in May 2023, attracted professional and non-professional archivists, researchers, and curators from different regions of the UK as well as from Germany and Israel. The workshop participants represented a diverse mix of archival institutions, from extremely small to those of a moderate size. All the participants deal with micro-archives on a regular basis and are familiar with the challenges of such material.

The workshop began with an introduction to the EHRI project, WP11 and EHRI's activities regarding micro-archival collections. Sandra Lipner, co-curator of WHL's exhibition *Holocaust Letters*, which was developed from micro-archival sources, led a tour of the exhibition, and discussed the process of putting it together, highlighting the challenges but, most importantly, the value and richness of these types of collections. Having seen an example of how micro-archival collections can be used by researchers and archivists once made more accessible, workshop attendees were introduced to the way the WHL is digitising documents while touring

the UK as part of their bespoke programme of family research and personal document collection preservation support, the *Recovery & Repair Project*.

A demonstration of WHL's portable digitisation kit was given. As already mentioned, often micro-archive owners are not ready to donate their collections and, thus, the only way for larger institutions to obtain copies, allow access to researchers, and ensure the preservation of the material is by first digitising it. Workshop attendees found this demonstration very relevant to their work since using similar kits to generate digital surrogates in situ could help with their mission to preserve micro-archival material in their respective regions. Digitisation is a very important step towards making a collection more accessible and it was also recommended in D11.2 as a potential EHRI service so T11.2 made it a priority to include this session in the workshop's agenda.

Subsequently, KCL conducted a demonstration of EHRI's Micro-Archive Publication Tool (MAPT). As discussed in the introductory section, micro-archives are not necessarily familiar with archival practices and tools, and it is challenging for them to delineate their requirements for new EHRI services without having seen some examples of what EHRI could offer them. Therefore, creating a tangible prototype for a service that could be showcased to micro-archive owners before soliciting further feedback from them was deemed necessary. To this end, KCL built MAPT, an experimental prototype that simplifies the process of publishing image-based micro-archives (MAPT is described in detail in Section 4). Presenting MAPT to the micro-archival community during the engagement workshop in London made it possible to receive more targeted feedback and helped us specify user requirements more precisely.

The workshop concluded with a focus group discussion on the potential usefulness of MAPT as well as other digital tools that could be deployed by EHRI to support micro-archive owners or larger institutions working with them. Details on the output of the focus group discussion are explained in Section 5.

T11.2 and potential future services were also discussed during the EHRI-3 General Partner meeting in Jerusalem in early June 2023 where T11.2 members had the opportunity to get updated with how the micro-archive engagement work is progressing in other EHRI-3 tasks. Specifically, having discussed some of the challenges faced by colleagues who work with micro-archives, a new service proposal emerged that had to do with designing and disseminating document preservation guidelines. Archivists within the EHRI network have recently noticed that non-professional micro-archive owners, in their efforts to protect their material, might accidentally harm it and lead to its faster deterioration. For example, there have been cases where documents have been laminated to protect them from damage. Such practices can be catastrophic for the material and EHRI could use its platform to raise awareness about this issue. WP11 members also agreed to keep showcasing and getting feedback on MAPT during future WP11 engagement workshops that are due to take place in late 2023/early 2024.

As envisioned by the task's description, the development of new services for micro-archival communities is taking place following a collaborative and iterative process. Since MAPT is still in a testing stage, this work is ongoing. More WP11 partners are planning to test MAPT during forthcoming engagement workshops and it is possible that service specifications might change as a result.

4 Micro-Archive Publication Tool (MAPT)

For the purposes of T11.2, KCL created a prototype for a tool that simplifies the process of publishing image-based micro-archives in a sustainable and standards-compliant manner.

4.1 Overview

This tool is intended to address the use case of a micro-archive where the holders wish to retain custody of the material, but also want to make it accessible in digitised form through a custom website, with standards-compliant harvestable metadata that can easily get ingested by the EHRI Portal and other aggregators.

Notwithstanding the complexities of digitisation itself (image capture, photographing, choosing appropriate formats etc) once images have been captured the process of making it available is not a straightforward one. Challenges include:

- Micro-archives will typically have some inherent structure, required to be reflected in how they are made available online.
- Generating metadata can be a complex process, involving numerous technical decisions and an understanding of the relevant conceptual standards.
- Images produced for the capture of physical material are often uncompressed and of higher quality than is practical to distribute online, and in many cases are not able to be distributed online due to rights concerns, including the custodian wishing to retain control of the source digital proxies.
- Any online publication involves the use of web hosting or integration with a very wide range of possible hosting services.

MAPT is a proof-of-concept for streamlining this process by:

- Providing a drastically simplified “wizard”-like interface to generating standards-compliant XML metadata that can get harvested by aggregators such as EHRI.
- Generating [IIIF](#) manifests that facilitate browsing the micro-archive as a structured artefact in any standards-compliant IIIF viewer software, enhancing its interoperability.
- Generating a simple but functional HTML page.
- Publishing the HTML, XML, and IIIF resources to Cloud-based web hosting that can be associated with any domain name and is easily replicable or re-hostable elsewhere.

4.2 Limitations & assumptions

MAPT is a **proof of concept** and as such has numerous technical and usability limitations. It has been created using the [Streamlit](#) platform which provides a convenient and versatile environment for the rapid prototyping and deployment of web-based applications and is already used by EHRI-3 in other capacities. A trade-off in using Streamlit is that the platform is not well suited to use-cases involving the processing of large binary datasets; however, for prototyping it was considered sufficient.

At present, MAPT also makes a number of technical assumptions about the nature of the material of which the micro-archive is composed. It assumes that:

- The micro-archive consists of material that has been digitally scanned or photographed.

- These digital scans or photographs have been organised in a file/folder structure reflecting the organisation of the source material.
- The material can be uploaded to a private Cloud-based storage location.

4.3 Using MAPT

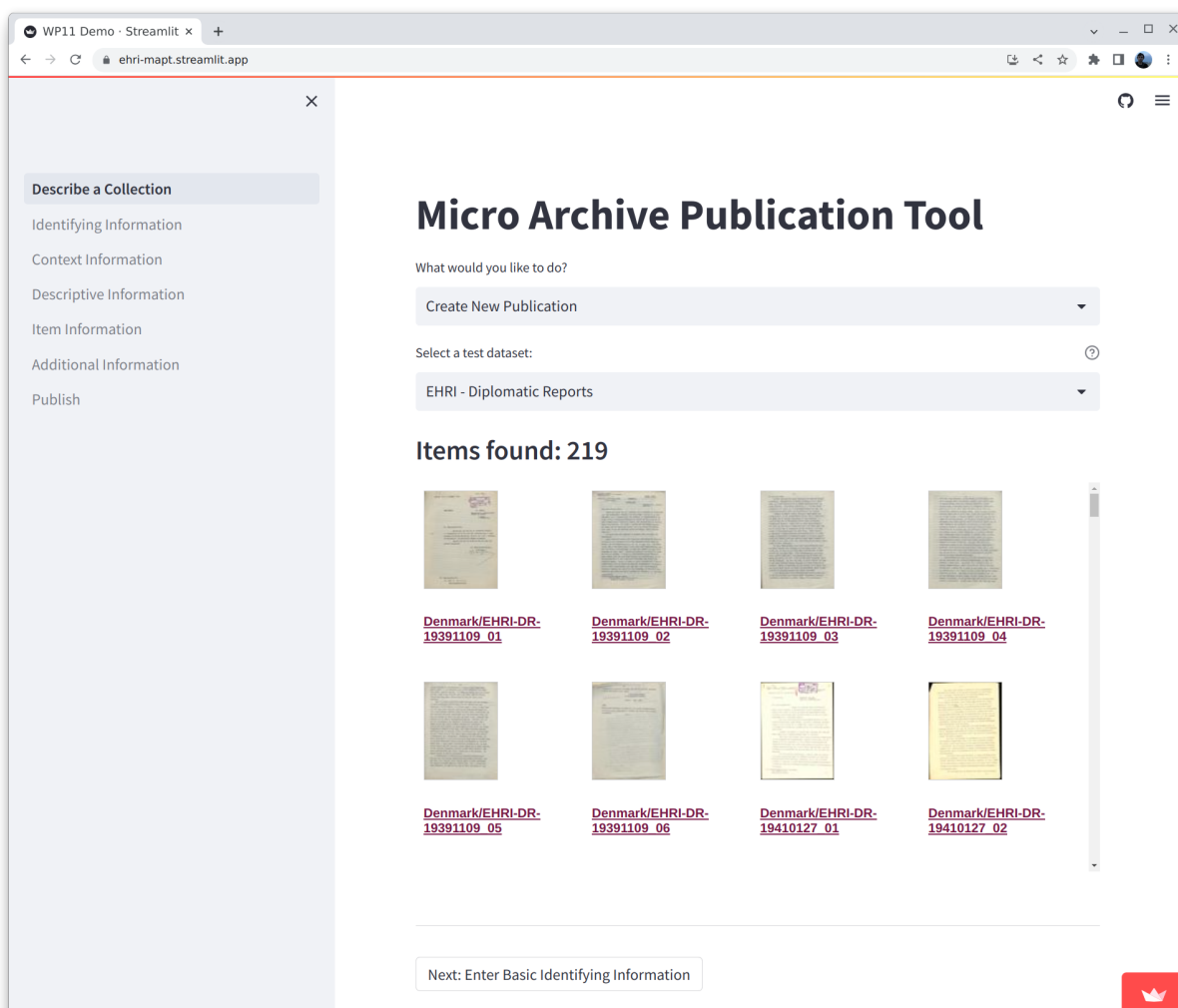
The tool is currently hosted online by Streamlit at:

<https://ehri-mapt.streamlit.app/>

At present, an example dataset has been selected via configuration. This demonstration data is sourced from either the [EHRI Diplomatic Reports](#) or [EHRI Early Testimony](#) online editions.

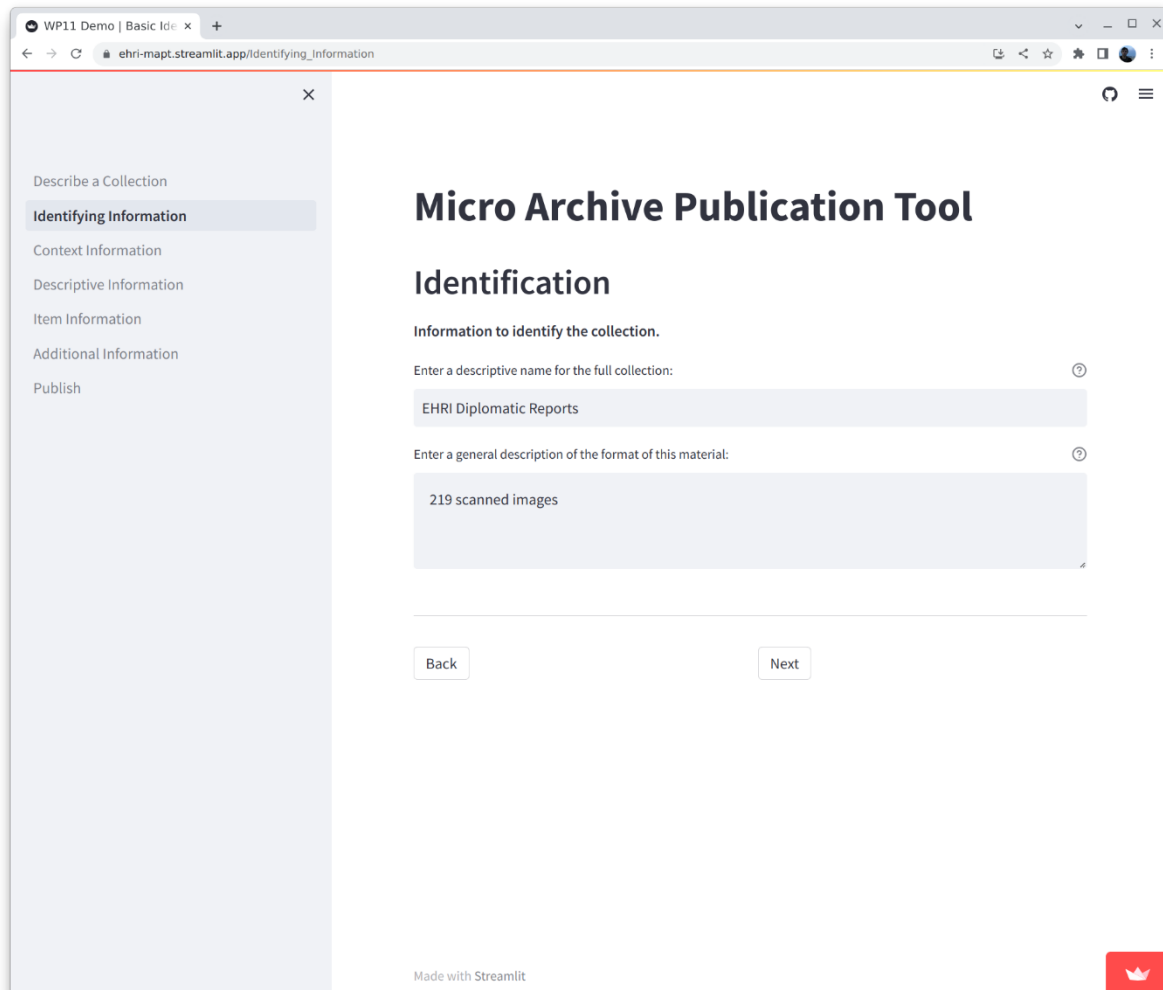
To add additional datasets requires uploading images, along with the file/folder structure in which they are placed, to a specific location on an EHRI-managed Cloud storage drive. This data is **not accessible to the public**. This location, and the credentials for accessing the data, are provided to MAPT via configuration and can be changed by the app administrator.

The home page of the application provides a selector for currently configured datasets, or an option to edit a previously created site. When a dataset is selected a preview of its files is shown:



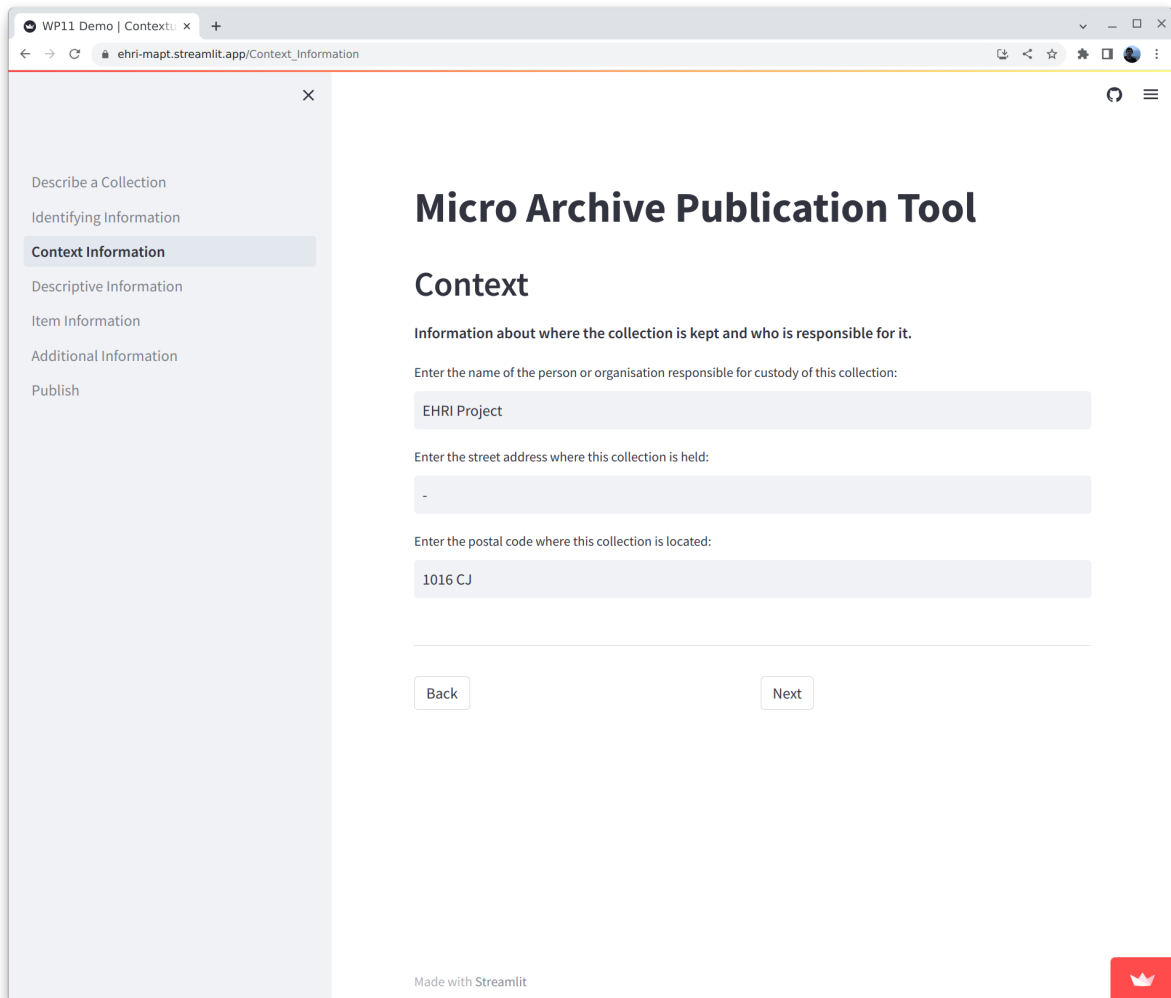
4.4 Describing the collection

Currently, the metadata fields that are provided are intended as a starting point only that could be developed further in future for different use-cases, with each page corresponding to a subsection of the General International Standards for Archival Description (ISAD(G)). The first page contains minimal identifying information, including the archive's title and the general extent and medium (which is pre-filled by the number of images present in the source dataset):



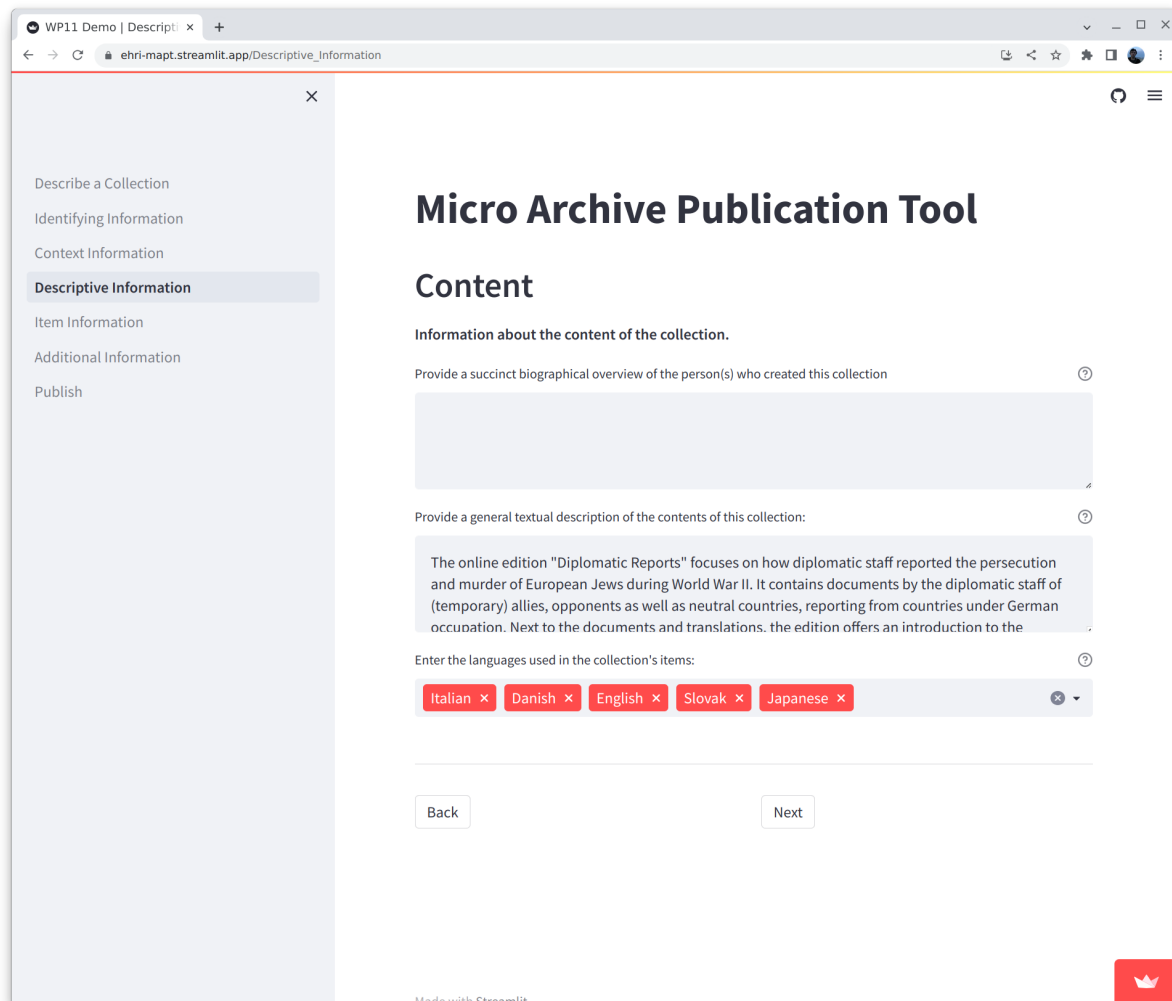
The screenshot shows a web browser window with the URL `ehri-mapt.streamlit.app/identifying_information`. The page title is "Micro Archive Publication Tool" and the current step is "Identification". A sidebar on the left lists the steps: "Describe a Collection", "Identifying Information" (selected), "Context Information", "Descriptive Information", "Item Information", "Additional Information", and "Publish". The main content area has the heading "Information to identify the collection." and two text input fields. The first field is labeled "Enter a descriptive name for the full collection:" and contains the text "EHRI Diplomatic Reports". The second field is labeled "Enter a general description of the format of this material:" and contains the text "219 scanned images". Below the input fields are "Back" and "Next" buttons. At the bottom left, it says "Made with Streamlit" and at the bottom right, there is a red Streamlit logo.

The second page contains information about the holder of the collection and their address:

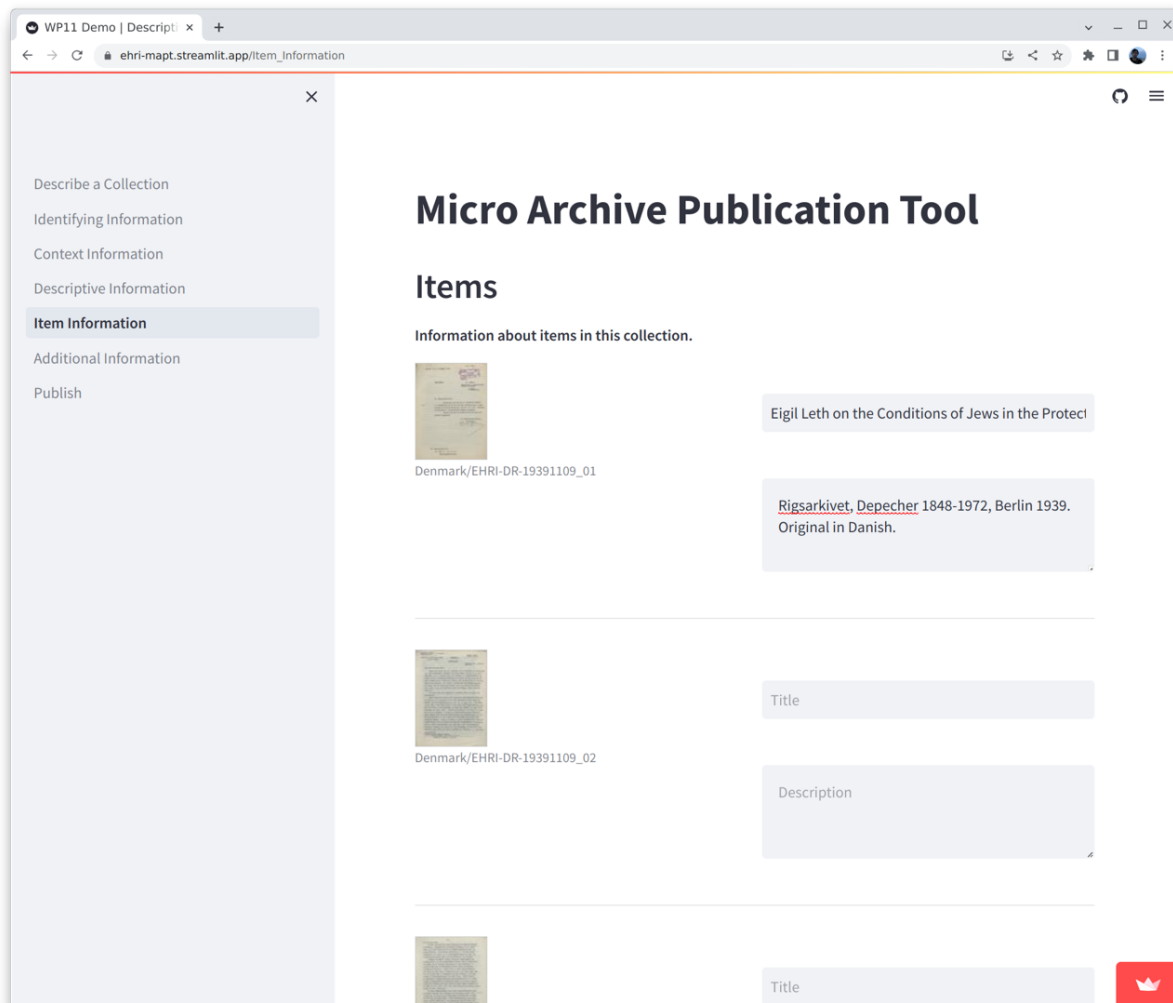


The screenshot shows a web browser window with the URL `ehri-mapt.streamlit.app/Context_Information`. The page title is "Micro Archive Publication Tool" and the current section is "Context". A sidebar on the left contains a menu with the following items: "Describe a Collection", "Identifying Information", "Context Information" (which is highlighted), "Descriptive Information", "Item Information", "Additional Information", and "Publish". The main content area has the heading "Context" and a sub-heading "Information about where the collection is kept and who is responsible for it." Below this, there are three input fields: the first is labeled "Enter the name of the person or organisation responsible for custody of this collection:" and contains the text "EHRI Project"; the second is labeled "Enter the street address where this collection is held:" and contains a hyphen "-"; the third is labeled "Enter the postal code where this collection is located:" and contains the text "1016 CJ". At the bottom of the form, there are two buttons: "Back" and "Next". The footer of the page includes the text "Made with Streamlit" and a small red crown icon.

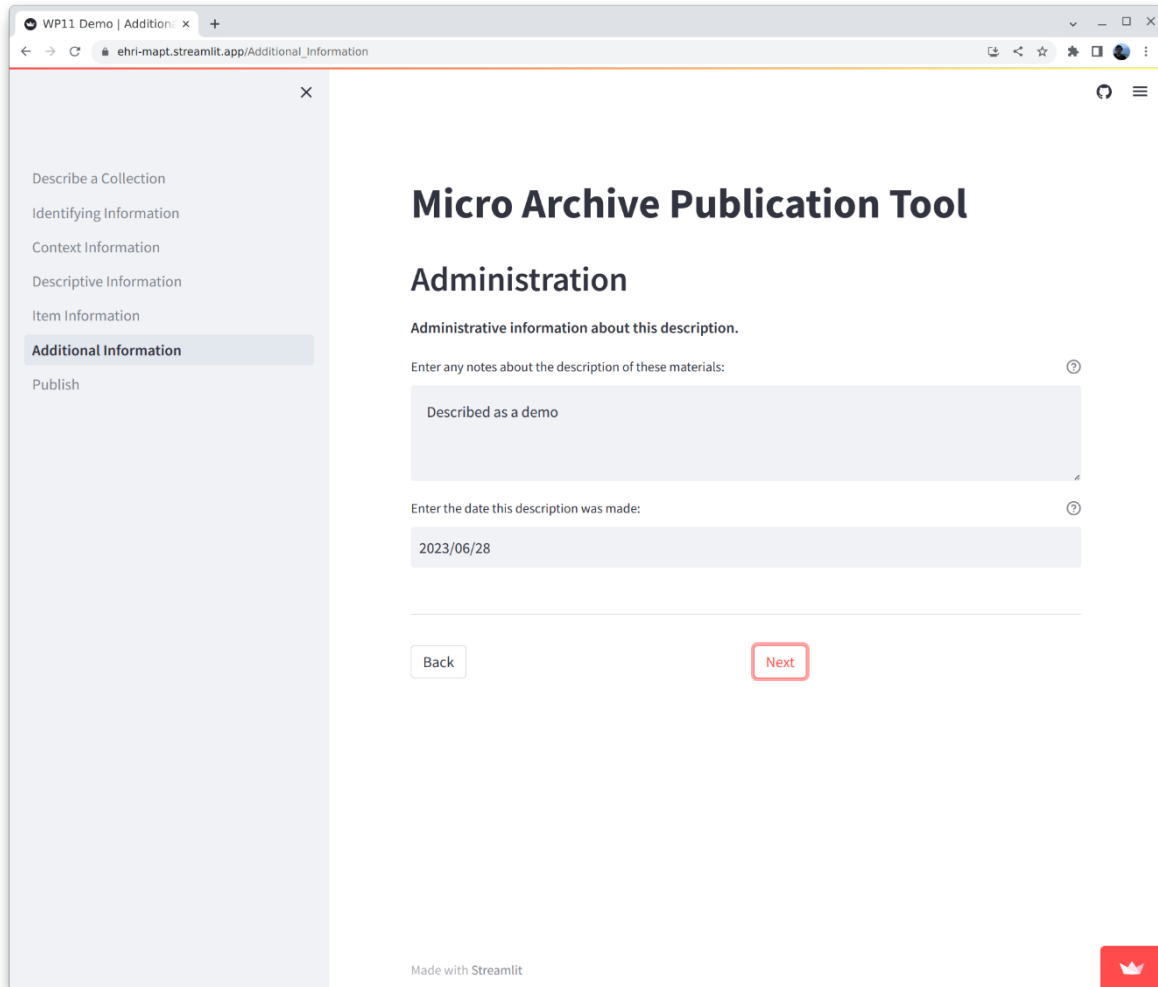
The third page provides fields for the description of the contents of the collection, and the languages used:



The fourth page provides the fields where each item can be individually described with a name and some descriptive text:



The fifth page provides fields for additional administrative information about the description:



WP11 Demo | Addition: x +

ehri-mapt.streamlit.App/Additional_Information

Describe a Collection

Identifying Information

Context Information

Descriptive Information

Item Information

Additional Information

Publish

Micro Archive Publication Tool

Administration

Administrative information about this description.

Enter any notes about the description of these materials: ⓘ

Described as a demo

Enter the date this description was made: ⓘ

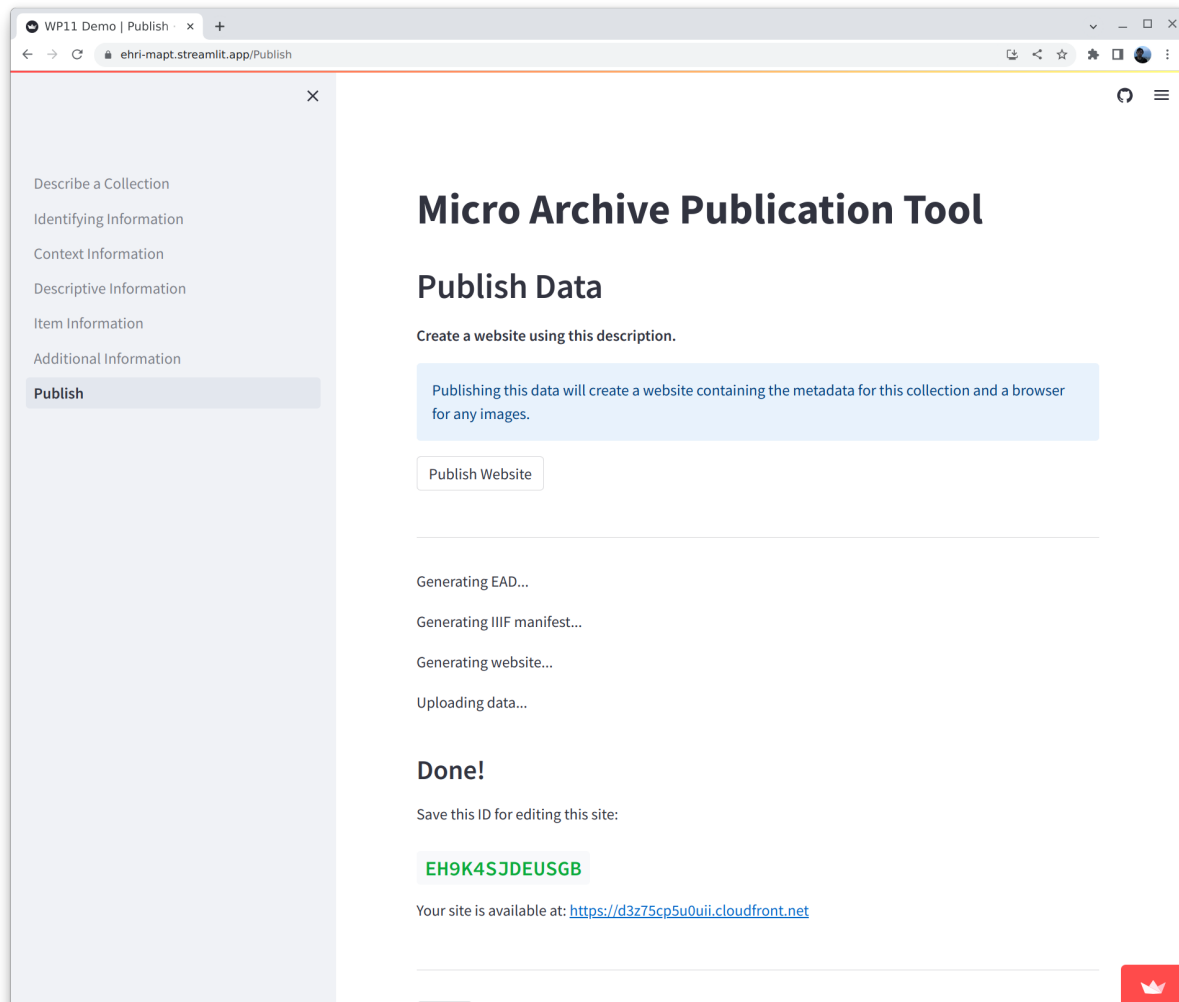
2023/06/28

Back Next

Made with Streamlit

4.5 Publishing the data

At present the “Publish” page contains only a “Publish Website” button, though the data to be published could also be offered for direct download to the user:



Behind the scenes, publishing a website involves generating the HTML, XML, and IIIF manifest data, then uploading it to a specific *private* Cloud storage location. Finally, an AWS CloudFront distribution is created that exposes the generated data at a new semi-random URL, e.g.:

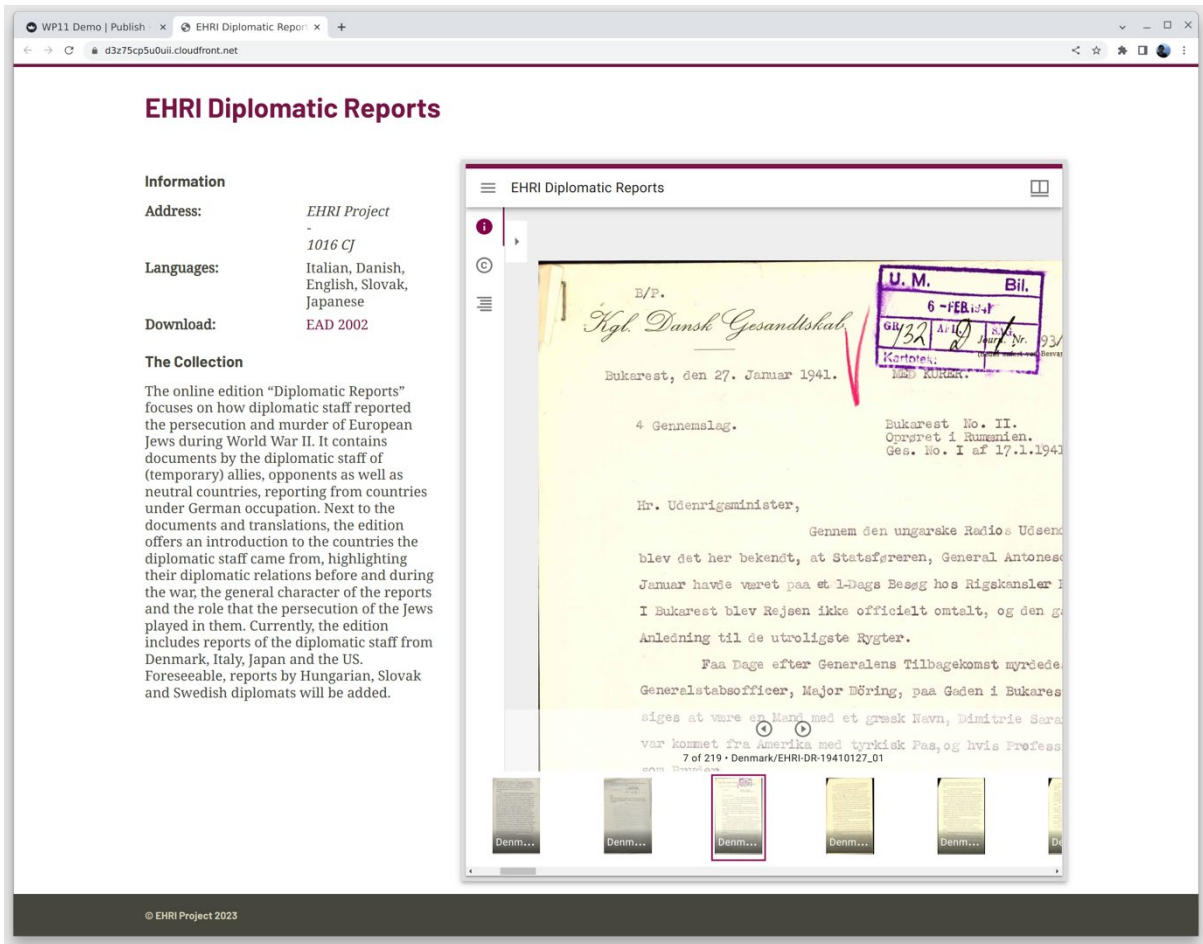
<https://d3z75cp5u0uii.cloudfront.net/>

This URL is not intended to be the final URL for a published micro-archive but provides a stopgap via which the material can be accessed online. To publish the site to an external audience the CloudFront distribution can be associated, via DNS, to a URL registered for the purpose.

Publication results in a *site key* being generated that is randomly generated and can be used to edit the previously provided information and republish the site if needed.

4.6 The output: a published website

A few minutes after publication the site will typically become available.



EHRI Diplomatic Reports

Information

Address: EHRI Project
1016 CJ

Languages: Italian, Danish, English, Slovak, Japanese

Download: EAD 2002

The Collection

The online edition "Diplomatic Reports" focuses on how diplomatic staff reported the persecution and murder of European Jews during World War II. It contains documents by the diplomatic staff of (temporary) allies, opponents as well as neutral countries, reporting from countries under German occupation. Next to the documents and translations, the edition offers an introduction to the countries the diplomatic staff came from, highlighting their diplomatic relations before and during the war, the general character of the reports and the role that the persecution of the Jews played in them. Currently, the edition includes reports of the diplomatic staff from Denmark, Italy, Japan and the US. Foreseeable, reports by Hungarian, Slovak and Swedish diplomats will be added.

EHRI Diplomatic Reports

E/P.
Kgl. Dansk Gesandtskab
Bukarest, den 27. Januar 1941.

4 Gennemslag. Bukarest No. II.
Oprøret i Rumænien.
Ges. No. I af 17.1.1941

Hr. Udenrigsminister,

Gennem den ungarske Radios Udsendelse blev det her bekendt, at Statsføreren, General Antonescu den 23. Januar havde været paa et 1-Dags Besøg hos Rigskansleren i Bukarest. I Bukarest blev Rejsen ikke officielt omtalt, og den gav Anledning til de utroligste Rygter.

Paa Dage efter Generalens Tilbagekomst myrdede Generalstabschefen, Major Döring, paa Gaden i Bukarest siges at være en Mand med et græsk Navn, Dimitrie Saraghi, var kommet fra Amerika med tyrkisk Pas, og hvis Professions

7 of 219 - Denmark/EHRI-DR-19410127_01

Some features of the published site include:

- An HTML presentation of the metadata provided by the user.
- A downloadable (or hot-linkable) XML EAD 2002 file.
- A viewer via which the IIIF manifest can be browsed.

Like other IIIF Presentation API v3.0 compatible viewers, it allows the user to:

- Pan and zoom individual images in a web-efficient manner.
- Navigate the structure of the collection, e.g., the folders in which the original images were arranged.
- View certain image-specific metadata, provided by the user.
- Share the link to the manifest, allowing others to view, compare, or combine its contents with other datasets, without exposing the source data.

4.7 How the viewer works

An EHRI-hosted [IIIF Image API v3.0](#)-compatible server is configured with private access to the underlying Cloud storage. When a user browses the collection via the manifest viewer, it requests web-optimised (e.g., JPEG) proxies of the source images on-the-fly. The source image scans, which may have a large file-size and potentially contain sensitive metadata, remain private and inaccessible to the web. Moreover, the system makes it possible to restrict

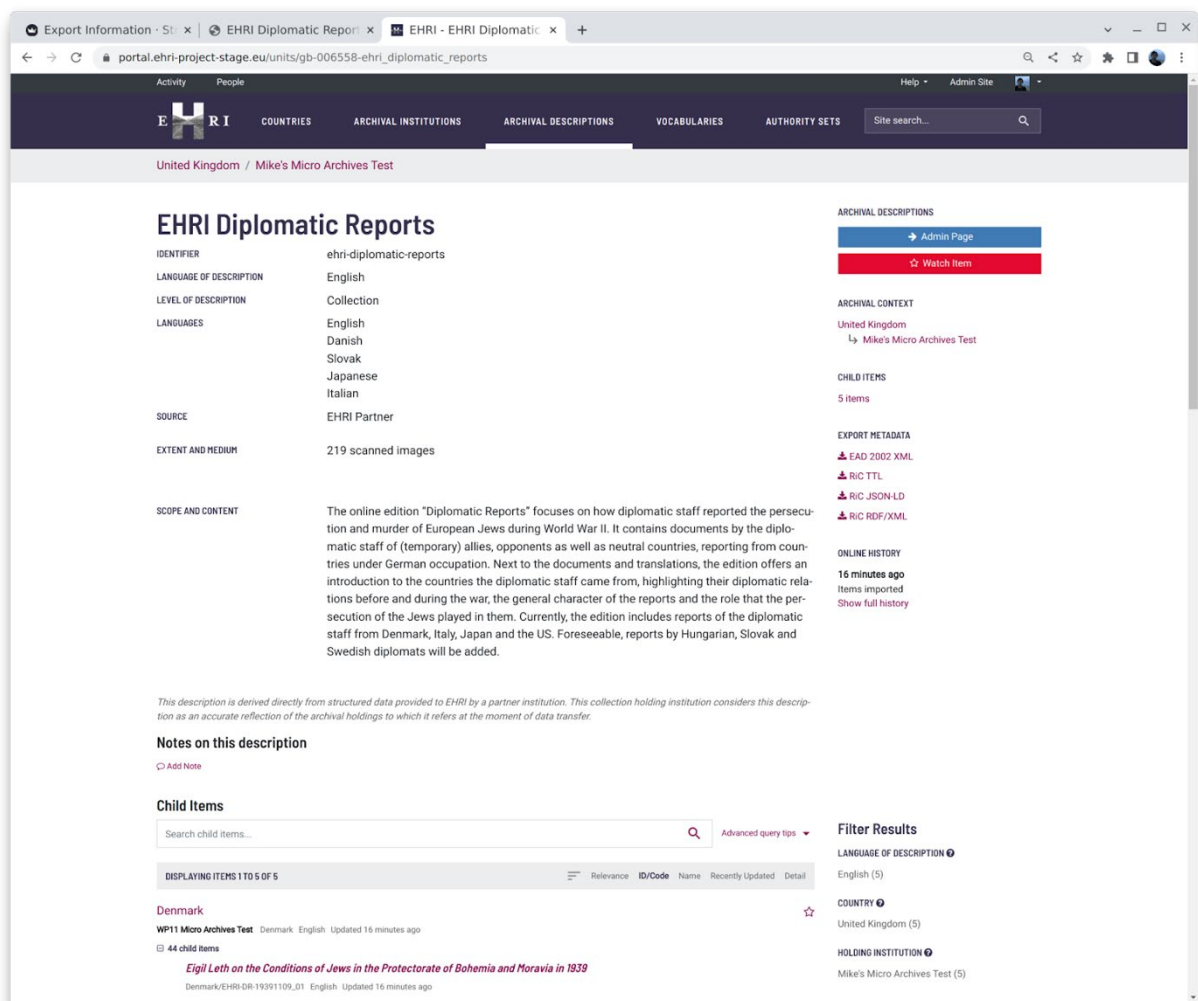
the quality of the publicly accessible version of the scan and to overlaid watermarks and other identifying information to discourage re-use without explicit permission.

This system works with any image collection uploaded to EHRI's storage and can easily be replicated by other infrastructures.

4.8 Harvesting MAPT-generated EAD metadata

The EAD XML generated by MAPT can be easily harvested by the EHRI portal. This requires that first an "institution" is created representing the micro-archive, which will contain its contents described as a collection.

The structure of the micro-archive, derived from its file/folder structure, will be reflected in the hierarchy of the collection harvested by the portal:



4.9 Limitations and future work

MAPT is currently a proof of concept and therefore could be further developed in the future, including:

- More analysis of the subset of archival metadata required for a "typical" micro-archive.
- Extending the metadata that is available for describing items.

- Allowing description of metadata-only collections, e.g., where there are no available digital proxies (in essence, a way to generate harvestable EAD-format metadata).
- Allowing metadata to be associated with intermediate levels of description, e.g., “folders” of images, which might constitute material of some shared provenance or purpose.
- Allowing multiple images to be “collapsed” into a single item, e.g., those that are scanned pages of a book or other multi-image item.
- Allowing more customisation of the generated HTML.
- More thoroughly vetting the generating of the EAD XML produced and adapting it to the more modern EAD-3.

5 Lessons Learned

Ever since the early T11.2-retaied internal meetings, the biggest lesson learned was that there is a consensus among archivists, historians, and researchers that in most cases, micro-archives will eventually open their collections through their cooperation with larger institutions that will either copy micro-archival collections or acquire the originals through donations. Usually, micro-archival communities have neither the resources (in terms of staff but also equipment) nor the expertise to catalogue, preserve and offer access to their entire collections according to international standards. This is something that has been mentioned during every WP11 meeting, EHRI-3 General Partner Meetings and at the engagement workshop that took place in May 2023.

EHRI members and workshop attendees representing larger institutions that work with micro-archives found it difficult to formulate any new guidelines or services that EHRI could offer to facilitate them in their work *specifically* with micro-archives, as opposed to their work in general. These institutions normally have the necessary equipment as well as expertise to perform their tasks and they also have existing practices in place to make their metadata harvestable by the EHRI Portal. This means that for them, once they can gain custody of a micro-archive through a donation and catalogue it, sharing high quality metadata with EHRI is not a problem. They argue that normally the hardest part of their job is to gain the trust of the owners of micro-archives and persuade them that donating their material or copies of it is the best way to ensure that it will be preserved for future generations as well as to reach a wider audience and have a greater impact by becoming accessible for research purposes. This means that the biggest challenge faced by EHRI-affiliated institutions when it comes to WP11 is not actually any lack of technical support or knowledge but rather the fact that trust-building and relationship-building take time and require emotional labour on the part of the staff performing this work (see also Regehr et al. (2022) on the empathic engagement and emotional response that is involved in archivists' work.)

Since supporting an institution in building personal relationships and trust with micro-archives is not something that can be realistically addressed in the form of a new tool or service by EHRI, T11.2 maintains that EHRI's most valuable contribution towards this work would be to continue the work started by WP4 in mobilising local representatives to map and reach out to local micro-archives and bring them in contact with larger institutions in their region that could guide and support them directly, in their native language, and with a better overview of the content and context of their collections. This is also consistent with the conclusion of D11.2 and is something that other WP11 tasks (and to a certain degree WP9) are already addressing. T11.2 believes that the future EHRI-ERIC National Nodes will be even better poised to achieve this goal.

Considering the consensus that in most cases micro-archives will share their data and metadata in partnership with larger institutions, T11.2 also acknowledged the fact that there will still be cases where micro-archive owners might opt for retaining custody of their material and it is exactly in these cases where a new EHRI tool or service might have the biggest impact. This was an important motivation behind the development of the MAPT prototype.

When workshop participants were asked for their feedback regarding MAPT, the majority appreciated the user-friendliness of the tool and the fact that extra effort was put into making its output easily harvestable and interoperable. The resulting website was considered as a very good incentive to persuade more micro-archives to share their metadata with the EHRI Portal. However, participants were also cautious about having non-professional micro-archive owners describe their own material because this might not only be unsustainable but also impractical. Indicatively, during the workshop a personal anecdote was shared by one of the attendees that sometimes, micro-archive owners might assume that, for example, a picture was taken at a certain concentration camp, but further research proved it was taken in another.

When it comes to making these collections accessible for research purposes, there is generally the expectation that an expert will be involved in cataloguing and indexing them so that people using such records will know whether or not they can trust their source and description. However, some researchers argued that they appreciate having access to descriptions written by micro-archive owners because, apart from the fact that this gives them a voice, the way they interpret their material can in itself be a research subject. Both sides agreed that family-written descriptions should be flagged as such so that researchers know the author(s) and context in which these descriptions were created or written. This will also help EHRI ensure that its credibility is not jeopardised by aggregating inaccurate descriptions.

Certain aspects of the work entailed in making a micro-archival collection as discoverable as possible post-digitisation, such as generating full-text transcripts of scanned documents and audio recordings, remain a challenge even for some larger institutions that lack enough human resources to pursue more of this work. T11.2 will consider offering AI-powered tools, such as OCR and ASR tools, to facilitate small as well as larger institutions with these kinds of tasks. Another remark that was made by one of the workshop attendees who is the curator of a UK-based micro-archive is that there is limited expertise in digitising 3D collection objects and perhaps EHRI and its partner institutions could offer a guide addressing this.

6 Conclusion

In the previous sections, this deliverable addressed the challenges and opportunities facing EHRI in terms of micro-archive engagement and the development of related services. What became clear through the work in T11.2 is that micro-archives often lack the resources and expertise necessary to comply with international archival standards. They usually also lack the equipment and know-how to digitise and preserve their material. Consequently, archivists, historians and researchers agree that most micro-archives will eventually cooperate with larger institutions for cataloguing their collections and making them more accessible. The most important and hardest challenge that professional archivists face when dealing with micro-archives is building and maintaining connections and trust with their owners, and this is not something that new EHRI services can address directly. However, EHRI's extended network of institutions and experts could be its greatest contribution to this work, by helping to facilitate such connections and to allocate the right people and resources to the right task. We expect that once the EHRI Central Hub and National Nodes are fully set up and operating, this work will become easier (at least for the micro-archives that are located in countries where a National Node will be created, and provided that the National Coordinators Committee will consider engaging with micro-archives as a priority). Furthermore, EHRI services might have an important role to play in the use-cases where micro-archive owners are committed to

retaining custody of their material, and to catalogue and preserve it themselves, and the MAPT prototype was built to cater to such cases.

Finally, T11.2 considers this task as still a work in progress and moving forward, we suggest that future work should focus on the following tasks:

- Keep testing the MAPT prototype during future micro-archive engagement workshops to receive feedback and make improvements.
- Build communication channels with other WPs and with the potential future EHRI-IP project to ensure that both existing and future services take into account the needs of micro-archives, starting with the WPs in charge of the newsletter and the blog.
- Propose that EHRI should strengthen its regional efforts to map micro-archives and assist in building connections between them and larger nearby institutions that can provide in-person support and guidance.
- Consider exploring state-of-the-art AI tools that will assist institutions of any size to make already digitised collections even more discoverable to maximise their impact and explore opportunities to build capacity in areas where there is lower expertise available, or where current methods require advanced skills, such as the digitisation of 3D objects.
- Consider sharing guidelines about document preservation to ensure that micro-archival material is not getting lost before experts can take action to preserve it.

7 References

- Erdösi, L., Zwilling, G., Wilhelm, M., Kovács, É., Zimmermann, N., Uiterwaal, F., Berkhoff, K., & Vanden Daelen, V. (2022). *Report on micro-archival communities and user requirements* (Deliverable 11.2). EHRI-3.
- Regehr, C., Duff, W., Aton, H., & Sato, C. (2022). "Humans and records are entangled": Empathic engagement and emotional response in archivists. *Archival Science*, 22(4), 563–583. <https://doi.org/10.1007/s10502-022-09392-5>