

EAD Conversion Tool

Developed for EHRI WP10 by [Ontotext](#)

User Guide

Document Date (23 Nov, 2017)

Table of Contents

Overview	3
Requirements	3
System Requirements	3
Licensing	3
Run the EAD Conversion Tool as a desktop installation	3
On Windows	3
On Mac OS	4
On Linux	4
EAD Conversion Tool HOME directory	4
Input directory	4
Mapping directory	4
Output directory	4
XQuery directory	5
Procedure	5
Select your organization	5
Select your files INPUT format	6
Select the transformation type	6
Preview/edit the mapping config file	7
Preview the input files	9
Start transformation	9
Explore the conversion report	9
Exploring the EAD validation inconsistencies	10
Validate the corrections	12
6 Validation tool	12

1 Overview

EAD Conversion Tool is a web-based data transformation and validation tool, created in the scope of the European Holocaust Research Infrastructure (EHRI) project. It can be used for transforming XML, XML-EAD1, and CSV data in a well-formed EAD 2002 format by mapping, correcting and validating it in accordance to the EAD 2002 schema guidelines and the harvesting/ingest workflow.

EAD Conversion Tool enables you to:

- Choose the mapping configuration file of your organization or use your own;
- Edit the mapping configuration to suit your needs;
- Use a custom transformation type;
- Convert your data to the EAD 2002 standard;
- Preview all mapping inconsistencies;
- Generate a well-formed EAD 2002 data file.

Supported formats:

- Input files (/input directory) – XML, XML EAD 1, CSV;
- Output files (/output directory) – EAD 2002;
- Mapping files (/mapping directory) – XLS, XLSX, Google Spreadsheet.

Of course, if you have any questions regarding this user guide and/or the EAD Conversion Tool, you can contact Francesco.gelati@arch.be or [ONTOTEXT](#).

2 Requirements

2.1 System Requirements

- Microsoft Windows 7 SP1, Windows 8, and Windows 10
- Linux
- Mac
- [Java 8](#) or later

2.2 Licensing

EAD Conversion Tool is available under [Apache License](#). It is free and open-source.

3 Run the EAD Conversion Tool as a desktop installation

The EAD Conversion Tool setup and running is easy and straightforward.

3.1 On Windows

1. Download the EAD Conversion Tool package from Ontotext FTP server -

ftp://ftp.ontotext.com/pub/EHRI/conversion_tool/, or from GitHub - <https://github.com/EHRI/ehri-conversion-tools>.

2. Double click the application file and follow the on-screen installer prompts.
3. Locate the EAD Conversion Tool on the Windows Start menu and start it. EAD Conversion Tool GUI (Graphical User Interface) opens at <http://localhost:8080>.

3.2 On Mac OS

1. Download the EAD Conversion Tool package from Ontotext FTP server - ftp://ftp.ontotext.com/pub/EHRI/conversion_tool/, or from GitHub - <https://github.com/EHRI/ehri-conversion-tools>.
2. Double click it and get a virtual disk on your desktop. Copy the program from the virtual disk to your hard disk applications directory, and you're set.
3. Start the EAD Conversion Tool by clicking the application icon. The EAD Conversion Tool opens at <http://localhost:8080>.

3.3 On Linux

1. Download the EAD Conversion Tool package from Ontotext FTP server - ftp://ftp.ontotext.com/pub/EHRI/conversion_tool/, or from GitHub - <https://github.com/EHRI/ehri-conversion-tools>.
2. Install the package with *sudo rpm -i* or *sudo deb -i* and the name of the downloaded package. Alternatively, you can double click the package name.
3. Start the EAD Conversion Tool by clicking the application icon. The EAD Conversion Tool opens at <http://localhost:8080>.

4 EAD Conversion Tool HOME directory

When started the EAD Conversion Tool automatically creates four sub-directories in its HOME directory for storing data and configurations.

4.1 Input directory

The */input* data directory is where you add the data files you want to transform.

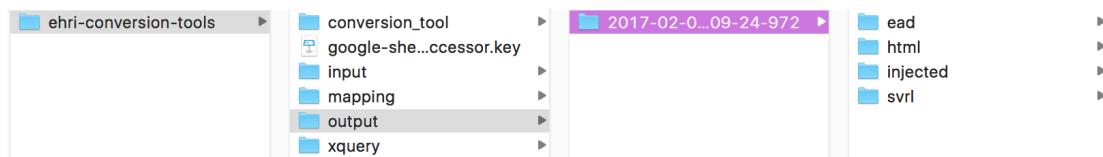
4.2 Mapping directory

The */mapping* directory is where you can add your own mapping configuration files, or in some cases, the edited default mapping config, after correcting the mapping inconsistencies from the validation results.

4.3 Output directory

The */output* data directory is where the EAD Conversion Tool stores all transformed data files. They are organized into sub-directories, which names reflect the time that they were created. Each sub-directory contains four other folders – /ead, /html, /injected, and /svrl. The ones of your interest are the first two folders, as they

contain the newly generated EAD 2002 files, as well as the results from the EAD validation in HTML format.



4.4 XQuery directory

The `/xquery` directory is where you can add a custom `.xqy` file to transform data files in a format different from the default EAD 2002 standard.

5 Procedure

To transform your data into a well-formed EAD 2002 format, follow the steps:

5.1 Select your organization

1. Put the data files you want to transform in the `~/input` directory.
2. Open <http://localhost:8080> in a browser.

EHRI FILE MAPPING

DOCUMENTATION

1 2 3 4 5 6

Please select your Organization

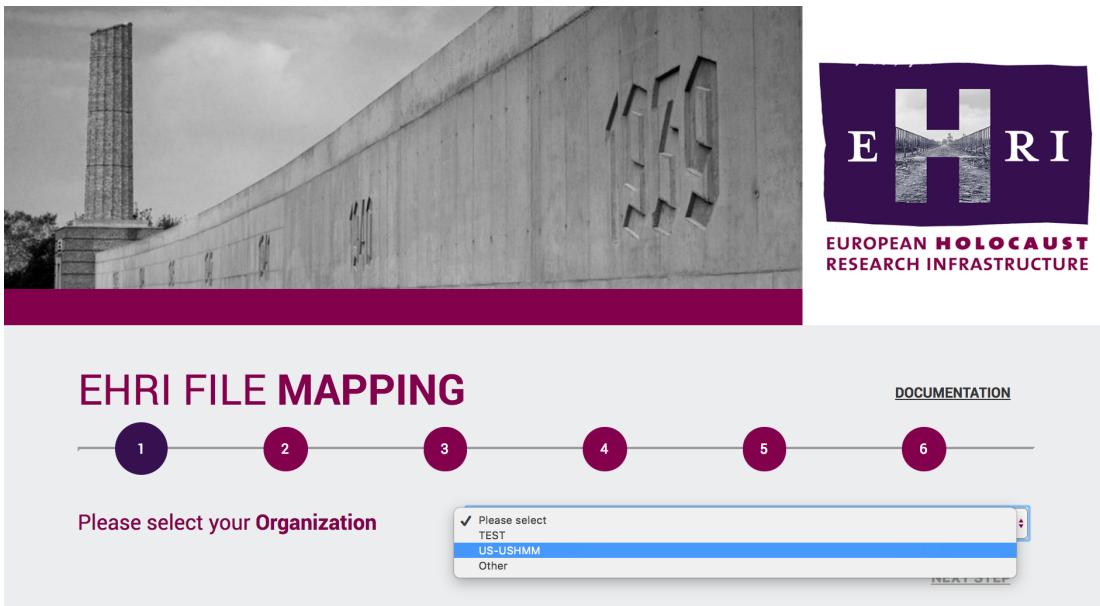
Please select

NEXT STEP

EHRI is funded by the European Union

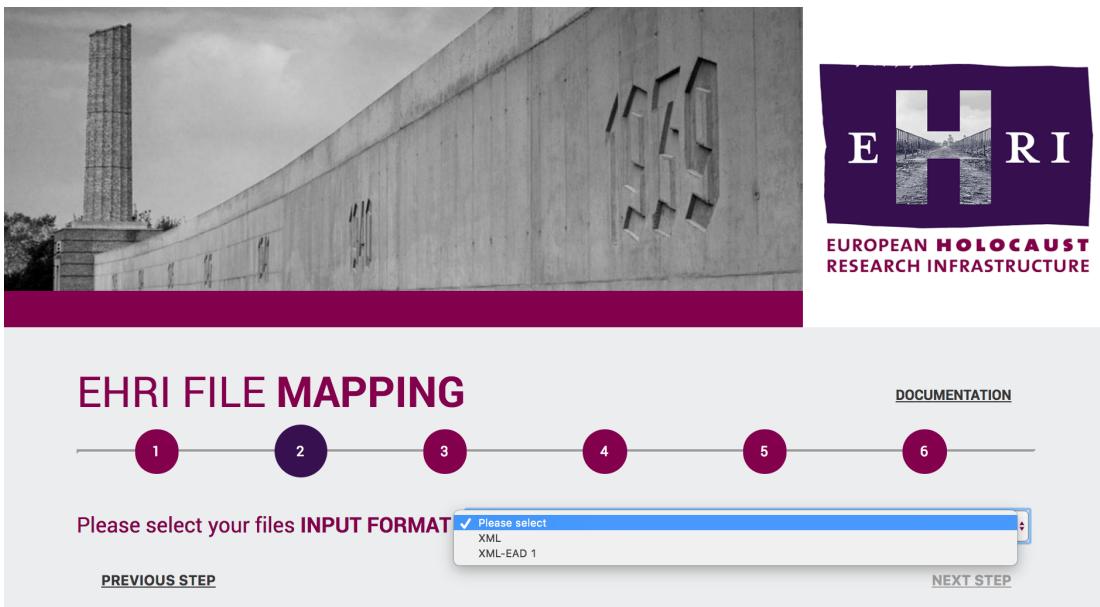
ontotext

3. Select your organization from the drop-down list or select “Other” for organization which is not presented and click NEXT STEP.



The screenshot shows the first step of the EHRI File Mapping process. At the top right is the EHRI logo. Below it is a purple bar with the text "EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE". The main title "EHRI FILE MAPPING" is centered above a horizontal progress bar with six numbered circles (1 to 6). Step 1 is highlighted with a blue circle. Below the progress bar, the text "Please select your Organization" is displayed, followed by a dropdown menu showing "Please select", "TEST", "US-USHMM" (which is selected and highlighted in blue), and "Other". At the bottom right of the page is a "NEXT STEP" button.

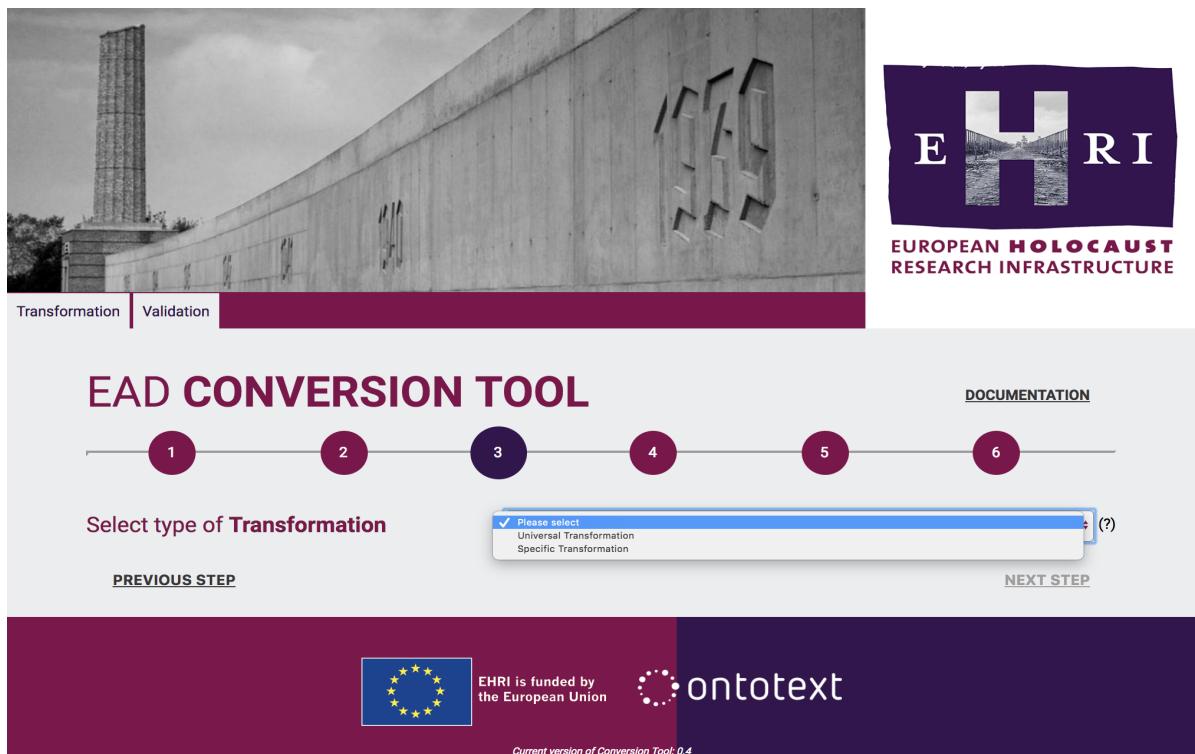
5.2 Select your files INPUT format



The screenshot shows the second step of the EHRI File Mapping process. It has the same visual structure as the first step, with the EHRI logo at the top right, a purple bar with the text "EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE", and a title "EHRI FILE MAPPING" above a progress bar. Step 2 is highlighted with a blue circle. The text "Please select your files INPUT FORMAT" is shown above a dropdown menu, which now lists "Please select", "XML" (selected and highlighted in blue), and "XML-EAD 1". Below the dropdown are "PREVIOUS STEP" and "NEXT STEP" buttons.

5.3 Select the transformation type

There are two types of transformation *Universal (default)* and *Specific*. Using the *Universal* one, you can transform your data files in the EAD 2002 format. If you want to use the tool for transforming data in other formats, you should create your own xquery transformation schemes and add them to the /xquery folder. Then, they can be used when the *Specific transformation* type is selected.



5.4 Preview/edit the mapping config file

All mapping config files are stored as Google Spreadsheets. Depending on your access rights, you can view or edit them, directly in the EAD Conversion Tool GUI or by clicking the *View Google Spreadsheet* link.

If you need to use a custom mapping, you can add it to the */mapping* folder and select it from the *Select local mapping file* drop-down list.



EHRI FILE MAPPING

[DOCUMENTATION](#)



[VIEW GOOGLE SPREADSHEET](#)

	A	B	C	D
target-path	target-node	source-node		value
1 /	ead	//doc		
2 /ead/	eadheader	.		
3 /ead/eadheader/	profiledesc	.		
5 /ead/eadheader/profiledesc/	creation	/str[@name="datetimemodified"]		"This EAD is created by EHRI on ", <date>
6 /ead/	archdesc	.		
7 /ead/archdesc/	did	.		
8 /ead/archdesc/did/	unitid	if (/str[@name="accession_number"]/text() != /str[@name="id"]/text()) then /str[@name="id" text()		attribute label ("accession_number"), text()
9 /ead/archdesc/did/	unitid	./str[@name="accession_number"]		attribute label ("former_accession_number")
10 /ead/archdesc/did/	unitid	./arr[@name="accession_number_add1"]/str		attribute label ("recordgroup_number")
11 /ead/archdesc/did/	unitid	./arr[@name="rg_number"]/str		attribute label ("subtitle"), text()
12 /ead/archdesc/did/	unittitle	./arr[@name="subtitle"]/str		

Select Local Mapping

Select local mapping file

[PREVIOUS STEP](#)

[NEXT STEP](#)

5.5 Preview the input files

The screenshot shows the EHRI File Mapping interface. At the top right is the EHRI logo: a purple square with a white stylized 'H' containing a black and white photograph of a concentration camp fence, with the text 'EHRI' to its right and 'EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE' below it. The main title 'EHRI FILE MAPPING' is at the top left. Below it is a horizontal progress bar with six numbered circles (1-6). The organization 'US-USHMM' is listed under 'Organization'. A section titled 'Input Folder Content' shows a single file named 'document.2016-11-22.xml'. Navigation buttons 'PREVIOUS STEP' and 'START TRANSFORMATION' are at the bottom.

5.6 Start transformation

This screenshot is similar to the previous one but with a darker background. It shows the EHRI File Mapping interface with the same progress bar, organization selection, and file preview. The 'START TRANSFORMATION' button is now highlighted in red at the bottom right. The EHRI logo is visible on the right side.

5.7 Explore the conversion report

The transformed EAD files are added to the `/ead` folder. The Conversion report shows the number of EAD files created, as well as the number of inconsistencies

(errors) found in each of them.



EHRI FILE MAPPING

[DOCUMENTATION](#)

1 2 3 4 5 6

Data was transformed Successfully!

From 1 files we created 9063.

Please take a look at the table below to find out the generated files and their errors

Conversion report

File Name	Number of errors
document_00000001.2016-11-22	11
document_00000002.2016-11-22	8
document_00000003.2016-11-22	11
document_00000004.2016-11-22	7
document_00000005.2016-11-22	6
document_00000006.2016-11-22	5
document_00000007.2016-11-22	12
document_00000008.2016-11-22	9
document_00000009.2016-11-22	9
document_00000010.2016-11-22	11
document_00000011.2016-11-22	6
document_00000012.2016-11-22	7

5.8 Exploring the EAD validation inconsistencies

1. Click a file name to check its errors.

The screenshot shows a user interface for an EAD conversion tool. At the top right is the logo for "EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE" (EHRI) featuring a stylized "H" and "E" with a landscape image in the background. Below the logo is a dark purple header bar with the word "Transformation" and "Validation". The main content area displays a document titled "doc-sample_irn515021" with a file location path. On the left, there is a sidebar with a maroon background containing two blue links: "Date of the Unit" and "Date of the Item". The main content area has a white background with a blue border and contains several sections of XML code. A red "X" button is located in the top right corner of the content area.

Each EAD generated file is presented in an html file. The navigation menu on the left lists all XML elements that do not comply with the EAD 2002 standard.

This screenshot shows an EAD generated file in HTML format. The left sidebar has a maroon background with blue "Date of the Unit" and "Date of the Item" links. The main content area has a white background with a blue border. It displays several XML elements with validation errors highlighted in red boxes. For example, the "Date of the Unit" element has an error message: "value of attribute \"type\" is invalid; must be equal to \"bulk\" or \"inclusive\"". Another error message appears below it: "value of attribute \"type\" is invalid; must be equal to \"bulk\" or \"inclusive\"". The "normal" and "type" attributes are shown with their values: "1920/1981" and "value" respectively, followed by the error message.

2. Click the EAD element to see its errors and correct them

For example, the picture below shows that the “Profile Description” element is not allowed, because there is a missing “eaid” element. In order to correct this error,

you must add this element to your XML input file. Depending on the validation errors, you can correct them in the input file, the mapping configuration, or the source code.

The screenshot shows the EAD Conversion Tool's validation interface. On the left, a sidebar lists categories: Profile Description, Archival Description, Date of the Unit, Date of the Unit, Conditions Governing Access, Acquisition Information, Arrangement, Biography or History, Biography or History, Scope and Content, and Conditions Governing Use. The main area displays two validation results:

- Profile Description:** An error message: "[ERROR] element "profiledesc" not allowed yet; missing required element "eadid"" is shown above a note: "This EAD is created by EHRI on 2017-02-02+02:00 based on the JSON file provided by USHMM on TODO: find out where to get this . This JSON file is constructed on a Catalog Record that was last modified on 2016-11-17 11:12:18."
- Archival Description:** An error message: "[ERROR] element "archdesc" missing required attribute "level"" is shown above a "Descriptive Identification" section containing fields for "ID of the Unit" (im515021) and "Label accession_number" (2004.273.1).

5.9 Validate the corrections

To validate the corrections, repeat the procedure and check the conversion report again.

6 Validation tool

To start a validation of converted files or of EAD files follow these steps:

- 6.1 Add the data files you want to validate in the ~/input directory
- 6.2 Go to <http://localhost:8080> in a browser
- 6.3 From the top menu select Validation

The screenshot shows the EAD CONVERSION TOOL validation interface. At the top, there are tabs for Transformation and Validation, with Validation selected. To the right is the EHRI logo: European Holocaust Research Infrastructure. Below the tabs, a progress bar shows steps 1 through 6. Step 1 is highlighted in red, while others are greyed out. The main content area is titled "EAD CONVERSION TOOL" and shows "Input Folder Content" with files: "Files", ".DS_Store", and "collections-formatted.xml". A "START VALIDATION" button is at the bottom. At the very bottom, there are logos for the European Union and ontotext, along with the text "EHRI is funded by the European Union" and "Current version of Conversion Tool: 0.4".

EAD Conversion Tool User Guide

6.4 To start the Validation, click the Start Validation button

6.5 Explore the Validation report

The screenshot shows the validation results for a single file named "collections-formatted". The report indicates 0 errors found.

File Name	Number of errors
collections-formatted	0

Below the table, it says "SHOWING 1 TO 1 OF 1 ENTRIES". Navigation buttons include "PREVIOUS", a page number "1", and "NEXT".

At the bottom, there are two buttons: "REPEAT VALIDATION" and "START NEW VALIDATION".

Logos at the bottom include the European Union flag and the ontotext logo. A small text note says "EHRI is funded by the European Union".