



EAD Conversion Tool

Developed for EHRI WP10 by [Ontotext](#)

User Guide
Document Date (5 February, 2018)

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.



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



EHRI FILE MAPPING

[DOCUMENTATION](#)

1

2

3

4

5

6

Please select your Organization

Please select
TEST
US-USHMM
Other

[NEXT STEP](#)

5.2 Select your files INPUT format



EHRI FILE MAPPING

[DOCUMENTATION](#)

1

2

3

4

5

6

Please select your files INPUT FORMAT

Please select
XML
XML-EAD 1

[PREVIOUS STEP](#)

[NEXT STEP](#)

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.



Transformation Validation



EAD CONVERSION TOOL

DOCUMENTATION

1 2 3 4 5 6

Select type of Transformation

Please select
Universal Transformation
Specific Transformation

PREVIOUS STEP NEXT STEP

 EHRI is funded by the European Union



Current version of Conversion Tool: 0.4

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](#)

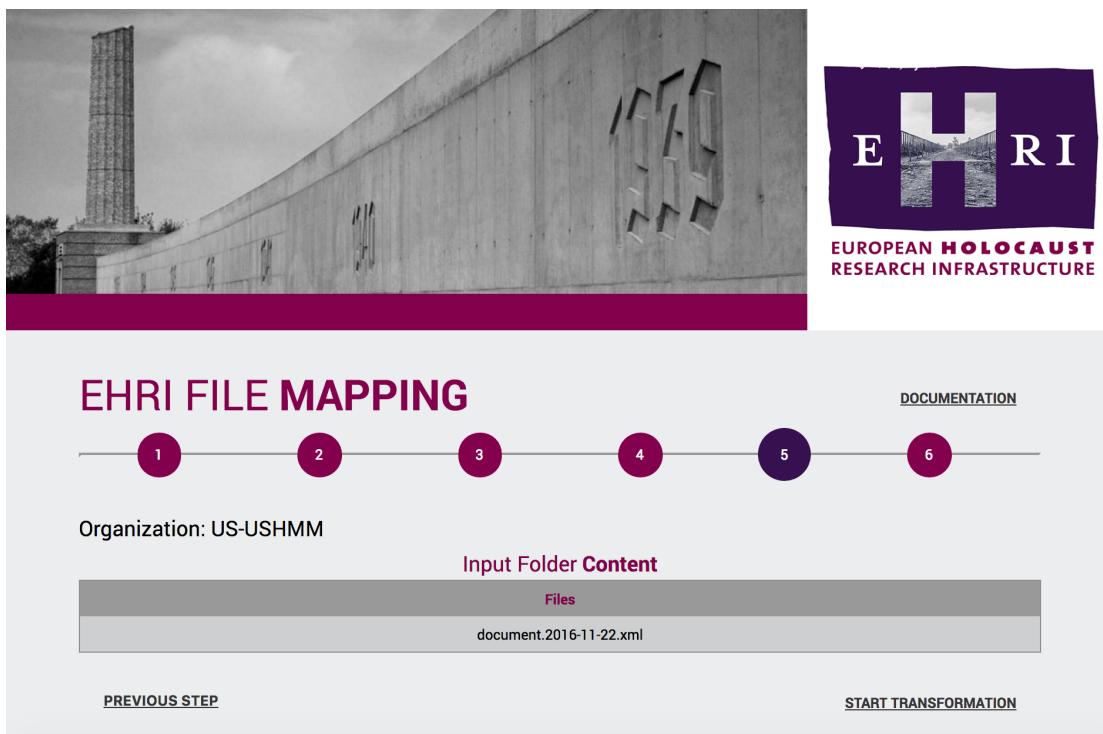
[VIEW GOOGLE SPREADSHEET](#)

	A	B	C	D
1	target-path	target-node	source-node	value
2	/	ead	//doc	
3	/ead/	eadheader	.	
4	/ead/eadheader/	profiledesc	.	
5	/ead/eadheader/profiledesc/	creation	./str[@name="date timemodified"]	"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"), tr
9	/ead/archdesc/did/	unitid	./str[@name="accession_number"]	attribute label ("former_accession_nu
10	/ead/archdesc/did/	unitid	./arr[@name="accession_number_add"]/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

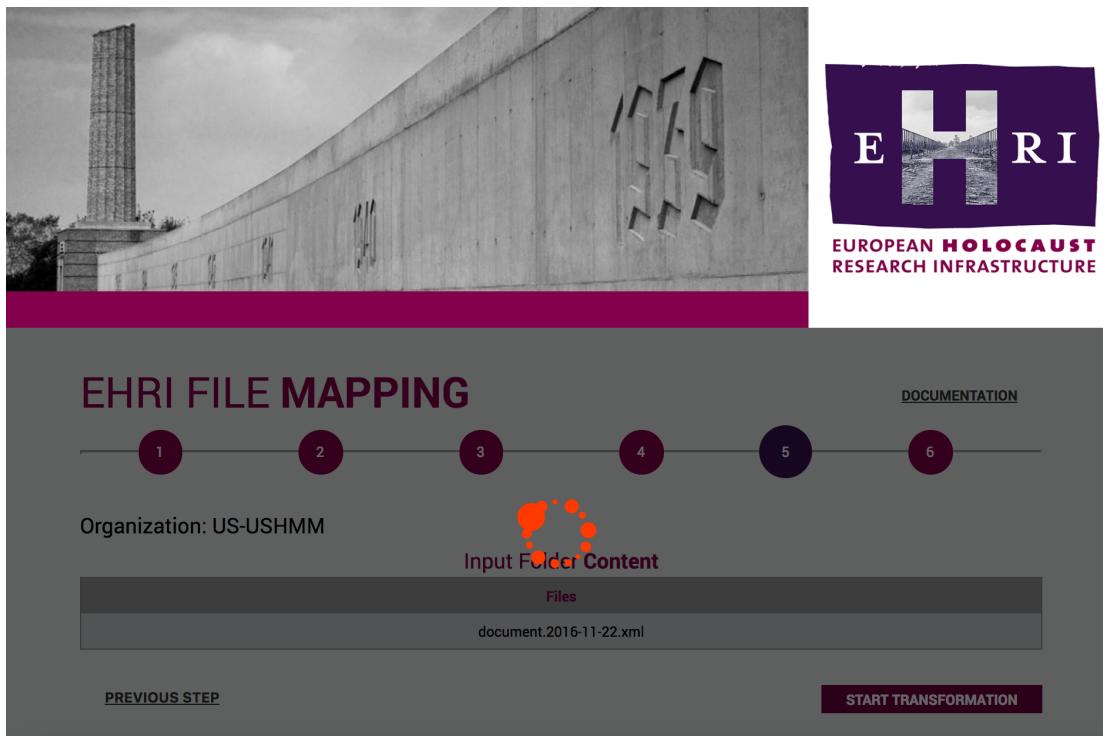
[NEXT STEP](#)
[PREVIOUS STEP](#)

5.5 Preview the input files



The screenshot shows the EHRI File Mapping interface. At the top right is the EHRI logo. Below it is a large image of a concrete wall with a large, three-dimensional 'H' sculpture embedded in it. A tall, thin, textured pillar stands to the left of the wall. The main title 'EHRI FILE MAPPING' is at the top left. Below it is a horizontal progress bar with six numbered circles (1 to 6). The organization 'US-USMHH' is listed. Under 'Input Folder Content', there is a 'Files' section containing the file 'document.2016-11-22.xml'. Navigation buttons 'PREVIOUS STEP' and 'START TRANSFORMATION' are at the bottom.

5.6 Start transformation



This screenshot is identical to the previous one, showing the EHRI File Mapping interface. The main difference is that the progress bar has moved to step 5, indicating the transformation process is underway. The rest of the interface, including the organization name, input file list, and navigation buttons, remains the same.

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.

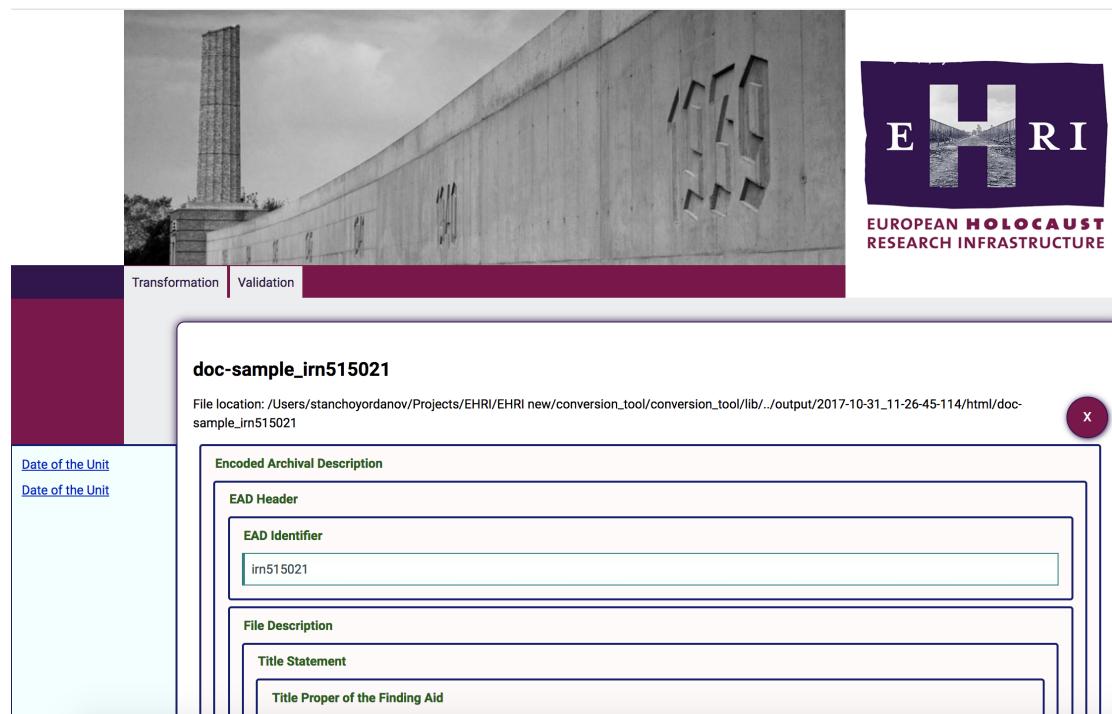
Please take a look at the table below to find out the generated files and their errors

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

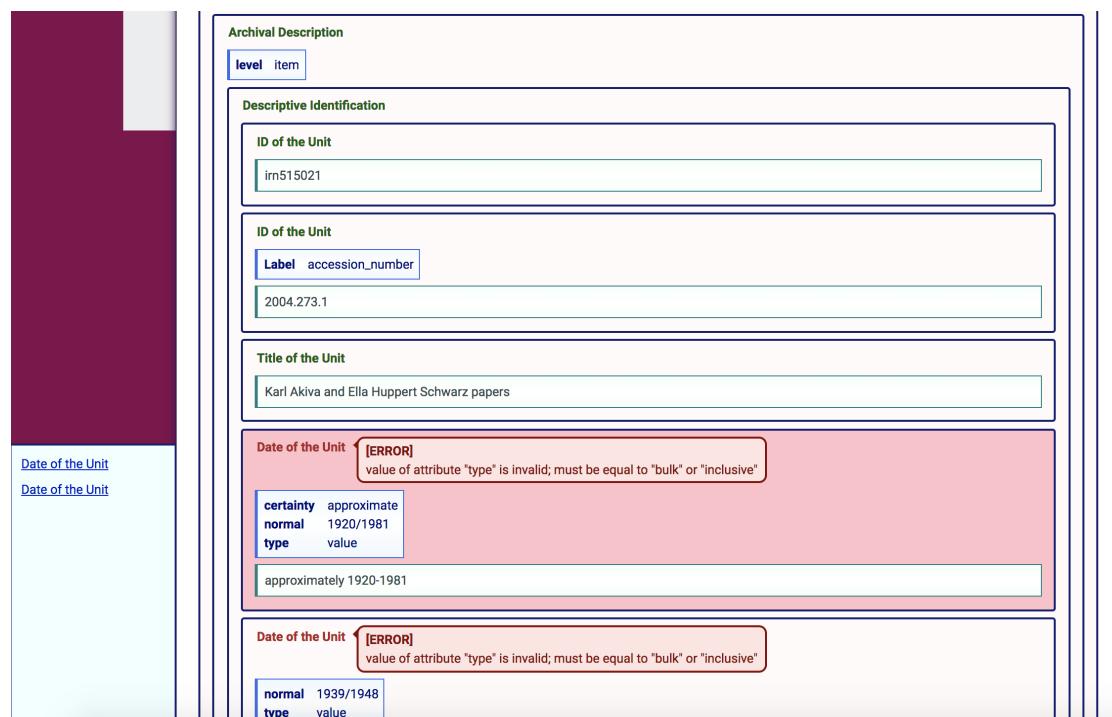
5.8 Exploring the EAD validation inconsistencies

1. Click a file name to check its errors.



The screenshot shows a web-based application for EAD conversion. At the top, there's a header with the EHRI logo and navigation links for 'Transformation' and 'Validation'. Below the header is a large image of a modern concrete building with a tall tower. The main content area displays an EAD XML document titled 'doc-sample_irn515021'. The XML structure is shown in a hierarchical tree view under the heading 'Encoded Archival Description'. The tree includes nodes for 'EAD Header' (with 'EAD Identifier' set to 'irn515021'), 'File Description' (with 'Title Statement' and 'Title Proper of the Finding Aid'), and other sections like 'Archival Description' and 'Descriptive Identification'. On the left side of the main content area, there's a sidebar with a dark red background containing two blue hyperlinks: 'Date of the Unit' and 'Date of the Unit'. A small circular 'X' button is located in the top right corner of the main content area.

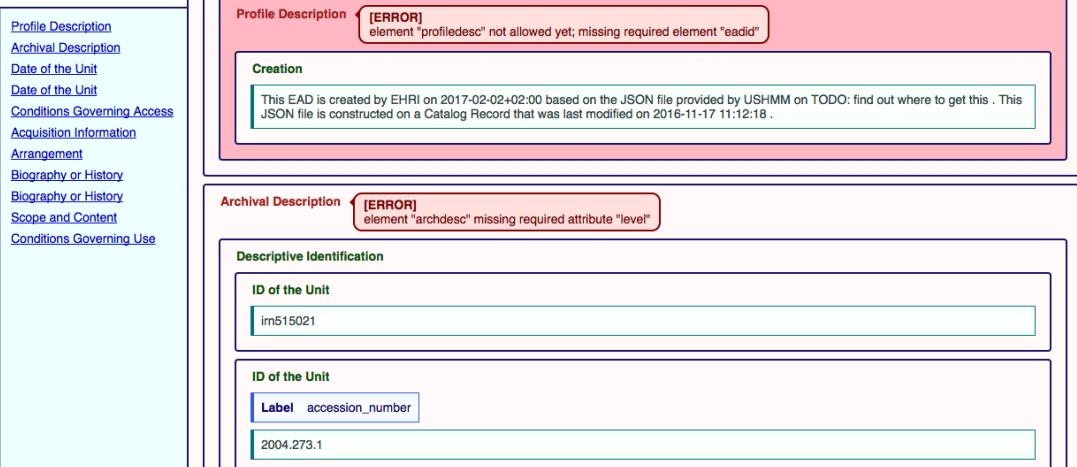
Each EAD generated file is presented in an html file. The navigation menu on the left lists all XML elements that do not comply with the EAD 2002 standard.



This screenshot shows the same EAD XML document from the previous screenshot, but with validation errors highlighted. The 'Date of the Unit' element has an error message: '[ERROR] value of attribute "type" is invalid; must be equal to "bulk" or "inclusive"'. Another error message appears below it: 'certainty approximate normal 1920/1981 type value approximately 1920-1981'. A third error message is shown for the 'normal' attribute of another 'Date of the Unit' element: '[ERROR] value of attribute "type" is invalid; must be equal to "bulk" or "inclusive"'. The EHRI logo and sidebar are visible on the left.

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, Conditions Governing Access, Acquisition Information, Arrangement, Biography or History, Scope and Content, and Conditions Governing Use. The main area displays validation results for three sections:

- Profile Description:** An error message states: "[ERROR] element "profiledesc" not allowed yet; missing required element "eadid"".
- Archival Description:** An error message states: "[ERROR] element "archdesc" missing required attribute "level"".
- Descriptive Identification:** Two entries for "ID of the Unit" are shown: "imr515021" and "Label accession_number" with value "2004.273.1".

5.9 Validate the corrections

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

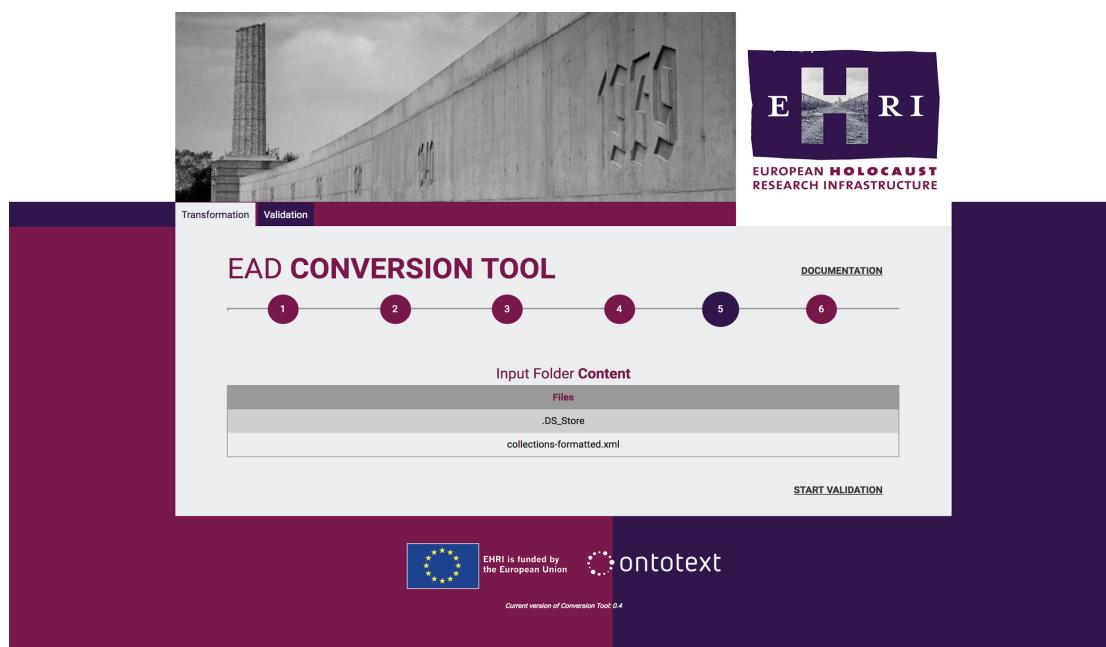
6 Validation tool

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



6.4 To start the Validation, click the Start Validation button

6.5 Explore the Validation report



The screenshot shows the user interface of the EAD Conversion Tool. At the top, there's a banner with the EHRI logo and a photograph of a modern concrete monument. Below the banner, a navigation bar has 'Transformation' and 'Validation' tabs, with 'Validation' being active. A title bar says 'EAD CONVERSION TOOL'. A progress bar at the top indicates steps 1 through 6, with step 6 being the current one. A success message 'File was Validated Successfully!' is displayed. Below this, a table titled 'Conversion report' lists one entry: 'collections-formatted' with '0' errors. Navigation buttons 'PREVIOUS' and 'NEXT' are shown, with '1' in the center. At the bottom, there are buttons for 'REPEAT VALIDATION' and 'START NEW VALIDATION'. The footer includes logos for the European Union and ontotext, and a note about the tool's version.

Transformation | Validation

EAD CONVERSION TOOL

File was Validated Successfully!

From files we created 1.

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

Show 10 entries Search:

File Name	Number of errors
collections-formatted	0

SHOULD 1 TO 1 OF 1 ENTRIES PREVIOUS 1 NEXT

REPEAT VALIDATION START NEW VALIDATION

 EHRI is funded by the European Union



Current version of Conversion Tool: 0.4