Ontotext AD
Sofia, Tsarigradsko 47A
tel +359 2 974 6160
fax: +359 2 975 3226
http://www.ontotext.com

EAD Converter

User Guide

Document Date (3rd Feb, 2017)

Table of Contents

1 Overview

EAD converter 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 EHRI guidelines and the harvesting/ingest workflow.

EAD converter 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 format;
- Preview all validation 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 Sheet.

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 converter is available under …. license. It is free and open-source.

3 Run the EAD converter as a desktop installation

The EAD converter setup and running is easy and straightforward.

3.1 On Windows OS

1. Download and unzip the EAD converter.
2. Click the *run.bat* file.
3. The EAD converter GUI automatically opens at http://localhost:8080.

3.2 On Unix OS

1. Download and unzip the EAD converter.
2. Click the *run* shell script file.
3. The EAD converter GUI automatically opens at http://localhost:8080.

4 EAD converter HOME directory

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

4.1 Input directory

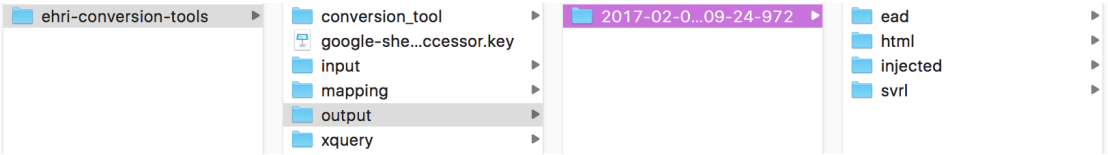
The */input* data directory is where you add the data files you want to transform.

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 validation inconsistencies from the conversion.

4.3 Output directory

The */output* data directory is where the EAD converter stores all transformed data files. They are organized into subdirectories, which names reflect their creation time. Each subdirectory 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

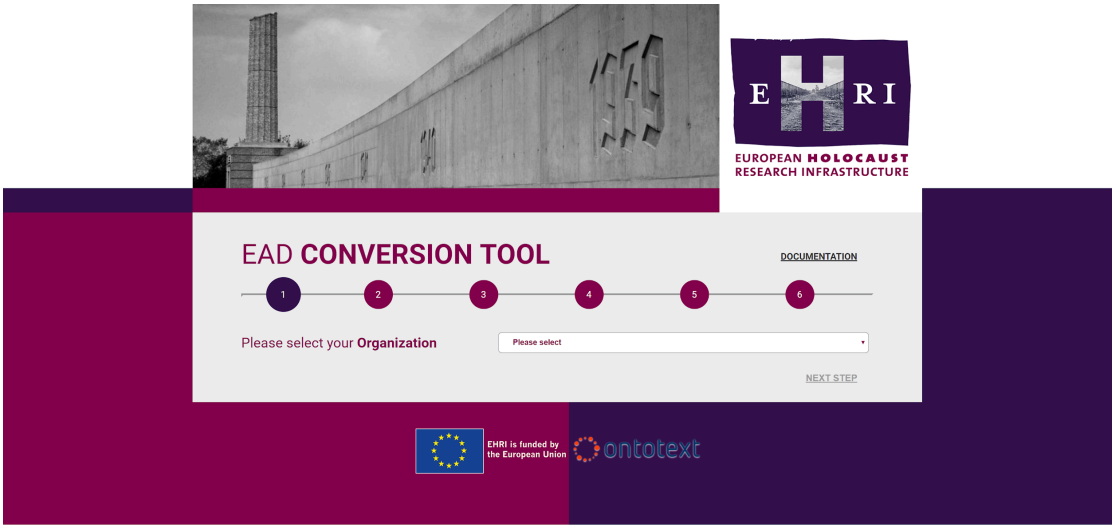
The */xquery* directory is where you can add a custom *.xqy* file to transform data files into a format different from the default EAD 2002 standard.

5 Procedure

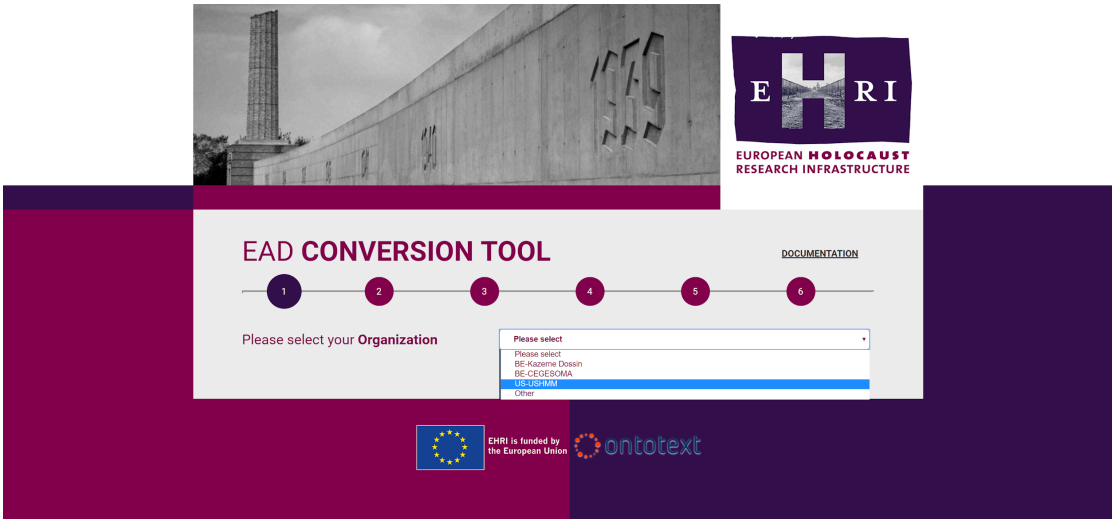
To transform your data into a well-formed EAD 2002 file, follow the steps:

5.1 Add the data files you want to transform in the *~/input* directory

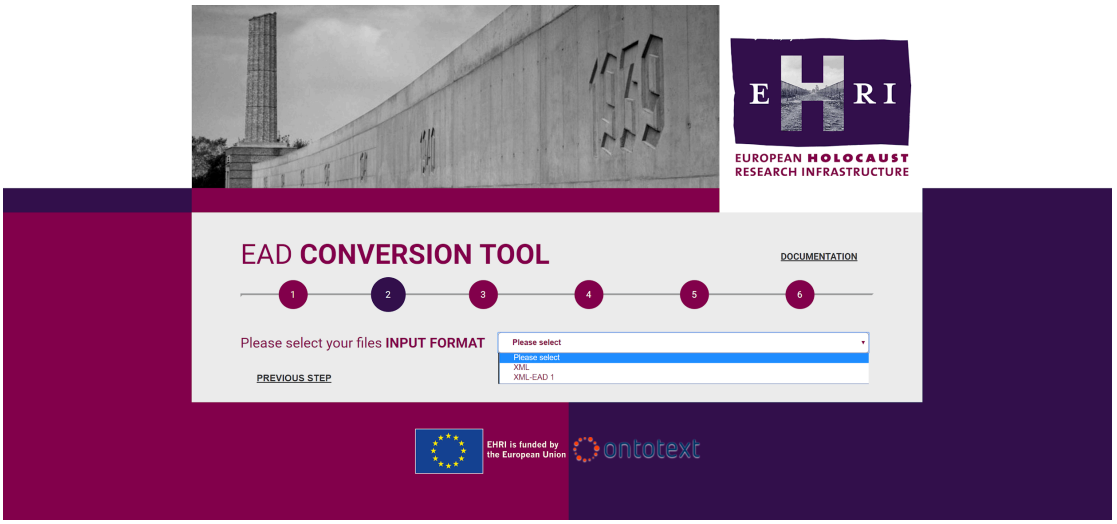
5.2 Open http://localhost:8080 in a browser



5.3 Select your organization from the drop-down list



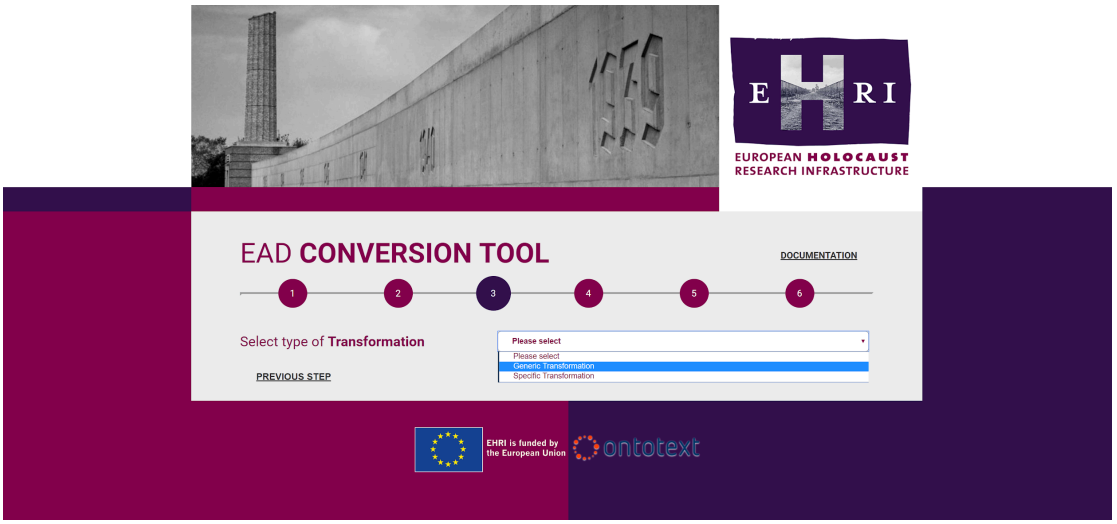
5.4 Select your files INPUT format



5.5 Select the transformation type

There are two types of transformation *Generic (default)* and *Specific*. Using the *Generic* one, you can transform your data files in the EAD 2002 format.

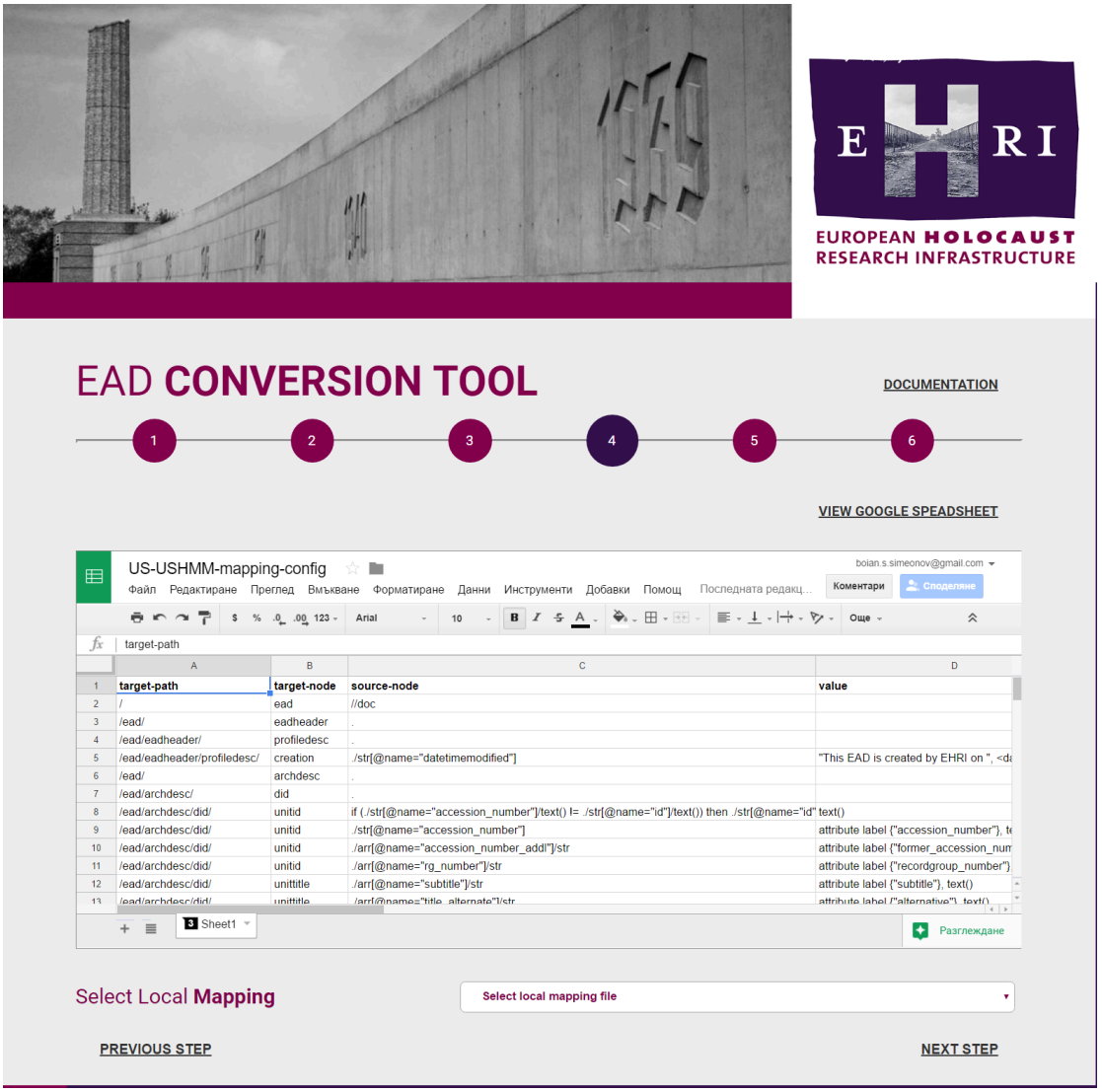
If you want to use the tool for transforming data in other formats, you should create your own xquery transformation schemes and add them to the */xquery* folder. Then, they can be used when the *Specific transformation* type is selected.



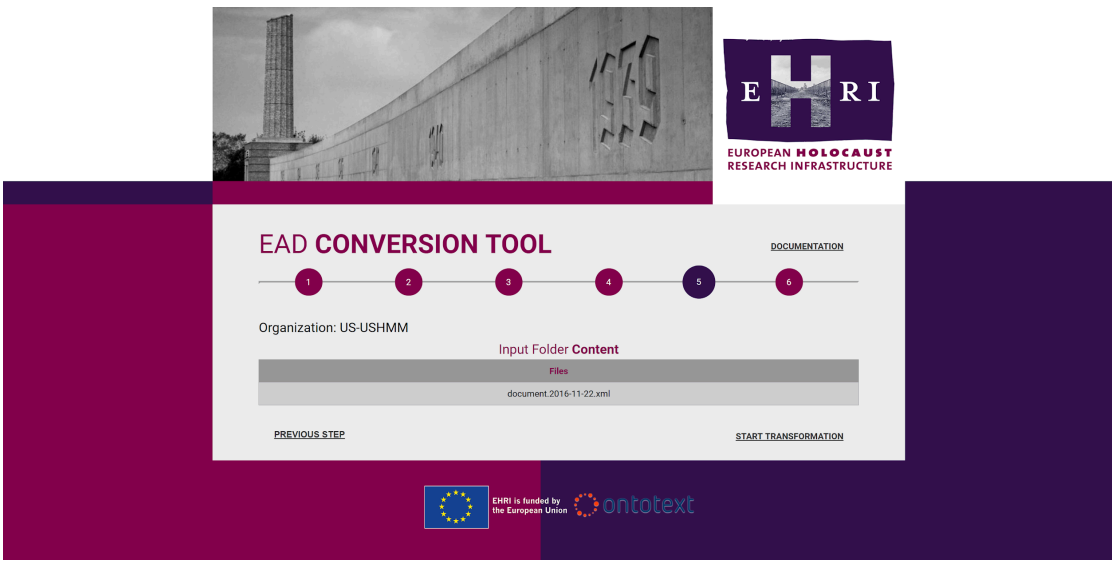
5.6 Preview/edit the mapping config file

All mapping config files are stored as Google sheets. Depending on your access rights, you can view or edit them, directly in the EAD converter UI 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.

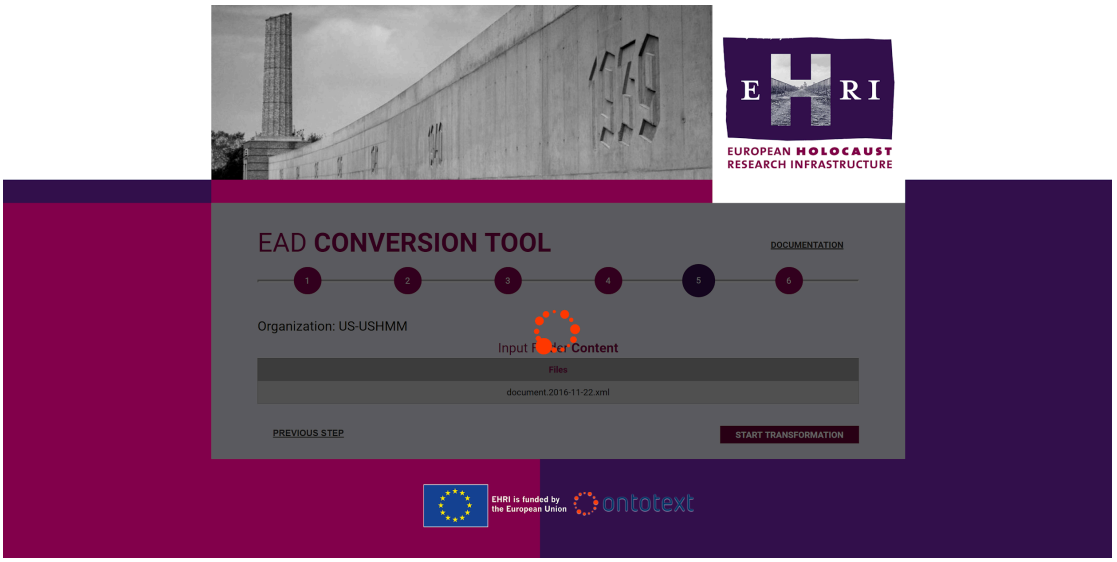


5.7 Preview the input files



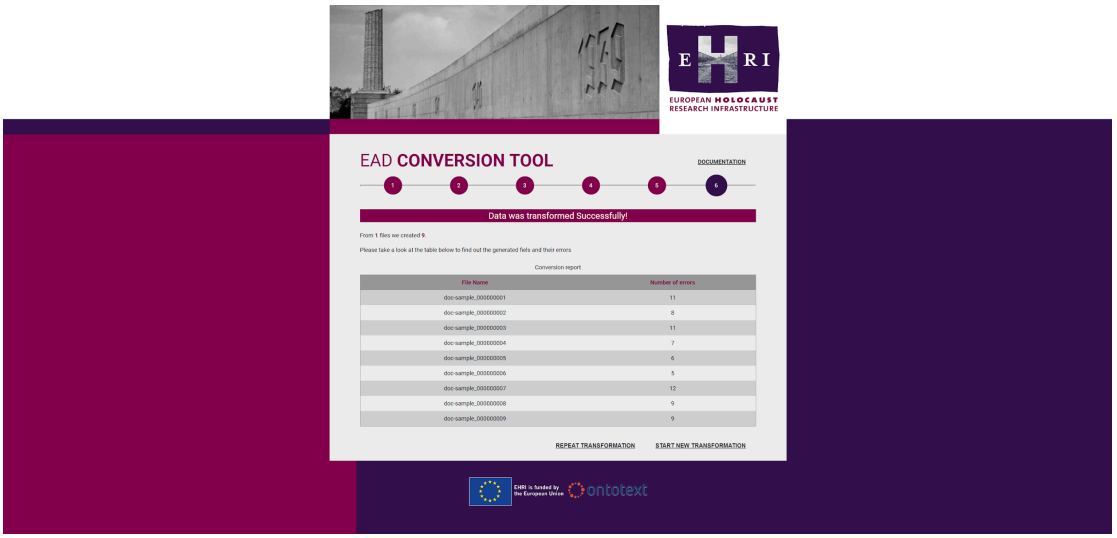
5.8 Start transformation

To start the conversion, click the Start transformation button.



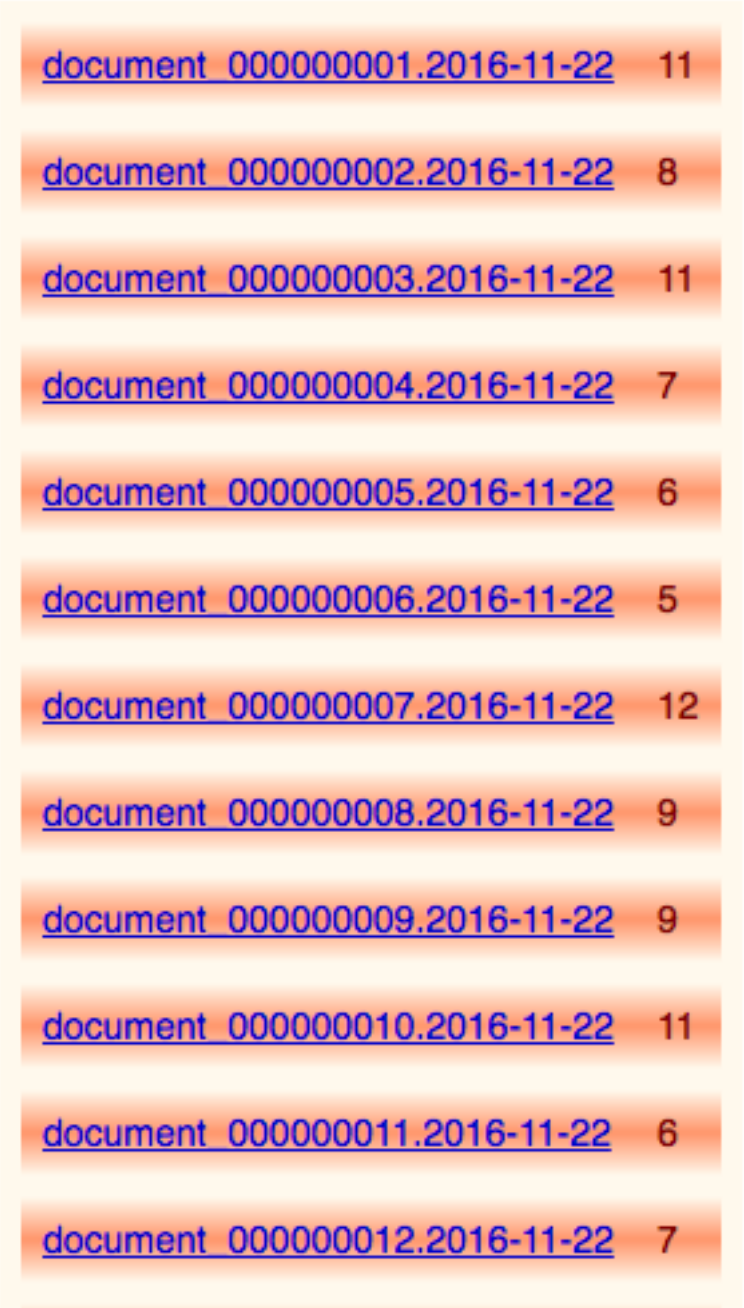
5.9 Explore the conversion report

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.



5.10 Exploring the EAD validation inconsistencies

1. Go to the *~/output/<timestamp>/html* folder and click the *index.html* file. It lists all files containing errors.

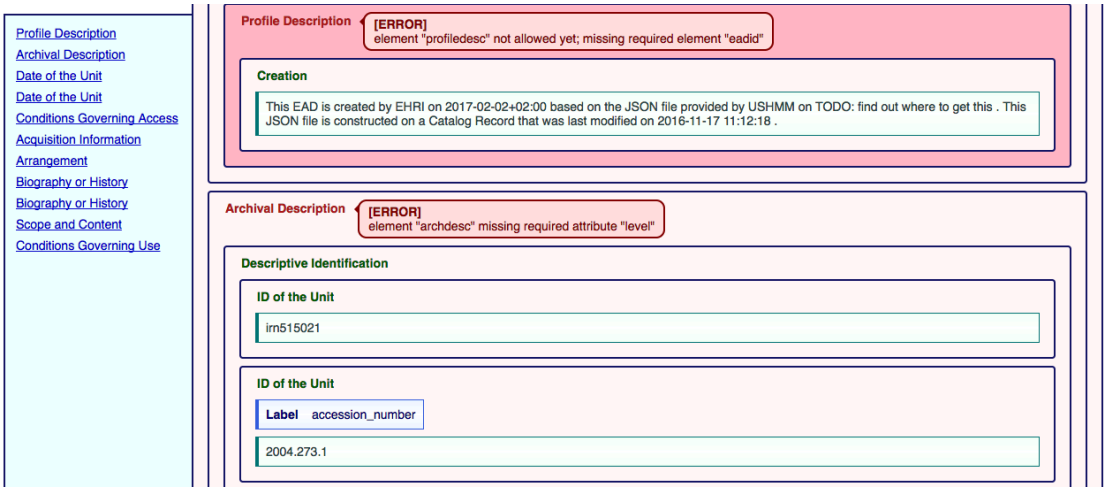


2. Click a file name to check its errors.

Each EAD generated file is presented in a user-friendly HTML format. The navigation menu on the left lists all XML elements that do not comply with the EAD 2002 standard.

3. 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 “eadid” element. In order to correct this error, you must add a “eadid” 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.



5.11 Validate the corrections

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