



## Use of the Metis Sandbox when publishing data via Europeana.eu

Archives Portal Europe (APE) Content Providers have an option within their Dashboard accounts to [convert their data to the Europeana Data Model](#) (EDM) in preparation to publishing them via Europeana.eu. This option will only be available for data that include links to digital objects and the conversion will only include those items within a collection that have such links. Any intermediate hierarchical levels and any items without links to digital objects will not be taken into account.

Before sharing the data with Europeana for publication, it should be tested via the [Metis Sandbox](#), Europeana's Content and Data Quality Checker, which also includes a preview option to see how the data would be displayed on Europeana.eu. This will be done by the central team at Archives Portal Europe Foundation (APEF) on behalf of the APE Content Providers. This document here outlines the constituent steps that are part of this action.

### 1. Create EDM ZIP file

Once the EDM conversion has been completed successfully in the APE Dashboard, a ZIP file including all EDM XML files for the converted fonds or collection can easily be created by initiating the "Download EDM" action. This can either be done by the Content Providers themselves, who would then need to forward the ZIP file(s) to the APEF team, or by the APEF team on behalf of the Content Providers after having been notified (via email to [info@archivesportaleurope.net](mailto:info@archivesportaleurope.net)) that the file(s) has/have been converted successfully.

### 2. Upload the EDM ZIP file to the Metis Sandbox

The APEF team will then upload the EDM ZIP file to the Metis Sandbox, which allows them to see if there'd be any issues when importing the data into Europeana.eu. This provides a last checkpoint before the effective delivery and also comes with a preview option to check how the data will look like in Europeana.

### 3. Settings upon upload

In the case of large collections with more than 1,000 files - the current maximum that can be processed in the Metis Sandbox, the APEF team will either split the EDM ZIP file into smaller batches to pre-process everything (especially when dealing with heterogeneous materials) or will apply a bigger "step size" to determine which records in the dataset will be imported. If the step size is  $n$ , then one in every  $n$  records encountered in the dataset will be imported, e.g., a step size of 1 (default) will import records 1, 2, 3, 4, ..., 1,000; a step size of 3 will import records 3, 6, 9, 12, ..., 3,000; a step size of 5 will import records 5, 10, 15, 20, ..., 5,000.



#### 4. Pre-processing in the Metis Sandbox

The Metis Sandbox will then pre-process the uploaded EDM files. As a first step, it will show the general statistics of how many files have successfully passed which pre-processing step from import and external validation via transformation, internal validation, normalisation and enrichment to media processing (for the digital objects linked in the metadata) and publication to Europeana's test environment.

#### 5. Reviewing the feedback from the Metis Sandbox

For each pre-processing step, in case any issues have been detected, the Metis Sandbox will provide information about which EDM files are affected by the issue found and will give an indication of what the reason for the issue might be. With this information, the APEF team can determine, whether the issue indicated can be solved centrally, e.g., when validation errors have been detected, which would fall back on the EDM conversion in the APE Dashboard, or would require changes in the source data conducted by the Content Providers themselves, e.g., any issues with the links to the digital objects.

The feedback from the Metis Sandbox also includes an overview of problem patterns detected in the metadata. These are data quality issues, currently specifically focused on the item's title and description as defined by the [EuropeanaTech Community's Data Quality Committee](#), e.g., with regard to the length of the title and description or with regard to repeated use of the same title for several items.

Last, the Metis Sandbox also provides information about which of the metadata and content quality tiers of the [Europeana Publishing Framework](#) would apply to each item in the dataset.

#### 6. Communication with the Content Provider

The APEF team will communicate a summary of the feedback from the Metis Sandbox to the Content Provider, highlighting any issues that might need their attention. If any of the issues detected prevent effective publication on Europeana.eu (e.g., when links to digital objects cannot be accessed), these will need to be fixed first and the process of the EDM conversion in the APE Dashboard and the use of the Metis Sandbox will need to be repeated with the updated data.

Other aspects such as the ones detected in the problem patterns evaluation might be inherent to the material and cannot be changed (e.g., when titles have been kept as submitted by the original records creator and cannot be amended in any way). Nevertheless, such inherent issues are useful information for the next step.



Similarly, any improvements with regard to metadata and content tiers might be aspects to work on before a future re-submission, but will usually not prevent publication (as long as data are not in tier 0).

Additionally, the APEF team will also share a link to the preview of the dataset in Europeana's test environment with the Content Provider, which usually makes it easier to assess whether all relevant information has been transformed and is displayed as expected. As the Metis Sandbox is a test environment, any data older than 7 days will be removed to better manage the space available. This also affects the previews, so checking these is a time-bound action.

#### 7. Communication with the Europeana Data Publishing Services team

When everything goes well and once the Content Providers have approved the data as previewed in the Metis Sandbox, the same ZIP file used here will be submitted officially to Europeana's Data Publishing Services team for processing. Publication of any new submissions will usually get scheduled into one of the team's sprints (2 to 3 weeks cycles) and the APEF team will be informed once the data have been published on Europeana.eu. Once publication is official, a link to the data on Europeana.eu will be sent to the Content Providers to complete the sharing process.

#### 8. Processing of data from known Content Providers and re-submission of data

When the same Content Provider submits further datasets, some of the steps above (especially steps 4 to 6) might not be required in that much detail as any systemic issues with this Content Provider's data should already be known from previous submissions and should have been dealt with before as necessary. The APEF team will of course still go through steps 4 and 5, but step 6 might mainly consist of sharing the link to the data's preview.

The same would apply to any data that has been updated, e.g., by adding further URIs for contextual entities or by including more links to digital objects or simply by correcting typos and other smaller aspects of the metadata, and is to be re-submitted. In this case, step 3 might simply apply the default step size and step 7 might mainly indicate anything that's new compared to the previous version of the data that was published on Europeana.eu.