

Archaeology in Europeana:

Quality assurance, enrichment and publishing



Our priorities

We want to make Europe's rich archaeological and architectural heritage easier for people to discover and more rewarding to use

Our emphasis is on **quality** of content and metadata over quantity

Monitoring the quality is an important part of the process



What do we mean by quality?

The Europeana Publishing Framework specifies requirements for content and metadata, these covers:

- The link provided to the content, the format and resolution of the content, or to the content in a viewer or web-page
- Rights labelling
- The metadata to include and how to support search by subject, place, date and agents
- Title and description
- Labelling the language of the metadata

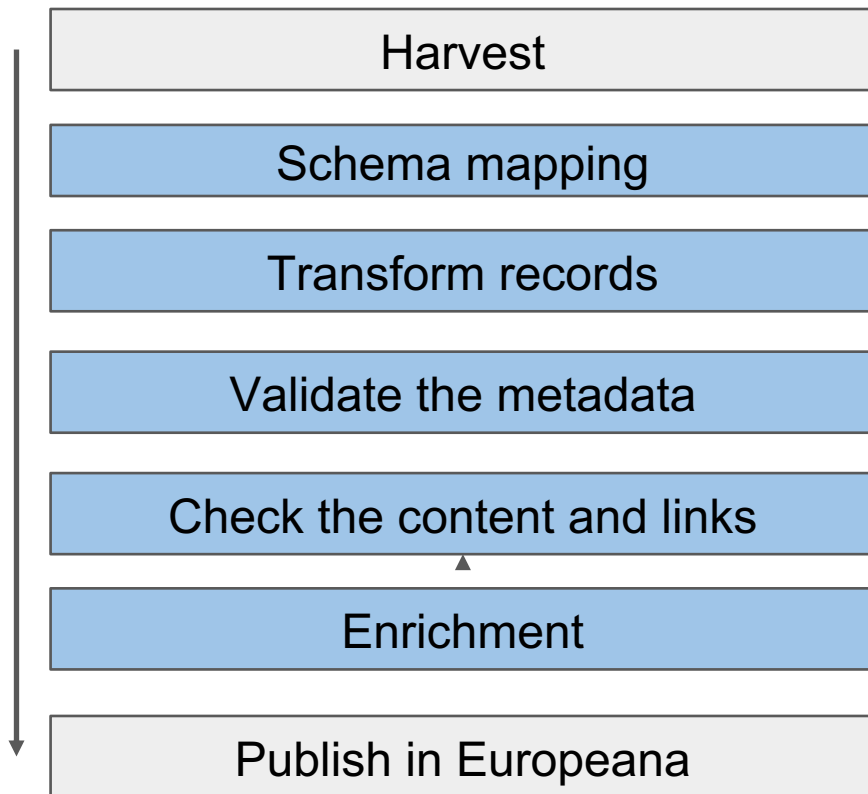
Monitoring happens throughout the content workflow

As you progress through the content work flow quality issues may be identified



Remember there are people at CARARE available to help

How do we monitor quality?



Actions: Mapping

MINT allows you to map the fields in your datasets to elements in the target scheme.

MINT will tell you if your mapping is valid according to the rules of the schema

This means that a field is mapped to every mandatory element

The screenshot displays the MINT Mapping interface for the dataset 'Beeldbank_EDM_10Apr2019'. The interface is divided into two main panels. The left panel, titled 'Search in input schema', shows a tree view of the source schema elements under the 'rdf:RDF' root. The right panel, titled 'rdf:RDF', shows a list of target schema elements under the 'edm:ProvidedCHO' root. Each target element has a corresponding input field for mapping, with a dropdown menu to select the source field. The status of each mapping is indicated by a green checkmark (valid) or a yellow star (warning).

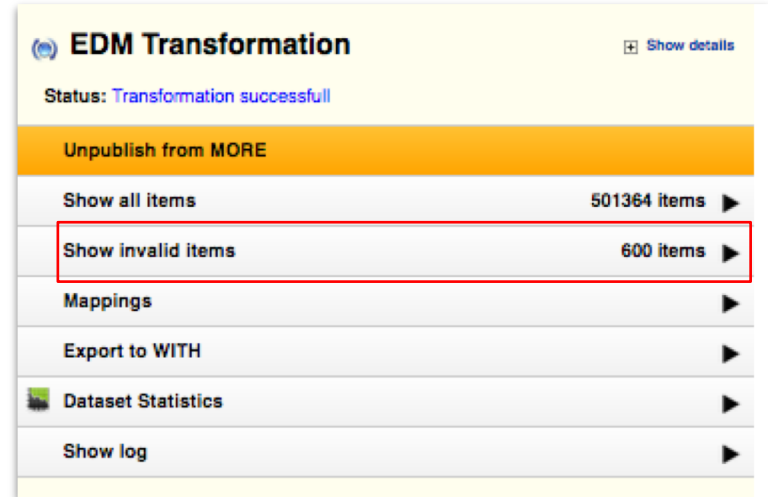
Target Element	Source Element	Status
edm:ProvidedCHO	structural	
dc:contributor	dc:contributor	Valid
dc:coverage	unmapped	Warning
dc:creator	dc:creator	Valid
dc:date	dc:date	Valid
dc:description	dc:description	Valid
dc:format	unmapped	Warning
dc:identifier	dc:identifier	Valid
dc:language	unmapped	Warning
dc:publisher	unmapped	Warning
dc:relation	unmapped	Warning
dc:rights	unmapped	Warning
dc:source	edm:dataProvider	Valid
dc:subject	dc:subject	Valid

Actions: Transformation

Your mapping produces an XSLT which can be used to transform all the records in your set to the target scheme

Validating the transformed dataset shows whether there are any invalid items

This means individual records miss mandatory data



The screenshot displays the 'EDM Transformation' interface. At the top, the status is 'Transformation successful'. Below this, there is a list of actions. The 'Show invalid items' action is highlighted with a red box, indicating 600 items. Other actions include 'Show all items' (501364 items), 'Mappings', 'Export to WITH', 'Dataset Statistics', and 'Show log'.

Action	Count
Show all items	501364 items
Show invalid items	600 items
Mappings	
Export to WITH	
Dataset Statistics	
Show log	

Actions: Review your dataset and the mapping

Reviewing the invalid items should show which data is missing

Common issues include missing data in a mandatory element. All records must contain:

- Description OR title
- Subject, type, spatial OR temporal
- edm:isShownAt OR edm:isShownBy
- edm:rights, edm:provider and edm:dataProvider

Sometimes it is possible to fulfil the requirements by editing the mapping, sometimes you will need to go back to your database



Think about the quality!

Common mistakes that affect data quality:

- Adding unuseful data to a field to fulfil a mandatory requirement - n/a, unknown, tbc. (whether in the mapping or the source database)
- Double mapping a field to fulfil a mandatory requirement



Try to add value and improve the quality

Process what you can now and make an update to add the missing records once they have been improved at home



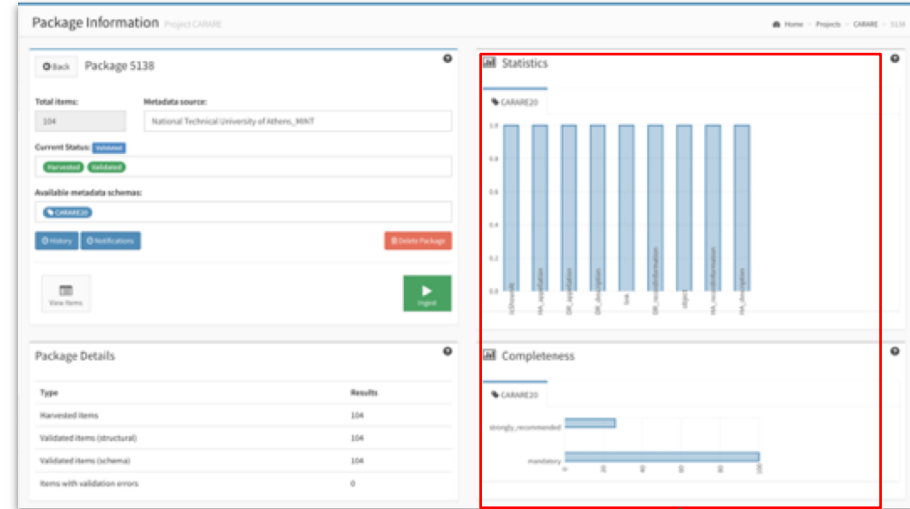
MORe quality monitoring

In the CARARE aggregation service content is ingested on MORe (either directly or from MINT)

Metadata is transformed to EDM on ingestion and records are validated against the EDM schema rules

Record completeness is assessed

EDM records are available to preview



MORe: Content analysis - images

Image Resolution

EDM

Resolution (WxH)	Count
1024x1536	6
1280x960	4
1536x1024	5
823x576	1
960x1280	15

Image Type Information

EDM

Image Type	Count
image/jpeg	31

Review the resolution and type of image-based content and their tier status

Consider enrichment

You can improve the quality of your metadata by running an enrichment service, for example to:

- Add language labels to your metadata
- Find place names
- Add geographic coordinates
- Run vocabulary matching services
- Add broad subject concepts

Available enrichment micro-services
List of available enrichment micro-services. Click on the "More info" link on each micro-service to get more information.

Service Name	Description	Status
Language	[EDM] Language Identification	Deactivate
GeoNorm	[EDM] Geo Normalization	Deactivate
LoGeo	[EDM] GeoLocation Enrichment	Deactivate
Geocode	[EDM] Geonames Geocoding	Deactivate
Rev.Geocode	[EDM] Geonames Rev. Geocoding	Deactivate
Vocabulary	[EDM] Vocabulary	Deactivate
Vocab. match	[EDM] Vocabulary Matching	Deactivate
Back Link	[EDM] Background Link	Activate
DaiGazeteer	[EDM] DaiGazeteer	Deactivate

Publishing

Once you've taken your dataset through all of the steps and you're happy with the results you can '**publish**' (or send the data to Europeana for ingestion)

This may not be the end of the quality monitoring story

- The project will be monitoring your content as it goes through to publication
- Ingestion by Europeana can also result in queries and requests for improvement



Publishing is a cycle

Publishing in Europeana is not a one-off event

The workflow has stages, checks and opportunities to go back and improve your content and metadata

Once you've published your collection successfully in Europeana, you can come back to CARARE and make an update to add new content or to make improvements.

Contact us! We're here to help

Remember CARARE has a helpdesk

We can give one-to-one support with metadata mapping, enrichment, use of vocabularies, geocoding, Europeana's publishing frameworks and rights labels

We can offer support with use of the aggregation services MINT and MORE, and with publishing to Europeana



Acknowledgements

These training materials were developed in collaboration with the Europeana Archaeology project and with funding by the European Commission under the *Connecting Europe Facility (CEF) Telecom Programme*

