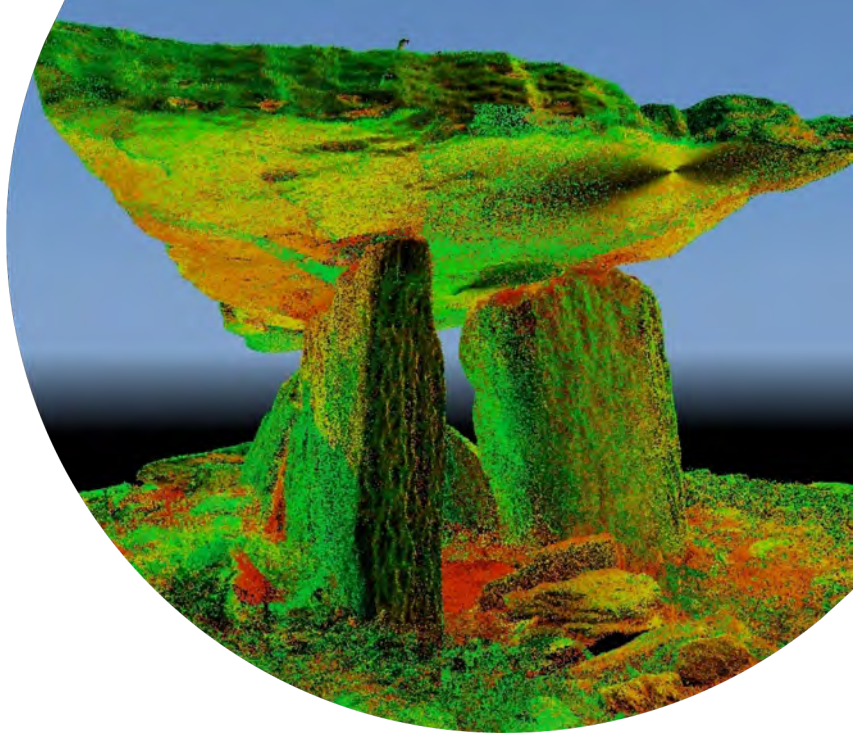
European databases in cultural heritage: making connections

Kate Fernie, CARARE

EuroWEB
21 November 2020

About CARARE



CARARE is a non-profit association whose main objective is advancing professional practice and fostering appreciation of the digital archaeological and architectural heritage. CARARE:

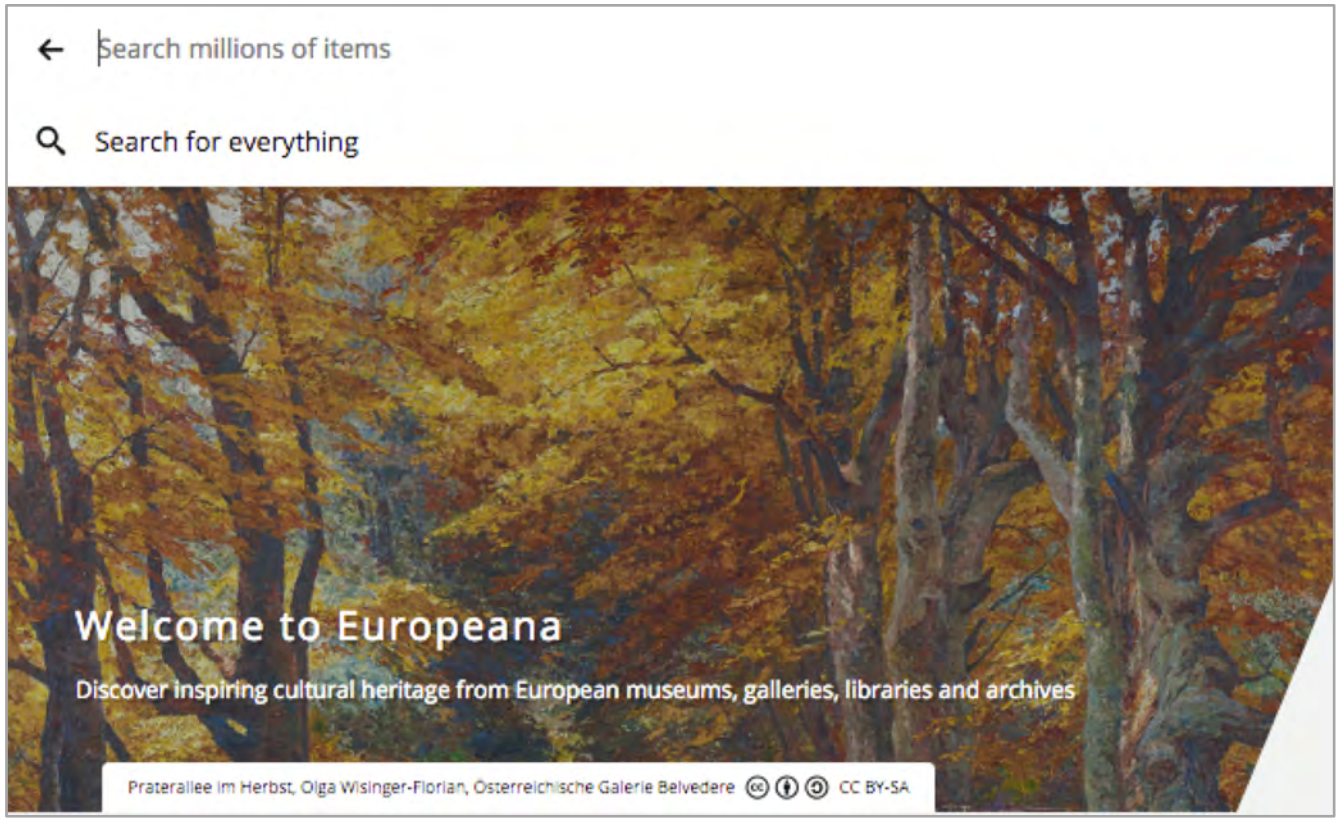
- Has a growing membership
- Provides advice, guidance and training
- Provides technical services to enable institutions to share their digital content with users via Europeana
- Is a partner in ARIADNE Plus

https://www.carare.eu

Introduction

This session will introduce:

- Europeana
- ARIADNE
- SEADDA
- Benefits and possibilities

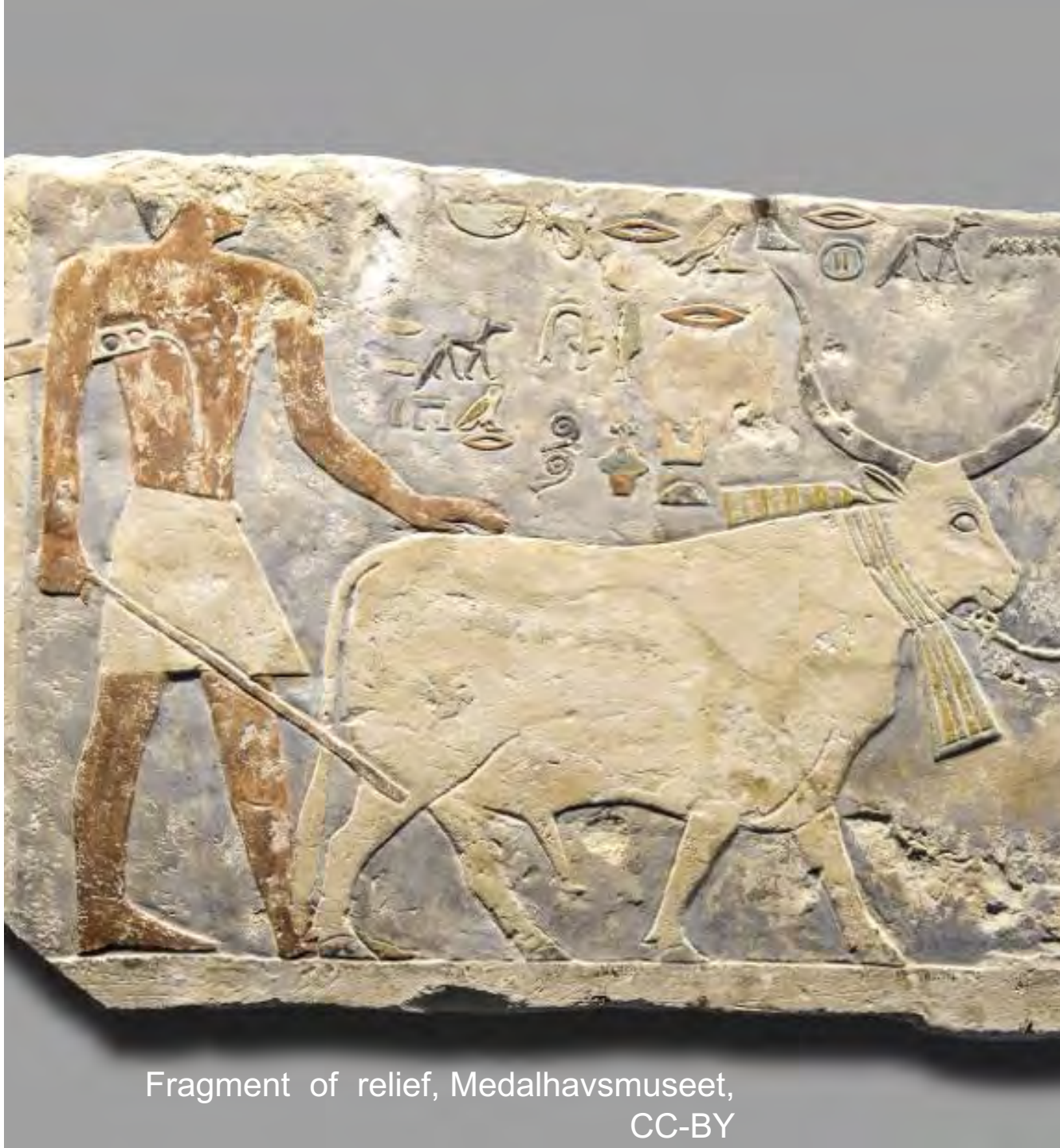


https://www.europeana.eu/

What is Europeana?

- Europe's platform for digital cultural heritage funded by the European Commission
- 50+ million items – photographs, videos, music, newspapers, text (books, reports, diaries, letters), spoken word and 3D
- Archaeology themed collections c. 1.5 million items

https://www.europeana.eu



Fragment of relief, Medalhavsmuseet, CC-BY

National Museum Stockholm,
Wiki loves Monuments, CC-BY-SA

Opening up collections

- Around 4000 cultural institutions, research institutions, private companies and others are sharing digital assets via Europeana
- All types of cultural heritage – objects, books, newspapers, archives, maps, drawings, photographs, buildings, monuments, natural history, costume
- Represented in 5 main formats - images, text documents, sound files, videos and 3D models

Oppidum ibérico de la Plaza de Armas de Puente Tablas, Jaen, Spain, Instituto de Arqueología Ibérica-Universidad de Jaén, CC-BY-NC-ND

What sort of data?

- Databases, Inventories, collection catalogues,
- Datasets, excavation archives, scientific data
- Reports, publications
- Media sets – photos, drawings, videos, 3D models, sound recordings, and more
- (Geographic information, GIS)
- (Vocabularies)

Background challenges

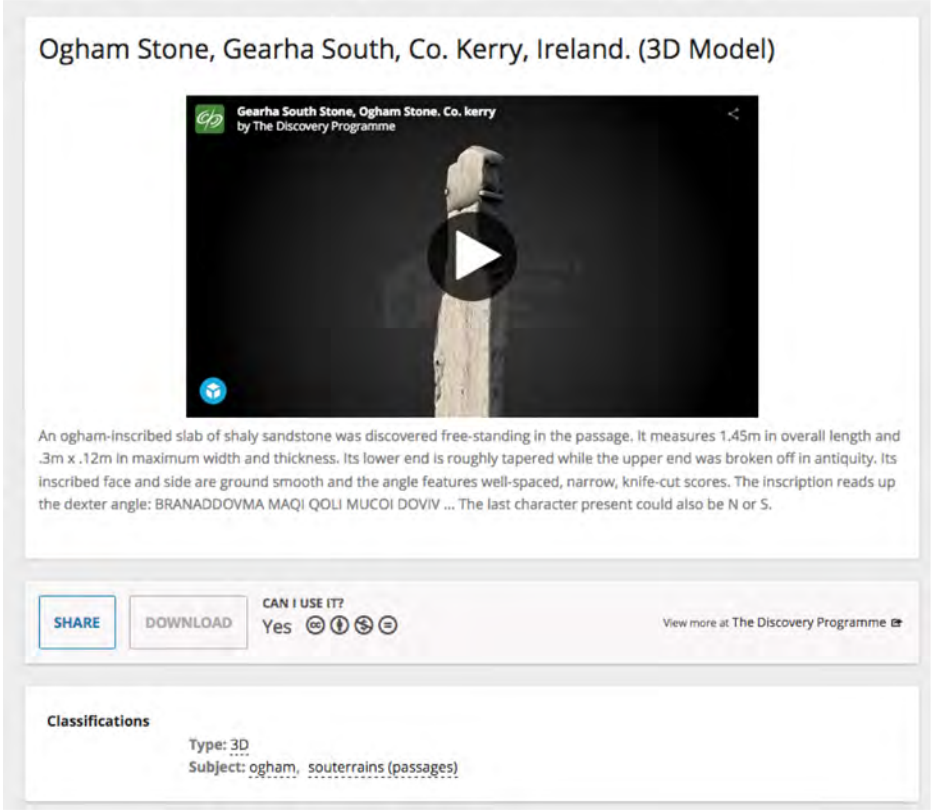
- Many different organisations and individuals are involved
- Heritage agencies, museums, archives, libraries and researchers all have different ways of describing objects
- There are different data standards
- Data is captured in many different languages, using different vocabularies, time periods and map systems

Geophysical survey at Wicina-osada obronna, Narodowy Instytut Dziedzictw

Europeana's Approach

Europeana has worked with partners on standardizing how it handles data and media formats. It has defined :

- A data model (EDM)
- A publishing framework with criteria for content and metadata
- A rights framework for content and metadata
- A data exchange agreement



https://www.europeana.eu/portal/en/record/2020741/OGHAM_CORCU_0001.html

Sharing data with Europeana: step by step

The basic requirements for sharing data with Europeana are these:

- Your content is digital and is published online (it can be pointed to by a web link)
  - The copyright status has been evaluated and how people may use the content is expressed in a rights statement/access licence
- Descriptions of the objects (metadata) is available and can be exported
  - The metadata is licenced for free reuse under a CCO licence

The next step is to identify an aggregator (like CARARE) which can help you to transform your metadata to EDM and share it with Europeana.

Aggregating metadata

Aggregators like CARARE work with institutions to help transform their metadata into the standardized formats used in web portals

- Metadata is provided in native formats or in formats such as Dublin Core
- We make crosswalks (or mappings) to enable the metadata to be transformed for ingestion by the portal

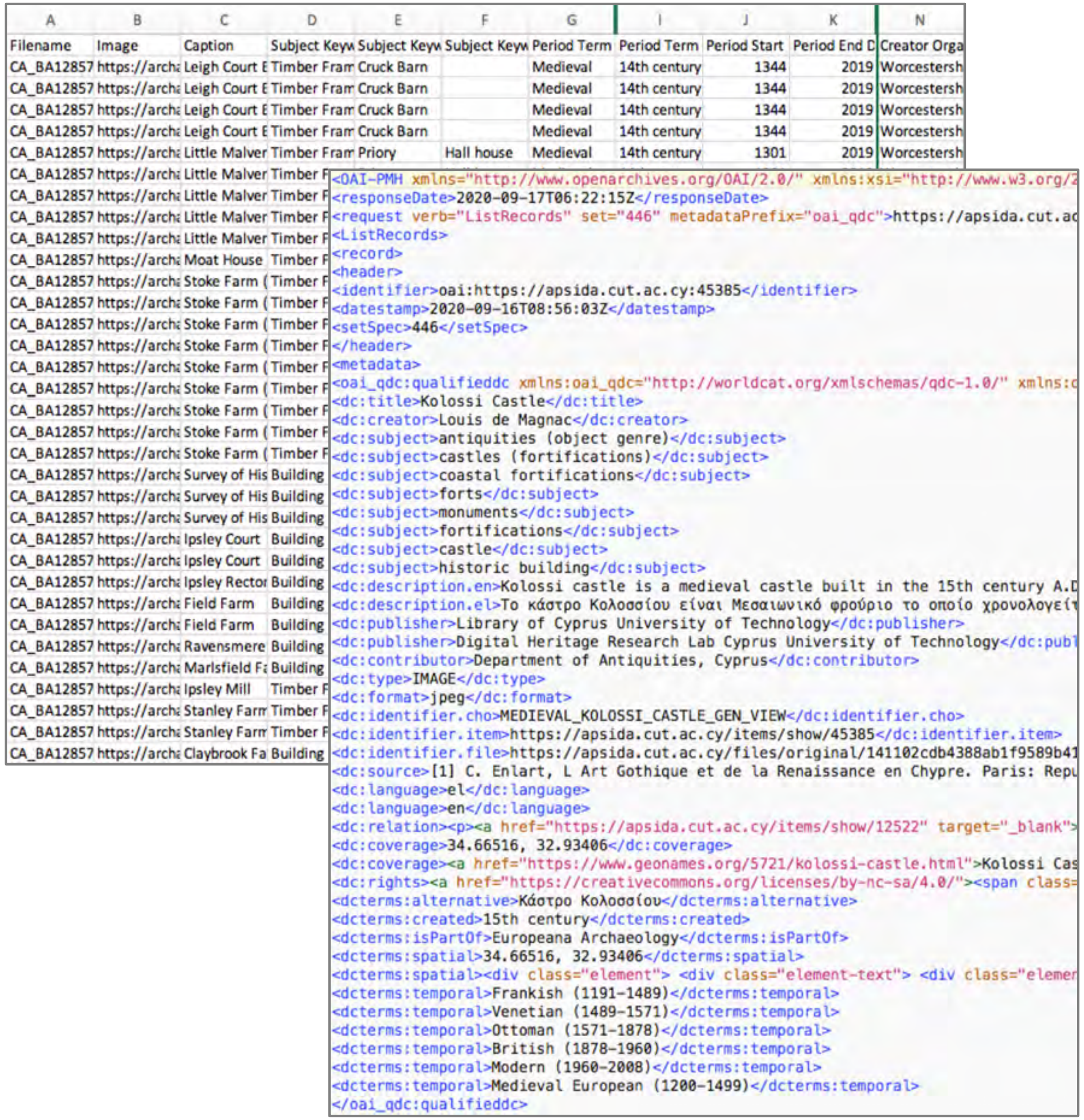
Metadata

Metadata is provided to us in a range of formats:

- XML records harvested from OAI-PHM repositories
- XML records in zip files
- CSV files

The data itself may be structured

- In a local schema
- In standard schema:
    - EDM, Lido, CARARE, Dublin Core, Omeka XML
    - EAD, MARC

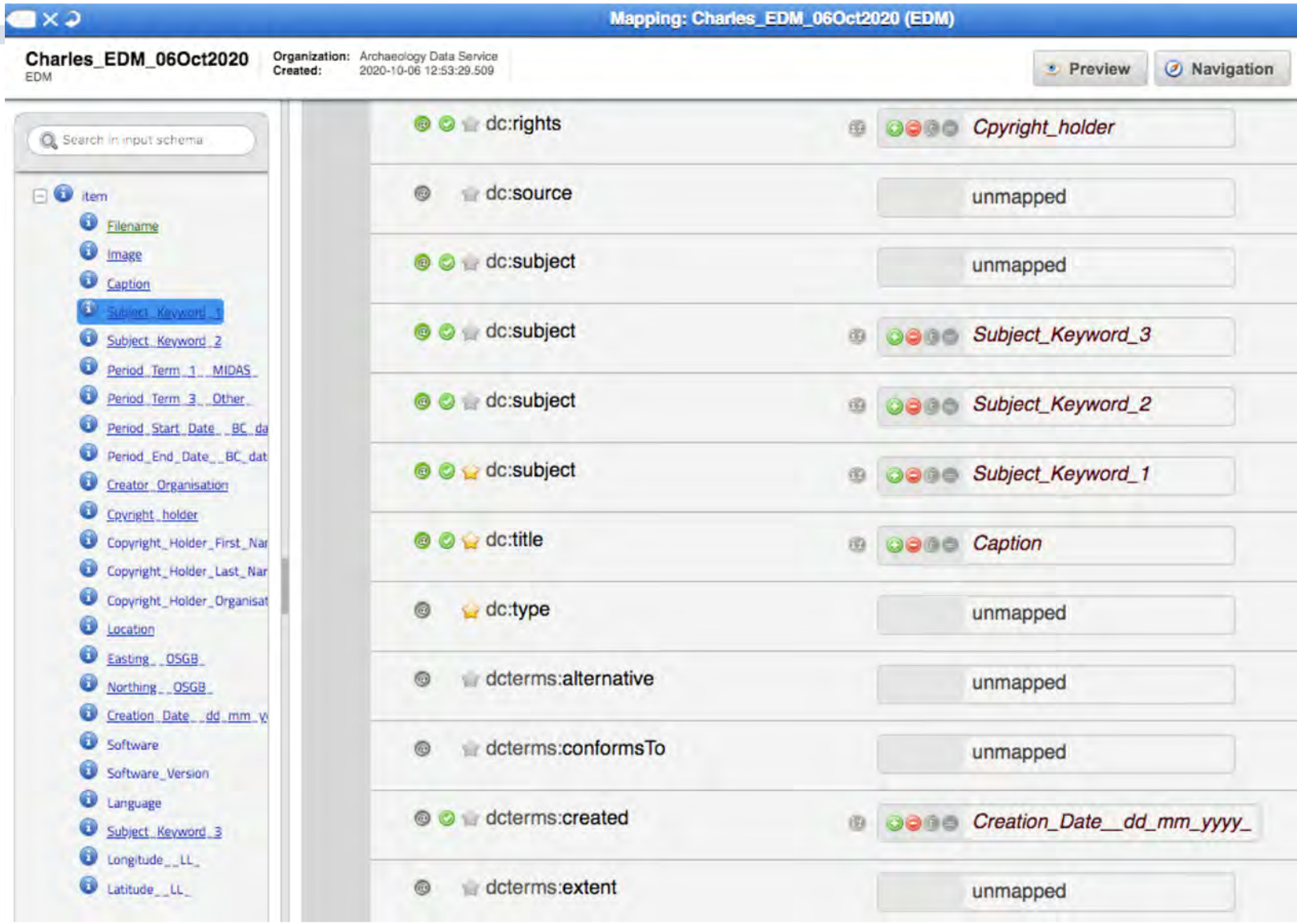


A	B	C	D	E	F	G	I	J	K	N
Filename	Image	Caption	Subject Keyw	Subject Keyw	Subject Keyw	Period Term	Period Term	Period Start	Period End D	Creator Orga
CA_BA12857	https://archa	Leigh Court E	Timber Fram	Cruck Barn		Medieval	14th century	1344	2019	Worcestersh
CA_BA12857	https://archa	Leigh Court E	Timber Fram	Cruck Barn		Medieval	14th century	1344	2019	Worcestersh
CA_BA12857	https://archa	Leigh Court E	Timber Fram	Cruck Barn		Medieval	14th century	1344	2019	Worcestersh
CA_BA12857	https://archa	Leigh Court E	Timber Fram	Cruck Barn		Medieval	14th century	1344	2019	Worcestersh
CA_BA12857	https://archa	Little Malver	Timber Fram	Priory	Hall house	Medieval	14th century	1301	2019	Worcestersh
CA_BA12857	https://archa	Little Malver	Timber F							
CA_BA12857	https://archa	Little Malver	Timber F							
CA_BA12857	https://archa	Little Malver	Timber F							
CA_BA12857	https://archa	Little Malver	Timber F							
CA_BA12857	https://archa	Moat House	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Stoke Farm (	Timber F							
CA_BA12857	https://archa	Survey of His	Building							
CA_BA12857	https://archa	Survey of His	Building							
CA_BA12857	https://archa	Survey of His	Building							
CA_BA12857	https://archa	Ipsley Court	Building							
CA_BA12857	https://archa	Ipsley Court	Building							
CA_BA12857	https://archa	Ipsley Rector	Building							
CA_BA12857	https://archa	Field Farm	Building							
CA_BA12857	https://archa	Field Farm	Building							
CA_BA12857	https://archa	Ravensmere	Building							
CA_BA12857	https://archa	Marlsfield Fa	Building							
CA_BA12857	https://archa	Ipsley Mill	Timber F							
CA_BA12857	https://archa	Stanley Farm	Timber F							
CA_BA12857	https://archa	Stanley Farm	Timber F							
CA_BA12857	https://archa	Claybrook Fa	Building							


```
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2
<responseDate>2020-09-17T06:22:15Z</responseDate>
<request verb="ListRecords" set="446" metadataPrefix="oai_qdc">https://apsida.cut.ac
<ListRecords>
<record>
<header>
<identifier>oai:https://apsida.cut.ac.cy:45385</identifier>
<datestamp>2020-09-16T08:56:03Z</datestamp>
<setSpec>446</setSpec>
</header>
<metadata>
<oai_qdc:qualifieddc xmlns:oai_qdc="http://worldcat.org/xmlschemas/qdc-1.0/" xmlns:d
<dc:title>Kolossi Castle</dc:title>
<dc:creator>Louis de Magnac</dc:creator>
<dc:subject>antiquities (object genre)</dc:subject>
<dc:subject>castles (fortifications)</dc:subject>
<dc:subject>coastal fortifications</dc:subject>
<dc:subject>forts</dc:subject>
<dc:subject>monuments</dc:subject>
<dc:subject>fortifications</dc:subject>
<dc:subject>castle</dc:subject>
<dc:subject>historic building</dc:subject>
<dc:description.en>Kolossi castle is a medieval castle built in the 15th century A.D
<dc:description.el>Το κάστρο Κολοσσίου είναι Μεσαιωνικό φρούριο το οποίο χρονολογεί
<dc:publisher>Library of Cyprus University of Technology</dc:publisher>
<dc:publisher>Digital Heritage Research Lab Cyprus University of Technology</dc:publ
<dc:contributor>Department of Antiquities, Cyprus</dc:contributor>
<dc:type>IMAGE</dc:type>
<dc:format>jpeg</dc:format>
<dc:identifier.cho>MEDIEVAL_KOLOSSI_CASTLE_GEN_VIEW</dc:identifier.cho>
<dc:identifier.item>https://apsida.cut.ac.cy/items/show/45385</dc:identifier.item>
<dc:identifier.file>https://apsida.cut.ac.cy/files/original/141102cdb4388ab1f9589b41
<dc:source>[1] C. Enlart, L Art Gothique et de la Renaissance en Chypre. Paris: Repu
<dc:language>el</dc:language>
<dc:language>en</dc:language>
<dc:relation><p><a href="https://apsida.cut.ac.cy/items/show/12522" target="_blank">
<dc:coverage>34.66516, 32.93406</dc:coverage>
<dc:coverage><a href="https://www.geonames.org/5721/kolossi-castle.html">Kolossi Cas
<dc:rights><a href="https://creativecommons.org/licenses/by-nc-sa/4.0/"><span class=
<dcterms:alternative>Κάστρο Κολοσσίου</dcterms:alternative>
<dcterms:created>15th century</dcterms:created>
<dcterms:isPartOf>Europeana Archaeology</dcterms:isPartOf>
<dcterms:spatial>34.66516, 32.93406</dcterms:spatial>
<dcterms:spatial><div class="element"> <div class="element-text"> <div class="elemen
<dcterms:temporal>Frankish (1191-1489)</dcterms:temporal>
<dcterms:temporal>Venetian (1489-1571)</dcterms:temporal>
<dcterms:temporal>Ottoman (1571-1878)</dcterms:temporal>
<dcterms:temporal>British (1878-1960)</dcterms:temporal>
<dcterms:temporal>Modern (1960-2008)</dcterms:temporal>
<dcterms:temporal>Medieval European (1200-1499)</dcterms:temporal>
</oai_qdc:qualifieddc>
```

Making a crosswalk

- MINT metadata mapping tool allows us to
- Map the elements in the native schema are mapped to EDM
- The mapping is then used to transform the provided data to EDM



Archaeology Data Service, The Charles Archive metadata

Europeana Data Model: a very quick intro

EDM:

- Is based on existing standards with extensions
- Is an RDF based model written in XML
- Follows Linked data principles
  - Data can be provided as resources with URIs not only as strings of text
  - Encourages the use of LOD vocabularies
- Accommodates the needs of different communities (libraries, museums, sound archives, etc)

Europeana record example

Title: Cover (food service)

Description: Roughly square cover made of different textiles pieced together and decorated with embroidery, mirrors and a cowrie shell edging…

Subject: Food and Feasting, covers (food service), cowrie shell, textiles

Coverage: Kashmir

Source: Horniman Museum

Identifier: nn14357

https://pro.europeana.eu/page/edm-documentation

Some challenges

- Metadata mapping is rarely straight forwards
- Metadata models are complex and reflect subtle difference in world view
  - Systems are designed to serve specific purposes – the metadata has a local context and audience
- How people record data in a system may differ subtly from one to another and can change over time
  - ➢ The provided metadata can be a bit messy
- A good metadata mapping can rise to these challenges and offers opportunities to enrich the data

Data wrangling

Mapping the metadata involves a process of structuring, cleaning sometimes adding to the provided metadata. For example:

- Adding information needed for the portal (e.g. name of the organization, language of metadata)
- Splitting multiple values (e.g. subject concepts listed in a single element)
- Cleaning data, e.g. trimming blank spaces, replacing/removing values like n/a inserted to complete in elements

Neolithic perforated object or macehead, London, UK
Portable Antiquities Scheme, CC-BY

Enriching the data

<dc:title lang="**es**">Necrópolis</dc:title>

Language identification

Provided Cultural Heritage Object

"Hornos de Peal, Jaén"

<dc:subject rdf:resource=http://vocab.getty.edu/aat/300054328/>
< dc:subject rdf:resource= http://vocab.getty.edu/aat/300000810/>
< dc:subject rdf:resource= http://vocab.getty.edu/aat/300305500/>

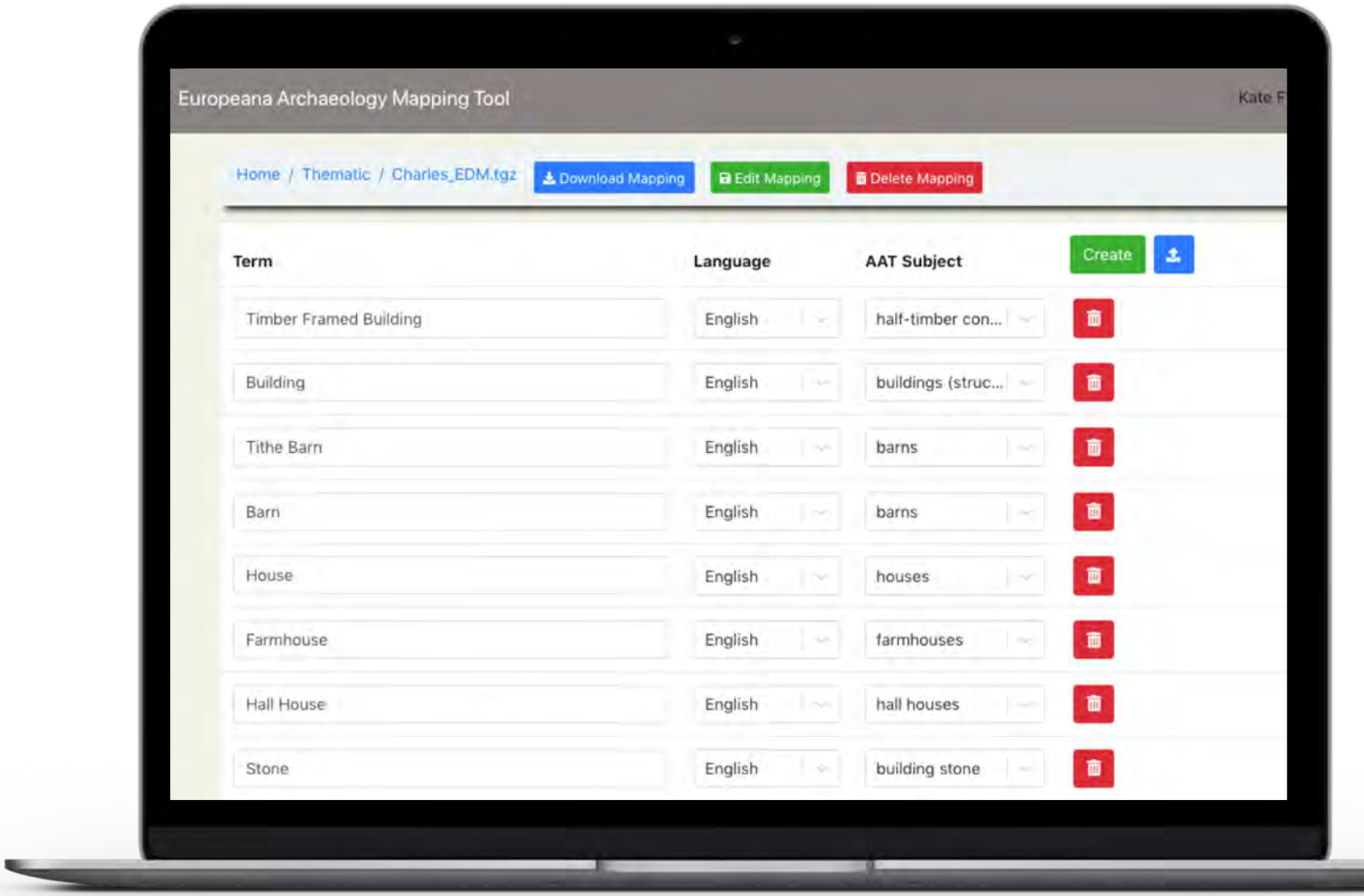
During mapping we can:

- Add broad subject terms from Linked Open Data sources, e.g. the Getty's Art and Architecture Thesaurus

Images: Instituto Universitario de Investigación en Arqueología Ibérica

Enriching using LOD

- Most provided metadata contains keywords in the national language
- We can ask organisations to map their keywords to the equivalent term in the Getty AAT (or similar) and use this to enrich the data
- Adds alternate language labels for the concept which improves discovery



Europeana Archaeology Vocabulary mapping service

Increasing interoperability: Place

- Spatial coordinates - converted from national mapping systems to the World Geodetic System (WGS84)
- We encourage the referencing of modern place names to Geonames which adds coordinates and regional context
- Historical place names can be referenced to Pleiades (https://pleiades.stoa.org/places) or Pelagios (https://pelagios.org/)

Increasing iteroperability: Time

Perhaps the most difficult concept:

- Periods have different time spans by region (e.g. the Roman period in Britain vs the Roman period in Italy)
- Different conventions are in use for expressing time periods
- Typically there is a lot of variation in the way dates and time periods are completed in databases (abbreviations etc)

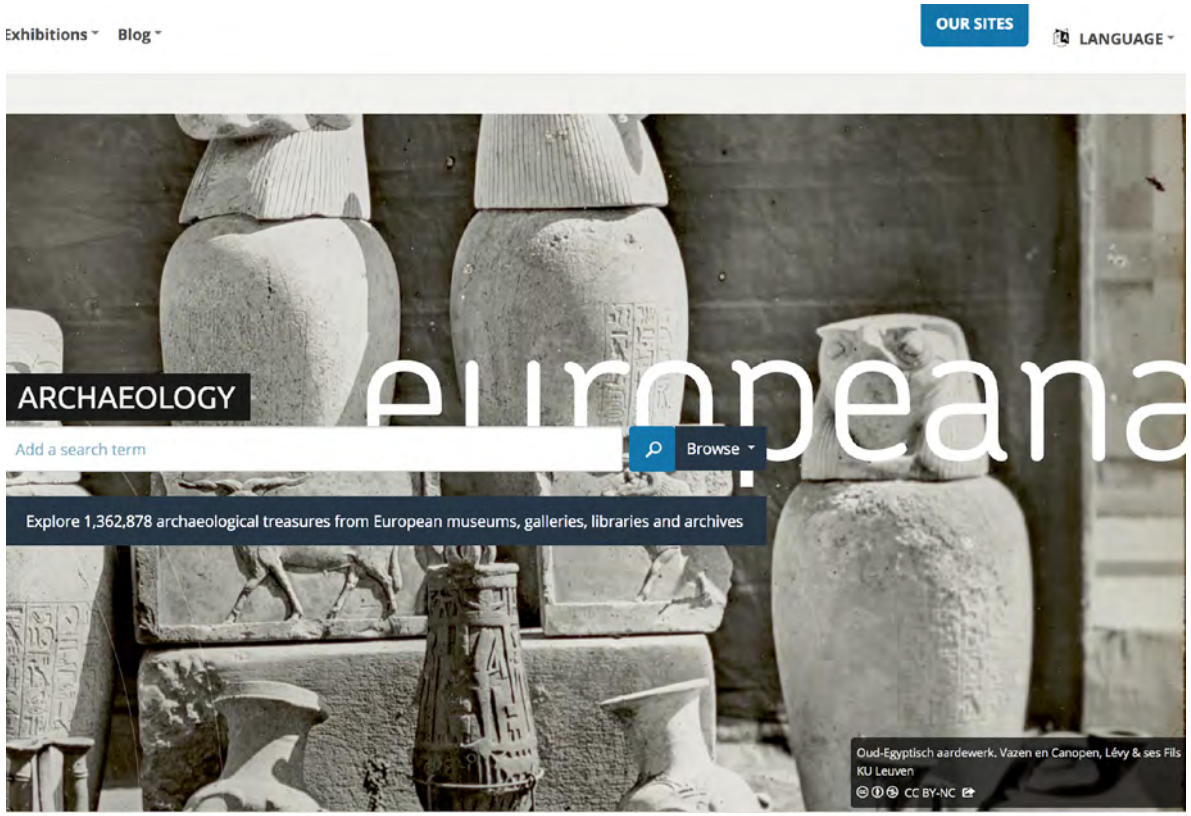
Using a controlled vocabulary with a start and end date for each term helps to improve discovery

- PeriodO (http://perio.do) is a linked open data gazetteer for historical periods which links when to where

Publishing in Europeana

The metadata that is provided to us

- Is mapped and transformed to a standard metadata format for Europeana
- Enriched where possible
- Provided to Europeana for publication



https://www.europeana.eu/portal/en/collections/archaeology

Catalog Services About

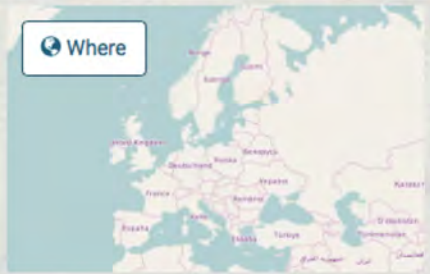
ARIADNE

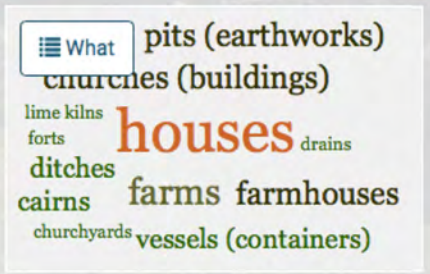
All fields | Search for resources in the Ariadne catalog ...

Welcome

ARIADNE brings together and integrates existing archaeological research data infrastructures so that researchers can use the various distributed datasets and new and powerful technologies as an integral component of the archaeological research methodology.

Browse the Catalog





ARIADNEplus

A research infrastructure for archaeology funded under Horizon 2020 ARIADNE plus aims to:

- Support the research community
- Create a searchable catalogue of research datasets
- Develop services and tools for researchers
- Provide guidance and training

https://ariadne-infrastructure.eu/

ARIADNEplus approach

ARIADNEplus has built on work on the CIDOC CRM to establish a data model (AO-Cat) which is based on the CIDOC CRM with extensions

- Data partners:
  - Provide metadata descriptions for their dataset(s)
  - Map their metadata to ARIADNE's Data Model (AO-Cat)
  - Map their subject vocabularies to AAT
  - Create Perio.do period definitions (linking periods to timespans and areas)
- Mapping tools are available from the project
- Personalised strategy for each data provider

ADCM TERM	FASTI MAPPING EXPLAINATION
dct:title	This was created using a combination of the site name and the season year e.g 'Baccano 2004'
dct:description	This used a standard phrase constructed using the site name and the season year e.g 'Fasti record for interventions in the year 2000 at Baccano'
dct:issued	The date from the database of when the season record was registered
dct:modified	The date that the most recent summary was registered on the database
dct:originalId	The Fasti season code, joined with Fasti and the itemkey from the database in this case the code for season is sea_cd e.g. fasti.sea_cd.AIAC_1
dct:identifier	The URL for this season resource on Fasti e.g. http://www.fastionline.org/season/AIAC_1
dct:language	Summaries in Fasti are stored in two languages, each of these is added to the record
dct:accessRights	All data in Fasti is under the https://creativecommons.org/licenses/by-sa/4.0/ license
dct:spatial	The latitude and longitude of each site are stored as numbers in the

Example of a mapping strategy

This case study describes the mapping from the FASTI online metadata to ARIADNE's data model

ARIADNE plus

Example

- **Is a: D1 Digital Object** (CRMdig)
- **Name:** “dFMROe Database of Coins”
- **Owner:** OEAW
- **Format:** Microsoft Access
- **Subject:** “Römische Münzen”
- **Coverage:** Austria, Romania
- **Temporal Coverage:** “Roman Period”
- [Other general information] | ...

- **Is a: D1 Digital Object** (CRMdig)
- **Name:** “FWC Roman and Medieval Coins”
- **Owner:** Fitzwilliam Museum Cambridge
- **Format:** Microsoft Excel
- **Subject:** “Roman coins” ; “Medieval Coins”
- **Spatial Coverage:** UK
- **Temporal Coverage:** “Roman Period”
- [Other general information] | ...

«Coin» (AAT) http://vocab.getty.edu/aat/300037222
«Roman» (AAT) http://vocab.getty.edu/aat/300020533

Training and support

Useful resources are available from:

- ARIADNE training hub: https://training.ariadne-infrastructure.eu/
- CARARE training hub: https://pro.carare.eu/training-hub/

Applying open/the Fair Principles to archaeology

Depositing project datasets in a digital repository

Data Science skills

Managing datasets from large archaeological projects

‘Transnational Access’

ARIADNEplus advertises calls for researchers to participate in individual training visits in archaeological data management to carry forwards their own research.

Sponsorship is available for participants

Calls should open again in 2021 when the Covid crisis allows:

https://ariadne-infrastructure.eu/transnational-access/

‘Short term scientific missions’

SEADDA is a COST action which focusses on archaeological data

https://www.seadda.eu/?page_id=1274

It offers opportunities for Short Term Scientific Missions for (young) researchers to visit host institutions, for details see:

https://www.seadda.eu/?page_id=1325

SEADDA
Saving European Archaeology from the Digital Dark Age

Conclusions

There are opportunities to improve the visibility of your content by sharing it with international portals such as Europeana and ARIADNE Plus

- This allows people to search for content across institutions and countries
- Licencing of the content allows it to be used by new audiences from education, research and the general public

Initiatives like CARARE, Europeana, ARIADNEplus and SEADDA all offer training and support for people working with digital data

Thanks for your attention

kfernie27@gmail.com

www.carare.eu

CARARE is a membership association and if you would like to find out more about joining us, please see: https://www.carare.eu/join-us/