



Introduction to Metadata and vocabularies

Kate Fernie, CARARE

This session will provide

- An introduction to metadata and its uses for recording monuments and buildings
 - A case study of the CARARE metadata schema
 - An introduction to data standards and the uses of controlled vocabularies and linked open data
 - Geo-Data and temporal data
-

Organising and recording information

- The use of computers in archaeology fieldwork and research has become routine.
- Data is easy to create but its always important to decide what information you need to capture.



Image © Oxford Archaeology (North)



Image © State Library of New South Wales 2015

Recording: the example of historic buildings

- Imagine that you've been asked to write a description of a building
- What would you do?
- Most people tend to come up with:
 - A list of questions to answer and
 - Types of information to record



Recording

- People often design recording forms, use spreadsheets or checklists

Address	Style	Materials	Date	Owner	---

Some of the challenges

- Imagine that several people are involved in your project and want to use the descriptions written by each other
 - Differences can occur that can make it difficult to find information, e.g.
 - The colour of the walls is described, but not the colour of the door.
 - People use different words to describe the colours, e.g. “vermillion”, “crimson”, “red”.
 - ...
-

Metadata: standards

- It takes some time to determine what metadata you need to capture for your research.
 - Very often the metadata created for research projects follow local conventions
 - This isn't a bad thing, but its helpful to follow standards whenever possible. They help to provide an agreed framework for recording.
 - If metadata is already standards compliant (all or in part), it is much less work later when you want to tell users what has been recorded and why.
-

Metadata: schemas, standards

Some standard ontologies in use in cultural heritage:

- The CIDOC CRM is a widely used ontology in the cultural heritage; this ISO standard is not specific to archaeology or to any one domain.
 - SPECTRUM and LIDO are examples of domain ontologies and cover processes specific to museums.
 - For archaeological sites, CIDOC core data standard for monument inventories.
-



Case study: CARARE

Background

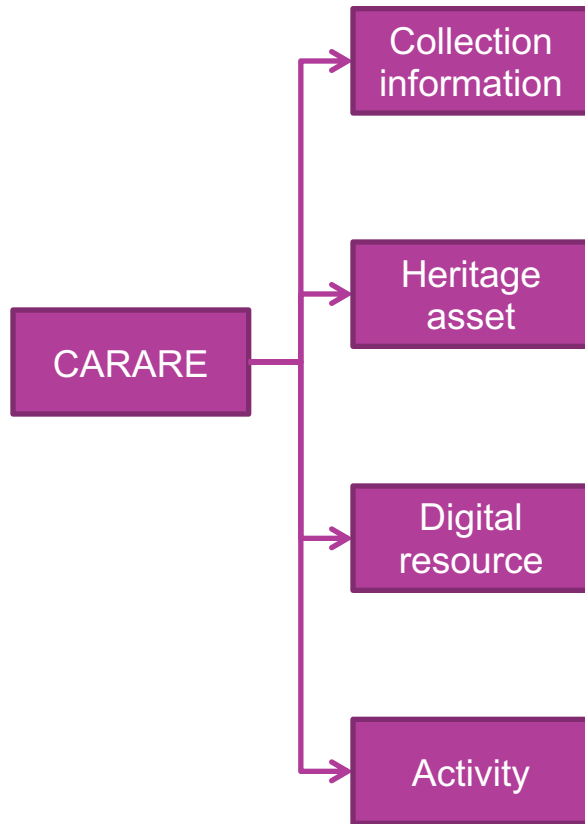
- CARARE is a network of archaeological and architectural heritage archives from across Europe.
- We began life as a project to make our digital content available to Europeana and to the public.
- We offer aggregation services and recommend good practices for content relating to archaeological monuments and historic sites

Case study: CARARE metadata schema

- Designed to bring together digital content for the archaeological and architectural heritage. The schema is based on:
 - CIDOC core standards +
 - MIDAS Heritage +
 - LIDO +
- It's compatible with the CIDOC CRM and also the Europeana Data Model



The CARARE Metadata schema



Heritage Assets

Monuments, buildings, landscape areas, artefacts and information sources such as images, documents, videos, audio recordings, 3D models ...

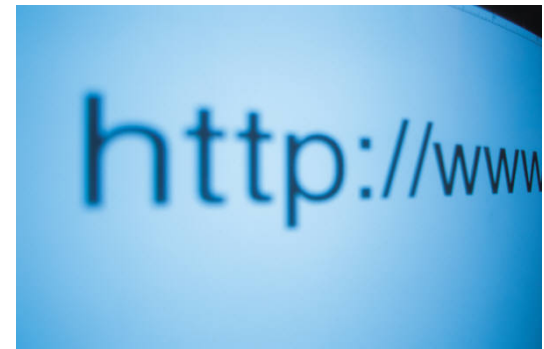
- Title, Description
- Characteristics
 - Type, Materials, Dimensions, Inscriptions
 - Spatial (place, address, map coordinates)
 - Temporal (date, time span, period)
- Actors
- References
- Relations



Digital Resource

Digital representations of the heritage asset (images, models, text files, etc.)

- Title
- Type and format
- Location of the resource:
 - URL, URI
 - Direct or to landing page
- Rights



N.B. There can be more than one digital resource attached to a heritage asset

Fieldwork, laboratory analysis, post processing, digital reconstruction, etc.

- Title
- Description
- Date
- Actors
- Type of event
- Methods and techniques
- General and specific purpose



How does this work?

1. The description of heritage asset (monument, landscape, building, etc.) is bundled together with its:
 - digital representations and
 - related events, and
 - Contextual information about people, places and subject concepts from controlled vocabularies
2. Relations between heritage assets support objects that are composed of other objects

How does this look in Europeana?

Heritage Asset 1



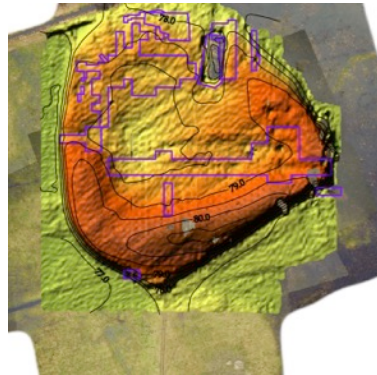
Wicina-osada obronna

Description:

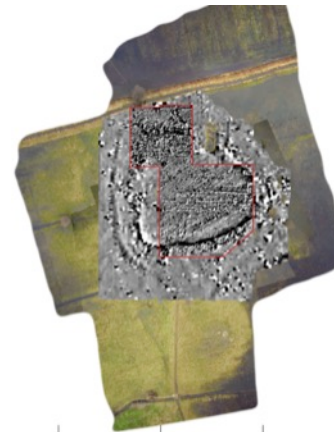
Grodzisko w Wicinie, gm. Jasień, woj. lubuskie zostało wzniesione przez ludność kultury lużyckiej, we wczesnej epoce żelaza, w tzw. okresie halsztackim C-D, który odpowiada przedziałowi czasowemu pomiędzy 700, a 450 lat p.n.e. Relikty założenia obronnego są jednym z najcenniejszych zabytków archeologicznych na terenie Polski. Posiadają bezcenną wartość dla dziedzictwa kulturowego naszego kraju oraz Europy. Nieprzypadkowo obiekt, decyzją Wojewódzkiego Konserwatora Zabytków w Zielonej Górze, w dniu 12 września 1968 roku, został wpisany do rejestru zabytków, gdzie figuruje pod numerem L-6/C.

Grody, grodziska, osady obronne to, pod wieloma względami, wyjątkowe stanowiska archeologiczne. Ich znaczenie dla dziedzictwa kulturowego i potencjał naukowy jest nie do przecenienia. W Polsce jest ok. 2500 tego typu obiektów – z tego 1725 wpisanych do rejestru zabytków – i jedynie 70 to pradziejowe grodziska z późnej epoki

Has Representation



Heritage Asset 2



Heritage Asset 3



Heritage Asset 4

Benefits of having a data standard

- Structured information which provides people with familiar concepts to work with
- Rich where the domain calls for it, e.g.:
 - Time – from earliest prehistory to modern dates
 - Space – place names, coordinates
- Provides a good framework for
 - several people working in a team, or
 - in our case interoperability between similar datasets captured in different countries and different languages



Subject concepts

- CARARE, ARIADNE and Europeana face specific challenges because the metadata records are provided in different languages
 - Specifying the use of thesaurii - lists of agreed terms with simple hierarchical relationships – is an important first step to improving access and interoperability between datasets
 - There are quite a lot of vocabularies available for archaeology
-

Vocabularies

These are some examples from the UK:

Archaeological objects

Archaeological Sciences

Building materials

Components

Event type

Monument Type

Maritime craft

Periods

Archives

These vocabularies were developed by the national agencies for England, Scotland and Wales

There are similar monolingual vocabularies available from France, Italy, Ireland, the Netherlands and other countries

Monument Type Thesaurus

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

- ▶ **AGRICULTURE AND SUBSISTENCE**
- ▶ **CIVIL**
- ▶ **COMMEMORATIVE**
- ▶ **COMMERCIAL**
- ▶ **COMMUNICATIONS**
- ▶ **DEFENCE**
- ▼ **DOMESTIC**
 - BACKYARD**
 - ▶ **BAKEHOUSE**
 - BEER CELLAR**
 - BROCH**
 - BURNT MOUND**

Thesaurus Browse

Monument Type Thesaurus

FISH TANK

[Home](#) ▶ [DOMESTIC](#) ▶ FISH TANK

Scope note:

A small brick structure attached to a house for keeping fish prior to cooking.

BT [DOMESTIC](#)

FISH TANK

RT [STORAGE TANK](#)

Identifier http://purl.org/heritagedata/schemes/eh_tmt2/concepts/83579

- CARARE has chosen to use the Getty Art & Architecture Thesaurus (AAT) as a central spine
- A set of archaeological subjects has been defined
- CARARE data partners make mappings between the subject concept in their native language and the concept in AAT
- This means you can search for a subject in Hungarian, and get results in German (English provides the glue)

Dolní Kounice

From Wikipedia, the free encyclopedia

Dolní Kounice (Czech pronunciation: [ˈdɔlɲiː ˈkɔʊɲɪtɕɛ]; German: *Kanitz*) is a small town in the South Moravian Region of the Czech Republic. It has around 2,400 inhabitants.



Dolní Kounice Panorama



- Geo names is a widely used international service internationally – provides coordinates for modern place names
- Lots of work has been also been done on Historic place names
- One of the best examples is the [Pelagios](#) project
 - Started life focussing on places in the classical world
 - Now extending to include historical places more generally
 - Links online resources to places in classical/historical world, aims to be machine readable and provides a map interface

Temporal Data

- When is always dependent on where (i.e. the iron age is different depending on where you are in the world)
 - This makes “when” by far the most difficult aspect of making archaeological data from different countries interoperable
 - CRMarchaeo has tried to deal with this using a concept called ‘space-time’ volumes
 - PeriodO uses ‘assertions’ to build consensus around temporal terms
-

Summary

- Metadata sits at the heart of what we do. Without good metadata preservation, access and interoperability aren't possible
- Metadata should be based on standards whenever possible
- Take time to decide what information you need to record and which standards are appropriate for your data
- Defining local standards for your project isn't a bad thing
 - Using agreed lists of terms is a simple first step to making your data more interoperable
 - Place is very important for most archaeological datasets
 - Record the assertions that you make for people who come later

Thanks for your attention

kfern27@gmail.com
