Digging into Archaeological Data and Image Search Metadata (DADAISM)

Christopher Power
Human Computer Interaction Research Group
Department of Computer Science
University of York
Overall Aim of DADAISM

- To create a usable and useful interactive system that will go beyond current approaches to support archaeologists in their tasks of finding, organising, relating and manipulating images and their metadata
- Why do we need to do this?
Images in Archaeology
Problems with Image Archives in Archaeology

• Searches that are overly specific result in too few results to be useful or too broad resulting in too much data being returned
• Broad lists of metadata provide too many opportunities to users to get lost in data sets
• Difficult to relate across datasets that use different metadata sets (some harmonisation approaches help with this)
• Missing metadata means that key images are not included in some searches
• Little or no support outside of form fill-in approaches for entering metadata
Possible solutions: Image Processing

- A lot of images in object datasets have similar features
- Those features are used to classify and reason about objects – meaning users want and need that information
- Those features often naturally overlap with the metadata sets
- Example: Roman Amphorae
Possible solutions: Image Processing (2)

• Because of these similarities, we can use image processing techniques to match the objects as being similar

• We can then automatically generate the metadata that is needed for searching, filtering and browsing

• We can do this immediately when the object is being catalogued if a picture is being available fixing the problem going forward or

• We can do this over the set of images to augment existing metadata
Possible solutions: Text Mining

- Archaeologists are already collecting and writing a huge amount of information.
- Grey literature reports have a wealth of information that could be at least linked to the images.
- Preferably they can be mined for more information to further augment metadata.
Possible solutions: Text Mining

• We can text-mine the grey literature for information about artefacts that are in image collections
• We can then recommend potential metadata for images based on what is contained in the grey literature documents
• These annotations can then be used to augment search further
Possible solutions: Instrumental Interaction

• As mentioned, many of our interfaces are heavily form fill-in interfaces

• A lot of focus has been on searching interfaces, but not on what do archaeologists want to do with the data when the system finds it for them
Possible solutions:
Instrumental Interaction

- We can study how archaeologists are currently working with archives and what kinds of activities they are doing with them.
- We can look at how we can use existing metadata, and the improved metadata produced from the other project components to improve the quality of browsing.
- We can evaluate our prototype with real users, iteratively as prototypes are available, to ensure that it is meeting their needs and goals.
Project Team

- University of York
  - Prof. Helen Petrie and Dr. Christopher Power (Computer Science) – Interactive Systems
  - Prof. Julian Richards (Archaeology) – Data provider and user engagement

- University of Saskatchewan
  - Dr. Mark Eramian (Computer Science) – Image processing

- University of Amsterdam
  - Prof. Maarten de Rijke (Computer Science) – Text Mining
  - Dr. Cees Snoek (Computer Science) – Image search
The key objectives of the DADAISM project will be the following:

1. A next generation interactive system that supports archaeologists in their research, specifically supporting new ways of doing exploratory search, organisation, relationship reasoning and labelling of image data

2. An augmentation of existing metadata about images (3 different sets of increasing heterogeneity) with content-related metadata that will improve the quality of search results that are returned to the users

3. Linking of images to other data resources such as researcher notes and grey literature archives

4. A set of general principles for applying the DADAISM approach over heterogeneous datasets, informing future practice in digital archaeology
Dissemination

• Academic publishing with four tracks:
  – HCI, Image Processing, Text Mining, Archaeology
  – Combined publications for pipeline into Digital Cultural Heritage journals (e.g. ACM Journal on DCH)

• Website and social media channels

• Recruitment of users for tools through existing ADS communities

• Hosting of data on ADS LOD services

• Hosting of interactive system in ADS service offerings
Questions?