

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

ads ARCHAEOLOGY DATA SERVICE

HOME ARCHSEARCH ARCHIVES LEARNING ADVICE OUR RESEARCH BLOG ABOUT US LOGIN

ALL | JOURNALS AND SERIES | GREY LITERATURE | PROJECT ARCHIVES | BIBLIOGRAPHIES | THESES

Lower Palaeolithic technology, raw material and population ecology

Gilbert Marshall, David Dupplaw, Derek Roe, Clive Gamble, 2002

Introduction
Overview
Query

Data copyright © University of Southampton unless otherwise stated

A · H · R · B
arts and humanities research board

Primary contact
Dr Gilbert Marshall
Department of Geography
Royal Holloway, University of London
Egham
Surrey
TW20 0EX
England
Tel: 01784 443569

[Send e-mail enquiry](#)

Resource identifiers
ADS Collection: 349
doi:10.5284/1000354
How to cite using this DOI

UNIVERSITY OF Southampton

Query
Please enter your search terms below press the query button to search the database.

Explanation of the search options

[Submit](#) [Reset](#)

Id no.	<input type="text"/>
Site name	<input type="text"/>
Country	<input type="text" value="ENGLAND"/>
Biface type	<input type="text" value="HANDAXE"/>
Raw material	<input type="text"/>
Edge profile	<input type="text"/>
Condition	<input type="text"/>
Percentage of circumference worked	<input type="text"/>
Museum	<input type="text"/>
Finder	<input type="text"/>
Weight (g) (min - max)	<input type="text"/> - <input type="text"/>
Length (mm) (min - max)	<input type="text"/> - <input type="text"/>
Breadth (mm) (min - max)	<input type="text"/> - <input type="text"/>
Breadth/Length (min - max)	<input type="text"/> - <input type="text"/>
L1/Length (min - max)	<input type="text"/> - <input type="text"/>
BA/BB (min - max)	<input type="text"/> - <input type="text"/>

ads ARCHAEOLOGY DATA SERVICE

HOME ARCHSEARCH ARCHIVES LEARNING ADVICE OUR RESEARCH BLOG ABOUT US LOGIN

ALL | JOURNALS AND SERIES | GREY LITERATURE | PROJECT ARCHIVES | BIBLIOGRAPHIES | THESES

Lower Palaeolithic technology, raw material and population ecology

Gilbert Marshall, David Dupplaw, Derek Roe, Clive Gamble, 2002

Introduction
Overview
Query

Data copyright © University of Southampton unless otherwise stated

A · H · R · B
arts and humanities research board

Primary contact
Dr Gilbert Marshall
Department of Geography
Royal Holloway, University of London
Egham
Surrey
TW20 0EX
England
Tel: 01784 443569

[Send e-mail enquiry](#)

Resource identifiers
ADS Collection: 349
doi:10.5284/1000354
How to cite using this DOI

UNIVERSITY OF Southampton

Query - Results
Records 1 - 50 of 1117
Pages: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | > >>
Click on the images or the links below to view more details for each biface.

	IRREGULAR FLINT HANDAXE (889g) Location: WARREN HILL ENGLAND Museum: BRITISH MUSEUM, LONDON, ENGLAND
	LINEAR FLINT HANDAXE (512g) Location: WARREN HILL ENGLAND Museum: BRITISH MUSEUM, LONDON, ENGLAND
	LINEAR FLINT HANDAXE (488g) Location: WARREN HILL ENGLAND Museum: BRITISH MUSEUM, LONDON, ENGLAND
	LINEAR FLINT HANDAXE (307g) Location: WARREN HILL ENGLAND Museum: BRITISH MUSEUM, LONDON, ENGLAND
	LINEAR FLINT HANDAXE (344g) Location: WARREN HILL ENGLAND

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

The screenshot shows the Archaeology Data Service (ADS) website. A red arrow points to the text 'currently 24852 reports available' in the search results. The website header includes the ADS logo and navigation links such as HOME, ARCHSEARCH, ARCHIVES, LEARNING, ADS-easy, ADVICE, OUR RESEARCH, BLOG, ABOUT US, and LOGIN. Below the header, there is a section titled 'Library of unpublished fieldwork reports' with an introduction and contact information. The main content area displays a grid of report covers, including one for 'Wessex Archaeology' and another for 'OLD STATE VILL'. The footer contains the ADS logo, 'THE UNIVERSITY of York', and social media icons.

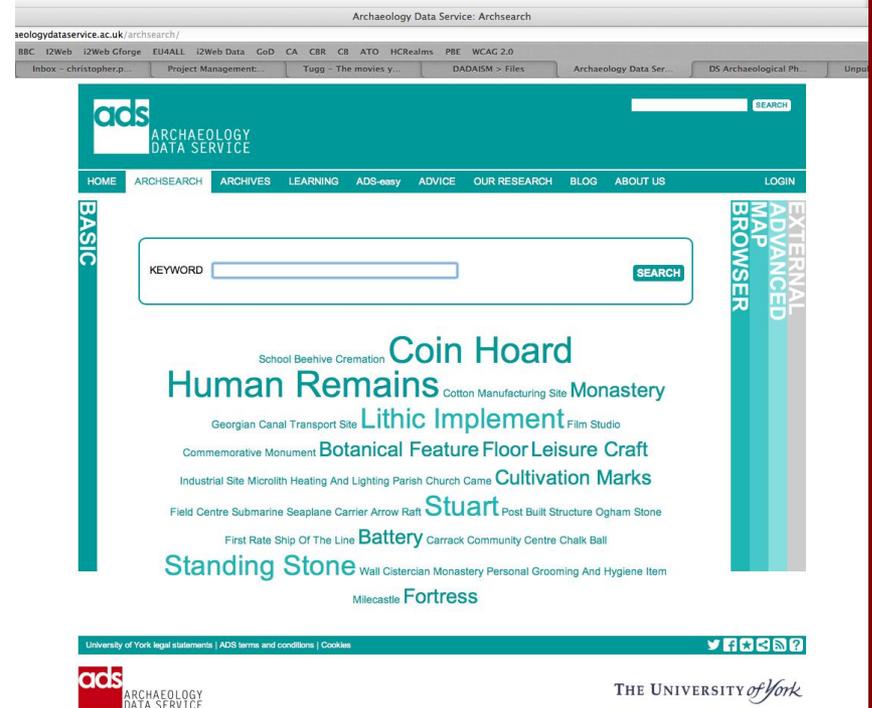
- Archaeologists are already collecting and writing a huge amount of information
- Grey literature reports have a wealth of information that could be at least linked to the images
- Preferably they can be mined for more information to further augment metadata

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

DADAISM Objectives

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?