

Improving Archaeologists' Online Archive Experiences Through User-Centred Design

CHRISTOPHER POWER, ANDREW LEWIS, HELEN PETRIE, KATIE GREEN, and JULIAN RICHARDS,
University of York

MARK ERAMIAN, BRITTANY CHAN, and EKTA WALIA, University of Saskatchewan

ISAAC SIJARANAMUAL and MAARTEN DE RIJKE, University of Amsterdam

Traditionally, the preservation of archaeological data has been limited by the cost of materials and the physical space required to store them, but for the last 20 years, increasing amounts of digital data have been generated and stored online. New techniques in digital photography and document scanning have dramatically increased the amount of data that can be retained in digital format, while at the same time reducing the physical cost of production and storage. Vast numbers of hand written notes, grey literature documents, images of assemblages, contexts, and artefacts have been made available online. However, accessing these repositories is not always straightforward. Superficial interaction design, sparsely populated metadata, and heterogeneous schemas may prevent users from working the data that they need within archaeological archives.

In this article, we present the work of the Digging into Archaeological Data and Image Search Metadata project (DADAISM), a multidisciplinary project that draws together the work of researchers from the fields of archaeology, interaction design, image processing and text mining to create an interactive system that supports archaeologists in their tasks in online archives. By adopting a user-centred approach with techniques grounded in contextual design, we identified the phases of archaeologists work in online archives, which are distinctive to this user group. The insights from this work drove the design and evaluation of an interactive system that successfully integrates content-based image based retrieval and improved metadata searching to deliver a positive user experience when working with online archives.

CCS Concepts: • **Human-centered computing** → *Empirical studies in interaction design*; • **Applied computing** → *Arts and humanities*;

Additional Key Words and Phrases: Archival search, archaeological information, exploratory search and discovery, content-based image retrieval, texture description, feature fusion, instrumental interaction, user-centred design

This work was funded by the Digging into Data Challenge with funding from JISC, AHRC/ESRC, NWO, and SSHRC.

Authors' addresses: C. Power, A. Lewis, and H. Petrie, Department of Computer Science, University of York, Heslington, York, YO10 5GH; emails: {christopher.power, andrew.lewis, helen.petrie}@york.ac.uk; K. Green and J. Richards, Department of Archaeology, University of York, The King's Manor, York, YO1 7EP; emails: {katie.green, julian.richards}@york.ac.uk; M. Eramian, B. Chan, and E. Walia, 176 Thorvaldson Bldg., 110 Science Place, University of Saskatchewan, Saskatoon, SK, S7N 5C9, Canada; emails: eramian@cs.usask.ca, brc591@mail.usask.ca, ewb178@mail.usask.ca; I. Sijaranamual and M. de Rijke, University of Amsterdam, Informatics Institute, Science Park 904, 1098 XH Amsterdam, The Netherlands; emails: {i.b.sijaranamual, derijke}@uva.nl.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

2017 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 1556-4673/2017/01-ART3 \$15.00

DOI: <http://dx.doi.org/10.1145/2983917>

ACM Reference Format:

Christopher Power, Andrew Lewis, Helen Petrie, Katie Green, Julian Richards, Mark Eramian, Brittany Chan, Ekta Walia, Isaac Sijaranamual, and Maarten de Rijke. 2017. Improving archaeologists' online archive experiences through user-centred design. *J. Comput. Cult. Herit.* 10, 1, Article 3 (January 2017), 20 pages.
DOI: <http://dx.doi.org/10.1145/2983917>

1. INTRODUCTION

An ever-increasing amount of digital archaeological data is being made available online through a variety of different web-based repositories. Each repository may contain thousands of resources, including everything from scans of handwritten notes, grey literature documents, photographs or sketches of sites, assemblages, and contexts, to images of individual artefacts.

Yet, many of these resources remain underused for a variety of reasons. In some cases, the content of documents is simply unavailable, needing either manual or semi-automatic transcription. In other cases, the content is available, but searching is not possible due to metadata being sparsely populated. These problems are compounded with archive collections often having non-standard or heterogeneous metadata schemas that make it difficult for users to guess what search terms may be relevant for their queries. All of these factors contribute to users struggling to find the data that they need within large archaeological archives.

While addressing all of these issues is important, it does not necessarily follow that just improving the metadata would result in a large increase in data usage in these archives.

In the last decade, there has been an increasing emphasis on producing systems that provide people with a good user experience. As users have increased their exposure to a range of technology options, through apps they can buy cheaply, or via online services and where they have the opportunity to explore for free before buying, they are developing a sophisticated understanding of what is “good” for them in technology. Technology will ultimately go unused if it is either unusable, where users cannot accomplish their goals, or where it delivers an overall negative experience. Online archives, similarly, need to be considering user experience as a primary outcome when delivering a new system, driven from the needs of users, or they risk any investment in improved data and metadata going to waste.

Interestingly, potentially due to the focus on improving search, there is a tendency to design systems with a *resource centric* approach, which is similar to the materials-based approach discussed by Elena et al. [2010]. That is, the design of the systems and their interfaces, such as those developed at the Archaeology Data Service (ADS), are driven from the structure of the data [Jeffrey et al. 2007; Charno et al. 2012]. When we look across the literature, we find many projects that deliver systems through this resource centric approach. For example, the Gazetteer created for the project relating to Lucanian heritage [Duplouy et al. 2014] was created for managing resources. This project has an excellent analysis of technical concerns and the requirements related to the materials that the system will work on and the resources needed to implement it. However, within the work there was no engagement with potential users to identify what are their key needs from the system. Another example is the ArkeoGIS system, which has concrete analyses about the types of information users will deposit into their system [Bernard et al. 2014], but no indication of what the processes were by which archaeologists would engage with the system.

While we critique each of the above, user goals were not a targeted outcome of the projects. Each of these systems are useful in of themselves and provide support to specific users. However, the resource centric approach leads to a lack of generalisation on how to build systems that support the broader tasks of archaeologists. As the systems are so closely tied to their data, it is difficult to extract lessons and interaction design patterns that could be applied to new archaeology systems.

We propose that in order to increase use of online archaeological archives and to make future systems easier to design, we need to build on the very broad literature about archive users [Marchionini 2006; Huvila 2007; Chapman 2010; Elena et al. 2010; Sinn 2012], and investigate how current systems are not meeting the needs of the archaeologists in research and professional practice. With this information, we can then involve archaeologists in the process to produce designs that are more supportive of the workflows found across a variety of different systems and practices. We then need to evaluate systems rigorously with real users, and use the lessons learned from all of this work to generalise design patterns that can be shared across the field.

In this article, we present the work of the Digging into Archaeological Data and Image Search Metadata project (DADAISM). The DADAISM project brought together researchers from the fields of archaeology, interaction design, image processing, and text mining to create a user-centred interactive system to support archaeologists in their tasks when working with digital image and document archives.

2. CONTEXTUAL DESIGN

There is a limited amount of work exploring the overall workflows that archaeologists undertake in working with online materials. As a result, there was a need to work with archaeologists to elicit the types of tasks they undertake when working with data in online archives.

We chose to use a contextual enquiry [Holtzblatt and Jones 1993; Holtzblatt 2009] approach to examine the archaeologist's work and draw out details in context. In this approach, the researcher encourages participants to undertake their normal tasks of work for the purposes of observing and understanding how and why the participant undertakes specific activities. Unlike a traditional interview setting, where the researcher drives the selection of topics, it is instead participants who are guiding the researcher through their routines. During the observation, the researcher will ask questions about the purpose behind particular actions and how it relates to the overall task. At times the researcher will also explain their interpretation of what is happening in order to confirm with participants that they are understanding the different activities appropriately. All interviews are then analysed for key common workflows and scenarios to feed into interaction design sessions.

2.1 Materials and Context

For contextual enquiry to be effective, it is often best to document activity in the place of work of the individual. In most cases, this was possible, with interviews being conducted in the offices of the participant on their own personal computers.

In four cases, participants reported that they do their work in a variety of spaces, including offices, in the field, and in libraries. For the enquiry sessions, three of those participants undertook their tasks in quiet offices at the University of York, and one in a quiet space in the university library. In each case, participants brought a variety of paper notes and notebooks with them, and accessed data remotely through shared drives or documents in cloud storage.

Participants used a variety of archaeology archives and web resources in their searches, including: the Archaeological Data Service archive, the British Museum website, the Portable Antiquity Scheme, and custom databases that were created from archaeological excavations.

In addition to notes taken with the interview schedule, all sessions were video recorded for later review and analysis.

2.2 Participants

Eight participants, five male and three female, were recruited as an opportunity sample through mailing lists and contacts of the Archaeological Data Service.

Participants were selected from across different groups of archaeologists including: six research archaeologists working in academic environments, one professional archaeologist who works in the field for private or public organisations, and one archaeology student.

The ages of the participants ranged from 25 to 41, with researchers and professionals having between 3 to 10 years working in the field of archaeology. Participants were self-described specialists in a variety of different topics including: small artefacts from Roman frontiers, lithic age antler and bone tools, flint and glass tools, Greek pottery, Middle Eastern pottery, landscape archaeology, and ancient colonisation.

2.3 Contextual Enquiry Procedure

The enquiry began with participants reading an information sheet and were asked to sign an informed consent form. The researcher then asked a set of warm-up questions exploring the participants' history in the archaeology field, the type of work they specialise in, and the types archives that they typically use. This was followed by a set of questions exploring the broad types of tasks participants usually undertook when using online archives, and some exploratory questions about their favourite and least favourite things about the archives themselves.

The focus of the enquiry sessions was on understanding how archaeology users from across the community work with image and document archives to achieve a variety of tasks. Normally, in contextual enquiry, individuals undertake their work in a one to two-hour session. However, the type of work that archaeology users do with the archives often extends out over a period of weeks or months. As a result, the contextual enquiry protocol was modified slightly and participants were asked to explore, with the researcher, a recent project that they had undertaken. The participants were asked to either extend that work if it was possible, or to retrospectively show how they began their work on the project. Participants were asked to explore key or unusual points in their process, in an adaptation of the critical incident technique used in other methodologies [Hartson and Castillo 1998; Serenko and Turel 2010].

Participants were asked to describe the project that they were about to show and explain what they were trying to accomplish. Then, the participants undertook their work in the archives, unguided by the interviewer, and were prompted for clarification and further discussion of key points. Each session ran approximately one hour.

2.4 Analysis

In order to contextualise and understand the different types of activities undertaken by the archaeologists, we undertook a scenario-based design methodology [Rosson and Carroll 2009]. We composed a series of scenarios that describe activities seen in the video recordings which were then rewritten as a smaller set of scenarios that captured the common activities of users. These scenarios were reviewed by archaeologists at the ADS in regards to how representative they were to their own experiences or the experiences of the ADS user base. After several rounds of review, a set of five master scenarios were used for ideation through the rest of the design process.

All participants were trying to find information about artefacts that related to the ones they were already working with. However, the final uses of the images were varied. Many of the participants were trying to collect together sets of similar artefacts in order to identify an artefact they had in hand, or in some cases, to pass off that identification task to another specialist team member. In these cases, there were often several types of analysis that would be conducted by other team members in parallel, with individuals contributing analyses to a central repository (e.g., shared spreadsheet or document). Other researchers were working on presentations and displays for public engagement. Several were working to find images for inclusion in publications or lectures.

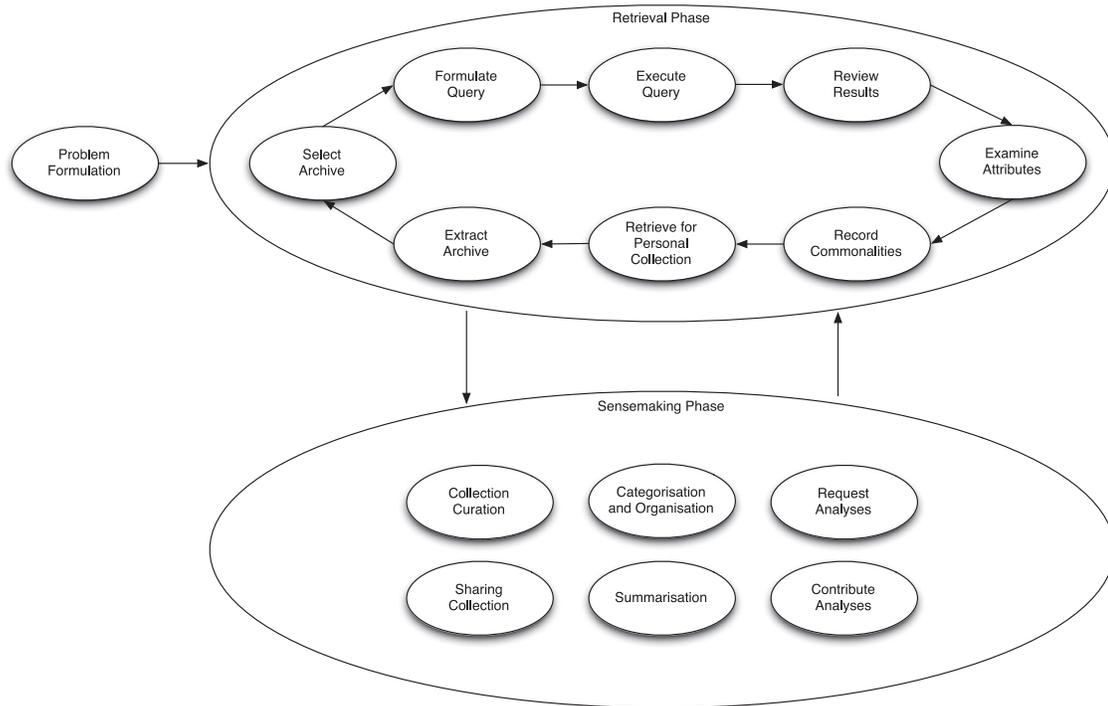


Fig. 1. Sequence model for archaeologists' work with online archives.

2.5 Design Insights

The analysis of the participants' sessions point toward several commonalities in workflow that can inform the design of our interactive system. A sequence model [Holtzblatt 2009] was constructed as an output of the scenario-based approach. The model represents the general progression of activities seen during the sessions. The model is shown in Figure 1.

The sequence model aligns well with the pattern of work for information described in other information-seeking models, such as Ellis and Haugan [1997], Marchionini [2006], and Bron et al. [2012], where the sequence is often represented as a cycle with a person moving from formulating a problem, to selecting sources and running queries, to results, and then extracting information for analysis and interpretation, with opportunities for feedback into early stages of the cycle to reformulate or change direction. See Bron et al. [2016] for a recent comparison of information-seeking models of the research cycle across a number of disciplines.

While at a fundamental level the journey is the same for archaeologists working with online archives, the emphasis on when the archaeologists transition between the different states in information seeking is distinctive. In almost all cases, our archaeologist participants experience two distinct phases of work: a phase of retrieving resources and a phase of sensemaking across those resources.

2.5.1 Retrieval Phase. In the first phase after query formulation, the participants collect together resources from across a large number of queries and result sets.

Participants would often formulate a set of keyword queries based around their knowledge of the artefact with which they were working, such as “flint blade” or “roman amphora,” in order to prime their search with some results.

After examining the results of those queries, they begin to collect metadata items to use in the next set of reformulated queries, often on paper outside of the search system. Sometimes, this resulted in successful follow-up searches, but there was a tendency for participants to append the keywords onto previous queries. This often resulted in overly specific queries that returned very small sets of questionable results to the participants.

The participants also often extracted metadata terms from the results of one archive to apply to other archives. This is interesting, because even though the metadata schemas are most likely different between the different archives, there seems to be an implicit trust that search systems will be able to find appropriate information based on the terms themselves.

When these attempts to find artefacts via metadata fail, participants often turned to the physical characteristics of objects they were searching for, such as trying a query for “green glaze” to try to bring up pottery objects from the archive. Two of the participants attempted to use Google image search to bring back images from across the web, potentially giving new avenues of exploration or new unknown archives. However, both of these participants lamented that these searches were seldom successful because the features that are used to identify and categorise artefacts generally do not use very general search algorithms. This issue is further complicated by many artefacts being shards or fragments of larger artefacts, meaning that the image search cannot accurately match artefacts.

These observations have several consequences for design. First, they demonstrate the importance of having complete metadata, and validate the idea that effort contributed to improving metadata availability will have an impact on overall success in archives. Second, any design solution should provide means for users to speed up that reuse of metadata in queries across multiple archives. Finally, it would be very beneficial for a system to allow users to perform searches on visual characteristics of artefacts, many of which are not recorded in typical metadata schema. This could be in the form of more specialised information beyond length and colour, or via an image search mechanism.

2.5.2 Sensemaking Phase. After collecting together large amounts of information, the participants move into a sensemaking phase where they begin to process the items collected in different ways in order to produce new insights [Klein et al. 2006a, 2006b].

There was less commonality across the participants in the method and order in which this phase was undertaken, often with participants doing many activities in parallel, before transitioning back into another retrieval phase. The following sections discuss the common activities to the Sensemaking Phase.

2.5.3 Personal Collection Curation. Perhaps one of the most surprising aspects of the participants’ sessions was the prevalence and size of personal collections of artefact images and documents that were stored in relation to any one project. During the search processes, participants often took resources from the web and stored them either in documents with research notes, or more commonly had hard drives or USB sticks with many hundreds or even thousands of pictures and documents. In many cases, a separate document or spreadsheet was kept with the web links to the items as well as key information like accession or collection numbers.

These images and documents are often stored in relatively simple structures to begin with (e.g., a folder labelled “Flint images”), with resources being added when there was even a slight possibility of them being useful. After the initial collection stage, there were phases of curating the images for duplicates or obviously irrelevant items, and then phases of organising and categorising the items into more elaborate schemas, usually through file folders.

In parallel with these activities, participants typically kept an ongoing set of research notes regarding why images were being kept, or particular features. These notes include relationships between the different resources as well as summarisation about why particular resources were being included in

the set. Where spreadsheets were used, those notes would be integrated with the information about the provenance of the resources.

2.5.4 Sharing and Collaboration. Most of the participants discussed sharing of their personal collections and their research notes for purposes of collaboration with colleagues. There was a large variety of different ways of sharing resources, with no clear style or theme emerging. Many people discussed sending research notes and archive (i.e., zip) files to colleagues through email, sending USB drives through the post, or sharing through cloud resources to request contributions from others. This, of course, leads to potential duplication or data loss problems, which are common in this type of asynchronous sharing.

Perhaps the most sophisticated structure that was seen during the sessions for collaboration was a spreadsheet for a single excavation that consisted of all of the identifying information for each resource from the site, including a web link where it could be found, and then individual protected columns for each collaborator to include their findings from different analysis methodologies. In this way, a single team was able to contribute all of their results to a single spreadsheet for each object of interest. It should be noted that this approach was one adopted for a specific excavation that took place as it progressed, not generated from a set of archive searches. However, it is an interesting model of sharing that might be integrated into designs.

3. INTERACTIVE SYSTEM PROTOTYPE

From the results of the contextual inquiry sessions, we proposed an interactive application that would support archaeologists across the two phases. For the Retrieval Phase, we looked at means of improving the initial searches, trying to take advantage of the visual characteristics of artefacts as well as proposing new ways to help users identify commonalities between the artefacts they are working with, and moving them to their next search tasks. In order to achieve this, we utilised image processing across specialist data sets, to allow matching of images in the archive to a query image provided by the archaeologist. Further, we employed text mining to extract content level metadata from documents that were previously inaccessible to researchers.

For the Sensemaking Phase, we focussed on the creation and curation of personal collections that could be shared with colleagues.

3.1 Data Sources

For purposes of creating the interactive system prototype, the project team identified several potential datasets from the ADS, which were good candidates for exploring the novel aspects of the interactive system.

We required one set of data that had a large collection of images of similar artefacts with good, reliable resource metadata, which could be used for ground truth against any image processing work undertaken in the project. We identified a robust flint biface artefact data set that met these requirements. Further, we had several document sets around flint bifaces that were underused due to the lack of content level metadata. The final datasets from the ADS that were used in the prototype were:

- J.J. Wymer Archive: A collection of transcribed notes from one of the foremost specialists in Palaeolithic archaeology
- Grey Literature Archive: A library of unpublished fieldwork reports with large amounts relating to a variety of different find types, including lithics.
- Lower Palaeolithic Technology, Raw Material and Population Ecology Archive: A database of 10,668 digitised images of 3,556 bifaces, as well as information on provenience, raw material, and standard

Table I. Methods Evaluated for Image-based Retrieval of Flint Biface Images

	Methods	References
1	Uniform local binary patterns	[Topi et al. 2000]
2	Orthogonal combination of linear binary patterns	[Zhu et al. 2013]
3	Segmentation-based fractal texture analysis	[Costa et al. 2012]
4	Global phase congruency histogram	[Kovesi 2000]
5	Angular radial phase congruency histogram (ARPCP)	[Chalechale et al. 2004]
6	Orientation-based phase congruency histogram	Our novel variation of Kovesi [2000] and Chalechale et al. [2004]
7	Gabor wavelet features	[Manjunath and Ma 1996]
8	Log-Gabor wavelet features (LGWF)	[Arróspide and Salgado 2013]
9	Binary texton features (BTF)	[Guo et al. 2014]
10	Fusion of LGWF and ARPCP	
11	Fusion of LGWF and BTF	
12	Fusion of LGWF and BTF	
13	Fusion of LWF, ARPCP, and BTF	

measurements. Entries have a time range from 1.5Myr to 300Kyr and include material from Africa, Europe, and the Near East.

3.2 Image-Based Retrieval

Given the examples of individuals trying to match visual characteristics of artefacts, we implemented a set of image-based retrieval algorithms that accept a query image as input, search a database of images for those that are most similar to the query image, according to some domain-specific criteria, and return a set of the most similar images. Many of our participants of the contextual enquiry sessions had experience either working with or teaching about biface artefacts of different types. We interviewed our participants regarding what characteristics they use to identify different types of biface artefacts.

From this, we hand-crafted numerical image descriptors that measure the similarity of images by extracting archaeologically relevant characteristics of the bifaces in the images. We evaluated 13 different methods for extracting texture information and compared their performance of pulling out shape and texture information to characterise similarity with respect to the query image. Shape information is extracted from images by segmenting the artefacts and extracting features like scale-normalised length and width of the object’s bounding box, scale-normalised area, and scale normalised breadths of the artefact at 20% and 80% along its length. These features are correlated well with “tool type” metadata from images in the database. Thirteen different texture descriptors were evaluated and all performed generally well as predictors of the “raw material type” metadata from images in the database. Table I lists the methods from the computer vision literature used to build our 13 texture descriptors.

These shape and texture descriptors are fused into a single numerical descriptor for the query image, which is then compared with the corresponding descriptors of images in the archive. We retrieve the top 100 images with the smallest differences from the query descriptor.

The best-performing methods for extracting shape and texture representation were able to retrieve sets of images with total metadata similarity that was, on average, 90% as high as the total metadata similarity of the 100 most similar images when using a leave-one-out cross-validation strategy against the ground truth metadata for the biface images. This means that our method mostly retrieves results that have high-likelihood of being highly relevant to users.

Table II. Token-Level Features for Automatic Metadata Extraction

Feature	Description
Word	The original word form
lower	The lower-cased form of the word
Length	The length of the word in characters
Has_cap	Does this word contain an upper case letter
Init_cap	Does this word start with an upper case letter
All_cap	Does this word consist solely of upper case letters
Has_digit	Does this word contain a digit
All_digit	Does this word consist solely of digits
Has_punct	Does this word contain punctuation characters (dashes, periods, hyphens)
All_punct	Does this word consist solely of punctuation characters
alnum	Does this word consists solely of letters and numbers
prefix	All prefixes of this word of length one up to four
suffix	All suffixes of this word of length one up to four
shape	The shape of the word: map upper case to "A," lower case to "a," digits to "0," and punctuation to "."
Summarised_shape	Same as shape, but collapse runs of identical consecutive characters: map "Aaa-AAA00." to "A a-A0."

3.3 Text Mining of Research and Grey Literature

The textual datasets in the archives have the same problems as the visual datasets with regards to the available metadata. To overcome the lack of descriptive metadata fields, a text-mining pipeline was developed to automatically extract salient phrases. Our approach builds upon the earlier work undertaken in the Archaeotools [Jeffrey et al. 2009] project and on work on the Contextualizing Media Research Data (CoMeRDa) toolkit [Bron et al. 2013]. The aim in both cases is to automatically extract the metadata fields to supplement the current manually added fields. The overarching guideline to decide what is extracted is the what, where, when, and who questions. Specifically, in this context, we are interested in what was found, where it was found, from what time period it originated, and who found it. This information is reflected in the existing item metadata fields, but due to the current fields being the result of a manual process, the coverage of existing metadata is sparse.

The automatic metadata extraction is cast as an information extraction task, and one that is strongly related to the more common named entity recognition (NER) task [Nadeau and Sekine 2007]. In the NER task, one is interested in extracting all proper nouns, denoting persons, locations, and organisations. The complicating factor for the metadata extraction is that the fields are mostly nominal entities, with the exception of the who and where fields, and as such, have a higher degree of variability (e.g., "worked flint," "charred chaff fragments," "mid-late iron age")

This information extraction task has a typical supervised machine learning setup. Using the manually annotated documents from the Archaeotools project, we trained a first order linear chain conditional random field (CRF) [Sutton and McCallum 2006] with the token level features from Table II. To enable the CRF to also take the context of the current word into account, we enrich its feature set by including the token level features of the two preceding and two following tokens.

The automatically extracted fields were then added to the existing metadata fields of the items in the datasets, if they existed, or were added as new fields where they did not exist before. This allowed us to have a more consistent set of metadata for each individual collection, as well as across the different collections, increasing the usability of the search interface.

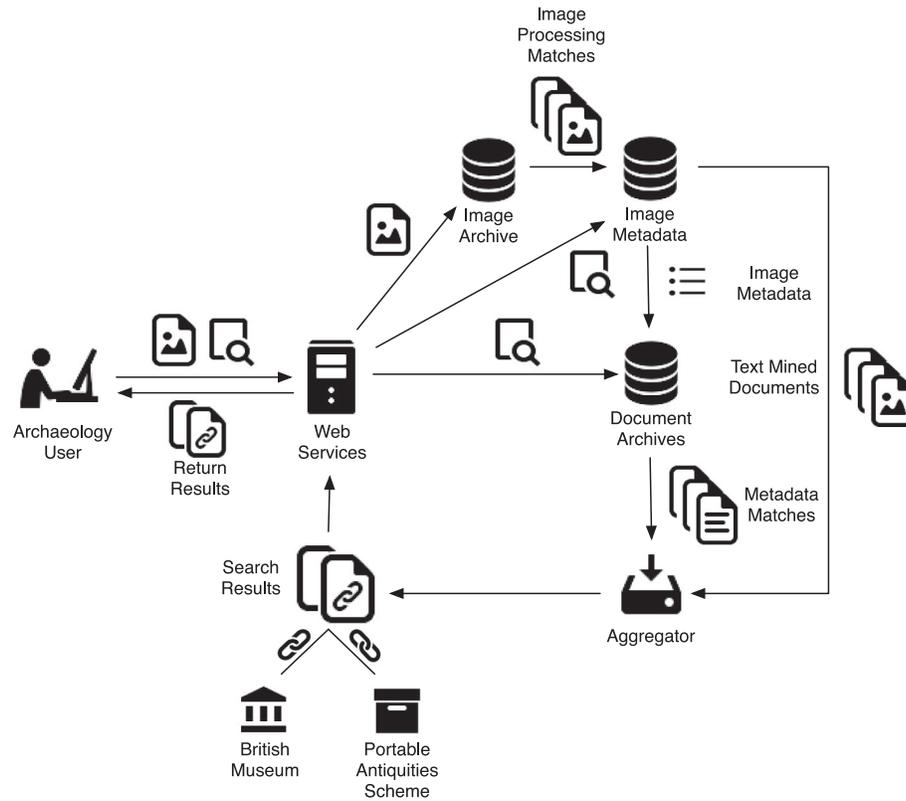


Fig. 2. The DADAISM Architecture.

3.4 System Architecture

We created the prototype as a Web2Py [Di Pierro 2011] application. The Web2Py framework allows the easy creation of multi-tier web applications with clear separation of interface, control, and data concerns.

The interface components were implemented using a combination of HTML5, CSS3, JQuery, and Python, and were connected to the data sources via a set of lightweight web services. This allowed us to create a web application that could sit largely separate from the data sources, meaning that the principles that guided the development for the flint dataset could be applied to any dataset that became available from the ADS with appropriate data and metadata, and pushes us further away from the resource centric approaches.

The architecture for the DADAISM system is presented in Figure 2. In this architecture, the archaeology user submits either an image to query the archives or a more traditional keyword search.

When a user submits a query image of an artefact, it is passed to an image processing component that matches visual characteristics of the artefact to similar images. The matched images form an image result set from which the shared metadata items are extracted. The shared metadata is then passed to a query engine that searches both the resource level metadata and text-mined content level metadata to identify documents from the grey literature and other research document datasets that may be relevant to the image. The document result set and the image result set are passed to an

aggregator component that passes all of the results back to the interface for presentation, along with a set of links that are created for the common metadata that trigger searches in popular archives, such as the British Museum¹ and the Portable Antiquities Scheme.²

When a user inputs text-based search, the process is very similar with the exception that the image processing component is bypassed, and metadata for all resources are used to retrieve items relevant to the terms.

3.5 Interaction Design

The prototype interface was created using the instrumental interaction design paradigm [Beaudouin-Lafon 2000, 2004; Beaudouin-Lafon and Mackay 2000]. Instrumental interaction emphasises the use of *domain objects*, which are built from the mental models of the users and often have analogues in real-world tasks similar to traditional metaphor-based interfaces. This design paradigm also encourages use of *interaction instruments* that allow the manipulation of domain objects by the users, often reifying multiple commands commonly performed together into one action.

For purposes of design, we identified four key domain objects that individuals work with through their tasks. The first is a *Query*, where the user asks for a set of objects matching particular criteria. The *Query* object in most archives is a text search box which, while useful for searching text in documents, often has low compatibility with what the archaeologist is trying to accomplish. The archaeologist users are often trying to translate what they see into words, which may not be entirely represented by metadata schema.

The second domain object that individuals work with is a *Resource*. Archive *Resources* are the complete set of the data object that the user is interested in (e.g., image of a flint tool, grey literature document for an excavation site) with the metadata describing both the data object (i.e., resource level metadata) and the contents of the object if available (i.e., content level metadata). Users are typically interested in viewing all of this information as a set and will want to see enlarged or enhanced versions of the contents of the resource alongside metadata. Resources are collected together, grouped and categorised by archaeologists, and they often have notes attached to them by the users that are dependent on the task the users are undertaking at the time. For example, a user may be working with images of flint objects as a result of an excavation, and record notes regarding the identifying characteristics. Later, the user may repurpose an image for a lecture and want to change the notes to be more closely related to the lecture material.

The third domain object is that of a *Result Set*, which is a collection of objects returned from the system to the user in response to a request.

Finally, one of the key domain objects for our design is that of a *Personal Collection*. Users need to be able to maintain a variety of collections of artefacts that are related to their own projects and resource needs. Due to the importance of these in maintaining an efficient workflow, the adding, removing, and organisation of objects needs to be closely associated with the Requests and Result Sets, as opposed to being a separate portion of the interface.

In order to accommodate all of these objects and their relations, we created the interface that is shown in Figure 3.

The interface is divided into two different panels, the Search panel and the Collections panel. Each panel isolates the user to particular tasks, with the Search panel providing the user access to data from the archives while the Collections panel is for maintaining and inspecting Personal Collections.

¹https://www.britishmuseum.org/research/collection_online/search.aspx.

²<https://finds.org.uk/>.

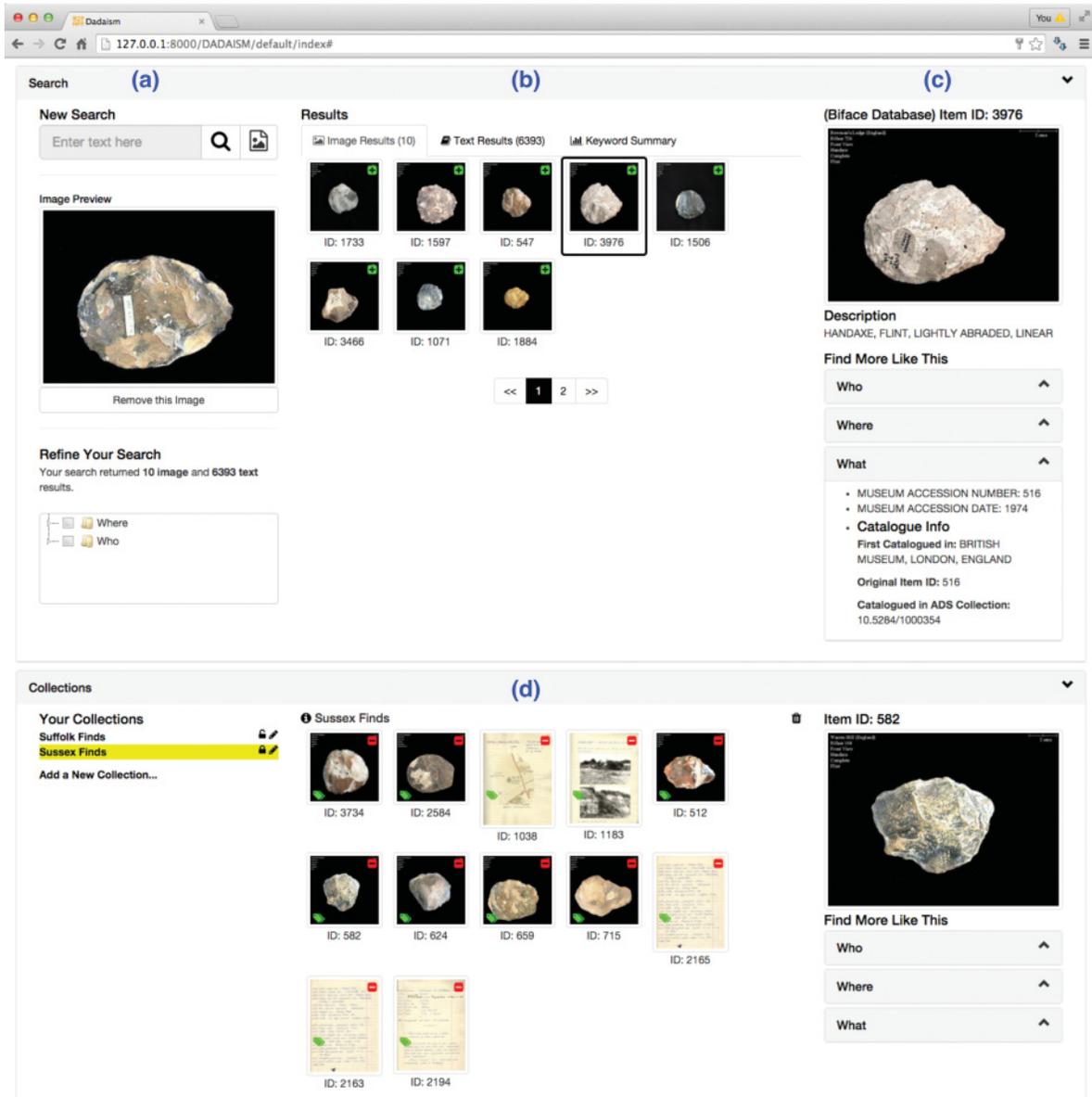


Fig. 3. DADAISM Interface capturing all of the key domain objects and their instruments.

The left hand side of the search panel (labelled (a) in Figure 3) manages the queries from the users. The users may either provide a set of query terms or they may upload an image. After a query has been executed and a result set appears, the users may refine their criteria through a series of filtered facets for where the object was found or who contributed the resources to the archive. This allows users to undertake the iterative refinement of their result set over time. This facet refinement instrument is tightly coupled to the result set so that it refreshes the result set when activated, thus providing immediate feedback, and with a low spatial offset to both the original query and the result set, keeping

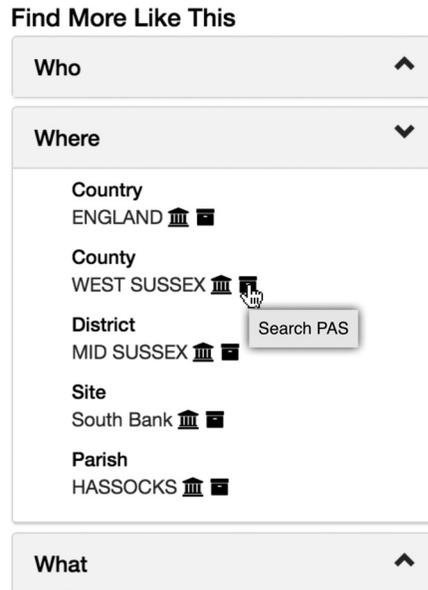


Fig. 4. Links that will generate new queries from common metadata.

with Beaudoin-Lafon's model of maintaining low degrees of indirection between instruments and the objects on which they operate.

The result sets are managed and explored in the centre of the search panel (labelled (b) in Figure 3). Each type of resource is divided into its own paginated tab, which are further subdivided by archive dataset if appropriate. Finally, a tab with counts of common keywords is provided to give archaeologist users an overview relationships between the datasets.

This panel has been instrumented in such a way that users can get additional information about a particular artefact. Selecting an artefact enlarges the resource image and will retrieve the key metadata associated that describe it. The metadata, as shown (labelled (c) in Figure 3), is instrumented such that users can click on the links for the British Museum and the Portable Antiquities Scheme to find more resources in those archives as shown in Figure 4.

Secondly, users can add items to their personal collections (labelled (d) in Figure 3). When they do so, the object immediately appears below in the Collections panel. The Collections panel allows the user to create their own named collections based on their projects. Objects are copied from the archive collection to the personal collection to allow users to build and categorise objects in ways that suit their activities. This area is presented consistently with the search area, with users being able to view their collections and interact with individual objects to get more information about them.

4. EVALUATION

We undertook a task-based user evaluation to explore the success of the design in improving access to archive images and documents. Participants undertook a representative task in the DADAISM interface, similar to those that were seen in the contextual inquiry sessions. Participants were asked to do the following task:

While out walking near Warren Hill in Suffolk, you discover a flint artefact on the ground. You decide that you would like to learn more about the artefact through the Archaeological Data Service

and their online archives. When you return home, you take a picture of your artefact, and you would like to compare it to other items in the archive. You want to answer the following questions:

- What type of artefact have you found?
- What is the approximate age of the artefact that you have found?
- Are there any similar artefacts that have been found in that area, and if so, what are some examples?

These questions were meant to encourage individuals to build their own collections, explore the different options available in the interface, and explore beyond into other archives. We did not record or count the accuracy of the answered questions as a measure in this evaluation.

For this evaluation, we collected a set of usability problems encountered by users and asked participants to complete a standardised usability questionnaire, specifically the Lavie and Tractinsky aesthetics and perceived usability scale [Lavie and Tractinsky 2004]. This scale captures five different characteristics of interfaces. Specifically, it captures the following aspects of user experience:

- Classical aesthetics: whether the website adheres to characteristics of clean and clear design, such as symmetry and grouping
- Expressive aesthetics: whether the website is creative and original in the opinion of the users
- Perceived usability: whether the users perceive that the site is easy to use and allows them to achieve their tasks
- Service quality: whether the users can trust the results the system provides
- Pleasure: whether using the site is a pleasurable experience

We were particularly interested in the first three, but also comment on the other two aspects.

4.1 Participants

Thirteen participants, six male and seven female, were recruited from across the cultural heritage community in the UK. While it was important to have experienced archaeologists represented in the set of participants, we also wanted to explore whether this interface would be appropriate for newcomers to the archaeology field, but who have experience with other types of archives.

Six of the participants were research archaeologists with 2–21 years of experience working in the area (Mean = 5; SD = 8), and there were a further four students of archaeology who were new to the field. There were three members of the cultural heritage community who work regularly with archives, including an archivist, a cultural heritage manager, and a student of conservation studies for historical buildings.

4.2 Materials and Setting

The participants used an instance of the DADAISM interface installed on a Macbook. The system was a local Web2Py server (version 2.14), with all participants using the Google Chrome browser (version 48).

All demographic and research data questionnaires were distributed by paper, along with information sheets and consent forms. Participants undertook tasks in a quiet environment without interruption, either in their own offices or in the Interaction Labs at the University of York.

4.3 Procedure

Participants were provided with an information sheet about the project and the study they were about to undertake and then an informed consent form. Participants were not given any training on the interface, as we were interested in their first-use experience. When the participants were done their

task, they completed the Tractinsky and Lavie scale and were asked for their three favourite and least favourite things about the interface.

After participants undertook the task, they were asked to participate in a retrospective verbal protocol [Van Den Haak et al. 2003]. In this protocol, participants watched a recording of their session with the system and explained what they were thinking and doing during the task. This allowed the researcher to understand what participants were thinking at the time, highlighting where and why they encountered problems. Participants were instructed that when they encountered something they considered to be a problem in the interface, they should describe the problem to the researcher and rate the severity of the problem on a scale from 1–5, where 1 is a cosmetic problem and 5 is a very major problem.

4.4 Results

Participants spent between 9 and 39 minutes using the interface (Mean = 18.5, SD = 9.5). A total of 140 usability problems were recorded across the 13 participants (Mean = 10.1, SD = 5.0).

From the outcomes of the evaluations, there were a number of very positive first impressions of the new interactive system, as well as a set of common usability problems that are of interest for the next phase of design of this system and future systems. The following sections discuss the outcomes of the evaluations.

4.4.1 Image Search and Query Complexity. Many of the participants were very positive about the ability to upload a photo to the system and do an image search. They were able to quickly enter into a set of search results and begin to explore within a few seconds of working with the system. This was a particularly positive aspect of the interaction with participants, finding the experience quite novel:

“I thought it was really good, actually, the image search; I’ve not come across that before in an online archive, so I thought that was really useful.” – Participant 5

Further, the search was, in general, felt to be quite accurate at bringing back a relevant set of image resources:

“I was really impressed with the speed and the accuracy of the image search. When that came up, I was like, ‘Whay! Look at that, there’s loads of things that match mine. That’s brilliant.’” – Participant 6

However, with the image search being present, there were a number of compromises that needed to be made in the interface. For example, we made the choice that once a user had chosen to do an image search, the text search would be de-activated to avoid confusion. Unexpectedly, most of the participants wanted to do further refinement by keyword in place of using the facets, thus combining the two different types of query.

There were also issues around participants wanting to do fairly complex text searches with advanced conjunctions or disjunctions in terms. We did not provide an advanced Boolean search functionality, as such query interactions are often error prone [Nahl and Harada 1996], and we felt that with the addition of the image search, there would be less of a need for this functionality. However, this specialist group of users is very experienced with advanced search capabilities, and found its absence conspicuous.

4.4.2 Visual Real Estate. One recurring theme throughout the usability problems relate to the amount of space available on the screen for each component. We built the application so that it was both responsively resized to the size of the browser window and optimal at a common full screen

resolution (1,366×768 pixels), with an attempt to use spacing and symmetry of the upper and lower half of the screen to build natural groupings of controls.

Many participants commented that they wanted to be able to reconfigure the screen in different ways, usually looking to fill the real estate with specific components, dependent on where their focus was at the time. For example, participants commented that they would like to expand the panel with result sets to fill the visible real estate and explore more items more quickly. In other cases, they wanted to be able to expand or even “pop out” individual artefacts from the right hand side of the interface, allowing them to see the images in a larger format.

4.4.3 Personal Collections. One important consequence of these real-estate issues was that the collections had a tendency to slip below the visible portion of the screen after the participant began exploring the result sets. If the participants did not notice the collections before searching, they had a tendency to miss that the functionality was available, making the icon to add a resource to collections perplexing. In order to address this, more feedback is needed to users on adding objects when that collection is out of the field of view. Both Participant 1 and Participant 16 commented on this in their retrospective sessions:

“Because it’s so cleanly designed—It looks like its cleanly designed to fit on that one screen, so I’ve assumed subconsciously that there’s nothing down there. So this whole idea of what I’m going to do with these things—how am I going to add them into my basket or collection?” – Participant 1

“I hadn’t realised that I can collect things, so I’ve been pressing the little plus, thinking it would go somewhere, but I hadn’t scrolled down to see that you can make your own collection here...”
– Participant 16

However, despite these initial setbacks with the personal collections, participants were generally positive about the ability to create and maintain their own personal collections with one participant saying:

“The idea of the collection is really useful. That you can add bits in there—pictures, images, and texts, I really like that idea. I can see that if I was doing research on a thing, a brooch or something that would be much, much better than trying to collate it all on Word, or even just downloading everything and putting it in a folder on my computer.” – Participant 16

It is worth drawing attention to the fact that the participant, who had not been part of the initial contextual inquiry session, discussed exactly the behaviours we had observed in those sessions. This emphasised to us that this system has merit in the workflows of archaeologists.

4.4.4 Metadata Presentation. Many of the participants were very positive about being able to get metadata terms summarised in a variety of different places. For example, in regards to the descriptive metadata under the enlarged images, one participant said:

“I like the description that comes up underneath with the keywords, because it was giving me immediately clues as to what to look for, similarities between certain artefacts.” – Participant 6

Further, the summary information regarding the objects which gave individuals an idea of where to search next was positively received by participants:

“That’s another thing I thought was quite good – Using the description from the Warren Hill artefact, I could find a lot of other artefacts.” – Participant 3

Table III. Mean and Standard Deviation for the Factors of the Lavie and Tractinsky Standardised Scale

	N	Mean	Std. Deviation
Classical aesthetics	13	3.74	0.39
Expressive aesthetics	13	3.23	0.48
Perceived usability	13	3.45	1.02
Service quality	13	3.28	0.85
Trust	13	3.13	0.93

The participants were all very positive regarding the access to the grey literature, with the ability to find information within the grey literature documents being very welcome as epitomised by the following participant comment:

“For the text search, this instant link to the grey literature is wonderful. This is so nice, rather than having to go through multiple databases and search for the grey literature—just going straight there... god, it makes a change. It’s fantastic.” – Participant 1

However, there were a number of issues that need to be looked at more closely through further design work. Users did not find the “Find more like this” section, with information about the object and links to external archives, as helpful as we expected. There were individuals who used this information successfully, such as Participant 1:

“This first [icon] that I clicked on took me straight to the PAS database which is fantastic, and took me to directly relevant finds, which was fantastic.” – Participant 1

However, many participants struggled with these options. Looking at the descriptions of the problems users were encountering, it appears that this related to two different and distinct issues. First, a number of participants discussed the unfamiliarity of the icons, and they tended to not explore them to identify them through tool tips provided. This likely is in relation to the deep engagement users had with the task, leading them to not interrupt their sensemaking tasks to work out ambiguities in the interface.

Secondly, and more importantly, there were a number of participants that questioned the contents of the information box. While they understood what the links did, the information we had provided, such as who contributed the resource, where it was found, or where it was housed, was not seen as helpful in moving the search forward. Improving this feature in the future will require user research in the future to understand what terms would be most useful to archaeologists in their tasks.

4.4.5 Standardised Questionnaire. Overall, the interface was rated above the midpoint for all of the aspects of the Lavie and Tractinsky scale, which was a very positive response from these participants, pointing to a usable and pleasing experience for the users. The means ratings and standard deviations for each factor are presented in Table III.

A set of one sample t-tests was performed on each factor of the scale, comparing the value to the midpoint of the scale. Participants’ ratings of Classical Aesthetics was significantly above the midpoint of the scale ($t = 6.7$, $df = 12$, $p < 0.00$), however, all other factors were found to not have statistically significant differences. The interface does appear to be aesthetically pleasing, which is excellent given complexity of the components on the screen. Unfortunately, we cannot make any further generalisations about the other factors.

5. DISCUSSION

In this article, we have presented the design for a new interactive system for archaeologists doing information seeking in online archives. In particular, we have focussed on the use of image archives of artefacts and related documents in grey literature and research notes.

The contextual enquiry sessions that we conducted with archaeologists as part of the user-centred process pointed to archaeologists having a very distinct way of working with these archives, specifically doing large amounts of rapid reformulation of queries and retrieval before moving to a sensemaking phase of integrating information across several different resources.

The interactive system we designed has successfully integrated both: the capability of image-based retrieval and improved metadata to provide access to documents that were previously difficult to work with in the archive. In general, participants responded positively to both of these aspects of the system, and we had very positive responses regarding the creation and sharing of personal collections.

Future work will investigate how metadata links can be better used to help archaeologists progress through the early retrieval stages, and will investigate other types of artefacts beyond flint where the image processing and text-mining approaches would be applicable. Further, we have only implemented very basic sharing functionality as part of this prototype. Due to the heterogeneity of behaviour that was observed in the contextual enquiry session, there is a need to do further user research into the collaboration activities that occur during the sensemaking phase of the information - journey before enhancing this design.

The user research conducted in DADAISM, and the subsequent research prototype that was designed and evaluated, provides a first step towards new interactive systems that better support the broader workflows of archaeologists in online archives.

ACKNOWLEDGMENTS

The authors would like to thank all of the participants for their time during the contextual inquiry sessions and the evaluation.

REFERENCES

- J. Arróspide and L. Salgado. 2013. Log-gabor filters for image-based vehicle verification. *IEEE Transactions on Image Processing* 22, 6 (2013), 2286–2295.
- M. Beaudouin-Lafon. 2000. Instrumental interaction: An interaction model for designing post-WIMP user interfaces. In *Proceedings of the 18th International Conference on Human Factors in Computing Systems – CHI’00*. ACM Press, 446–453.
- M. Beaudouin-Lafon. 2004. Designing interaction, not interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. 15–22.
- M. Beaudouin-Lafon and W. E. Mackay. 2000. Reification, polymorphism and reuse: Three principles for designing visual interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. 102–109.
- L. Bernard, D. Ertlen, and D. Swartz. 2014. ArkeoGIS: Merging geographical and archaeological datas online. In *Proceedings of 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology (CAA’14)*. Archaeopress, 401–406.
- M. Bron, M. de Rijke, J. van Gorp, A. Vishneuski, F. Nack, and S. de Leeuw. 2012. A subjunctive exploratory search interface to support media studies researchers. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR’12)*. ACM, 425–434.
- M. Bron, J. van Gorp, and M. de Rijke. 2016. Media studies research in the data-driven age: How research questions evolve. *Journal of the Association for Information Science and Technology* 67, 7 (July 2016), 1535–1554. (Online June 2, 2015.).
- M. Bron, J. van Gorp, F. Nack, L. B. Baltussen, and M. de Rijke. 2013. Aggregated search interface preferences in multi-session search tasks. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR’13)*. ACM, 123–132.
- A. Chalechale, A. Mertins, and G. Naghdy. 2004. Edge image description using angular radial partitioning. *IEEE Proceedings on Vision, Image and Signal Processing* 151, 2 (2004), 93–101.

- J. C. Chapman. 2010. Observing users: An empirical analysis of user interaction with online finding aids. *Journal of Archival Organization* 8, 1 (February 2010), 4–30. DOI: <http://dx.doi.org/10.1080/15332748.2010.484361>
- M. Charno, S. Jeffrey, C. Binding, D. Tudhope, and K. May. 2012. From the slope of enlightenment to the plateau of productivity: Developing linked data at the ADS. In *Proceedings of the 40th Annual Conference of Computer Applications and Quantitative Methods in Archaeology (CAA'12)*, Earl Graeme et al. (Eds.). Amsterdam University Press, 217–223.
- A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina. 2012. An efficient algorithm for fractal analysis of textures. In *Proceedings of the 25th SIBGRAPI Conference on Graphics, Patterns and Images*. IEEE Computer Society, 39–46.
- M. Di Pierro. 2011. web2py for scientific applications. *Computing in Science & Engineering* 13, 2 (March 2011), 64–69.
- A. Duploux, V. Capozzoli, and A. Zambon. 2014. An inventory of Lucanian heritage. In *Proceedings of the 42nd Annual Conference on Computer Applications and Quantitative Methods in Archaeology (CAA'14)*, F. Giligny, F. Djindjian, L. Costa, P. Moscati, and S. Robert (Eds.). Archaeopress, 333–340.
- T. Elena, A. Katifori, C. Vassilakis, G. Lepouras, and C. Halatsis. 2010. Historical research in archives: User methodology and supporting tools. *International Journal on Digital Libraries* 11, 1 (March 2010), 25–36. DOI: <http://dx.doi.org/10.1007/s00799-010-0062-4>
- D. Ellis and M. Haugan. 1997. Modelling the information seeking patterns of engineers and research scientists in an industrial environment. *Journal of Documentation* 53, 4 (October 1997), 384–403. DOI: <http://dx.doi.org/10.1108/EUM0000000007204>
- Z. Guo, Z. Zhang, X. Li, Q. Li, and J. You. 2014. Texture classification by texton: Statistical versus binary. *PLoS ONE* 9, 2 (2014), e88073.
- H. R. Hartson and J. C. Castillo. 1998. Remote evaluation for post-deployment usability improvement. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI'98)*. ACM Press. DOI: <http://dx.doi.org/10.1145/948496.948499>
- K. Holtzblatt. 2009. Contextual design. In *Human-Computer Interaction: Development Process*. CRC Press.
- K. Holtzblatt and S. Jones. 1993. Contextual inquiry: A participatory technique for system design. In *Participatory Design: Principles and Practices*. 177–210.
- I. Huvila. 2007. To whom it may concern? The users and uses of digital archaeological information. In *Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA 2007)*, Axel Posluschny, Karsten Lambers, and Irmela Herzog (Eds.). Archaeopress, 1–8.
- S. Jeffrey, W. Kilbride, S. Waller, and J. Richards. 2007. Thinking outside the search box: The common information environment and archaeobrowser. In *Proceedings of the 35th International Conference on Computer Applications and Quantitative Methods in Archaeology (CAA'07)*, Axel Posluschny, Karsten Lambers, and Irmela Herzog (Eds.). Dr. Rudolf Habelt GmbH, 2006–2011.
- S. Jeffrey, J. Richards, F. Ciravegna, S. Waller, S. Chapman, and Z. Zhang. 2009. The archaeotools project: Faceted classification and natural language processing in an archaeological context. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 367, 1897 (2009), 2507–2519.
- G. Klein, B. Moon, and R. R. Hoffman. 2006a. Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems* 21, 4 (2006), 70–73.
- G. Klein, B. Moon, and R. R. Hoffman. 2006b. Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems* 21, 5 (2006), 88–92.
- P. Kovesi. 2000. Phase congruency: A low-level image invariant. *Psychological Research* 64, 2 (2000), 136–148.
- T. Lavie and N. Tractinsky. 2004. Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-computer Studies* 60, 3 (2004), 269–298.
- B. S. Manjunath and W.-Y. Ma. 1996. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 8 (1996), 837–842.
- G. Marchionini. 2006. Exploratory search. *Communications of the ACM* 49, 4 (April 2006), 41. DOI: <http://dx.doi.org/10.1145/1121949.1121979>
- D. Nadeau and S. Sekine. 2007. A survey of named entity recognition and classification. *Linguisticae Investigationes* 30, 1 (2007), 3–26.
- D. Nahl and V. H. Harada. 1996. Composing boolean search statements: Self-confidence, concept analysis, search logic, and errors. *School Library Media Quarterly* 24, 4 (1996), 199–207.
- M. B. Rosson and J. M. Carroll. 2009. Scenario based design. In *Human-computer Interaction*, Andrew Sears and Julie A. Jacko (Eds.). CRC Press, 145–162.
- A. Serenko and O. Turel. 2010. Rigor and relevance: The application of the critical incident technique to investigate email usage. *Journal of Organizational Computing and Electronic Commerce* 20, 2 (April 2010), 182–207.

- D. Sinn. 2012. Impact of digital archival collections on historical research. *Journal of the American Society for Information Science and Technology* 63, 8 (August 2012), 1521–1537. DOI: <http://dx.doi.org/10.1002/asi.22650>
- C. Sutton and A. McCallum. 2006. An introduction to conditional random fields for relational learning. In *Introduction to Statistical Relational Learning*. 93–128.
- M. Topi, O. Timo, P. Matti, and S. Maricor. 2000. Robust texture classification by subsets of local binary patterns. In *Proceedings of the 15th International Conference on Pattern Recognition*. 935–938.
- M. Van Den Haak, M. De Jong, and P. J. Schellens. 2003. Retrospective vs. concurrent think-aloud protocols: Testing the usability of an online library catalogue. *Behaviour & Information Technology* 22, 5 (2003), 339–351.
- C. Zhu, C.-E. Bichot, and L. Chen. 2013. Image region description using orthogonal combination of local binary patterns enhanced with color information. *Pattern Recognition* 46, 7 (2013), 1949–1963.

Received April 2016; accepted August 2016