

The Source to Output Repositories Project: Archaeology

Issued to:
Graham Pryor
StORe Project Manager

Project: **Project StORe (Source to Output Repositories)**

10th July 2006

By Daniel Hull

University Library and Archives
University of York
Heslington
York
YO10 5DD
UK

Tel: +44 (0)1904 321134 / 433954
e-mail: djh121@york.ac.uk

ISSUE DETAILS

	Name	Signature	Date
Prepared by	Dan Hull		
Checked by	Graham Pryor Elizabeth Harbord Stuart Jeffrey		
Issue Status			
Purpose of issue			
Authorised for issue by			
Word Ref.			

AMENDMENT DETAILS

Revision No.	Date	Affected Pages	Date Change Made	By (initials)
1				

Contents

Executive Summary	4
List of Figures	6
List of Tables	6
1) Project Aims	7
2) Background: Repository Development in Archaeology	8
2.1 Source repository: The Archaeology Data Service	8
2.2 Output repository: White Rose Consortium ePrints	8
3) Summary of Significant Observations	12
3.1 Identities	12
3.2 Project aims	12
3.3 Source data	13
3.4 Source repositories	13
3.5 Metadata	14
3.6 Data access and sharing	14
3.7 Output repositories	14
3.8 Support	15
4) Questionnaire Analysis	16
4.1 Selection of survey constituency	17
4.2 Questionnaire responses	17
4.3 Identities	17
4.4 Project aims	18
4.5 Source data	19
4.6 Metadata	24
4.7 Output repositories	28
4.8 Preferred routes to output repositories	30
4.9 Information support	32
5) Interview Analysis	33
5.1 Selection method	33
5.2 Interview technique	33
5.3 Identities	34
5.4 Project aims: Source to Output links	35
5.5 Project aims: Output to Source links	36
5.6 Project aims: additional features	38
5.7 Further suggestions	39
5.8 Source data	41
5.9 Metadata	43
5.10 Data access and sharing	45
5.11 Output repositories	47
5.12 Information support	48

6) Additional Issues	50
6.1 Identities	50
6.2 Project aims	50
6.3 Source data	51
6.4 Source repositories	51
6.5 Metadata	55
6.6 Data access and sharing	55
6.7 Output repositories	56
6.8 Open Access	57
6.9 Charging policies	58
6.10 Thesauri and Word Lists	59
6.11 Information support	59
Bibliography	60
Appendix 1: List of interview questions	61
Appendix 2: Scenarios	65

Executive Summary

The StORe Project commenced in September 2005, with the aim of developing ways of enabling repositories of published reports and papers to interact directly with repositories of source data from which they are derived. This report constitutes Workpackage 2 in that development, and presents the results of an online questionnaire sent to researchers, and a series of one-to-one follow-up interviews. The current report comments only on the results from the discipline of archaeology, though some comparison is made with the rest of the survey constituency. A summary of significant observations deriving from this survey can be found in section 3 of this report.

List of Figures

1	Pie chart showing the content of the White Rose ePrints IR by subject area	9
2	Pie chart showing the subject areas represented in White Rose IR content at York	10
3	Proportions of archaeology content in the White Rose IR compared to staff numbers	10
4	Bar chart showing the numbers of questionnaire respondents for each community	11
5	Bar chart showing the file types produced by archaeologists, in descending order	22
6	Bar chart showing the file types produced by archaeologists, in descending order	22
7	Bar chart showing the relative importance of different metadata types for archaeologists	25

List of Tables

1	The identity of questionnaire respondents for archaeology and the total survey constituency	18
2	Response to Question 2: attitudes towards source to output linkage	18
3	Response to Question 3: attitudes towards output to source linkage	18
4	Attitudes to source to output linkage by identity	19
5	Attitudes to output to source linkage by identity	19
6	Source data formats produced by archaeologists	20
7	Free text entries indicating further source data formats	20
8	Source data formats produced by the survey constituency as a whole	21
9	Number of respondents who produce a combination of data formats	21
10	Data formats by source repository for archaeology	22
11	File types produced by repository for archaeology	22
12	Stages at which archaeologists assign metadata	24
13	Metadata types which archaeologists would like assigned to their data sets	24
14	Free text additions to a question on metadata types	25
15	Levels of repository support by stage at which metadata is assigned	26
16	Levels of repository support, by personnel who enter the metadata	27
17	Forms of support/guidance received by repository users	27
18	Non-users of output repositories	28
19	Non-depositors in output repositories	28
20	Free text responses indicating output repositories used	28
21	Repository types used by archaeologists	29
22	Search levels employed by archaeologists	29
23	Free text responses commenting on search levels used	29
24	Search levels used by output repository type	30
25	Preferred routes into output repositories	30
26	Routes into output repositories pursued by ADS depositors	31
27	Line graph showing routes into output repositories pursued by ADS depositors, versus non-depositors	31
28	Archaeologists who have not deposited in an output repository	32
29	Types of information support received by archaeologists	32
30	Free text responses indicating types of information support	32
31	List of interviewees, indicating their role, date interviewed and the method used	34
32	Free text responses indicating attitudes to source to output linkage	50
33	Free text responses indicating attitudes to output to source linkage	50
34	Number of archaeologists who produced a combination of data formats	51
35	Source repositories used by archaeologists for data deposition	51
36	Free text responses indicating additional source repositories used by archaeologists for data deposition	52
37	Free text responses indicating the stage at which archaeologists assign metadata	55
38	Reasons for accessing others' research data given by archaeologists	55
39	Methods by which archaeologists access others' research data	56
40	Levels of intermediation provided to archaeologists	59

1) Project Aims

The StORe Project 'seeks to address the area of interactions between output repositories of research publications and source repositories of primary research data. Through user surveys the functionality required by researchers in both types of repository will be determined, with the aim of identifying options for increasing the value of using primary data in source repositories as well as at the point where researchers submit to or download papers from output repositories. This two year project is multi-disciplinary in scope, embracing seven different scientific domains: archaeology, astronomy, biochemistry, biosciences, chemistry, physics and social sciences. With the aid of the survey results, general principles for middleware development to link source and output repositories together will be formulated, and a business analysis will be performed. Consequent to this analysis a pilot demonstrator will be developed in one of the domain areas. A full and extensive evaluation of the project will be carried out in order to inform JISC of the best options for future development in this area.' (Pryor 2006)

Within archaeology, some specific repository issues were already apparent prior to the launch of the StORe Project, which Workpackage 2 had the potential to elucidate further. In particular, although source repository development had developed rapidly since the creation of the Archaeology Data Service in 1997, the issue of output repository development has only recently begun to be addressed. There are very few discipline specific output repositories available within the discipline, with the exception of the ADS itself (which, though largely a source repository, can also be thought of as providing output material), the Council for British Archaeology, and the agglomeration service provided by the HEIRNET (Historic Environment Information Resource Network) portal, HEIRPORT. This lacuna has begun to be addressed recently with the inception of the OASIS Online Access to the Index of Archaeological Investigations project, and the AHRC-funded LEAP project (Linking Electronic Archives and Publications). In spite of these recent developments, it is still not clear to what extent researchers in archaeology desire access to output repositories. Following developments in source repositories, the StORe Project offers the opportunity to reflect more broadly on the attitudes held by archaeologists towards online stores of information within their own sub-discipline. For example, what apprehensions might archaeologists have about depositing in or using such stores in their research? What directions would they like such repositories to take in the future? What technical facilities could be developed and funded which are not currently in existence? How could their own research activities in particular benefit from the linkage of online publications to archives of data? What specific technical difficulties do they regularly face which inhibit their research?

Therefore, in addition to the remit defined by the StORe Project, it is hoped that the questionnaire and interview results presented here will be of use to repository managers in archaeology more generally, as they seek to build and enhance the services they offer in the future.

2) Background: Repository Development in Archaeology

In order to provide some context for the ways in which researchers currently use both source and output repositories, the following paragraphs outline the history and usage of each. An important difference between the user community of source and output should be pointed out; the source repository, the Archaeology Data Service (ADS), hosts data produced by (potentially) any archaeologist or archaeological project based in the UK. By contrast, the output repository, the White Rose Consortium ePrints service, although available for use by anyone, only holds output material produced by members of the universities of Leeds, Sheffield and York research communities. Consequently, the history and usage of each differs somewhat, as will be seen in the following description.

2.1 *Source Repository: The Archaeology Data Service (ADS)*

The ADS was formed in 1997, with the aim of supporting ‘research, learning and teaching with high quality and dependable digital resources’ (Richards 1996). Its creation was prompted by two considerations; the first is that there had been a great proliferation of digital data in archaeology throughout the 1990s in particular, but that insufficient guidelines, safeguards and facilities had been put in place to ensure that this data was as safely stored and as appropriately accessible as possible. The second consideration was that ‘archaeology is in a special position in that much of the creation of its data results from the destruction of primary evidence’ (Richards 1997). Consequently, it was formed with a view to acting as a central resource for cataloguing, preserving and providing access to dispersed digital data generated by archaeologists working within the UK. Such data consists of ‘text reports, databases (related to excavated contexts or artefacts, for example), images (including aerial photographs, remote sensing imagery, photographs of sites, features and artefacts), digitised maps and plans, numerical datasets related to topographical and sub-surface surveys and other locational data, as well as reconstruction drawings’.

In many cases, project archives consist of many different elements, generated over a broad span of time by a range of different researchers with varied affiliations and intentions. From the point of its creation, provision of storage and access facilities has necessitated liaison between ADS and a range of both central and local archaeological bodies, from individual data creators working in universities, commercial units or singularly, to local government Sites and Monuments Records (SMRs, now referred to as Historic Environment Resources, or HERs), the National Monuments Record, the OASIS project and others. Collaboration with such agencies has taken the form of negotiating deposition of project data, but also coordinating with other digital resources in the form of Portals, so that users can access the ADS archives from elsewhere, or access a broader range of sources from the ADS website.

Records held by the ADS derive from both HE-based research projects and the results of developer-funded work carried out by commercial archaeological companies. Often, however, projects are conducted by researchers from both parties. Likewise, use of the ADS appears to be fairly broad within the archaeological community. For the quarter February to April 2006, there were 2,714,894 successful requests made of the website, with an average of 1.029 gigabytes of data downloaded per day. Previously, analysis of the servers requesting access to the ADS website has suggested a wide range of organisations within the archaeology community (Kilbride and Winters 2001).

2.2 *Output Repository: The White Rose Consortium ePrints Repository*

The WR ePrints repository was formed in 2004 as part of the JISC-funded SHERPA project, with the aim of giving ‘access to peer-reviewed papers by authors from the Universities of Leeds,

Sheffield and York' (<http://eprints.whiterose.ac.uk/> 2006). Its broader remit is to, 'reduce access barriers to research output from the White Rose Consortium partners, provide an OAIPMH (Open Archives Initiative Protocol for Metadata Harvesting) compliant service, ensuring records are readily searchable and retrievable, bring together material currently distributed across institutional networks, and to provide a multi-disciplinary open access platform.

Its development thus far has been managed by a team of one representative from each White Rose University Library and the Repository Officer, Rachel Proudfoot. The approach was to identify copyright cleared content initially, and then to contact all academic departments and individuals systematically, in order to encourage deposition in the repository. At the same time advocacy for the repository idea was pursued through University committees and with senior University managers. Certain issues have hampered this development. For example, the fact that the Universities of York and Sheffield do not have institutional publications databases has slowed the process of identifying suitable publications. A further issue has been some resistance from academics about depositing in Open Access repositories. Reservations centre on the ambiguities of copyright arrangements and the desirability of having identical pagination in published and repository versions.

For archaeology, take-up has been slow but steady, with currently 51 articles in the repository. As the following charts demonstrate, this represents a relatively small proportion of the total content of the repository.

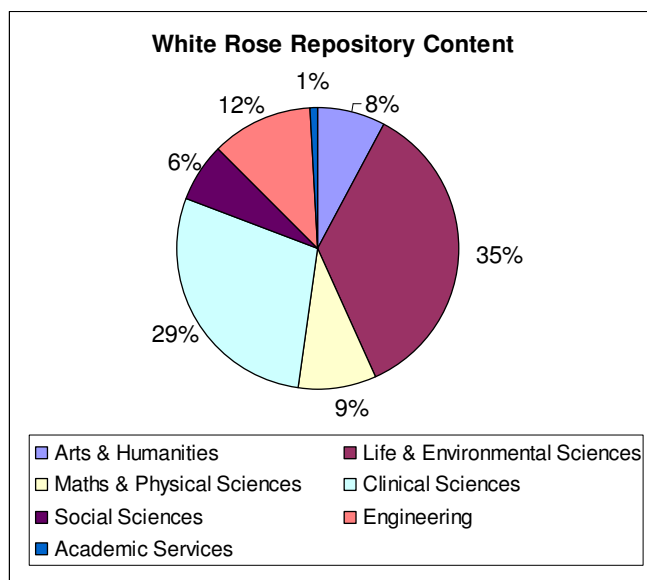


Figure 1: Pie chart showing the content of the White Rose ePrints IR by subject area

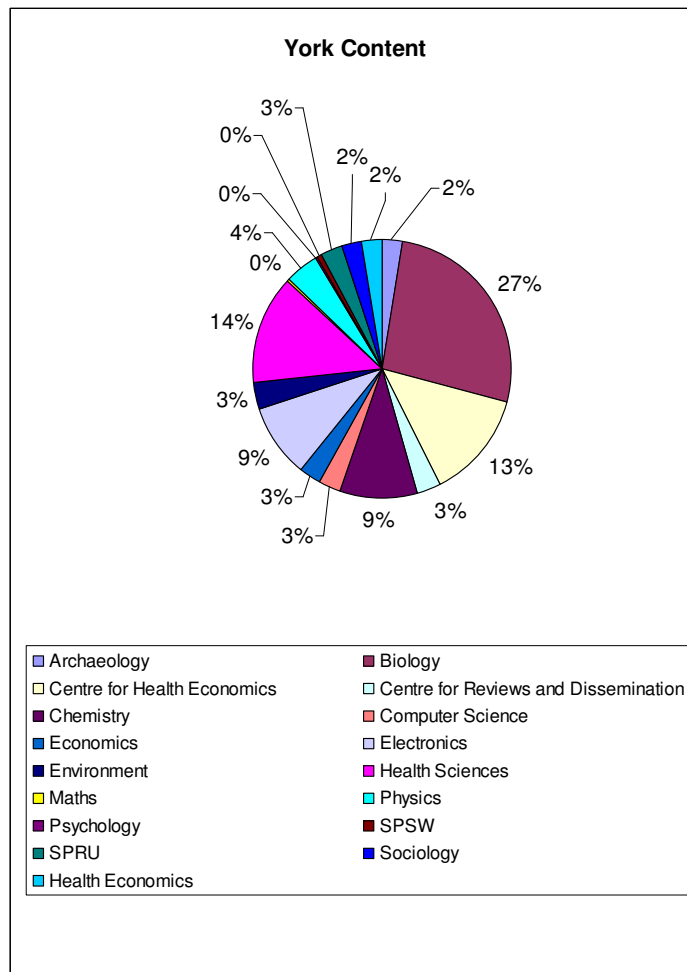


Figure 2: Pie chart showing the subject areas represented in White Rose IR content at York

However, this small proportion may in part be a reflection of the small size of Archaeology departments compared with, for example, Biology. The point could also be made that the range of different publication formats pursued by archaeologists is generally broader than for a subject like Biology, where publication in a relatively small number of journals is the norm. For archaeologists, this broad range of different publications makes clarifying copyright arrangements prior to deposition in the WR ePrints a more complex task than for some other disciplines. Although the ROMEIO database does much to clarify copyright arrangements for potential paper depositors, it is currently less well equipped to provide guidance for archaeologists.

Furthermore, within the White Rose Consortium, only Sheffield and York have full archaeology departments. Leeds has one archaeology member of staff. If the above figures are ‘normalised’ by calculating the number of articles according to the number of FTE academic members of staff in the three departments of York, Sheffield and Leeds, this figure does not seem so low:

Total academic staff at Leeds/Sheffield/York	3575
Total archaeology academic staff	48
Percentage of academic staff	1.34
Percentage of articles in the White Rose ePrints IR	2.00

Figure 3: Proportions of archaeology content in the White Rose IR compared to staff numbers

Although this number would appear relatively low, it is anticipated that many more academics are willing to contribute to the WRC ePrints, but just have not yet done so. Archaeology content could thus rise with continued efforts to raise awareness and talk academics through the process and requirements of deposition.

At the time of writing this report (July 2006), archaeology papers had been downloaded around 2,500 times from the WR ePrints (Rachel Proudfoot *pers comm.* 19.7.06).

As Kilbride and Winters have noted, usage statistics have to be treated with some degree of caution (2001). Cache settings can reduce the number of 'true' of hits, for example. Moreover, hits do not give an indication of how far the content of websites is *understood*, nor does it indicate how the information obtained there is actually *used*.

It should also be pointed out that although these two repositories formed the initial focus of the archaeology component of the StORe Project, it became apparent when calculating the potential survey constituency for archaeology that we would have to look well beyond the White Rose region. Consequently, the majority of the questionnaire and interview participants are not specifically users of the White Rose IR, though they were questioned on the issue of institutional repositories more generally.

3) Summary of Significant Observations

What follows is a summary of significant observations from the questionnaire **and** interviews, arranged according to the themes specified by the StORe Project Manager as part of **Workpackage 2: A Survey of Researcher Use of Repositories**. It should be born in mind that each of the observations listed below can only be properly understood within the context of the more detailed summaries within sections 4 and 5 of this report, and are not necessarily reliable conclusions when taken in isolation. This point is exacerbated by the relatively small rate of returns in the questionnaire phase, so caution must be exercised when seeking to use these results to argue for trends within the archaeological research community as a whole.

3.1 *Identities*

A total of 721 questionnaires were sent to archaeologists throughout the UK, 680 of whom were HE staff, postgraduates or research associates, and 41 were non-HE, based in local government departments, commercial companies and the national Portable Antiquities Scheme.

A total of 65 responses were received, representing 9.0%, 72.3% were HE, and 27.7% of whom were from the non-HE sectors. This may seem a disappointing turn out in comparison to a number of previous surveys of archaeologists (for example, the 29.3% response rate of the CBA's Publication of User Needs Survey; Jones et al 2001) but can be explained by the vacation timing of the questionnaire launch, a degree of 'survey fatigue' and perhaps an inherent lack of interest in the subject matter. Within the context of the StORe Project as a whole, the 9-10% achieved by most subject disciplines was in fact enough to provide an interesting spread of results.

The largest group of respondents defined themselves as 'University academic staff' (55.4%).

The interviews were carefully targeted to include five individuals who had not responded to the questionnaire, in order to ensure that a broad impression of different repository experiences was gained. 16 were interviewed in all.

It was clear from the interview phase that significant overlaps exist between different sectors in archaeology, and especially between deposition and use of HE and non-HE data sets.

3.2 *Project aims*

Two observations can be made about the notion of source to output repository development. Firstly, respondents from archaeology seem far more enthusiastic about the issue of source to output repository linkage (in either direction) than the overall survey constituency. Secondly, within archaeology there is a slight preference towards output to source linkage, rather than source to output linkage: 60.0% selected 'significant advantage to my work' with reference to source to output linkage, and 64.6% output to source.

The interviews revealed that the reasons for this enthusiasm centre on the potential for improved speed within the research process. The slight preference for the output to source direction can be explained by a desire to deconstruct the work of others in order to scrutinise their methodology and research process. Occasionally, researchers said that they incorporate others' data into their analyses. Supporters of source to output linkage cited the possibilities of enhanced research profile as the reason for their enthusiasm.

A minority of interviewees expressed pessimism about the likelihood that efficient and comprehensive source linkage could be managed without high levels of cost. One respondent actually stated that such linkage would be a distraction for the researcher.

Universal support was voiced for the construction of an index whereby researchers could discover who had made use of their own data sets. A more mixed reaction was produced when archaeologists were asked about a dataset knowledgebase, with most appearing supportive though some expressing concerns about the management of such a resource, and one respondent warning of the possibilities of 'information overload' for the repository user through the FAQ and ratings facilities.

Similarly, widespread support was expressed during interviews for the notion of linking between related resources, and especially between different papers within output repositories.

When asked about suggestions for further improvements to both source and output repositories, many useful suggestions were forthcoming. These can be summarised as follows:

- Greater presence of books and book chapters in output repositories
- Ability to export search results from output repositories directly into bibliographic software such as Endnote
- Geographical search facilities for both source and output repositories
- Ability to continue refining and sorting search results beyond the initial search
- A reduction in authentication requirements
- Consistent interfaces, so that researchers do not have to keep 're-learning' every few months
- With source repositories, the ability to download data sets and manipulate them yourself, rather than relying on restrictive in-built interfaces
- Reduction of charges, especially for journal articles returned by Google, and geographical data such as maps
- Greater aggregation of online resources, especially an increased presence of 'hubs'
- Links to technical guides for non-experts clickable from both source and output material
- Greater presence of academic blogs, gathered together in usable, searchable formats
- Digitisation of older journal articles, especially those in county-based archaeological journals

3.3 *Source data*

Both the questionnaire and the interviews confirmed that archaeologists tend to produce highly complex data sets, often with many different data formats and sometimes (though rarely) large file sizes. These are sometimes, though not always, linked in the form of a Geographical Information System, which often forms a part of the way archaeologists store, analyse and present their research data. 74.4% of StORE respondents (throughout the whole survey constituency) who produce a GIS as part of their research are archaeologists. Indeed, in the file types described, archaeologists tend to produce more maps, plans, plots and images than other disciplines.

3.4 *Source repositories*

The questionnaire phase highlighted high levels of awareness of source repositories among archaeologists. 64.9% of archaeologists had already either deposited with the Archaeology Data Service, or were 'intending to do so soon'. The interview phase suggested that although many use the ADS, the diversity of interests within the discipline results in a wide variety of other source data being sought, downloaded and used online. Indeed, only 54.7% of those who had submitted data to an online source repository had done so with ADS.

3.5 *Metadata*

High levels of awareness of metadata issues were also indicated by the survey, and many researchers stated that it was a subject of concern for them. During the interviews, many expressed frustration at the difficulties of searching accurately and reliably for resources. This is due either to problems over differences in keyword usage, or by inadequate information about the structure and provenance of data sets. With regard to the former concern, there was a high degree of enthusiasm for the employment of standardised word lists and thesauri. It would appear also that such concerns are beginning to impact on the assignment of metadata within archaeology, with high levels of awareness among depositors of data sets: 66.2% of whom decided on and assigned metadata themselves. The questionnaire showed that the support and advice offered by repository staff was in part responsible for the good levels of awareness among depositors. However, among non-depositors, such an awareness is not quite as high, with some rather vague definitions of metadata preferences emerging from the interviews. It would appear that advocacy by repository staff could be stepped up to provide reassurance about the process of deposition: many expressed concern during interviews that data deposition, and especially the assignment of metadata, is a time consuming and complex process, thus deterring them from doing so.

Some suggested the importance of an accompanying ‘narrative’ or free text explanation to accompany the data.

3.6 *Data access and sharing*

The questionnaire showed that most archaeology respondents are happy to share their data widely, with 64.9% of researchers saying that they either had already, or that they intended to, deposit in the ADS, while a further 13.8% had or intended to deposit with another source data repository. However, the interviews implied that this impression is deceptive, with significant levels of concern being expressed about the notion of providing data for public access. Although around half of the interviewees were generally happy to deposit data which would be made generally available, as long as deposition did not precede publication of an associated research paper, the other half had serious concerns. These can be divided roughly into two camps: those that are happy to share data with colleagues, but *not* to make this available publicly for various reasons, and those that have issues with the idea of data sharing at all. Of the former, some voiced concerns that levels of access control should be maintained, especially with regard to geographical precision within the data, in order to avoid illegal looting of archaeological sites. Others, especially those that work as part of larger teams or abroad, feel they cannot deposit online because they do not solely own the data which has been collected. For a small minority, the notion of sharing data at all is problematic because of the risk that peers may publish results prematurely, or misunderstand the data set and publish undue criticism.

Very few researchers appeared aware of any procedures that could be put in place to control access to their data, besides personal vetting on a case-by-case basis.

3.7 *Output repositories*

The questionnaire results seemed to imply that archaeologists do not make as much use of output repositories as researchers in other disciplines. 23.1% of archaeologists claimed *not* to use output repositories to gain access to published papers, compared with just 8.1% for the survey constituency as a whole. Moreover, 41.5% of archaeologists claimed not to deposit in output repositories, compared with just 20.4% as a whole. Of those archaeologists who *do* use output repositories, it

would seem from the questionnaire that roughly equal use is made of institutional, discipline and publisher repositories.

However, these results are misleading if taken at face value: the interviews suggested that many researchers initially misunderstood the broad definition given to output repositories by the StORe Project, and that in fact greater use is made of them than previously supposed. All researchers interviewed had used output repositories, and many had deposited publications within them in some form. Very few made use of an institutional repository, and almost none had deposited in one, either because they were not aware of their existence, were unsure of deposition arrangements or because of concerns over copyright. A similar discrepancy emerged between questionnaire and interview results with regard to the routes archaeology researchers take to papers held within output repositories. The questionnaire implied that many different routes were taken by researchers, whereas during interviews, almost all claimed to use simply Google or a direct link from their library catalogue.

With regard to search methods, the great majority of archaeologists prefer to use a simple method of searching. Many described the fact that they were happy to browse through several pages of results, and welcomed the distraction of finding other papers of which they were not already aware. When searching for something specific, many researchers turn instead to an advanced search, and often use a discipline-specific repository for this.

All researchers in archaeology expressed enthusiasm for the Open Access movement, many on the grounds that it accorded well with the broader ethic of knowledge-sharing within academia. There was some, though not universal, appreciation of the potential impact of this on publisher revenue.

3.8 *Support*

Most researchers in archaeology were not aware of the support available to them, and relatively few make much use of either online help or the advice of professional repository staff. This lack of engagement with support mechanisms appears to have two strands. Firstly, there are those who 'battle through' any problems they encounter, and operate within a mindset of self-sufficiency. Thus, this category of researcher claims to acquire all they need from repositories. Secondly, there are those who do not seek help because they have a generalised mistrust of the support network available to them, thereby believing that 'there's no point' in asking. This latter group often give up when they encounter problems finding what they need.

When asked about metadata issues, it was clear that those who had asked for help when depositing resources in online had benefited a great deal from doing so, and that they could carry this enhanced awareness into future repository use.

4) Significant Observations from the Questionnaire

4.1 *Selection of survey constituency*

When deciding to whom the StORe questionnaire should be sent within the archaeological research community, the following criteria were considered:

- Constituency had to be national, rather than specific to the White Rose region, as users of the specified source repository (the ADS) would be dispersed throughout the UK.
- The archaeological research community is relatively small, with only 26 UK HE departments, and an average of 27 HEFCE paid staff per department.
- An initial inspection of staff composition using departmental websites suggested that the typical make-up of an archaeology department usually goes well beyond full time staff, with part time staff, post-doctoral researchers, research students, research associates and technicians all contributing in inter-linked ways to archaeological research, and depositing in and downloading from repositories.
- It would be difficult to reach 100% coverage of postgraduate students, as many are not listed on departmental websites.
- Although research groups exist in archaeology, these were not considered a reliable enough method to disseminate the questionnaire.
- ADS users are also to be found beyond HE staff, within the commercial, local government and museum sectors. Preliminary interviews suggested that the research activities of these sectors are often difficult to distinguish from purely University-based research, and that significant overlap can be identified.

Given these criteria, it was decided to disseminate the StORe questionnaire as follows:

- To target all of the staff, researchers, postgraduate students and technicians within all HE archaeology departments within the UK. This provided a total of 721 potential respondents.
- To send the questionnaire to each of these individually, rather than relying on a 'trickle-down' technique via heads of department or research groups.
- To target a subset of archaeological researchers from all sectors within a predefined sub-region of the UK, in order to provide a meaningful comparison with staff based purely within HE. This sub-region was defined as the White Rose area (ie Yorkshire), which provided a total number of 41 potential respondents (5.7% of the overall constituency).

Given the 'total' nature of questionnaire dissemination, there was a risk that the StORe survey might be perceived as generic and thus unimportant. Advocacy was therefore attempted through representations at the Standing Committee for Archaeology (SCFA) meeting, and through follow-up emails to heads of department. A reminder was then sent to every potential respondent two weeks after the initial email was sent.

4.2 Questionnaire response

The questionnaire responses for archaeology, in comparison to each of the other repository communities, can be broken down as follows:

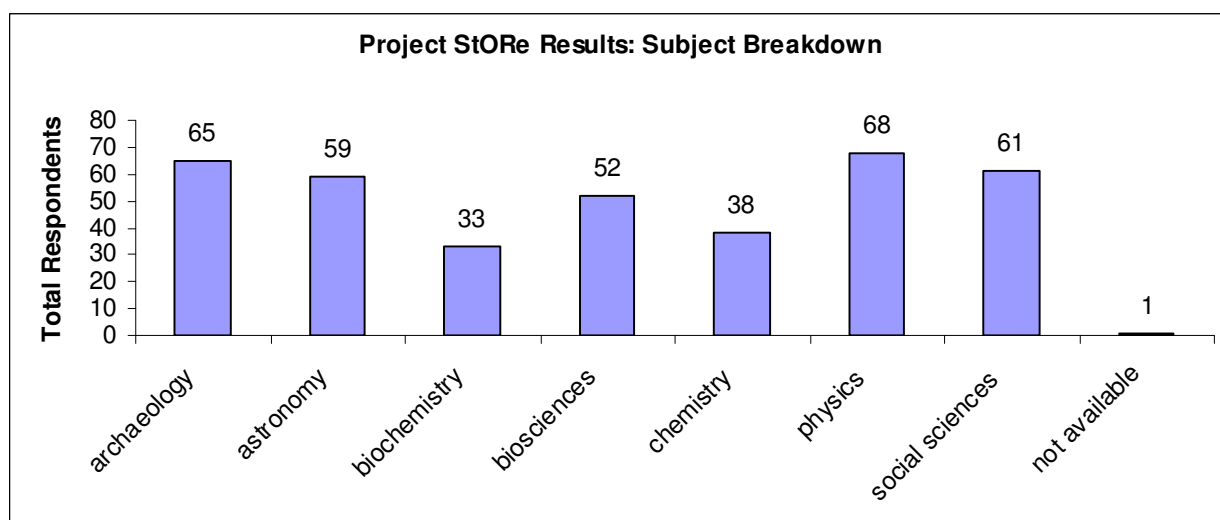


Figure 4: Bar chart showing the numbers of questionnaire respondents for each community

Although this figure may seem relatively high compared with most of the other repository communities surveyed, it represents only 9.0% of the total. This was somewhat disappointing after the comparatively high ‘turn out’ in previous surveys of the archaeology community. However, this could be explained in part by a degree of apathy towards surveys given the high number of them in recent years. Also, given that the questionnaire was sent out towards the end of the spring term and going into the Easter break, it seems likely that a certain proportion of staff would have been away from their email.

It should be noted that in the final stages of the questionnaire analysis, one of the archaeology respondents was transferred to the discipline of social sciences, leaving just 64 respondents for archaeology. However, this change was made too late to change the analysis carried out below.

4.3 Identities

The archaeology respondents can be further broken down as follows:

Total questionnaires sent: 721

Non-HE archaeologists: 41 (5.7%)

HE staff/postgraduates/research associates: 680 (94.3%)

Total responses received: 65 (9.7%)

Non-HE archaeologists: 18 (27.7%)

HE staff/postgraduates/research associates: 47 (72.3%)

More specifically, archaeology respondents identified themselves according to the categories set within the questionnaire as follows (overleaf):

Identifier	Total for archaeology	% Archaeology responses	Total for all subjects	% All respondents
University academic staff	36	55.40%	197	52.30%
University research assistant	8	12.30%	31	8.20%
Postgraduate student	3	4.60%	83	22.00%
Undergraduate student	0	0%	0	0%
Contracting researcher	5	7.70%	18	4.80%
Independent researcher	1	1.50%	7	1.90%
<i>Other</i>	12	18.50%	41	10.90%

Table 1: The identity of questionnaire respondents for archaeology and the total survey constituency

It can be seen from this that because of the way in which the archaeological community was targeted, it had a higher number of contracting researchers, but a lower number of postgraduate students, when compared with the survey constituency as a whole. This may have had a bearing on the responses given, as outlined further below.

4.4 Project aims

In response to question 2, when asked, ‘Source repositories contain primary research data. If a standard feature of such repositories was the ability to identify and link to the publications that had been developed from these data, how advantageous would you find it?’, the results were as follows:

	Archaeology Results		All Results	
Significant advantage to my work	39	60.0%	166	44.0%
Useful but not of major significance	21	32.3%	140	37.1%
Interesting but not particularly useful	2	3.1%	41	10.9%
Of no interest to me	0	0%	10	2.7%
Not sure at this point	3	4.6%	17	4.5%
<i>Other</i>	0	0.0%	3	0.0%

Table 2: Response to Question 2: attitudes towards source to output linkage

In contrast, in response to question 3, when asked, ‘How advantageous to you would it be if it were possible to go directly from within an online publication (electronic journal article or other text) to the primary source data from which that publication was developed?’, the results were as follows:

	Archaeology Results		All Results	
Significant advantage to my work	42	64.6%	175	46.4%
Useful but not of major significance	19	29.2%	151	40.1%
Interesting but not particularly useful	3	4.6%	29	7.7%
Of no interest to me	0	0%	8	2.1%
Not sure at this point	0	0%	5	1.3%
<i>Other</i>	1	1.5%	9	2.4%

Table 3: Response to Question 3: attitudes towards output to source linkage

In general terms, these results imply two things. Firstly, respondents from archaeology seem far more enthusiastic about the issue of source to output repository linkage (in either direction) than the

overall survey constituency. Secondly, within archaeology there is a slight preference towards output to source linkage, rather than source to output linkage.

In response to Question 2 (source to output linkage), the archaeology responses can be further broken down by identity as follows:

	University academic staff	Research assistant	Postgraduate student	Contracting researcher	Independent researcher	Other	Totals
Significant advantage to my work	24	5	1	4	0	5	39
Useful but not of major significance	9	3	2	1	1	5	21
Interesting but not particularly useful	2	0	0	0	0	0	2
Of no interest to me	0	0	0	0	0	0	0
Not sure at this point	2	0	0	0	0	1	3
Totals	37	8	3	5	1	13	67

Table 4: Attitudes to source to output linkage by identity

In response to Question 3 (output to source linkage), the archaeology responses can be further broken down by identity as follows:

	University academic staff	Research assistant	Postgraduate student	Contracting researcher	Independent researcher	Other	Totals
Significant advantage to my work	24	6	2	4	0	6	42
Useful but not of major significance	10	2	1	1	1	4	19
Interesting but not particularly useful	1	0	0	0	0	2	3
Of no interest to me	0	0	0	0	0	0	0
Not sure at this point	0	0	0	0	0	0	0
Other	1	0	0	0	0	0	1
Totals	36	8	3	5	1	12	65

Table 5: Attitudes to output to source linkage by identity

It could be suggested from this that University academic staff make up the vast majority of those in favour of source to output linkage (in both directions), but that interest is also spread throughout research assistants, postgraduates and contractors. It is perhaps unsurprising that those more in favour of output to source linkage are research assistants and postgraduate students, since it is these groups who are more likely to have to interrogate data sets in detail.

4.5 Source data

In terms of the source data produced by those within the archaeology community, in general terms the following data types are produced:

File types	Frequency	No.
Audio:		0
Databases:		54
Derived data:		13
Drawings, Plots:		46
Gene/protein sequences:		0
Geophysical data:		23
Images:		50
Instrument data:		11
Photographs:		53
Plans, Maps:		44
Qualitative questionnaire data:		7
Quantitative questionnaire data:		6
Radiographic data:		3
Raw data:		17
Remote sensing:		11
Spectra:		2
Statistical data:		21
Synthetic data:		5
Telemetry:		0
Text-based files:		40
Topographical data:		27
Video:		3
Other (<i>please specify</i>):		4

Table 6: Source data formats produced by archaeologists

The free text entries were also revealing here, as they identified additional file types as follows:

Analytical (compositional and isotopic) artefact data.
Chromatograms and mass spectra; elemental analysis of ceramic fabrics; stable isotope analysis data.
GIS data [in fact provided as a format type in the following question].

Table 7: Free text entries indicating further source data formats

For the sake of comparison, when the survey results for all disciplines as a whole are observed, the following results are produced:

Files types	Frequency	No.
Audio:		14
Databases:		183
Derived data:		126
Drawings, Plots:		216
Gene/protein sequences:		42
Geophysical data:		27
Images:		195
Instrument data:		91
Photographs:		115
Plans, Maps:		59
Qualitative questionnaire data:		37
Quantitative questionnaire data:		39
Radiographic data:		11
Raw data:		159
Remote sensing:		15
Spectra:		124
Statistical data:		129
Synthetic data:		48
Telemetry:		5
Text-based files:		168
Topographical data:		32
Video:		27
Other (<i>please specify</i>):		32

Table 8: Source data formats produced by the survey constituency as a whole

In general terms then, archaeologists produce a higher proportion of maps, plans, plots and images than the research community as whole. This is in part due to the often geographical and spatial nature of archaeological enquiry, and also to the growing accessibility of Geographical Information System packages. Indeed, of the 39 researchers who professed to creating CAD/GIS electronic data files, 29 (74.4%) are archaeologists.

This propensity to make use of GIS within archaeological research in part explains the complexity of many archaeological research projects, which seem to be more complex than within the research community as a whole. For example, when asked in Question 6, ‘Are the data you generate sometimes a combination of different data formats’, the results were as follows:

	Archaeology Results		All Results	
Often	34	52.3%	143	37.9%
Sometimes	22	33.8%	134	35.5%
Rarely	6	9.2%	57	15.1%
Never	1	1.5%	24	6.4%
Potentially	1	1.5%	11	2.9%
Other	1	1.5%	8	2.1%

Table 9: Number of respondents who produce a combination of data formats

When the file types produced are analysed in relation to the different repository communities within archaeology, the following results are produced:

	ADS	NERC	UKDA	Alexandria Archive	Historic Environment Record	English Heritage/ NMR	Other	TOTALS
Audio:								-
Databases:	17		1	1	1	2	9	31
Derived data:	6	1	1				1	9
Drawings, Plots:	19			1	1	2	5	28
Gene/protein sequences:								-
Geophysical data:	15			1		2	1	19
Images:	20			1		2	5	28
Instrument data:	3						2	5
Photographs:	19			1	2	2	4	28
Plans, Maps:	21			1	2	2	2	28
Qualitative questionnaire data:	3					1	2	6
Quantitative questionnaire data:	2					1	2	5
Radiographic data:								-
Raw data:	7	1					1	9
Remote sensing:	7					2		9
Spectra:								-
Statistical data:	8	1		1		1	1	12
Synthetic data:	2							2
Telemetry:								-
Text-based files:	14	1			1	1	3	20
Topographical data:	15			1	1	2	1	20
Video:				1			1	2
Other (<i>please specify</i>):								-
TOTALS	178	4	2	9	8	20	40	

Table 10: Data formats by source repository for archaeology

If the information on file types produced by archaeologists is then condensed into a bar chart, the relative quantities of each file can be seen:

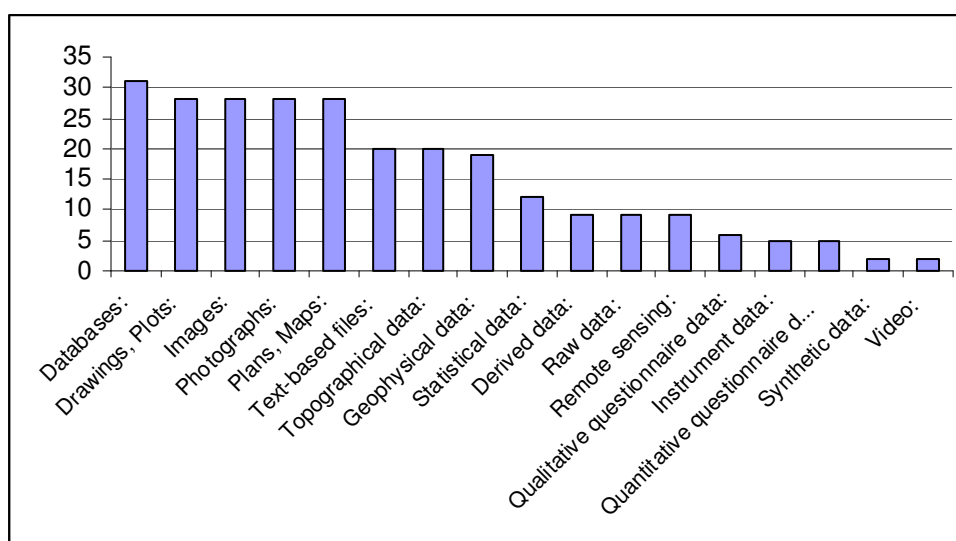


Figure 5: Bar chart showing the file types produced by archaeologists, in descending order

In terms of the specific formats used by archaeologists who submit data to repositories, the following is revealing:

	ADS	NERC	UKDA	Alexandria Archive	Historic Environment Record	English Heritage/ NMR	Other	TOTALS
CAD/GIS:	13			1	1	2	2	19
Extensible mark-up language (XML):	2						3	5
Database files (e.g. Access, MySQL):	14		1	1	1	2	9	28
Flat files (e.g. FITS):								-
Hypertext mark-up language (HTML):	4			1			4	9
Image files (e.g. .jpg, .tif, .bmp, .gif):	16			1	1	2	4	24
Plain text (.txt):	5	1			1		2	9
Portable document format (.pdf):	9			1		1	6	17
Rich text files (.rtf):	7			1		2	3	13
Spreadsheets (e.g. Excel/.xls):	13	1	1	1		1	7	24
Statistical software:	4	1	1	1		1	2	10
Tables/catalogues:	10					1	2	13
Word processed files (e.g. Word/.doc):	17	2		1	1	1	5	27
Other (<i>please specify</i>):						1		1
TOTALS	114	5	3	9	5	14	49	

Table 11: File types produced by repository for archaeology

One of the conclusions which could be drawn from this is that archaeologists frequently place their data in database formats, but that still relatively few use formats which are easily searchable online (such as XML).

If those file formats are condensed into a bar chart format, the relative frequencies of each can be seen:

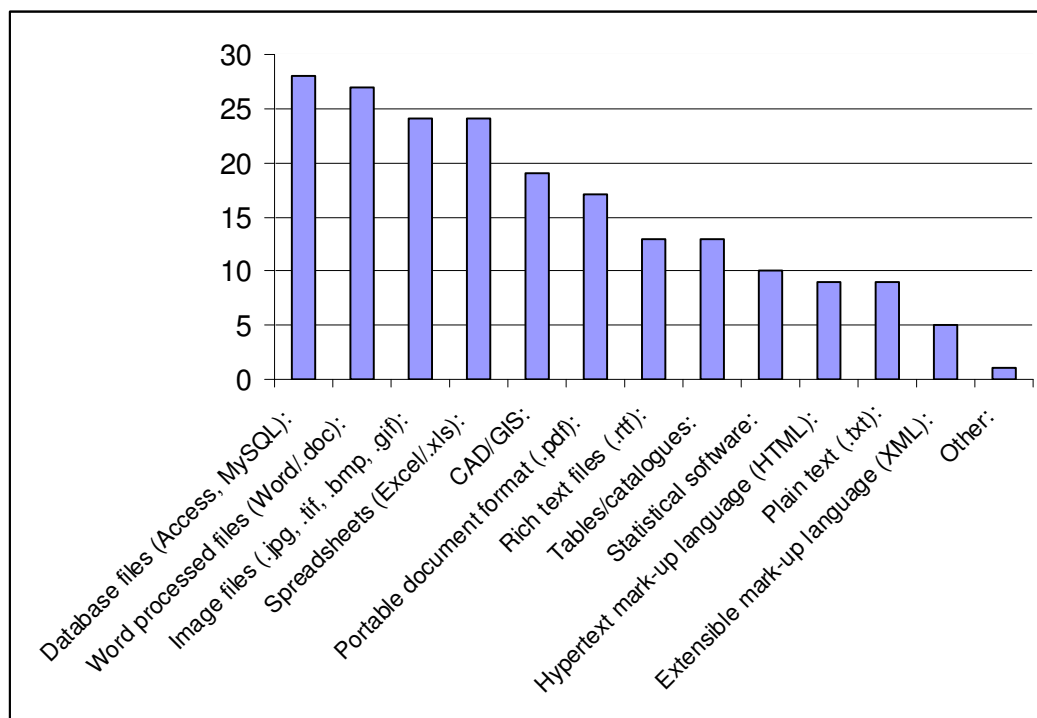


Figure 6: Bar chart showing the file types produced by archaeologists, in descending order

This study of file types may be useful in informing repository managers of the file types and formats generated by the archaeological community. Of the source repositories used by archaeologists, 54.7% of submissions were to the ADS, with a surprisingly high number of other repositories specified by respondents (most of which were not given as a choice in the questionnaire).

It would seem likely that large numbers of archaeologists are willing to submit data to the ADS, but that they have not yet done so. When asked in Question 8, ‘How often do you submit data to the Archaeology Data Service’, only 19.3% had already done so, but a further 45.6% had not yet done so but were ‘intending to do so soon’.

4.6 Metadata

An impression was gained from the questionnaire of high awareness of metadata issues among archaeologists, with 66.2% choosing the option that they themselves ‘decide which metadata terms to use and assign them’.

Of those archaeologists who have submitted data to the Archaeology Data Service, they primarily assign their metadata at the following stages:

Metadata stage	Instances
Prior to data creation	1
As part of the indexing process for source data files	3
During file saving	5
When submitting data to the repository	10
After submission of my data to the repository	-
No metadata are assigned	1
I am not certain of the stage at which the metadata are assigned	1
Other (mostly ‘it depends on the project’)	3

Table 12: Stages at which archaeologists assign metadata

It can be seen from this that most ADS depositors assign their metadata as they prepare it for deposition. This indicates that archaeologists often do not assign metadata until they contact the ADS and receive prompts and advice. After they have acquired this advice and prepared an archive once for digital deposition, archaeologists may then in the future assign their metadata with greater awareness and at an earlier stage in the research process.

In response to Question 9, ‘would you indicate what types of metadata you consider it important to assign to your data’, archaeologists answered as follows:

Metadata Item:		No:
Project title:		57
Project description:		43
Project reference numbers/identifiers:		42
Author/data creator name(s):		58
Title of data set:		43
Subject keywords:		48
Funding source:		19
Publisher:		27
Dates of project:		35
Date (e.g. of data creation):		37
Format (e.g. PDF or HTML):		41
Other (please specify):		11

Table 13: Metadata types which archaeologists would like assigned to their data sets

When ranked according to preference, the results can be displayed in a bar chart as follows:

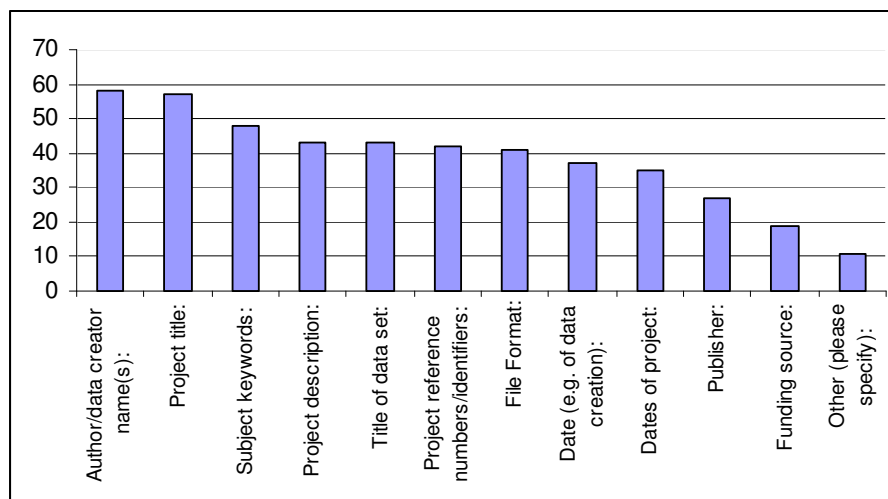


Figure 7: Bar chart showing the relative importance of different metadata types for archaeologists, in descending order

From this, it could be suggested that though there is fairly general recognition of the importance of a whole range of metadata, preference expressed by data creators suggests a need for general information such as the author and project title, slightly less so for more specific details such as subject keywords and a project description, but rather less demand for publisher and funding source details. Further suggestions deriving from free text entries include the following:

Archaeological period, artefact type, artefact material, conservation method
Description of GIS processes applied, min/max co-ordinates, cell resolution for raster data, etc
Geographic location, period range
Material examined, chemical class, analytical technique
Object description, primary material, findspot (very important)
Three dates are useful: date of project, date of creation of digital resource and date of creation of original data. The latter is generally only appropriate for retrospectively created data, but may be of value to many in the humanities

Table 14: Free text additions to a question on metadata types

It is revealing to examine the attitude of researchers towards metadata assignment in comparison to the levels of support offered by repository managers. A cross-tabulation of Question 10, ‘At what stage are metadata assigned to your data?’, with Question 25, ‘What assistance in your use of repositories is provided by a librarian or other knowledge management support?’. The results are as follows:

Levels of support:	Metadata assignment:	Prior to data creation	As part of the indexing process for source data files	During file saving	When submitting data to the repository	After submission of my data to the repository	No metadata are assigned	I am not certain of the stage at which metadata are assigned	Other	Totals
Provision of documentation (guidance notes, fact sheets, etc.)	4	11	8	10	2	2	5	2	44	
Formal training and documentation	3	6	6	5	2	2	0	1	25	
Online or telephone help	3	8	5	8	4	2	3	2	35	
Assistance with the structuring of specific searches	1	1	2	3	0	0	1	2	10	
Assistance with the conduct of searches	2	1	2	3	0	0	1	3	12	
Full intermediary service (e.g. the conduct of searches and organisation of results)	1	1	1	0	0	0	0	0	3	
Unknown	15	11	16	8	1	3	5	3	62	
Other	0	0	1	0	0	0	0	1	2	
Totals	29	39	41	37	9	9	15	14	-	

Table 15: Levels of repository support by stage at which metadata is assigned

The most striking conclusion to be drawn from this analysis is that the majority of respondents do not know what levels of support are available to them when dealing with repositories. Another conclusion is that those who are assigning metadata either prior to or during submission of their data to a repository are also those who are aware of the levels of support available to them, and have easy access to such materials as guidance notes and fact sheets. This could be further extrapolated to imply that the greater the levels of support provided by repository managers, the more integrated the process of metadata assignment becomes. This conclusion is further supported by a cross-tabulation of Question 11, ‘Who assigns metadata to your research data’, with Question 25, ‘What assistance in your use of repositories is provided by a librarian or other knowledge management support?’. The results are as follows:

Level of support:	Meta data personnel:	I decide which terms to use and I assign them	Research colleague(s) assign metadata on the team's behalf	Research support staff assign metadata on the team's behalf	Research support staff assign metadata on the team's behalf	Metadata are assigned by library/information services staff	Metadata are assigned by the repository administrators	Metadata are generated automatically	It is not known who assigns metadata	Other	Totals
Provision of documentation (guidance notes, fact sheets, etc.)		18	2	2	1	5	2	3	2	35	
Formal training and documentation		11	2	1	0	4	1	0	2	21	
Online or telephone help		14	2	2	0	6	1	1	2	28	
Assistance with the structuring of specific searches		4	1	0	0	0	0	0	0	5	
Assistance with the conduct of searches		5	1	0	1	0	0	0	0	7	
Full intermediary service (e.g. the conduct of searches and organisation of results)		1	0	0	0	0	0	0	0	1	
Unknown		21	7	6	0	1	4	4	6	49	
Other		0	1	0	0	0	0	0	1	2	
Totals		74	16	11	2	16	8	8	13		

Table 16: Levels of repository support, by personnel who enter the metadata

This would suggest that those who assign their own metadata are also those who are aware of the support that is available to them. There is something of a correlation therefore between metadata awareness and the support offered by repository managers.

However, it is also clear that relatively few repository users, when seeking advice and help, receive guidance from an individual. When asked in Question 24, 'Do you receive support and/or guidance in your use of output repositories?', only 10.8% declared that they had received personal support:







Documentary support:		7.7%	5
Personal support provided by an intermediary:		10.8%	7
Repository-enabled support:		21.5%	14
No support is provided:		24.6%	16
Unknown:		26.2%	17
Other (please specify):		9.2%	6

Table 17: Forms of support/guidance received by repository users

As can be seen, the majority stated either that no support was provided, or that they did not know what support was available.

4.7 *Output repositories*

In general, archaeological researchers do not seem to make as much use of output repositories as other disciplines. This can be illustrated by the following figures:

Those who do *not* currently **use** output repositories to find information for use in their research:

Archaeology Respondents	All Respondents
23.1%	8.2%

Table 18: Non-users of output repositories

Those who do *not* currently **deposit** their research publications in an output repository:

Archaeology Respondents	All Respondents
41.5%	20.4%

Table 19: Non-depositors in output repositories

This would seem to indicate a marked under-use of output repositories by archaeologists. This impression of under-use is further reinforced by some of the ‘output’ repositories specified in free text boxes. When asked in Question 19, ‘Which kind of output repositories do you use to find and retrieve information for use in your research?’, a sample of the responses is as follows:

Archaeology Data Service.
Don't recognise any in your list.
No comment.
Not really clear about what this is asking, and how it relates to the few works which I have consulted electronically.
Portable Antiquities Scheme.

Table 20: Free text responses indicating output repositories used

The Portable Antiquities Scheme repository is not an output repository, as it provides primary information about archaeological objects and does not hold publications. Although the Archaeological Data Service provides access to both data and publication (in addition to ‘grey’ literature which may be considered as both source data and output), it is unclear whether the respondents who specified it in this context did so intentionally, or because they misunderstood the question. The other responses given here indicate a degree of confusion about what constitutes an output repository.

Of those archaeologists that *do* use output repositories, their preferences were as follows:






Institutional:		47.7%	31
Discipline:		49.2%	32
Publisher:		40.0%	26
None:		23.1%	15
Other:		10.8%	7

Table 21: Repository types used by archaeologists

When using output repositories, respondents indicated in Question 23 ‘What level of searching do you normally find sufficient when using an output repository?’.







Simple - author, title, keyword, date:		53.8%	35
Advanced, using a range of fields and identifiers:		21.5%	14
Employing Boolean logic:		1.5%	1
Using a subject thesaurus or subject headings:		1.5%	1
No preference:		13.8%	9
Other (please specify):		7.7%	5

Table 22: Search levels employed by archaeologists

It can be seen then that the vast majority of output repository users within archaeology prefer to use a simple method of searching. One aspect which emerged often during the interview phase (discussed below), but which is also reflected in a free text comment, is that users may often adapt their search method according to (a) the reason why they are carrying out a search in the first place and (b) the level of success encountered in the first search. For example, some of the free text responses commented that:

A combination of simple, and if this doesn't get the answer advanced searches.
Advanced, using a range of fields and identifiers: This should also include some level of geographical specificity.
It depends on what I'm looking for and the options available. I use the minimum that I can.
It very much depends what I'm looking for. If I want to identify a specific mass spectrum it requires a much more detailed search than looking for a paper on a general topic.

Table 23: Free text responses commenting on search levels used

If the search method is cross-tabulated with the type of output repository being used, some indication of an adaptation of search method according to the circumstances is given:

	Institutional	Discipline	Publisher	None	Other	Totals
Simple	22	16	18	4	4	64
Advanced	8	12	5	2	0	27
Boolean logic	1	1	1	0	0	3
Using a subject thesaurus or subject headings	0	0	0	1	0	1
No preference	0	2	1	7	1	11
Other	0	1	1	1	2	5
Totals	31	32	26	15	7	

Table 24: Search levels used by output repository type

It could be implied from these results that when researchers are using a discipline-specific output repository, they are more likely to use an advanced search, since presumably they are looking for something quite specific. While using institutional or publisher repositories, users seem more likely to perform a keyword or author search, in order to make use of a broader, more generalised set of returns.

In response to Question 23a, 'What further options, features or functionality would enhance your level of searching?', very few responses were received. This aspect of the survey was served rather better during the interview. However, it is useful to note that one respondent commented that they would like to have the 'ability to generate a saved set of links to favourite publications/journals for ready access, rather than having to go through a long-winded hierarchy of options'.

4.8 Preferred routes to output repositories

Archaeologists stated that the routes they would normally or prefer to take into a publication held within an output repository are as follows:












Via a known repository's URL:		27
Via an Open URL resolver:		4
Via a library catalogue that links directly to an article in a repository:		34
Via a library subject page:		13
Through a publisher's online service (e.g. ScienceDirect):		18
Directly through a specific journal's own web site:		23
Through an author's personal web page:		16
From a link provided in an e-mail, CD-rom, USB drive etc:		15
From an Internet search engine (e.g. Google):		40
Through a subject portal service (e.g. Entrez):		6
I have no normal or preferred routes:		18

Table 25: Preferred routes into output repositories

As can be seen, the majority of output repository users reach a publication via a search engine, or via a link from a library catalogue. However, what is also demonstrated by this analysis is that a whole range of routes are taken by users, with a surprisingly wide spread across the choices given in the questionnaire.

If this analysis is refined somewhat to reveal only those archaeologists who actively deposit in the ADS, it reveals the following results:

Via a known repository's URL:	10
Via an Open URL resolver:	1
Via a library catalogue that links directly to an article in a repository:	16
Via a library subject page:	6
Through a publisher's online service (e.g. ScienceDirect):	5
Directly through a specific journal's own web site:	7
Through an author's personal web page:	7
From a link provided in an e-mail, CD-rom, USB drive etc:	6
From an Internet search engine (e.g. Google):	17
Through a subject portal service (e.g. Entrez):	3
I have no normal or preferred routes:	4

Table 26: Routes into output repositories pursued by ADS depositors

If the two sets of results, for archaeologists who have not yet deposited source data with the ADS and for those that have, are compared directly, the results are revealing. In order to make the results more directly comparable, the ADS depositors have been increased proportionally to represent the same overall numbers as non-ADS:

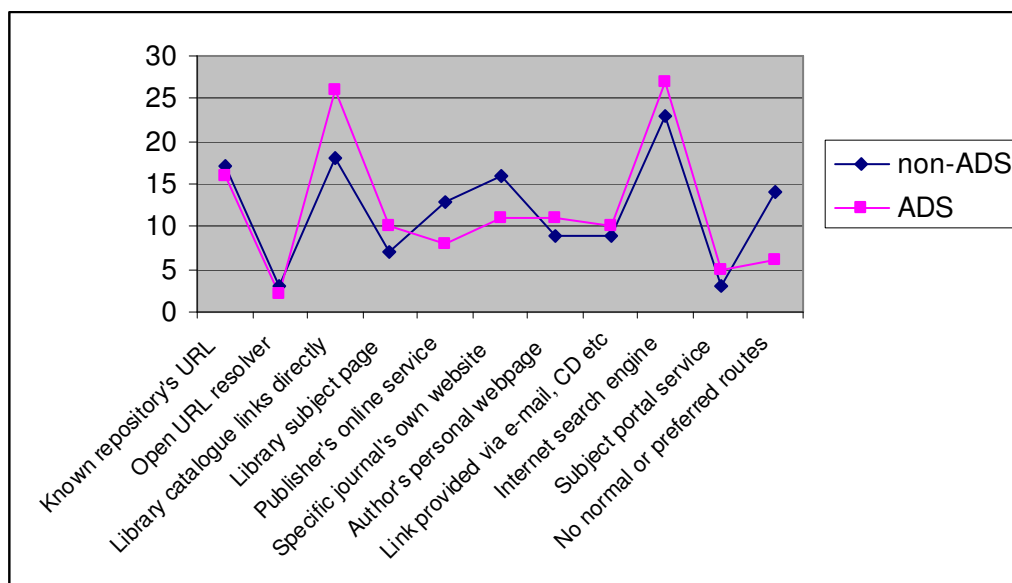


Table 27: Line graph showing routes into output repositories pursued by ADS depositors, versus non-depositors

It can be seen from this that those who have already deposited source material in the ADS are more likely to reach output material by linking directly from a library catalogue. They are also slightly more likely to link directly from an email, and to use a search engine. Those who are not ADS depositors are more likely to go directly to a publisher's or a journals website, or expressed no preference regarding a 'normal' route.

Another interesting observation to emerge from the questionnaire results is that those who have previously deposited in a source repository are also more likely than the average archaeologist to deposit in an output repository. The results for those who have *not* deposited a publication in an output repository are as follows:

All Archaeologists	Archaeologists who have deposited source data
41.5%	34.6%

Table 28: Archaeologists who have not deposited in an output repository

The percentage of archaeologists who deposit in source repositories can be seen to be lower, perhaps indicating that a willingness – and a knowledge of how – to deposit source data goes hand-in-hand, to an extent, with output publications.

4.9 Information support

Finally, in order to analyse the levels of support available, Question 23 asked, 'Do you receive support and/or guidance in your use of output repositories?'. The answers given by the archaeological community were as follows:







Documentary support:		7.7%	5
Personal support provided by an intermediary:		10.8%	7
Repository-enabled support:		21.5%	14
No support is provided:		24.6%	16
Unknown:		26.2%	17
Other (please specify):		9.2%	6

Table 29: Types of information support received by archaeologists

Of the 'others' specified, the free text comments are revealing, as follows:

Coaching by subject librarian.
Depends on the repository, and how much I need the information.
I haven't yet tried to output any of my publications myself.
My institution provides some support in how to search output repositories but it is quite general.
Not sure I understand. I rely on my own skills and occasionally phone / email for assistance.

Table 30: Free text responses indicating types of information support

It could be concluded from this that researchers in archaeology have a mixed experience of the levels of support available to them. The majority (50.8%) state either that no help is provided, or that they are unsure what support is available to them.

5) Interview Analysis

5.1 *Selection method*

It was decided that in order to allow sufficient time for analysis and report writing, but to provide the maximum amount of qualitative and contextual information possible, around 10 interviews would be conducted. In fact, it became clear once the interview phase had commenced that rather more would be possible, so a total of 16 were conducted in all. Two of these interviews were in fact carried out in January/February 2006 during a preliminary interview phase (specific to archaeology), which was intended to help inform and shape the issue covered in the questionnaire. Since most of the issues covered by these preliminary interviews were in fact very similar to those asked during the June 2006 phase, the results are included here.

Initially, the intention was to select interviewees on the basis of those who volunteered as part of their questionnaire response. This produced a total of 16 candidates. However, on examining the expertise and institutional background of each of these, it became clear that this sample would not be adequately representative of the archaeological community as a whole. These 16 contained a high proportion of researchers with a keen interest in digital repositories or in computer applications as a whole, and thus were likely to give answers which may potentially provide a biased representation of the way researchers regard, use and anticipate repositories in archaeology.

As a result of this realisation, only 11 of the 16 volunteers were selected for interview. It was decided that approaching some researchers who either filled in a questionnaire but did not volunteer, or who in fact had not engaged with the StORe Project at all, would be interesting and beneficial to the project as a whole.

It was also realised that the original survey strategy for the archaeology component – to questionnaire nationally but interview only in the White Rose region – would not be possible. Only two researchers from the Universities of York, Leeds or Sheffield volunteered for interview, plus one individual working within the commercial sector. Three archaeologists as a whole would have provided very limited information, whereas many of those who volunteered were from institutions elsewhere in the UK. Consequently, it was decided to employ a national (rather than a regional) remit for the interview phase also. However, in order to retain *something* of a White Rose focus, the five non-volunteer interviewees were selected from the University of York.

A further criterion employed during the selection of interviewees was to ensure that a broad range of researchers was selected, in order to gain a genuinely broad impression of researcher behaviour and attitudes within archaeology. The following categories were selected for interview:

- Post-doctoral fellow
- Lecturer
- Senior lecturer
- Postgraduate researcher
- Local government historic records officer
- Portable Antiquities Scheme Finds liaison officer
- Commercial contractor

5.2 *Interview technique*

As already stated, a preliminary interview phase was carried out in January/February. This contained a comparatively brief set of questions, but which covered most of the main issues addressed in the later interviews. These questions were used for interviewees 001 and 002. During the June 2006

phase, a set of questions was formed in the basis of guidance provided by Project Manager Graham Pryor, in combination with issues arising from the questionnaire phase. These questions were used for interviewees 003 to 007. However, it was soon realised that the order of the questions was non-intuitive, and presented problems when interviewees began providing information in detail which was not required until later in the interview. This made the transcribing of the interviewees a complex task. A third question set was then devised which, although addressed exactly the same issues as before, did so in an order which was more intuitive and provided a clearer structure to the answers given. This new question set was used for interviewees 008 to 016.

A copy of each of the two question sets is provided in Appendix 1.

A combination of telephone and face-to-face interviews was carried out. It was decided that face-to-face interviews represent the best information collection strategy as a range of sensory information beyond verbal alone can be gathered. However, many interviewees expressed a preference for telephone interviews as they can be more flexibly arranged and take rather less time than a personal visit.

In all cases, having acquired the consent of the interviewee, the interviews were recorded. In one case (Interviewee 012), the recording was not retained for technical reasons.

A full list of interviews is as follows:

No.	Date	Origin	Role	Type	Duration
001	23.1.06	University of York	Senior Lecturer	Face-to-face	25:22
002	07.2.06	University of York	Senior Lecturer	Face-to-face	37:03
003	09.5.06	University of Bournemouth	Lecturer	Telephone	35:00
004	09.5.06	University of York	Lecturer	Face-to-face	48:00
005	15.5.06	Portable Antiquities Scheme	Finds Liaison Officer	Telephone	55:00
006	18.5.06	University of York	Post-doctoral fellow	Face-to-face	25:00
007	18.5.06	University of York	Lecturer	Face-to-face	28:00
008	24.5.06	University of Bradford	Researcher	Telephone	29:02
009	25.5.06	North Yorkshire Heritage Unit	Historic Record Officer	Telephone	35:09
010	25.5.06	University of Nottingham	Lecturer	Telephone	23:29
011	26.5.06	University of Bradford	Senior Lecturer	Telephone	44:10
012	01.6.06	Glasgow Museum	Researcher	Telephone	38:00
013	06.6.06	University of Sheffield	Post-doctoral fellow	Face-to-face	44:22
014	06.6.06	University of Sheffield	Researcher	Face-to-face	22:44
015	08.6.06	University of Birmingham	Lecturer	Telephone	43:51
016	08.6.06	Malton Archaeological Practice Ltd.	Commercial contractor	Telephone	16:30

Table 31: List of interviewees, indicating their role, date interviewed and the method used

What follows is an exposition of the main issues arising from the interviews.

5.3 *Identities*

The most obvious aspect to emerge from the interviews was the sheer breadth of archaeological specialisms within the research community. This breadth can be very roughly characterised by a spectrum, from a humanities focus at one end with an interest in historical and theoretical issues, and a science focus at the other with a particular emphasis on the use of techniques deriving from biology, chemistry and physics to understand the environmental context within which humans have

dwelt in the past. A third category can be added to the range of respondents interviewed, consisting of those with an interest in what may broadly be termed ‘heritage management’. These may be involved either in the ways in which physical heritage is perceived and dealt with in the present, or more specifically in the logistics of how archaeological remains are investigated and the results stored and disseminated. Of those interviewed, two (Int 009 and 012) expressed a particular interest in data management:

‘I’m the Historic Records Officer, which means I’m responsible for managing the collections here of historic environment data. This includes archaeology, buildings, battlefields and other things.’ (Int 009)

‘More recently [I have specialised in] information management, repository usage, project management and, more broadly, heritage management.’ (Int 012)

These three broad categories of archaeological research have clear ramifications for some of the more specific issues addressed in the interviews, and will be returned to below.

What is less clear from the identities expressed by researchers is the extent to which their specific status had an affect on the way they engaged with repositories. For example, it is not necessarily the case that a senior lecturer would be less inclined to use fast-developing digital resources than a younger, postgraduate researcher. For example, Interviewee 011 (a senior lecturer) declared an awareness of a wide range of both source and output repositories, and was also developing methods for metadata gathering at the point of recording. Interviewee 008 (a postgraduate researcher), on the other hand, declared that:

‘I don’t think the ADS would have anything relevant, and I didn’t know that NERC had such a system [a source repository] up and running.’ (Int 008)

This issue will also be returned to below.

Finally, it was found during interview conversations that within the archaeological research community, a large degree of overlap exists between commercial, local government and the academic sectors. Each use each others’ data and publications in a complex manner, and many individuals play a role – or have played a role in the past – in multiple sectors. This sense of overlap is in part reflected in the usership of the ADS. It also has ramifications for the development of repositories in the future because, if HE institutions are to play a significant role in this, research will not be properly served unless some of the needs of other sectors is taken into account also.

5.4 *Project aims: Source to Output links*

In terms of the specific StORe project aims, of addressing ‘the area of interactions between output repositories of research publications and source repositories of primary research data’ (Pryor 2006), this produced some interesting answers. Most researchers were very enthusiastic about such links. When asked, ‘if a standard feature was the ability to identify and link directly to the publications developed from source data, how advantageous would you find it?’, interviewees expressed views such as:

‘I think that sounds fantastic.’ (Int 008)

‘Yes I would, but it often doesn’t happen yet.’ (Int 001)

‘There are quite a number of standalone website papers out there at the moment that do exactly that, and yes, that is very useful to be able to take advantage of the power of the computer in a non-linear way.’ (Int 003)

‘It would have an enormous impact, yes.’ (Int 004)

‘It certainly would be an advantage to my research...When something is electronically done and you can cross-reference, it’s a wonderful resource and a joy to use. It’s just that this is not always the case just yet, and it’s very hit and miss.’ (Int 005)

‘I think it would be incredibly valuable, both in a data sense in terms of getting a sense of what is available, but also in a literature sense because you’d [be able to] have a lot of things there in front of you...’ (Int 006)

‘I think it probably would, as I suspect it would be a way of discovering publications of which I wasn’t otherwise aware – he says speaking from a position of complete ignorance! But it strikes me that this might be a useful thing.’ (Int 015)

‘Too damn right it would, yes!’ (Int 014)

5.5 *Project aims: Output to Source links*

In spite of the apparent enthusiasm voiced above, interviewees seemed to express a slight preference for links in the other direction (in concurrence with the questionnaire results, see above). When asked, ‘if you navigate directly from within an online article or text to the source data from which it was derived, how advantageous would you find it?’, the following answers were given:

‘I think that anything which makes getting access to primary or even ‘boiled down’ data easier is a good thing. It’s currently quite difficult to publish large amounts of data – Internet Archaeology is about the only way at the moment. It would be admirable to be able to link up papers and large data sets more easily. It is interesting to be able to publish in a different way, where you can enable links back to primary data without having to worry about providing too much detail in the text. This makes articles more readable.’ (Int 002)

‘It would be fantastic if you could do that, yes. It is always a problem...that in all the papers we are reading you never get the raw data, so you can never check the analysis or use their data to compare with your own. I think what I actually end up doing is not looking at other people’s data because it’s so rarely accessible.’ (Int 007)

‘I know it’s the case that if you read journals like I think it’s *Nature* or *Science* that they have the article itself which is then supported by online documentation which provides greater detail and more data relating to the specific points made in the article. I guess that’s largely related to issues of space and storage, and the small number of words which are allotted to articles in those kind of journals. I find that very useful if you’re interested in the detail of a specific argument, and being about to track and go back to the primary data itself and examine that in more detail is very useful. If that’s available online, you can do it virtually instantaneously.’ (Int 010)

‘...frequently when you’re examining a publication, you don’t have that finer detail that we require for when we’re going into the field. The levels of [archaeological] deposits, the detail of what you’re encountering at whatever level. That would great if you could go back to the archive and have a look.’ (Int 016)

The reasons for this preference centre on the potential for improved speed within the research process:

‘Specifically, it’s the ease of the links that is attractive, and the ability to interrogate the data to see why they are saying what they’re saying.’ (Int 004)

‘It’s partly the speed of the research, because you would hopefully have a good range of source material in front of you, literally at the click of a button...’ (Int 006)

‘I think it’s the convenience, but it would also be nice if with a lot of different projects and papers that I’ve read if you actually had access to the data. I know that you can always get in touch with the person who’s done the work, but quite often it’s still not that easy to get to the data itself because it’s only really now that people are beginning to think about how to archive data properly.’ (Int 007)

As well as the apparent convenience in an immediate sense, these expressions of enthusiasm for easy, electronic ways of linking from output to source repositories probably link to a broader theoretical movement in archaeology which has been developing since the late 1980s. This movement, termed by some ‘post-processualism’, advocates that any set of material remains from the past can be interpreted in a range of potentially ways. Furthermore, it has also encouraged archaeologists to deconstruct the work of others in order to understand how the process and broader circumstances of their work has led to their eventual conclusions. This dual emphasis, on multivocality and epistemology, has arguably contributed to an enthusiasm among researchers in archaeology to reach beyond the published text and into (at least a version of) the ‘original’ data set.

Having said this, some respondents were either less obviously enthusiastic about the possibilities of source to output linkage (in either direction), or had some genuine concerns. For example, one respondent commented that:

‘I can see that it might be [advantageous]. The sort of work I mostly do doesn’t usually require me to look at tables of small artefacts found in excavations or anything like that, but I could see that it would be useful sometimes.’ (Int 015)

Others, though admittedly a minority, had more specific concerns about the management of such facilities. For example,

‘I would doubt that every data set could confidently be provided with comprehensive list of every publication associated with it...It would be a real ‘own goal’ if JISC were to make such links available, but then not be able to manage them properly.’ (Int 012)

‘...I imagine from a development point of view it could be quite complex and expensive and time consuming, and especially in order to maintain it I imagine it’s an enormous task – to get someone, or a team of someones, to maintain all of those two-way links between every data archive and every publication.’

For one interviewee, it was not management that was an issue, but that such links might actually be a distraction to the conduct of their research:

‘I’ve found it often quite good that you have direct access to the data, and you can try to interrogate the data yourself and see if you would come to the same conclusion. In a way, that is purely curiosity driven, because it’s not such that I want to actually re-analyse the data. If somebody has done this then most of the time I’m happy enough that it’s been done and I wouldn’t bother doing it again. So in a way the incentive is not a necessity but a nice feature. Maybe sometimes I should not waste my time doing all these nice things, rather I should do something else. In a way, it could be more destructive than useful and therefore I think it’s a good question to ask. I would say that linking back to the original data – the only *real* reason I can think of, other than re-processing which I don’t think is a good idea [in geophysics], is to check whether somebody has done it right. And there, one should probably simply rely on the peer-review process, whereby if it has been published then we have to accept that the people have used the right methods in order to process the data. So maybe, although it may sound like a really good idea, in terms of really doing it – I would say on reflection it may not be as necessary as I originally thought it would be useful.’ (Int 011)

5.6 *Project aims: additional features*

With regard to these concerns, some were expressed about the difficulties of achieving the breadth of new developments mooted by the StORe Project. For example, when asked about the desirability of a ‘dataset knowledgebase’, many responded with enthusiasm that was either generalised, such as ‘absolutely, as links are often hit-and-miss or difficult to find’ (Int 001), or somewhat non-committal as in, for example:

‘If by doing this they can really guarantee that links won’t go down, in principle this would be a good thing.’ (Int 002)

Others were keen on the potential of a dataset knowledgebase to reduce the number of broken links, but more sceptical about other functions:

‘Yes, that would be useful, because that would mean that keeping a link live isn’t reliant on just one person – you could change your ISP or whatever and it wouldn’t affect such links. FAQs and ratings: perhaps less so, though it would mean that you could see which journals are highly recommended and which not, as well as which ones are referred to more often.’ (Int 008)

In common with previous concerns about source to output linkage, some are keen in principle, but foresee management issues on the horizon:

‘Links are of variable reliability. If a ‘dataset knowledgebase’ improves the reliability and the maintenance of these different electronic links between different data sources, then I guess it would be useful, depending on how much time and energy would be required to develop and maintain it.’ (Int 010)

Others were less obviously enthusiastic in their response to the description of a ‘dataset knowledgebase’. For example:

‘I’m sure it would have its uses.’ (Int 006)

'Is the dataset knowledgebase basically an index? I can see this being useful, but I can also see that with such a service, the links would probably break more easily.' (Int 011)

'The idea of a dataset knowledgebase sounds useful enough, as it can certainly be frustrating when links go down. However, it depends on how well such a resolver is designed and managed, and it really depends on who would be paying for such a resource to be maintained in the long term. The ideas of FAQs good also, although I'm not sure that a resolver, or Middleware, is the best place to locate such a service. That, and ratings by users, would be better maintained by someone else, but the question is, 'who?' Perhaps the British Library or someone similar are the best people to maintain a service like that.' (Int 012)

'With electronic resources, there can often be a problem with information overload. If there's too much information, in terms of FAQs or other information to read before you can actually access data, often makes what should be a rather simple task a much longer process.' (Int 005)

These responses, like a number of others, may in part be due to the uncertainty with which the facility was described to the interviewees in the first place. As an interviewer, I felt unable to describe sufficient technical details to inform researchers how such a facility was to be put in place, and how it would be managed, which thereby had a likely impact on their enthusiasm for such a service.

By contrast, the more straightforward concept of linkage between related resources was received with enthusiasm. In particular, the idea of being able to link directly between a publication and other articles cited in that publication would be of benefit to all of the researchers interviewed.

5.7 *Further suggestions*

Many researchers provided useful ideas about how repository functionality could be enhanced, either when asked directly or elsewhere in the interview. Some suggestions focused on the notion of aggregation:

'I think a central area where you go to get directions to repositories for materials that you might want would be good...If there was a central resource, a hub or whatever, where you could click on 'medieval village' or 'artefacts' or 'bones' or whatever, and then you could be given a few examples of what you might expect to find somewhere, then that would be good.' (004)

'It sounds silly, but I've always found it quite useful if I'm reading something with which I'm not very familiar, to have more general summaries on particular topics available. So, for example, if I'm reading something on Bio-molecular Archaeology, there would be a link offering more information on that topic. If you've got a quick reference there of basic terminology, the same thing that you often get in kids' books, that would be enormously helpful. It sounds very simple, but the few times I have encountered such a service it has been immensely helpful, just because I thought I understood a term and I thought it meant something else and in fact it was being used much more specifically. I think that that would enhance inter-disciplinary research as well.' (Int 005)

'I think it [a dataset knowledgebase] *could* be a useful tool if there was any way of going into and saying, I've read this article, is there anything else which is related.' (Int 011)

Others had specific issues about the content of repositories:

'I find the most frustrating aspect of many data repositories is their lack of broad coverage of different kinds of primary publications. For example, searching bibliographic databases will often give you a great return on journals, but nothing at all that's published in books or edited volumes and conference proceedings. They're just missed out entirely in that electronic literature so I find that a major drawback.' (Int 010)

'I think blogging is becoming more and more useful – there are some out there now that are not just people's boring diaries, but actually, why I've gone in this direction, I've seen this, why I've done that. I don't whether that can be made to work?' (014)

The idea of links between related resources was especially well received, partly because of its ability to replicate, in a rapid and convenient fashion, an aspect of the research process already followed by many:

'I can certainly see the value of being able to do that [link between related resources]. The normal way of reading something and finding an interesting reference contained within it and then having to hunt that separately – being able to click straight through to it would be really quick and easy and useful. Apart from anything else you don't waste time, if it turns out to be as useful as it looked as if it might be, then you just click away again, rather than wasting an afternoon trying to hunt it down then finding that it was a waste of time after all.' (Int 015)

The notion of presenting publications and their source data in adjacent windows was less well received, and a number of respondents commented that it was not essential for them:

'I can see its potential usefulness – whether I'd use it I'm not sure. There's the headache factor to be taken into account. I can see in theory that it would be quite a nice thing to be able to do...Where one is reading a publication where it really matters that you actually have the dataset which is being referred to in front of you, then yes I can see that that [the facility of displaying two datasets at once] would be extremely important. Where you can read a publication without having the dataset in front of you, but knowing that you can go and refer to it, I probably wouldn't click through and read it as it's not immediately necessary.' (Int 015)

'I'm not quite sure how opening up source and output together would be any more advantageous than simply opening up a window. In the electronic world, you're not dealing with a static frame, so I don't know how having data open up on the same screen as being essential. The ability to have a link and to actually do that laterally: to see the data then to get back to where you were – that's very valuable.' (Int 003)

Finally, the perception of there being a need amongst researchers to identify published (and pre-published) papers that have made use of their source data was very well received by interviewees. One commented, for example, that:

'I can imagine people with bigger egos thinking, 'Oh, I wonder how many people have cited me today'! But I can also see it being useful to see how then, if you're going to do

another follow-up publication on the data and you want to have an easy way of seeing who had written on your data and said either it's good or bad or useful for their research or not, then to construct an argument in your next research output publication, that would be useful to help construct a bibliography of comparative material yes.' (Int 013)

Another saw academic potential in being able to track and understand *why* other researchers were searching their source archive, and how they were doing this:

'That would be brilliant. I think there is already a facility out there that enables you to identify what people are searching on in your database. So if, for example, over a four month period you could throw up whether people were looking at the flint archive or the environmental archive or whatever. If you could maybe pick out particular keywords that they're looking at, I would see that as a possibility for generating an article stating that others have been looking at this or that, and developing arguments around *why* they've been looking at particular keywords.' (014)

5.8 *Source data*

The evidence from the interview phase about the types and formats of files produced by researchers was very much in agreement with the questionnaire results. One issue which it would be worth emphasising here is that many researchers appeared not necessarily have thought about the individual file types that they produce during a project – they tend to think instead in terms of the tasks carried out throughout the project as a whole. It was often only when prompted that they considered what file types they had generated. This reveals the extent to which project archives are often complex combinations of different file types, generated at different stages of the project. One interviewee responded, for instance, that 'the files deriving from my research are far too many and too complex!' (Int 010). A common occurrence within archaeological projects is for many different file types and data collection exercises to be combined within a single Geographical Information System (using programmes, for example, like ArcGIS or MapInfo). These often contain topographical data, satellite and map data tiles, distribution maps deriving from database records and georeferenced images deriving from geophysical prospection. It is clear that such systems are not necessarily static collections of data, but may be added to and altered continually throughout the life of a project, and even amended after the deposition of an archive in a repository.

Source repositories

The questionnaire phase of this survey suggested that although awareness of source repositories in archaeology, and especially the Archaeology Data Service, is generally high, the number of active depositors is quite low. This was borne out by the interview phase also, which

When asked how frequently researchers deposit data within a source repository, one interviewee commented as follows:

'At the moment, very little. That's because I haven't yet finished producing any data of my own. I know that the database which I'm currently producing using Portable Antiquities Service data, and from the Early Medieval Corpus which I'm currently cleaning up, will be stored probably through ADS in some form or another. We're also going to be producing a big Internet Archaeology paper, so that will link to the ADS archive.' (Int 006)

This sense, of their being an intention to submit, but that research was not currently at a stage which would allow this, was common to a number of other researchers. Many produced comments such as, 'I'm not quite sure what my two bosses' views are - they seem to think that this project may go ahead and be widened using more data so, yes, it would make sense if it was accessible in the future' (Int 008), or that ...

In common with the questionnaire, the interviews suggested strongly that many intend to deposit in future:

'We are going to start putting them all online, but we haven't started that process yet. We have found that our work is not getting into the public domain for other people to access.' (Int 016)

Another interesting aspect to emerge from the questionnaires was the extent to which archaeologists submit data to digital archives, and even manage archives themselves, which were not included within the choices provided in the questionnaire. These include such resources as:

- Portable Antiquities Service
- Early Medieval Corpus
- Historic Environment Records (HERs: formally known as Sites and Monuments Records)
- National Monuments Record
- English Heritage Geophysical Survey Database
- Royal Commission on the Ancient and Historical Monuments of Scotland
- DSpace
- University of Bournemouth repository

In common with the questionnaire results, it has been demonstrated that there are high levels of awareness of source repositories in archaeology. However, in spite of this awareness, the fact remains that around half of the interviewees had not yet deposited any of their own data in a repository.

Others felt that their research did not produce 'data' as such, and that their project archives would not merit inclusion in a repository. One researcher feared that 'in my case I probably don't produce 'data'. The nearest I would get to it would be tables and maybe the odd diagram' (Int 015) – in fact, when further prompted, this researcher went on to demonstrate that they generate both photographic images and maps as part of their work.

There appear to be two further, and fairly direct, reasons for this lack of deposition. The first is fears over the amount of time it takes to assign the correct metadata to a particular dataset, and the second is an occasional (but nevertheless apparent) reluctance among researchers to share their data publicly. These two issues will be dealt with in the next section.

With regard to use of source repositories, two statements can be made. The first is that many researchers are not immediately aware of what constitutes a source repository, or at least do not use this term to describe them. It was quite frequently the case during interviews that respondents claimed not to make much use of source repositories, but then later demonstrated quite an active use of and interest in them. Perhaps partly because of the way in which source repositories were described to them, some interviews equated them simply with the Archaeology Data Service, rather than recognising that a whole range of other resources could also be classified as such.

The second statement which can be made is that *all* of the interviewees made some use of source repositories, but for various different reasons. For some, it is to acquire data specifically for the

purpose of explicit comparison with their own research, and often in a methodological sense. Interviewee 003, for example, stated that:

‘If someone’s used an approach, you might want to actually go back and look at their approach to see if it’s applicable in a different regional, temporal and spatial setting. It can often be used as a template.’ (Int 003)

For Interviewee 008 too, seeking out others’ data sets was a matter of understanding how their research processes had been conducted: ‘There are different people using different data and trying different techniques, and I need to know how they reached their conclusions’ (Int 008).

Some researchers go beyond seeking out data sets for methodological comparison, and actually make use of others’ data in their own analyses. Interviewee 006 is engaged in a project based almost entirely on combining and analysing data sets drawn from online source repositories. After first ‘bringing in geographical data from Landmap and Batholemew into a GIS’, he then imports artefact information from the Portable Antiquities Service and the Early Medieval Corpus. This then forms the basis for studies of the ‘archaeology of economics, exchange and trade, and production’ (Int 006).

Other researchers use source repositories in a less incisive, analytical sense, preferring instead to use them more ‘lightly’ for the sake of curiosity and generalised comparison. The possibilities of having large, varied data sets online means that such comparisons can be embarked upon in a fairly unstructured fashion. Interviewee 004, for instance, was content that, ‘I may lose the thread of what I was originally looking for, but I am quite happy where I end up’.

For researchers involved in teaching, the majority of those interviewed expressed a desire to use source data in a pedagogic role, usually deriving from their own research but sometimes from the research of others.

5.9 *Metadata*

The questionnaire results suggested that there are high levels of awareness among archaeologists of the need for metadata. Even where interviewees professed to be ‘only vaguely familiar with this term, they still appreciated the importance of keeping ‘a record of why things are the way they are and how they’re being modified’ (Int 010). This understanding seems very often to be born of frustration with the data sets of other researchers, which are not always associated with sufficient metadata to allow a proper deconstruction of their content. Frequently, researchers expressed the need for greater levels of metadata...

Some researchers commented on the inadequacy of metadata provided in association with datasets:

‘There are certainly very strong issues in terms of the re-use of data where the metadata record isn’t as good as it should be.’ (Int 011)

‘The amount of reading I have to do in order to find out one small process in what others have done with a data set – often five or six pages – can get a bit daunting. If metadata is done properly, it’s fantastic. When it’s not, it can be a real problem, yes.’ (Int 008)

For a number of archaeologists, there are difficulties in finding data online due to inconsistency in the terms used across different repositories:

‘...what is often missing is the application of common wordage as keywords which would bring up that data. Instead, projects sometimes choose obscure terms for their metadata. They are not the same terms that I would expect to use.’ (Int 014)

Yes, I do often get frustrated – there are problems with all sorts of catalogue entries when they’re put together by none specialists in certain fields. They are not recording the kind of information that you might like, which is easily fixed if you can actually go and look at the data and make your own additions to their catalogue notes. (Int 013)

‘Within the Portable Antiquities Service, [there’s the problem] of keywords within itself – there’s such variation and such a lack of standardisation that it’s very difficult to be sure that you’ve got everything you want.’ (Int 006)

One repository manager struck a more optimistic note, implying that metadata issues are being addressed by the discipline. When asked whether metadata was a problem in archaeology, he stated that:

‘I think the situation now is a lot better than it used to be, and many people now seem aware of the importance of metadata. They are aware that essentially, if metadata doesn’t exist to support their data, then their data doesn’t exist at all – it just isn’t visible.’ (Int 012)

For a small but active minority, metadata generation was not only a priority, but the importance of its collection throughout the research is forming an important part of investigative processes. For example, one interviewee stated that:

‘This is a real problem. We are looking at ways of developing in-lab automated recording systems which will provide much fuller information for users of our datasets in the future.’ (Int 001)

However, the assignment of metadata at such an early stage in the research process appears to be rare, as the majority of researchers who assigned metadata did so either at the point of submission to a repository, or as they were indexing material at the end of a project. In common with the questionnaire results, a strong correlation exists between the application of metadata and the levels of support provided. When not prompted by a list, such as that provided in the questionnaire, many were vague about precisely what categories of metadata needed to be assigned to their own work. Interviewee 007 stated that she ‘would need a lot of guidance’ when assigning metadata, for example. Others were less concerned with the ‘straightforward’ information of author, date, subject keywords etc, but were keen to emphasise the need for some form of contextualising free text narrative, as others would need to understand the kind of approach that we’re taking in our work’ (Int 015).

‘Other researchers will need to know the basic problems with the original data sets – that’s fundamental. They also need to know how I’ve approached it and what I’ve done with it.’ (Int 006)

Such information would seem difficult to fit into standardised metadata schemes, but could conceivably be improved a great deal by source to output linkage, so that the associated publication explains much of the broader context of a particular data set:

'I would follow the protocols – such as Dublin Core – so that the data is as accessible as possible. I think for what we would want to deposit, that would be enough, especially if the archive is then linked through to the publication.' (Int 014)

Given these generally high levels of awareness then regarding metadata, the most significant factor preventing researchers from submitting metadata, and indeed from depositing source data as a whole, is not necessary ignorance of the issues. In fact, it is more likely to be the amount of time required (or at least, perceived to be required) to complete the relevant information. One researcher was not alone in both appreciating the importance of metadata, but also stating that it has 'taken a back seat compared with the task of actually collecting and recording the data' (Int 010). For another archaeologist, the task of submitting a 'grey literature' report to the 'Online Access to the Index of Archaeological Investigations (OASIS), held by the ADS, was 'an absolute nightmare' (Int 016), and that 'it's such a cumbersome way of uploading information'. Another archaeologist feared that:

'Metadata is something that makes me think of a huge amount of more work! With the thin sectioning that I do, I could do images of them all but then I know that I'd have to do explanations with each and include all the metadata. It's an enormous thing really!' (Int 007)

This fear seems to be associated with a broader perception of repositories, that the whole process of deposition is time consuming. When asked whether they had submitted any information to an online source repository, one researcher replied:

'Never. I have no reason not to, but the pressures of time. It sounds pathetic, but it's another additional job that you don't have to do.' (Int 004).

This would seem to be a fundamental problem with data deposition, and relies on information support being provided by repositories in order to reassure researchers that deposition need not necessarily be an arduous process, as long as metadata is collected as an integral part of the research project.

5.10 *Data access and sharing*

The questionnaire suggested that most researchers in archaeology would be happy to share their data widely. 64.9% stated that they either had already, or that they intended to, deposit in the ADS, while a further 13.8% had or intended to deposit with another source data repository. However, the interview phase revealed striking contrasts in researchers' attitudes towards sharing data. These contrasts may in part be due to the discrepancy between the questionnaire respondents and those chosen for interview, with the former being generally keen to engage in the subject of online repositories, and the latter a more genuinely mixed constituency.

Many interview respondents voiced concerns about sharing their data with others. These can be grouped under two broad categories. Firstly, there are those who are happy to share data with colleagues, but who would rather not make such material available publicly. For example, a number of respondents were enthusiastic about the possibilities of exchanging data with colleagues for the sake of comparison, on a one to one basis, in order to 'exchange our ideas informally, get feedback and thoughts and criticisms' (Int 010). However, this data is often made available on a case-by-case basis, to research associates, friends and colleagues:

‘It depends who it is [that has made the approach] and whether you know them very well. Often it’s the case that you have to build up the relationship first by knowing people through conference attendance, before there’s this kind of intellectual ‘back and forth’. It’s also largely distributed on the understanding that it’s not published and it shouldn’t be cited in formal publications – it’s more just for letting people know what’s out there.’ (Int 010)

Data is shared in this way via email attachments and portable drives, and occasionally via internal networks. To researchers such as these, the idea of making their data available to all online is not a possibility either now or in the future. The reasons often given for this are varied. For two respondents, the fact that they were working abroad at the behest of another nation’s government means that making data about that country freely available would be problematic:

‘...the material that we’re working with, whilst we have some of the intellectual ownership of that, I think a large proportion of that lies with the cultural heritage organisations in the countries where we work. For example, in the Emirates or Iran, I think they would regard themselves as the people who have rights over the distribution of the data that arises from research in those countries.’ (Int 010)

‘It’s out of my hands really because it’s essentially not my data – the data belongs to the Egypt Exploration Society and I’m working at their behest. So, it’s not up to me really, it’s not my data. The arrangement that you enter in to on a lot of the Egyptian projects that you work on, because the permits go through the Egyptian government, and everyone’s essentially working for the Egyptian government and you’re working on a project for them, so it’s really up to the director of the project or the organisation what happens, so any paper notes and electronic notes that I take I give copies of them all – and then I will eventually give the originals to the Egypt Exploration Society for the current project I’m working on.’ (Int 013)

Even once a project is completed, some of these archaeologists felt that the organisations and governments with whom they were working abroad would not view the idea of public access to the data positively.

For other archaeologists, particularly those working within the UK, the issue of the illegal looting of sites which they are working on is perceived as a major problem. In spite of the extensive liaison process undertaken by the Portable Antiquities Service and others in recent years, some archaeologists still feel that revealing the location and details of archaeological sites which have not yet been fully investigated would be irresponsible. For example:

‘...what are people going to do with the data? I still have a very strong feeling of access by metal detectorists. I recently had several sites where I did not deposit the information with the SMR (the local Sites and Monuments Record) either, because obviously everybody can go there and access the information. We had instances on one particular site where we know that people used our data to [illegally] dig up the site – obviously that’s pretty appalling. So, there are certain issues with making our geophysics data publicly available. One could make a distinction between making research data publicly available, and making it available to other researchers. That’s one of the slight problems with ADS that, when one deposits with them, it’s publicly available.’ (Int 011)

One of the local Finds Liaison Officers for the Portable Antiquities Service explained the formal restrictions which are placed on the data which they hold in order to tackle this problem:

‘There are different levels of access built into the PAS database. Level 1 is the public access that is freely available on the web. This does not display all of the grid references (only 4 figures are given), so the finest resolution one can get at this level is one square kilometre. It may not give information about the date found and the finder, so not all of the information is given out. This information is available all over the world, and a user is not required to register or to sign in...Level 2 is open to those who are actually doing the recording. Level 3 is an administrative level. Researchers have to apply for Level 2 information to be able to extract what is sometimes considered sensitive information like findspot locations and full grid references which are not open to the public, so that certain areas are not pinched by metal detectorists. We don’t want to encourage treasure hunting.’ (Int 005)

Another category of non-depositor are those who view the whole notion of data sharing problematic. These are very much in the minority, but their attitude seems to be one of extreme reluctance to make their data widely available:

‘I would worry that if I went to ADS with my corpus of data, you do open yourself up to a lot more potential criticism I suppose, not just from archivists but from people who can just delve in and say, ‘What a ridiculous sample’...The idea of someone diving in to a dataset and just pulling apart the foundations of someone’s research is not an attractive proposition...I’ve never said, ‘Here’s all my data, help yourself!’. I think there’s a problem with that probably. Somebody would then write a paper on it – they’re bound to.’ (Int 004)

For other researchers, indeed for around half of those interviewed, stated that they had no problem at all with sharing data online. Most of these, however, felt that a published synthesis should appear either first, or in conjunction with the data archive:

‘I don’t think I’d worry about people publishing prematurely because if you publish what you want to at the same time as archiving your data, then you’re not going to worry about it – that’s how I would want to do it.’ (Int 007)

This kind of approximate, time-related embargo was common to many researchers, who would be happy to archive in a source repository, and thus make their data widely available, as long as they had published a synthesis first. These researchers were often those who were the most enthusiastic about source to data linkage.

5.11 *Output repositories*

The questionnaire revealed that, in contrast to source repositories, archaeological researchers seemed comparatively unaware of source repositories. 23.1% claimed never to have consulted an output repository and 41.5% had never deposited a publication in one. This impression, following the interview phase, in fact appears to be misleading, and derives in large part from a misunderstanding about what actually constitutes an output repository. Almost none of the interviewees (with the exception of Interviewee 001) had ever used an institutional repository such as the White Rose ePrints, although a small number were aware of their existence. For the majority, if a broader definition of an output repository is used (to include such resources as publishers websites, discipline specific repositories and online journals), they had in fact all used an output repository but did not term it as such. In a number of cases, as the interview moved along it became apparent that the interviewee had often deposited publications online, but did not think of such deposition as having anything to do with output repositories.

The reasons why the majority had not engaged with institutional repositories are threefold. Around third of interview respondents were unaware of such a service, either because it does not yet exist (as at the University of Bradford), or because they had not been made aware of it. Secondly, although some were aware of an ePrints service at their institution, they were unsure of how to proceed in depositing. Issues such as which version of a paper to submit, what format and how to incorporate figures all arose. It may be that time pressures are also a factor in resolving these uncertainties. Finally, a small minority are concerned about the issue of copyright, and have not found a way, or do not know how to go about, clarifying these. In some cases, although researchers were enthusiastic about the idea of Open Access (attitudes to which are addressed in more detail, below), they could also appreciate that publishers would have some concerns. This apparently contrasting attitude was common among interviewees. For example:

‘I think it’s a good idea from a researcher’s point of view – I can understand why publishers would be pissed off about it!’ (Int 013).

The issue of paying for the resources provided by output repositories was also raised by a number of interviewees, and consequently will be discussed further below.

The questionnaire indicated that researchers reached the papers provided in output repositories via a number of different routes. In fact, the interviews suggested a narrower range of options pursued. The majority use one of two routes: either Google (Scholar or Home), or through a link from their institution’s library catalogue. Many researchers appeared to like the large numbers of ‘hits’ that each route provided, as it offered the opportunity of browsing through papers of which they were unaware. In order to pursue this kind of searching, most respondents use a simple keyword search. However, when they are looking for something more specific, they would use an advanced search.

A small number of researchers stated that they were frustrated with the level of searching allowed by output repositories, and wanted to be able to refine their searches more accurately, as well as to save certain sets of search criteria:

‘Sometimes there are limitations where you can only filter results once. But if you can keep repeating that process until you feel it’s useful, that’s quite helpful, though it isn’t always the case at present.’ (Int 005).

5.12 *Information support*

It was very clear from the interview phase that the majority of researchers do not seek help when dealing with output repositories. Responses can be divided into those who feel relatively self-sufficient – even when they do encounter a problem – and those who seem to have little faith in the support network available even if they do encounter problems and cannot solve them alone. Into the former category fall the following quotations:

‘There often isn’t any help available at all.’ (Int 016)

‘I think I usually give up before I ask for help, unfortunately!’ (Int 008)

‘Often the help files are pretty unhelpful – they’re often written by someone who knows exactly what they’re doing. They don’t necessarily think that someone may be coming in not knowing how the system works. In general though I can get to what I need.’ (Int 006)

'No matter how expert you are, there are always specific tricks, traps and techniques that will provide better answers. These are not static organisations or tools. Six months ago I may have used a repository, but come back and find that the interface has completely changed.' (Int 003)

Into the latter category fall those who do not ask for help, but appear confident and self-sufficient in their use of repository resources:

'...I find most of the web interfaces these days reasonably straightforward. As long as there's a search box, it usually produces something which one can interpret reasonably well.' (Int 011)

'I generally find it fairly easy...I don't think I've experienced any particular problems. In general it seems okay and I don't think I'm missing anything – the only problems occur when I try to do things too quickly then the system may crash. But then all that means is that I'll have to come out and go back in again, which is only a bit of extra time and not anything particular to worry about.'

Others, though a small minority, feel they can regularly call on a professional advisor as and when they need help. These often seem to be those who are involved in repository management or digital archiving in some way themselves, and who are perhaps therefore more aware of the resources and networks for intermediation available to them:

'The PAS has its own ICT advisor...He's is only a phone call away, and is very helpful.' (Int 005)

Some researchers were quite specific in the problems which they regularly encountered and had to ask library staff to help them with. For example:

'The biggest problem and frustration I would have when using online repositories is the amount of authentication required, so often I forget passwords and have to have them renewed. If there are too many log-ins required, it can be very frustrating!' (Int 012)

Some useful further suggestions emerged from the question, 'what functionality is missing from the output repositories you have used?'. A brief list of these suggestions is as follows:

- A geographical search facility – akin to the ArchSearch facility provided for source data through ADS – would work well for output publications too. Users could click on a geographical region and be shown all of the papers available which deal with that area.
- The ability to export search results directly into Endnote (as is currently the case with the Web of Science).
- Greater presence of books and book chapters within output repositories.
- Digitisation of older journal articles, especially those in county-based archaeological journals

Beyond the specific remit of the StORe Project, a number of additional issues arose as a results if the interviews which it is worth presenting in the following section.

6) Additional Issues

This section presents those issues covered by the questionnaire and interviews which have not already been described in preceding sections.

6.1 Identities

All of the relevant information on this issue has been dealt with sections 4 and 5 of this report.

Project Aims

When asked in Question 2, ‘Source repositories contain primary research data. If a standard feature of such repositories was the ability to identify and link to the publications that had been developed from these data, how advantageous would you find it?’, the following free text responses were produced by respondents:

Archaeology generates mixed assemblages, so publication is often dispersed across several specialist literatures.
I often generate datasets which I am asked to share, or seek to share others' data.
It would occasionally be extremely useful to be able to access the raw data, but only exceptionally, relative to the number of articles I read. If this pattern is more general, one has to wonder if the infrastructure development, and constraints on data format, actually will repay the costs.
Many output repositories present summary data only; one of my main interests in synthesis of data from specific periods/geographical regions so primary data are crucial.
The large amount of data generated through archaeology is a big problem, and often detracts from publications. If we could archive data and link to publications this would provide a means for greater accessibility of original datasets as well as resulting in more manageable publications.
The usefulness would be directly linked to how well this was designed.
This would be a great help but it would require the links to be robust, consistent and accessible. Links that were fragile or were not consistently addressed to the right material, or required me to enter another set of logins would be counter-productive.
This would be an immense help to my research - but it would be no good just to have the framework set up, and the ability to add the details about published research - I would never get round to it. The publications (old as well as new) would have to be added centrally as well.

Table 32: Free text responses indicating attitudes to source to output linkage

When asked in Question 3, ‘How advantageous to you would it be if it were possible to go directly from within an online publication (electronic journal article or other text) to the primary source data from which that publication was developed?’, the following free text responses were produced:

A link would mean researchers could work directly between the publication and the data with little hassle to them.
Already thinking in these terms - initial project will be to have a textbook with a url to data.
Also as a journal editor, I feel that this would be of considerable advantage to readers of the journal, and the contributors.
As above. The trick is surely to have both in one place rather than having to link across. I'm not sure the distinction between source and output is very clear or coherent.
As long as the data is presented clearly, for example, translated out of project abbreviations.
I would rarely want to drop everything and re-analyse the data right then. More likely, I would finish the article,

and then access the data to explore a particular issue. So a direct link would be of little use.
Papers generally (ought!) to publish the data to support the inferences made in that piece of research, but cannot publish all the peripheral data that would enable a spin-off investigation to be pursued.
This would be a significant advantage when trying to identify what I am seeing in my own data and raw data is rarely available for comparisons. This is especially important as my Department is not able to purchase the latest libraries of data which are available and because archaeological materials are rarely included in these libraries.
When looking at artefact catalogues it would be useful to be able to find out more detail on specific objects or assemblages which may not have been discussed in detail in a report.
Will depend upon the nature of the project - could be useful in seeing the detail of archaeological reports.

Table 33: Free text responses indicating attitudes to output to source linkage

6.3 Source Data

In support of the commentary given in section 4 above, when asked in Question 6, ‘Are the data you generate sometimes a combination or group of different data formats?’, the results were as follows:







Often:		52.3%	34
Sometimes:		33.8%	22
Rarely:		9.2%	6
Never:		1.5%	1
Potentially:		1.5%	1
Other (please specify):		1.5%	1

Table 34: Number of archaeologists who produced a combination of data formats

6.4 Source repositories

When asked in Question 7, ‘To which source repositories do you submit your data’, it was clear that the majority who responded submitted to the ADS.






Source Repository	No.
Archaeology Data Service:	 29
Brookhaven National Laboratories:	0
CERN:	0
Genbank:	0
National Crystallography Service:	0
NERC Data Centres:	 2
Protein Structures Database:	0
SuperCOSMOS Science Archive:	0
UK Data Archive:	 2
UniProt:	0
None:	 29
Other (please specify):	 20

Table 35: Source repositories used by archaeologists for data deposition

However, the free text responses suggest that a number of additional source repositories were used to archive data:

Alexandria Archive http://www.alexandriaarchive.org/ .
County Council Historic Environment Records, designated museum archives.
dSpace.
English Heritage / NMR.
http://ausgegraben.org .
I intend in future to make geophysical data available from the BSA web site.
Integrated archaeological database (university of York).
Many of my primary data are subsumed within excavation reports, outlines of which are passed onto OASIS.
None as the ADS do not hold this kind of data and science archives are not usually interested in archaeological data. Also it is not encouraged as it never seems to occur to anyone that this kind of data would be useful and I think there is a fear of allowing others access to your raw data. I have to say that I had not even really considered submitting my data to a central repository.
Portable Antiquities Scheme database (www.findsdatabase.org).
Put online on my own website.
RCAHMS.
Unsure of at present, depends how project pans out.
Will be Archaeology Data Service.
Will be doing so, for a current project, just beginning. Have not yet done so.

Table 36: Free text responses indicating additional source repositories used by archaeologists for data deposition

The most obvious comments to include here from interviews are those that relate to the usefulness or otherwise of specific source repositories that respondents had used. This form of feedback, though anecdotal and somewhat brief, may provide some constructive criticism for repository managers to act upon. The comments presented here are only those which refer specifically to the source repository dealt with directly by the archaeology component of the survey, namely, the Archaeology Data Service.

A number of researchers provided feedback on the ADS in the early stages of the interview phase of this survey. Consequently, a question was then added to the interview questions: ‘What are your views on the service provided by the ADS?’. The responses received can be grouped analytically into five approximate categories:

To begin with, two comments were generally complementary of the services provided by ADS, but provided little specific detail:

‘I do use it a reasonable amount. Obviously, the bulk of my research at the moment has been elsewhere, but I think certainly now that there’s more grey literature going online, there’s more archives going online it is becoming an increasingly useful tool. As time goes on, I think it will be invaluable.’ (Int 006)

‘I think it is really good. I do like looking through it, particularly as it’s always being added to. I think sometimes the front page is quite difficult. I now know how to get to

ArchSearch, but previously it was difficult to get to, especially from the AHDS pages. In general, it's a nice site and it's well laid out.' (Int 007)

Secondly, the largest number of comments came from those for whom the ADS has no direct relevance, either because they work abroad and the ADS does not currently hold collections that are in their geographical region, or because the researcher is unfamiliar with the ADS and merely *perceives* that it holds nothing of relevance:

'I don't work for it is the problem – I haven't used it. A terrible admission to make! I just haven't looked at it, I don't know whether I'm missing something enormous or not. But I think also all my stuff that I do read, I never read, 'This article wouldn't have been possible without the ADS', so I don't think I must go there and see what they've got. It could be that people are drawing their data from it, but I've never ever come across a reference to it. Bear in mind that I come from history as a background not archaeology...they've given me some advice on checking grants. I know they're a rich and valuable resource, but I just haven't used them. Actually I have used it to see what they had around the area of Hutton, but didn't find anything that related to my period. It's perhaps partly that I've never had any funded projects, so have not yet been obliged to submit anything: I don't take anything out because I don't put anything in. A project or a bursary which would provide for a project directly linked to the data which ADS provide would bring it into the public domain slightly more, but I'm sure that's the last thing they've got money for! I've been to presentations about HEIRNET or HEIRPORT or whatever it's called, and have gone away thinking I really must use it, but just never have.' (Int 004)

'I've never made use of it. I wasn't aware of it until I came to the UK. There doesn't seem to be much material in there which relates very closely to the research that I'm doing. Thirdly, there's the point that I wouldn't feel very comfortable depositing information in that without the complete approval of the collaborators in the host countries where we work, and I think they would have substantial issues about raw data from these excavations being deposited in there, especially without a very clear guidance on rules governing access to that data were.' (Int 010)

'I haven't actually looked at the ADS for Egyptological things – I did once because we had someone from them come and talk to us about it on an away day we had so I looked straightaway after that while it was fresh in my mind, but it tended I thought to be mainly British and non-British in the sense that when someone had gone to them they would archive it, but there was nothing of interest to me. It looked really interesting if you were working on British or English archaeology, but at the time I looked at it I didn't think there was anything there that linked into specifically what I was working on in Egypt.' (Int 013)

'I haven't really used it, no. I think I've gone there for a couple of things, but it was some time ago now. I think it's mostly an issue of relevance. Also, I don't carry around in my head a list of useful sources, and to the extent which I do, the ADS wouldn't necessarily be on that list. I guess it's also laziness.' (015)

'We don't use it a lot, no we don't. It doesn't crop up that often for us [in searches], and a lot of the information that we require is already readily accessible [without having to go to the ADS].' (Int 016)

Thirdly, there are those respondents who voiced criticism of the ADS, but appear not to have used it in some time, and whose views in fact derive from older versions of the ADS:

‘It’s a couple of years since I used it and I’m not sure if I can remember that far back! I was looking for some bone assemblages around Oxfordshire, and it was very useful for that. But then again, although the sites were mentioned there I had to actually go and find the references for myself – there was no automatic link-through.’ (Int 008)

‘Some aspects of it are quite useful. I do get frustrated sometimes with the interface. It seems almost topological in the way you have to go through things. Because the ADS has got several different types of interfaces, there isn’t a unified approach, that is sometimes irritating. I use it to have a quick flick round. I don’t find it useful at the moment as a great tool. Part of the problem is there is no real graphical reference of the temporal or spatial coverage.’ (Int 003)

Fourthly, some respondents had specific comments on either the search interface provided by ADS, or on particular resources for which they searching but without success:

‘I found the other day a digitised microfiche supposedly available on the ADS, but the link did not work. You really do have to make sure that such links are always live, otherwise it defeats the object of digital resources.’ (Int 002)

‘I think something to consider with the PAS repository, but also the ADS, is the issue of depositor versus user. Often, it’s very easy to input data, but as a user trying to extract data it can be frustrating...I have sometimes found it the case with ADS, that although I know that something is on the ADS site, but I can’t seem to find a way of getting to it easily. I always wondered if that was an issue of designing for people to input and import data, but they have considered the complications for people and for making it user friendly. Often though, the ADS geographic interface I’ve found very useful, especially when it’s used in combination with a filters that you can use repeatedly. Sometimes there are limitations where you can only filter data results once. But if you can keep repeating that process until you feel is useful, that ‘s quite helpful, though it isn’t always the case at present. There are currently a lot of limits on what you can filter, and that’s very frustrating from a user point of view.’ (Int 005)

Lastly, a number of researchers discussed aspects of the technical interface provided by ADS:

‘I think it’s brilliant – it’s really very, very good...I have seen other services and I’m quite happy to say officially that what EDINA offers on a far larger budget is certainly not better than what the ADS have put together. I think the ADS have shown the way in many different areas of online repositories. I think the way ADS still is maintaining the cutting-edge of the technology is really very good - with the limited resources that they have available it is amazing what ADS has been able to do. So, full endorsement!’ (Int 011)

‘It is important for ADS to keep attracting further data sets and to continue to build the resources it has, without necessarily changing its interface or facilities too much so that users have to keep learning new skills. The architecture could be retained, but with increased numbers of data sets available.’ (Int 012)

‘I have the impression that a lot of things on the Archaeology Data Service seem to be just dumps for people’s reports. They haven’t got a publisher so they’ve just dumped all

of this data on there, and they have not put it in any way, shape or form that makes it interesting. There is one on there that I had a look at where you could click on the archive and have a look at the cemetery data – you could look at the position of the burials and how it related to the rest of the site. That’s the sort of thing - you need to be able to fully interrogate what’s in there I think. Another impression I have with the ADS is that people [who deposit] have got funding to produce an electronic archive. And, yes, they produce an electronic archive but it’s often very half-hearted. I don’t think they are fully aware or capable of producing an archive that is really a valuable entity in itself. They are really just seeing it as putting data into a publication in the same way as you would put material into a print publication as opposed to an electronic publication. There are cases where depositors clearly know they have to produce an electronic archive and they do, but it’s not very imaginative, let’s put it that way! It can be just ‘ticking a box’, and I think a lot of information is actually lost that way.’ (Int 014)

6.5 Metadata

Although most metadata related results have already been covered in sections 4 and 5 of this report, it may be worth presenting the following results. When asked in Question 10, ‘At what stage are metadata assigned to your data?’, the following free texts responses were received:

As part of the data.
Depends on overall project manager / director.
Don't know.
During compilation of the project archive ready for deposition.
In part before, but it partly depends on specific task being undertaken.
Metadata is created and managed throughout the lifecycle of the data creation process and presumably post deposition also.
What i do with my files now is different from what I would/should do were they to be designed for a repository.

Table 37: Free text responses indicating the stage at which archaeologists assign metadata

6.6 Data access and sharing

When asked in Question 12, ‘Why might you wish to access the research data generated by other research programmes?’, the following results emerged:







Reason:		No.
To understand the broader context and orientation of my research:		48
To test the uniqueness and validity of my research objectives:		25
To test the uniqueness and validity of their research objectives:		20
To identify useful contacts:		26
To access data that are useful or necessary to my research:		63
Other (please specify):		7

Table 38: Reasons for accessing others’ research data given by archaeologists

In terms of access technique, when asked in Question 13, ‘How would you normally access the research data of other researchers?’, the following results were produced:







Reason:		No.
I do not normally access others' research data:		10
Through the exchange of data held on portable media:		37
By access to networked file servers at my own institution:		12
By access to networked file servers at other institutions:		8
Through online access to source repositories:		31
Other (please specify):		15

Table 39: Methods by which archaeologists access others' research data

6.7 Output repositories

With regard to output repositories, all of the relevant questionnaire results have already been covered in section 4. However, it is worth presenting the comments received during the interview phase about the output repository most relevant to the survey, the White Rose ePrints institutional repository. It should be born in mind that interviewees were not asked specifically about this repository, and that the following comments emerged in an ad hoc fashion when researchers were asked about other issues, such as ‘Which kind of output repository do you use when looking for information to draw on in your research?’.

WHITE ROSE ePRINTS

The most striking element of feedback to be provided by the interview phase of the StORe Project is that very few archaeology researchers currently appear to make use of the WR service, either for depositing or downloading papers. This situation does not appear to be specific to this repository alone: for researchers outside the White Rose region, this lack of awareness is equally true for their own institutional repositories. However, some researchers did comment specifically on the WR ePrints and their views are issued here:

‘I generally don’t make PDFs available through my own website, nor have I engaged very helpfully with the White Rose project either. Everything they were asking for were articles which could not be provided as PDFs due to copyright. The pre-PDF copies I refused to give them on the basis that they would go without tables, without proper drawings where relevant, and without the proper page numbers – they would just look a mess, unprofessional in that format. All you’d be giving would be a piece of text which they couldn’t cite. Alright, they could go on and find a copy of the proper article, but this seems like a lot of work for no great advantage.’ (Int 002)

‘Institutional repositories are very important to me. They really do help the research process, although I usually find articles in a repository not through the repository itself, but by using Google Scholar. Pagination really doesn’t matter if it’s only the word version of a final manuscript and not the publisher’s PDF – I just need to know what’s out there.’ (Int 001)

‘I will admit that the White Rose Repository did approach me and ask if I would like to submit something, but because of time pressures I haven’t yet done so... With the White Rose Repository, they could make it easier for me to contribute simply by providing a link to an example of what they want so I know what to give them. I think

they just want a cleaned up word document by with the images also, but you're not used to depositing in an online format you don't usually have the two together in a format that works. The prospect of marrying the two up, I didn't have time to do that at the time...[With regard to copyright] I assume that something like the White Rose Repository has its guidelines about how people are supposed to reference it, and obviously I would be happy the more people read it. Whether the publisher would be delighted that everyone could get it free, I don't know.' (Int 004)

'I was consulted [about submitting to the WR ePrints]. There were several papers that were identified as ones that I had written which could perhaps have been deposited in there, and they were asking me whether I had any problems with depositing with them. I don't, but they [the White Rose ePrints manager] were still trying to work out if they could get permission for those [from the publisher]...[With regard to copyright]: they're published anyway, so it doesn't bother me. It's great because it means they are more widely available I suppose. In some ways I don't understand why journals would allow that [papers being deposited in repositories] to be the case, because they have subscriptions and if they are giving permission for people to deposit elsewhere it makes those papers more freely accessible. So I don't understand how it works, but I'm quite happy for it to happen.' (Int 007)

6.8 *Open Access*

Although not an element in the questionnaire phase, often, when the subject of output repositories was being discussed in interviews, the issues of charging policies and the Open Access movement were raised by the interviewee. This section and the following one will now present the relevant comments produced by researchers on these two related topics.

With regard to the Open Access movement, almost every interviewee voiced enthusiastic support for such a movement. For some, this enthusiasm was unqualified:

'I think it's a jolly good idea – certainly within the academic community where the exchange of information and research results is central to what we do. I think that it's grand notion!' (Int 015)

'With the research that we do, it's important that your work is made as freely accessible to everyone and anyone as possible. Particularly people outside academia or people who aren't in an academic institution and so don't have subscription to some of these journals, but still have an interest in archaeology, I think it's more a problem for them accessing some of these papers, so [Open Access] is better. And for students as well, sometimes it's not so easy for them to access things.' (Int 007)

'Without a doubt, I'm a fan of the Open Access movement. I do believe that data that has been collected using public money, or that public money has played a significant role in, should be made available: not necessarily immediately, but I do think that that whole issue of charging for access to data has to change.' (Int 003)

Other researchers displayed an appreciation of the sensitive issues of copyright and the income of publishers. For example:

A minority of researchers felt it was important to assert that in spite of the enhanced availability provided by the Open Access movement, they felt that conventional print forms of books and journals still had a strong role to play in their research:

'I like books and I like to have them, so I think it would be a bad thing if publications were *only* available online. I think that's the wrong way to go – people can't read a whole book of the computer screen. Some people try to tell me that you can but I completely disagree with that. I don't want to sit and read a 300 page monograph off the screen, I'd rather sit and read it from a book. I think it's useful to have it available online, but I don't know of anyone that doesn't print them and stick them in a filing cabinet somewhere.' (Int 014)

I don't tend to actually read literature online. I look up journals that are online, that I don't actually read them online. I'll print them off and read them. I don't actually read them off a screen! If I can find it online, that's nice and convenient, and then it's a case of printing it off and then I've got a hard copy which I can refer to. Particularly if you're comparing how something is done by one person, in comparison with how another person does the same thing elsewhere, it's quite useful to have the two texts side by side, which at the moment isn't *that* easy on a computer screen... I'm a little bit old fashioned in that I actually like the feel of something solid, of turning paper and all that sort of thing'. (Int 015)

6.9 *Charging policies*

A number of respondents during the earlier interviews expressed opinions about the charging policies adopted by repositories. Consequently, a question was included in all subsequent interviews which asked, 'How do you feel about having to pay for certain resources from repositories online?'

'I don't think it frustrates me – obviously one has to pay for books and journals, so the idea of paying for access to material isn't a particular problem from that point of view. It does raise the issue of making choices as to what to subscribe to and what not to subscribe to, but that's only the same as one would with printed material anyway. So from that point of view, it's not a major consideration I don't think.' (Int 015)

'I don't see why you can't have a sort of Napster system, where you pay ten pounds a month and you can download anything you want, but you can't burn it and you can't print it – that would give you constant access.' (Int 006)

'I'm quite surprised how often documents that I thought we would have to pay for are freely available on the internet.' (Int 016)

For some researchers, the central issue with regard to charging policies is where the funding for the original research has come from:

'I would also say that some of the issues of charging have made some data providers basically unusable. For instance, the Ordnance Survey's data we just can't access. It's important to have resources either free or at a nominal charge. But the fact of the matter is that this is a big issue. [With the Ordnance Survey] that was government information paid for by the tax payer, and in this country we have to pay outrageous amounts of money to be able to use that, and are extremely limited what you can do once you've acquired the data. You can then go to a site such as the USGS and get their

data free of charge, so it is actually forcing staff to have projects outside the UK, because of the problems of data sourcing.’ (Int 003)

6.10 *Thesauri and Word Lists*

Some enthusiasm was voiced during interviews for increased use of standardised thesauri and word lists, in order to make different data sets compatible and as searchable as possible. Some of the comments received are as follows:

‘The problem with that database [the Petrie collection] – and with other ones – is how they’ve classified things. The same thing classified by different people could be called, for example, ‘metal’, ‘bronze’, ‘copper’, ‘copper alloy’, and the database doesn’t allow you to easily transfer. So, some things will come up under ‘metal’, but they wouldn’t come up if you searched under ‘bronze’. It’s one of those database problems where you’ve got to have a set nomenclature, which is mutually understandable to everyone.’ (Int 013)

‘Standardised terms for artefacts are hard to put into practice. Even with the standardisation as it is today, there are still problems. Coins come to mind: I could put in ‘fourth century’, but if someone has put in only ‘House of Theodosius’, then that object could be missed out when you run a search. That puts a lot of responsibility on the recorders and the data inputers to select the appropriate term so that it will be chosen in searches in the broadest sense.’ (Int 005)

6.11 *Information support*

Most of the relevant results from both the questionnaire and interview phases have been included in sections 4 and 5, above. However, it is worth simply laying out the results for Question 25 which asked, ‘What assistance in your use of repositories is provided by a librarian or other knowledge management support?’:

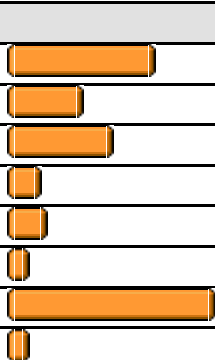
Option:		No.
Provision of documentation (guidance notes, fact sheets, etc.):		25
Formal training and documentation:		12
Online or telephone help:		18
Assistance with the structuring of specific searches:		5
Assistance with the conduct of searches:		6
Full intermediary service (the conduct of searches and organisation of results):		1
Unknown:		35
Other (<i>please specify</i>):		2

Table 40: Levels of intermediation provided to archaeologists

Bibliography

Condrón, F., Richards, J., Robinson, D. & Wise, A. (1999) *Strategies for Digital Data: Findings and Recommendations from Digital Data in Archaeology: A Survey of User Needs*. York: Archaeology Data Service.

<http://eprints.whiterose.ac.uk/> (2006) *White Rose Consortium ePrints Repository*. Page consulted 12.7.2006

J.M. Morrell Library, University of York (2002) *ARCHway – Gateway to Resource Sharing in Archaeology*. University of York: Unpublished Report.

Jones, S., MacSween, A, Jeffrey, S., Morris, R. & Heyworth, M. (2001) *From the Ground Up: The Publication of Archaeological Projects: A User Needs Survey*. York: Council for British Archaeology.

Kilbride, W. & Winters, J. (2001) 'Observing the game: What can access statistics really tell us?', In Z. Stančić and T. Veljanovski, *Computing Archaeology for Understanding the Past: CAA 2000*. Oxford: British Archaeological Reports. International Series 931: 339-345.

Appendix 1: Interview Questions

The two scripts below provide details of the questions used to structure conversation during the interviews. As described in Section 5 (above), a revised question order was used for interviews 008 onwards, though the same questions were used.

INTERVIEW TRANSCRIPT SUMMARY (1)

Name	
Organisation	
Role	
Contact Number	
Date Interviewed	
Time Started	
Time Finished	
Telephoned From	
Duration	

Identities

1. What are your specific fields of interest?

What would a typical research process be for you? How do you go about discovering literature and data produced by others, and how is that incorporated into your own research?

Project Aims

5. Source repositories: if a standard feature was the ability to identify and link directly to the publications developed from these data, how advantageous would you find it?

6. Similarly, if you could navigate directly from within an online article or text to the source data from which it was derived, how advantageous would you find it?

Why in particular would such links be advantageous?

Source Data

7. What kinds of electronic source data do you produce, and in what formats? For example: CAD/GIS, XML, database files, spreadsheets, html, images, plain text, word processed files, PDF, statistics files etc.

9. How frequently are the data you generate a combination or group of different data formats? So, for example, a GIS...

10. How frequently do you submit data to ADS, or indeed any other source repositories?

11. How would you normally access the source data generated by other research programmes?

Source Repositories

12. How frequently do you submit data to any of the following source repositories?

<ul style="list-style-type: none">• Archaeology Data Service
<ul style="list-style-type: none">• NERC Data Centres
<ul style="list-style-type: none">• UK Data Archive
<ul style="list-style-type: none">• Other (<i>please insert</i>)

Metadata

13. What types of metadata do you consider it important to assign to your data? Project, project description, project reference numbers/identifiers, author/data creator, title of data set, subject keywords, funding source, publisher, dates of project, date (of creation), format, other.

14. Do you regularly assign metadata do your own datasets? At what stage?

15. Who assigns the metadata?

Data access and sharing

16. What measures do you use to make your research data available?

17. Why, in particular, would you feel the need to share your data with others?

18. What factors would discourage you from sharing your research data?

19. What kind of formal restrictions do you apply to the release and/or access to your research data?

20. In practice, what actual measures or processes do you use to control access to your data? (*note this is asking about practical measures specifically*) for example: *an approved list of data users, users click on an approved url which has been sent to them by email, password authentication, data on standalone computers etc., no formal restrictions.*

Output repositories

21. Which kind of output repository do you use when looking for information to draw on in your research? *institutional, discipline (ADS grey literature library) and publisher repositories.*

22. Would you use an output repository when seeking information for use in teaching? (*keep the focus on the three main types: institutional, discipline and publisher*)

23. Have you deposited your own research publications in an output repository? Which one?

24. What routes do you normally take into the contents of output repositories?

25. When you are using an output repository, what level of search do you normally find provides you with satisfactory results? For example, simple, advanced, Boolean logic, subject keywords, geographical interface?

Support

26. Do you find yourself requiring advice in using output repositories, either from individuals or from online links or advice?

27. What form does this advice take? Is it effective? What could be improved?

Reprise of project aims

28. Having now considered both source and output repositories, and how they might relate, what functionality do you consider to be missing from the source repositories you have used?

- Source repositories

29. One of the aims of this project is to develop effective links between repositories of source data to repositories of published papers because we believe there is a need amongst researchers to identify published (and pre-published) papers that have made use of their source data. In what way could you identify with that perceived need?

30. Linking to source data from output repositories will require that an adequate range of metadata is applied to the source data that will persist over time. Do you foresee this as a problem in archaeology? What solutions might there be here?

Reprise of project aims

31. And, similarly to the question that focused on source repositories, what functionality is missing from the output repositories you have used?

- Output repositories

32. We are considering building an interface for output repositories that would let you, as a depositor, associate newly deposited publications with the data from which they are derived. In what way might that be of benefit to you or, indeed, others? Genuinely important, or just a 'nice to have'?

33. A number of new operations could be supported within an output repository. For example:

- The automatic creation of links between related resources
- The automatic embedding of source repository data
- The presentation of relationships (i.e. showing publications and their source data in adjacent windows)

How do think these could meet your needs? What other features might you expect to be advantageous?

Reprise of project aims – potential solutions

34. One option we are considering for making two-way links between source and output repositories is the development of an online service known as a '**dataset knowledgebase**' – a central store which holds information about which source data relates to which publication or report. This might also have additional features such as quality assessments or ratings from users, frequently asked questions (FAQs) about specific sets etc. What is your opinion of the value of such a concept and are there specific issues you might want it to address?

35. Some data repositories are open to all enquirers while others are password protected. If we are expecting to design links that will provide access from open repositories to controlled repositories, we shall need to devise some level of validation and temporary access rights. What degree of access validation do you think is appropriate when users are accessing data sets?

Additional questions

36. How do you feel about having to pay for certain resources from repositories online? Do you feel this is justified?

37. What are your views on the service provided by ADS?

INTERVIEW TRANSCRIPT SUMMARY (2)

Name	
Organisation	
Role	
Contact Number	
Date Interviewed	
Time Started	
Time Finished	
Telephoned From	
Duration	

Identities

1. What are your specific fields of interest?

What would a typical research process be for you? How do you go about discovering literature and data produced by others, and how is that incorporated into your own research?

Source Data

7. What kinds of electronic source data do you produce, and in what formats? For example: CAD/GIS, XML, database files, spreadsheets, html, images, plain text, word processed files, PDF, statistics files etc.

9. How frequently are the data you generate a combination or group of different data formats? So, for example, a GIS...

Source Repositories

12. How frequently do you submit data to any of the following source repositories?

<ul style="list-style-type: none">• Archaeology Data Service
<ul style="list-style-type: none">• NERC Data Centres
<ul style="list-style-type: none">• UK Data Archive
<ul style="list-style-type: none">• Other (<i>please insert</i>)

Accessing others' data

How often do you access the data generated by other research programmes?

Why might you wish to access the source data generated by other research programmes?

11. How would you normally access the source data generated by other research programmes?

Metadata

13. What types of metadata do you consider it important to assign to your data? Project, project description, project reference numbers/identifiers, author/data creator, title of data set, subject keywords, funding source, publisher, dates of project, date (of creation), format, other.

14. Do you regularly assign metadata do your own datasets? At what stage?

15. Who assigns the metadata?

30. Do you view the application of metadata as a problem in archaeology? What solutions might there be here?

Data access and sharing

16. What measures do you use to make your research data available to others?

17. Why, in particular, would you feel the need to share your data with others?

18. What factors would **discourage** you from sharing your research data?

19. What kind of formal restrictions do you apply to the release and/or access to your research data?

20. In practice, what actual measures or processes do you use to control access to your data? (*note this is asking about practical measures specifically*) for example: an approved list of data users, users click on an approved url which has been sent to them by email, password authentication, data on standalone computers etc., no formal restrictions.

Output repositories

21. Which kind of output repository do you use when looking for information to draw on in your research? *institutional, discipline (ADS grey literature library) and publisher repositories.*

22. Would you use an output repository when seeking information for use in teaching? (*keep the focus on the three main types: institutional, discipline and publisher*)

23. Have you deposited your own research publications in an output repository? Which one?

24. What routes do you normally take into the contents of output repositories?

25. When you are using an output repository, what level of search do you normally find provides you with satisfactory results? For example, simple, advanced, Boolean logic, subject keywords, geographical interface?

Support

26. Do you find yourself requiring advice in using output repositories, either from individuals or from online links or advice?

27. What form does this advice take? Is it effective? What could be improved?

Project Aims

5. Source repositories: if a standard feature was the ability to identify and link directly to the publications developed from these data, how advantageous would you find it?

6. Similarly, if you could navigate directly from within an online article or text to the source data from which it was derived, how advantageous would you find it?

Why in particular would such links be advantageous?

29. One of the aims of this project is to develop effective links between repositories of source data to repositories of published papers because we believe there is a need amongst researchers to identify published (and pre-published) papers that have made use of their source data. In what way could you identify with that perceived need?

31. What functionality is missing from either the source or the output repositories you have used? Can you suggest any improvements?

32. We are considering building an interface for output repositories that would let you, as a depositor, associate newly deposited publications with the data from which they are derived. In what way might that be of benefit to you or, indeed, others? Genuinely important, or just a 'nice to have'?

33. A number of new operations could be supported within an output repository. For example:

- The automatic creation of links between related resources
- The automatic embedding of source repository data
- The presentation of relationships (i.e. showing publications and their source data in adjacent windows)

How do think these could meet your needs? What other features might you expect to be advantageous?

Reprise of project aims – potential solutions

34. One option we are considering for making two-way links between source and output repositories is the development of an online service known as a '**dataset knowledgebase**' – a central store which holds information about which source data relates to which publication or report. This might also have additional features such as quality assessments or ratings from users, frequently asked questions (FAQs) about specific sets etc. What is your opinion of the value of such a concept and are there specific issues you might want it to address?

Additional questions

36. How do you feel about having to pay for certain resources from repositories online? Do you feel this is justified?

The ADS

37. What are your views on the service provided by ADS?

Are there any other issues which are relevant and that you would like to discuss?

Appendix 2: Scenarios

Archaeology Scenario 1

Title	A researcher looks for data sets using online resources
Author	Dan Hull
Narrative	<p>A researcher wants to write an article on the way in which metal objects are deposited in Iron Age graves in the north of England. He therefore wants to find as many relevant data sets as possible to draw into his research. To begin with, he looks for all the relevant publications he can find on the subject using Web of Knowledge, searches in his University library catalogue, and through discussions with colleagues. Having found out from these sources the important excavation sites for his topic, he types the names of these sites into the ArchSearch facility at the Archaeology Data Service. This returns some useful hits, some of which provide access to the data sets themselves for him to look at, others just the details of the excavation project (leaving him to contact the researchers direct). For the ones with details, he downloads images of metal objects and inspects them, plus looks at a plot of where and how the finds were deposited. On the off chance that he's missed some important excavation sites, he clicks on specific counties in the interactive map function of ArchSearch, to see what data sets might be there.</p> <p>He knows that details about some Iron Age burial sites would also be in the local Historic Environment Record for each county, so he looks for a link from the ADS. Although some links are present, in other cases he has to travel directly to the HER in person.</p> <p>Having built up a reasonable body of data, he embarks on writing the article.</p>

Archaeology Scenario 2

Title	A researcher wants to deposit a recently generated data set in an online resources
Author	Dan Hull
Narrative	<p>A researcher has recently completed an excavation and survey project at an early Medieval settlement site in the south of England. She wants to deposit an archive of the project in an online resource, partly because of the potential for enhanced profile for her research, and partly because she wishes to publish an article about the work, and there are too many images and tables to include in the publication alone.</p> <p>After careful negotiation with her research collaborators, it is agreed which elements of the archive can be released for deposition in a repository, as long as this happens simultaneously with the publication of an article (to prevent other researchers publishing aspects of the archive before her). Having made this decision, she sets about deciding on what metadata should be provided alongside the data set. This seems a daunting task, but she telephones the Archaeology Data Service and receives some useful advice. She then sets about preparing the images, plots, databases of finds and a spreadsheet of stratigraphic contexts for submission. She spends some time reducing the findspot co-ordinates in the data set to 4-figure (rather than 6-figure) grid references, as some of her collaborators are worried about metal detectorists taking an interest in the site. She also makes sure that the terms used in the databases conform to commonly agreed thesauri, so that other researchers can easily search for what they need. The data set is then sent to the ADS. She knows that the article she has published on the site will go into the next</p>

edition of her chosen online journal in September, so she asks ADS not to release her data set publicly until then.

Both the ADS and the journal editors have agreed to provide an electronic link so that the article and the data set are mutually accessible.

Keen to disseminate her results as widely as possible, she then sends out the url to both the data set and the article to colleagues in her department and in her research field.