

From elements to elephants: a review of progress in providing online access to the United Kingdom's archival information

Amanda Hill, University of Manchester

In 1998 the National Council on Archives published a forward-looking paper called *Archives On-Line: the establishment of a United Kingdom archival network*.¹ This paper recommended "...the creation of a network that aims ultimately to provide access from a single gateway to *all* archival catalogues in the United Kingdom", hoping to "... avoid the problems of piecemeal investment in systems and practices which, in the event, may prove to be mutually incompatible".

Unfortunately, the single gateway pictured by this report did not immediately materialise. Instead, in the years following the publication of *Archives Online*, the archival community in the United Kingdom developed a number of collaborative projects which have all been putting the information held in paper catalogues online. National projects include A2A (for England),² the Scottish Archive Network (SCAN),³ Archives Network Wales (ANW)⁴ and the Archives Hub⁵ (covering records held in universities and colleges throughout the UK). Funding sources for these projects have been varied, reflecting the variety of institutions in which archive collections are to be found. There are also a number of collection-mapping projects, many of which originated in the higher education sector and were funded originally by Research Support Library Programme grants.⁶ Many of these latter projects have a subject focus, but the largest, AIM25,⁷ is geographically based, describing archives held in educational and learned institutions in the London area.

The impact of these different services has been significant, as the evaluation activities undertaken in the course of the last five years have shown.⁸ Usage is high, and the services have been successful in attracting users who have not previously used archive catalogues and record offices. Of those who filled in the 'new user' survey on the A2A site in 2000-2002, over 60% had not used archive catalogues before.⁹ The success of the individual services has, however, to be balanced against the frustrations of their users, who frequently complain that there are now simply too many such websites, with sometimes overlapping remits and content. This is also a problem for the custodians of the archives, who may find that there are several copies of their descriptions available in a variety of online services. This increases their exposure, but can cause headaches when the descriptions need to be updated or deleted.

¹ *Archives On-Line: the establishment of a United Kingdom archival network*, National Council on Archives, 1998, available online at <http://www.ncaonline.org.uk/materials/archivesonline.pdf>

² <http://www.nationalarchives.gov.uk/a2a/>

³ <http://www.scan.org.uk/>

⁴ <http://www.archivenetworkwales.info/>

⁵ The Archives Hub is funded by the Joint Information Systems Committee (JISC) and is available at <http://www.archiveshub.ac.uk/>

⁶ <http://www.rslp.ac.uk/projects/>

⁷ <http://www.aim25.ac.uk/>

⁸ Several of these are available from <http://www.archiveshub.ac.uk/introduction.shtml>

⁹ *A2A: Access to Archives Report, April 2000 – March 2002*, A2A Central Team, E-Access Department, Public Record Office, 2002

Plans for providing an overarching search mechanism for the principal online archival resources have been developed by The National Archives and a wide range of partners as part of the 'Linking Arms' programme.¹⁰ This plan reflects the conclusions of the Government-commissioned Archives Task Force, which had the creation of an Archives Gateway as its principle recommendation.¹¹ There is widespread support for the concept of the Gateway amongst archivists, users and potential users of archives, but securing a financial commitment to develop and maintain such a service is proving more difficult.

The effectiveness of an Archives Gateway will depend to a large extent on the quality and comprehensiveness of the finding aids which are made available for searching. Standardisation is crucially important in the world of union catalogues, and the rapidity of the adoption by the UK archival community of the General International Standard for Archival Description (ISAD(G))¹² has been notable. Indeed, conformance to ISAD(G) was the first requirement of National Council on Archives' *Interoperability Protocol*, published in November 2003.¹³ This document was drawn up with the future Archives Gateway in mind:

The aim of this protocol is to facilitate interoperability between these strands of the archival network in the UK. The protocol seeks to provide a standard of minimum conformity and full compliance, covering data structure, content and technical matters, to which existing strands of the current archival network have committed themselves and which newly-developed resources should adopt.

Interoperability Protocol

Other standards covered by the protocol include those for data exchange formats, subject indexing and name authorities. Encoded Archival Description (EAD), has become the accepted XML standard for encoding the elements of electronic finding aids in recent years and is now widely used all over the world for the exchange of archival data. Proprietary database systems are widely used in the UK, and these are increasingly offering EAD export and import facilities, in response to demand from their users.

Subject indexing has been another area where the archival community has seen rapid change over the last five years. Archivists in the UK have generally not been trained in subject indexing as part of their postgraduate training, and yet subject access to archival resources is vitally important in an online context, where there are no friendly staff members to guide researchers to collections which match their area of interest. The UK Archival Thesaurus project (UKAT) was developed during 2003 and 2004 to meet a need for a thesaurus which met the particular requirements of archivists in this

¹⁰ More information on Linking Arms can be found at <http://www.nationalarchives.gov.uk/partnerprojects/linkingarms/>

¹¹ *Listening to the Past, Speaking to the Future*, Report of the Archives Task Force, MLA 2004, available online at <http://www.mla.gov.uk/action/archives/00archives.asp>

¹² *ISAD(G) : general international standard archival description*, International Council on Archives, Committee on Descriptive Standards, 2000, available online at <http://www.icacds.org.uk/eng/standards.htm>

¹³ The Interoperability Protocol can be found at <http://www.ncaonline.org.uk/materials/interoperabilityprotocol.pdf>

country.¹⁴ The full thesaurus is now available for download from the UKAT project's website in SKOS-Core RDF format.¹⁵ It should be noted, however, that it is unlikely that every archive repository in the country will adopt UKAT for subject indexing: a significant number of university archives, for example, use the Library of Congress Subject Headings, often in line with the practice of their university library. It might be that technology will provide the solution to the use of multiple thesauri, as it does with the 'Subject Finder' facility of the Archives Hub.¹⁶ The results of a search in the Subject Finder on the word 'textiles' are shown in Figure 1 [subjectfinder.jpg], illustrating how terms from different thesauri are returned together, to give the user the best chance of finding appropriate collections.

Despite overall adherence to the structure of ISAD(G), there are significant differences between the data of the existing online catalogues which will have an impact on the experience of the users of the Archives Gateway. Most significant, perhaps, is the proportion of full finding aids online, compared to summary descriptions of archival collections. In February 2005 the numbers of finding aids in those two categories within the main archive network strands looked like this:

| Resource | Multi-level finding aids | Collection-level descriptions |
|----------------------------------|---------------------------------|--------------------------------------|
| A2A | 57,000 | 30,000 |
| AIM25 | | 6,000 |
| Archives Hub | 50 | 18,800 |
| Archives Network for Wales (ANW) | | 5,100 |
| Scottish Archive Network (SCAN) | | 20,000 |
| Total | 57,050 | 79,900 |

Table 1: Number of full and collection-level finding aids in UK online archive networks

As this table shows, more than half of the 134,850 finding aids currently available through the major archive network websites are at collection-level only. Although some of these descriptions are very full, they obviously cannot contain as much detailed information as a full archive catalogue. Research carried out for the LEADERS project, based at University College London, showed that 76% of users of archive services are looking for names of individuals, families, organisations or places.¹⁷ The needs of these users are not being fully met by the current resources.

The other major group of users, forming 19% of the LEADERS project's respondents, are searching for records relating to particular topics. This group is overwhelmingly comprised of professional and academic researchers, whose needs are best met by providing information about subject strengths in collections. Unfortunately, subject indexing is extremely variable across the existing archive networks: some descriptions

¹⁴ <http://www.ukat.org.uk/>

¹⁵ The SKOS-Core guide is available at <http://www.w3.org/2001/sw/Europe/reports/thes/1.0/guide/>

¹⁶ The subject finder can be found at <http://www.archiveshub.ac.uk/searchsub.shtml>

¹⁷ Anna Sexton, Chris Turner, Geoffrey Yeo and Susan Hockey, 'Understanding Users: a prerequisite for developing new technologies' in the *Journal of the Society of Archivists*, Apr 2004

have no subject index terms attached, others very few, as the table below demonstrates.

| Resource | Finding aids | Subject index terms | Terms per finding aid |
|----------------------------------|---------------------|----------------------------|------------------------------|
| A2A | 87,000 | 26,744 | 0.3 |
| AIM25 | 6,000 | 65,867 | 11 |
| Archives Hub | 18,850 | 42,711 | 2.26 |
| Archives Network for Wales (ANW) | 5,100 | 11,168 | 2.1 |
| Scottish Archive Network (SCAN) | 20,000 | 0 | 0 |

Table 2: Number of subject index terms in UK online archive networks

Providing a subject index search across these resources will not produce reliable results, meaning that only simple keyword searches are planned for the first stage of the Archives Gateway. In the future, however, it may be possible to retrieve unindexed records in response to a subject search, by comparing keywords associated with a term from those records which are indexed, and then feeding those keywords back as search terms to the records which do not contain subject index terms.

Further funding may well be required to bring our existing online finding aids up to the standards required for a truly useful Archives Gateway, although improvements to the quality of electronic catalogues are already being carried out by the repositories who own them. One important step which should be taken now to make the Gateway more viable is to regularise the way in which the archive networks are populated with data. The best solution would be for each archive repository to own and maintain one 'master description' for each of its finding aids. Providers of archive networks then need to put appropriate procedures in place for regularly updating their copy of the finding aids from their contributing repositories, to ensure that their content does not become increasingly out of date. This will also guarantee that any improvements to catalogues which are made by the holding repositories will be mirrored in the networked version: this is often not the case in the current network strands. It should also be possible to make the content exposed in this way available to other resources, such as regional or subject-based gateways or portals.

Duplication of content will be a real problem in the Archives Gateway. Some catalogues are already available in more than one of the archive networks, often in slightly different forms. Data originally created for the Archives Hub, for example, has been provided to a number of other projects. Once these are drawn together into a combined gateway it will be extremely difficult for users to know which of the duplicate descriptions is the most current or comprehensive. This will be less of a problem if all the network services are able to draw their information from a master copy in the manner described above, as exact duplicates would then be easy to identify and to hide from display.

The Archives Hub is attempting to address some of these issues with its move towards a distributed model, where repositories are able to host their own EAD descriptions

within a local version of the Archives Hub software (known as a Spoke), while still making them available to the main Archives Hub service, for users to search. The Hub will gather information from the Spoke each night, so that its data is never more than 24 hours out of date.¹⁸ In this way, responsibility for maintenance of the electronic finding aids rests firmly with the repository, which also benefits by having web and Z39.50 access to its own data. It will also be possible to provide access via SRW (search/retrieve web service) in the next version of the software.¹⁹

Uncatalogued collections form another problem area in relation to the creation of an Archives Gateway. The Operation Logjam project, carried out in the north-west region of England during 2003, surveyed 30 of the area's main archive-holding organisations. The project team found that, on average, 29% of archive collections held by repositories in the region are unavailable for research because they have not been catalogued.²⁰ It is likely that a similar proportion of uncatalogued material is to be found (or rather, not to be found) in the rest of the UK. The Logjam report estimated that it would take 65 archivist years to catalogue just those collections identified as of highest priority. The estimate for eliminating the entire cataloguing backlog in that one region was 299 years of professional work. For an archival gateway to be truly representative of the nation's archives, work on prioritising the cataloguing of these backlogs and making the results available online is crucial.

Perhaps all this is yawningly obvious for DigiCULT readers, particularly to those who have had to deal with combining library catalogues, but for many archivists the issues surrounding standardisation, subject indexing and union catalogues are quite new. The EAD/Data Exchange group of the Society of Archivists works hard to raise awareness of these issues within the profession, running training courses for newly-qualified archivists and providing information to other resources on its webpages.²¹

The development of standards for the description of content and for the exchange of archival data has been essential in the journey towards the creation of national and international archives networks. Much progress has been made in the UK towards bringing information about archives online, but there is still a lot to do before the 1998 vision of *Archives Online* is realised. It is in the provision of training, tools, procedures and support for staff in the UK's archive repositories that the biggest challenges for the domain still remain. Putting these key elements in place will help to ensure that the resources that are built from them in the future will be vibrant, living stores of the nation's memories: big grey elephants, rather than expensively-produced white ones!

¹⁸ More information about the Spokes software is available from <http://www.archiveshub.ac.uk/arch/spokes.shtml>

¹⁹ The SRW maintenance agency is the Library of Congress : <http://www.loc.gov/z3950/agency/zing/srw/>

²⁰ Janice Tullock and Alexandra Cave, *Logjam: an audit of uncatalogued collections in the North West*, North West Museums, Libraries and Archives Council, 2004. Available online at <http://www.nwmlac.org.uk/News/PDFs/Logjam1.pdf>

²¹ <http://www.archives.org.uk/groups/eaddataexchange/usefullinks.html>