1 Background

Archives On-Line and its impact

In 1998 the National Council on Archives published *Archives On-Line: the establishment of a UK archival network*, which set out a vision of a single online point of access to all archival catalogues in the UK.

The report stated that “...the growing tendency of educational users at all levels from primary schools to higher education to regard the Internet as a preferred tool for resource discovery makes it increasingly vital that detailed information about archival holdings should be accessible in this way; a need that will be felt the more strongly as libraries place the Internet at the disposal of every member of the public.” The report also noted that it was important that the “National Archives Network” should capture as much data as possible: critical mass was required to ensure it became *the* tool of resource discovery, not just one of many.

The objective of the report was widely welcomed by the professional community, although the report’s preferred model for delivering this service (an independent managed network) was quickly abandoned in favour of an Internet-based service.

One of the difficulties foreseen in *Archives On-Line* was that the diverse funding arrangements of the different constituencies within the archives domain would make the organisation and funding of a single national initiative to achieve this goal difficult. The report therefore recommended that a series of independent "implementation projects" should be developed that could be funded in different ways; but that these projects should work within a common framework of principles that would mean they had enough in common to be joined up at a later date.

Publication of the report was quickly followed by the announcement of a series of such implementation projects. These were:

a) *Access to Archives* (A2A) – a collaborative programme catering mainly but not exclusively for archives of local repositories in England, co-ordinated by the Public Record Office (PRO), and funded to date mainly by the Treasury Invest-to-Save budget, the Heritage Lottery Fund (HLF) and the PRO;

b) *The Archives Hub* – a collaborative programme for Higher Education Institutions in the UK, funded by the Joint Information Systems Committee (JISC) of the Higher Education Funding Councils;

c) *Scottish Archives Network* (SCAN) – a collaborative programme for archive services in Scotland, co-ordinated by the National Archives of Scotland and funded by the HLF;
d) **Archives Network Wales (WAN)** – a collaborative programme for archive services in Wales, co-ordinated by the Archives Council Wales and the National Library of Wales, and funded by the HLF;

e) Historical Manuscripts Commission’s **National Register of Archives** online indexes and **National Name Authority Files**;

f) Several smaller projects, of which the most important are **AIM 25** and **GASHE** (*Gateway to Archives of Scottish Higher Education*) funded by the JISC and designed to capture data about the archives of particular types of institution, mainly in further and higher education;

g) Retroconversion projects for a single institution, several of which were already underway or completed before the NCA report was published, including those of the British Library, Public Record Office, National Library of Scotland, Public Record Office of Northern Ireland, Essex Record Office, Surrey Record Office etc.

(Note on terminology: **Retroconversion** is the umbrella term used to describe the conversion of existing paper catalogues into machine readable form through scanning, marking up and re-keying. **Retrocataloguing** is the term used to describe the new electronic cataloguing of items for which there is no existing catalogue; **Digitisation** usually refers to the electronic scanning or reproduction of original archival material).

Representatives of these projects have continued to meet at irregular intervals to discuss standards for minimum conformity and other matters affecting compatibility. They have also conducted a user evaluation study of the major projects, funded by Resource. Most recently, the projects have agreed that the time has come to consider how these initiatives (some of which are approaching the end of their funding) can best be linked up to create the one-stop-shop for archival catalogues which **Archives On-Line** envisaged five years ago.

**Users/New users**

Existing strands of the National Archival Network have proved very popular with users. The analysis of statistics for the use of A2A shows that 1.3 million searches of the system have been made since May 2001 and in that same period 2.7 million catalogue pages were viewed. Over 100,000 searches and 200,000 catalogue requests are consistently made every month. The PRO’s 1901 Census proved phenomenally popular – over 30 million users tried to access the site on its launch day.

These strands are drawing in new users of archive material. Currently, 65% of first time visitors to the A2A site have never searched an archival catalogue (in any format) before. 85% are interested in researching their family history. 7.7% of users consider themselves non-white in comparison with 2% of visitors to local authority record offices.
2 What remains to be done?

Responding to what users want; evaluating user and non-user needs

In 2002, the National Archive Network Users Group (representing the main projects outlined above) undertook a small user evaluation study of the extant network projects. The study concluded that the projects appealed to a diverse audience and offered great potential for further development. All users experienced some problems with navigation and expected development to concentrate on improving useability, searching, navigation and imaging. Users are looking for a common gateway which offers layered and tailored interfaces. Young users, in particular, expected to see a greater emphasis on design and interaction.

The challenges ahead

The implementation projects have made an excellent start with the process of creating the National Archives Network, but there is a great deal still to do.

a) Retrospective conversion. The foundation of the project is the capture of catalogue data in electronic form. The implementation projects have made different levels of progress with this task, and it is hard to estimate what percentage of the total volume of lists has now been converted. The best estimate is that there are about 4,500,000 pages of catalogues in the UK in total. A2A will have captured over 700,000 pages by the end of its current (second phase); perhaps 40% of the pages in its constituency?

b) Collection level or multi-level description? One of the key ways in which the various implementation projects vary is whether they present the full catalogue of an archive, or just a ‘collection level description’ which includes the introductory and contextual information and a summary of the content. A2A provides access to full catalogues; the Archives Hub a mixture of the two but mainly collection-level descriptions; the Scottish and Welsh projects collection-level only (though Scotland is considering multi-level description for a successor project). Because it is a feature of archives that they often include unpredictable stray documents, which a user would not look for in the collection, collection-level descriptions do not reveal the full variety and utility of the collections they describe.

c) Clean data or dirty data? The original Archives On-Line report envisaged that alongside the retroconversion process would run a programme to improve old catalogues which had not been written with modern data structure standards and content expectations in mind. With this in mind, Archives On-Line identified the total cost of retroconversion as about £33-£38.5 millions. There has been a divergence between the projects about their approach to this issue. A2A has sought to capture the biggest possible volume of data, and has eschewed all data cleaning processes to minimise the cost per page of retroconversion. The projects that have undertaken collection level description have typically involved much more catalogue improvement. Indeed, in some cases, the existing listings have lacked much of the information typically included in a collection level statement, and these have been generated from scratch. The cost per page of the Archives Hub in particular has been much higher than that of A2A for this reason.
d) **Bringing data sets together.** Clearly there will be many technical issues involved in making a single service out of the existing projects, but inconsistency in the quality and level of detail of the catalogue data included will need to be explained and if possible addressed.

e) **Central or dispersed datasets.** A key issue to be resolved in creating the “National Archives Network” will be whether the data from the implementation projects is brought together on a single massive server, or whether it continues to be held in a dispersed way, as at present. A central server makes the searching of the data quicker for users and technically easier to accomplish; but the varying nature and quality of the data may cause technical problems. Dispersed data avoids the need to overcome these technical problems, and enables projects to retain control over their own information, but requires searches to run in a more complex and slower way. Currently, professional opinion favours dispersed data storage.

f) **One service or many?** Should there be a single point of access to all this data, however it is stored? Do different audiences value different presentational styles, through a variety of websites from which the linked datasets can be searched? The balance of professional opinion seems to favour a plurality of approaches, enabling customers to have more choice; but this needs to be tested; do users in fact want a single gateway – a national archives portal through which this plurality of multilayered access routes can be delivered?

g) **Sustainability.** Each of the implementation projects has different arrangements for the long-term sustainability of their dataset. In most cases, this is dependent upon a national or academic institution supporting the service after the cessation of project funding. There are, however, no copper-bottomed guarantees that sites will be actively maintained and further developed in the long term, if additional funding is not secured. Given the public investment that has been made in (a) creating the catalogues in the first place and (b) providing Internet access to them, we need to identify the funding for the indefinite sustainability for this service.

h) **Future cataloguing.** Although our inheritance of existing catalogues is very large, new collections continue to flow into archives all the time and most have large backlogs. Collections continue to be catalogued as fast as resources allow. Where services are generating new catalogues in electronic form, it may be relatively simple to add these to electronic networks. We need mechanisms to ensure that this happens, however, and we also need ways of continuing capture data from institutions that are still only able to supply hard-copy catalogues.

i) **Digisation.** In the English local authority sector around 50 of the 130 archive services (37%) are currently undertaking or planning programmes of digitisation\(^1\); however over half of such offices report that their overall capacity to digitise their holdings is poor. Although undoubtedly all archive services have recognised the tremendous potential, there are considerable barriers preventing services from developing digital resources, including funding problems, lack of core skills, staff time and satisfactory technological infrastructures. This is particularly the case with

\(^1\) Our shared past: Developing 21st Century Archive Services – Phase Two, Archival Mapping Project Board (Public Record Office), 2001
smaller archive services. Many archive services have been involved in the 152 People’s Network Content creation strand digitisation projects (funded through the New Opportunities Fund), and as with A2A, the establishment of regional partnerships is a useful development in its own right in addition to the resources which this one-off programme has released.

j) **Cross-domain issues.** The creation of Resource raises obvious questions about the desire of users to search for information on particular topics across different formats – books, periodicals, archives, works of art, and museum objects. In 1998 retroconversion of archival catalogues seemed a large enough task to contemplate, without taking on the additional technical complexity of cross-domain searching. Now, however, some University Libraries are building systems that cross-search between library and archival catalogues, and there are commercial products in development that are intended for applications such as this. Moreover, the Full Disclosure initiative has given the elimination of cataloguing backlogs and the online accessibility of collections information a higher priority in the Library and Museums domains, as well as in Archives. It may be timely to reconsider the role of cross-domain searching, without losing focus on the task of completing the retroconversion task.

k) **“What the users really want is ‘the stuff’!”** Archives On-Line took the view that the resources would never be available for the comprehensive digitisation of all the UK’s archival resources, and only through the catalogues could one demonstrate the breadth and value of archives to the online community. It cannot be denied, however, that having that found something of interest exists from a catalogue site, the first thing users want to do is to be able to see it. There thus needs to be a close relationship between the thinking behind the National Archives Network project and the national strategy for digitisation.

**What can the ATF do for the electronic National Archives Network?**

The vision of a National Archives Network is one of the archives domain’s success stories of the last five years. Enormous strides have been made towards translating into reality what was then just a theoretical possibility. Had the governmental and funding structures of the UK been more supportive of grand projets, it might have been possible to approach the initiative in a more effective way, but the archives domain was in no position to break a cast of thinking which is so long and firmly established.

The ATF needs to build upon what has been achieved, and the commitment from institutions and funders which has been secured to date. It would be helpful if it could recommend:

* The establishment of an agency to direct and co-ordinate the future development of the network; to directly manage key components of it (perhaps built on A2A) and to actively sustain it for the future. This could be an agency will responsibility only for the NAN, or it could have broader responsibilities for promoting the Full Disclosure project and cross-domain searching. It will need to be representative of the professional and user communities served, and of the funders who have supported the creation of the network. In either case it will need sustainable (not project) funding. It needs to be recognised that there are new costs...
for the archives domain in sustaining an online service, even if the development costs continue to be met from project funding. These are part of the costs of the Government’s e-government agenda, and could come most fairly on a pro-rata basis from the appropriate parent department of state (LCD, DFES, DTLR, DCMS and the Scottish, Welsh and Northern Irish executives). This inter-departmental initiative could perhaps be appropriately ‘owned’ by IDAC.

- **Establish and reinforce the principle of maximum quality and comprehensiveness in the content of the National Archives Network.** Collection level description and ‘dirty data’ are a start, but the project needs to continue to achieve the full presentation of multi-level catalogues of a consistent standard. The continuum expressed by *Full Disclosure* between retroconversion and retrocataloguing needs to be acknowledged, and the need for the elimination of cataloguing backlogs to follow (or run in parallel with) retroconversion to be urged on funders.

- **Establish and support the principle of a managed and co-ordinated approach to future digitisation projects**, which would enable a consistent pattern of linkages to be constructed between the NAN and the related images. (Digitisation probably needs a separate paper to justify this.)

- **Access** The creation of the Archives Task Force gives us a clear opportunity to address the issue of the creation of a single search mechanism (point f) above). The Historical Manuscripts Commission’s ARCHON Directory and Portal go a long way to providing contact details for record repositories in the United Kingdom and also for institutions elsewhere in the world which have substantial collections of manuscripts and providing a reference point for information about archival resources and projects that are planned, ongoing or completed.

Is there in fact a clear mandate for the creation of a one-stop shop **national archives discovery portal** through which a plurality of multilayered access routes can be delivered? What structures need to be in place? Who is currently working on what and how might the existing initiatives be drawn together? What targets might be set on such a goal?

Is our ultimate vision of flow of archival information (catalogues, images, digitised documents, digital records) which takes account of all the opportunities offered by digital networks, offering opportunities for exploration – historical, personal, social, corporate – to the broadest possible range of people wherever they can use it – the home, the classroom, the office, delivered through single gateway of discovery?

---

**National Council on Archives (Nick Kingsley)/Resource**

January 2003

---

2 Adapted from *British Archives – The Way Forward*, National Council on Archives, 2000, p7