

NEWS

NDAD Newsletter #6 - September 1999

A new newsletter design: news is on this page, with links to the usual selection of interesting articles on data archives and related topics. Please send any comments to r.davis@ulcc.ac.uk.

- **News from Guilford Street**, including:
 - [System & Project News](#)
- [Conferences and Seminars](#)
- [Staff News](#)
- [Access to Archives](#) - A report by Peter Garrod on the *Access to Archives* workshop, held at the PRO in April.
- [The Great Britain Historical Geographical Information System](#) - A report by Patricia Sleeman.
- [Managing the corporate memory](#) - Report by Richard Davis of a recent seminar at the PRO
- [Personnel Profile](#) - Sally Hughes, NDAD Data & Applications Specialist

Project News

New Datasets Online

Three new datasets have been available online since April. These are:

Department of Health:

[Anatomy Dataset](#) (CRDA/21)

This dataset incorporates data gathered between 1992 and 1995 under the provisions of the 1984 Anatomy Act and the anatomy regulations of 1988 for the acceptance and disposal of bodies donated for anatomical research.

Ministry of Agriculture, Fisheries and Food:

[Coastal Protection Survey of England](#)

Details of England's flood and coastal defences from surveys undertaken between 1993 and 1997. It is a natural counterpart to the [Welsh Office Coastal Survey](#) (CRDA/6).

Welsh Office:

[Survey of Contaminated Land in Wales](#)

Information about sites in Wales assessed as 'contaminated' (with toxic or potentially toxic materials) during a survey of 1987/88.

New User Registration Procedure

Users can now register online and gain instant access to NDAD datasets. Users will be automatically allocated a User ID and a password, and their accounts will be active immediately. Facilities for users to change their password and update the contact information they have supplied have also been enhanced. For more information see the [Registration](#) section of the website.


First Annual Review

The first NDAD Annual Review Meeting (incorporating the Quarterly Review Meeting) was held on 24 March 1999. The meeting was chaired by Andrew McDonald, Head of the Records Management Department at the PRO. Richard Blake was formally introduced as the new contract manager for NDAD (he had previously been involved with the NDAD project as deputy contract manager. Farewells and thanks were tendered to NDAD's first contract manager, Susan Healy, who was reluctantly (she said!) moving on to new challenges within the PRO. The meeting concentrated on reviewing the progress achieved and problems and issues encountered: the PRO congratulated and thanked ULCC for its achievements with NDAD over the past year, and looked forward with optimism to the coming year.

Amongst the possible improvements discussed were enhancements to the search/data analysis facilities, and to the user registration system [*see above*].

Conferences, Seminars, Visits

Visit by Dutch National Archives

 We were happy to welcome Hans Hofman and his delegation from the [Dutch National Archives](#), who visited us as part of their trip to the PRO to gather information about approaches to electronic archives. Kevin and Ruth made a presentation about the archival and technical aspects of NDAD, which was greatly appreciated, as was the Friesian liqueur that the visitors presented to us as a token of thanks.

Informasjon 99 in Norway

Kevin Ashley was invited to talk on database preservation and access at 'Informasjon 99', a conference held in Stavanger, Norway which dealt with a wide range of topics to do with preservation, dissemination and understanding of information. Contributions came from archivists, records managers, librarians, information scientists, telecommunications specialists, data librarians and data archives.



The preservation of large data collections is of particular interest in Stavanger, it being home to a large number of oil companies all of whom have massive investments in Petabytes of data from seismic surveys (a Petabyte is a million Gigabytes). Due to the fact that the rest of the presentations were in Norwegian, Kevin finds it difficult to describe them in any detail.

Training Day



On June 30th, staff from NDAD and the PRO's EROS team presented at a joint seminar to the Railway Heritage Committee. This unique body, [charged by statute](#) with preserving Britain's unique railway heritage in all its forms, was [established in 1994](#). It is having to develop policy to deal with the increasing amount of railway records which exist only in digital form, and the seminar was designed to inform the committee of current work in this area, both in preservation of materials and public access to them.

Conferences

Assistant Archivist **Patricia Sleeman** has recently returned from the annual meeting of the [Society of American Archivists](#), in Pittsburgh, 24th-29th August. She attended a workshop on managing electronic records and took part in many discussions and presentations on archival practice and developments. We hope she will describe the event in more detail in the next newsletter.

Papers about NDAD and associated issues of electronic archives will be presented at the following conferences during September and October:

- **Society of Archivists (Irish Region)**

Trinity College, Dublin: 7th - 10th September 1999

Covers issues of relevance to conservators, records managers and archivists, developments within archival professions and expectations for the future. **Kevin Ashley** will present the paper *Preserving the History of Government Computing: Social and Technological Change*.

- **DRH99**

King's College, London: 12th - 15th September 1999

Digital Resources in the Humanities is a major forum for the creators, users, distributors and custodians of digital resources in the Humanities. **Kevin Ashley** will present a paper provisionally titled *Database Preservation: Reconciling access and authenticity*. Read the [account](#) of our visit to last year's DRH conference in Glasgow in NDAD News #4.

- **DROs Conference**

Carden Park, Cheshire: 6th-8th October 1999

Annual conference for UK Government Departmental Records Officers organised by the Public Record Office. NDAD staff will be on-hand to demonstrate the system and discuss matters pertaining to the appraisal, transfer and preservation of datasets.

- **DLM Forum '99**

EC Conference Centre Charlemagne, Brussels: 18th - 19th October 1999

A multidisciplinary forum organised by the European Commission, inviting representatives from government, archive services, research and industry to discuss the problems of management, storage, conservation and retrieval of machine-readable data (*données lisibles par machine*).

↑ Staff

Departures

In recent months we have had the sad task of saying farewell to two respected colleagues, Ruth Vyse and Fiona Latham.

As University of London Archivist, **Ruth Vyse's** contribution to the project has been invaluable. Kevin Ashley writes:

Ruth's involvement with NDAD dates from the early days of the tendering process in 1996, and its sound archival base owes a great deal to her work and understanding of the issues. She played a key role in every aspect of making the service a reality, from archival theory to publicity material, recruitment of the NDAD team to dealing with depositors. When the work has been hard, Ruth's positive demeanour has never failed to lift our spirits. We will miss her greatly, but she has left a sound foundation for us to build on.

Personal commitments have taken Ruth to Walsall, and we're sure the archive service there is already benefiting from her considerable talents.

Fiona Latham joined NDAD in December 1997 as Senior Archives Assistant. Her work included handling user registrations and support, as well as cataloguing and scanning, and her adeptness at art and design meant she was much in demand for preparing displays and publicity. She has gone to seek out new life and new civilizations in Australia, where she is no doubt having a wonderful time

in the sun!

Profiles of both [Ruth](#) and [Fiona](#) have featured in previous newsletters.

Arrivals

Kate Bradford Kate Bradford is the new Archives Assistant on the NDAD team. She has a degree in Conservation of Library and Archive Materials from Camberwell College of Art, and has previously worked in the Archives Department of the Institution of Mechanical Engineers. Her work included cataloguing, preparing exhibitions and promotional material which she looks forward to being involved in at NDAD.

Charlotte Smith, a student on the MA Archives and Records Management course at UCL, recently completed a six week research project funded by the Arts and Humanities Research Board. The project evaluated ISAD-G as a framework for the archival description of electronic datasets. The resulting report will form part of Charlotte's MA, and also be the basis for a set of guidelines which will be sent to the International Council on Archives, and an academic research article to be submitted to the Journal of the Society of Archivists, Archivaria or American Archivists. In September Charlotte will start work in the Archives of the National Gallery.

Access to Archives

Peter Garrod, NDAD Assistant Archivist

Contents

1. [Introduction](#)
2. [Minimum Data Standards](#)
3. [Implications for NDAD](#)
4. [Postscript](#)
5. [Notes](#)

↑ 1. Introduction

I attended a workshop, "Access to Archives", which was held at the [Public Record Office](#) (PRO) on 29 April 1999 to discuss some of the issues surrounding the creation of a UK National Archival Network. It followed on from a report authored by Nick Kingsley of Birmingham City Archives, [Archives On-Line: The Establishment of a United Kingdom Archival Network](#), which was published last year by the National Council on Archives (NCA).

The Kingsley report proposed that archivists should take advantage of the Internet to create a network linking information about records held at archive repositories throughout the UK. Data taken from repositories' catalogues of their collections would be mounted on regional or national servers, and linked to a common access gateway allowing searches across the network. An interlinked National Name Authority File would be created to allow for the consistent searching of information linked to the names of persons, families, places and corporate bodies.¹ The function of the network would be to link metadata about archives, not digitised images of documents: in the words of the report, "the sheer scale of the UK's archival holdings makes any scheme for their *wholesale* digitisation impractical in the context of existing technology".²

The main obstacles to the creation of a National Archival Network are:

1. The need to raise the necessary funds (Kingsley estimated that the infrastructure costs alone of establishing the network would be at least £7.6 million).³
2. The need to retrospectively convert archival catalogues- many of them paper based- to digital form.
3. The need to ensure that data submitted to the network conforms to nationally and internationally agreed standards, in particular the [General International Standard Archival Description](#) and the [International Standard Archival Authority Record for Corporate Bodies, Persons and Families](#) (both issued by the [International Council on Archives](#)), and [Rules for the Construction of Personal, Place and Corporate Names](#) (issued by the NCA).
4. The need to decide on a common format for making archival data available on the Internet. Kingsley highlighted Encoded Archival Description (EAD) as the most likely format (for further information about EAD, see [Patricia Sleeman's article](#) in NDAD News #5).⁴

Links in this Article

-  ["Establishing a UK Archival Network"](#)
-  [International Council on Archives](#)
 -  [ISAD\(G\)](#)
 -  [ISAAR\(CPF\)](#)
-  [NCA Rules](#)
-  [JISC](#)
-  [PROCAT](#)
-  [Royal Commission on Historical Manuscripts](#)
-  [ARCHON](#)
-  [NDAD: Thesaurus](#)
-  [NDAD: Coast Protection Survey of England](#)

Since the publication of the Kingsley report the Heritage Lottery Fund (HLF) has agreed to provide £2.9 million over 3 years to fund a Scottish Archival Network (SCAN) linking "top level" catalogue information from participating Scottish repositories.⁵ At the "Access to Archives" workshop it was revealed that a bid to establish a Welsh Archival Network (WAN) had also been submitted to the HLF. We were told that these are "First Stage Implementation Projects" (FISPs) to create a National Archival Network, with "Access to Archives" (A2A) being the name of the proposed FISP covering repositories in England. An A2A Steering Group had been set up with the PRO in the lead to prepare a bid to the HLF for funding to create an English network. The project would include the retrospective conversion of catalogue data from participating repositories to digital formats compatible with national and international standards. The resulting database would be made available on the World Wide Web with EAD as the display format.⁶

The A2A workshop brought together representatives from around 90 English archive repositories to examine the issues surrounding A2A and gauge the willingness of repositories to participate in the proposed network.

Margaret Proctor (communications officer of the A2A Steering Group) explained that a modularised approach was being taken to the development of the National Archival Network. This involved SCAN, WAN, A2A, the HE Hub (a [JISC](#) funded initiative to make collection level descriptions of archives in the higher education sector available on the Internet), the creation of a National Name Authority File, and possibly other regional and sectoral initiatives which would come together in a National Archival Network. The A2A Steering Group wanted the network to be as inclusive as possible, although the need to submit a viable bid to the HLF might mean that the initial participation of repositories from some sectors (e.g. the business sector) would not be possible. One issue would be to identify "priority" archives for inclusion in the network. This would be based partly on the needs of users, and partly on the availability of catalogue information in formats which allow the easy export of data. **David Crook** explained that A2A's bid to the HLF for funding was being prepared by a task group (of which he was the convenor) which aimed to submit an initial bid in June 1999. This bid was being made in the name of the Public Record Office as the lead organisation, though he stressed that A2A should not be seen as a PRO project. The HLF should decide on the initial bid by the end of 1999; if it was accepted, it would be necessary to prepare a more detailed bid with the aim of final acceptance by the end of 2000. A not-for-profit company would probably be created to manage the network. As the bid would be for a sum of well over £100,000, HLF rules would require partnership funding of at least 25% from participating organisations. Some of this could be "in kind" contributions, e.g. voluntary labour. **Meg Sweet** of the PRO outlined the proposed minimum standards for submitting data to the A2A network (see Section 2, [Minimum Data Standards](#)), and explained that EAD was being recommended as the format in which data should be made available. Meg and **David Thomas** spoke about the PRO's experience of converting paper based finding aids to EAD as part of the [PROCAT](#) project to make the PRO's catalogues available online. Meg and **Bill Stocking** discussed other retroconversion initiatives related to PROCAT, including an initiative to incorporate data into PROCAT relating to public records held at repositories like the national museums and galleries, the Post Office and BT archives. The PRO is also working with a group of 6 local record offices holding deposited public records, to look at issues surrounding the conversion of catalogue information to EAD for online access.

Participants in the workshop were generally positive about the prospects for A2A. However, there was concern that archive repositories would find it difficult to locate partnership funding at a time when many are struggling to maintain existing services. It was agreed that repositories who were interested in participating in the network would feed their comments to the A2A Steering Group, and that the Steering Group would give an indication of the likely costs (in terms of partnership funding) to participating repositories by August/September 1999.

↑ 2. Minimum Data Standards

During the workshop **Meg Sweet** set out proposals for the minimum standards that data submitted to the network would have to meet. I found this the most interesting part of the workshop, partly because it had the greatest implications for NDAD (see Section 3, [Implications for NDAD](#)), and partly because it is one area where A2A might have a direct impact on the working practices of archivists. Archive services which want to participate in the network will need to ensure that their catalogues conform to the network's standards. This could potentially impose a degree of uniformity in archive cataloguing which so far has been lacking.

Meg explained that the aim was for catalogue data to conform to the [General International Standard Archival Description](#) (ISAD(G)), the [International Standard Archival Authority Record for Corporate Bodies, Persons and Families](#) (ISAAR(CPF)), and the NCA's [Rules for the Construction of Personal, Place and Corporate Names](#). Data would have to follow the principles of "multilevel description" set out in ISAD(G): i.e. catalogues should proceed from the general to specific and should provide information which is appropriate to the level of description; each level of description should be linked to the level above it; and information given at a higher level of description should not be repeated at lower levels. ISAD(G)'s terminology for naming levels of description should also be followed. These levels range from the **fonds** (all of the records created or accumulated by a particular organisation, person or family) at the top, to the **item** (an individual document) at the bottom. Meg proposed that the highest level of description of any catalogue data submitted to the A2A network should include at least the following:⁷

- Reference code ☐ ("country code as specified in ISO 3166 followed by repository code as specified by HMC [the [Royal Commission on Historical Manuscripts](#)] followed by local repository specific code")
- Title ☐ ("a formal or given title that encompasses all the records being described")
- Creator(s) ("individual name(s) of the organisation(s) or individual(s) responsible for creating and accumulating the records being described")
- Abstract ("a brief summary of the scope and content of the material being described")
- Creation dates ☐ ("the date(s) of creation of the records being described")
- Extent and form ☐ ("the number of records being described" and their "broad physical form(s)")
- Access conditions ("any conditions that restrict or affect access to the records being described")
- Level of description ☐ ("the hierarchical level, in ISAD(G) terminology, of the material being described")

The elements marked ☐ are equivalent to the 5 elements of description which ISAD(G) considers to be "essential for international exchange of descriptive information".⁸ The other elements are equivalent to "optional" elements in ISAD(G) (i.e. elements of description which can be used if they are considered to be appropriate to the material being described and the level of description), except for the "Abstract" element. This is not found in ISAD(G), but submissions have been made to the International Council on Archives (ICA) that it should be included in the mandatory ISAD(G) elements, together with the other "optional" elements listed by Meg. The "Abstract" element corresponds to an <abstract> tag which was recently added to EAD (see [Patricia Sleeman's article](#) in NDAD News #5). Meg explained that a decision about the inclusion of this element was not expected until the relevant ICA committee met in October 1999. It had been felt that an "Abstract" element was needed because the "Title" element might not adequately reflect the subject matter of a document.

Meg also proposed that in any data submitted to the network, names of persons and corporate bodies should be formulated according to the NCA's [Rules for the Construction of Personal, Place and Corporate Names](#). NCA Rules should also be followed for corporate, personal and place names

where these are used as access points (i.e. for indexing purposes). Subject indexing should attempt to follow the terms used in the *UNESCO Thesaurus*.

3. Implications for NDAD

The workshop was thought provoking because NDAD would almost certainly want to participate in a National Archival Network if it becomes a reality. We are in a better position than many repositories to submit data to the network:

- All of our finding aids are in electronic form and are already available on the World Wide Web. We do not have a legacy of paper catalogues which would need to be converted to digital form.
- Our catalogues have been structured using section and sub-section headings, most of which are mappable to ISAD(G)'s elements of description. We have also added elements not found in ISAD(G). This has been necessary because ISAD(G) is a general standard, and does not include all the elements that are needed to adequately describe electronic datasets.
- We have created an online subject index based on the *UNESCO Thesaurus* (see [Finding aids search - keyword selection](#) for further details).
- We are working to develop an index of corporate, personal and place names in accordance with the standards set down in the NCA's [Rules for the Construction of Personal, Place and Corporate Names](#) and the [International Standard Archival Authority Record for Corporate Bodies, Persons and Families](#).

However, if the National Archival Network does indeed develop as envisaged by the A2A Steering Group, we may need to modify our catalogues in order to submit data to the network. The most fundamental change would be to convert our catalogues from HTML to EAD, if the EAD format is adopted by the network. Regardless of the A2A project, conversion to EAD is something which we will need to examine as EAD is the emerging standard for Internet delivery of archival metadata (see [Patricia Sleeman's article](#) in NDAD News #5).

We may also need to incorporate an "Abstract" element and a "Creator" element into our top-level catalogues (i.e. our Series Catalogues). However, the "Abstract" element may not be necessary, as all of our Series Catalogues include a Scope and Content section which describes the subject matter of a series of datasets. The "Creator" element listed by Meg is the equivalent of ISAD(G)'s optional "Name of creator" element. We haven't employed this element of description in our catalogues, on the grounds that the creating body is identified elsewhere, specifically:

- In the Administrative Histories which are linked to our Series and Dataset catalogues
- In the "Administrative context" section of our Series Catalogues, which is mappable to ISAD(G)'s "Administrative/biographical history" element
- As the first component of the title at the head of each catalogue page, though in some cases this will be a generic title referring to a continuing administrative function (e.g. "Agricultural Departments") where we have adopted this as the overall title of the linked Administrative History.

"Abstract" and "Creator" elements, if required, could be added to our catalogues as part of a future conversion from HTML to EAD. The final change would be to modify the references which we assign to datasets and documents, to include a prefix consisting of the UK country code as specified in ISO 3166,² and the number assigned to NDAD in the Historical Manuscript Commission's [ARCHON](#) database. Our prefix would thus be GB 1874, which would be inserted before our existing references in data exported to a National Archival Network. Aside from the prefix, the remainder of our references would not change (e.g. the Series reference for the [Coast Protection Survey of England](#) datasets would become GB 1874 CRDA/10). Country + repository prefixes are a requirement of ISAD(G) and are something which we would want to consider as part of the long-term development of the NDAD, regardless of the future of A2A.

↑ 4. Postscript

An update on the A2A project by Margaret Proctor was published in the July 1999 issue of the *Society of Archivists Newsletter* (no. 122). It reports that in the wake of the workshop at the PRO and a similar event at Lancashire Record Office, the A2A Steering Group decided not to submit a large bid for a central project to the Heritage Lottery Fund (the original projection had been for a central entity costing £3.75 million over 3 years). Instead, a scaled down bid for £1.258 million to cover the staffing and infrastructure costs of a central project was submitted by the PRO at the end of May 1999 to the government's Invest to Save Budget. The editorial costs of conversion will now be covered by separate regional and local bids to the HLF. It is hoped that this will reduce the amount of partnership funding which participants in the network will need to raise. Even if the bid to the Invest to Save Budget is unsuccessful, the PRO will still launch a more limited archival network using data from its own AD2001 project, and data from the British Library, national museums and galleries and the PRO's Places of Deposit project.

Further details about the A2A project and updates on its progress can be found in the [Access to Archives Project](#) pages of the PRO's web site.

↑ 5. Notes

1. Nick Kingsley, [Archives On-Line: The Establishment of a United Kingdom Archival Network](#) (London: NCA, 1998), sections 5.5-5.14 and recommendation 7; sections 4.18-4.23 and recommendation 5.
2. *Ibid.*, section 2.12.
3. *Ibid.*, Executive Summary.
4. *Ibid.*, sections 4.1-4.6.
5. *Society of Archivists Newsletter*, January 1999 (no. 116), p. 5.
6. *Society of Archivists Newsletter*, May 1999 (no. 120), p. 1.
7. Quotations are taken from handouts provided by Meg Sweet at the A2A workshop.
8. International Council on Archives, [General International Standard Archival Description](#) (Ottawa, 1994), section I.5.
9. BS EN ISO 3166-1:1998, *Codes for the representation of names of countries and their subdivisions. Country codes*.

The Great Britain Historical Geographical Information System

Patricia Sleeman, NDAD Assistant Archivist






As part of the National Name Authority File Project a group of archivists was invited to a presentation on the Great Britain Historical Geographical Information System ([GBHGIS](#)) given by **Professor Humphrey Southall** and **Ian Gregory**. GBHGIS is part of the Great Britain Historical Database Project (GBGIS), located at Queen Mary and Westfield College London and Queen's University Belfast and which holds a large database of British nineteenth and twentieth century statistics. It is available from and operated by the History data service (HDS), part of the Arts and Humanities Data service ([AHDS](#)) based at Essex University and was launched in April 1998, emerging from the strong tradition of historical demography which exists at Cambridge, Essex and Edinburgh universities.

The Great Britain Historical GIS project uses GIS which is linked to their database and permits many of the statistics to be mapped at county or more local levels. The GBHGIS records changing boundaries, assembled using maps and textual sources or documents which record change. The system stores alterations in boundaries as well as dates of the change so that it can construct an accurate map of the district for any date. It can be used as a base map generator, creating maps for a wide range of dates, types of unit and parts of the country. Boundary data such as a base map of registration districts and pre-1911 civil parishes (by autumn 1999) are available to researchers from UKBORDERS, hosted by [EDINA](#) at the Edinburgh University Data Library. At present only Mark 1 data are available (see below). The project also disseminates GIS based tools and methodology among the communities of UK historical researchers.

The project developed out of a database created by Humphrey Southall onto which a mapping front-end i.e. GIS, was placed by Ian Gregory using the GIS package, ARC/INFO. The database contains over 100 tables and some 10 million records but is constantly growing as a result of the work of an input team in Belfast. As the GBHDB contains statistics which were recorded for large numbers of separate geographical units once or twice a year, a system had to be constructed which tracked these changes. To plot these changes the project relied on maps and various reports of changes in boundaries (see below). It aims to cover the whole of the British Isles but at the time of writing has complete coverage for parishes in England and Wales only. Further work remains to extend coverage for parishes in Scotland, Northern Ireland and the Irish Republic.

Ian ran us through the Mark 1 GIS, showing the Poor Law Unions and Registration Districts of England and Wales, c.1840-1910. The system built by the project is capable of producing accurate information coverages for any date to the exact day, for any county or collection of counties in England and Wales. It is also capable of linking this to any suitable attribute data set (census data, Poor Law statistics, data from the Registrar Generals Reports, etc.) held in a relational database management system (RDBMS) for analysis and mapping. The construction of this system consisted of gathering information on the date and effect of boundary changes from maps and official reports; digitising a master map, adding boundaries that had been abolished by changes and giving all features and attributes to allow a coverage to be extracted for any appropriate date, area, and type of

Links in this Article

-  [GBHGIS](#)
-  [AHDS](#)
-  [EDINA](#)
-  [GBHGIS Demo](#)
-  [GIS Guide to Good Practice](#)

unit; writing software to allow easy data retrieval based on a 'point-and-click' menu interface and writing software and creating gazetteers (look-up tables) to allow the map data to be linked to as wide a selection of attribute data as possible.

The main source for digitising the maps came from the second edition of the 2 miles to the inch county administrative area maps published by the Ordnance Survey between 1907 and 1910. This was an extremely useful edition for the project as it showed both Poor Law Unions and Local Government units. Other maps include the 1/2" maps, 1901-10. London parish boundaries have been established from 6" maps. A project for the centre for Metropolitan history used 25" maps for City parishes. As non-digital maps obviously only give a snap shot of changes at a certain time textual records were also used such as the Annual Reports of the Registrar General (1896 onwards) and the Registrar General's Decennial Supplements (1861 to 1896). These list the districts affected, the area transferred, some measure of the population involved, and the exact date of the change. Other sources include earlier ESRC projects which had already begun constructing a database of these changes up to 1901, Youngs, Local administrative units of England, RHS 1979, a place name gazetteer for the 1881 census enumerators' books and pre-census taxation sources like the Public Record Office's E179 project. Additional information came from various relevant tables in Parliamentary Papers. For changes prior to 1861 a statement of the names of the several Unions and Poor Law Parishes in England and Wales was used which gave dates for the creation of each Union and subsequent changes. This source was also used to establish when Unions behaved differently from Registration Districts. As a crosscheck, census reports from 1851 to 1911 were used to check which Registration Districts had been affected by changes in the inter-censal periods, and the population involved. The 1900 two miles-to-the-inch, 1888 4 miles-to-the-inch and 1851 12.5 miles-to-the-inch sketch maps were also used to help interpret changes.

Once a map was digitised, a county would be transformed onto detailed National Grid co-ordinates using features such as churches and railway junctions to provide reference points. It was then joined onto the existing counties and rigorously checked to ensure logical consistency at every date. The local government districts, namely County and Municipal Boroughs and Urban and Rural Districts were then digitised and joined onto a separate coverage in a similar way. In a GIS a spatial feature like a line (an arc to use the terminology) has a linked attribute table holding data about it. In relation to boundaries, information such as the county that the arc is in, and whether an arc is or is not a county boundary is provided. The arc is also given start and end dates to identify changing boundaries. ARC/INFO stores areas (polygons) as a series of lines enclosing a label. Each label point holds the attributes for the entire area, including the name of the area, the county it is in, and start and end dates. More complicated types of boundary change are also dealt with such as old areas disappearing, new areas being formed, and areas changing name. The storing of name is what gives the GIS its real power as it means that the map data can be linked to any data table that has the same names allowing mapping and spatial analysis to be performed.

To use the system and to allow easy extraction of data a software package was written which uses point-and-click menus to specify the type of unit required (Poor Law Unions, Registration District, Union County, Local Government units, or Administrative County), the date and the area covered (from the individual county to all of England and Wales). The program then extracts the appropriate arcs (boundaries), label points, and other features and builds an ARC/INFO polygon coverage from them. The new coverage can then either be stored on the workstation for further manipulation by ARC/INFO or downloaded in a range of formats. The final stage involves linking the GIS to a relational database running under Oracle (which apparently links easily to ARC/INFO) to allow querying of, and map production from, attribute data through the GIS. This has required building gazetteers so that the polygons on the GIS can be joined to data in the database, matching by place name. While there is a gazetteer available, Humphrey Southall described the project's plans to develop a more specialised gazetteer, of the ancient parishes of England and Wales which would extend the chronological range of the BHGIS back to the early nineteenth century and earlier when evidence from contemporary maps is partial and unreliable. It would also serve to link relevant data

already computerised by academic projects over the last decade. By doing this the GBHD could produce parish information online, which would provide: a simple statistical history of any town or village selected by the user; a specialised gazetteer for England and Wales providing locations for all civil administrative units from Civil parishes upwards since 1876 and all reporting units identified in a range of systematic surveys from the 13th century onwards and each located within a hierarchy of units and by grid reference. A demonstration system can be seen at <http://www.geog.qmw.ac.uk/gbhgis/gaz/start.html>. Several sources which have already computerised have been identified which could be used and the structure of such a gazetteer which would consist of several tables held in a relational database system.

It was agreed after the presentation that there was considerable scope to explore future collaboration between BHGIS and archival National Name Authority File for British place names. It was agreed that Ian Gregory would copy the parish lists for Hampshire, Northamptonshire and London for further investigation by the county records offices. Several people from county record offices were especially interested in whether once the project is completed (depending on funding) they could obtain a subset of the data at county level. The gazetteer would be extremely useful for archivists as well as family and local historians. It was suggested that they would be in a position to up-date it when necessary and Essex could maintain it and migrate it whenever required. Another point raised were the problems in preserving GIS. At the moment the GBHGIS deposit their GIS with Essex but in a format which does not easily allow it to be revived in its real GIS format.

1. 'An organised collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyse, and display all forms of geographically referenced information', taken from 'GIS Guide to Good Practice' on the AHDS web-site (<http://ads.ahds.ac.uk/project/goodguides/gis/sect72.html#gis>.)

Managing the corporate memory

Richard Davis, NDAD Data & Applications Specialist

On Thursday May 13th, the British Computer Society's Document Imaging Group and the Public Record Office jointly hosted a half-day meeting on digital archiving at the PRO in Kew.

In the first presentation, *Issues of Accessing Electronic Records Over Long Periods*, **Ian Macfarlane** from the PRO discussed the PRO's future role and the lifecycle of electronic records. A review of 1995 had identified that, as hard copy diminished and electronic records became increasingly important, the PRO needed to address the problems and challenges associated with preserving electronic data in order to continue to fulfil its remit to record the background to evidence based policy decisions.

The basic life cycle of electronic records was identified as having four key stages, analagous with the life cycle of paper records: created, filed, semi-archived (typically at 3 years), archived (30 years).

Among the many issues that needed to be addressed were identified the needs for ...


- provision of public access to the records
- dealing with records from different computer systems
- archiving important discussions taking place via email
- topic-based search facilities
- version control
- secure, long-term off-line or nearline storage
- reviewing the PRO's appraisal procedures for electronic records (the PRO currently takes 2% of paper records from government departments)


Among the projects that resulted from these deliberations were EROS, targeting electronic records in office systems, and NDAD, dealing with datasets from government departments. In approaching these issues, the PRO decided that, in addition to developing plans for appraisal, safekeeping and disposal, it also needed a policy on corporate records and inventories of application and record collections.

The recent white paper *Modernising Government* set the year 2004 as a target for all public documents to be in electronic form. As part of the *Invest To Save* budget the PRO and CCTA were jointly awarded the Electronic Records Management Systems project intended to define a common statement of organisational requirements for ERMS. The benefits of developing such a system would include making IT systems more customer-driven, cheaper and quicker procurement, and a reduction of the financial and technological risks of such projects. This program has a parallel in the European Union's DLM Forum.

In the longer term, the PRO must address other preservation issues such as the need for more permanent media. 'Etched' CDs in titanium and toughened glass are being considered for some applications, while for others the approach adopted by NDAD (tapes are regularly sampled, copied and replaced automatically) clearly has a great deal of in-built longevity and security. The PRO expects to be able to move data to new physical media as and when necessary, and will also ensure

Links in this Article

 [Quantor Ltd](#)

 [National Archives Learning Curve \(PRO\)](#)

 [CEDARS](#)

 [Leeds File Archive](#)

that, wherever possible, archived electronic records are stored using stable, documented ISO or similar generic formats (e.g. TIFF, PostScript, ASCII, SGML). Another option for providing access to files from a wide variety of computer systems would be to use viewing software capable of rendering documents and data created by different systems (QuickView, in Windows 95, is an example of such a viewer program).

In addition, the PRO's Online Electronic Catalogue (PROCAT), through the media of both the Internet and the Government Secure Intranet (GSI) will consolidate holdings within the PRO, NDAD, EROS and Departmental systems. This in turn will enhance access to government information, and make a significant contribution to current Freedom of Information initiatives.

The second presentation was by **Bernard Harris** of [Quantor Ltd](#). Quantor was formerly part of [NCR](#), and now is an independent company focused principally on micrographics solutions to data storage, including COM and COLD: Computer Output to Microfilm, and to Laser Disc.

Mr Harris also emphasised that the life-cycle of documents included an active and inactive stage. He pointed out that in spite of the increasing trend towards using electronic documents and data storage, *paper* was still in very widespread use, and its archiving still requires careful consideration. With the ever growing use of electronic and magnetic-based media, micrographics are sometimes considered to have limited use, but they remain a valuable archival tool, particularly for preserving copies of paper documents.

New technologies do not invariably supplant older ones, and therefore mixed-technology solutions are often desirable. Particularly in areas of legal admissibility and evidential weight, micrographics still have many advantages over electronic records. COLD is a good medium to use for the distribution of data, possibly in combination with electronic methods; microfilm images, once created, can then be digitized if necessary, even as part of an on-demand system. Micrographics still have a significant role to play in archival solutions.

Next, from the PRO, **Caroline Shenton** discussed and demonstrated some of the PRO's recent adventures on the Internet. Information about genealogy at the PRO's web site is always immensely popular: this will receive a major boost when the 1901 census goes online. This project has recently gone to tender, and its goal is to make data from the 1901 census available on the World Wide Web as soon as it is open to the public in 2002. Furthermore, the PRO plans to have a full-featured searchable catalogue system on the Web by 2001, although the basic catalogue data may be available much sooner.

The PRO was heavily committed to the UK Government initiative, the *National Grid for Learning* and has already used the Web to create a number of interactive online presentations of interest to teachers and students, based on historic documents and documents (such as the sinking of the Titanic and the Domesday Book). As a further example of the Web's potential to assist "exploring history through original documents", and the PRO's own [National Archives Learning Curve](#), Caroline showed us copies of government correspondence from the 1960s, chronicling Harold Wilson's unsuccessful attempts to buy a colour television for 10 Downing Street, contrary to the wishes of the security services who believed this would pose a security threat.

The PRO is committed to using the Web to expand its user base and offer virtual exhibitions using high quality well-packaged multimedia.

The final presentation was *Solving the problems of long term storage of digital data*, by **Dr David Holdsworth** of Leeds University. [Dr Holdsworth](#) discussed the many challenges of preserving important data which needed to be confronted, such as those of

- deteriorating media
- obsolete or malfunctioning reading equipment
- obsolete data formats
- proprietary systems

- intellectual access
- intellectual property.

[CEDARS](#) is a project which addresses these issues, aiming to address the strategic, methodological and practical issues and will provide guidance for libraries in best practice for digital preservation. One methodology currently being reviewed is the Open Archival Information System (OAIS) is a model for creating descriptive packages for digital data, comprising discrete packages of Archival information, Submission information and Dissemination information: it has been submitted to ANSI and ISO.

Dr Holdsworth described the importance of migrating important data to new media and formats as they appear. At the [Leeds File archive](#) ongoing migration is integral part of software design. For archival and historical reasons it may also be desirable, where possible, to preserve the data's original byte stream and rendering software. It is frequently possible to create, on new computer systems, emulators of older computers, and programs designed for those systems can then be run as if in their original environment. Dr Holdsworth referred to emulator projects which he hoped to complete one day, including creating an emulation of ICL's GEORGE 3 operating system.

As storage and processing capabilities continue to increase at an exponential rate, the costs of keeping today's data with tomorrow's technology are minuscule, but indexing and cataloguing are essential if that data is to be of use to researchers in the future. "Such is the advance of technology that the cost of storing the data from yesteryear trickles away to nothing if you move your data with the technology, and do not rely on being able to read the media from yesteryear."

The seminar ended with questions to the presenters, and discussion of matters such as the importance of keeping important data, including images and texts, in standardised formats, and as separate as possible from potentially fragile or proprietary rendering technologies (e.g. Java, Javascript, Windows). The importance of catalogues and the appraisal process was emphasised yet again, reminding me of something Tom Lehrer said 40 years ago: "Life is like a sewer: what you get out of it depends on what you put into it."

Personnel Profile

Sally Hughes is one of the NDAD team's Data and Applications Specialists. Sally has worked with a number of our early datasets, consulting with government departments, analysing the data received and preparing descriptions of the data for inclusion in NDAD online catalogues.





My post-school life started with three years at Westfield College, University of London, studying pure maths (with English as a subsidiary subject). Sadly, Westfield is no longer based in Hampstead as it was in my time (it merged with Queen Mary College in 1989 and is now centred on the Mile End Campus in the East End - described by [QMW](#) as "an historic, vibrant, multi-cultural area of London" but to my mind missing some of the advantages of the former location, including its proximity to Hampstead Heath and Golders Hill Park).

After graduating, I worked for a year before doing a one-year Postgraduate Diploma course at the College of Librarianship Wales (CLW) at Llanbadarn Fawr, Aberystwyth (CLW is now The Department of Information and Library Studies of the University of Wales). I much enjoyed my time in Aberystwyth, particularly the joys of walking along the sea-front when the sea was at its wildest.

I then spent two years working in the library and information services of the Institution of Mining and Metallurgy in Portland Place (handy for shopping in Oxford Street). Having by then become interested in the developments in the applications of computers in libraries and information science, I applied for a job in the Information, Statistics and Computing Section of the Food Science Division of MAFF. This was my entry point into the world of computers and it was an exciting time: computers had only relatively recently become available for use in offices and introduction within the Civil Service was strictly controlled and limited. Luckily the head of the section was particularly good at 'working the system' (justifying the procurement) and the first machine, a Wang, (I think it was a 2000T - perhaps the first personal computer worthy of the name?) was obtained under the guise of being a laboratory calculator. I joined when this machine was about to be upgraded to a multi-user Wang 2200 MVP (with 4 terminals). Since most of the operating system was "hard-wired", the Wang's o/s and interpreted Basic occupied only about 3K of the total 64K memory. The Wang was used for storage and analysis of food-related surveys, monitoring research projects, and what was perhaps one of the first systems in government for control of registered files. For the latter, a free-text system was written, with stop-list and using bit manipulation; this seemed amazingly quick and its performance would perhaps even today be more than acceptable. All software was written in-house, including a communications program which enabled the Wang to be used as an intelligent terminal for accessing on-line information providers such as Lockheed Dialog and for the received information to be stored on disk.

In the mid 1980s MAFF implemented a department-wide computer strategy, based on PRIME minicomputers. Food Science started with a 750 and progressed over the years to a PRIME 9955 II. During this time, I was involved in operating and administering the minicomputer, supporting users (including their use of the Grapevine broadband network) and developing applications, using mainly Doric/Henco Info and later PRIME Information (Pick-like DBMS) and the STATUS free-text software. I was involved in developing systems to assist with the scientific aspects of some of the food-related 'incidents' which the Ministry had to deal with, including the scare concerning

Links in this Article

-  [Ministry of Agriculture, Fisheries and Food](#)
-  [Queen Mary and Westfield College](#)
-  [ICI: History of IC House, Millbank](#)
-  [British Computer Society](#)

contaminants in wine and also MAFF's monitoring of the fallout from the 1986 Chernobyl disaster. In the early 1990s, MAFF implemented its office automation strategy, MAIDEN (MAFF Integrated Desktop Environment for the Nineties); this was based on Siemens Nixdorf UNIX servers. Soon after this I moved within MAFF to the Information Technology Directorate to work as a business analyst.

Over the years I worked in a number of MAFF offices, starting in Great Westminster House (Horseferry Road); when this was due to be demolished, we moved along the road to Nobel (& Ergon) House, Smith Square (an impressive building, adjoining the ICI HQ - the main [doors](#) of the latter are worth stopping to admire if you are down that way). From there I moved to Victory House, Kingsway, then, when that was vacated by MAFF, I had short stints in Ergon House (again) and St Christopher House in Southwark Street. During my years with MAFF I used a number of computers, operating systems, programming languages & software packages and developed a number of applications; this has proved a useful background for my work with NDAD (as has my knowledge of the workings of the Civil Service). I was also fortunate that MAFF funded my attendance at a course leading to the [British Computer Society](#) examinations, as a result of which I became a MBCS.

I applied for the NDAD Data and Applications post at ULCC because it sounded such an interesting new project, and offered the opportunity to combine my early interest in information science with my knowledge and experience in IT. I joined ULCC in December 1997 and have much enjoyed the work, collaboration with the other data specialists and archivists, and even dealing with government departments! (My knowledge of MAFF has come in useful: I have had a fair bit of contact with MAFF whilst working on datasets being archived by that Department). The working environment at ULCC is a little less hectic than at MAFF (although I wouldn't want to give the impression that we don't work hard!) and I have the luxury of having only a short walk to work, which means that I have had more time and energy to devote to other interests, including London (its buildings and history - I can recommend the history of London evening class at Birkbeck College), gardening, and walking.