

NDAD News

No. 2
April 6th 1998

[Project Development Manager's Report](#)

[Dataset News](#)

[Our First Dataset](#)

[FAQ #1: User Registration](#)

[Personnel Profile: Ruth Vyse, University of London Library Archivist](#)

Just in time for Easter, this is the second NDAD newsletter.

Now that the system has successfully completed trialling by the Public Record Office, NDAD staff are working hard to get more datasets online. In this issue we include some details of those datasets that are in the proverbial pipeline!



In *Our First Dataset* an NDAD data & applications specialist talks about her work on a dataset from the Museums & Galleries commission. We hope you will be interested to find out how the datasets get from government departments to **you**.

We kick off our *Frequently Asked Questions* series with answers to some common questions from our ever-growing band of users about the NDAD User Registration process.

Our *Personnel Profile* continues with a profile of Ruth Vyse, the University of London Archivist who has worked extensively on the project since its inception in 1996.

And finally ... a Happy Easter to all NDAD users!

[CRDA News](#)

[NDAD News](#)

[NDAD](#)

[NDAD Home Page](#)

NDAD Project Development

Kevin Ashley reports on a momentous few months in the lives of NDAD staff at the University of London Computer Centre, bringing together a system which uniquely combines the expertise of archive and information specialists.

This report is being written shortly after a significant event in the lifetime of this new service for central government records. On March 24th 1998 the services and systems put in place by ULCC and the University of London Library for the NDAD project were formally accepted by the PRO, bringing to a close an intense and sometimes difficult period of development and trialling, and heralding the start of what we hope will prove to be a valuable and popular service for all those with an interest in digital records.

The confirmation of the new service took place a few days before the introduction of the [PRO's new web site](#), which is already showcasing examples of the work being undertaken by them on the AD2001 project. This and other developments in the work of online archives - whether providing access to online catalogues of traditional archive material, or to online archive material itself - are illustrative of the rapid changes taking place in the way archivists are providing access to the material in their custody. The services to be provided by NDAD will bring together work in three of these areas - online searchable catalogues, preservation of material which is inherently digital in nature, and online access to archival sources.

Most of the services NDAD is offering, and the technologies behind them, are already on offer in one form or another elsewhere in the UK and the world. Online access to data is a service which has been offered by many academic computing services for a number of years. Indeed, ULCC began doing this almost 30 years ago, with key social science resources such as the Census Small Area Statistics and the Family Expenditure Survey being accessed through statistical packages such as SPSS and SAS. Online catalogues of material have been provided by libraries in the academic and public sectors for some years. Online archival catalogues are still something of a novelty, although (as mentioned above) they are being rapidly made available. Preservation of digital material is also an activity which has been the concern of many groups for almost as long as digital material has existed. Much of this has taken place within industry (the seismic and pharmaceutical industries being particularly active in this area) and within specific scientific disciplines which involve large amounts of unique data, such as pictures from weather satellites, or data which is expensive to re-create, such as that from atomic and nuclear physics. In almost all cases, these data sources are being preserved by the bodies which created and continue to use them. A smaller number of centres, of which the outstanding example is the [Essex Data Archive](#), have focussed on collecting information from a variety of sources to form archives of interest to one or more research groups.

What we believe is new about what we are doing at NDAD is the synthesis of each of these elements in a single service. We are preserving and cataloguing acquisitions to archival standards, providing online access to the catalogues, and linking these directly to the preserved information itself. The infrastructure delivering the online service has been sized so as to be able to handle individual datasets of many gigabytes in size, and a total data capacity of many terabytes. In preserving the databases and associated material which arrives with them, we must focus not only on the short-term needs of the academic researcher of today, but on the wider needs of others who may want access today, and more particularly on the needs of future generations. Material which we are accessioning is not being preserved for 7 years, 30 years or even 100 years, but for the foreseeable future. To ensure that it is comprehensible and accessible in the distant future requires

us to put a great deal of thought into the catalogue descriptions which you read, and into the invisible meta-data we create which allows our system to know how to process the data and deliver it to you. It is this meta-data which will be used to aid in translating the databases to future software environments and systems. The meta-data, along with the catalogues and dataset documentation which arrives from departments, will be all that future generations will have to enable them to make sense of the archive being created today. Even now, much of the information we need to properly describe a database is not necessarily available from any written source, but must be gathered through verbal interview with those directly involved with its creation and use.

The team we have gathered together to make this possible is itself a rare mixture of talents.

- The project archivists - currently headed by Ruth Vyse, whose profile appears [elsewhere in this issue](#) - are the key players in the decisions about how to organise the holdings of the archive, how to catalogue and index those holdings so as to make them accessible and comprehensible, and guide the decisions of the various computing specialists on data processing so as to ensure the archival integrity of the holdings.
- The archive assistants provide backup to the archivists in the organisation and preservation of material, which includes the digital scanning of paper documents received. They also provide general administrative support to the project, as well as coordinating the provision of help and advice to the service's users. They are often the first point of contact for users contacting our helpdesk.
- The data specialists act as the interface between the archival and computing worlds. Each of them has a sound background in areas related to database design and use, statistical analysis of tabular data or user interfaces to online data and catalogues. Their role centres on ensuring that the data received from departments is what it purports to be, producing meta-data descriptions of the data, converting databases into a form suitable for long-term preservation and access, and validating the converted data both to ensure that its description is correct and that the conversion process has preserved its integrity. Most are also involved to some extent in the production of the web site and some of the software tools used on it.
- The systems support staff are responsible for the construction and smooth running of the various software and hardware systems on which the whole service runs. This includes the tape robot in which all the data and scanned documentation is ultimately stored, the hierarchical storage management system which keeps track of the data within it, and the web servers which deliver web pages to the end user.

The staff profiles which will run in each future issue of the newsletter will allow you to meet - if only virtually - each of those involved in making this service possible.

ULCC has been privileged to be able to bring together a team covering all these skills of which we can be extremely proud. We are still as excited about the possibilities of what can be done as we were at the project's inception. We hope we can enable you to share in that enthusiasm.

Kevin Ashley

Work in progress

Datasets: Currently Available

This is a list detailing the data that have been received by ULCC from government departments. Archivists and data specialists at ULCC are currently working extremely hard to bring these datasets to you in the near future.

Title: **Digest of Museums and Galleries (DOMUS)**
Department: Museums and Galleries Commission (MGC)

Title: **Form 7 (Schools Census)**
Department: Department for Education and Employment (DfEE)
Currently the Census for 1976 is available. Data for other years will follow, but access to data from later years may be subject to certain restrictions.

Title: **British Crime Survey**
Department: Home Office (HO)
This dataset covers 1996. Data access restrictions may apply.

Title: **Metropolitan Police Crime Statistics**
Department: Metropolitan Police Service (MPS)
This dataset covers the period 1990-1996. Data access restrictions may apply.

Title: **Coast Protection Survey Database**
Department: Ministry of Agriculture, Fisheries and Food (MAFF)
Holds information about the 1994 statistics of coast protection defences in England, that identified where work was required, etc.

Title: **Form 7 (Schools Census) 1993**
Department: Department for Education and Employment (DfEE)
Data access restrictions may apply.

Title: **Coastal Survey of Wales**
Department: Welsh Office
Review of Welsh coast and sea defences, January 1998.

Behind the scenes at the museum

The story behind NDAD's first on-line dataset

The first dataset to be made available on NDAD was the Digest of Museum Statistics (DOMUS), provided by the Museum and Galleries Commission (MGC). *Sally Hughes*, the Data & Applications Specialist responsible for this dataset, agreed to answer some questions about what was involved in getting the data from the MGC and onto the NDAD system.

Tell us a little about the background to the dataset.

As I understand it, the main aim of the DOMUS is to add to the body of information available about UK museums and galleries. The Museums and Galleries Commission carried out some work in the early '90s analysing the market potential for museums and art galleries and this identified the need for more information and in particular for accurate statistical data about the museum sector. DOMUS was launched by the MGC in Spring 1994, following a pilot study in 1993. Forms are sent out annually to all museums which are part of the MGC's Registration Scheme.

How much data was there?

DOMUS is a fairly small dataset compared to others that will be held on NDAD. The total size of the 8 files sent to us by the MGC was less than 3 MB. This covers 1753 museums and galleries. Later in 1998 we are expecting a further two files, covering two special surveys: (i) care of collections and (ii) public access. (Since 1996 the DOMUS survey has included a set of questions on a topic particular to that year.) After that, we are expecting to receive annual accessions of DOMUS data.

What media were used?

The data was sent to us on 2 floppy disks. This made it fairly simple for the MGC to send the data to us and for me to deal with the data at this end. The only slight complication was de-compressing the files, but this was no problem once I had tracked down the software needed to do this.

What was the format of the data that was sent?

The DOMUS system was developed in FoxPro for the MGC by a consultant. It was therefore easier for them to simply send us the FoxPro files. Where we cannot read the DBMS proprietary format we would ask the Department to export the dataset in a generic format, such as comma-separated variable length records (CSV).

What did you do once the dataset arrived? What procedures did you follow?

When you are processing a dataset, there are various procedures to be followed, in particular to ensure the security and integrity of the data, but the two main aims are:

1. to convert the data into standard formats, suitable for on-line browsing and for producing copies for researchers to use on their own computer systems;
2. to describe the structure of the dataset and format of the fields in sufficient detail to enable researchers to search and analyse the data effectively.

(A third aim is to identify and document any errors or anomalies in the original data, but this is really part of describing the dataset).

I imported the data into Microsoft Access, as Access can read FoxPro files. The first task was to understand the data, including the relationships between the files. As Access is pretty flexible and

has powerful built-in facilities to sort, group data etc, I was able to look at the data in some detail and this helped in deciding what needed to be done to produce the data in the final required form (in this case it was fairly clear that the most efficient format would be CSV) and to detect errors in the original data.

There was little work required on the data itself. Perhaps the most significant task was separating out the Institutional and personal addresses; this had to be done as the latter are 'closed' for 30 years, i.e. the personal names and address are not to be viewable via NDAD.

Export of the data in comma-separated format was straightforward as this is one of the facilities provided by Access. However I did have to do this more than once to get the dates exported in the right format, as Access defaulted to exporting dates with two-digit years! Perhaps this demonstrates that even Microsoft is not immune to the Year 2000 problem!

Did you encounter any other problems or difficulties?

The main problem was caused by the Postcode being used as the link field between three of the files. Presumably the MGC DOMUS system is more forgiving of errors in key fields (e.g. blank and duplicate postcodes) than Access is! It did require some perseverance to get round these errors in order to pull together the Institution addresses with the other organisation details. I worked closely with our Assistant Archivist, Peter Garrod, since most of a dataset catalogue, including the description of the dataset's background and context, is produced by the project's archivists.

Other problems were fairly minor and if all datasets are this easy and the Departments as co-operative, we won't have anything to complain of - but somehow I suspect that not all of the datasets will be as amenable to analysis and conversion! I would like to thank Bethan Hurst of the MGC for all her help and for putting up so graciously with all our phone-calls asking for yet more explanations!

Are you pleased with the results?

Yes I am, particularly as the content of the dataset is interesting to me (I am not sure that I will come across many more of such direct interest after all, when going on holiday I will be able to look up what museums in the area have collections of interest to me!).

On a more serious note, it's good that NDAD's first dataset is one that will (when we get the next acquisition) allow researchers to carry out trend analyses of the data which cannot be done on the original system. This is because although the MGC system holds some annual data (i.e. each year, there is a new set of data covering the 'changeable' data such as staffing and visitors numbers, opening hours and financial details), the other data (i.e. the 'reference' data which changes less frequently such as the name of the museum, address, types of collections) is over-written each year. Storage of each year's data on NDAD means that it will be possible to, for instance, compare the financial data with data of the same date on the types of collections held at the Museum.

Are there any particular aspects of the processing of this dataset which will be useful when handling future datasets?

Apart from general increased familiarity in following the procedures to process and document datasets, the main thing that springs to mind is the need to be able to ask about errors and anomalies in the data without giving the impression that one is being critical of the design of the original system, in other words to be tactful! Hopefully we accomplished this but perhaps we ought to ask one of our Departmental contacts to put forward *their* views in the next issue of the NDAD newsletter so we have a better idea of what's it like being on the other side and in particular what we should avoid doing next time round! After all, it should be a learning process for us, the Departments and the PRO.

Frequently Asked Questions

1: User Registration

This is the first in a planned series of articles answering frequently asked questions about aspects of the NDAD system. Since most users of the system will need to register, *User Registration* seemed a very good place to start.

I only want to read the extremely interesting information which you have provided about UK government departments. Do I need to register?

You are free to view any of the pages on the web-site. Only if you wish to view the data itself must you register, as this part of the system is protected. You won't be able to look at the data until you have a username and password. Otherwise, unless you are challenged for a password, you are free to view anything at the web-site.

What happens when I fill out the on-line registration form?

When you submit a completed on-line registration form, the information you provide is stored on our system. If you have provided us with a valid email address then you will receive an email message confirming this. We will then send you the registration confirmation form, and a letter telling you your username and password.

So I can use the system straight away?

No. You will be able to use the system only when we have received your signed registration confirmation form.

But I have already filled out the on-line registration form. Why must I fill out a second form?

Because we require a signature before we can authorize you to view the data! The form we send you will be pre-filled with any information you have already submitted using the on-line registration form, so all you need to do is check this is correct, sign the form and return it to us.

What will you use the information for?

The information is strictly for our own internal use, so that we can contact you, send you NDAD-related information, and send you any data you may request. If you order any chargeable services from us, then we will obviously need to invoice you, too!

The information will not be supplied to anyone else. From time to time we may analyse the details we hold on an anonymous basis, to identify usage trends.

I am a non-UK national and/or a non-UK resident. Is that a problem?

There is no restriction on registering users of any nationality to use the services provided. We ask users to supply this information for the purposes of producing statistical information for the Public Record Office, and so that we can be aware of any communications issues that may arise.

Since I first registered I ... have moved house / have changed my email address / want to change my password. How do I do this?

All of these eventualities have been anticipated! Go to the main [registration page](#) and follow the links to the Change User Details form or the Change Password form. Of course you must be a registered user in order to use these functions!



National Digital Archive - Datasets

User Registration

Please use this form to apply for registration as a user of the National Digital Archive - Datasets. When you submit your application, you will immediately receive confirmation via email (if you have specified an email address).

Any information provided in this form will be used by the University of London and the Public Record Office solely for the purpose of communications regarding the National Digital Archive - Datasets and related services. It will not be used for any unrelated purpose or passed to third parties.

If you need any assistance with the NDAD registration process, you can email [Registration](#), or you can [contact us](#) by phone or by post.

First Name:
 Last Name:
 Title:
eg. Dr, Mr, Dr
 Email Address:
leave this field blank if you have no email address
 Nationality: UK National
 Other

Other UK National enter nationality above
 Home Address:



NDAD Change User Details

Use this screen at any time to change your account details.

When you are happy that your details are correct, click the **Change Record** button to save your changes. If you make a mistake, the **Reset Fields** button will reset all fields to their original values.

User ID: **crmd1**
 Email:
 Address:
 Correspondence Address:
 Phone:
 Fax:



National Digital Archive of Datasets

Change your Password

Please use this form to change your password. This should only be used by registered users of the NDAD archive.

If you need any assistance with the NDAD registration or password process, you can email [Registration](#), or you can [contact us](#) by phone, by post or in person. If you have forgotten or inadvertently compromised your password please contact us as soon as possible by phone or email.

Change password for crmd1

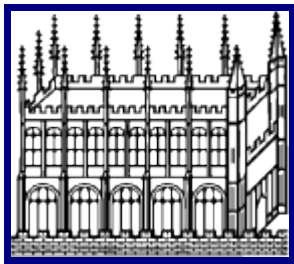
Type your new password into both text fields and press "Change"

New Password:
 Type it again:

Personnel Profile

2: *Ruth Vyse*

Ruth Vyse is the University of London Archivist. She has played a key role in the project team from the creation of its initial bid to the PRO in late 1996, and has ensured that the service developed has integrated sound archival principles with its computing technology. Ruth has spent the last 16 months balancing her extensive work on the project with her many responsibilities at the University of London.



I did my archive training at University College London. My first archive post in 1975 was as Assistant Keeper in the University Archives at Oxford. As the only full time member of staff I had a range of duties including dealing with enquiries, accessioning and cataloguing archive material, indexing and ensuring the preservation of older and more recent records from a range of university departments. My office and one of the storage areas for the archives was in a seventeenth century tower forming part of the Bodleian Library. This was where the archives had been kept since the appointment of the first Keeper of the Archives

in the 17th century. My office had an 18th century plaster ceiling and a very good view of the Radcliffe square area of Oxford. The post provided a chance to gain experience of a wide range of archival duties and to deal with documents from the 13th century to the present day.

Whilst at Oxford I was secretary of one of the regional groups of the Society of Archivists and convenor of a working group on methods of listing. Part of the latter's work was incorporated subsequently into the textbook *Manual of Archival Description*.

From Oxford I moved to Barnsley in South Yorkshire, where I was responsible for setting up an archive service in the Central Library. I carried out a range of duties including promoting the service via talks to local groups, a local history fair, exhibitions and assistance with the publication of a book of photographs on Barnsley and a video on the area. The archive service took in records from many sources including those of local organisations, local firms and non-conformist chapels. One of the exhibitions I mounted dealt with the linen industry in Barnsley, the town's major industry in the first half of the 19th century, before it was replaced by coal. The post provided experience of a variety of archive tasks including outreach activities, conservation and dealing with non-paper material such as photographs and film. It also increased my knowledge and understanding of local history.

My next move was to the University of London Library where I have been since 1994. I am responsible for the archives of the central university and the manuscript holdings of the library. The last three years have presented many challenges to both the central university and the library. A major part of my work has been to survey and take in records as the central administration moved offices and adapted to a changed role. I have also carried out surveys of current and semi-current records preparatory to establishing a records management programme.

At the end of 1996 ULCC asked the University Library to join it



in a bid to the PRO to manage the preservation of datasets from central government departments. I was asked as University Archivist to provide guidelines on cataloguing since the PRO wanted the contractor to carry out this task in addition to the preservation of the datasets. When ULCC and ULL were awarded the contract in Summer 1997 I was seconded for six months for four days a week to the project. Although I had dealt with non-paper records before working on the project I had not dealt with datasets. I have enjoyed the opportunity to see how archival cataloguing techniques can be applied to this kind of record. I have welcomed the chance to work with and learn from data and systems specialists. A fruitful collaboration has been established already between all groups particularly in the area of describing both the administrative background to datasets and the data which they contain. I think that there is tremendous potential as the work progresses for crossing what has sometimes been a divide between archivists and those in the computer field and showing how both can work together fruitfully.