

# NDAD News

Newsletter of the UK National Digital Archive of Datasets

---

No. 5, March 30th 1999

---

## [Project News](#)

News from the project and other recent events.

## [EAD UK Users' Day](#)

Patricia's report on this well-attended event at the PRO.

## [Towards an NDAD FAQ](#)

Fiona's collection of Frequently Asked Questions about NDAD, with answers!

## [Personnel Profile](#)

Richard's turn to release the balls.

# NDAD Project News

[EAD UK Users' Day](#)  
[NDAD lecture to UCL students](#)  
[Visit by staff of Essex Data Archive](#)  
[Presentation to JISC](#)  
[History of ULCC](#)  
[Welcome](#)

---

## EAD UK Users' Day

On 29th January 1999 the PRO hosted the EAD (Encoded Archival Description) UK Users day at Kew and several members of the NDAD team attended. Although the NDAD system does not directly use EAD, it is one of a number of developments in archival standards which the NDAD team are following with interest. Patricia Sleeman describes the event in more detail [elsewhere](#) in this newsletter.

---



## NDAD lecture to UCL students

On 27th January, Ruth and Kevin gave a presentation to students from the [UCL School of Library, Archive and Information Studies](#). The presentation covered the acquisition, cataloguing, preservation of and on-line access to digital datasets, and the students received a tour of the hierarchical storage management facility. The course tutors commented on how valuable it was for the students to be able to see an actual, working electronic records service.

---



## Visit by staff of Essex Data Archive

Staff from the [Data Archive](#) at the University of Essex in Colchester visited NDAD in January. Kevin explained and demonstrated the system to them, and it was a good opportunity for members of the NDAD team to discuss the similarities and the differences between these two important digital archive systems. We hope it will be possible to arrange a return visit, by NDAD staff to Essex, during the Spring.

---



## Presentation to JISC

On 4 February 1999 Ruth Vyse gave a short presentation on NDAD's use of the UNESCO thesaurus for subject indexing of catalogues to a Joint Information Systems Committee ([JISC](#)) seminar. She is to serve on a task force to look at the use of the UNESCO thesaurus, in conjunction with work by Janet Foster on subject indexing, for subject indexing of collection level descriptions of archives and manuscripts.

---



## History of ULCC

During February, the University Library exhibition hall hosted a small exhibition, "Thirty years of London computing - a brief history of ULCC". Photos and exhibits chronicled the history and work of The University of London Computer Centre since its creation in 1968 as a central facility for academic computing in London, later becoming a regional and then a national centre for computing networking resources. The exhibition is now finished, but you will shortly be able to see it online on the new, improved ULCC web site, currently under development.

---



## Welcome Ben and Richard

Two new Unix systems administrators have joined ULCC to support and develop the systems used by NDAD and by other projects at ULCC.

Ben Wheeler first used a Sinclair ZX81 when he was 5, "got a computery degree and did computery jobs at BT and Ford before getting a computery job here. Lives in a house with nine computers and one person. Refutes suggestions that computers are his whole life." (*You forgot to mention that Nokia communicator that you always carry with you. Newsletter Ed.*)

Richard Booth started life in the real world as an engineering geologist for a large international consultancy in South London, and went on to do an MSc. He became involved with work on geotechnical databases and user administration on Unix systems, and decided that he enjoyed working with computers a great deal more than engineering. Richard lives in Sussex and enjoys windsurfing.

# EAD UK Users' Day

On January 29th 1999, the [PRO](#) hosted the EAD UK Users Day , focusing on issues of current concern to the existing and 'putative' EAD community. NDAD Assistant Archivist *Patricia Sleeman* describes the events and discussions that took place.

---

## EAD purpose and history

Encoded Archival Description is an SGML/XML Document Type Definition, a set of rules for defining and expressing the logical structure of archival finding aids in electronic format. These rules are applied using tags (or electronic markup) embedded in the electronic finding aid. Designed for on-line presentation of archival finding aids, EAD is hierarchical, portable and capable of data exchange. Because it is implemented using both the internationally defined Standard Generalized Markup Language (SGML, ISO 8879), and the emerging Extensible Markup Language ([XML](#)) of the World Wide Web Consortium ([W3C](#)), EAD applications can be developed independently of proprietary software.

The EAD DTD is intended for inventories and registers of any length, describing the full range of archival holdings, but its extensible design will allow it to incorporate other types of finding aids (such as guides to repositories) and the development of ancillary coding standards to accommodate archival authority information. Such related or complementary standards include: Text Encoding Initiative (TEI); Guidelines for Electronic Text Encoding and Interchange; the MARC formats; and the International standard archival authority record for corporate bodies, persons and families (ISAAR CF). The EAD encoding scheme does not define or prescribe intellectual content for finding aids, what it does define is content designation and it is compatible with ISAD(G), the key internationally accepted General Standard Archival Description.

Development of EAD began in the University of California, Berkeley Library in 1993 under Daniel Pitti, the principal investigator for the project. After the initial Alpha and Beta versions, in 1998 Version 1.0 of EAD was released. Accompanying the release of the Version 1.0 EAD DTD is a completely revised and updated EAD Tag Library, compiled by the Society of American Archivists EAD Working Group and published in August 1998 by SAA, and which is expected to be available on-line in Autumn 1999.

The PRO were among the main pioneers in the development of EAD, hosting an EAD implementors day in May 1997 as well as an EAD open-day with Daniel Pitti in December 1997 and they provide the British representative on the EAD working Group responsible for the release version of EAD and accompanying documentation.

## EAD developments and sources

Meg Sweet of the PRO (one of two international representatives on the working party converting the beta version to version 1.0) described the main differences between the most recent versions. In particular the <add> element can now be used at any level of description, and the <abstract> element (which has been recommended to be mandatory in ISAD(G)), is intended to provide a brief statement of information about biography, history, scope and content in the first or second screen seen.

One of the problems noted was that EAD is too rich for the internet and requires special browser plug-ins such as the Panorama SGML viewer in order to view an EAD document. The development of Extensible Markup Language (XML) offers a solution: XML is (in many respects) a restricted form of SGML, created so that richly structured documents could be used over the web. Although SGML offers a more flexible and proven solution for the creation and long-time storage of complex documents, in many organizations XML will become the standard format for web delivery. The developers of EAD delayed the launch of Version 1.0 of EAD so that an XML-compatible version could be produced: the fact that the next generation of web browsers will include support for XML offers the prospect of low-cost client software for accessing data and documents distributed using XML, although, inevitably, it seems unlikely that Microsoft's and Netscape's implementations will be 100% compatible with each other or with the [W3C recommendation](#).

Bill Stocking of the PRO followed Meg's interesting presentation with a guide to sources of information about EAD, in particular the Library of Congress site (<http://lcweb.loc.gov/ead/>).

## **Authority records**

Lesley Richmond from the University of Glasgow emphasized the importance of recording authority information to help archivists in the future and in relation to EAD. An authority record defines an entity in a structured and standard way, to establish an unambiguous link between it and other entities. Authority records must be consistent and stable across space and throughout time. ISAAR (CPF) (International standard archival authority record for corporate bodies, persons and families) is an example of such a standard developed by the ICA (International Council for Archives). Lesley called for a web-based network of archival authority records, to act as a focal registration point for a multitude of contributions: in essence a centralized reference point for authority records. The development of AAR DTD (Encoded Archival Authority) was noted.

It was noted that as EAD has been primarily developed in an American context, this had caused certain problems in non-English speaking countries such as France, owing to translation problems. As a result it is desired that the development of EAA DTD be more international compared to that of EAD. Lesley encouraged archivists from all repositories to record authority information and mentioned that the PRO are currently mounting authority records for organizations for which they hold records. Glasgow is already contributing to the National Authority Files project and we were shown some of the records which they had compiled. Lesley suggested that the creators if guided could be coaxed into using authority records when creating their own records. She also spoke about the dramatically changing patterns of public use of archives urged archivists to take the lead in providing contextual information for our resources.

## **Z39.50**

Paul Watry of Liverpool University gave us an overview of the EAD attributes for Z39.50, an information retrieval protocol which enables a researcher to perform complex searches and manipulation of data. He also emphasized the importance of establishing a relationship between MARC-AMC and EAD. The Z39.50 Implementation Group (ZIG) is currently drafting more formal guidelines for the attribute sets. The need for the provision of attribute sets (labels or tags for data that everyone agrees on) for XML/SGML applications is important to enable Z39.50 to work. Z39.50 attributes are usually expressed in numbers, and EAD elements can be mapped to many of the ZIG standard attributes.

HTML, Dynaweb and Z39.50 are all mechanisms for electronic delivery; the time is ripe to develop

links between EAD and Z39.50. Bib-1 could be replaced with the world standard MARC, as the majority of EAD elements are also in MARC-AMC. However, bringing about such convergences would not be easy, and raised many practical problems. Who has the authority to speak for the EAD/Archives community? How can we ensure co-ordinated development among different communities such as CIMI, TEI, MARC, EAD and DDI? Should any if all of the attribute sets become official standards, and if so, under what structure?

Compatibility between EAD and other standards for archival description was mentioned many times during the day. EAD's relationship with Dublin Core was considered potentially problematic. Dublin Core is a metadata element set intended to facilitate discovery of electronic resources. Originally conceived for author-generated description of Web resources, it has since attracted the attention of formal resource description communities such as museums, libraries, government agencies, and commercial organizations. Its central feature is the building of an interdisciplinary, international consensus around a core set of 15 metadata elements. However, certain compatibility problems exist between Dublin Core and EAD: in Dublin Core, for example, the author element identifies the author of the document, whereas in EAD it refers by definition to the author of the finding aid. There seemed to be a general consensus that in many cases Dublin Core is simply not specific enough for many archival needs.\*

## **Retrospective conversion**

After a tasty lunch during which we had the opportunity to look at the PRO's TOPCAT, various presentations on projects for converting and generating EAD were given by representatives from the University of Liverpool, Glasgow University and the PRO.

Julie Allinson from the University of Liverpool described a retrospective conversion of a finding aids for the Calendar of the Cunard papers. Different levels of description were marked up on a hard copy using different colours to represent the tags needed: the documents were then sent for offshore keying. In another exercise, a finding aid held in a relational database was used, and EAD generated by applying tags to the fields selected in a query.

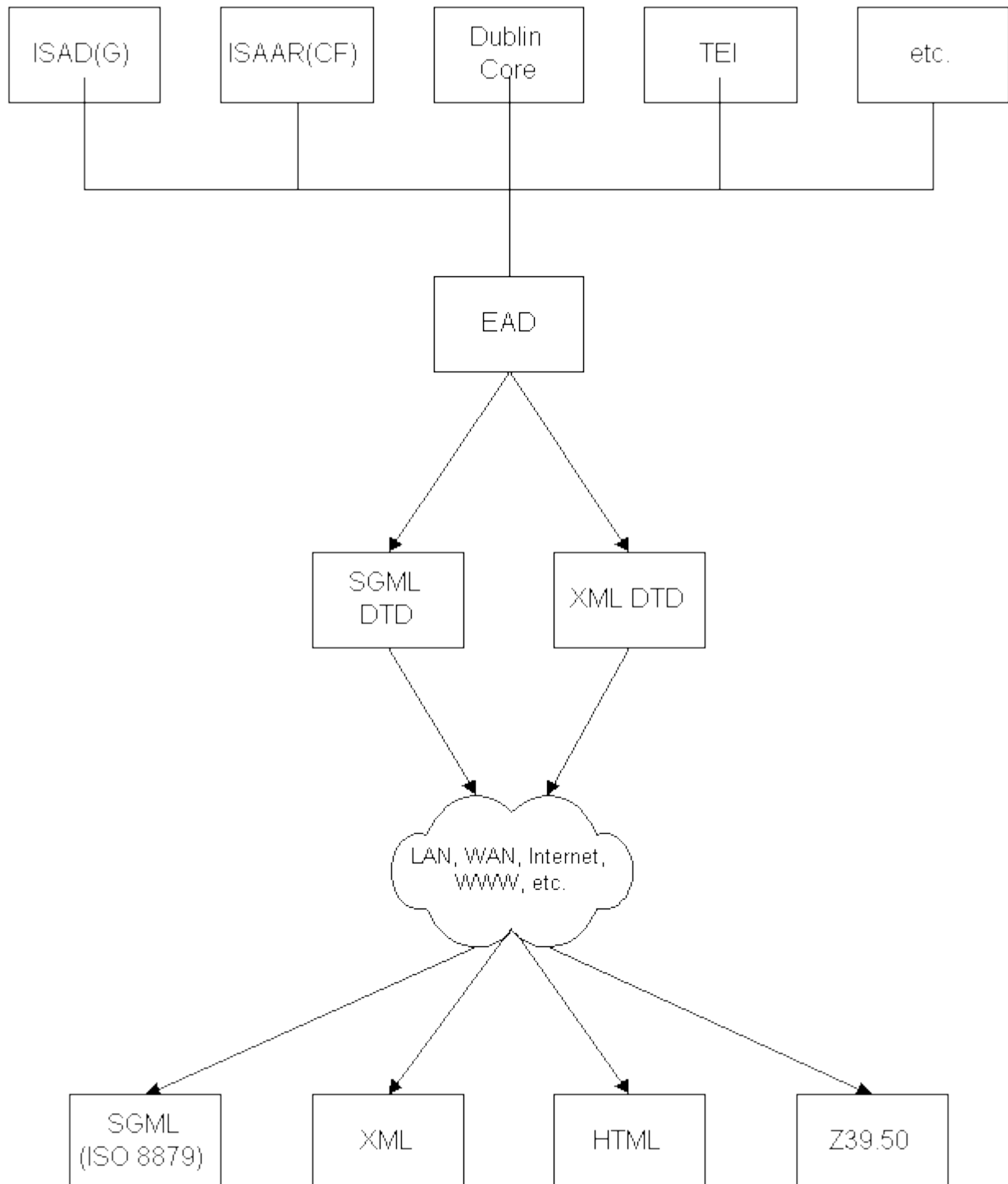
Matthew Hillyard of the PRO showed us TOPCAT, the PRO's retrospective EAD conversion project for their top-level descriptions. 350,000 pages of lower level information has already been converted (PROCAT) and is already available to researchers. The information in TOPCAT had been taken from the PRO's printed guides, entered into an Oracle database, and subsequently converted to EAD. The project is expected to be launched in April.

Pete Johnson of the University of Glasgow demonstrated a Microsoft Word template, using paragraph styles corresponding to the ISAD(G) elements, and character styles used to identify sub-paragraph level EAD elements such as dates and cross-references. Word cannot nest styles, but nesting can be implied through judicious use and naming of styles. A Word macro then processed the styled document and added tags to generate a complete EAD-encoded document.

## **Conclusion**

Summing up, David Thomas discussed how EAD could be the basis of an archive network. Being XML compliant, EAD could act as the display medium over the web; the future lies in developing standards for data content and Z39.50 content: consistency in these areas is extremely important. EAD is still in its infancy and will grow in many unexpected ways in the future, as technological advancements and the emergence of new web browsers and protocols will undoubtedly influence its development. It will also be subject to scrutiny by archivists all over the world who hold very different collections. It is likely to remain work in progress, dynamic and alive.

\* Sorting out existing or potential relationships between these many different standards and initiatives, each with its own intended application and target audience, can be a very perplexing process. The diagram below illustrates our conception of the place of EAD among these standards and the technology used to deliver electronic finding aids and data using computer networks.



# Towards an NDAD FAQ

As NDAD celebrates its first anniversary, Fiona Latham summarises some of the queries that the NDAD Helpdesk has received from its users since its launch in March 1998.

---

- [What is NDAD?](#)
  - [What data does NDAD have?](#)
  - [Do you have information about my family history?](#)
  - [Does NDAD have military records?](#)
  - [How do I register with NDAD?](#)
  - [Why do you need my address on the registration form?](#)
  - [How much does it cost to use the NDAD service?](#)
  - [How do I change my password?](#)
  - [My password does not work - what do I do?](#)
- 

## What is NDAD?

The National Digital Archive of Datasets (NDAD) exists to preserve and provide access to an emerging category of public record - databases and other forms of structured data created by Government departments.

NDAD is run by the [University of London Computer Centre](#) (ULCC) and the [University of London Library](#) (ULL) on behalf of the [Public Record Office](#) (PRO).



## What data does NDAD have?

NDAD is a service that holds computer datasets which were used, or continue to be in use in British Government Departments and other similar bodies that are subject to the Public Records Act.

A 'dataset' is a computer file or related set of computer files, which is organised under a single descriptive title and is capable of being described as a unit in the finding aids. It may comprise one or more accessions. The information held within a dataset can naturally be represented as a series of tables containing columns for particular types of information, and rows for each instance of the data.

A current list of NDAD's datasets is available at the [List of Catalogues by Department](#).

As well as holding the data itself, NDAD provides essential metadata which describes the relationships between data items, helps the systems present data on-screen to users and allows for export of data to modern software systems. NDAD also contains documents relating to these datasets, covering such areas as system design and uses made of the data.

For further information about the data and finding aids refer to the [Accessing Data](#) information page.

The archive is not primarily for collections of electronic documents - an area addressed by the PRO's [EROS project](#).



---

## Do you have information about my family history?

No - NDAD does not currently hold data about genealogy and family history. Our website offers a service that holds data from British Government Departments such as surveys and databases. A list of our databases can be seen in NDAD's current [list of catalogues by department](#)

The [Public Record Office](#) (PRO) web site has a link on its home page to a section called 'Readers' which provides [Information for Genealogists](#)

The Family Records Centre holds the following records

- Indexes of births, marriages and deaths in England and Wales since 1837
- Indexes of legal adoptions in England and Wales since 1927
- Indexes of births, marriages and deaths of some British citizens abroad since the late 18th century, including deaths in the two World Wars

If you are interested in these records, you can phone the Family Records Centre (0181 392 5300) or write to them at:

The Family Records Centre  
1 Myddelton Street  
London EC1R 1UW

Alternatively, you could look at the [earl web site](#) which is the on-line starting-place to find information about materials in public libraries which will help you trace your family history in Britain and Ireland.



---

## Does NDAD have military records?

No - if you are looking for military records I recommend that you visit the [Public Record Office](#) website; or email your query to their [reader services division](#)



---

## How do I register with NDAD?

Full details of how to register with NDAD are detailed on NDAD's [Registration help pages](#).



---

## Why do you need my address on the registration form?

We need your full postal address in order for us to process your registration. This is so that we can then send you a letter enclosing your username and password and your pre-filled form to check and sign and return to us. The information you provide to us in your registration form is strictly for our

own internal use so that we can contact you, send NDAD-related information, and any data you may request.



## **How much does it cost to use the NDAD service?**

It is **free to register with the Archive**. Once registered, you will be able to access the data online without any costs.

Any costs you incur will result from the use of our [chargeable services](#) (e.g. ordering copies of data, booking a Consultancy or paid research). You will be notified of these costs in the form of an invoice before you receive the service/items so you will have the option to cancel if you decide not to proceed with the order.



## **How do I change my password?**

Once you have received details of your registered user password, you can change your password to something more memorable and secure using the [Password Change Form](#). (You will be prompted to supply your original password before you are given access to the actual Password Change Form). The form will ask you to enter your new password twice. As soon you have successfully changed your password your browser will prompt you to enter it again. We advise that you use this form as soon as you receive details of your username and password from the Registrar. Thereafter, it is also advisable to change your password every six months or so. When choosing your new password please remember to use a combination of numbers and letters, using both uppercase and lowercase letters, avoiding dictionary words and variants on your name. If you require further assistance with changing your password, or if you should lose or compromise your password, please [email the Helpdesk](#) or telephone the Helpdesk on 0171 692 1212.



## **My password does not work - what do I do?**

Firstly you need to ensure that you are typing your password exactly as it appears in our letter to you that details your username and password. The username and password are always written in lowercase. The username always features a number at the end of a series of letters. The password issued to you consists solely of letters. If, after testing your password, you continue to be unsuccessful in gaining access to the data, please [email the Helpdesk](#) or telephone the Helpdesk on 0171 692 1212.



# Personnel Profile

Apart from gently encouraging his colleagues to produce interesting newsletter articles and trying to impress on them the concept of a *deadline*, *Richard Davis's* main contribution to this edition of the newsletter is the following account of how he came to find himself in just such an enviable position.

---



I left school in 1985 and started work at the [Metropolitan Police](#) Department of Computer Services in [Putney](#), which benevolent organisation kindly undertook to school me in the use of [ICL](#) mainframes and databases. On arrival I was given a small pick-axe and a stack of punch-cards and set to work in the stygian gloom of the subterranean Cobol mines. From adamantine boulders of syntax I was taught to extract valuable lines of code and fashion them into intricate and unlikely systems of structured programs. One such system was the vast and trunkless Crime Statistics System. My colleagues were an eclectic and good-natured crew, and on one day a year the upper floors of the building held a commanding view of a small section of the Boat Race course.

In 1988 I went to ICL Defence Systems, to work on a procurement system for the Ministry of Defence in [Southwark](#), using the same systems and tools I knew from the Metropolitan Police. At about this time I also remember using an electronic information system in my local library, which combined searchable catalogues, local information, electronic mail: it too used technology I was familiar with, and I remember thinking how much more stimulating it must be working on such a system as that, rather than, say, one forecasting Army spending.

The following year I went to [Warwick University](#) to study English Literature. I had many excellent tutors, who helped make my studies enjoyable and successful, and besides the usual fare of such a course, I also got to study Italian literature and spend a pleasant summer learning the language in [Siena](#). I passed my three years at Warwick largely unmolested by computers; however, when essay deadlines were pressing, I was always in demand to help coax friends' purple prose from departmental PCs or their own quirky Amstrad word processors.

Sated (for a time) with study and impecuniosity, I had interviews with several government departments, even the secret service. Luckily, whatever MI5 thought of me is still classified, and I returned to Scotland Yard, this time in the Management Services Department based in [Marylebone](#). I worked on several projects (including a fascinating study of [hand-cuffs](#)), but most of all enjoyed working on a small Apple network for the (self-styled) Forms Unit Corporate Identity Team.

My next stop was the Metropolitan Police Forensic Science Laboratory (MPFSL, also known as SO7), near [Lambeth](#) Palace. Ultimately, I supervised a small group of staff supporting the scientists' many and varied computer systems: the work was a typical pot-pourri of systems administration on Novell, Unix and Windows, user support and training, installation, upgrades, repairs, data protection, and so on. Sponsored by enlightened managers, I enrolled on a part-time [MSc Computing Science](#) course at [Birkbeck College](#). The course was thorough and interesting, consolidating much of my experience and introducing many new (to me) aspects of computing and IT, particularly the emerging Web.

In 1996, I changed employers without changing my position or location: the MPFSL was annexed by the Birmingham-based Forensic Science Service ([FSS](#)). The then Home Secretary visited for the occasion, and everybody at the lab was given a framed certificate. After this merger, I worked on integrating systems from the two organisations, and on installing new kit, such as a capricious DEC

Alpha server for the office automation systems, and a network of Apple Power Macs for the National DNA Database. More interesting, though, was setting up an NT-based intranet server, and developing C programs to convert data for a CD-ROM database of forensic science abstracts, published by the FSS in partnership with [SilverPlatter](#) (but I wish I'd known Perl at the time).

In the Autumn of 1997, my interest in information systems and databases rekindled, the chance to work on the emerging Computer Readable Data Archive project (now NDAD) at [ULCC](#) seemed too good to miss. That's why I have spent the past year in [Bloomsbury](#), helping stoke the boilers of the SS NDAD, working on some of the many tasks necessary to present and preserve these important government datasets. And wasn't my joy complete, to find among my first tasks has been the rediscovery of something very familiar to me, re-examining that very same old Metropolitan Police Crime Statistics System which is where we came in, and which will shortly be available on NDAD. Instant karma, indeed.