



Datasets News

Welcome to Datasets News from the National Digital Archive of Datasets.

Following the recent information review, we are pleased to announce that full access to open datasets has been restored. NDAD collections can again be viewed, queried and downloaded directly online. 10MB are free of charge, and larger downloads can be ordered.

NDAD provides an important service to researchers across many fields and disciplines, and has for a long time been unique in the UK in offering this online querying facility for a wide range of archived datasets, as well as catalogues and related documents, seamlessly integrated with the datasets to provide a fully digitised service.

To view the datasets, use the Browse NDAD button: <http://www.ndad.nationalarchives.gov.uk/> on the homepage, or the Browse link: <http://www.ndad.nationalarchives.gov.uk/access/> in the NDAD menu near the top-right of every page.

By interrogating NDAD's datasets online, it is possible to analyse and research information gathered from a wide variety of departments and subjects (from Ancient Woodlands to Accidents in the home, Bats to Beer, and Crime to Coastal Defences). NDAD therefore provides an important service to researchers across many fields and disciplines.

Editor, The National Archives

National Digital Archive of Datasets now accessible to the public

NDAD is now fully accessible to the public at www.ndad.nationalarchives.gov.uk/about. It contains quantitative, qualitative and geographical data created by central government departments and agencies, as well as associated reports and documentation. The datasets date from the early 1960s, when government began to use computers more extensively and to store digital data on a larger scale. Some of the data, however, originates from the early twentieth century. Unavailable elsewhere, this material provides a unique insight into social, economic and environmental change, as well as

government policy. They are of special interest to researchers and students in the social sciences, statistics and contemporary history, but also more generally.

Databases have become central to government recordkeeping, so the NDAD material provides an important record of government decision-making. Accordingly, the datasets reflect the changing concerns of the government over the recent past, particularly the last twenty years. They cover a wide range of subjects, with some emphasis on education, health, environment and crime. The better-known data sets include the British Crime Survey, the Countryside Information System and the North Sea



Geographical Information System, but the collection has much more to offer the researcher. Access to NDAD via the internet is free. An online querying facility enables the selection, manipulation and analysis of data and information. Data can be downloaded to the user's computer for more sophisticated analysis by means of statistical packages.

NDAD supports its users with search tools, finding aids and descriptive documentation, which assist searching and provide valuable contextual information. They explain the methodologies behind data collection and the reasons why particular datasets were constructed. Thesaurus-based indexes further assist effective use.

The factual data stored in NDAD provides a valuable resource and can be reused for numerous research purposes, sometimes unrelated to the reason they were originally created.

Dan Gilfoyle, Academic Liaison Coordinator, The National Archives

Just a Flèche Wound

(Fencing injuries in the Home & Leisure Accident Surveillance System)

For this article I've chosen to investigate whether the NDAD archive has any sources of information about the sport of fencing. I began fencing a few years ago after trying it for the first time during a family holiday to Center Parcs. The Essex County Fencing Union website

<http://www.essexfencing.com/>, describes fencing fights as characterised by flurries of speed, highly aerobic movement, and lightning-fast blade movements.

Whoever wrote that obviously hasn't seen me fence!

It has to be said that Fencing is not a sport in which Britain is particularly strong in international competition having won just the one Olympic gold medal (Gillian Sheen in women's foil at the 1956 Olympics in Melbourne). A brief examination of the National Lottery Awards

<http://www.ndad.nationalarchives.gov.uk/CRDA/39>

database held by NDAD indicates that between 1995 and 2006 the sport received slightly more than £2.3million in lottery funding. Over a similar period the sport of cycling received nearly £26.5million, clearly that's why the cyclists win all the Olympic medals!



Given the nature of NDAD's holdings I didn't expect to find a considerable quantity of data on fencing but one angle that seemed worth pursuing was that of fencing related injuries and the safety of the sport. Although a combat sport, fencing is extremely safe, largely due to the protective kit worn by fencers. The chapter on fencing in *Epidemiology of Sports Injuries* by Dennis J Caine et al (1996), reports that there has been very little research on fencing injury rates and patterns but suggests that major injuries or injuries that could cause a lengthy absence from the sport appear to be relatively



rare. <http://www.exra.org/FencingChptr.htm> In recent years, Fencing's national and international governing bodies have introduced new kit safety regulations designed to protect fencers from potentially serious injury that could be caused by broken blades. The British Fencing Association's safety guidelines can be seen here.

http://britishfencing.com/British_Fencing.asp?PageID=100

One of the datasets held by NDAD is the Home & Leisure Accident Surveillance Systems

<http://www.ndad.nationalarchives.gov.uk/CRDA/58>

(HASS/LASS). This was a system for collecting information about home and leisure accidents that was set up by the Consumer Safety Unit of the Department for Trade and Industry (DTI) in 1976. It was discontinued in 2003. The main aim of HASS/LASS was to provide information that supported the prevention of injuries sustained by consumers in avoidable accidents in the home, or in leisure pursuits. The data demonstrated this by showing the nature and size of problem areas; allowing trends to be monitored; highlighting where in-depth research was needed; and measuring the success of initiatives to solve identified problems. NDAD holds a dataset containing data on around 5 million personal accidents gathered between 1978 and 2002. I consulted this dataset by carrying out a series of searches looking for occurrences of the words "fencing" and "sword" in the details of individual accidents.

As might be expected, many of the search results for "fencing" describe accidents with what can be described as the garden variety of fencing. A number of these could perhaps have been lifted straight from the script of a 1970s BBC sitcom. For example "mending garden fencing-hit finger with hammer" and "running away from a dog-caught finger on fencing."

Where the fencing related accidents actually involve swordplay there is far more emphasis on the play. These would seem to range from the fairly mundane – "playing fencing with brother using wooden swords was hit on hand with one" through the downright dangerous - "fencing with billiard cues his brother poked his eye with it and cut eye" to the bizarre - "banishing [sic] cutlery, made fencing gesture-penetrated thigh." I think that descriptions of these injuries do indicate that play fencing can be risky, especially when brothers are involved. My advice would be either to invest in some foam swords (much safer!) or find the address of your local fencing club.

http://www.leonpaul.com/acatalog/Childrens_Gifts.html

http://britishfencing.com/British_Fencing.asp?PageID=39

Using the HASS/LASS dataset, I did locate records of about 60 accidents related to fencing as a sport. Very few over a period of 24 years which does I think, underline the safety of the sport and also supports the conclusions of the authors of *Epidemiology of Sports Injuries*. In fact some of the injuries whilst fencing related, had nothing to do with the sport itself, for example, a car accident on the way to a fencing competition and an injury sustained in an impromptu game of football during warm up before a fencing class. One thing that does stand out is that the descriptions of many of the minor injuries suffered whilst fencing include comments such as "all protective gear worn". Only one of the injuries described appears to me to be serious although a couple certainly bring tears to the eyes: "fencing competition and opponent ran straight into victim hit face/ mask pushed into face squashing nose" and "hit in groin by sword whilst fencing" – ouch...

Jim Jamieson, NDAD, Archivist

Flèche: An attack in which the aggressor leaps off his leading foot, attempts to make the hit, and then passes the opponent at a run.



Record of Scheduled Monuments (RSM)

This is a fascinating and complex dataset which was used at the offices of English Heritage since about 1990 (when it was first set up) to around 2004, when the current snapshot was taken by NDAD. Our snapshot went live in January 2008:

<http://www.ndad.nationalarchives.gov.uk/CRDA/64/DS/1/detail.html>.

Monuments - or archaeological remains - are a fascinating subject, although this particular dataset may not be the first primary source you should look at if you want information on stone circles, barrows and cromlechs of the British Isles. You may prefer to start with published gazetteers and sources (many of which have been derived from the RSM in any case) that would be more informative and accessible. The RSM sees monuments primarily as legal entities. Certain archaeological sites in the UK are protected by law; English Heritage administer the formal process of scheduling (and rescheduling) these monuments in order to protect them. This was a lengthy process and involved correspondence with local authorities as well as the Secretary of State. To support this, the dataset doubled up as a workflow system, meaning that many of the tables and fields simply hold pro-forma data to be used in form letters; or were used to log stages of completion in the process.

Another caveat is that the unit of description, as it were, may not always be a single monument. A scheduled monument is really just a legal construct, which doesn't always reflect archaeological classifications. This construct may be multi-period and may include all or part of more than one class of monument. For these purposes, the Department worked with separate entities such as 'archaeological items', 'constraint areas' and 'management zones', all of

which mean little to the countryside Rambler visiting a famous site like Avebury.



English Heritage also had dealings with owner-occupiers of sites, largely through Field Monument Wardens making formal inspection visits to the monuments to ensure that they were being well looked-after. This data is in the RSM dataset, and in some ways constitutes the most unique part of the record. However under the FOI act, the names and address of the owner-occupiers are closed for 85 years. Many of the texts of the reports are also closed for the same reason, but a fair number are open too.

Nonetheless, an NDAD user can extract useful and interesting data from this dataset. As it's such a large and complex dataset (75 tables), the remainder of this article gives you a few suggestions to help you hack through the thicket of information, most of which will probably not interest you. We can disregard system tables, as they were simply in place to help the database work, or run batch commands. We can also disregard workflow tables and fields, unless you're particularly interested in following the progress of a single case.



This leaves us with a number of tables which contain information about ancient monuments. But just to make it even more interesting, the system contains two distinct types of monument data. There are monuments scheduled under the Monuments Protection Programme (MPP), and other monuments scheduled under something called the 'Old County Number' (OCN) system. MPP and OCN even have two different numbering systems. All of this reflects evolution in the way the Department used to work. For their own reasons, they chose to include both 'new' and 'old' records in the single database system, expressing this as two different 'data domains'. While these domains may have been clearly separated in the original Oracle system, it is harder to discern in the NDAD rendering of the dataset where all the tables have been 'flattened out'. (We have however tried to untangle the situation in our descriptive catalogues).

For our purposes, let's just interrogate the MPP data domain. For all the reasons stated above however, there isn't a single table or field where you could search for, say, all the monuments in a single county that interested you. But start with table 33, and query the MONUMENT_NAME field. This worked perfectly for one of my personal favourite sites, The Rollright Stones, which turns out to be monument number 28160. You could also retrieve records which contain the phrase "stone circles", and if you see one in the resulting hit-list which you want to follow, note down the Monument Number. For this exercise, let's say I'm interested in 22450, the Oddendale concentric stone circle.

(Incidentally table 33 contains 42 fields, most of which are administrative trigger dates and casework fields; you may want to limit the display to filter these out).

Using that number 22450, query table 19, which will tell you the archaeological items associated with the monument. Archaeological items are specific to individual, archaeologically defined sites, and English Heritage recorded interpretative descriptions and an assessment of importance for each archaeological item contained within a scheduled monument. This search will yield the item number(s). For Oddendale, there's only one - it's 125387.

Now you've got the keys you need. This Item Number 125387 can be used to learn things about the Monument from the other tables. For example:

- Query Table 3, which tells you the National Grid reference for the Monument, its height, and its class.
- Query Table 50 to learn which historical period it has been assigned to. Oddendale turns out to represent both Bronze Age and Neolithic periods.
- Query Table 32 to learn something about the stability and vulnerability of the Monument, and the land use in the surrounding area. This query shows us that Oddendale is reckoned by EH as three separate 'management zones'. Although these are related to the Archaeological Item records, they do not necessarily match any of the other spatial zones defined as constituting a monument.
- You could also query table 69, which tells you the printed sources that relate to the monument, although you'd also have to query table 65 to learn what the sources are. This tells me that I could find something about Oddendale on page 178 of the *Trans Cumb & West Antiq & Arch Soc. Old Ser*, Volume VI. (At this point you would probably have to start looking for a good reference library.)
- Querying Tables 21 and 20 would tell you what events had taken place, and what



they were. These events are almost exclusively to do with observational visits from official inspectors. For example, with Oddendale a field observation took place on 18th March 1992.

Finally, we come to Table 71. This table holds the lengthy descriptive text reports, and constitutes the most interesting and unique part of the dataset. Just three things to bear in mind:

- Over 500,000 records in the table are closed, for reasons as above.
- Some of the texts in this table aren't reports at all - just standard texts for use in pro-forma letters.
- Retrieving the data to make it display properly is a bit difficult.

The latter problem is caused by the way the database was built. Each line of a report is held as a *single record*, so you will need to access a number of separate records in order to get the full report, and then sort them into numerical order to make them display as a coherent text. (Further advice on how to do this is included in our catalogues).

I couldn't get anything for Oddendale this way, but I did better with The Rollright Stones. To demonstrate, run this query:

```
TEXT_CODE = Monument Description
TEXT_IDENTIFIER = 28160A
VERSION = 4
```

This yields 69 records. Download this from NDAD as a CSV, sort it by the SEQUENCE_NO field, and you get a fine report whose first paragraph reads like this:

The monument includes a stone circle, portal dolmen, standing stone, round
cairn and a ditched round barrow which form the core of the megalithic sites
known collectively as the 'Rollright Stones', situated 900m north east of
Little Rollright on the Oxfordshire - Warwickshire border and which lie within
three areas of protection. The monument was one of the 26 monuments protected

in the original Schedule for the 1882 Ancient Monuments Protection Act, and
was taken into state care in 1883. Although partly obscured by trees and
hedges today, the monuments would all originally have been inter-visible. The
Rollright Stones lie on the scarp of the Cotswolds on the crest of a ridge
running from south west to north east. They have attracted later activity
including burial mounds, an Iron Age settlement immediately to the north east
(which is the subject of a separate scheduling) and an Anglo-Saxon cemetery.

As you will see, the designers of this aspect of the database were intent on imitating the look and feel of a word-processed document, or a page from a book. This means that some records are left *blank*, in order to act as paragraph separators!

Your only problem is how to connect this text description with the monument you're interested in. Table 71 does it through the owner-occupier field, which relates to Table 48 - whose data are either closed or become anonymous for 85 years. (The way I found the text identifier in the example above was through searching for 'Rollright' in the field TEXT_LINE).

In summary:

- To best access data in the dataset, you need to find the unique ID fields for your monument, and look them up in other tables.
- Use Table 33 to find the Monument Number, and Table 19 to find the Item Number.
- If all you want to do is query the reports, then search Table 71 as suggested above.

Ed Pinsent, NDAD, Archivist

For further information on Scheduled Monuments.

<http://www.englishheritage.org.uk/server/show/nav.1369>



“THE LONGER, THE BETTER” A celebration of long-term datasets in commemoration of the 300th anniversary of the birth of Robert Marsham, FRS

*"The value of long-term **environmental datasets** is being increasingly recognised in the light of climate change and other drivers of environmental change. Robert Marsham FRS, was one of the pioneers of long-term data gathering when he started recording the "Indications of Spring" on his Norfolk estate in 1736, subsequently continued by his descendants until 1958. This meeting aims to review the various uses of long-term datasets for understanding our environment and to discuss the issues surrounding their maintenance and development."*

<http://tinyurl.com/9w29cw>

In late October, I attended an extremely interesting and informative seminar sponsored by The Linnean Society of London and The Centre for Ecology and Hydrology, in commemoration of the 300th anniversary of the birth of Robert Marsham FRS (father of British phenology, and very much the David Attenborough of the 18th century!)

[\(http://www.robertmarsham.co.uk/\)](http://www.robertmarsham.co.uk/)

To my delight as a Digital Archivist, presentations and panel discussion focussed around 'A celebration of long-term datasets'. Hundreds of years after Marsham's own seasonal recording of nature's distinct events, our present day global warming has prompted an even greater need to know how the natural world is responding to climate change, particularly in order to predict consequences for the future. So amongst the rich extent of their value, long-term datasets are shedding

important light on environmental change by helping us to predict, measure and manage that change. They also serve as convincing indicators that climate change is affecting us *now*.

Long-term environmental data has hidden and far-reaching value. The day explored key issues surrounding their repurposing, curation and preservation. The meeting attracted a wide range of practitioners, such as researchers, data custodians/creators, as well as practicing ecologists and policy makers who use the data within the government and private sectors. It was also a rare opportunity to network and hear firsthand the key research concerns and interests that our potential 'NDAD users' have, when it comes to long-term environmental data.

In his introduction and talk on Robert Marsham, Professor Tim Sparks commenced with an interesting example of unique long-term data, the "Oak first leafing: Surrey 1950-2008". It's apparently the only data of its kind in the UK and was collated by one independent ecologist ('Jean' who is now 80 years old!). Professor Sparks drew to our attention, what an amazing contribution one person can make. The data is unique in the length of the period it covers, and it was collated under her own volition, without government support. Much crucial long-term data would not be available today if not for the efforts of amateur and independent data collectors. The panel discussion later reflected on the need to engage independent data creators in standards and data sharing.

NDAD observation:

.... In light of his obsession with trees, had he lived today, I'm confident that Marsham would have been fascinated by NDAD's own Ancient Woodland Inventory, an inventory of woodland sites over 2 hectares in England which are thought to have been in existence since at least 1600 AD!

<http://www.ndad.nationalarchives.gov.uk/CRDA/43>



Long-term environmental datasets have current and untapped potential. They also have specific value and uses within the context of a diversity of research domains. Examples include the specific value of long-term datasets relating to phenology; distributions; population modelling; monitoring for pollution; monitoring by conservation agencies; marine resources; and land use. Impressed by the wealth of useful information presented at this seminar, it's my intention to individually explore some of these specific research contexts over coming issues of the NDAD newsletter. In the meantime, I hope to address some of the common themes that were prominent or frequently addressed by the majority of speakers during this celebratory event.

Did you know?

The longest human record comprises "1300 years of Cherry flowering at Kyoto". Unfortunately, the last 20 years has witnessed the most drastic change and drop in flowering over 1300yrs of data collection!



VALUE OF LONG-TERM DATASETS?

"What is the point in having long-term datasets?"

Professor Fitter raised this crucial question, one which is uppermost in the minds of digital archivists. He also considered some key questions concerning 'the value, serendipity (unforeseen value), and loss of long-term data'; while providing excellent examples of how long-term data is being used to monitor climate change e.g. data showing rapid increases in carbon dioxide.

Among the many good reasons identified for creating and maintaining long-term datasets, Professor Fitter noted:

- Trends are more discernable with long-term data ('you need data to span at least 50 years sometimes to understand long-term trends and discern changes')
- Rare or extreme events tend to determine shifts in behaviour
- Long-term data can be used to detect cycles (seasonal/long-term)
- Keystone species have long life spans
- Without such datasets, we're unable to predict future questions and concerns!

The value of long-term datasets for ecological science, and as an evidence base supporting decision-making by conservation policy makers, was recognized by all speakers. The unique value and uses of long-term datasets within their specific research domains was also discussed. For example, the value of long-term data for:

- Distinguishing short-term from long-term trends;
- Demonstrating how species can respond differently to the same environmental pressures;
- Helping to expose serious issues in habitat management and the decline in species;



- Studying the dynamics of significant population changes over time;
- Assisting regulation/compliance, identifying hazards and risks, temporal/spatial variation, and resilience;
- Supporting strategic management, meeting international obligations, educating the public and government about the environment, and offering advice to government departments.

Long-term data can play a vital role in assisting conservation agencies with monitoring, assessment, reporting, target/priority setting, and to serve as vital evidence for government policy, according to The Chief Scientist for Natural England, Dr Tom Tew. The social and economic trends derived from this data are just as important as the environmental data. Natural England are starting to measure how people engage in the environment, and the ensuing health benefits.

REPURPOSING DATA

Archived environmental datasets have been put to unforeseen uses, particularly in relation to climate change. These datasets can be reused for numerous research purposes, sometimes unrelated to the reason they were originally created. This is often the case for many of the central government datasets currently archived and made accessible via our NDAD service.

<http://www.ndad.nationalarchives.gov.uk/CRDA/43>

'Whale catches over time' was one data source noted by Professor Fitter. While the data was initially collected by commercial whaling companies to increase profits and returns, this same data can now be used to demonstrate the impact of excessive harvesting and changes to the whale population.

The 'Park Grass Experiment, Rothamsted' is another example of reuse. Originally collated 100 years ago, as part of efforts aimed at

deciding how best to fertilize grass to increase yields, this data is now used 100 years later for a number of different purposes. The archived samples of park grass, continuously collected at regular intervals since 1856, can be used to examine the dynamics of park communities. From grass samples, you can extract microbe DNA and study microbial communities over space and time.

While some of this Victorian survey data is published, according to Professor Chris Thomas, much of it isn't brought together to enable people to go out and undertake repeat surveys, and continuous monitoring. (Hence, a real need for repurposing and cross searching of data). Information from repeat surveys [such as those we hold at NDAD] is very important. Professor Thomas advocates compilation of the surveys/data in order to make this a reality, to study colonization patterns based on long-term data where colonization extinction dynamics can also be seen. From long-term historical data, you can deduce that species are expanding or declining.

Dr Tom Tew spoke of how long-term data (originally collected for other reasons) has been re-used by Natural England to demonstrate the need for a dynamic site-based approach to its conservation monitoring. Subsequently, when speaking on monitoring marine resources, Dr Simon Jennings flagged up a number of interesting media articles, all showing data being used for a very different purpose than the original survey design intended. For example, as an unexpected value of the original scheme, the long-term UK data on plankton has revealed evidence of major shifts in distribution, phenology and abundance.

In a presentation on the value of long-term data for population modelling, Professor John



Harwood and Dr Stephen Baillie alerted us to the untapped potential of cross searching and integrating data across datasets. The trends in survival rates and productivity of populations of the British grey seal is one example.

SUSTAINING THE RECORD: PRESERVING DATA OVER TIME

"How do we sustain the record?" (Another fundamental question raised by Professor Alastair Fitter). Many datasets have lasted 50 or more years, but some data has been lost. He argued that we need to consider:

- Issues of stability and resilience, by encouraging ecologists to think long-term
- Providing incentives, like funding regimes
- Working with and supporting independent naturalists (who do it for the love of it) and creating new networks to engage them
- Creating new experiments (but considering them carefully)
- Sustaining existing records and maintaining existing long-term projects
- Aligning with research council priorities and capitalizing on existing resources like the Environmental Change Network
- Archiving data and samples!

NDAD observation:

While some data archives were mentioned during the day, interestingly the transient nature and technological obsolescence affecting born-digital datasets was not explored as a key concern. Whereas paper-based datasets have lasted for hundreds of years, we are now challenged by legacy systems and datasets disappearing at a rapid rate, and so we need preservation strategies and standards in place (not just for data creation and collation), to ensure data can be maintained and made meaningful over the long term.

As most presentations predominantly drew upon long-term datasets created and maintained in paper form, I took this

timely opportunity to highlight to fellow attendees that NDAD was developed as the government initiative it is, (i.e. as part of a digital preservation strategy) to ensure that structured born-digital data from central government departments is salvaged, preserved, and made accessible over the long term.

How do we present a case for preserving long-term datasets to funders? In their talk on population modelling, Professor John Harwood and Dr Stephen Baillie argued that *'we need to work together'* and to further examine if we are *'enabling organizations, via modern technology, to do their job faster and better'*. Dr Simon Jennings also raised the ongoing challenge of government short-term projects conflicting with the need to focus on the long-term in order to generate time-based series.

How do we sustain the long-term data record? *"How do we mainstream a healthy natural environment into political thinking?"* There was general consensus that we need increased public understanding of the importance of long-term data, and that reporting on data can help increase government and public attention to the issues. For example, the increasing public resonance resulting from *'Our dying sea'* (<http://www.independent.co.uk/environment/our-dying-sea-680091.html>) - an article that was based on monitoring long-term data from the 1800s. Increasingly, media headlines are drawing upon the monitoring of long-term data, and this same data is being used as indicators for government; essential to leveraging public funds. How we can come up with an indicator of something like 'ecosystem resilience', as well as being able to measure it?

CONCERNS AND FUTURE CHALLENGES:

There are gaps in research and gaps in historical data. Professor Chris Thomas noted that while distribution datasets are helpful for monitoring climate change, there is a *"lack of availability of*



good 'change in land use data' and lack of availability of data on environmental layers". Speakers and attendees alike reaffirmed that researchers would like more data on land use. Dr Simon Jennings spoke of the difficulties in obtaining data relative to the terrestrial environment. Professor Chris Thomas argued that, as we need more evidence to convince government of trends in climate change, we therefore need more datasets!

As part of the concluding panel discussion, the following concerns and future challenges were raised by attendees:

- The nature of funding keeping long-term datasets going;
- The level of enthusiasm for keeping datasets going;
- Detective work needed to make systems and data more interoperable; the need for more metadata standards (SPIRE and mapping GIS data concern); are Ordnance Survey's commercial drivers getting in the way of interoperability?
- Curatorial issues (who keeps the data when it's so scattered?), and the need for more joined up thinking;
- The need for promoting similar standards for data storage;
- What should drive the creation of long-term datasets?
- Make economic imperatives of government to create these long-term datasets.

Closing comments of the day attempted to provide solutions to some of these challenges and concerns:

- *"Local natural history societies should openly report on their websites what they're collecting data about."*
- *"We need to persuade people they are important, their data is important; and that it*

won't be important at all if they don't curate and maintain their records". We need data and metadata to be deposited.

- *"We need to offer support where there is a gap of information in dataset documentation and ask custodians about data before it's too late"*
- *"Tools for capturing data are extensive (unlimited storage for free!)"*
- *"People need a reason to rescue data..."*
- *"Connect people with the way their data is used"*
- *"Local authority/BDN (Biodiversity Network) i.e. County offices available for people to submit a diversity of data to (as places of deposit)."*
- *"If a decision is made to discontinue a survey or throw data away (as happened with soil sample data), this should be exposed to the wider community".*
- *"We need to stop the data losses!"*
- *"We need to build a climate change momentum by being better at communicating the importance and wealth of data we have – communication is the key."*

NDAD observation:

As a fellow attendee, I added that central government datasets are public records and vital evidence stemming from The National Archive's legal mandate, to preserve and make accessible, central government datasets. [The National Digital Archive of Datasets (NDAD), operated by ULCC on behalf of TNA, preserves



and provides online access to archived digital *datasets and documents* from UK central government departments]. I reaffirmed that it's important that data creators archive not just data but also crucial metadata such as dataset documentation in the form of field reports, data dictionaries and data models; in order to help us preserve and make the data meaningful over time. Long-term environmental datasets currently held by NDAD include:

- British Bats (CRDA/17)
<http://www.ndad.nationalarchives.gov.uk/CRDA/17>
- Ancient Woodland Inventory (CRDA/43)
<http://www.ndad.nationalarchives.gov.uk/CRDA/43>
- International Designations - European database ('Natura 2000') (CRDA/57)
<http://www.ndad.nationalarchives.gov.uk/CRDA/57>

By interrogating NDAD's datasets online, it's possible to not only analyse and research environmental data, but also information gathered from a wide variety of departments and subjects (from Ancient Woodlands to Accidents in the home, Bats to Beer, and Crime to Coastal Defences).

The work of Robert Marsham has survived to inspire and inform modern day eco-conscious environmentalists from across the world, but today's Robert Marshams are posed with a much greater risk of vital data being lost to posterity, in the face of uncertain funding, attitudinal barriers, and rapid technological change. This seminar however represents a significant step towards harnessing and promoting the multi-faceted value of long-term environmental data, as well as helping to address some of the many challenges facing its ongoing curation and sustainability.

Joanne Anthony, NDAD, Archivist

Allied Sources:

A summary report of the seminar, as recorded by some of the keynote speakers:

<http://www.linnean.org/fileadmin/events2/events.php?detail=127>

Full Programme of the day's event:

<http://tinyurl.com/9w29cw>

Images used in this issue under Creative Commons:

WC fencing 2006 foil team by Wisze (retrieved from

<http://www.flickr.com/photos/wisse/272559014/>)

Avebury by pl_jakub (retrieved from

<http://www.flickr.com/photos/75227967@N00/182418718/>)

Sakura in front of temple gate by Kimtaro (retrieved from

<http://www.flickr.com/photos/kimtaro/422166129/>)

