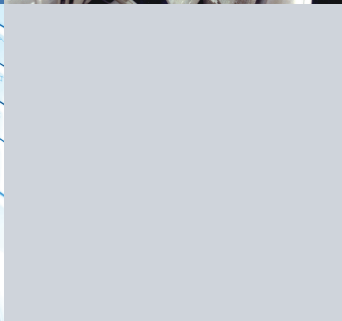
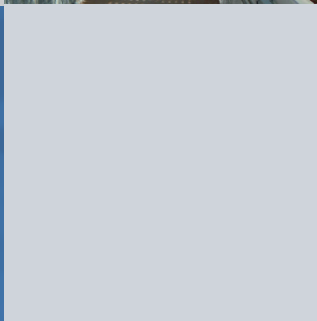
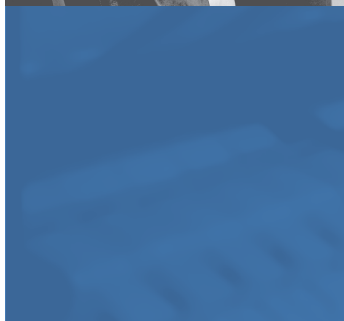
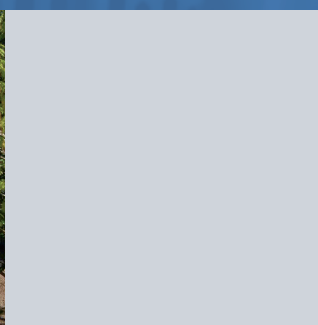




40<sup>th</sup> UKDA  
1967-2007



**ACROSS THE DECADES**  
40 years of data archiving





# CONTENTS

Introduction	1
Origins	2
Environment	4
Profiles	8
Acquisitions	10
Processing	14
Dissemination	18
Support	22
International	26
Research	30
Family	34
Future	36
Timeline	ibc

[www.data-archive.ac.uk/ukda40](http://www.data-archive.ac.uk/ukda40)

Note: Although the name of the UK Data Archive has changed over time, for simplification it will be referred to as 'the Archive'.







# INTRODUCTION



The research of social scientists affects us all at every stage of our lives. Their findings influence government policies on childcare, education, wages, health, and pensions. For much of this research, social scientists are dependent on infrastructural support of a quite different nature than that needed by other scientific disciplines. Whereas a particle physicist needs a large collider and an astronomer needs a large telescope, a social scientist very often needs access to large collections of quality data. Data painstakingly collected from surveys, questionnaires or interviews for one study, can be analysed again for an entirely different study, providing that accurate records have been kept.

Recognising the need to preserve and curate high-quality data for secondary analysis, the Social Science Research Council (SSRC) (as the Economic and Social Research Council (ESRC) was then called) supported the establishment of the Archive in 1967, and has provided the long-term commitment of funds, vital for its success. The ESRC also leads the way in providing data for the Archive by investing in large-scale studies to create

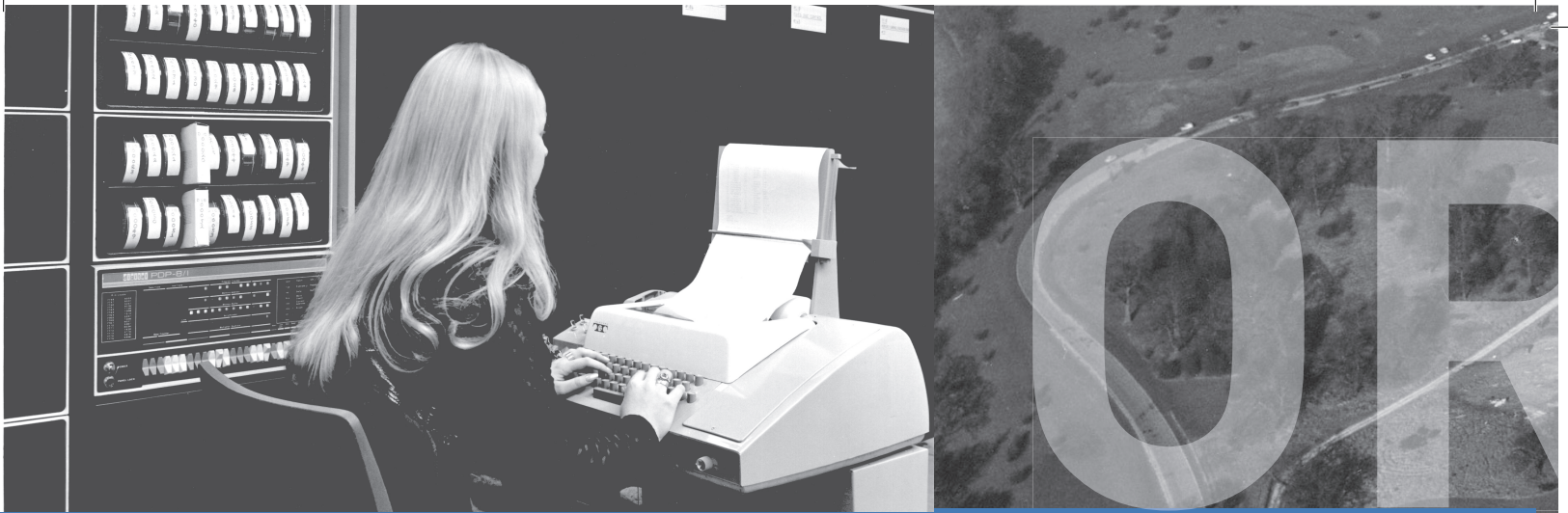
valuable datasets, and by requiring all data collected with ESRC funds to be offered to the Archive for secondary analysis.

The Archive has proved more than worthy of this investment. Throughout its history, it has remained at the cutting edge of developments to become an international leader in the field of management and preservation of data resources. As a result, the Archive is well placed to take advantage of any opportunities created as technology continues to develop at an increasingly rapid pace. This publication celebrates some of the Archive's many achievements to date, and the services it provides to the social science community in the UK and beyond. Today's Archive is a credit to all the directors and staff who have worked hard to ensure its success, and to the University of Essex for its support in hosting the Archive over the past forty years.

**Professor Ian Diamond**

*Chief Executive, Economic and Social Research Council*





# ORIGINS OF THE ARCHIVE

Funding for the SSRC Data Bank, as it was then called, commenced on 1 October 1967. However, the origins of this enterprise predated this by several years. In December 1963 the Social and Economic Archive Committee (SEAC) was established to investigate and propose solutions to the problem of sharing information about social surveys and the data generated by them. SEAC was hosted by Political and Economic Planning (PEP) – the British policy think-tank established in 1931 and the forerunner of the Policy Studies Institute. It was well supported and well connected, receiving funding from the London School of Economics (LSE) and the fledgling Social Science Research Council (SSRC), as well as industry and government.

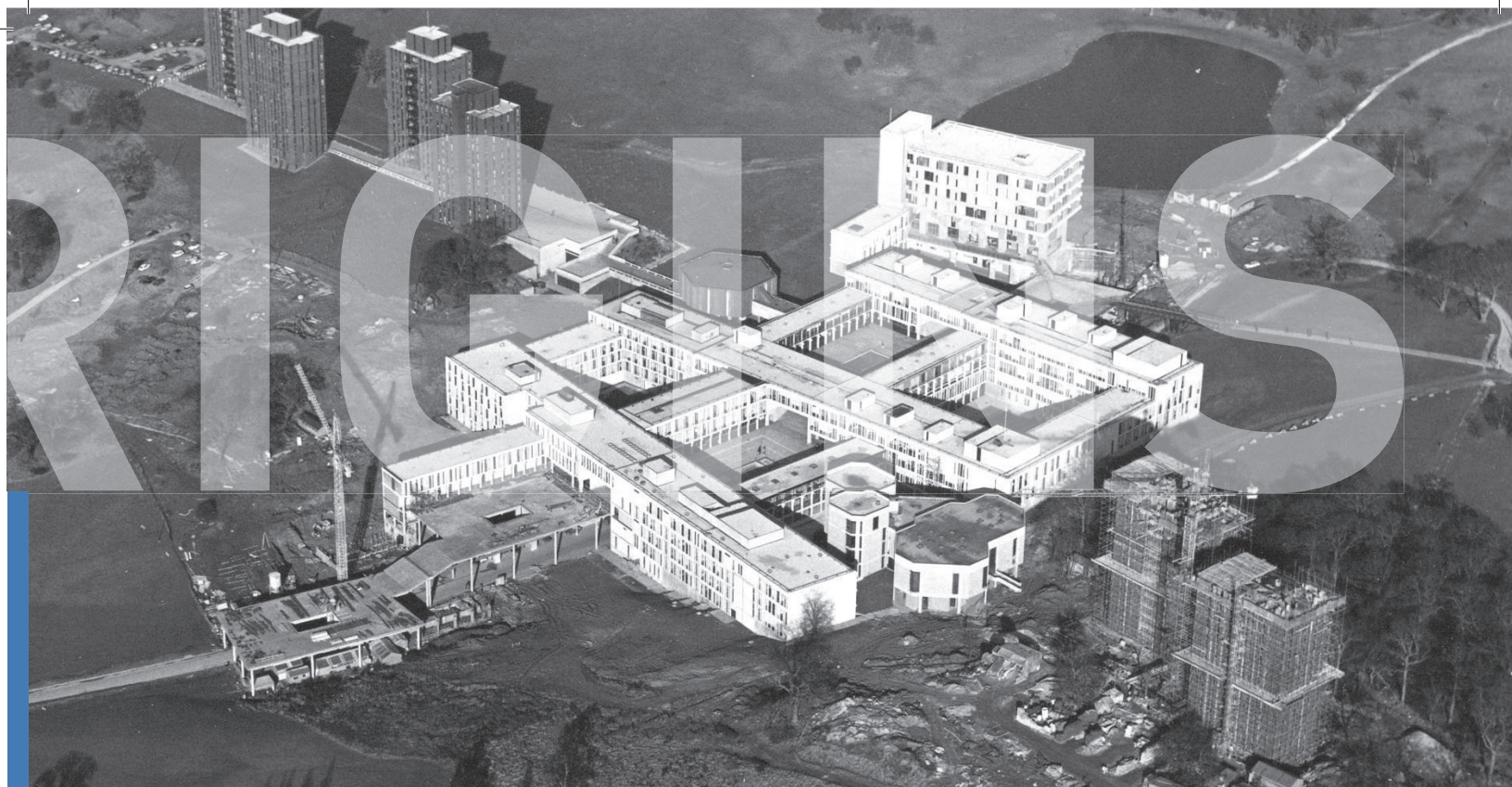
SEAC was particularly concerned that costly and time-consuming survey work was being replicated through poor communication between researchers and that data were being sold to the US, and therefore 'lost' to British researchers. To meet this concern SEAC compiled an inventory of survey data which could be made available for secondary research. Interestingly, this concentrated on market research and other commercially produced data rather than university-based work. Also within SEAC's brief was to explore options for the location of an archive dedicated to social and economic research. The sub-committee for this work comprised John Johnson (Manchester), Philip Abrams (Cambridge), Michael Young (first chair, SSRC), Richard Lipsey (Essex) and John Madge (PEP). In March 1966, they invited proposals for the establishment of an archive.

At a meeting on 23 May 1966 the sub-committee considered three proposals – from the University of Essex, PEP and from the SSRC itself (Strathclyde University had been considered a candidate, but did not submit a proposal). The proposal from PEP was not realistically worked out in any detail, its case resting primarily on the argument that such a national resource should be located in London. The SSRC's bid, from Michael Young himself, was more interesting.

Young's bid arose from a conversation that Young had with Stein Rokkan (Professor of Comparative Politics, Bergen) following the 3rd Conference on Social Science Data Archives held at PEP in London earlier in 1966. Young's proposal for the SSRC to run the archive directly is intriguing, especially in light of forty years of history. First he felt that, as one of the functions of an archive should be to provide methodological advice to prospective SSRC applicants, it would carry more weight if it were run by the funders. Second, he believed that the SSRC would be more likely to attract and retain more highly skilled employees. Third, he thought that, as an archive should hold 'ecological data' – government data drawn from the census and surveys such as the Family Expenditure Survey – the SSRC would be better placed than a university to acquire these. Fourth, he argued that the SSRC would be 'neutral' in servicing and training the university sector, ensuring that it was used by all.

The main drawbacks to Young's proposal were the higher costs of locating the facility in London, and the lack of computing facilities at the SSRC. To counter the second issue Young tried to establish a collaboration with the LSE through Claus Moser (who was very much against the archive being housed at Essex), in the hope that the computing activities could be situated there, but this came to nothing.





Despite Young's credentials, the sub-committee came out in favour of the proposal from the University of Essex, which was judged to address all the stated criteria. It could be established quickly, and provided extremely good value for money, given that the University offered to meet the costs of five staff (not all full-time) and provide both accommodation and computing time.

In response to Young's submission, the sub-committee noted that the Archive "once recognised as a truly national centre, would in any case develop the functions Dr Young had in mind ... by existing, the archive would introduce an element of self-discipline into the design of future programmes". It also recommended that "a very close relationship, short of direct control, between the archive and the Social Science Research Council is highly desirable".

After this decision Essex was invited to submit an application to the SSRC. This was sent on 10 October 1966, with Allen Potter, then Head of the Government Department, named as the Principal Applicant. The application requested a total sum of £33,000 over five years, including £4,500 per annum for staff costs, £500 for travel and £1,000 per annum for magnetic tapes. A year later, the SSRC Data Bank started work, and the rest, as they say, is history. As Eric Roughley, a key employee of the Archive for some 25 years from its first day of operation, put it on the occasion of its 20th anniversary – "the bold decision to establish an archive in 1967 has been resoundingly vindicated".



**'I regard this development in the social science field as probably the most important of the next generation.'**

ALLEN POTTER, *Evening Standard*, 31 January 1967







# THE SOCIAL SCIENCE ENVIRONMENT –

Established in 1967, the Archive was first named after the Social Science Research Council (SSRC) – the source of its initial (and ongoing) financial support. Since then, its name has changed no less than five times, each change reflecting the social science environment at the time. Even the Archive's birth name, the SSRC Data Bank, was not the name first conceived.

Early proposals were for a 'Social and Economic Archive' but Michael Young, then Chairman of the SSRC, thought the Archive should bear the Council's name to instil confidence in the Archive's longevity. And according to David Allen, his colleague at the SSRC: "Young disliked the word 'archive', feeling that it had a dispiritingly musty ring to it". He preferred 'data bank' – a term that unfortunately acquired somewhat sinister overtones soon after.

In 1972, to increase the Archive's appeal to potential depositors, it became the Survey Archive. 'Data' reappeared in the Archive's name in 1982, reflecting

the broadening range of its data holdings as social science became less survey-orientated. In the eighties, government antipathy towards the social sciences led to pressure on the Council to drop 'science' from its name. It became the Economic and Social Research Council (ESRC) in 1984, forcing the Archive to change its stationery for the second time in only two years.

The next name change followed government calls for stronger links between academics and the users of their research. The Archive responded, partly by increasing its support for teaching and learning, and was awarded direct funding from the Joint Information Systems Committee (JISC) in recognition of this new emphasis. With a broader support base, came a more generic name: 'The Data Archive'.

The latest (but perhaps not last) change to the UK Data Archive in 2000, reflects the increasingly global nature of social science research and the Archive's strong international focus.



## 1960s – THE DATA BANK

Although the 1946 Clapham Committee recommended that public funding of social science research in universities expand, by the early 1960s total research funding for the social sciences from all sources stood at only £5 million and there were still as many historians teaching in Oxford as there were sociologists teaching in universities in Britain as a whole.

The expansion of universities following the Robbins Report of 1963 and the establishment of the Social Science Research Council (SSRC) in 1965 led to the expansion of social science research within universities and it was within this context that the SSRC Data Bank was set up at the University of Essex in 1967.



## 1970s – THE SURVEY ARCHIVE

The 1970s was a period of growth in empirical social science research. By the mid-1970s approximately £50 million per annum was being spent on social research in Britain, half in universities and the rest in 'in-house' government research and independent research units.

In its early years the Data Bank experienced difficulties in populating its collection due to the high standards it required from deposits, an immature culture of data sharing, and the restrictions on use attached to certain studies particularly government surveys. The turning-point came in the early 1970s when the Government Statistical Service enabled government surveys to pass to the Survey Archive, as the Data Bank had been renamed in 1972.

## 1980s – THE SSRC/ESRC DATA ARCHIVE

The Survey Archive was renamed the SSRC Data Archive in 1982 in order to reflect the fact that a broader range of data resources, more than just surveys, was now being collected and stored.

# CHANGING TIMES, CHANGING NAMES

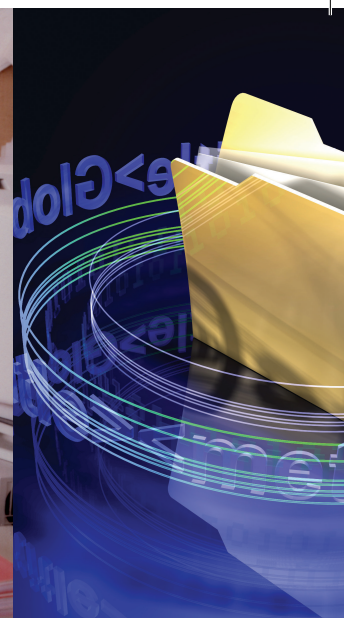


‘The Archive was referred to as ‘the Rutherford Laboratory of the social sciences world’ in a House of Lords debate on the Research Councils on 21 November 1984.’



‘Data is deposited with the Bank on a wide range of topics including such intriguing questions as “Did you go on a school visit to a coal mine?”

Wyvern, 16 February 1968



In the early 1980s government spending on universities in general and social science research in particular, was cut. As part of this process the Rothschild enquiry was set up to review the ‘scale and nature’ of the SSRC’s work. Despite government antipathy towards the SSRC, the resulting report was supportive of the Council whilst recommending a stronger focus on empirical research and research considered to be of ‘public concern’. Whilst Government accepted Rothschild’s main recommendation – that the SSRC continue – a consequence was that the SSRC was renamed, becoming the Economic and Social Research Council (ESRC) in January 1984. This resulted in a second name change to the Archive in two years – ESRC Data Archive!

Whilst the 1980s could be seen to be a low point for the social sciences, in retrospect, pressures on funding had both negative and positive impacts on the Archive – less was spent on primary data collection, yet this in turn encouraged increased secondary use of research data and a greater acceptance of data sharing. The 1980s also saw the Archive branch out through its involvement in a number of large co-operative data-orientated projects – key of which were the Domesday Project and the Rural Areas Database. This set a trend which has continued up to the present.

### 1990s – THE DATA ARCHIVE

The 1993 White Paper on Science and Technology led to an emphasis on wealth creation and the need to establish closer and deeper partnership between the academic community and users of its research. In line with this, the 1990s witnessed an extension of activity by the Archive. In 1992 the History Data Unit was formed as a specialist unit within the Archive, becoming part of the Arts and Humanities Data Service (AHDS) three years later (and as a result renamed the History Data Service).





‘The creation of Britain’s first memory bank on the computer at the University of Essex is a tremendous feather in the cap of the University...Hitherto the fame of the town of Colchester has rested upon the past, specifically its Roman background. From now on it will rest equally, if not more so, on the University.’

Colchester Gazette, 7 February 1967

In 1996 direct funding from the Joint Information Systems Committee (JISC) was received in recognition of the support provided by the Archive for teaching and learning. This led to a further name change with the dropping of the Council preface to become The Data Archive, together with a new colour scheme and logo being selected to co-ordinate with the new image of the University of Essex.

## 2000s – UK DATA ARCHIVE

This latest period started with yet another name change. In order to reflect both its UK-wide remit (post devolution) and the importance of its role within the international data network, it became the UK Data Archive (UKDA).

Qualidata, a specialist unit initially set up in the Department of Sociology at Essex in 1996 to acquire and support the secondary use of qualitative materials, merged with UKDA in 2001, both streamlining procedures and facilitating greater cross-over between qualitative and quantitative approaches.

The UKDA's portfolio was also further expanded by the creation of the Census Registration Service (CRS) in 2001. Although it had provided census data to researchers for 20 years prior to this, the creation of CRS marked the UKDA becoming a member of the ERSC/JISC-funded Census Programme to a range of UK census outputs.

The changing nature of the Archive's role within the social science community and the requirements of its users were reflected in the ESRC's green paper on 'Data Archiving and Data Resources' (2000) and resulted in the creation of a major new initiative in the form of the Economic and Social Data Service (ESDS) which came into operation in 2003. This formulated changes already underway focusing on a more user-driven service with specialist support and data enhancement for key data resources through a distributed service, based on collaboration between four centres of expertise: the UKDA and Institute for Social and Economic Research (ISER) at Essex, and the Cathie Marsh Centre for Census and Survey Research (CCSR) and Manchester Information and Associated Services (MIMAS) both located at Manchester.

In recognition of its position in disseminating and preserving an increasingly diverse collection of government data, from 1 January 2005 the UKDA became a designated Place of Deposit for public records for The National Archives (TNA), thus making the deposit of materials a legal requirement for the first time, and thereby ensuring the supply of key social surveys for future research.

Most recently, in the 40th anniversary year, together with ISER, the UKDA moved into a new purpose-built social science research centre costing £6.5 million.





# DIRECTOR PROFILES

The evolution of the Archive over the last 40 years, presented each successive director with different – sometimes quite unexpected – challenges. Their success in meeting those challenges has played a major part in shaping the Archive celebrated today.

The first director, Allen Potter, arrived prepared to tackle the challenge of setting up a new computerised archive, capable of storing large quantities of data, only to find that his major challenge (and that of his successor, Ken Macdonald) was actually finding any data to put into an archive. When the third (and to date the longest serving) director, Ivor Crewe, joined the Archive, he thought that acquiring major datasets and handling the ambitious plan of recording and categorising every single question asked in every single survey, were challenge enough. But, as he recollected later, he found himself “taken completely by surprise by the miniscule number of actual users of the Data Archive” and very quickly reprioritised his agenda to focus on boosting user numbers.

Under Howard Newby, the Archive rapidly expanded in size as a result of a number of new projects, so much so that the Archive building had to be modified

to fit in more staff. New projects and growth were also a feature of the Archive under Catherine Hakim and Denise Lievesley. Having come to the Archive from outside of the world of academia, both also brought new perspectives and worked hard to build stronger ties to external stakeholders, especially data-producing government departments.

The importance of external stakeholders still remains today but, as the community the Archive serves has become more diverse – one of the many challenges facing the Archive’s current director, Kevin Schürer, is to ensure that the network of activities in which the Archive is involved connect to provide the desired synergies. The accelerating pace of change in internet technology presents yet another challenge, as Schürer explains: “The Archive must be proactive rather than reactive – anticipating and preparing for new technology before it becomes established so that we are ready to roll out the solutions as soon as they are needed”. Indeed, expecting the unexpected was what most attracted Schürer to the Archive. “It is a wonderful time to be director because the Archive is facing a period of great change and I have the opportunity to influence that change”, he says.



## ALLEN POTTER

**March 1967 – September 1970**

Professor Allen Potter had been on the staff of the Department of Government at the University of Essex since 1965 when he was appointed the first director

of the SSRC Data Bank, also serving as a University Pro-Vice-Chancellor from 1969-70. He left Essex in 1970 to become Professor of Politics at the University of Glasgow, additionally serving as Vice-Principal 1979-82, a Governor of the Glasgow School of Art 1979-82, and as a member of the US-UK Educational Commission 1979-84. Retiring in 1984, he has subsequently lived in Worcestershire.

## KEN MACDONALD

**October 1972 – September 1974**

Prior to his period as Director, Ken Macdonald was Senior Lecturer in Government at Essex. His research interests included the development of computer packages for data analysis, social and political theory, and methodological issues relating especially to the study of social mobility, electoral behaviour and industrial relations. After two years as Director he moved to Oxford to work on the Social Mobility Project where he is currently Lecturer in Applied Social Studies and a Fellow of Nuffield College.



## IVOR CREWE

**October 1974 – October 1982**

Professor Sir Ivor Crewe first came to Essex in 1971 as a lecturer in the Department of Government. From being Principal Investigator for the British

Election Study in 1973 (which he continued until 1981) he moved to the Archive in 1974 to become its longest-running director, being in post a total of eight years. Moving back to Government, he served as Head of Department from 1984 to 1988 becoming the University’s Pro-Vice-Chancellor Academic in 1992 and Vice Chancellor in 1995. As Vice Chancellor he served as President of Universities UK from 2003 to 2005, receiving a knighthood for services to higher education in 2006.





### HOWARD NEWBY

**October 1983 – September 1988**

An undergraduate, postgraduate and later lecturer in Sociology at Essex, Professor Sir Howard Newby moved to the Archive as Director in 1983, a post which he held for five years. He left to become Chair of the ESRC in 1988 and later Chief Executive in 1994, moving to Southampton University where he was Vice Chancellor for seven years, during which period he also served as President of Universities UK. In 2001 he was appointed Chief Executive of the Higher Education Funding Council for England and in 2005 became Vice Chancellor of the University of the West of England. He was awarded a CBE in 1995 for services to social science, and a knighthood in 2000 for services to higher education.

### CATHERINE HAKIM

**April 1989 – August 1990**

Prior to being appointed Director in 1989, Dr Catherine Hakim worked as Principal Research Officer in the Department of Employment's Social Science Branch, where she was engaged in research, briefing and advice on labour market issues including unemployment, small firms, self-employment, home working, labour flexibility and women's position in the labour market. Leaving to join the London School of Economics as a Senior Research Fellow, she has subsequently published a large number of research papers and books focusing on labour market trends, women's issues and research methods.



### DENISE LIEVESLEY

**October 1991 – December 1997**

Professor Denise Lievesley began her career as an official statistician specialising in survey sampling. Later she became the Director of the International Statistical Institute in the Netherlands, before moving to Essex as Director of the Archive in 1991. She left in 1998 to become Director of Statistics at United Nations Educational, Scientific and Cultural Organization (UNESCO), where she established its new Institute for Statistics. Whilst at UNESCO she served as President of the Royal Statistical Society from 1999 to 2001 and is currently President of the International Statistical Institute. She returned to the UK in 2005 as Chief Executive of the English Information Centre for Health and Social Care, based in Leeds.



### KEVIN SCHÜRER

**February 2000 – present**

Professor Kevin Schürer worked at the Cambridge Group for the History of Population and Social Structure, University of Cambridge, before moving to the Archive in 1993 on a part-time basis as head of acquisitions and data processing. He left to join the Department of History at Essex in 1995, only to return to the Archive as Director in 2000. Formerly President of the International Association for History and Computing and currently President of the Council of Social Science Data Archives, he is also an Academician of the Academy of the Social Sciences and UK representative for the European Strategy Forum for Research Infrastructure working group in Social Science and Humanities.





## ACQUISITION OF DATA

According to an article published in The Times on New Year's Day, 1967, Britain was "losing to America a steady stream of vital information essential for social survey work" because, unlike the USA, Britain had nowhere to store the data.

Indeed, the results of an earlier feasibility study provided convincing evidence that there was a large amount of data just ready and waiting for an archive to hold them. But, when the Archive was established, most of the potential depositors identified in the study found reasons not to hand over their data. Some claimed to be still working on them, others had developed an almost parental attachment to them, but most seriously, there appeared to be legal obstacles to acquiring the large datasets collected in government social surveys. Three years after it was established, the Archive only held a small random collection of surveys.

Fortunately, the seventies saw a change in policy for the government departments based, according to Ivor Crewe, less on any concern for the Archive's goals, and more on convenience. Severe job cuts had increased civil servants' work-load, Crewe explains, "so the business of getting a dataset to some academic let alone a student, was not a priority for them".

Since then, the quantity and diversity of data stored in the Archive has gone from strength to strength. This is due, in no small part, to the activities of the ESRC – not only for maintaining their faith (and funding) but also for supporting the creation of large datasets such as the British Household Panel Survey (a study that consistently features in the top five of the users' hit parade).

Forty years on, the quality of the Archive's data results in high demand both nationally and internationally and, it seems more than appropriate that international demand is dominated by researchers from the USA.





At the first meeting of the council for the newly-formed SSRC Data Bank, held on 8 February 1968, the council urged that contact be made with “industrial firms” who were seen as “an important source of data”, and recommended that the “data collection should not be restricted to materials relating to individuals and that data relating to firms should be included”. Acting on this, much effort was placed in the early years of the Archive on acquiring opinion poll data generated by leading market research companies – especially National Opinion Polls (NOP), Market and Opinion Research International (MORI), and Gallup. Although several studies were being worked on simultaneously, formally the first dataset to be acquired by the Archive and made available to others was Village Life in Hampshire, 1965, a study undertaken by Mass-Observation, on behalf of Hampshire County Council. However, it did not prove an easy task for the Archive to persuade data creators to deposit, and many of the early datasets acquired were generated by academics working in universities and related research centres. The first catalogue produced in 1968 listed only six datasets – as David Allen, the research council’s link person for the Archive later commented: “The cupboard [of studies] was embarrassingly almost bare when the time came for Mother Hubbard in the shape of the SSRC to peer in at the end of the initial ... five years”.

In the late 1970s a significant change was brought about with the proactive acquisition of government-generated data – especially large-scale, nationally-representative, repeated surveys. The first breakthrough came with the aggregated small area statistics for the 1971 census, which, as director at the time Ivor Crewe remembers, was a “mark of government recognition and approval” and was in part a result of relieving government departments of “the burden of providing data to bona fide users”. In this sense, this reinforced the Archive’s position as a conduit between data creator and data user. This beginning was followed by the acquisition of the General Household Survey, the Labour Force Survey, and the Family Expenditure Survey, together with a growing range of other, smaller, government surveys. With these acquisitions, usage both increased and widened, reflecting the broad spectrum of research issues which could now be analysed. Government-generated data are now a critically important part of the Archive’s collection, especially given the availability of series of survey data collected over a period of more than 20 years.







‘In the early days ... a £1 was transferred between the depositor and the University of Essex... When I joined the Archive many years later, the transfer of the pound was no more but when a study had been processed and released for analysis, we would send a copy of the Agreement back to the depositor and affix a 10p stamp – clearly some tough negotiations had taken place to reduce the sum by 90p.’

Susan Cadogan, Senior Acquisitions Officer, UKDA

In developing its current acquisition strategy the Archive is informed by the ESRC-led National Data Strategy and works in close collaboration with The National Archives (TNA) in order to put in place a coherent joined-up policy to ensure that key materials for the understanding of society are available for research, now and in the future. A critical partner in achieving this goal is the UK Office for National Statistics (ONS), with whom the Archive now has a contract and concordat, the result of a close and long-term working relationship. The strength of this relationship is reflected in the trust and confidence placed in the Archive as the custodian of data for which the creators may have concerns about satisfying confidentiality promises made to respondents, as well as the upholding of intellectual property rights and adherence to data protection issues. Just as the last decade has witnessed heightened user demand for research data of increasing detail, so concerns over the protection of anonymity have grown. Striving to maintain a balance between these competing pressures, the Archive has recently developed, in conjunction with ONS, a framework for access to more sensitive data, available under Special Licence, which allows registered users access to more detailed and potentially disclosive data under stricter conditions of use.

Another important source of data for social science research has been that created by the academic community, for the academic community. In this respect, over the past 40 years the SSRC/ESRC has itself been a significant sponsor of data creation projects. Major surveys in this category include: the British Election Survey, the British Household Panel Survey, the Millennium Cohort Study, as well as the UK contribution to the European Social Survey. The latest in this formidable line of data investments is the UK Longitudinal Household Study, which, with a sample of some 40,000 households, will be the largest survey of its kind in Europe. The Medical Research Council (MRC) has also contributed to the collection with important data such as the National Diet and Nutrition Survey.

In recent years the ESRC has invested in the licensing of data in addition to the funding of data collection. This policy has enabled data access agreements to be brokered for data collections which would otherwise have been inaccessible to most researchers due to commercial constraints. This trend was started with ESDS providing access to important international macro data series (aggregated data such as those held by the Organisation for Economic Co-operation and Development (OECD), International Monetary Fund (IMF) and World Bank) via its partnership with MIMAS, and has more recently been extended to micro data (individual-level data) such as the LatinoBarometers.

The Archive changed its name from the ‘Survey’ to the ‘Data’ Archive back in 1982 in order to reflect the fact that its remit had widened to all machine-readable data in the social sciences rather than just survey data. Although many still think of the Archive’s collection consisting mostly (if not entirely) of numerical survey-orientated databases, the collection has become more and more diverse over the years in terms of theme and file type. This is reflected most strongly in the creation of the specialist History Data Service in 1992 and the merger with Qualidata in 2001. Both have expanded the collection to contain non-numeric, textual, image and mixed methods collections. These changes reflect the evolving nature of the Archive’s acquisition and collections development policy, which needs to be continuously refashioned to keep pace with the changing data requirements of the social science research community. Web sites are already being archived for future use, the challenge is preparing for the future generations of data – podcasts, blogs, CCTV.







## DID YOU KNOW?

- 1972**
  - 72 per cent of households have refrigerator
  - 66 per cent washing machine
  - 42 per cent telephone
  - 37 per cent central heating
  - 52 per cent car or van\*
- 1974**
  - first inventory of holdings published and distributed
- 1975**
  - 450 studies in UKDA
- 1977**
  - Top 5 studies accessed
    - National Opinion Polls National Political Surveys
    - British Election Studies
    - Family Expenditure Survey (FES)
    - National Child Development Study
    - Relative Deprivation and Social Justice, 1962
- 1979**
  - 1,400 studies in UKDA
- 1981**
  - 74 per cent of households have colour TV
  - 23 per cent black and white TV
  - 93 per cent refrigerator
  - 49 per cent freezer
  - 23 per cent tumble drier
  - 4 per cent dishwasher\*
- 1982**
  - 2,000 studies in UKDA
- 1987**
  - Top 5 studies accessed:
    - General Household Surveys (GHS)
    - 1981 Census
    - FES series
    - Central Statistical Office (CSO) Databank
    - Agricultural Census of England and Wales
- 1991**
  - just 5 per cent of households do not have colour TV
  - 68 per cent have video cassette recorder (VCR)
  - 21 per cent home computer
  - 55 per cent microwave oven
  - 82 per cent central heating\*
- 1993**
  - 3,000 studies in UKDA
- 1997**
  - Top 5 studies accessed:
    - British Household Panel Survey (BHPS)
    - FES series
    - GHS series
    - Labour Force Survey (LFS)
    - National Sample from the 1851 Census of Great Britain
- 2001**
  - 42 per cent of households have satellite, cable or digital TV
  - 80 per cent CD player
  - 40 per cent access to the internet at home
  - 98 per cent telephone, including 70 per cent with access to at least one mobile phone
  - 92 per cent central heating
  - 28 per cent more than one car or van\*
- 2007**
  - almost 6,000 studies in UKDA
- 2007**
  - Top 5 studies accessed (orders, downloads, Nesstar browsing):
    - LFS series
    - GHS series
    - FES/Expenditure and Food Survey (EFS) series
    - BHPS
    - British Social Attitudes (BSA)

\* [www.statistics.gov.uk/lib2001/viewerChart4750.html](http://www.statistics.gov.uk/lib2001/viewerChart4750.html)



# PROCESSING AND PRESERVATION

Today, most computer users find it hard to imagine life before downloads or compact disc technology. But forty years ago, the computer storage medium was paper. Data were recorded using holes punched into paper tape or punch cards and required special tape or card readers (often prone to chewing up the data as they were fed in).

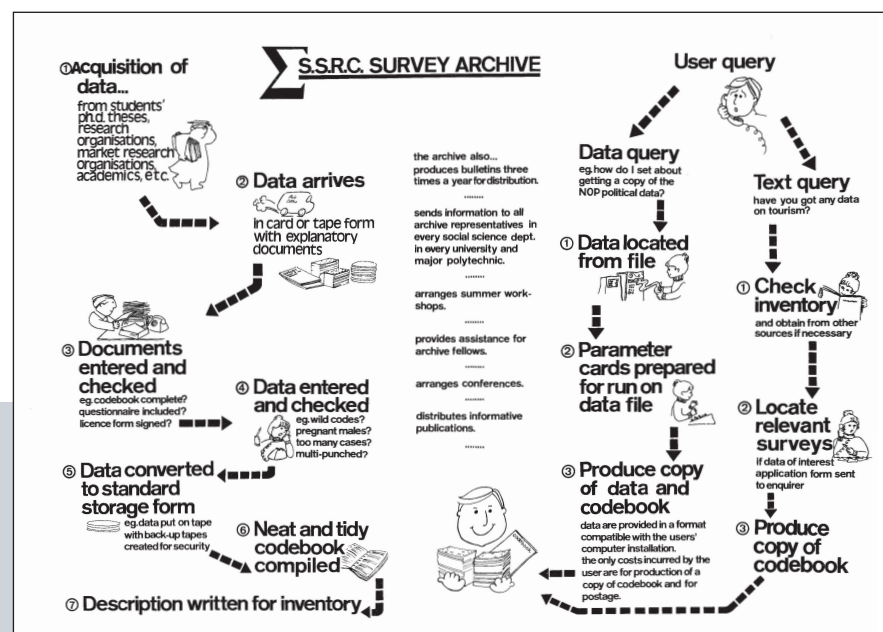
The Archive received data in this format for over twenty years – primarily as stacks of 80 byte punch cards. In the seventies and eighties most of the Archive's data were backed up on large reels of magnetic tape, typically holding up to 40 megabytes (MB) of data. Some studies, particularly the big government surveys, were deposited at the Archive with this media. The racks upon racks of magnetic tapes that built up over this period made the Archive less than popular with the University of Essex Computer Department that stored them. Often large datasets of complex data arrived at

the Archive in a format quite specific to the in-house computer programs used to create them. The material required time-consuming processing before it could be provided to users in a form suitable for secondary analysis.

The Archive still strives to make data analysis as simple as possible for users, developing innovative technological solutions to ensure that users can access data in the most convenient formats and media types. Even though storage media and file formats used in the early years are now obsolete, the Archive's preservation strategy has ensured that none of the datasets collected have been lost. Most of the Archive's data are now stored on Super Digital Linear Tapes (SDLTs). These can hold up to 800 gigabytes (GB) of data or the equivalent of 10,000,000,000 punch cards – laid end to end, these cards would circumnavigate the world, not just once, but 160 times!

## SOME THINGS NEVER CHANGE

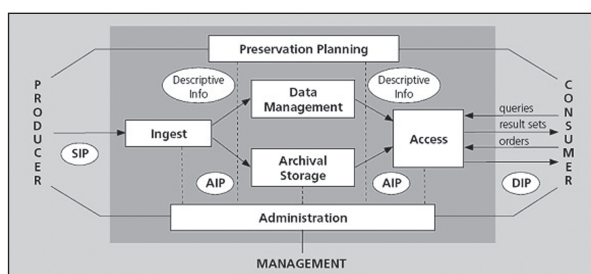
"The greatest misconception about survey archives is the belief ... that when data ... arrive at the archive their transfer is complete". This quote from Allen Potter, the Archive's first director, is taken from an article in New Society published in 1969, yet it is in many ways timeless. Preparing data both for use by others unrelated to the collection process and for long-term access through a managed preservation policy, can be arduous, time-consuming and in some respects unrewarding, in as much that it is often taken for granted by the user community. The 'ingest process' – as it is now referred to – consists of several distinct but inter-linked steps. A valiant attempt to capture these in diagrammatic form was made in 1976 and is reproduced here.





Whilst this flow diagram may seem somewhat politically incorrect thirty years on, what is striking is the fact that in many respects the representation of the ingest process (essentially on the left hand side of the diagram) is conceptually little different from today.

The internationally-recommended standard for the process management and organisation of digital archives is now the Open Archival Information System (OAIS) reference model, which was first developed by the Consultative Committee for Space Data Systems and is now being applied to digital repositories of all kinds. What is very heartening and reassuring is that the Archive measures up to the OAIS very well and is compliant in nearly all regards, and that a system of internal processes which evolved over time out of practical necessity fits well with the theoretical gold standard of the OAIS.



## OTHERS DO...

Although the conceptual nature of the various ingest activities may not have changed significantly, technological change has had a major impact on the way in which the ingest tasks are undertaken. The preparation of data for users was highly complex in the 1970s, and was considerably reliant on access to computer time, something which is inconceivable today. As Eric Roughley commented: "At this time [1972] the Archive remained a small, and rather cosy, organisation. The eight or nine of us would meet each morning to decide which job would be run that afternoon on our two-hour dedicated time slot on the University's ICT 1909 mainframe". The later switch to a DEC PDP-10 computer improved matters, but changing the electricity supply to this machine in the summer of 1976 meant a downtime of over a week, leading to considerable delays in the production of data and codebooks for users. Compare this to 2007 when the entire Archive moved building and relocated its server room with just one hour of downtime for users!

The multiplicity of data formats and file size also created problems in the 1970s, as illustrated by this quote from the Bulletin, the Archive's newsletter, relating to processing of the Family Expenditure Survey (FES) in 1976: "The FES is a huge data set that usually requires two or three magnetic tapes per survey.... The data are stored ... in a format suitable for

the Department of Employment computer programmes, but not, however, in a format that allows analysis with the kind of computer programmes available to our users. That is, one cannot apply SPSS or TSP or OSIRIS to these data without a considerable amount of pre-analytical processing. What this meant, in effect, is that we could not simply convert the data tapes to our own in-house computer code ... [as] ... it would render the data virtually inaccessible to all but the most skilled programmer....[w]e designed a system which allows the user to determine whether she/he is interested in household or individual data.... Then given a set of variables requested by the user, we shall be able to supply a data set that can be immediately input into any of the popular statistical programmes. Thus all that users will have to do is complete a specially designed user request form, on which they describe the variables of interest. This form will serve as the input to our extraction programme". Where things have not changed is in the belief that, regardless of how data are supplied to the Archive, they should be made available to researchers in a form which suits their requirements. In this regard one size will never fit all, so now, as then, processing has to employ the latest technical solutions to continuously ensure that users can choose from a range of formats and media types when receiving the data.

## THIS IS YOUR LAST CHANCE!

"This is your last chance!" announced the Bulletin in 1990, noting that: "The University of Essex will be replacing its mainframe computer this Summer and, in consequence, we shall no longer be able to handle seven-track magnetic tapes, paper tape or punched cards". In 1967 when the Archive was established no-one really thought about the consequences of having to keep the data files that were being collected and stored, accessible over time. As technological change, both hardware and software, renders computer files unreadable, and as magnetic-based storage media deteriorate over time, digital preservation has now become a major topic in its own right. In the past 40 years the Archive has witnessed a succession of changes in both storage media and file formats. As someone recently observed: "if digital preservation did not exist, the Data Archive would of [sic] had to invent it!". Over the years the Archive has had to respond and adapt to changes by developing a flexible yet pragmatic policy to preservation – for example, by holding data in non-proprietary migratable formats across multiple storage media. There is no simple solution to the problem of digital preservation. It is more about minimising risk of data loss or redundancy rather than trying to double guess what the future holds. The comforting thing is that in its 40-year history the Archive has yet to 'lose' a dataset – let us hope that this track-record can continue!

Postscript: the last chance has not exactly passed. The Archive still maintains magnetic tape and punch card readers and endeavours wherever possible to 'rescue' important data collections held in old formats or on obsolete media.

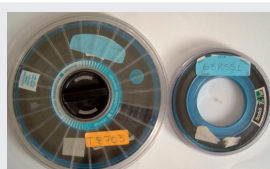


‘As a new member of staff in 1978, my induction started with a lesson on the mysteries of the punched card, followed by a session at a floor-standing computer terminal which noisily emitted realms of paper charting its communication with the university’s huge mainframe computer. In the secretary’s office, the desks were piled high with large magnetic tapes waiting to be packaged in protective containers to be shipped off to waiting researchers.’

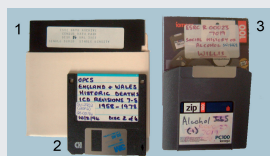
Kathy Sayer, Senior Census Registration Service Officer, UKDA



80-column punched cards and 6-row paper tape. Both generally used in the UKDA for receiving data pre-1990. Punched cards were used as the primary means of data dissemination until around the same time.



Magnetic Tape. Scotch (Seven inch) (c.1981) and Graham Magnetics (Twelve inch) (c.1979). The left hand tape held 1600 bits per inch, meaning that the whole tape would contain some 40 Megabytes (MB) of data. These tapes were used predominantly for data storage, but some studies would have been deposited at the UKDA with this media.



Floppy Disks. (1) unbranded 5.25" disk used by the UKDA to distribute its Census Data Pack to schools in the mid 1980s. This type of disk was introduced in 1976. (2) 3.5" disk of the type first introduced 1984; (3) a Zip Disk, first introduced in 1994 and generally considered obsolete by 2003. The UKDA still receives data on this form of media. The first floppy disk was sent out from the UKDA in October 1981.



Optical Disk – usually known as a CD, introduced on the market in late 1982. Not used by UKDA as dissemination media until 1991. In 1993 the ESRC gave a special grant to the UKDA to purchase a “writable” compact disk workstation”. The cover shows the first bespoke CD published by the UKDA in 1991; the CD on the right was the form in which most data was disseminated until around 2000 when the download service was launched.



Three different magneto-optical disks. Used by Data Archive for backups from the late 1980s. Each holds around 650 MB of data.



Quarter inch cartridge. Media offered by a local authority in 2005 as part of a study on access to media. Note the use of detailed file information on the cartridge.



Data cartridges: Sony 8mm data cartridges (left) and Maxell HS8/112 data cartridges (both c.1989). Used for the delivery of 1991 census data. The Sony 8mm data cartridges contained 90 metres of tape with a capacity of 1.3 Gigabytes (GB) of uncompressed data per tape and a read/write speed of 0.6 MB per second. This tape format was superseded by 1993.



SDLT (Super Digital Linear Tapes) These magnetic data storage tapes were first developed in the early 1980s. The UKDA currently uses these media as part of its preservation strategy. The most recent of these tapes store 800 GB and can be written at the rate of 60 MB per second.





## BUT WHAT DOES IT MEAN?

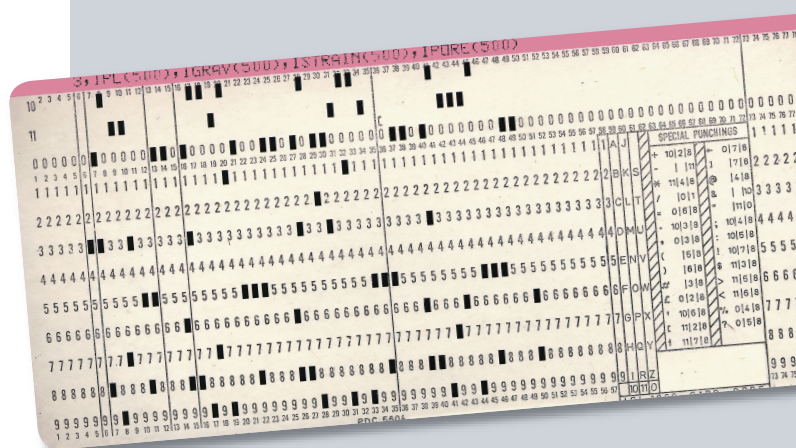
Data, it might be claimed, are only as useful as the metadata which support them. Data cannot be properly analysed and interpreted without adequate accompanying documentation (or what is nowadays often termed 'metadata'). In terms of social science data this could take the form of codebooks, technical manuals describing sampling techniques, explanations of derived variables, application of weightings, and so on – in essence, all that is needed to explain the provenance of the data and the methodologies used in their creation. From the 1960s and the days of punch cards onwards, through the development of statistical software packages and virtual data environments at the turn of the 21st century, the Archive has endeavoured to strike a balance between producing metadata and documentation files in user-friendly formats, and keeping them in a platform-independent state for preservation and long-term usability.

One major project, undertaken during the 1990s, involved the scanning and transfer of paper documentation for over 3,000 datasets (the vast bulk of the UKDA's collection at that time), produced both suitable image preservation format and easy to use, universally-readable Adobe Portable Document Format (PDF) files. This enabled the easy distribution of documentation via the web-based catalogue, therefore maximising the amount of free-to-download information at users' finger-tips.

A major advance witnessed over recent years, but with much potential still to be realised, is the development of internationally recognised metadata standards. These greatly facilitate the transfer of information between systems (for example, across countries) and likewise help in addressing the problem of preserving materials over time. One particular development of importance has been the establishment of the Data Documentation Initiative (DDI). The Archive has been a long-standing member of and contributor to this organisation, and which has produced an eXtensible Markup Language (XML)-based interoperable standard. One advantage of the DDI approach is that it retains both data and metadata within the same XML structure.

'The Data Archive is repeating its 1984 special offer on data supplied on floppy disks. The special offer charge for any dataset ordered (by 30 September 1985) on floppy disks will be only £9.95, which will include the handling-charge, the cost of one disk and postage and packing.'

Bulletin No. 31, May, 1985



## DID YOU KNOW?

The first deposit and issue of data from the Archive was using punch cards.

Indeed, in 1968 the Essex County Standard described the Archive as being "a large filing system – a battery of steel cabinets each contains 30 or more trays with 2,000 cards to a tray. The present storage space is for 2 million punched cards, and the material is duplicated onto magnetic tape on spools, taking up to 2,400 ft of tape".

If punch cards were still in use, the latest version of the British Household Panel Study would take up some 6 million cards, while the entire collection would extend over 12,300 million punch cards!

# DATA DISCOVERY AND DISSEMINATION

There is little point in storing and preserving data if no-one knows what is there or how to access it. So the Archive has always made it a priority to ensure that users can find and retrieve data in the most simple and convenient way possible.

The Archive's first catalogue was a basic inventory produced in 1969 using the printer in the Computer Centre at the University of Essex. It generated just eight requests for datasets that year, but as the Archive had only six users and a very limited range of data to choose from, it is fair to say that the catalogue itself was a resounding success. As the Archive grew, so did the task of producing the inventory. The next catalogue, published in 1972, took two members of the Archive's staff over a year to collate. Updated inventories were then released regularly until the Archive launched its computerised bibliographic retrieval system in 1986, making paper catalogues redundant – just as well because even then, the catalogue had grown to a heavy two volume tome.

The Archive was quick to take advantage of internet technology, launching its first web site in 1994 and developing a searchable web-based catalogue. Advances in information and communication technology also revolutionised the dissemination process. With the launch of UKDA download in 2001, users can order and download their chosen datasets from the online catalogue at the click of a mouse. A far cry from the earlier cumbersome process of ordering data by mail (using paper registration forms), and waiting patiently for the data to arrive in the post on portable media such as punch cards or CDs.

The Archive's users have embraced the changes and in 2006, the online catalogue generated over 49,000 requests for datasets, the vast majority as downloads and only 95 as CDs (there were no requests for punch cards).

‘The SSRC Data Bank ... is claimed to represent a revolution in information retrieval techniques.’

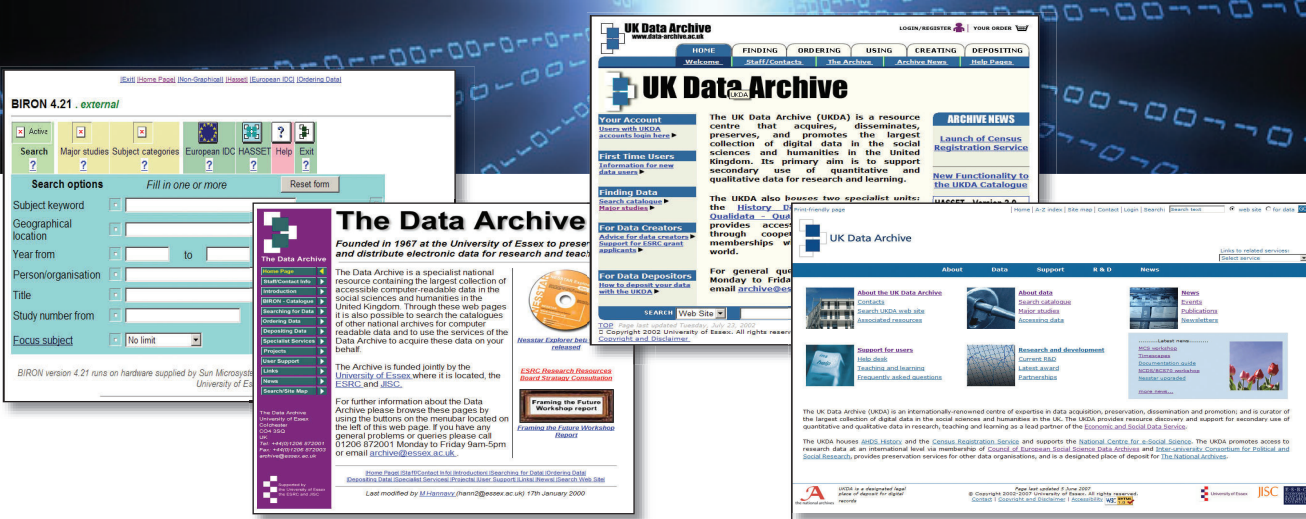
## TALK ABOUT REVOLUTION

Being formed in 1967 – the year of revolution – the Archive obviously predates the world of the web, e-science and cyber-infrastructures, and even email for that matter, by decades. Consequently, in terms of the Archive's history, it is only fairly recently that the point has been reached whereby most researchers and users will interact with the service wholly through the internet. Web-based delivery has revolutionised the entire industry of data-based services and changed them beyond recognition, but, for a service born 40 years ago, access to resources and the dissemination of those resources has not always been so straightforward.

It was not until May 1984 that the UKDA launched a service for the exchange of material on floppy disks. Prior to that, data were supplied on tape as the 'hard' medium of choice and occasionally via telnet in what might be thought of as the e-medium of its day. Even then, supply via telnet (and later ftp) required a technical expertise on the part of the remote user. Indeed, it was not be unusual for users to require the assistance of colleagues in computing services to 'pull' data across the network using telnet protocols. But even with these technical obstacles in place, dissemination via telnet remained a faster and cheaper option than copying data to a hard medium and then distributing the material by post.







In the early 1990s, in collaboration with the Archive, MIDAS (as it was then) at Manchester began to host remote access to a selected number of large-scale government surveys. Rather than requesting copies of the data to be sent from Essex, researchers could logon to the national computing service at Manchester and conduct their analyses using standard software such as Scientific Information Retrieval (SIR) or Statistical Package for Social Science (SPSS). The early 1990s was also when the Archive started to supply data on CD – which it described in a release to users in 1991 as something “which can be played on most desktop computers, ... providing an excellent (and inexpensive) vehicle for distributing large amounts of stable data”.

## SPINNING A WEB

A significant turning point, of course, was the introduction of the web. The Archive developed a web-based presence fairly early, launching its first web site in 1994. Initially, like a lot of web services at the time, it was limited in scope, essentially allowing users to search the holdings using the Bibliographic Information Retrieval ONline (BIRON) catalogue and, for the first time, place orders online. At this time, all users were still required to print off and sign a user undertaking outlining their responsibilities as a user each time they placed an order.

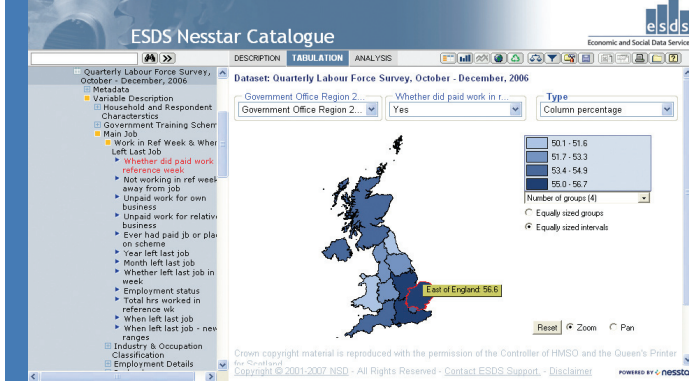
The arrival of the new millennium saw further technological advances and in November 2000 a new suite of web pages

were launched. Users were now able to register with the Archive online – paper registration forms, processed manually and subject to postal delay/misdirection, became a thing of the past – removing a significant barrier to use. Requests for data became instantaneous and users were (and still are) able to order data using a ‘shopping-basket’ facility.

A year later users were not only able to request data instantaneously, but could browse and download selected data at the click of a mouse. Nesstar, the internet-based software, permitted users to browse, visualise and undertake exploratory analyses on survey data. Now the Nesstar Catalogue has been linked to the Council of European Social Science Data Archives (CESSDA) data portal allowing UK-based researchers to locate, access and browse data from across several European countries through a common interface.

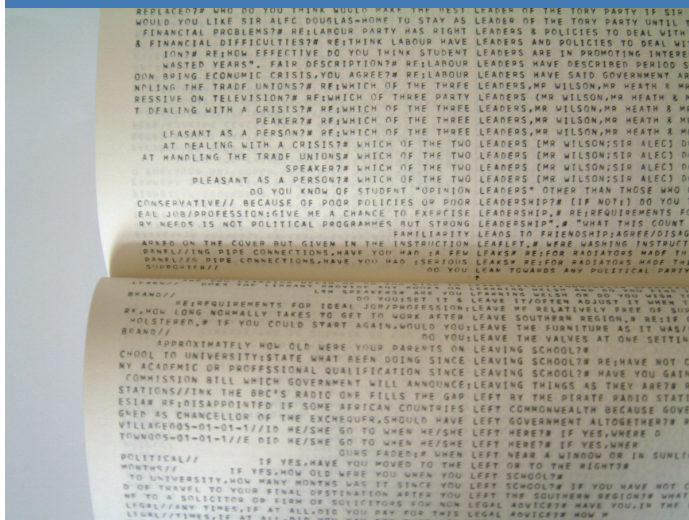
2001 also saw the launch of UKDA Download – an online data download facility for registered users, offering data in SPSS, Stata, tab-delimited and Rich Text Format (RTF) format. This download facility remains a core part of the user experience today with all new studies, whether numeric, textual or multi-format, routinely prepared for instant download, and all back catalogue datasets prepared for download as and when they are ordered. Restricted datasets can also be disseminated with full security controls – a significant improvement over the previous situation where restricted datasets would have to be delivered via ftp or offline media.





## NESSTAR

The map, produced using Nesstar, illustrates the advances made in the presentation and dissemination of metadata since the 1970 paper catalogue record. The user is now able to browse and analyse the metadata online.



‘...the operator even wore white gloves...’

Printing an early copy of the catalogue

In 2003, an online browsing system for qualitative data called ESDS Qualidata Online was developed in order to compliment the Nesstar system for survey-type data. This allows free-text interview transcripts to be searched in a structured manner and also allows related qualitative research materials (such as audio files and images) to be embedded for download.

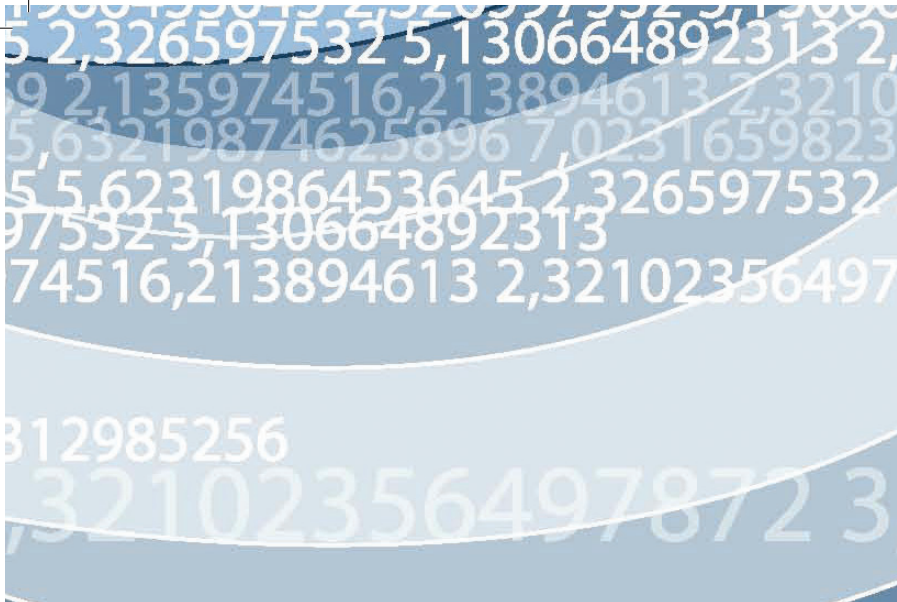
In hand with the introduction of these web-based technologies which speed the process of delivering data to researchers significantly, perhaps one of the greatest achievements in making access to data quicker and easier has been the introduction of the one-stop registration service. First introduced in 2002 and subsequently modified to include Athens authentication, this allows researchers and users to register once and agree standard terms and conditions of use online and then access as many data collections as they wish. This is a huge advance from the days when researchers had to complete, sign and send a data request form for every dataset required.

## IT'S THERE SOMEWHERE...

Providing information about the content of the data collection held by the Archive has always been a high priority. In the early days of the Archive it was thought that the main finding aid should be in the form of a comprehensive (paper-based, of course) cross-referenced list of questions derived from surveys. This was to include questions asked in all social surveys, including those not held by or accessible from the Archive. The KWIC (keyword-in-context) project as it was known was hugely ambiguous for its time and proved to be almost too burdensome to create and cumbersome to use. As David Allen later commented: “Few who consulted KWIC discovered what they were hoping for, more were confused by it and even found it a source of merriment (‘colour’, for example, lead one not to questions on racial problems but to a survey on paint)”.





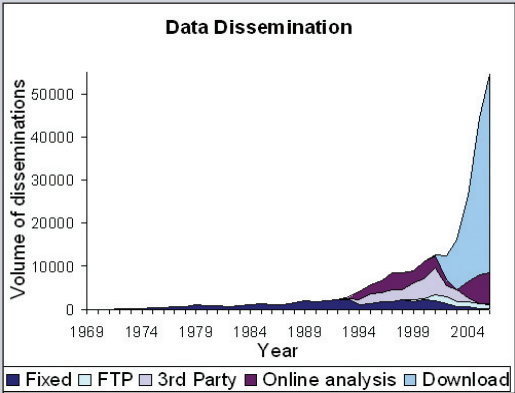


KWIC was relatively short-lived and replaced by a more conventional catalogue or inventory of data holdings, updated on a regular basis. The first of these was produced in 1969 and is remembered with due reverence by Eric Roughley (Deputy Director of the Archive from 1967 until his retirement in 1992): “With due ceremony, the Computing Service’s printer was cleaned and prepared, a new ribbon inserted and, in a most stately fashion, a wedge of print-out would be carried over to the Computing Centre where we gathered around to observe the new catalogue being printed off. The whole exercise was conducted in a manner more like a masonic ritual -the operator even wore white gloves! The pristine product was then distributed to our subscribers (six) who paid £3 ...for the privilege”. Later editions of the catalogue were printed, bound and published – the last such publication of this kind being a weighty two-volume tome.

An important change was implemented in 1981 when the Archive formerly adopted a standard study description for survey data developed collectively by a group of European social science data archives. Common description has subsequently facilitated exchanges of information between archives and cross archive information retrieval. In January 1986 the UKDA launched its first computerised bibliographic retrieval system, known as BIRON, developed under the leadership of Bridget Winstanley. A thesaurus of terms used for the indexing of its humanities and social science datasets known as Humanities and Social Science Electronic Thesaurus (HASSET) had also been developed building on the 1971 UNESCO thesaurus. Still under continuous development and renewal, this has subsequently been implemented in a number of informational retrieval systems around the world, and in a condensed form (the European Language Social Science Thesaurus (ELSST)) has been translated into eight languages to date, with others planned. Alongside BIRON, in 1995 the Archive released its first web-enabled catalogue that allowed cross-searching of other European archives collection descriptions. This resource was one of the first web portals found in the social sciences. The UKDA eventually said goodbye, sadly, to BIRON in December 2002 as advances in technology enabled much faster, and more flexible access to the catalogue via the Archive’s web site.

### DID YOU KNOW?

Changing modes of data dissemination, in particular the introduction of download facilities, have had a dramatic impact on the volume of data supplied to researchers. Considering the entire period 1967 to 2006, the final year alone accounted for some quarter of all data supplied, with half of all the data ever supplied by the Archive being distributed in the last three years!



HASSET is a thesaurus reflecting the subject content of the UKDA holdings which was developed by the UKDA in 1986 as an aid to resource discovery. It has been continuously expanded and updated since then, and has become the pre-eminent tool for the organisation and retrieval of social science data.

#### Terms in HASSET in June 2007

geographic	2,714
subject	4,481
synonyms	3,750
total terms	10,945

In the year 2005/06 the overall number of datasets delivered to users rose by nearly 20 per cent over the previous year, from 41,134 to 49,169.



## SUPPORT – GOING OUT TO USERS

Helping users to make the most of the Archive has always been a high priority, but for the first seven years, this support was mostly in the form of detailed guides.

Little had been done to inform potential users about the existence of this valuable resource until Ivor Crewe took up his directorship in 1974, by which time the Archive's lack of use had become a critical concern. In an effort to boost user numbers, Crewe gave talks up and down the country promoting the Archive and its collection with an almost evangelical zeal.

He also edited the first of what was to be a series of 75 issues of the Bulletin – a tri-annual publication providing a continuing source of up-to-date information about the Archive and any datasets it had recently acquired. The current newsletter, UK Databytes, is a quarterly publication highlighting the Archive's new developments, but much of the communication with users is now through email, mailbase lists, and the web site.

Over the last ten years, there has also been an increased emphasis on support targeted to specific groups of users, including workshops that focus on specific datasets and training on methodology. According to Kevin Schürer, this type of support is all the more vital as the Archive attracts an increasing diversity of users: "The number of users from disciplines outside economics and sociology is increasing and, most encouragingly, we are seeing a significant increase in use by people outside academia", he says.

The increased focus on outreach has paid dividends and the Archive now has many thousands of requests, not only for the data in its growing collection but also for help and support. Much of the guidance the Archive provides is now available online and, in 2006, users downloaded 807,797 PDFs giving advice on data use as well as data management and preparation.





The services offered by the Archive have always been more than simply data in – data out. In addition to all the data processing and ingest activities, the support of researchers in their use of the data, together with the promotion of the collection has always been central for the Archive over its 40-year history. Indeed, when the Archive was established there was a clear emphasis on the provision of data analyses to researchers who, were not in a position to conduct the analyses themselves. (In fact, it was originally hoped that this service would generate revenue to support other in-house tasks – an expectation never realised.)

Up until the mid-1970s support for users was rather limited by today's standards, and efforts in some respects were concentrated on designing and producing detailed bespoke user guides rather than promoting the data directly. This changed with the appointment of Ivor Crewe as Director in 1974, who, keen to increase the number of users, made concerted efforts to, as he later put it, “take the Archive out on the road”. Talks were given up and down the country about what the Archive did and in 1975 the first issue of the Bulletin was produced, replacing the much shorter, ad hoc and infrequent newsletter. The Archive and its holdings were also promoted through the annual Summer School in Quantitative Methods held at the University of Essex.



In the 1980s, as the collection of major government surveys held by the Archive began to grow, attention turned to organising regular User Group meetings focusing on such series as the General Household Survey (GHS), Family Expenditure Survey (FES) and Labour Force Survey (LFS). These User Group meetings played a major role in supporting the needs of researchers by providing substantive and technical support as well as bringing together both government data providers and users to share information, ideas and experiences. Under ESDS, these continue to be a major component in researcher support and are heavily subscribed to both from within and outside the academic community. In 1986 the first weekend workshop was held (for the extra-keen) on ‘Teaching Quantitative Data Analysis’. Although the experiment was short-lived, the spirit which prompted it still lives on and flourishes in the now highly successful ESRC-funded Research Methods Festival.



Comments received from participants at the Health Data Workshop on 12 March 1998

- ‘Workshop was very good – interesting, pitched at right level, length and good venue.’
- ‘Very interesting – all sessions. The information should be useful in our work.’
- ‘Extremely useful – got something out of all sessions. Well-balanced programme.’



Although it has been high on the Archive's agenda for a long time, since the formation of the specialist History Data Service in 1996 and the Economic and Social Data Service in 2003, the Archive has placed an even higher priority on targeted, focused user support. This has taken multiple forms: events for general awareness raising, workshops focusing on specific themes or datasets, user groups, and training on methodology relating to the use of Archive-held data. The Archive equally collaborates with other expert groups, centres and information providers to host days of interest. In 2006 alone, the Archive organised or contributed to over 160 user-support events – as many in a single year as in the first twenty years!

The Archive has also devoted attention to harnessing the enthusiasm of data professionals, computing staff and librarians across the UK who provide a useful service as local agents on behalf of the Archive. This was particularly important in the days when punch cards and magnetic tapes holding data had to be sent somewhere, and forms had to be signed and collated. These were the network of Official Representatives or ORs. Although the original purpose and role of the OR has changed, such contacts still provide an important function in promoting the work of the Archive locally and helping local researchers with any queries. The Archive continues to support local representatives and work closely with data librarians in universities.

## COMMUNICATION

The Archive produced its first Newsletter in 1968, shortly after being founded. This was little more than a broadsheet, typed onto a single page and copied. This was replaced in 1975

with the production of a fuller length Bulletin (all back copies of which are available in PDF format from the Archive web site). Its main purpose was “to provide a continuing source of up-to-date information about the Archive and to act as an informal supplement to the Inventory”. Indeed, of the 32 pages of the first issue, 26 were taken up with listing names and institutional details of Organisational Representatives, new holdings and collections held in foreign data archives. The Bulletin was a tri-annual publication which included longer articles on methodology in the use of data materials held by the Archive, book reviews and news items. Due to increased use of email, mailbase lists and the web for disseminating information the last issue was produced in September 2000. From 1984 the Archive also produced a regular GHS Newsletter to help support the user group. This ceased in the 1990s. The Bulletin was replaced in 2001 by a quarterly newsletter, UKDatabytes (renamed UK Databytes in 2003) focusing on news items providing readers with an alternative way of keeping abreast of developments. UK Databytes currently has a postal circulation list of approximately 2,750, with many more readers accessing it online.

## BRANDING

The outward image of the Archive has changed quite dramatically over its lifetime with a transition right through the colour spectrum from red to blue via purple and green! Logos have also morphed radically reflecting the input from new directors and teams as they impose their own style and vision to fit the intended profile of the organisation, or just to keep up to date with emerging styles.







## USER GUIDES

Helping users to get started has always been one of the Archive's challenges. In the 1980s paper guides were prepared on the use of data in teaching and learning. The 2000s have seen the development of a series of user-friendly searchable online guides that both introduce and give step-by-step advice on how to use various services. These are also used in hands-on workshops.

Other guides provide a more in-depth exploration of issues pertaining to using data by research themes, software packages or data analysis techniques. Thematic or subject-based help in finding data are always popular, where an overview of the subject, examples of key datasets, and links to UK centres of expertise in the respective areas, are collated. The Archive has also prepared guides to good practice in the areas of data usage and sharing, data documentation and preservation, and case studies of secondary analysis.

## HELP DESK

The Archive provides a dedicated help desk for general questions about finding, accessing and using data, including help with the most popular data handling and analysis software. There are currently over 45,000 registered users and in 2005-06 the help desk answered over 3,000 queries.

## SELECTED EVENTS THROUGH THE DECADES

### 1978 LOCAL DATA FACILITIES: ARCHIVE REPRESENTATIVES' CONFERENCE

This conference focused on the organisation and maintenance of local data facilities, such as might be established by a department's store of data.

### 1982 DATA ARCHIVE SUMMER SCHOOL

Participants attended the survey design course taught by Martin Collis and Denise Lievesley, while the survey data analysis course was taught by Eric Tanenbaum.

### 1996 TOMORROW'S ARCHIVE

The 30 year history of the Archive was celebrated with a name change to The Data Archive. Ron Amann, then Chief Executive of the ESRC opened the day's presentations.

### 1999 FRAMING THE FUTURE: PAINTING A DIGITAL LANDSCAPE FOR SOCIAL SCIENCE RESOURCES

This workshop brought together major players in the use of key research resources to take forward practical discussions on operability.

### 2003 ESDS LAUNCH

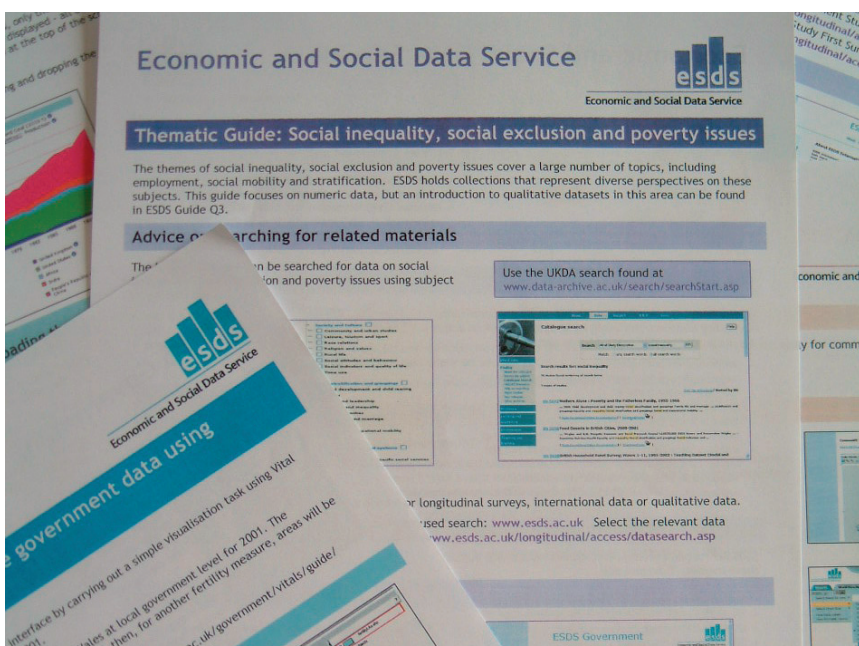
The Economic and Social Data Service was officially launched in 2003. Speakers included John Pullinger of the ONS, Ian Diamond, and Len Cook, the National Statistician.

### 2007 COOKING NUMBERS AND EATING WORDS: USING DATA TO INVESTIGATE FOOD, LIFESTYLE AND HEALTH

This was held as part of the ESRC Festival of Social Science. The conference presented evidence and fostered debate on empirically-derived data sources providing evidence on the social benefits, constraints and implications of food, health and lifestyle.

'The SSRC Data Archive has established a 'postbox' on the University of Essex DEC-System 10 computer to enable people with access to the Essex computer either via PSS or the telephone to send 'electronic mail' directly to the Archive'

Contacting the Data Archive 'On-Line' – Bulletin 21 February 1982







## INTERNATIONAL PRESENCE

When the Archive was established, other data archives already existed in the USA, the Netherlands and Germany, although an article in *The Times* (1 January 1967), suggests that the German archive did not have much of a head start: “The German archive at Cologne contained a survey conducted for an oil company interested in petrol consumption. The survey contained questions about people’s weekend travel and holiday activity”. Although media reports of the day portrayed the international archives, particularly those already established in the USA, as competition for the fledgling British archive, the Archive’s activities have always focused on international co-operation.

In June 1976, an informal alliance (the Council of European Social Science Data Archives or CESSDA) was set up between the Archive and the six other European data archives in existence by then, to promote co-operation as well as the unrestricted exchange of data between nations. This alliance has grown and, with the addition of Romania and Luxembourg in 2001, now boasts over 20 member countries.

The Archive has made significant contributions to CESSDA’s research and development projects and Kevin Schürer is currently CESSDA’s elected president – the first time that a director from the Archive has held this position.

The Archive has also built strong links outside Europe: the Inter-university Consortium for Political and Social Research (ICSPR) in Michigan, USA, is now one of its main international partners, enabling the Archive’s users to access datasets in the oldest and largest social science data archive in the world.

The Archive’s international standing attracts an increasing number of delegates seeking information and guidance. The visitors’ list includes Canadians, Australians, Russians, South Koreans and even a group from Kyrgyzstan. This year, in January alone, the Archive hosted delegations from the South African Human Sciences Research Council and the Chinese Academy of Social Science.



Prior to the creation of the Archive in 1967, a small number of data archives already existed in other countries – the USA, Germany and the Netherlands. Thus from the start the Archive was part of an, albeit informal, international network. (In fact, one of the – potentially spurious – reasons given for creating the Archive was to “stop the present flow of survey material to ‘banks’ in other countries” – Evening Standard, 31 January 1967.) Since the 1970s the Archive has engaged with a number of international initiatives in order to foster co-operation on key archival strategies, procedures and technologies; to encourage the exchange of data and technology across national boundaries; and to promote the acquisition, archiving and distribution of electronic data for research and teaching.

The Archive is a member of CESSDA and of the International Association for Social Science Information Service and Technology (IASSIST) through which it engages in international collaborative projects on issues such as data sharing, metadata and social science thesauri. The Archive is also the UK national member institution of the USA's national social science and historical data archive, ICPSR in Michigan, as well as the International Federation of Data Organizations (IFDO). The Archive's involvement in cross-national research and development projects has made a significant contribution to new developments in data preservation and dissemination, metadata standards, software for web browsing, data discovery and data delivery.

UKDA staff have always kept abreast of social science data developments and represented the UKDA at many key international forums across the world. Equally, the UKDA has seen many visitors from all over the world come to learn how the UKDA works as a leading centre of expertise in data sharing, preservation, and use. In 2006 the number of such visits reached an all-time peak of 80.

## CESSDA PORTAL

COUNCIL OF EUROPEAN SOCIAL SCIENCE DATA ARCHIVES

### CESSDA

In June 1976 the directors of the then seven national European data archives, including the Archive, met in Amsterdam to discuss international data transfers. This led, chiefly at the instigation of Stein Rokkan (the Norwegian political scientist, who was also influential in convincing Michael Young, the first chair of the SSRC about the need for a data archive) to the formation of CESSDA, which above all agreed a protocol for the administration of data transfers between national organisations – later to become the CESSDA Trans-border Data Access Agreement.

The Council's General Assembly meets annually and now represents 23 national member organisations. An annual CESSDA Expert Seminar has also taken place since 1987, concentrating on practical information related to data archiving and aimed at developing the careers of junior and middle-ranking archive staff. A major achievement has been the development of the current CESSDA Data Portal, replacing its predecessor, the Integrated Data Catalogue. This provides a seamless interface to datasets from several social science data archives across Europe, building upon the work of a number of EU-funded projects. In 2005 the Archive director was elected President of CESSDA (the first time the position has been held by the UK) and in 2006 CESSDA was identified by the European Strategy Forum on Research Infrastructures (ESFRI) Roadmap process as a major European research infrastructure.



‘Now that the number of social surveys is rapidly increasing, the Data Bank offers a wonderful opportunity for comparative research over time and across nations.’

Professor Townsend, reported in Wyvern, 10 Feb 1967



## ICPSR

One of the Archive's main international partners is the ICPSR in Michigan, USA – the oldest and largest social science data archive in the world. The Archive first became affiliated with ICPSR in 1975, when it was agreed that it would act as a national distribution agent for ICPSR in the UK. All academic users in UK universities, polytechnics and research institutes, as well as users in central and local government and not-for-profit organisations, are entitled to obtain ICPSR datasets through the Archive. In recent years regular staff exchanges have taken place between the Archive and ICPSR, and the director has been a long-time attendee of ICPSR Council meetings as the invited European (CESSDA) representative.

## IASSIST

IASSIST is an international organisation of professionals working with information technology and data services to support research and teaching in the social sciences. There are 300 IASSIST members, including a number of Archive staff. Members come from a variety of workplaces, including data archives,

statistical agencies, research centres, libraries, academic departments, government departments, and not-for-profit organisations. An IASSIST conference has been held annually since 1974 and the Archive has long been an important and well-represented contributor.

Archive staff have been represented in a variety of elected and appointed positions over the years, including Administrative Committee members, European Regional Secretaries, and chairs of various committees. Melanie Wright is the current president of IASSIST and Louise Corti was IASSIST conference programme co-chair in 2007. Both Melanie Wright and Louise Corti are Associate Directors of the Archive.

## THE ARCHIVE AS AN INTERNATIONAL SOURCE OF GUIDANCE AND EXPERTISE

Since the 1970s the Archive has provided guidance to a host of visitors from overseas on issues such as data archiving, data management and preservation, cataloguing and metadata, and user support. One of the earliest of these visits was in 1975 from the Director of the Social Science Computing Laboratory at the University of Western Ontario, Canada, who visited the Archive as part of a European fact-finding tour. This was followed shortly afterwards by a visit from the Director of the Steinmetzarchief in Amsterdam. In more recent years visitors have been received from the Human Sciences Research Council, South Africa, the Chinese Academy of Social Science, South Korea, Taiwan, Hong Kong, Russia, Kyrgyzstan, Singapore and Australia, to name but a few. Most come on fact-finding missions, often related to establishing a new data archive, or just to learn more about what the Archive does and how it does it.







‘The UK Data Archive is an exemplar of success in the international data archive community...Their dedication to provide the highest quality services along with their over-arching commitment to standards in data preservation, security, and interoperability truly make the Archive a world-class operation.’

Diane Geraci, Librarian for the Social Sciences, Harvard University



## ARCHIVE VISITING FELLOWSHIP (1975 ONWARDS)

In 1975 the UKDA inaugurated its Visiting Fellowship scheme. By 1978 there had been nine Fellows from Canada, the United States, Italy and New Zealand.

In May 1977 Professor Samuel Kernell of the Department of Political Science, University of Minnesota, was the Survey Archive Visiting Fellow. His research interests lay in public opinion and electoral behaviour in the USA, on which he had published a number of articles in the *American Political Science Review*, *Public Opinion Quarterly*, and the *Journal of Politics*.

He spent his time at Essex on a comparative analysis of short-term electoral change in the USA, Britain, France, West Germany and Japan, using data lodged in the Archive.

# RESEARCH AND DEVELOPMENT

The Archive's research and development activities have always focused on developing the tools and products on which its data preservation and dissemination service depends.

Although vital for the core services it provides, additional funding must be sought to cover research and development costs, and the Archive has been particularly successful over the last five years in securing awards from sources including the ESRC, the European Commission and JISC.

As the nature of data becomes more global, there is an increasing emphasis on collaborative international projects such as the recently completed MADIERA project. This project, involving archives from seven countries including the UK, aimed to break down language barriers in searching for data by employing a multilingual thesaurus, based on HASSET – the thesaurus developed over the last 20 years at the Archive. The

authors of the first version of HASSET, Ivor Crewe and his colleague Bridget Winstanley, could scarcely have imagined that their brain-child would not only have stood the test of time, but would be translated into seven different languages (soon to be ten).

It is often the case that the technology developed for one specific purpose will have wider application in the future. Kevin Schürer expects this to be true of the technology developed for Histpop, a recently completed JISC-funded project. This project produced a web site with the complete collection of British historical population reports from 1801 to 1937, providing a comprehensive resource for historians. But, as Schürer explains, "it has also taken the Archive into the new and important area of delivering images online within a searchable browsable content". Schürer forecasts an increasing demand for visual and indeed audio-visual data, such as CCTV. "The whole nature of what researchers need in terms of data is changing and we need to be prepared for that", he says.

Over the years the Archive has been extremely successful in securing a number of awards in addition to those for running the core services in order to conduct, what is broadly termed research and development. Income has come from a number of sources, including the ESRC, the European Commission and the JISC. The Archive works in collaboration with organisations worldwide, including partner archives, data libraries, research centres and national statistical institutes, to achieve the goals and objectives of each award. The research and development activities have always aimed to enhance the Archive's services, consolidate its position as an internationally respected organisation and spearhead new developments in data preservation and dissemination. In particular, the main goal is to develop tools or products that can then be rolled out or implemented in the main services. In recent years this has resulted in the development of software for web browsing, data discovery and delivery, as well as enhanced data products.

## ACQUIRING THE FAMILY EXPENDITURE SURVEYS (1975-76)

In 1975 the SSRC agreed to meet the cost of transfer to the Archive of the Family Expenditure Surveys (FES) from 1961 onwards. This involved conversion of very early tapes to a more up-to-date format, as well as some data standardisation, since methods of coding the survey data were radically different for the period 1961-67. It was noted at the time that: "the cost of punching or copying alone for the period 1957-73 would be some £62,500. In addition there would be very considerable expense, at present uncosted, in moving and controlling the original documents in preparing explanatory materials and in liaison with the Archive" (Office of Population Censuses and Surveys, April 1975). Thus, unfortunately, due to what were seen as prohibitive costs, the earlier pre-1961 data were not rescued, in retrospect, perhaps a rather short-sighted decision.



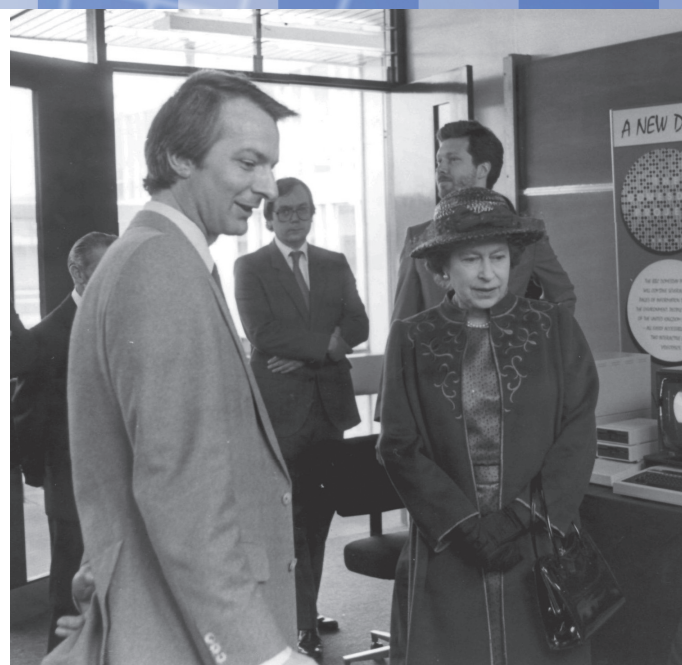


### THE DOMESDAY PROJECT (1985-86)

At the beginning of October 1985 the Archive reached agreement with the BBC to be a major contractor for the provision of data for a new Domesday Survey which was undertaken from 1985 to 1986. The BBC launched the publication of the discs on 26 November 1986 to coincide with the 900th anniversary of the Domesday Book, using then up-to-date interactive video disc technology. The discs contained a vast amount of information including 24,000 maps, 250,000 pages of textual information, 60,000 photographs, 1 hour of moving film and sound and 200 megabytes of data. The Archive supplied over 3,200 datasets for inclusion on the discs which were intended for use in teaching and research, including schools. The Queen was shown the Domesday Project by the then director Howard Newby, during her visit to the University of Essex in 1985.

### NESSTAR AND FASTER (1996-2002)

In 1996 funding from the European Commission, under the Information Engineering part of the Telematics Programme, supported the project known as Networked Social Science Tools and Resources (NESSTAR). The aims of this project were to further internet developments of value to the European data archives and their users, with key components being an integrated yet distributed catalogue of data holdings, with further modules devoted to data browsing, data visualisation, simple analysis, data subsetting and downloading. The Archive's main partners in this enterprise were the Norwegian Social Science Data Service and the Danish Data Archive, with the Archive in a co-ordinating role.





‘WEC-funded projects provided a stimulating environment for working with international colleagues and exploring how we could use new technologies to develop improved services for our users.’

Simon Musgrave, Acting Director, 1998-2000



In September 1999 the Archive began contract negotiations with the European Commission for a new research proposal called Flexible Access to Statistics, Tables and Electronic Resources (FASTER) – to take forward some of the developments within NESSTAR, aiming to develop a robust architecture for the dissemination and use of statistics.

As a result of these projects, in late 2000 the Archive launched its first Nesstar product – ‘Nesstar Light’ for online data discovery, browsing, tabulation and download. Subsequently, additional versions of the Nesstar software have been released, with the latest version (3.5) allowing tabulation of aggregate (cube) and micro data, regression and correlation, recoding and grouping of variables.

#### MADIERA (2002-06)

The European Union (EU)-funded Multilingual Access to Data Infrastructures of the European Research Area (MADIERA) project aimed to develop and employ a multilingual thesaurus to break down language barriers in the discovery of key social science data resources. As a result the multilingual thesaurus European Language Social Science Thesaurus (ELSST) was developed as a follow-up to the Language Independent Metadata Browsing of European Resources (LIMBER) project. This has subsequently been used within the CESSDA data portal to provide a seamless interface to datasets from social science data archives across Europe. The distributed catalogue is currently searchable in nine languages (German, Danish, Greek, English, Spanish, Finnish, French, Norwegian and Swedish).

#### CHCC (2001-03)

The Archive contributed in three areas to the Collection of Historical and Contemporary Census Data (CHCC) project. The project’s main objective was the development of learning and teaching resources achieved by improving accessibility to the primary data resources and by developing an integrated set of learning and teaching materials.

The History Data Service improved access to the Historical Census Collection by making data from census enumerators’ books and data from the published census reports available online. This work led indirectly to the historical census download service, now managed by AHDS History and currently accessed by more than 650 users annually.





The Archive's role was to develop a stand-alone Census Resource Discovery Service (CRDS) which is still available and employs metadata schemas in order to index its dataset and learning resource records. The storage and resource discovery of the metadata are centralised but the entry system is distributed, with each CHCC partner entering their own metadata.

Finally the Archive was responsible for the evaluation of the project outputs. This survey-based evaluation work included testing by students and a tutor-led case study of the service.

### **TRAMSS (1998- 2000)**

The Teaching Resources and Materials for Social Scientists (TRAMSS) project aimed to place exemplar data analyses in a substantive context by introducing data sources and methods via research questions. Geared towards research students in quantitative social science research, the resource guided users through finding, extracting and applying analysis complex models to data. It is still in use by teachers and has subsequently been updated.

### **JISC X4L (2002-03)**

The JISC Exchange for Learning Project (JISC X4L) project built, piloted and documented the evaluation of a set of data-based resources for use in teaching in political science courses within higher and further education institutions. The resources were based on existing Archive holdings that were re-purposed to suit the pedagogical needs identified by the tutors authoring and piloting the teaching resource for their existing courses. This included obtaining permission from the Home Office to create an abbreviated customised version of the British Crime Survey for online analysis, which has subsequently been used as a major resource in its own right.

### **ARCHIVE DOCUMENTATION DIGITISATION PROJECT (1996-98)**

An ambitious project to scan the entire paper documentation holdings of the Archive (around 600,000 pages) aimed to provide machine-readable documentation for all datasets within 18 months. Tagged Image File Format (TIFF) images were produced, which were then converted into PDF files and combined to make comprehensive user guides to accompany the data.

Hiring teams of temporary staff who worked in shifts 24 hours a day, the project was successfully completed within a year, leaving time to digitise the administrative and deposit- process material associated with studies. As a result, documentation for all studies within the Archive can be delivered to researchers in machine-readable form (indeed, all data documentation is openly accessible, browsable and downloadable via the Archive web site). In addition, internally, staff can view processing and background information about any study onscreen, without the need to handle original materials. Digitisation is now a seamless part of data and documentation processing, although these days most material is received by the Archive in a computerised format.

### **OHPR (2004-07)**

The JISC-funded Online Historical Population Reports (OHPR) project undertook a major digitisation exercise to capture and mount on the web a complete collection of British historical population reports from 1801 to 1937. With almost 200,000 page of images, from all British censuses from 1801 to 1931, and a complete run of annual reports of the Registrar General until 1920, the web site, called histpop, is a comprehensive resource for students of population history. Additionally, the histpop site contains a vast quantity of contextual material, including essays and working documents from the General Register Office.

### **QUADS (2005-06)**

The ESRC Qualitative Archiving and Data Sharing Scheme (QUADS) aimed to develop and promote innovative methodological approaches to the archiving, sharing, re-use and secondary analysis of qualitative research and data, in all of their disparate shape and forms. A range of new models for increasing access to qualitative data resources, and for extending the reach and impact of qualitative studies was explored. The scheme also disseminated good practice in qualitative data sharing and research archiving.



## THE FAMILY OF SERVICES

A significant development over the 40-year history of the Archive has been the transition from an organisation running a single service to one managing a family of services. The Archive started as a department holding one grant from the SSRC in support of the Data Bank. Despite name changes, this essentially remained the situation for much of the first two decades. The change really got underway in the mid-1980s when, under the directorship of Howard Newby, the Archive began to attract project funding for other activities, most notably the Domesday Project and the Rural Areas Database. This process has continued and extended through the 1990s up to the present, in particular with a string of successful European Commission-funded projects (NESSTAR, FASTER, LIMBER and others), most of which have been orientated toward research and development work and the advancement of technical solutions.

In the last decade, however, in addition to separately funded, usually short term projects, the Archive has developed a number of discrete, yet interlinked, services. The first of these was the History Data Service, which started life in 1992. In 2001, Qualidata, a project originally located in the Department of Sociology at Essex moved over to the Archive, later to become ESDS Qualidata. The following year the Census Registration Service was formed as part of the ESRC/JISC census programme, and in 2003 the new distributed Economic and Social Data Service was launched in partnership with Manchester. The latest additions to the expanding portfolio of services have been the data support service for the Rural Economy and Land Use programme (2005) and the Nesstar support service (2007).

Thus the Archive is now a family of services – each with separate funding streams, accountability, objectives and deliverables – but linked by a single mission – the support of high quality research, through the discovery, delivery and preservation of data.



## HISTORY SERVICE

In January 1992 the History Data Unit, funded initially by the British Academy, the Leverhulme trust and eventually jointly by the AHRC and JISC was established as a specialist unit within the Archive. Three years later, the Arts and Humanities Research Council (AHRC) was set up as a distributed organisation comprising a managing executive and five service providers, with the Archive hosting the service for history. At this time the History Data Unit was renamed the History Data Service and formally became a part of the AHDS. In October 2003 AHDS 'service providers' became 'subject centres' and all the subject centres of the AHDS were renamed, with History Data Service (HDS) becoming AHDS History. The funding situation for

AHDS History is now uncertain, but the Archive retains a commitment to providing curation and preservation services for the historical community.

## CENSUS SERVICE

The Census Registration Service (CRS) project to provide a one-stop census access and registration service for UK higher and further education began in 2001. The service, launched in 2002, provided an integrated, seamless, user-friendly, one-stop registration system for access to all the varied census resources from the 1971, 1981, 1991 and 2001 decennial censuses. Registration for the various census products and materials was provided using a flexible web-based user interface, whilst still maintaining the access controls and authentication required by the participating census offices. In 2004, the service was rolled out to include a one-stop registration service for ESDS, allowing registered users access to all ESDS data services through the one-stop Athens Single Sign On. Development of the CRS is continuing, with the imminent arrival of Census.ac.uk which, whilst continuing to provide a seamless point of access to census data, will also offer many other services such as centralised searching across all census resources, help and support for users, census guides, workshops and online training materials.

## RURAL ECONOMY AND LAND USE PROGRAMME DATA SUPPORT SERVICE

The Rural Economy and Land Use Programme Data Support Service (RELU-DSS) provides a one-stop shop for RELU researchers and staff to gain information and guidance on issues surrounding data sharing and preservation, and information about third party data sources. Equally, the RELU-DSS provides information and recommendations to the RELU management bodies on key data management issues that relate to RELU projects.

The work of the RELU-DSS provides the relevant ESRC and Natural Environment Research Council (NERC) data centres (the Archive and the Centre for Ecology and Hydrology (CEH) respectively) with sufficient information to enable planning for management, longer-term preservation and maximised re-use value of RELU datasets.

A co-ordinated service has been established between the Archive and the CEH, with the lead being taken by the Archive. The Archive provides the first point of email and telephone contact and hosts the RELU-DSS information portal. The RELU-DSS provides reactive and proactive advice on data management to award holders and applicants and offers expert guidance to the RELU programme on data management issues for RELU projects.

## QUALIDATA

The ESRC Qualitative Data Archival Resource Centre - Qualidata - was set up in the Department of Sociology at the University of Essex in October 1994 to facilitate and document the archiving of qualitative data arising from research, whilst also drawing the academic communities' attention to its existence and potential. In 2001, Qualidata staff joined the Archive, consolidating the vision of a strategy to set up a specialist qualitative unit within the Archive. Initial action focused on merging the acquisitions sections of the two centres and then incorporating Qualidata catalogue records into the Archive catalogue. In 2003 Qualidata became ESDS Qualidata, one of the specialist services of ESDS.

## ECONOMIC AND SOCIAL DATA SERVICE

The Economic and Social Data Service is a national data archiving and dissemination service which came into operation in January 2003. The service is a jointly-funded initiative sponsored by the ESRC and the JISC.

The ESDS is a distributed service, based on a collaboration between four key centres of expertise: UK Data Archive (UKDA), University of Essex; Institute for Social and Economic Research (ISER), University of Essex; Manchester Information and Associated Services (MIMAS), University of Manchester; Cathie Marsh Centre for Census and Survey Research (CCSR), University of Manchester.

These centres work collaboratively to provide preservation, dissemination, user support and training for an extensive range of key economic and social data, both quantitative and qualitative, spanning many disciplines and themes. The ESDS provides an integrated service offering enhanced support for the secondary use of data across the research, learning and teaching communities.



# THE FUTURE



“Prediction”, observed Niels Bohr, the Danish Nobel prize-winning physicist, “is very difficult – especially about the future”. As this review of the Archive has demonstrated, change has been an important aspect of its forty-year life - indeed, the last few years have witnessed a number of rapid and sweeping changes. Thus, it would be naïve to suggest that one can predict with any certainty what the next decade, let alone the next forty years, will bring.

However, a number of trends are already underway, and it is probable that several of these will continue. The data-using communities which the Archive serves will certainly expand. This is not just about expanding numbers - which in itself presents major challenges - but the growing heterogeneity of the user community. Not only is the Archive serving academic researchers from a growing range of disciplines, but increasingly Archive supplied data are being used in learning and teaching at all levels. This expansion is mirrored outside of the academic sector, and is being driven by the increased availability of materials available for online browsing and analysis. In short, the Archive's user community is becoming more diverse, and will continue to do so. This is to be welcomed but it will also present new challenges, as diversity will inevitably give rise to differential expectations of support and data delivery.

Hand in hand with greater diversity, individual users will in future place greater demands for information drawn from across traditional disciplinary boundaries. This will entail not just a greater mixing of qualitative and quantitative data sources, but hopefully a breaking down of disciplinary divides by resorting to methods and data normally thought to be the preserve of others. In this scenario, the Archive will need to ensure that it provides appropriate bridges and gateways to subject-specific collections held by specialist centres, in particular in the humanities, the natural and environmental sciences, and within genetics, life sciences and epidemiology.

Related to this is the fact that the world of data is becoming increasingly fragmented and distributed. Gone are the days (if they ever existed) when a single data centre could claim to hold all the research data critical even for a specific subject. Thus a growing challenge for the Archive will be to ensure that the research communities that it serves can locate and access the data they require regardless of where they are held. This will require working with data creators, repositories and others to ensure that adequate metadata and finding aids are created and joined-up access and authentication procedures are in place.

As researchers increasingly aspire to work within a global data network, where data can be located and moved across international borders in a transparent fashion, the Archive will

need to work to ensure that wherever possible international barriers are broken down, by creating an interoperable distributed data infrastructure. A further challenge will be to ensure that data are secure, are not subject to misuse and that trusted mechanisms are in place whereby sensitive data can be used for research without infringing growing confidentiality concerns. This can only work through co-operation and trust.

While an increasingly distributed data landscape poses problems for resource discovery and access, of greater concern is the impact on sustainability and long-term preservation of data resources. Research projects and other data creators may well have a short-term desire or requirement to disseminate the material they hold, perhaps via institutional repositories, but they rarely have the resources, expertise or longevity to ensure the data are accessible and usable in the long-term. While institutional repositories will grow in importance and use, the need for centralised repositories for unwieldy or complicated research data, which demand painstaking work to prepare for preservation, will not disappear. Thus, to maximise research investments and to retain data resources for re-use, the Archive will need to play an increasingly important role in managing the data lifecycle and digital preservation. Also envisaged are mixed models, where, in some cases, access is provided by researchers and institutions in the short-term within a distributed setting, with the Archive taking longer-term custodianship within a process of informed records management.

All of these future challenges will require tools – tools for processing, discovery, access, authentication, security, dissemination, analytical, and preservation – as well as standards to power these tools. In the past, out of necessity, the Archive has largely invented or developed such tools itself. In the future, given the magnitude and interdisciplinary nature of the tasks ahead, it will need to work increasingly with others, within the e-Science, cyber infrastructure community and beyond.

Developing these tools must also reflect the fact that the nature of data is changing. The traditional boundaries between ‘data’ and ‘publication’ are being eroded. A future challenge will thus be to provide seamless access to research resources rather than data per se, whereby users can move easily from research sources and outputs as the gap between the two narrows and blurs.

Quite how far along these various roads the Archive will have travelled within the next forty years is open to speculation, but I am sure the journey will be eventful and exciting!

**Professor K. Schürer**  
Director, UK Data Archive



#### 1967

SSRC Data Bank created  
Allen Potter, Director  
Floppy disc invented at IBM  
Sexual Offences Act legalises homosexuality

#### 1968

Intel founded  
First release of SPSS  
Grosvenor Square demonstration, 200 arrested  
Students revolt at Essex University

#### 1969

Fire on campus destroys wooden building housing administration and clerical records of the Data Bank

#### 1973

Labour Force Survey starts

#### 1974

Ivor Crewe, Director  
Ceefax teletext introduced  
Two general elections in UK

#### 1975

450 studies in collection  
Affiliated with Inter-University Consortium for Political and Social Research (ICPSR)  
First edition of the Survey Archive Bulletin  
Sex Discrimination Act/Equal Pay Act

#### 1976

Council of European Social Science Data Archives (CESSDA) established  
Race Relations Act  
Apple Computer founded

#### 1981

Introduction of MS DOS 1.0

#### 1982

Survey Archive becomes SSRC Data Archive  
2,000 studies in collection  
Unemployment reaches 3 million  
Falklands War

#### 1983

Howard Newby, Director  
Introduction of CD

#### 1984

SSRC Data Archive becomes ESRC Data Archive  
Launched service for exchange of material on floppy disks  
One pound note withdrawn

# ACROSS THE DECADES OF

#### 1970

Development of UNIX started  
Formation of Office of Population Censuses and Surveys from General Register Office and Government Social Survey Department  
General Household Survey started

#### 1971

UK Census  
First email program  
First release of SAS

#### 1972

SSRC Data Bank becomes Survey Archive  
Ken Macdonald, Director  
Second catalogue released  
Unemployment reaches 1 million

#### 1977

Family Expenditure Study workshops  
Summer school in Survey Design and Analysis

#### 1978

Day seminar: Measuring the Quality of Life  
Archive Representatives Conference: Local Data Facilities

#### 1979

1,400 studies in collection  
First spreadsheet for personal computers – VisiCalc

#### 1980

Unemployment reaches 2 million

#### 1985

Domesday project  
Visit by Her Majesty the Queen  
First release of Microsoft Windows  
Sinclair C5 launched  
Ozone layer discovered

#### 1986

Purchased optical disc drives and a Micro Vax II computer  
'The Independent' newspaper launched

#### 1987

First annual CESSDA Expert Seminar  
Bridget Winstanley and Marcia Taylor won British Library Research and Development award

**1988**

Archive's 21st celebrations: seminar on 'Information Needs in the Social Sciences to the year 2000'

Tree planting

**1989**

Catherine Hakim, Director

WWW prototype built at CERN

**1990**

Paper copy of the ESRC Data Archive Catalogue cost £40

**1991**

Denise Lievesley, Director

ESRC Data Archive's first Compact Disc

Linux kernel born

Omnibus Survey

**1996**

ESRC Data Archive becomes The Data Archive

Digitisation project

NESSTAR project begins

Hotmail launched

**1997**

Five staff presented papers at IASSIST conference in Denmark

AMASS V4.5.8 hierarchical file management system acquired

Extension of offices and new telephone system

ERSC re-structured

**1998**

Launch of Great Britain Historical Database Online

**2001**

Download service begins

Census Registration Service

First issue of UK Databytes

COSMOS, AMRADS

iPod released

UK Census: 0.7 % of population recorded as Jedi!

**2002**

MADIERA, Go-Geo, GeoXwalk, MetaDater

**2003**

Start of Economic and Social Data Service (ESDS)

JISC X4L

1987 1989 1991 1993 1995 1997 1999 2001 2003 2005 2007

# OF THE UK DATA ARCHIVE

1988 1990 1992 1994 1996 1998 2000 2002 2004 2006

**1992**

Introduction of Windows 3.1

**1993**

3,000 studies in collection

BIRON available for external users

PDF 1.0 released

History Data Unit formed at ESRC Data Archive

**1994**

UKDA becomes web-based resource

Netscape Navigator replaces Mosaic

**1995**

HASSET available via web

Pentium Pro released

**1999**

Users able to browse, perform simple analysis and download certain selected data directly from the web site

Minimum Wage introduced

**2000**

Kevin Schürer, Director

The Data Archive becomes UK Data Archive

Introduction of 'shopping-basket' ordering facility

LIMBER, FASTER, Geo-projects MetaNet, CHCC

**2004**

ESDS one-stop registration service

DARP, OAIS/METS, VPS, OHPR

Mozilla Firefox released

**2005**

UKDA designated Place of Deposit for public records

Report on OAIS/METS standards

SAFARI UKDA, SQUAD, QUADS, SToRE

**2006**

RELU, DEXt, Census Portal

Data Catalogue

**2007**

Almost 6,000 studies in collection







UKDA is a designated place of deposit for digital records



University of Essex

JISC



UK Data Archive  
University of Essex  
Wivenhoe Park  
Colchester  
Essex  
CO4 3SQ  
UK

Tel: +44 (0)1206 872001  
Email: [archive@essex.ac.uk](mailto:archive@essex.ac.uk)