FEATURES

5 Samples of Anonymised Records from the 1991 Census for Great Britain
   by Angela Dale

13 Data Libraries as Vending Machines; Or, What We Can Learn From Arthur Dent
   by Laura Guy

20 Sober Ways, Politic Drifts and Amiable Persuasions; approaching the Information
   Highway from the Dusty Trail.
   by Ken Hannigan

27 MERGING CULTURES: Danish Integration of Academic Data Service into
   Traditional Archive System
   by Per Nielsen

36 Disseminating Data From Longitudinal Surveys: Issues Facing the Survey of Labour
   and Income Dynamics
   by Maryanne Webber

News

4 IASSIST President

42 IASSIST '96 Conference
Message from the President

Dear Colleagues,

Since taking office last May, I have been doing a lot of thinking about IASSIST and the role IASSIST can play in our professional lives. There are many changes taking place within the data world, largely fueled by politics, economics and enormous changes in technical capabilities. In order for IASSIST to move forward in this turbulence, we need to rethink the goals and mission of the organization. This evaluation can only take place successfully if there is input and participation from IASSIST members. While there will be a number of opportunities for member involvement in setting a future course, the first opportunity will be in early 1996. Along with your dues renewal notices, we will also be sending out a questionnaire, designed to survey members in their workplace, in terms of professional involvement and status, and in terms of adjusting to technological upheaval. Using a mail questionnaire is always a risky business, since the usual return rates are low. I hope you will take the time to respond as your replies will be important in determining what IASSIST does in the future, and it will be important for all of us to understand the specific nature of our profession. The results of the survey will be presented at our next annual conference, in Minneapolis, May 1996. I look forward to seeing you then.

Best Wishes,

Libbie Stephenson
ihe5dta@mvs.oac.ucla.edu
(310) 825-0716
Samples of Anonymised Records from the 1991 Census for Great Britain

by Angela Dale
Census Microdata Unit,
University of Manchester

Introduction
For the first time in a British census, the 1991 statistical output included Samples of Anonymised Records (SARs). Known as Census Microdata or Public Use Sample Tapes in other countries, SARs differ from traditional census output of tables of aggregated information in that abstracts of individual records are released. The released records do not conflict with the confidentiality assurances given when collecting census information since they contain neither names or addresses nor any other direct information which would lead to the identification of an individual or household. Essentially three per cent of records have been released in two samples. The SARs offer users the freedom to import individual-level census records into their own computing environment and the ability to produce their own tables or run analyses which are not possible using aggregated statistics.

Background to the release of the SARs
Requests had been made for SARs to be released from previous censuses in Great Britain. The principal stumbling block in the past had been an argument as to whether SARs could be considered a statistical abstract for release under Section 4.2 of the Census Act 1920 at the request and expense of user(s). Furthermore, in the past, requests for SARs had failed to reach a compromise between those (often geographers) wanting fine grain areal detail and those (often sociologists and demographers) wanting fine grain detail on other variables such as occupation.

The 1991 Census White Paper (Her Majesty’s Government 1988), however, announced:

“The Government intends that results from the 1991 Census should wherever practicable be made available in a convenient form to meet users’ needs”

Legal advice having been received that SARs could be deemed statistical abstracts, the White Paper went on to say:

“Requests for abstracts in the form of samples of anonymised records for individual people and households ... would also be considered, subject to the overriding need to ensure the confidentiality of individual data”.

The Economic and Social Research Council (ESRC) set up a working party to negotiate with the Census Offices and present a formal request. Their report, presented to the Census Offices in 1989 (subsequently published as Marsh, Skinner et al. 1991) concentrated on the benefits of releasing SARs, the uses to which they would be put, and also an assessment of the confidentiality risks involved in releasing SARs.

The request was mentioned by Ministers during the debate on the Census Order in Parliament at the end of 1989. Having considered the request, the Registrars General for England and Wales and for Scotland announced in July 1990 that they had agreed in principle to the release of SARs from the 1991 Census. There then followed detailed work by the Census Offices and ESRC in developing the statistical specification. An independent technical assessor, Professor Holt (University of Southampton), was appointed to advise the Registrars General on the confidentiality aspects and to write a report to Ministers. Following receipt of the report it was announced in March 1992 that two SARs from the censuses in England and Wales and in Scotland would be produced and released to ESRC. Similar SARs for Northern Ireland have also been made through an ESRC purchase. These allow the production of harmonised SARs for the whole of the United Kingdom.

Details of the SARs
Two SARs have been extracted from the GB censuses:

1 a two per cent sample of individuals in households and communal establishments; and
2 a one per cent hierarchical sample of households and individuals in those households.
The two per cent SAR has finer geographical detail and the one per cent SAR has finer detail on other variables, thus providing a solution to the conflict between users' demands discussed above.

The two per cent individual SAR contains some 1.12 million individual records (1 in 50 sample of the whole population enumerated in the census). It was selected from the base which lists persons at their place of enumeration. Details are given as to whether or not the person was a usual resident of that household, and if so (and enumerated in a household) whether they were present or absent on census night. The following other information is given for each sampled individual:

- details about the individual ranging from their age and sex to their employment status, occupation and social class;

- details about the accommodation in which the person is enumerated (such as the availability of a bath/shower and the tenure of the accommodation) or, if they were in a communal establishment, the establishment type (hotel, hospital, etc.);

- information about the sex, economic position (in employment, unemployed, etc.), and social class of the individual’s family head; and

- limited information about other members of the individual’s household (such as the number of persons with long-term illness and numbers of pensioners).

In effect, all the census topic variables listed are on the file; the only exceptions are variables either suppressed or grouped to maintain the confidentiality of the data. In all, there are about forty pieces of information about each individual, and the size of the raw data file, before any new variables have been derived and before any data compression techniques have been applied, is around 80 megabytes.

The one per cent household SAR contains some 240,000 household records together with sub-records, one for each person in the selected household. Information is available about the household’s accommodation together with information (similar to the two per cent sample) about each individual in the household and how they are related to the head of the household. The raw data is supplied as a hierarchical file in non-software specific character format (one line of information about housing and household, followed by one line of information about each individual in the household).

The full details of the information provided in both SARs are given in the Codebook and Glossary files produced by the Census Microdata Unit. Table 1, however, provides summaries by describing the information collected on the census form, the detail of coding of that information on the census database, and in how much detail that information is being released in the SARs.

The sampling procedure used
Census data goes through two separate coding processes. The easy to code information such as housing details, sex, date of birth, and country of birth is processed for all forms (100 per cent). The harder to code information such as occupation and industry is only processed for 10 per cent of forms. Both SARs were drawn from the 10 per cent sample so that they contain information from the whole of the census form. A detailed description of the sampling scheme for the SARs is given in Dale and Marsh (1993, chapter 11).

Confidentiality protection in the SARs
The census offices in some European countries have refused to release microdata because they believe, on the basis of research such as that conducted by Paass (1988) and Bethlehem et al. 1990), that the risks of disclosing information about respondents’ identities are too high. Much of this work is concerned with how many people have unique combinations of census characteristics which would make them open to identification. The Economic and Social Research Council Working Party which negotiated the release of the SARs took the view that uniqueness was only one part of a four-stage process of disclosure: data in the microdata file would have to be recorded in a compatible way to that in an outside file, the individual in an outside file would have to turn up in a SAR, the individual would have to have unique values of a set of key census variables and the matcher would need to be able to verify this uniqueness. Rough estimates of the size of risk at each stage were made; when cumulated, the risks of disclosure appeared very low; multiplying the various probabilities together, the working party concluded that the risk of anyone in the population being identifiable from their SAR record were extremely remote; their best estimate was something of the order of 1 in 4 million. (For more details of such calculations, consult Marsh, Skinner et al. 1991, Marsh, Dale and Skinner (1994) and Skinner, Marsh et al. 1992.) The arguments put forward were important in persuading the census offices to release the SARs suitably modified to protect anonymity where this was
felt at risk. In this section the various disclosure protection measures taken are described.

Sampling as protection
The low sampling fractions of the SARs offer a strong source of disclosure protection for sensitive data. It not only reduces the actual risk that a particular individual can be found in the census output, but it probably has its greatest effect by reducing the chances that anyone would make the attempt at identification by this means. The two SARs (a one per cent sample of households and a two per cent sample of individuals) are sufficiently small to offer a great deal of protection; the samples do not overlap so that the detailed household or occupational information available on the household file cannot be matched with the detailed geographical information available on the individual file.

Restricting geographical information
One of the key considerations which may affect the possibility of disclosure of information about an identifiable individual or household is the geographical level to be released (i.e how much detail is given about where the person was enumerated). The full census database holds information at enumeration district level (about 200 households or 500 persons in each ED) and even at unit postcode level (about 15 households). If released, such detailed geography would obviously pose a confidentiality risk. Empirical work and comparisons with SARs released in other countries showed that a sensible level for release would be areas equivalent to large local authority districts for the individual (2%) SAR.

To be separately identifiable, the decision was taken that an area had to have a population size of at least 120,000 in the mid-1989 estimates. The primary units used were local districts; only one geographical scheme was permitted, or smaller areas could be identified in the overlap, say between a local district and a health district. A population size of 120,000 is slightly higher than the lowest level of geography permitted in the US SARs (100,000), but it still has the advantage of allowing all non-metropolitan counties in England and Wales, most Scottish regions, all London boroughs (except the City of London), and all metropolitan districts to be separately identified.

Smaller local authority districts (under 120,000 population) were grouped to form areas over 120,000. Several rules were used to decide how districts should be amalgamated where this was necessary. First, the integrity of county/Scottish region geography was always maintained, where possible. Secondly, districts which achieved the minimum population threshold on their own were left intact, where possible; and smaller areas were grouped with each other. Thirdly, grouping was done on the basis of contiguity. And finally, if there was a choice left once the above criteria had been met, areas were grouped on the basis of their apparent social and historical similarity.

The one per cent household SAR, because of its hierarchical nature (i.e. statistics about the household and all its members), is more of a disclosure risk. For this reason it was decided that, for this SAR, the lowest geographical detail revealed would be the Registrar General’s Standard Regions, plus Wales and Scotland. The only exception is that the South East is split into Inner London, Outer London, and the Rest of the South East Region.

It should be noted that the order of records in both SARs has been re-arranged before the Census Offices release them. This is to prevent any possible tracing of individuals or households back through a region or district.

Suppression of data and grouping of categories
Some alterations have been made to the data to reduce the number of rare and possibly unique cases. The extent to which the variables on the local base have been either suppressed entirely or modified by grouping small categories before release in SARs is shown in Table 1.

Information which is unique in itself, such as names and addresses, has been omitted altogether; (technically these variables have not been suppressed since they are never put on the computer). Precise day and month of birth have been suppressed.

The thresholding rule
The degree of detail permitted on other variables was the subject of a thresholding rule which ensured that the expected value of any category at the lowest level of geography on any file was at least 1. The threshold, when operationalised, dictated that a category must have 25,000 cases in it in the GB file before it could be released on the individual SAR, or 2,700 cases before it could be released on the household SAR.

With some other variables, the smaller categories have been grouped, either across the entire range of the variable or only at the extremes (a process known as "top coding"). The rule used to decide the level of detail to be released was to group information categories to a sufficient detail so that, on average, the expected sample count would be at least one for each
category of each piece of information for the lowest geographical area permitted on each SAR.

Some justification for restricting attention to the distribution of the univariate categories of each variable in turn was given by Marsh et al (1994). They demonstrated that the risk of an individual having a unique combination of values of a set of variables could be predicted with a high degree of certainty simply from knowledge of their membership of rare categories of each variable taken singly. The precise cut-off at an expected value of 1 was set at a value sufficiently high to give reasonable protection of anonymity.

The rule was applied to each census variable. Expected counts were obtained by using 1981 Census frequency counts (supplemented by more recent surveys, for example the Labour Force Survey) at the national level for the whole population. To obtain expected counts, the count of 1 per category per SAR area was grossed up to the national level:

\[ C = \frac{1}{X} \times \frac{Y}{Z} \]

where

- \( C \) = expected count at the national level
- \( X \) = sampling fraction (1/50 for individual SAR and 1/100 for household SAR)
- \( Y \) = national population (56 million)
- \( Z \) = smallest geographical area population (120,000 for individual SAR and 2.1 million (East Anglia) for household SAR

Thus 25,000 and 2,700 were the two thresholds used for the individual and household SARs respectively. In theory, a small amount of random noise could have been added to certain variables in a manner analogous to the procedure adopted for the small area statistics. A technique similar to this has been used in the 1990 US Census for example: geography has been subject to a degree of perturbation by switching a small number of similar households between nearby areas (Navarro et al. 1990). However, the natural levels of noise in the data, combined with the analytical difficulties of minimising bias to both measures of location and spread by such techniques in a multipurpose file led to perturbation not being implemented in any form for the SARs.

Grouping of variables

When expected frequency counts fell below the threshold, categories were grouped. With some variables, grouping was only required at one end of the distribution: thus rooms were top-coded above 14 and the number of persons in the household was top-coded above 12. Two variables were both grouped and top coded; with age, 91 and 92 were grouped, 93 and 94 were grouped and 95 and over was top-coded; with hours of work, 71-80 hours per week has been grouped and the rest top-coded above 81.

When variables were not measured on a numeric scale, judgments had to be made about which categories to put together. Classifications for census data are often hierarchical. For example, for the Standard Occupational Classification there are 371 unit groups, 77 minor groups, 22 sub-major groups, and 9 major groups. In cases such as these, small categories could be amalgamated to the next level in the hierarchy. In other cases, detailed advice was sought from subject experts about how the groups should be formed.

In the case of three variables in the two per cent individual SAR, it was deemed necessary to further group categories, even though they contained numbers which fell above the threshold: occupation, industry, and subject of qualification. As a result of advice received from the Technical Assessor, occupation was reduced from the 220 categories proposed (out of a possible 371) to 73; similarly industry was cut from a possible 334 to 60 and subject of educational qualification from a possible 108 to 35. (Almost full occupational detail remains on the one per cent household SAR, however.)

There were other factors which determined the detail to be released:

- Categories of occupations and industries in the public eye were grouped further than mathematically necessary to guard against disclosure; for example, actors/actresses and professional sportsmen/women;

- Large households were seen as a disclosure risk in the household sample. Applying the frequency rule to size of household, a large household in the 1981 Census was estimated to be one of 12 persons or more. Consequently, only housing information is given for households containing 12 or more persons. No information about the individuals in the household is given.
<table>
<thead>
<tr>
<th>Item</th>
<th>Household (1%) sample</th>
<th>Individual (2%) sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of categories (maximum*)</td>
<td>No. of categories (maximum*)</td>
</tr>
<tr>
<td>Geographical area of enumeration</td>
<td>12</td>
<td>Standard regions of England (with split of South East into Inner London, Outer London)</td>
</tr>
<tr>
<td>amalgamated to form areas over and Rest), Wales and Scotland</td>
<td>120,000</td>
<td>Housing/household information</td>
</tr>
<tr>
<td>Accommodation type</td>
<td>14 (14)</td>
<td>Detached, semi-detached or flat in a commercial or residential building; converted or not self-contained accommodation in a shared house or flat</td>
</tr>
<tr>
<td>Availability of amenities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 bath/shower</td>
<td>3 (3)</td>
<td>Exclusive, shared or no use</td>
</tr>
<tr>
<td>1 inside WC</td>
<td>3 (3)</td>
<td>Exclusive, shared or no use</td>
</tr>
<tr>
<td>1 central heating</td>
<td>3 (3)</td>
<td>Full, part or none</td>
</tr>
<tr>
<td>Cars (number of)</td>
<td>4 (4)</td>
<td>0, 1, 2, 3 or more</td>
</tr>
<tr>
<td>Floor level (lowest), of accommodation (Scotland only)</td>
<td>7 (101)</td>
<td>Basement, ground, 1st/2nd, 3rd/4th, 5th/6th, 7th to 9th 10th or higher</td>
</tr>
<tr>
<td>Number of household (accommodation) spaces in dwelling</td>
<td>4 (35)</td>
<td>Top coded: 4 or more</td>
</tr>
<tr>
<td>Number of persons (enumerated) in household</td>
<td>12 (99)</td>
<td>Top coded: 12 or more</td>
</tr>
<tr>
<td>Number of residents in household</td>
<td>Derivable</td>
<td>4 (99)</td>
</tr>
<tr>
<td>Number of dependent children in household</td>
<td>Derivable</td>
<td>2 (99)</td>
</tr>
<tr>
<td>Number of pensioners in household</td>
<td>Derivable</td>
<td>2 (99)</td>
</tr>
<tr>
<td>Number of persons with long-term illness in household</td>
<td>Derivable</td>
<td>2 (99)</td>
</tr>
<tr>
<td>Number of persons in employment in household</td>
<td>Derivable</td>
<td>3 (99)</td>
</tr>
<tr>
<td>Number of rooms</td>
<td>15 (19)</td>
<td>Top coded: 15 or more</td>
</tr>
<tr>
<td>Description</td>
<td>Value</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Number of persons per room</td>
<td>Derivable 5</td>
<td>Ranging from less than 0.5 to more than 1.5</td>
</tr>
<tr>
<td>Tenure</td>
<td>10 (10)</td>
<td>Owner occupier or rented (public sector or private)</td>
</tr>
<tr>
<td>Wholly moving household indicator</td>
<td>2 (2)</td>
<td>Yes (all resident household members are migrants from the same address) or No</td>
</tr>
<tr>
<td>Individual information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>94 (111)</td>
<td>Single years 0 to 90, 91/92, 93/94, 95 and over</td>
</tr>
<tr>
<td>Status in communal establishment</td>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>Type of communal establishment</td>
<td></td>
<td>Not applicable</td>
</tr>
<tr>
<td>Country of birth</td>
<td>42 (102)</td>
<td></td>
</tr>
<tr>
<td>Migrants _ distance of move (km)</td>
<td>13</td>
<td>5, 10, 20 and 50 km bands; top coded above 200 km</td>
</tr>
<tr>
<td>Distance to work (km)</td>
<td>8</td>
<td>10 km bands; top coded above 40 km; 0.9 km band split 0-2, 3-4 and 5-9</td>
</tr>
<tr>
<td>Economic position primary</td>
<td>10 (12)</td>
<td>Employee, self-employed, unemployed, student, retired etc.</td>
</tr>
<tr>
<td>Economic position secondary</td>
<td>7 (10)</td>
<td></td>
</tr>
<tr>
<td>Economic position of family head</td>
<td></td>
<td>Derivable</td>
</tr>
<tr>
<td>Ethnic group</td>
<td>10 (10)</td>
<td></td>
</tr>
<tr>
<td>Family head indicator</td>
<td>2 (2)</td>
<td>Yes or no</td>
</tr>
<tr>
<td>Family number</td>
<td>5 (5)</td>
<td>Used to identify individual’s family</td>
</tr>
<tr>
<td>Family type</td>
<td>8 (8)</td>
<td>Married or cohabiting couple family with or without children or lone-parent family</td>
</tr>
<tr>
<td>Gaelic language (Scotland only)</td>
<td>5 (8)</td>
<td>Ability to speak, read or write Gaelic</td>
</tr>
<tr>
<td>Hours worked weekly</td>
<td>72 (99)</td>
<td>Single hours 0_70, 71 to 80, 81 or more</td>
</tr>
<tr>
<td>Industry of employees and self-employed</td>
<td>185 (334)</td>
<td>Mainly third digit (groups) of 1980 SIC</td>
</tr>
<tr>
<td>Category</td>
<td>Code</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Limiting long-term illness</td>
<td>2 (2)</td>
<td>Yes (individual has illness) or no</td>
</tr>
<tr>
<td>Marital status</td>
<td>5 (5)</td>
<td>As household sample</td>
</tr>
<tr>
<td>Migrant - geographical area of former residence</td>
<td>13</td>
<td>Standard regions of England (with split of South East), Wales, Scotland, outside GB</td>
</tr>
<tr>
<td>Occupation</td>
<td>358 (371)</td>
<td>Mainly unit groups of 73 (371) 1990 SOC Mainly minor groups</td>
</tr>
<tr>
<td>Number of higher educational qualifications</td>
<td>3 (7)</td>
<td>0, 1, 2 or more</td>
</tr>
<tr>
<td>Level of highest qualification</td>
<td>3 (3)</td>
<td>Higher degree, first degree, above GCE A-level</td>
</tr>
<tr>
<td>Subject of highest qualification Classification</td>
<td>88 (108)</td>
<td>Mainly third digit of Standard Subject Classification 35 (108) Mainly second digit of Standard Subject</td>
</tr>
<tr>
<td>Relationship to household head</td>
<td>17 (17)</td>
<td>8 (17)</td>
</tr>
<tr>
<td>Resident status</td>
<td>3 (3)</td>
<td>Present resident, absent, resident, visitor</td>
</tr>
<tr>
<td>Sex</td>
<td>2 (2)</td>
<td>As household sample</td>
</tr>
<tr>
<td>Sex of family head</td>
<td></td>
<td>Derivable</td>
</tr>
<tr>
<td>Social class</td>
<td>8 (8)</td>
<td>As household sample</td>
</tr>
<tr>
<td>Social class of family head</td>
<td></td>
<td>Derivable</td>
</tr>
<tr>
<td>Socioeconomic group</td>
<td>19 (20)</td>
<td>As household sample</td>
</tr>
<tr>
<td>Term-time address of students and school children</td>
<td>4</td>
<td>Inside or outside region of usual residence</td>
</tr>
<tr>
<td>Transport to work (mode)</td>
<td>10 (10)</td>
<td>As household sample</td>
</tr>
<tr>
<td>Visitor _ geographical area of residence</td>
<td>13</td>
<td>As household sample</td>
</tr>
<tr>
<td>Welsh language (Wales only)</td>
<td>5 (8)</td>
<td>Active use of (speak, read or write)</td>
</tr>
<tr>
<td>Workplace</td>
<td>5</td>
<td>Inside or outside region of usual residence</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inside or outside SAR area of usual residence</td>
</tr>
</tbody>
</table>

* The maximum number of categories as available on the full census database.
given.

- Geographical information for such items as workplace and migration (address one year before census) has been heavily grouped. This is because of the high likelihood of uniqueness of such information when used in conjunction with area of residence.

**Dissemination**

The licensing and distribution of the SARs is the responsibility of Manchester University who have a contract with the ESRC. The SARs may be used for both academic and non-academic purposes. All Higher Education Institutions (HEI) are required to sign an End User Licence Agreement which makes the HEI responsible for those members of their institution who are using the data. Users within each institution must be either members of staff or students and must sign a further individual registration form which contains a binding undertaking to respect the confidentiality of the data. Specifically, users have to guarantee not to use the SARs to attempt to obtain or derive information about an identified individual or household, nor to claim to have obtained such information. Furthermore, they have to undertake not to pass on copies of the raw data to unregistered users, and the Census Microdata Unit has the responsibility of auditing their use of the data. They must sign a statement that they understand that the consequences of any breach of the regulations on the part of any user in a specific institution can lead to the withdrawal of all copies of the data from that institution. Non-academic organisations sign a similar End User Licence Agreement and undertake not to allow the data to be user other than by their employees.

The data is free for the purposes of academic research; to get the data free the researcher must be doing the research in an institution qualified to receive an ESRC award, and the research must be funded either by the Universities Funding Council or one of the Research Councils. When the data is used either by those outside the academic sector or by researchers in universities for sponsored research, a charge is made for the data. In order to encourage a high volume of usage of a product whose advantages may not yet be well appreciated in Britain, these charges are being kept extremely low; an entire national SAR can be bought for £1,000 + VAT, and subsets of a county or local district for £500.

1 Paper presented at IASSIST 21st Annual Conference May 9-12, 1995, Quebec City, Canada.

**References**


12
Data Libraries as Vending Machines; Or, What We Can Learn From Arthur Dent

by Laura Guy 1,2
University of Wisconsin-Madison

"...technology causes trouble. As a major agent of change it intrinsically, not accidentally, dislocates and distresses established relationships and forces economic, political or social change."

Abstract
Technological change has had a tremendous impact on how we do our jobs. It not only has affected how we organize and provide access to information, but how our users conduct their research. This change has created new challenges for our profession, not the least of which is wondering if it will make us obsolete by replacing us with knowledge-based systems. The nature of these changes is discussed and we fantasize a bit about the data library vending machine. Finally, we look at our users and how we might best continue to provide them with the services that they need.

The Vending Machine Analogy
When I was a little girl we would visit my grandfather where he worked in one of the state office buildings in St. Paul, Minnesota. In the basement of this building there was a little cafeteria--a lunch counter. It was staffed by one or two people, and they sold sandwiches, soup and drinks.

I'm sure that similar lunch counters existed in office buildings all around the country. But if you go to one of these buildings now you will likely see a bank of vending machines.

What happened? Although it may be overly simplistic, it appears that the people lost their jobs to technology. What were the reasons? Vending machines may be cheaper. They take up less space. They are on duty 24 hours a day, seven days a week. They may be more efficient. They don't require vacation time or a health care plan. They don't complain about work conditions.

In the last few years, we have seen vending machine technology advance. Some of them can talk. They can take $1 bills and even $5 bills and make change. They dispense hot soup and cold salads.

The questions that we are dealing with are as follows: Can systems be developed that provide users access to data without the need for data librarians? To what extent can what we do be replaced by vending machines--that is--data vending machines? What are the tasks that can or should be automated or eliminated by technology, and after that happens, will there be anything left for us to do?

Coping with Change
I doubt that there is one of us who isn't simply breathless at the speed of the technological change that we are experiencing. Today's leading edge technologies quickly become commonplace. It's likely that as professionals our primary task over the next decade will be coping with this change. Much of the transformation is evolutionary in nature rather than fundamentally discontinuous: the change builds on itself. The speed and breadth of the transformation we are experiencing creates interesting challenges as well as opportunities. As the old axiom goes: "God protect me from living in interesting times."

These are indeed interesting and exiting times for us, brought in part on the following changes:

1. Change caused by technological advances in hardware:
   Client/Server technology: the architecture of the computer systems we use is changing rapidly. Hardware is becoming more "personal" and "portable." Personal devices are connected to powerful servers that are part of a distributed information system.

   Storage Media capacity: Multi-gigabyte local storage capacity is becoming commonplace. Advances in disk technology and compression rates facilitate the storage of vast amounts of information on-line and provide interactive access.

Summer 1995
**Network capacity:** The development of Ethernet and Token Ring networks enhances connectivity and provides extremely reliable data throughput. Today's hardwired networks will become tomorrow's wireless networks, enabling users to connect to anywhere from anywhere.

**Speed capacity:** Over the last few decades we have seen the doubling of raw CPU power every 18-24 months. Today's 20 and 50 MIPS machines will become tomorrow's 1,000 MIPS machines.

**Memory capacity:** We now have 16-megabit memory chips available, and soon we will see 64-megabit chips. Future computers will be able to store hundreds of thousands of typed pages in a computer's main memory and millions of pages on local disk drives.

2. **Breakthroughs in Software:**

New software architecture: The hardware revolution has enabled the development of easier-to-use software. This software in turn encourages the creation of collaboration tools and work group environments. The interfaces are more powerful, and applications are moving from the stand alone model to intelligent workflow. We see multi-tasking and multi-media capabilities.

The virtual environment: Graphic techniques and animation are transforming the way we visualize information and complex computations. Artificial Intelligence modeling and virtual reality coupled with enhanced visualization capabilities allow users to explore and interact with a virtual environment. If you think this sounds a bit farfetched, watch one of the America's Cup programs on ESPN. Software developed by Silicon Graphics incorporates a variety of measures such as wind speed and direction, compass readings, boat location, speed and course information into an astounding real-time graphic display.

Document-based systems: The concept of "document" has become more complex with the advent of hyperlinked data, text, graphics, video, sound, and so forth. The tools that operate on these documents also have become more sophisticated. Information becomes document-based when documents exist separately from the applications that create and operate upon them.

Object-oriented systems: Suited to modeling complex problems and processes, these advanced systems have the ability to self-update and communicate output in a variety of manners (voice, visual, etc.). Software tools are more modular and applications more flexible and powerful.

3. We see a paradigm shift from the data processing model to the information processing model as described by Ronald Weissman of NeXT*. In the new model information becomes content enriched, existing in an environment of "creativity" and "ambiguity" (an example might be data in a spreadsheet as opposed to information on the World Wide Web; the first is static, one-dimensional, unambiguous and incomplex, the second is dynamic, multi-media, possibly ambiguous and capable of presenting complex subject matter). In more concrete terms, we see documents linked with abstract, index, and bibliographic information, and numeric data linked with meta-data. Such documents may be more accessible and, for the general public, perhaps more captivating.

4. We've seen changes in our users: scholarly research methods have evolved into what Michelson/Rothenburg call network-mediated scholarship*. Scholarly communication and collaboration, as well as the broader research process, have undergone significant transformation. Many of the changes our users are experiencing are in large part technology driven.

5. We are experiencing changes in levels of connectivity. For those of us who are "wired" there is an enhanced ability to access, analyze, disseminate and communicate information instantaneously and without regard for distance.

6. There is an increasing amount of information published in electronic form (for example, the growth of government information) and a growing number of formats for electronic records and information such as e-mail, CD-ROM, magnetic tape, word processor publication, dial-up services, on-line services, G.I.S., spreadsheets, relational databases, floppy disks, and bulletin board systems.

7. Recently there has been a hypermedia revolution and accompanying it the concept of the nonlinear document (and the thought process behind it--which is not that new*!). We see multidimensional data that integrate diverse formats of information. Recall the information processing and the data processing models mentioned above. The new paradigm includes a growing complexity of systems, increasingly sophisticated applications, and a plethora of document types.
It will be essential for us as librarians and archivists to project and assess the importance of these changes. Clifford Lynch warns that in the 1980's "many research libraries...thought that users' needs for access to online database searching were substantially overstated." By not meeting the challenge of a transforming environment we risk making our libraries irrelevant and ourselves obsolete. And while we must remain agile, at the same time we need to be wary of developing systems and new services that may be poorly matched to the needs of our users, poorly designed or poorly implemented.

**Decentralization of Information Resources**

In the last several decades we have seen the transformation of traditional, paper-based, largely manual information systems into automated electronic systems. The obvious example is the library card catalog, which was transformed in the early 1980's by the development of online public access catalogs (OPACs).

We have experienced a more recent evolutionary change into a distributed or decentralized information environment. For example, by the late 1980's OPACs were beginning to be made available through the Internet. Later, to obviate the need to learn new user interfaces to access the various OPACs, the Z39.50 standard was developed. The standard is based on client/server technology and greatly facilitates network information access.

This evolution towards distributed resources closely follows the evolution of the Internet itself. The development of the early ARPANET in the late 1960's and early 1970's had a strong economic basis because of the great expense of computers; it enabled resource sharing. As technology became cheaper, the need to centralize (cost share) decreased.

In the 1980's the financial need for groups to come together to share computing resources decreased. The death of centralized mainframe computing followed closely the advent of minicomputers and micros. Increased networking capabilities allowed individuals to link to other individuals for reasons other than cost-sharing and without any regard for physical proximity. In the 1990's the dumb terminal attached to a mainframe computer has been replaced by the stand alone "scholar's workstation," which has powerful CPU, lots of disk space, a CD-ROM drive and an Internet connection.

A familiar manifestation of this decentralization has been the proliferation of information resources on the Internet. The sharing of information has been greatly facilitated by the development of tools such as anonymous FTP, the Wide Area Information Server (WAIS), and the World Wide Web. Research projects that collect data such as the National Survey of Families and Households (NSFH) now have the technological ability to cheaply and easily become their own access providers. Individuals now have the capability to become resource centers. Bill Goeffe's *Resources for Economists on the Internet*, is a good example of what one person can accomplish using technology no more sophisticated than what can fit on a desktop.

Information decentralization causes librarians quake in their shoes. And rightfully so: as the anarchistic nature of the Internet intrudes into information systems there is a recognized lack of standardization, centralized authority, and access control. There is no institutional control over individuals like Bill Goeffe, who may in the twinkling of an eye forsake his resource and allow it to lapse. Similarly with the NSFH, one might ask what happens to the project's data when the funding ends or the research group disbands?

In a decentralized or distributed environment access becomes independent of location. The physical location of a resource has little meaning. In fact, something that virtually appears as a single resource might physically exist in separate parts in disparate locations. From the user's viewpoint it is not really important where the resource lives*. In this new information model our old notions of control require reformation. Who is responsible for a resource that consists of multiple copies dispersed among multiple locations? New forms of control may be required for insuring the continuation of important information resources.

**Data Libraries as Vending Machines**

As librarians and archivists we exist to serve users and preserve information. If we are to be replaced by vending machines it will be because they serve users and preserve information better than we do. They would be cheaper, replacing staff and facilities with computer hardware and software; they would be easier to use, enhancing the scholar's workstation and providing access to needed information from the desktop; they would be faster, allowing for the instantaneous access to any information at any time; they would be decentralized, maximizing connectivity through resource sharing; and, they will be without boundaries, providing access to users no matter where they may be.

Most data users go through a process that consists of four separate steps:

---

*Summer 1995*
1. identification of potential data sources
2. determination of the usefulness of the data
3. obtaining access to the data
4. obtaining analyses from the data

Any data vending machine environment would have to play a role in each of these steps. It would have to assist users possessing varying levels of experience as they make their way through a research sequence that is not necessarily linear; for example, it may not be until the user reaches step 4 that they realize that the data do not meet their needs. The system would have to deal with the lack of standardized terminology to describe data and the lack of standards for formatting and storing data.

The data library vending machine would have to be smart. It would need to be capable of a multitude of sophisticated activities such as conducting reference interviews, answering questions, conducting searches, and assisting with problems. It would need to be capable of serving a diverse group of users that spans a wide range of abilities, including computer knowledge, communication skills, and research expertise. It would need to be able to instruct as well as assist, and preserve as well as make accessible.

Who Are Our Users and What Do They Want?
This is an obvious assertion that we need to know who our users are and how we can best meet their information needs. To be sure, they are a diverse group, but in general we can talk about them in terms of what they are usually involved in: the research process. This process can be divided into five parts:

1. identification of sources
2. communication with colleagues
3. interpretation and analysis of data
4. dissemination of research findings
5. curriculum development/instruction of next generation

The research process has undergone a tremendous amount of transformation. These changes in turn have an impact on how we, and others associated with the research process, do our jobs. The research paradigm includes a diverse group of agents: researchers, publishers, computer specialists, vendors, librarians, archivists, and professional organizations. Information technology has influenced changes in all of them.

In the last several decades end-user computing has become more convenient, cheaper, faster, more powerful, and easier to use with sophisticated interfaces and the advent of interactive environments. Technology has provided increased connectivity that enhances the research process through the expansion of access and the facilitation of communication.

As librarians we are more commonly serving a "remote clientele." Users frequently do not need or wish to "come through the door" to receive assistance. This change, although gradual, impacts directly on the way we provide our services. Much of the communication is electronic in nature. The need for remote access to documentation as well as data is a part of our users' growing demands and expectations for timely and adequate service.

We've seen a transformation in the capabilities of our users. Some are very technologically sophisticated and have access to the most powerful computing resources. These people, typically faculty in my environment, prefer to have minimal interaction with the library. They want what they want when they want it, and prefer to require little human assistance.

I can imagine this group working well in a "data library vending machine" environment of the type I believe is practical within the constraints of today's technologies: an environment that includes FTP or WWW accessible data and meta-data, intelligent searching, extraction, and analysis capabilities.
Others users, often but not always students, have varying degrees of computer sophistication and varying access to resources. Many have very little experience with FTP or the World Wide Web. They do not have the computer resources required to, for example, download large files off the Internet. Many lack a basic understanding of the nature of numeric data.

I have difficulty thinking of this second group as a potential vending machine clientele. The front end that would be needed to teach these people what they need to know in order to use any given data set is beyond my imagination.

We are tied inextricably to the information needs of our constituency and must always monitor these needs carefully. It is critical that we work with our users and the other agents involved in meeting their needs (for example information providers and computer scientists) to continue to push forward information management.

The "Expert Systems" Fear

Expert systems, sometimes called "knowledge-based systems" are a type of computer program that uses the knowledge-based techniques of Artificial Intelligence. Simon Hayward described them as computer programs that represent knowledge and apply expertise to manipulate that knowledge and to achieve solutions. Over the last decade or more the fear of being replaced by expert systems has sounded in many places. By-in-large, this has not come to happen. Certainly we will see, coming out of Artificial Intelligence, the development of intelligent agent technology that will provide aids for locating, evaluating, analyzing and interpreting information. But, for data librarians the fear of being replaced by sophisticated expert systems that interface data and meta-data with users is a concern that may, for the foreseeable future, be unwarranted. The omniscient, omnipotent and omnipresent data library vending machine described above will probably not exist for a long time.

Nevertheless, we must consider: To what extent can what we do be replaced by expert systems? Alternatively, what is it that we do would we like to see eliminated by technology? What might we do with the extra time that will facilitate our users and enhance our role in the research process? And, can this process be viewed positively rather than as a threat?

An Informal Users' Survey:

During the months of March and April, 1995, I conducted an informal survey of users of the DPLS. I kept track of the complexity of their needs, their level of sophistication, and their access to resources. The question I wanted to answer was the following: which of these users would operate well in a data vending machine environment?

I found that 30% of our users would be well-served by the vending machine data library. They are experienced and familiar enough with numeric data that they need very little help to use it; these people just need to know where the data are. As more data are made available publicly on the Internet (for example LABSTAT, the PSID, and the Penn Tables) we will have less frequent contact with these users.

The other 70% are the users who don't know what data sets are available or which data they want to use. They don't have the experience working with data to understand what they need to do to access it or how statistical software works. There are many who simply like to come in and talk about their ideas and appreciate whatever type of feedback they might get. Some users who have little experience using computers are scared and need comforting and hand-holding. Some of our users prefer to look at paper-based resources. There are those who don't have a clue as to how to do secondary analysis.

The continued growth of CD-ROM publishing, and the development of increasingly sophisticated interfaces will change the above percentages. For example, compare an Internet-based interface that merely provides a raw data set (like FTP) with an interface that allows people to do SAS or SPSS runs interactively in real-time on a remotely stored data set (such as can be built today on the World Wide Web). Or compare the latter with an on-line data center system that provides all meta-data associated with a multitude of data sets, including variable-level information, instruments, methodology discussions, bibliographies, and so on.

The percentages will also be changed by the users themselves. Over the years we've seen an increased recognition among scholars of the importance of quantitative analysis skills, and a spreading of this recognition to less traditional fields such as history and education. But as the sheer numbers of users doing quantitative research increases, they are coming to us more technologically sophisticated because they are exposed to computers at an ever earlier age.

The new technologies described above promise to eliminate some of the more tedious parts of our jobs such as tape
rollovers and extracting data for users. This will free us up to deal with important issues such as proper levels of documentation, the structure of the information systems providing data and meta-data, and bibliographic, abstracting and indexing problems. It allows us to devote more time to assisting users, teaching, and developing standards and policies. Its likely that in the process we will redefine what it is that we do and what it means to be a "librarian."

We need to ask ourselves what does the word "library" mean? is a library just a physical place or might it become more? An environment? A mind-set? A virtual world? Or, could it become less? A computer attached to the Internet? We must also ask ourselves as user demands and expectations grow, can we provide? 

So What Does Arthur Dent Have To Do With All This?
When I first started to think about this panel I remembered something that I had read:

"After a fairly shaky start to the day, Arthur's mind was beginning to reassemble itself from the shell-shocked fragments the previous day had left him with. He had found a Nutri-Matic machine which had provided him with a plastic cup filled with a liquid that was almost, but not quite, entirely unlike tea. The way it functioned was very interesting. When the Drink button was pressed it made an instant but highly detailed examination of the subject's taste buds, a spectroscopic analysis of the subject's metabolism and then sent tiny experimental signals down the neural pathways to the taste centers of the subject's brain to see what was likely to go down well. However, no one knew quite why it did this because it invariably delivered a cupful of liquid that was almost, but not quite, entirely unlike tea." The Hitchhiker's Guide to the Galaxy"

There is a real danger that the vending machine data library will almost, but not quite, entirely be inadequate.

The Human Factor
In the vending machine data library might the human factor be missed? It is important to keep in mind that in electronic systems the human factor plays a very important role. In the words of one science fiction author, it is the human factor which gives these systems their "heart." Thus, it will be humans who must deal with pertinent issues such as standards development, information integrity, accountability, and responsibility; it will be humans who realize the importance of metadata as an essential supplement to standard bibliographic approaches; it will be humans who design the systems, who implement policies and develop the tools and criteria by which the systems will operate; it will be humans who will develop new descriptive systems, finding aids, navigational aids and informational hooks that are suited to the constantly changing electronic environment and user demands.

There are equations where the human factor hasn't been missed, for example, in bowling alleys there once were "pin boys." It is a cold hard fact that if an entity can be adequately replaced by technology under the current ësystem it probably doesn't deserve to survive. Interestingly, at the University of Wisconsin, human-staffed delicatessens are proliferating. Why? The profit motive? The need for Human Interaction? Could it be that we still do need to have that perfect cup of tea?

Conclusion
We live in a time when concepts like "unique" and "multiple" are becoming obscure, as are "library" and "archive." Our survival and transformation as a profession depends on how we respond to the changes we are experiencing and the new paradigm in which we find ourselves. It is important that these changes not be perceived as threats but as opportunities, and that we work to turn what might be weaknesses into strengths. As the information infrastructure becomes stable and established, focus will shift to areas that are open to contributions that we can play an important role in making. For example:

- Meta-data engineering for better methods and tools for describing information.
- Development of new descriptive systems and finding aids.
- Development of access tools to facilitate navigation and information retrieval.
- Development of improved user interfaces.
- Development of new governance and control mechanisms over information.
Standards for document and data management dealing with diverse areas such as scanning, text encoding, and storage, and retrieval issues.

Insuring the continuation and widening of the information infrastructure.

Teaching colleagues and users.

Copyright, intellectual property, privacy, and public use issues.

Promotion of the archival mandate and the protection and preservation of electronic information.

We must temper "carpe diem" in an environment of decreasing funds, private sector competition, and the need for us to develop advanced skills and expertise. In light of the changes we are experiencing, it is clear that there are many challenges ahead if we are to remain a viable and useful profession. These challenges will require innovation, agility and deep understanding of our environment and the needs of our users. Defining these challenges may help us to continue and grow as a profession; meeting them will definitely lead us to live in interesting times.

1 Paper presented at IASSIST 21st Annual Conference May 9-12, 1995, Quebec City, Canada.

2. I would like to thank Derek Zahn for his substantial contributions to this paper.


9. That is, until it disappears.


Sober Ways, Politic Drifts and Amiable Persuasions; approaching the Information Highway from the Dusty Trail.

by Ken Hannigan
National Archives of Ireland

The term "National Archives" usually conveys an image of a large organisation with a staff numbering several thousands, as in the National Archives of Canada or the United States, or several hundreds, as in most of the national archives in Europe. It should be made clear from the outset, however, that the National Archives of Ireland must be considered on a much smaller scale. Ireland is a small country on the periphery of Europe with a small population (just over 3.5 million in the Republic of Ireland) and the National Archives of Ireland in Dublin can be seen to reflect the size of this population base. Not only are we smaller than most national archives, we are smaller even than the specialist divisions of many national archives. We are smaller, for instance, than the Center for Electronic records in the US National Archives.

Our total staff numbers 35; our total professional staff numbers 13. We are, therefore, comparable in many ways to some of the state archives in the United States. In fact on the evidence of Richard Cox's recent study, The First Generation of Electronic Records Archivists in the United States, there are many points of similarity between the situation obtaining in state archives in the United States, and the situation obtaining both in the National Archives of Ireland and among the archival profession generally in Ireland.2

The National Archives of Ireland has existed under this name only since 1988 when the National Archives Act (1986) came into effect, though the constituent parts of our organisation, the State Paper Office and the Public Record Office of Ireland, have existed separately since 1702 and 1867 respectively and have been part of a de facto amalgamation since the late nineteenth century. The National Archives Act has radically transformed the role of our organisation, however, and has given us responsibilities similar to those of the National Archives of Canada and Australia. We now have a thirty year rule of access for government records, and no such documents may be disposed of without the written consent of the Director of the National Archives. Our Act placed an enormous burden on us, with accumulations of documents dating from the beginning of the state and before, and formerly not covered by legislation, having to be processed. At the same time as our responsibilities have expanded so dramatically, our traditional business has also been increasing significantly. We have an annual readership of 17,000. This may not be huge by the standards of most national archives (according to a recent notice posted on the "Archives" listserv, the number of people accessing the New York State Archives gopher in January 1995 was 17,000, the same as our readership for the whole of last year) but we have experienced a huge increase in public access in a generation amounting to a tenfold increase in the last twenty-three years.

Our user profile is very different to that of a data archives or library. Over 50% of the readers' tickets which we issued in the first three months of this year were issued to people undertaking genealogical research on their own families. This statistic has a bearing on the sort of service we must provide and how priorities are addressed. Tourism is Ireland's second largest industry (after agriculture). There is a huge Irish diaspora in North America, Australia and the UK, and it is from this that most of the tourist traffic comes. The roots factor is an important element in all of this and we are, whether or not we would wish to be so, part of the roots industry. Some 37% of our readers come from abroad, most of them tracing their roots, and they form a constituency which we must be careful to service.

Apart from genealogists, amateur and professional, the remainder of our readers are divided between academic researchers, local historians, teachers and trainee teachers, and a considerable body of legal searchers.

As to the documents being produced, the emphasis here is also heavily on genealogy. The household returns of the 1901 and 1911 censuses (which are, respectively, the earliest Irish census for which full household returns are extant, and the latest census for which the household returns are open to inspection) accounted for 42% of all documents produced to readers in the first quarter of this year. The 1901 census alone accounted for 26% of all documents produced in this period. Far behind the census, the next largest categories were:

modern departmental records (22%)
eighteenth and nineteenth century State papers (11%)
and
testamentary records (7%)

Like most national or state archives, we must face two ways at once. We are expected to provide a service to a research
public which is largely composed of genealogists, and we
must provide a service to Government, to appraise its records
which must be authorised for disposal or accepted for
transfer. We must balance our obligations to the research
public and to government with our obligations to a third
constituency-posterity. We must preserve an adequate
record of our own time and continue to preserve the records
of previous ages which have been entrusted to our care.

We have 13 professional archivists on our staff. This is a
small enough number, but in relative terms these 13
constitute a sizeable proportion of the professional body of
archivists in Ireland. Total membership of that body at
present numbers 67. Increasingly, candidates for jobs in
archives are required to have a post-graduate qualification in
archival studies. In the past twenty-five years archivists have
professionalised, indeed it could be said that it is only in the
last twenty five years that the profession has been defined in
Ireland. Most current holders of the Diploma in Archival
Studies are graduates of the only archives school in Ireland,
that in University College Dublin, and so that school, to a
large extent, controls entry to the profession. However, many
of us in mid-career, particularly in the state sector, have no
specialist archival qualification. We are all arts graduates,
however, most with history degrees.

Because of our low numbers, there is no separate Irish
professional organisation for archivists; we form an Irish
region within the Society of Archivists, the bulk of whose
members are in the United Kingdom. Our professional focus
and contacts, therefore, have tended to be with our
colleagues in the United Kingdom with whom we have much
in common. And so to the dusty trail.

Dust is certainly a metaphor with which traditional archivists
in Ireland and the UK are familiar, though not, perhaps,
entirely comfortable. Dust and decay are an essential part of
our popular image and this image is one of the problems
which we face in approaching the superhighway. It is likely
that a word association test administered to the average
person in the street in Ireland would result in a string such as
"archives, dust, decay, dead, buried". "Buried in the
Archives" is a phrase we frequently hear used in relation to
documents, or even in relation to ourselves as archivists!
Thus the following statement which a national daily
newspaper in Ireland recently published as part of an
interview with one of the country's leading popular
composers is probably fairly representative of popular
attitudes:

"I honestly do believe that merely sticking with the past is
for archivists. Forging new forms for the future, on the other
hand, is for the living."

Well, certainly there is a sense in which archivists are seen to
be, if not actually dead, then as having escaped from life. We
hear frequent reports of people being told by career guidance
counsellors or teachers that a career in archives is an option
for those of a shy retiring nature, or timid disposition, who
might find an alternative, teaching, for instance, or career
guidance counselling, perhaps, too hard on the nerves.

We tend to have a cobweb-enshrouded image largely based
(as Richard Kesner has identified it) on the popular notion
that archivists are antiquarians, that we are a little removed
from everyday life. We are not entirely blameless in this
regard. Some of us have cultivated the image of the
antiquarian, perhaps many of us are attracted by this self-
image and have even been attracted to the profession by it.
So there may be something of a self-fulfilling prophesy at
work here, as the world of traditional archives has attracted
those who have consciously not wanted to be part of a
thrusting, aggressive, brash, profiteering, macho world. We
are mostly history graduates; we are people who put
posterity above profit and power.

The world of archives is also a very stable one. Within the
archival profession in Ireland today most us who have been
there for ten years or more are doing the same jobs which we
were doing ten years ago - and in the same organisations.
Few of us have experienced anything else in our professional
lives. It is not typical of the organisations with which we do
business, the organisations for whose records we are
responsible. It is certainly not typical of the IT people with
whom we come in contact but who disappear out of our orbit
again with bewildering speed. This stability has left many of
us locked into practices and perspectives which are anti-
dynamic. And as most of us are burdened by the daily
demands of keeping a public service going and overwhelmed
by backlogs of unlisted and unappraised records, it is
frequently not until systems break down that we consider
change.

There is a large element of this present in our response to
computers. There was a time, not so long ago, when
archivists could get away with a statement like "I know
nothing about computers" and even make this sound like a
virtue. We were helped in this by the fact that our favourite
constituency of readers - historians - by and large also tended
to spurn computers. It is of course no longer fashionable for
archivists to admit that they know nothing about computers.
Even the most obsturately antiquarian of us have by now
realised that computers are, or should be, essential tools of
the trade. But we are not yet really at home with them. We
have not as a profession come fully to terms with the impact
of automation. It is a fact that the largest special interest
group within the Society of Archivists is the IT Group, but
within that group to date we have tended to concentrate very
narrowly on a single aspect of computerisation, and the most
popular events organised by that group are software
demonstrations. We are terribly interested in learning how
computers can help us to continue doing the things we have
always done in the ways we have always known and loved.
We have come to the conclusion that computers are probably
a good thing, we certainly want to know a little more about
them, but really, we are not technical people and we still tend
to revel a little in the fact. These attitudes put us at a considerable disadvantage in coming to terms with the wider aspects of computerisation.

Automation has implications for the specialised functions of "traditional" archives in three main areas.

Firstly there is the question of automating the archival tasks, accessioning, repository management, and so on. This should not pose any difficulties for traditional archives. We are basically talking about stock control here, something which is eminently suited to automation.

Secondly there is the obligation to provide an efficient and reliable service to readers and potential researchers, including the obligation to provide and disseminate information about our holdings. We are in the information business, though we do not all see it this way, and computers are tools for information management.

Thirdly there is the increasingly worrying question of what to do about the records generated by computers. These three aspects cannot be divorced; our failure to come to terms with the first two leaves us ill-equipped to deal with the third.

Many people from outside the world of archives, and even some archivists, are surprised at the failure of archives in Europe to automate more rapidly. In a recent issue of The American Archivist, Ronald Weissman expressed astonishment at finding a newly-created series of handwritten finding aids at the new State Archives in Florence. There would be no difficulty in finding similar instances of archives all over Ireland and the UK tenaciously holding on to the old methods.

There are two main problems which we face in automating and which partly, though not totally, explain our slow progress. One is quite simply the question of resources. It seems that archives everywhere are low on the priorities of governments and funding agencies. The country will not grind to a halt if the archives fail to function efficiently. The business of archival management does not generally attract large-scale commitment of resources. Our very modest degree of computerisation in the National Archives of Ireland has been achieved in a piecemeal manner and without the benefit of a specialist IT unit.

The second problem attaching to automation is potentially more difficult to resolve. Effective automation of archives demands consistent descriptive standards, ideally ones which are universally accepted. Unfavourable comparisons are frequently made between the extent of our computerisation and that obtaining in even fairly modest county libraries where users see the benefits of online catalogues and bar-coding systems. There are of course some fundamental differences between archives and libraries, though these are not perceived by an impatient public, despite the efforts of both the professional librarians and professional archivists to delineate the two professions. The fact that our collections come ready-made, that rather than being a continuous series of single-level items our collections sometimes involve complex arrangements, and that retention or recreation of the original order is a cardinal rule of archival description, these have all posed problems for traditional archives the world over in their attempts to computerise their services and exchange information on their holdings. Despite some heroic, some would say quixotic, efforts, there is no universally accepted standard of archival description in Europe or even within the Society of Archivists in the UK and Ireland, nor is there any widely-used or agreed software for archives such as the Dynix system for libraries. To computerise the archives is to plough a lonely furrow.

There have undoubtedly been some fairly sophisticated archival automation systems in Europe. The Public Record Office in London has since the nineteen seventies operated a computerised ordering system which is still far ahead of what is available in most other archives in Ireland or the United Kingdom. The current updating and extension of that system will put the PRO very much ahead of the field again. In France, computerisation allows not only for online searching of finding aids but also for remote access and advance ordering, something which is made possible by widespread use of the Minitel videotext system in that country, a degree of use unparalleled in any other European country (France's Minitel system accounted for 87.41% of all European videotext terminals in 1993). The Historical Archives of the European Union in Florence has, since 1993, provided online access to its database finding aids on the European Commission's Echo co-host. Spain is also well advanced towards linking its various state archives in one network which will allow remote access to all of them.

Online access to finding aids is still very much the exception rather than the rule for European archives, however, and most computer-based projects have tended to be exclusive to each institution. There has been little or no co-ordination among or between archives, no sharing of information other than what is already available over publicly accessible channels, no cross-fertilisation. The systems are mostly not compatible with each other and do not lend themselves to the sort of inter-institutional exchange of information that is now the norm for libraries. There is a commitment at high level to do something Europe-wide about automation and there is in existence a group of experts, comprising the heads of all national archives in the European Union, charged with coordinating archival policy and practice including archival automation, but a large part of the problem is that the senior managers, the heads of archives, who are attempting to formulate common policies in this area, are in general themselves not terribly comfortable with technology and, therefore, not sure what it is they wish to do. Despite a commitment to harmonisation and co-operation at the top, there has been little contact or co-operation among archives.
and archivists further down the hierarchy across national and linguistic boundaries.

A major part of the problem in Europe is also of course the difficulty of language. One indication of this is evident on the Internet. The archives listservs in North America are not paralleled in Europe (though a small “Archives and the Internet” discussion group has just this year been established within the IT group of the Society of Archivists in the UK and Ireland and may develop into a listserv. [Author’s note: since this paper was presented, the “Archives and the Internet” discussion group has become a very vigorous forum for exchange of information among archivists in the UK and Ireland.]). While there has been criticism of the American “Archives” listserv from within the profession in the United States, it represents a very useful forum of over 2000 archivists exchanging information on matters of common concern. The US and Canadian listservs are certainly of considerable benefit to those of us who access them from outside North America. It is significant, though perfectly understandable, that those subscribers to the listservs who are outside the United States and Canada are mainly in the English speaking world, and predominantly in Australia and New Zealand. On the “Archives” listserv there are, for instance, only two subscribers from Germany (the country which accounts for 28% of the IT market in Europe) and none from France which, in terms of archival automation, is arguably the most advanced of the larger countries in Europe.

It is obvious that the Internet as a whole is still overwhelmingly a North American phenomenon. But this area is developing rapidly in the UK and Ireland. In Europe the number of computers directly accessible on the Internet has doubled every year for the last three years, but in Ireland within the last year, the number has tripled, and all the signs are that this is continuing to mushroom. The tendency until now within Ireland and the UK has been for Internet access to come mainly from the academic community. It is not common for government employees to have access to the Internet as part of their work, so there is no “.gov” element in our addresses. High telephone charges in Europe compared to those in the United States and the disparate nature of the telephone systems, which have coincided fairly rigidly with national boundaries, have inhibited access to the Internet by private individuals. Also household computer ownership in Europe is only about a third of that obtaining in the United States. Nevertheless, just by looking around one can see that things are changing. The fact that the next version of Microsoft Windows will come bundled with an Internet access program (Microsoft itself functioning as an internet access provider) will almost certainly result in a huge new wave of Irish and UK connections from outside academia. For those archivists who connect, there will probably be a gravitational pull, at least initially, towards North America rather than into Europe. Despite commitments to further cooperation and harmonisation in Europe, it is likely that the real dynamic will exist, for the moment, on the Internet.

Given that there has been no listserv for archivists in the UK and Ireland, presence on the American “Archives” listserv is probably a reasonable guide to the number of archivists who are using the Internet in these countries, and the number of archivists who are on the Internet in these countries is probably in turn something of an indicator of the extent to which archivists have themselves embraced the new technologies [Author’s note: since this paper was presented, the “Archives and the Internet” discussion group has become a very vigorous forum for exchange of information among archivists in the UK and Ireland.]. Relative to the size of their populations, Australia and New Zealand are leagues ahead of the UK, and Ireland hardly figures. In this context it may also be significant that more than half of those appearing on the “Archives” listserv with UK addresses are in university archives rather than state or official archives.

The Internet has huge potential for satisfying one of our primary needs, the need to disseminate information on our services and holdings to potential readers, and particularly to that diaspora of roots enthusiasts which we must cultivate. Some traditional archives have already started to run gophers or to put up Web pages. Although our own computerisation is not very far advanced, we have considered it important to establish a presence on the Internet and now have some pages on the World Wide Web by courtesy of a neighbouring third level college which has kindly afforded us space on their server. There is clearly going to be growing demand for us to provide more and more information online. There will be growing pressure from Europe to service a free information market to match that being developed in the United States. Academics will surely soon start demanding that we use the available resources to improve access for them. It is rather surprising that they have been so reticent to date. In the light of the statistic of 17,000 people accessing SARA’s gopher in January this year, we await with some trepidation the consequences of our own heads appearing above the parapet of the superhighway.

As “traditional” archivists, we have much to learn from the pool of available knowledge on the Internet in many areas, but particularly in relation to the problem of electronic records, which represents one of our greatest challenges, if not our greatest challenge, but has as yet has caused very few ripples to appear on the surface of the archival waters in Europe.

We in Ireland have a National Archives Act as strong as most comparable archives acts and one which gives us statutory powers in respect of digital data. Our Act specifically defines “Departmental records” to include magnetic tapes and discs, optical or video disks, and other machine-readable records. In fact there has been some debate over whether our Act, in specifying types of media, such as tapes and disks, has rather missed the point and concentrated on the medium rather than the message (a major part of the problem being of course that you can happily preserve mountains of disks and tapes but this will
not guarantee that the data remain accessible). However, we are confident that such definition is not exclusive and we regard the terms "files" and "other documentary or processed material" mentioned in our act to be media-transparent. It is the message that we are charged with preserving.

The main problem however, is not one of definition, it is the problem of what we do to give effect to our Act. We have not yet managed to seriously address the challenge posed by electronic records, but we are not alone in this regard. Although traditional archives in Europe are aware of the challenge posed by digital data, progress to date in addressing it has been very slow. According to a recent study presented to the Canberra conference on electronic records last November, no national archives in Europe has yet got beyond the stage of holding the output of anything other than database systems, and many of us have not even got that far. In the United Kingdom and Ireland the strongest player on the archival field and therefore the one that leads the way in many respects, the Public Record Office in London, despite a number of high level studies of the issue going back over twenty five years, has yet to decide a policy on electronic records. Things now seem to be moving in Britain, however, with the appointment in late April 1995 of an Information Manager in the Public Record Office specifically charged with the task of developing a strategy for handling electronic records, and the appointment of a powerful committee of senior officials to ensure that he functions with the necessary support. It also seems certain that a formal decision will be made that archival electronic records in the form of structured datasets will be lodged with an existing agency rather than in the Public Record Office itself and that preservation of digital data will continue to be outsourced. In fact the existence of the ESRC Data Archive in Essex as the de facto place of deposit for official electronic archives in the United Kingdom has probably allowed the Public Record Office the luxury of time on this issue. Most of the large datasets which might have been identified for preservation by the Public Record Office have probably been preserved in Essex.

Elsewhere in Europe surprisingly little has yet been achieved. Per Nielsen has outlined exciting developments in Denmark which may offer a blueprint for some other countries. Of the other National Archives in the European Union, it seems that only those in Finland, France, Germany and Sweden have themselves accessioned electronic records and these mostly consist of datasets. The National Archives of the Netherlands, however, has taken the initiative in attempting to bring the question of electronic records onto the archival agenda in Europe.

Traditional archives seem to have suffered a paralysis in confronting this issue which has presented them with problems of two types. Firstly there are obvious problems associated with the preservation and future accessibility of such records - instability of storage media necessitating regular migration of data, rapid hardware and software obsolescence. There is no need to recite these to an audience of data archivists. It is possible that we in Ireland have already lost some of the large datasets created in our large information-gathering departments. We do not know, and our very preliminary efforts to find out, based as they are on our own ignorance of systems, have been inconclusive to say the least. The responses we have received have tended to be blandly reassuring, disturbingly so in the context of what we know to be the practice of some of these agencies in relation to their paper records. Given that we do not yet know how we are going to address this problem, we have not yet probed too deeply. That said, we have found the level of response to our preliminary questionnaires to be disappointingly low, the lack of response indicating, perhaps, a belief among IT managers that we are not there to help them.

The second area of concern for traditional archives relates more to what has been termed the second generation of electronic records, the records of the electronic office, and to what has been called the distributed environment in which electronic records are being created. Alongside the spread of computers has gone the breakdown of central file registries and filing systems. Everyone creates their own documents and files them on the hard drives of their PCs or on personal directories or even on floppy disks. We find a multiplicity of systems, a multiplicity of software packages being used on them, a multiplicity of drafts and duplicates being stored in them. Finding our way through this maze will be a colossal task.

The traditional practice of traditional archivists, appraising records when the records have reached the end of their life-cycle is clearly not appropriate in the case of electronic records. If we wait until the records cease to be current or until they are released into the public domain in thirty years, or even twenty years, time there may be nothing left to appraise. There is a coincidence of developments here which is alarming. The last twenty five years or so, a period which has seen and is continuing to see the transition from paper to digital records, is also the period which has seen a generation of archivists professionalise. We are in the process of climbing into our professional fortresses and pulling up the drawbridges behind us, making it more difficult for those from other than a very narrow spectrum of training to enter the profession. But it is ironic that this generation of archivists, which has been so careful to professionalise, to define standards, may be the generation which will fail most spectacularly to leave behind a record of its own time.

The options for traditional archives faced by the problem of what to do about electronic records are threefold. We can decide to use existing data archives and libraries as places of deposit and even perhaps develop an organisational link with these archives along Danish lines; we can try to establish our own data archives as an integral part of the existing archives;
or we can insist that archival electronic records be maintained by the creating agencies, with our organisations providing an inspectorate to ensure that such records are adequately catered for by the creating agency. It is unlikely that the deposit of official digital data with an existing data archives will be the strategy followed in Ireland, despite the fact that this seems to be about to happen in the United Kingdom. There are various reasons why this is unlikely to be our route but the strongest one is that there is no such entity as a data archives currently existing in Ireland. As to our becoming a data archives, it has to be asked, and it has been asked, if it is at all appropriate for "traditional" archives to accession electronic records other than as a last resort? Would we be placing ourselves on a treadmill to maintain access to these records, something which may be done only by relegating other aspects of our responsibilities? Would the archives be able to administer whatever privacy laws may regulate access to such data in the future? With so many systems current throughout the organisations for whose records we are ultimately responsible, would we have to become museums of software and hardware systems? The last question scarcely bears thinking about. As it is, we "traditional" archivists can barely master our own software and hardware.

There is a compelling logic to the arguments advanced by David Bearman and Margaret Hedstrom in favour of a non-custodial approach by traditional archives to such records\(^19\). The fact that the Australian Archives will now opt for this kind of approach, as set out in recently published guidelines, will weigh heavily in its favour with those of us who have yet to make a decision in this area\(^20\). This question will be addressed at European level in the Spring of 1996 when a major multidisciplinary forum will be called in Brussels to be attended by representatives of archives as well as IT specialists from throughout the European Union. This meeting will be held under the auspices of the European Commission and will attempt to co-ordinate policy on machine readable records. It seems very likely that this forum will be influenced by decisions taken by the Australians and by the very forceful arguments emanating from Pitsburgh.

Yet there is a huge caveat which must be entered here, as Edward Higgs has recently warned elsewhere\(^21\), our previous experiences with some of the agencies which would have to become custodians of archival data do not inspire total confidence. Yet it seems at the moment that, even with this caveat, local retention is the only practical option open to us in Ireland - though this of course may change. As mentioned above, traditional archivists in Ireland are not computer people. We in the National Archives do not at present have the resources to manage these records. It is unlikely that we will be given them in the short term, not on the sort of scale that would make the job feasible, and there is little merit in embarking an a project with a better than even chance of failure.

It is simply very difficult to force the issue of electronic records onto the archival agenda or indeed onto any agenda. Few people are interested. There is no pressure group or no constituency outside the archives demanding that something to be done about electronic records.

Historians in Ireland have not seriously begun to use such records (some of them are now engaged in setting up databases of economic statistics or even online textual databases, but they have not yet begun to lobby on behalf of existing machine readable records). The late John Blackwell who addressed the Amsterdam conference of IASSIST in 1985 made some attempts to raise the issue in Ireland, but seems to have met with little support\(^22\). If we were to close our reading room in order to stocktake, were we to withdraw a heavily used series of records and substitute microfilms, we could be fairly sure of a loud and unfavourable reaction from our research public. But if we choose to do something which will actually result in catastrophic consequences, if we ignore electronic records, no one will notice for a long time. No one outside the world of archives is currently lobbying about electronic records. This is something that we in the archives have to worry about for the moment on our own, sure in the knowledge that if we continue doing nothing will have left a shameful legacy.

We must seek allies in attempting to give electronic record keeping a higher priority. There are some developments which indicate where we might find these allies. Freedom of Information legislation is imminent in Ireland. There is a strong political commitment to this at present and the legislation currently promised looks set to be a far-reaching measure with radical effect. There will be major consequences both for the archives and for the holders of official information. For the archives, Freedom of Information, together with Data Protection, may eventually supplant the National Archives Act and the 30 year rule as the regulator of access. There are, anyway, moves in Europe to have the norm for access reduced to twenty five or twenty years\(^23\). The gap, therefore, between current records and non-current records is likely to diminish. As for the information-creating agencies, they will have to be more accountable for the information they create and hold, in whatever form it is held. Something like the traditional registry system will have to be reinstated, but perhaps with routes of access from the outside world. And this system will of course have to encompass electronic records. Perhaps a Government Information Locator System may be used in the future as a route into unpublished official information or archival information, or at least into the finding aids for such information, and may ultimately support a gateway for online access to archival electronic records.

Whether traditional archives become non-custodial regulators of electronic records or custodians of such records, or, more likely, become a combination of both, we will clearly have to acquire the knowledge and skills which will allow us to make intelligent and correct decisions on the
scheduling of such records. Given our background and training and what has been to date an unimpressive track record with computers, it is unlikely that we traditional archivists will easily turn ourselves into electronic archivists. No-where within the profession in Ireland at the moment are there the skills required to tackle this job. We are, however, greatly heartened by the news that one of the staff of the Center for Electronic Records at NARA, Mark Conrad, has been selected under the Fulbright scheme to spend the next academic year teaching in the Archives Department of University College Dublin. This is a hugely significant development in terms of archival formation in Ireland and we may soon see the emergence of a generation of Irish archivists with some skills in the management of electronic records. Perhaps we in the traditional archives also need to make more radical plans now for a period of transition, and look outside our traditional recruiting pool to train new archivists for a new age. We should, to the extent that we can, encourage into the profession some from a technical rather than an arts background. And certainly “traditional” archivists must seek to forge stronger links with the data archivists and librarians, for it seems that we are now on the same road, having travelled to it from very different starting points.

1 Paper presented at IASSIST 21st Annual Conference May 9-12, 1995, Quebec City, Canada.


3 The Irish Times, 10 February 1995, p12.

4 Quoted in Cox, op. cit., p.40


8 ibid.

9 List of subscribers to “Archives” and “Arcan-L” listservs supplied on 26 and 28 April 1995.


11 Emerging Technologies, p.5

12 The address for the National Archives of Ireland Home Page at the Dublin Institute of Technology is <http://147.252.133.152/nat-arch/>.


14 ibid.


16 See Per Nielsen, “Merging Cultures: Danish Integration of Academic Data Services into a Traditional Archival System” elsewhere in this volume.

17 Higgs, op cit.


21 Higgs, “Information Highways”


MERGING CULTURES: Danish Integration of Academic Data Service into Traditional Archive System

by Per Nielsen
Danish Data Archives

ABSTRACT:
The historical outline of the Danish Data Archives as an academic service facility is outlined. The reasoning underlying a recent and globally unique organizational affiliation of the DDA, viz. to the group of traditional archives, is presented; and the stages of the merging cultures process are outlined. The appropriateness of the archives integration is demonstrated in a presentation of projects that were not feasible in the old university affiliation of the DDA. An outlook towards future projects is also given.

Background and history of the DDA
The Danish Data Archives (DDA) is slightly older than IASSIST. Founded in 1973, this author joined the DDA in 1974 - early enough to be there (in Toronto) when IASSIST was established as a “grass-roots organization” of individuals working in or using data archives, data centers, data libraries, or what these academic service facilities were named in each country or state. In the following, we shall refer to such installations as Data Organizations (DOs).

To some extent, IASSIST was set up as the response from the Old Boys' Network to the claim of the 1968-generation of more influence or power; to some extent IASSIST was set up to bridge the gap between the (predominantly male staffed and dominated) European Academic Data Archives and the (astonishingly female influenced) North American Data Libraries. IASSIST represented a merging of cultures according to generation, gender, and geography.

Be it as it may: IASSIST has survived with astonishingly small adjustments in a changing environment, borne by the enthusiasm and energetic work of (especially North American) individuals. During the same period, many DOs have undergone substantial changes. This report provides an overview documentation of some of these changes in a small country (Denmark, 5.2 million inhabitants) and refers in the form of parallels to changes in a number of other countries, predominantly in Europe.

The DDA was established on April 1st, 1973, after several years of preparation within the Danish Social Science Research Council (SSRC), as a feasibility project dealing with archiving and servicing problems related to three major data types:

1.1.1. Political and Social Survey Data, i.e. questionnaire-collected research data resources in a de facto anonymous form. This was the “typical DO activity”, known from e.g. the ICPSR, the Zentralarchiv, and the ESRC Data Archive.

1.1.2. Economic Time Series and to some extent regional data, both in terms of contents and methods (harmonization, adjustments to regional changes, etc.) and computer handling systems. This area, especially the regional data aspect, was known from Norway, where the NSD was started in 1971.

1.1.3. Population Register Data, i.e. identifiable data on individuals. There were no known DOs active in this area, but it was expected to be central in the future.

Needless to say, it was the advent of the computer and the challenges inherent with its use that was the rationale behind the project. During the Feasibility Project Period (1973-1976), the staff (predominantly engineers!) were occupied with all the technicalities of the computer age; there was, unfortunately, less knowledge (or even ignorance) vis-a-vis the substantive issues within the research disciplines potentially contributing data to and using data from the project.

It is symptomatic for the situation that a Steering Committee consisting of former researchers, now research administrators (viz. the SSRC chairman, an organization professor from a Business School; the Director of the Danish National Institute of Social Research (ISR), a government research facility of considerable magnitude and influence; the Director of Danmarks Statistik (the Danish Central Statistical Office, CSO); and the Director of the National Archives (also heading the Provincial Archives) would establish the project with almost exclusively engineers as staff members. It illustrates the attitude that the computer age was still so young that only technical specialists were able to deal with the matter. Technicians were the priesthood of the time.

When the author of this article (an economist by training, but rather a sociologist by practice) was accepted as a staff member (February 1st, 1974), he was the first non-engineer in full-time employment as an academic staff member within areas 1.1.1 and 1.1.3 above (there were economists in 1.1.2,
which lived its own life); only one half-time student had a social science training. Many years were to pass until the technical education and skills were considered the "side product" and the social science background was the focus of the staff qualifications.

By the end of the three-year feasibility project, the first "culture clash" emerged within the staff. The few social scientists felt that the (many) engineers were not appropriately contributing to the development of the organization and, especially, to its integration in the research milieu of the universities and other schools of higher education within the social sciences (broadly conceived). Already at this early stage did we (the economists) demand that all staff members reported their time spent on different (detailed) subprojects; of course, the time-use statistics calculated showed that there were too many engineers and a lack of social scientists if we were to fulfill the plans defined by the Steering Committee.

1.2. Interim SSRC Period of Transition 1976-1978

Given that the Danish SSRC (contrary to the situation in e.g. Norway and the UK, where the SSRCs finance the DOs to a great extent even to-day) had a formal limitation on the period of time in which the Council was allowed to run projects (three years), negotiations were carried out by the mid-seventies to find the lasting host of the DDA. Little by little, it was realized that a strong base in a social science research environment was more important than the technicalities; therefore, negotiations were carried out with the relatively large, public "midwife-institutions" (the ISR, the CSO, and the National Archives) to urge one of these established organizations to adopt the techno-baby. Given that not enough breeding monies were offered to keep the baby alive at its present size, the negotiations failed.

Internally, partly because the SSRC gradually shrunk the money sack and partly because the technicians ran projects according to their own interest rather than to the benefit of the baby (shown by the subproject time registration referred to above), a change in staff policy had to take place. More social science trained staff were employed when vacancies appeared (which they did frequently, because highly qualified computer people were in high demand everywhere); and, more importantly, the scope of the DDA was narrowed considerably: Both the Time Series Subproject and the Population Register Subproject (1.1.2 and 1.1.3 above) were abolished, and only the Survey Subproject (1.1.1) was kept in the final model.

Furthermore, the second "culture clash" emerged, this time involving external agents: The understanding and confidence between the DDA Director (an engineer) and the SSRC members (social scientists) deteriorated; and, in 1977, the directorship moved to the social science side when the former director returned to concentrate on his own private consultancy firm.

The final organizational belonging of the DDA ended up being decided by opportunistic political/bureaucratic considerations rather than substantive research concerns: The Ministry of Research and Education (whose minister happened to come from and be elected MP in Odense!) found it relevant to support the smaller university centers rather than the big universities; consequently, Odense University was urged (it cost them money!) to take the baby into custody.


Formally by April 1st, 1978 (five years after establishment), the DDA was moved to Odense; the physical move took place at the turn of the year 1978/79. Looking in the rear-view mirror, this turned out to be the beginning of the consolidation decade, the happy childhood of the baby: After initial fighting over relative budget sizes, the DDA ended up in a stable and acceptable economic situation.

Organizational, the DDA was set up with a double reference structure: On one side, as employees of the University, the DDA had to follow the rules of the Rector and Board of the University. On the other side, the DDA had an external Board of Overseers (five persons) who took care of the more narrow inspection of the activities and the development of the organization. In practice, to be honest, the DDA Director and the staff took most of the strategic decisions during this decade of consolidation; the baby was free to mature according to its own qualifications and cumulation of experience. More and more, the DDA staff identified with the "DO culture" (acquired and supported from international cooperation on many different levels and in many different projects) rather than anything else. (This is the kind of "data archive movement" culture that has kept IASSIST going strong for so many years.)

All was well; nobody questioned the relevance of the DO culture or the utility of the DDA activities, and most staff members considered the Odense University affiliation a permanent one. But alas! · The centre-right governments of the mid-eighties saw it as their major task to shrink the public sector, and the universities had reductions in their budgets at the same time as there was in increase in student enrolments. Universities had to critically inspect their resource allocation; and, needless to say, the eyes of the Odense University administration fell on the DDA during that process: The university demanded that the 3 academic staff members of the DDA (all with titles of associate professors) should participate in the normal social science curriculum of the university, teaching in the same amount of time as all other professors at the university.

The DDA staff argued that (1) formally, the DDA was an institute with national coverage, not an institute with special contribution to Odense University; (2) the teaching obligation of the academic staff was fulfilled in national
training programs rather than in the Odense University curriculum; (3) the Ministry of Research and Education gave the budget of the DDA directly, exactly in order to make sure that the archive could fulfill its national obligations.


In fact, this was the fourth “culture clash”, viz. between more and more strangled university administrators and DDA’s relatively “anarchistic DO culture” (in the best sense of the term). When it turned out that the “stubborn rector and top administration” of the University were not willing to listen to the arguments of the DDA director and staff, we told them that we had to discuss the situation and our future with the DDA Board of Overseers. Of course it was annoying and frustrating for the university top management to see that their subordinates did not just obey orders (which they were supposed to do under the “visible management model” which was in fashion).

The DDA director and academic staff told the university administration that we would opt for a review process - if the DDA Board was in agreement. Fortunately, the Board Members were in agreement; they were even enthusiastic about such a step, because the Evaluation or Review Mania had floated over the country as a politically correct measure in the years of budget cutting.

A review committee of six established researchers was set up (nominated by three research councils - social science, humanities, and medicine - and three important research institutions). The review committee report was generally favourable seen from the viewpoint of the DDA Board and staff; they presented a number of recommendations among which the organizational ones are of interest in this context:

The uncertain leadership structure should be abolished; it had been inadequate right from the outset and was critical in times of crisis.

The DDA should be relocated institutionally, and six possible solutions to the organizational setting were proposed for the DDA Board to further negotiate. (Odense University was not among the institutions recommended; they had been so negative in the review process that they disqualified themselves in the eyes of the review committee.)

After discussions with the involved research councils (for social science, humanities, and medicine) and the major research milieus within the same disciplines the Board could start negotiating a final placement for the DDA, now an adolescent. There were three organizational belongings that were considered interesting, viz.:

1.4.1. The DDA as a unit within the Danish CSO

(Danmarks Statistik). It took only one meeting to be turned down: The Director of the CSO held that the two cultures could not be merged, especially due to two incompatible phenomena: (1) Where the DDA had always tried to push their (anonymized) data on as many users as possible, the CSO had the principle of keeping their (identifiable) data strictly within the organization itself. (2) Where the DDA had always succeeded in keeping their services free of charge, the CSO tried to earn a big fraction of their total budget by user payments. [Needless to say, the DDA interest in a CSO placement was exactly to change that big organization in a more service-oriented direction.]

1.4.2. The DDA as a unit within the Danish Computer Center for Research and Higher Education (UNI*C). The UNI*C Director was interested; she felt that the center should add substance to its predominantly technical services, and they were under transformation so that the integration would be feasible at short notice. The DDA could choose between Copenhagen and Aarhus if they were to go for that model. The transaction was bureaucratically simple, because the DDA would stay within the realm of the same government department, viz. the Ministry of Research and Education.

1.4.3. The DDA as a unit within the National Archives. Here again, the DDA Board was met with relatively open arms (i.a. because the outgoing Director of the Danish National Archives had been functioning two periods (6 years) on the DDA Board during the mid-eighties). The DDA could stay in Odense, because the archives were spread over the country anyway. The transaction was bureaucratically more complicated, because the DDA would have to change government department, moving from the Ministry of Research and Education to the Department of Culture. There was an incalculable risk of losing money during such a transfer.

All in all, we were quite satisfied with the negotiations. Getting a “yes” in two out of three proposals is not all that bad! Several rounds of negotiations were carried out with the management of the two possible hosts; I think it is fair to simplify the matters to the following decisive elements that distinguished the two:

**Continuity:** Because the DDA could continue its activity in Odense in the National Archive-model, there was no risk of loss of professional capacity in that model. There was a risk of losing substantial parts of the “DO culture” in a geographical move - and thus a risk of assimilation with the new culture.
(maybe even annihilation of the “DO culture”) rather than integration into the new culture with DDA’s own cultural identity relatively intact.

Permanence: The National Archives, being several hundred years old already, and being one of the only institutions mentioned in the Constitution, will survive new centuries. UNIC, on the other hand, was already undergoing severe changes in business plans - in transition from being predominantly a mainframe host to having a wider agenda: Mainframe host (parallel processors and other very expensive equipment), facility management host, network administrator, and value added services agent.

Substance: The major argument, however, was that the substance dealt with in the traditional archives and in the DDA was the same: Both are information agents, the major difference being the data-carrier - which will change in the traditional archives anyhow. Many avenues of DDA development were more easily passable in the National Archives model than in the UNIC model.

The choice having been made, only the bureaucratic work remained; and even though this process took considerably longer time than expected it ended successfully: As of January 1st, 1993, the DDA was a unit in what had, in the newly enacted Archives Act, been named the Danish State Archives (SA). We could thank our Board Members (whose assignment period had twice been prolonged with one year because the transition took so long to carry through), and we were cast in the arms of a new host

1.5. Independent Unit in the Danish State Archives Group from 1993

The “anarchistic DO culture” had to be integrated into the “bureaucratic civil servant culture” according to the decisions taken. As always when you move in with new people, there was some reluctance and cautiousness from both sides: From the DDA point of view, we insisted on staying separate for some time to secure (reassure) the independence; we were not going to be “swallowed” by this, as we considered, somewhat “dusty” system ten times larger than we were.

The entry avenue was paved with a number of lucky circumstances: (1) A new Director of the National Archives entered the arena a couple of years before us, and he came from the university and research circles, too; (2) yet another unit had been adopted in the State Archives only three months before us; (3) a modernization process had been started within the archives themselves. Partly due to these circumstances, the entry into the New World (which is a very Old World!) was successful and seems to develop to the benefit of both sides. Before looking into that, however, we shall make a short digression to a description of our new “family” and then return to a specification of the potentials of the new affiliation of the DDA.

Short Description of the Danish State Archives

Before the advent of the Archives Act of 1992, the State Archives were referred to as “The National Archives and the Provincial Archives” - most of which were century-old. In the Archives Act of 1992, the State Archives (SA) was defined as a group; we shall very briefly introduce these institutions and the rest of the archives complex in the country.

The Danish State Archives have less than 200 man-years at their disposal; quite a considerable number of the employees, furthermore, are not regular employees; rather, they are unemployed or disabled persons undergoing training or rehabilitation programs on behalf of social authorities.

The staff-size of the Danish State Archives in comparison with the size of the state administration that they serve is considerably lower than in the other Nordic countries, a fact which has been demonstrated to the politicians again and again.

2.1. The National Archives

The National Archives (Rigsarkivet) and its predecessors (i.e. Geheimearkivet, the Secret Archives) date back some 400 years. The institution is located face-to-face with the Danish Parliament (Folketinget). With approx. 80 man-years available, the National Archives is obliged to make an appraisal of all documentary material in central government and archive what is deemed necessary from legality considerations and to document the present for future researchers.

The National Archives is divided into an Appraisal Branch (incl. a private archives unit, a military archives unit, and an MRDA unit) and a Servicing Branch; also, the institution hosts the Secretariat of the whole group of the Danish State Archives.

2.2. The Provincial Archives

There are four Provincial Archives. Their purpose is to provide archival facilities for government agencies spread over the country. Also, voluntarily, the county and municipality administrations may deposit their archives with the provincial archives; however, they have to pay. Even so, they have to abide by the principles for appraisal defined by the State Archives (formally: The Director of the National Archives).

Three of the Provincial Archives (for Zealand and the other islands east of the Great Belt, in Copenhagen; for the island of Funen in Odense; and for Northern Jutland in Viborg) are exactly 100 years old here in the mid-nineties. The fourth Provincial Archives, that of Southern Jutland in Aabenraa, is only about 60 years old. It was established some years after
the Referendum in 1920 which brought Southern Jutland back under the Danish Crown; it cooperates closely with archives in Schleswig which remained German as an outcome of the Referendum.

Two Provincial Archives (in Copenhagen and Viborg) are “big” (approx. 35 man-years), two others (in Aabenraa and Odense) are small (approx. 10 man-years).

2.3. The Danish National Business History Archives
Founded as an independent state-financed institution in the fifties, the Danish National Business History Archives tries to reflect all aspects of business life: It holds archives from firms and business units as well as from organizations (employers’ organizations, employee’s organizations, private organizations and associations) as well as from individuals with a certain standing.

Needless to say, before as well as after the entry of the Danish National Business History Archives into the State Archives Group (entry per October 1st, 1992), there has been a need to define the functional dividing lines between that institution and the private unit within the National Archives.

Opposite the major volume within the National Archives and the Provincial Archives, the Danish National Business History Archives has to rely exclusively on voluntary depositing of material (much like the DDA); they have no legal claim that donors shall archive their administrative remains.

The Danish National Business History Archives has less than 15 man-years at its disposal; within that frame, it also serves as a municipality archive for the city of Aarhus where it is situated.

2.4. The Danish Data Archives
The DDA entered the “family” on January 1st of 1993; it had about 10 man-years of staff-time at its disposal in the operating budget when entering. Due to the uncertainties regarding affiliation in the late eighties and early nineties, it had become extremely difficult to attract research grants to augment the total level of activity.

Needless to say, there are donors of computer archives that may either deposit at the National Archives (MRDF unit) or at the DDA; we shall refer to the “functional integration” in some detail below.

2.5. Other Archive Groups (not in the Danish State Archives)
Outside the “family”, a number of archive institutions are of interest in terms of collaborative projects (private archival material) as well as because they rely on the definitions of the SA in terms of appraisal (City and Local Archives). The major groups are:

2.5.1. The National Library: As per tradition, many private papers (especially from writers, artists and other actors in the cultural realm) end up in the National Library (next tobour to the National Archives in Copenhagen).

2.5.2. The Labour Movement’s Library and Archives: Financed by the Labour Unions, this Library and Archive documents the labour movement in Denmark and is thus also predominantly in the private archives sector.

2.5.3. The City and Local Archives: According to the Archives Act of 1992, counties and municipalities have an obligation to keep their records according to the decisions taken by the Director of the Danish National Archives; however, they do not have to deposit the records with the Provincial Archives. More than a dozen of big city municipalities have established City Archives with a professionally trained archivist (usually a historian) as the head. In many minor municipalities, the Local Archives have been staffed only with amateurs in the past. From the Archives Act of 1992, however, Local Archives have to be part of the Municipality Administration and professionally managed; otherwise, the records shall be deposited with the Provincial Archives of the relevant region (paid for by the municipality).

3. Advantages and Disadvantages of Archives Integration
From the national viewpoint, the Archives Act of 1992 explicitly regulated that all public authorities shall deposit their archives in a “professional” archive institution. This is, of course, an important step in the direction of securing future historical research at all levels, the national, regional, and local.

In this section, however, we shall return from the digressional “family description” and look at the advantages and disadvantages of the integration of the Academic Service Facility (the DDA) into the Traditional Archive System (the SA). Without doubt, the viewing angle is that of the DDA - due to the fact that the author is placed there, and because that is the “natural” IASSIST platform for evaluation.

3.1. The “Laissez-faire Period”
As already touched upon above, the first year or so in the new family was characterized by a “laissez-faire” state of affairs in the sense that all parts did what they used to do without much interference. It was a period of gradual confidence-building. However, the period was also one where the activities of all units in the SA were thoroughly documented in a 3-volume Action Plan.

When the DDA entered the SA, they were in the middle of this documentation process; so they could immediately add the DDA resources, products, and services to those of the
other SA-units so that the final report presented to the Ministry of Culture provided an overview of the whole new group of the Danish State Archives.

Based on the SA Action Plan 1994-1998 that was published in three volumes by the end of 1993, a so-called Performance Contract was undersigned between the Ministry of Culture and the SA in 1994. The idea is that the archives get more resources (approx. 10 man-years) in return for specified improvements in performance (efficiency, servicemindedness, productivity). The first Performance Contract is running in the period 1995-1996, only; however, it is anticipated that a new contract be designed for the period 1997 through 2000 by the end of 1996.

During the “laissez-faire period” there were not many advantages or disadvantages of the new host situation. Life went on pretty much as in the past; the DDA was left with the same resources and the same tasks as under Odense University. However, on the positive side, this generated confidence that the SA system was not going to “swallow” the DDA; on the negative side, some resources had to be spent on statistical reporting and planning activities that were not immediately to the benefit of the DDA and our “traditional” user clientele.

3.2. The Integrationist Period
Gradually, as the work with the Action Plan 1994-1998 proceeded, it became necessary to define what was labelled “functional integration” (in fact meaning specialization) within the SA Group. In short, this means that, opposite to the century-old tradition, not all units can upkeep all the specialties of the archival business.

For instance, all the production and distribution of micro-film and micro-fiche will take place at one “virtual unit” (which happens to be located within one physical unit, viz. the Provincial Archives in Viborg). Similarly, the conservation activities are being collected in another “virtual unit”, in this case spread over 2-3 physical units. Furthermore, we work with the notion of “specialist archives/archivists”, meaning that one unit (and one archivist within that unit) is the SA specialist vis-a-vis one type of authorities (e.g. police authorities, county archives, hospital patients’ files).

Turning to the MRDF material, there are two centers in the SA system: The DDA takes care of everything from the “private sector” (incl. research). Also, the DDA is responsible for research remains from many public authorities (e.g. the ISR and an institute for clinical epidemiology) and for a number of semi-public institutions (e.g. the Cancer Register, which is now being moved from the de jure private Danish Cancer Society to the public realm).

It took tough negotiations to define these functional division lines between the DDA and the MRDF unit of the National Archives. A fifth “culture clash” appeared between DDA’s service-oriented activity, international orientation, and informal contact methods on one side and the MRDF-unit’s acquisition-oriented activity, relative isolation, and formal contact methods. Furthermore, the MRDF unit of the National Archives was stuck with very old equipment whereas the DDA has been trying to be at the technical frontline.

So what’s the difference, the sceptic might ask; hasn’t the DDA held the Danish Omnibus Surveys, the Danish Welfare Studies, the Danish Time Budget Data, and other material from the Danish ISR all the time?

Yes! - But there is a difference, and the difference is two-sided: Firstly, the DDA now holds not only survey materials that are de facto anonymous as before; the DDA can now hold materials that are registers according to the Danish Acts on Public and Private Registers. Secondly, with respect to public authorities, the DDA is not dependent on the willingness of the agent to understand the importance of archiving; if the Director of the National Archives and the Director of the Data Surveillance Authority agree that a register shall be archived, the DDA staff can collect that register from the data owner in a capacity as an archive authority.

Even in terms of research registers (especially medical registers) there was some reluctance to give very sensitive patient information to an archive that was a university institute. Being part of the “official archive system” improves the chances that single researchers and research groups are willing to deposit their materials. As a consequence, more data materials will be available from the DDA for future research under the new model.

The advantage for the DDA (or rather for our traditional user clientele) is that more research relevant information will be available for secondary analysis. The disadvantage, seen through the glasses of the DDA staff, is that more tasks are placed on our shoulders without a corresponding inflow of personnel resources. Furthermore, the DDA senior staff is heavily involved in tasks (appraisal of computerized stuff from public authorities that are not immediately of interest for our research users; modernization of the other units in the technical sense, incl. establishment of a new version of “their” archives database on a new platform) that make life busier without augmenting the service level towards our primary users.

3.3. The Immediate Future
Like in many other countries, the politicians and the broader public are very interested in the so-called “information society”. In Denmark, a Government Committee Report (“The Information Society in the Year 2000”) was published in the autumn of 1994. It was immediately followed by the
establishment of a separate Ministry of Research under which the national IT-strategy was located (in accordance with the recommendations in the Bangeman Report from the European Commission which appeared a few months earlier than the Danish Info-2000 Report). So, in March of 1995, the Government produced its annual IT-plan "From Vision towards Action: The Information Society in the Year 2000", which in some respects looks like the Clinton/Gore initiative in the direction of Information Superhighways, in other respects is encompassing a lot more due to the special character of the Danish society.

The Government IT-Plan for 1995 and the SA Performance Contract with the Ministry of Culture require a lot of decisions from the SA. Just to mention a single challenge with a long-range perspective: Before mid-1995, the SA is going to define the rules and procedures that we deem necessary in order to allow the authorities to adopt the practice of "the paper-less office" from the beginning of 1996 (paper-less, because incoming paper-mail is scanned and saved (e.g. in a TIFF-format or equivalent), and where in-coming e-mail as well as outgoing mail of all types are saved in searchable format, e.g. in the SGML-format - with a well-defined DTD - or in other expectedly long-term viable formats).

3.4. The Longterm Perspective

The advantage for the DDA of the placing within the State Archives system is, of course, that we "archived the institution" within a long-term viable institutional structure, forming part of the national information strategy. As many IASSISTers will realize (more or less horrified!), the whole *raison d'etre* of many data libraries may vanish within a very foreseeable future due to the fact that end-users can download their research resources directly from the producers or other facilitators - on a global scale.

In the near future, academic data service organizations will face a strong competition from private and quasi-public vendors trying to monopolize their services not unlike the way that many (European) CSOs have done in the past. The information society involves rapid institutional changes even to the information specialists.

To establish a condition with Freedom of Information (and equal access) is no longer a question of some academic institution-building, only. National, and in turn international (in Europe e.g. within the European Union) information strategies will be developed from the political level, and they will severely influence the survival conditions for most of our academic service institutions.

4. Projects Facilitated by Archives Integration

The functional integration of the academic data service and the traditional archive system has already had an impact on the "palette" of activities of the DDA. Below, we shall touch upon a few projects that are facilitated by this integration.

4.1. The Source Entry Project

Like the other Scandinavian countries, Denmark has excellent demographic sources. In order to ease the access to those sources that may account for so much as 80% of the use of traditional archival material, the traditional archives have had large projects (in part jointly with the Mormon church) producing films and fiche with these sources. The film/fiche versions of the sources have in turn been distributed to City and Local Archives, thus releasing the increasing pressure on the SA reading rooms.

It goes without saying that such demographic sources invite computerized treatment. And, indeed, many amateur historians and genealogists (organizationally cooperating within the association DIS-Danmark) have been entering a lot of these sources into computer programs. These source entry initiatives, however, were scattered in coverage, differing in quality, and more often than not non-transferable because of technical limitations.

In 1992, DIS-Danmark formed a Cooperation Committee for Source Entries (Danish acronym: SAKI), and several staff members from the Danish State Archives (incl. Hans Hans Jørgen Marker from the DDA) were invited to serve on that Committee. During less than one year's work (1992-1993), this Committee completed a set of recommendations (called the SAKI Model) for the creation of machine readable source editions of structured sources (published in a special issue of the DDA quarterly newsletter *DDA-News*). The recommendations should secure higher quality of the products from this huge amateur project.

In order to improve the transferability of data, a special Source Entry Program (KIP) is offered to people who want to serve as source entry personnel. Furthermore, the DDA serves as the central archiving facility and distributing service for all these computerized sources. Finally, a Coordination Committee (KOKI) keeps track on who is doing what to avoid duplication of effort.

At the DDA, the computerized sources are standardized and documented. And the DDA can supply copies of sources (usually in paper-form, but if needed also on film or microfiche) free of charge to people who are willing to and capable of making contributions to the program. By the end of April, 1995, more than 5.1% of the 1845 Census is available, with the 1787 Census in second place (3.4%).

In a not too distant future, such frequently used demographic source material as censuses and church registers may be available in data form as well as in the form of scanned images (based on the film/fiche versions). This will revolutionize the nature of use of such sources and open a lot of new projects: Person recognition and family reconstitution based on neural networks; automatic movements up and down family trees in a graphically based environment; etc.
4.2. The Computerization “Rightsizing”
As mentioned above, the “old SA-family” used technical equipment from the mid-eighties (mini-computer technology from Norsk Data); the system is completely closed from the outside world, because this was considered necessary to secure confidentiality at the time of installation.

In August-September 1995, all units within the SA (except the DDA which will be on that platform already) get new client-server equipment after specifications laid down in a group where the DDA has held the chairmanship. This means that the SA-units will be able to benefit from the resources on Internet and other communication networks, and it implies that a strategy can be adopted where the descriptions of the materials in the archives can be brought to the users electronically.

DDA is heavily engaged in a rescue operation where an existing (hierarchically organized) archival data base is going to be transferred to the client-server environment and entered into a relational data base system (viz. MS NT SQL Server).

Although these technical cooperation projects have drained resources from the DDA, they do hold a perspective for the future: Because the DDA has a longstanding experience with user contacts (the MRDF unit in the National Archives has only served about a dozen users since its inception in the early seventies; the DDA has several hundred user requests a year), we may well be designing the user interfaces for the whole SA “family” in the future.

4.3. The Register Research Facilitation
As an initiative of the Danish Research Foundation, a Working Committee (where the author of this article was a member) has been defining a model that might facilitate the use of personal registers (incl. registers in the CSO) for research purposes. The recommendations of the Committee, to establish a Register Research Center, adjacent to but independent of the CSO, and to establish a Register Archival Facility at the DDA) are being implemented right now.

The DDA could not have played an active r̃le in this project without having an authorization to hold identifiable personal records. The idea is, furthermore, to take on a medically trained staff person to make sure that the many registers in hospitals and medical departments be rescued, to the benefit of contemporary and future research.

To people outside Scandinavia a comment may be relevant: The Danish society is administered almost completely via computer registers; the citizens are registered with the CPR-number (Central Personal Number) as the unique identification code. This implies that all types of personal information may be merged in research projects, and this is of great importance - so far especially within medical research. This register-based research potential is considered to be unique for the Scandinavian countries.

4.4. The Government System Contacts
The placing of the DDA within the Danish State Archives seems to have brought us closer to the Government system than we were under Odense University. This implies that the DDA has been represented on numerous Committees and Working Groups where “the future is designed.”

This being said, we still try to keep the “anarchistic DO culture” as our life-style and the equal access to information as our distribution principle. To do so is facilitated by a comment from the political system in the report leading to the Archives Act of 1992: The leading principle in the administration of the Archives Act is going to be to secure “the greatest possible openness.”

5. Closing Notes: Merging Cultures
In less than a quarter of a century, a lot of “culture clashes” have been experienced by the DDA - internally and in the contacts with the outside world. My guesstimate goes that we are now going to see a “reverse process” - a merging of cultures where there are not so many “computer-nicks” or research-discipline monopolists who claim their superiority. So much information will be readily available that technical and human network-building as well as inter-disciplinary sensibility, cooperation and understanding will be much more important than media-oriented or discipline-based exclusiveness.

In the Danish case, the merging cultures are visible in two respects already demonstrated in the project descriptions above: Firstly, from being a “traditional” social science data archive holding survey data relevant for the political and social sciences, the DDA is rapidly moving into a position where historians and medical researchers are added as new user groups. Secondly, since we had to abandon the Population Register Data subproject (cf. 1.1.3 on p. 3) in the mid-seventies, the activities have not included register data. There is no crucial difference in method analyzing survey or register data; the two should complement each other rather than being seen as two different approaches. More often than not, register research projects will contain a process where subpopulation data held by the researcher have to be merged with register data held by some public authority; therefore, it seems logical to have the services and the data resources collected in one place - in a small country which cannot afford to have several, discipline-specific data service organizations.

On the Danish data arena, only the economic time series (incl. the regional data, cf. subproject 1.1.2 above) are not yet incorporated in the service “palette” of the data service unit; and, to be honest, I think that they should not be! - Time series data should be available from the main producers, viz. the CSOs. Needless to say, they will be entered into the archives for historical research in due time;
but as far as contemporary research is concerned, the time series data should be distributed by the producers - and if they introduce obstacles, we should concentrate our energy on removing these.

The major reason why I find that contemporary (economic) time series and regional data are unappropriate in academic DOs is that they are constantly changing - in the course of time (new weekly/monthly/quarterly/annual figures should be added) and because of changes in administrative regions (which necessitates a backward harmonization).

In conclusion: The technological development will have a crucial effect also on the institutional landscape a decade from now. We already face the rapidly changing conditions of our activity brought about by the Internet and WWW services; so far, we (the DO personnel) can feel easy at the frontier because we know more about these advanced technical information interchange facilities than most of our users. But take care: New generations of users are entering the professional scene; they know "the computer age" because they already grew up in it, and they will ask for services in terms of selective information facilitation that we are not yet able to produce.

There are plenty of challenges for IASSISTers for the next couple of decades. After that, many of the IASSIST pioneers can sit back in their homes, living on their pension schemes, communicating with each other about the rapid-changing world and the oddities of the younger generations.

The topics old people always communicated about ... - but we shall be in the favourable position to communicate electronically and globally!

1 Paper presented at IASSIST 21st Annual Conference May 9-12, 1995, Quebec City, Canada.
Disseminating Data From Longitudinal Surveys: Issues Facing the Survey of Labour and Income Dynamics

by Maryanne Webber
Statistics Canada

I. INTRODUCTION
The Survey of Labour and Income Dynamics is one of several new longitudinal household surveys being mounted by Statistics Canada. Like the others, SLID is preparing for the release of its first round of microdata. The dissemination of microdata from longitudinal surveys poses several challenges. The purpose of this paper is to outline these challenges and some of the measures being proposed to deal with them. The paper begins with a brief overview of the survey content and design as context, but the main purpose of the paper is to provoke discussion on general dissemination issues, using SLID as a case study. The intended audience is research librarians and others who will play a role in the dissemination process.

II. OUTLINE OF THE SURVEY
SLID is designed to track the experiences of individuals in the labour market, their level and sources of income and changes in family life over a period of six years. The first panel began in 1993, with labour and income information collected from about 31,000 persons aged 16 and over. A second panel will begin in 1996, doubling the sample size. In 1999, when the first panel ends, a third one will begin. This approach of rotating, overlapping panels ensures that the sample remains representative.

During the six years, 13 interviews are conducted. A preliminary interview is done when a panel first starts up, to collect background demographic, education and work experience information. One year later, an annual cycle of labour and income interviews begins. Every January, information on the person’s labour market activities throughout the previous year is recorded; in May, income sources and amounts for the previous year are collected.

A summary list of variables from the survey and a chart depicting the main types of information are presented in appendix. Major research areas will range from employment and unemployment dynamics and labour market transitions linked to the life cycle, to job quality, workplace inequality issues, family economic mobility (dealing with shifts in income level), low income dynamics (or flows into and out of poverty), demographic events and the relationship between work and education. Researchers are expected to come from many disciplines.

III. DATABASE SIZE AND COMPLEXITY: THE MAIN CHALLENGE

By household survey standards, the SLID database will be large and complex. Even with our best efforts to make it approachable, researchers will need to make an “up front” investment of time and effort to come to grips with it. Why is this so?

Number of variables and hierarchical structure
Perhaps the most fundamental reason is the size of the dataset and its internal relationships. As a rough estimate, there are 500 distinct variables in the full dataset, without taking the time dimension into account. This means that events, spells, variables collected annually and variables collected as many times as applicable are all counted only once — and there are many such variables in the dataset.

Hierarchical relationships abound in the data. A person can have several employers and information is collected on up to six jobs per year. There may be several work absences from each job. Over time, even if a person does not change employers, he or she can have several occupations, wage rates and work schedules. The survey will also yield information at the household and family level. Because of the hierarchical nature of the survey content, we are processing the data in a relational database environment and are also proposing to use a relational database for the microdata output.

Time dimension
Like all longitudinal surveys, SLID users will need to grapple with the time dimension. From the time perspective, we can distinguish different types of variables. First, variables like gender, year of birth and ethnic origin, are fixed. If an error is detected these variables may be corrected but otherwise they do not change over time. Next, there are annual variables, such as weeks worked during the year and investment income. For these variables, the reference period is by definition the calendar year. Thus, for a full panel, there will be six observations for each record. There are also cumulative variables, like years of schooling, years of work experience and number of children where, depending on the respondent’s activities or circumstances, the values may or may not require updating each year. Finally there are dynamic variables which relate to spells. The duration of a spell may range from a week to several years. SLID’s content includes many variables expressed as spells and, to facilitate analysis, spells that cross the beam between two reference years (for example, an unemployment spell that begins in November and ends the following March) will be
linked up on the database. In effect, the dataset that will ultimately look like the information for a six-year period was collected retrospectively at the end of the six years, as opposed to being a series of unrelated snapshots.

Units of analysis
Another factor that adds to the learning curve — and this again is due to the hierarchical properties of the data — is that there are many possible units of analysis. The person is the basic unit. In addition to being the appropriate unit for many types of research focused on the individual, the person will also generally be used for studies of the family. Because family composition can change over time, the definition of family poses some sticky problems in longitudinal research. One can however define the person as the unit of analysis and develop typologies to characterize the person’s family circumstances over the study period.

The person-job is a unit of analysis used with data from labour market surveys with a one-year reference period, like the Survey of Work History and the Labour Market Activity Survey. We expect that researchers will also use the person-job for SLID studies. This unit of analysis came about as a way of handling the fact that a person may have several jobs, concurrently or consecutively, during a one-year period. Instead of using complex and arbitrary assumptions to select a main job for the year, all jobs are included and weighted using the respondent’s sample weight. Sometimes they are further weighted by annual hours worked, so that part-time jobs lasting one month are given less weight than full-year, full-time jobs.

Some studies will use spells as the unit of analysis. For example, if a person is unemployed for two separate stretches during the study period, the two spells of unemployment will be included, both receiving the respondent’s sample weight. Demographic and other characteristics can be treated as attributes of the spell. Similarly, researchers may use transitions as a unit of analysis. Some transitions can be identified from dynamic variables, when one state ends and another begins. Some data users will no doubt want to develop definitions of transitions tailored to a particular study. For example, it should be possible to use SLID to study work-to-retirement transitions or job promotions. But since these are complex processes, there is no variable or flag on the database identifying these events. Rather, the user will need to look at a range of variables and explicitly define the event of interest.

IV. TOOLS TO HELP RESEARCHERS GET STARTED
The survey staff are very aware of the challenge data users face in getting started. It is incumbent on us to develop tools and user support strategies that increase data accessibility. What are these tools and strategies?

Database design
Because of the size and complexity of the data, a data model was developed. This is a device for structuring the survey content and giving explicit expression to the relationships in the data. The development of the data model was done following two important principles, both of which were intended to aid the data user.

First, variables were defined in keeping with the survey’s content objectives, rather than as a simple reflection of the questions and response categories used in data collection. The survey questions are designed to accommodate data collection, and are often not that useful as analytical variables. For example, to collect one content item, there may be several different questions addressed to various subgroups.

Second, the decision to collect data annually was based on respondent recall and other operational considerations. It was decided that this feature of the data collection operation should be transparent in the output variables (except of course in cases where annual observations make sense from a content point of view). The data for a six-year panel should look like they were collected once covering the full six-year period.

These principles required a significant “front” design and development effort but hopefully they will pay off in downstream benefits to data users who would otherwise have to recreate “seamless” data from a series of snapshots.

Software to retrieve data from database
We are planning to provide a public-use microdata file with front-end software that, at a minimum, allows users to select variables and subpopulations of interest, for specified timeframes. These smaller datasets can then be downloaded into a flat file for further analysis using whatever software the user chooses. There will also be easy ways of producing simple frequency counts from the full dataset, to help users define their study populations.

CD-ROM
The public-use microdata file will be available on a CD-ROM. This will hopefully increase data accessibility.

Major reference products
There are three types of documentation in the works: technical documentation of the database content and structure; a user handbook; and research papers providing detailed documentation on specific topics.

The main SLID database is being designed with the technical user documentation — variable names, descriptions, definitions, algorithms for derived variables, code lists and user notes — as an integral part. This documentation is being stored in a relational format, so it is possible to extract parts and produce customized reports. Microdata users will be able
to access the documentation electronically as it will be imbedded in the product.

A handbook or "friendly" user guide is also being developed. This should be of interest to users of custom tabulations as well as to actual and potential microdata users. After the first few editions, this publication will probably stabilize and enjoy a relatively long shelf-life — perhaps we will re-issue it every six years to coincide with the completion of a panel.

Finally, SLID has a general purpose research paper series. Since 1992, we have produced about 15-20 of these reports each year. We are beginning to use this series as a repository for detailed information on specific variables, for example, the composition of "roll-up" categories for mother tongue and ethnic origin.

Workshops
To get started, some users may be interested in participating in a workshop. We are quite sure that there will be interest in a workshop on the content and structure of the database. We have already been asked by a few groups to do workshops of this type and have agreed.

There may also be interest in analytical techniques appropriate for use with these data.

Sharing information on research in progress
Throughout the survey development process, decisions and issues have been documented in the quarterly newsletter, Dynamics. While there will still be developments to communicate in coming years, we expect that the role and content of Dynamics will gradually shift, hopefully becoming a forum for exchange on research underway outside as well as inside Statistics Canada. It is very beneficial for the survey staff and the Agency to be aware of data uses (as well as research not being done because of the lack of a few key variables). Short research summaries in Dynamics would keep us up to date and could supplement whatever other exchange mechanisms exist among researchers in a particular field.

V. CONFIDENTIALITY
Longitudinal surveys in general face a challenge because the events and transitions that they document — and that are central to their analytical potential — may make it difficult to disclose the identity of respondents. Moreover, when the first wave is released, it is impossible to know what patterns of change over time will be common or rare. Several years down the road, this means that we may need to reconsider the content of the public-use file as the data from successive waves build up.

In SLID's case, there are difficult trade-offs between geography, family information and labour market detail. The data are supposed to meet the needs of researchers in a range of disciplines and to allow analysis of the interactions that exist between labour market behaviour, family circumstances and income. This makes it very difficult to protect confidentiality without "short-changing" any particular user group.

The search for solutions is very lively. Research is under way on techniques for quantitatively assessing disclosure risk and on alternatives to suppression and collapsing. Other statistical agencies are being consulted on their approaches. An attempt is being made to prototype a remote access system, which would allow researchers to write and test their programs off-site and telecommunicate them to us so we could execute them against the full database. We are also investigating the possibility of licensing researchers to use a middle-level file for a specified purpose, following stringent rules regarding access, security and disposal. There is enough concern and energy being devoted to this issue to hope that solutions will emerge.

In the meantime, we are defining the content of a public-use microdata file that would be screened using the usual Statistics Canada procedures. Several analytically interesting derived variables are being added to the file to reduce the impact of missing detail. Here are a few examples:

* several occupation typologies;

* the relevant low-income cutoff, or a measure showing family income as a ratio of the relevant LICO;

* a derived variable showing the link between occupation and major field of study.

Hopefully, variables such as these will help researchers to proceed with their work even if some of the very detailed information (like 4-digit occupation) is not on the public-use file.

We also face a dilemma with respect to family information. On the main base, it is possible to link up family members (and previous family members) but, to provide this capacity on the public-use file, it would be necessary to reduce the amount of labour market information. As a compromise, we are proposing to include a good range of family variables, but only for a subsample of respondents. This means that researchers have access to more variables on the public-use file and, should they require results for the full population, the same program can be re-run against the full data base. These measures will ensure that, even if some variables are missing, the public-use file will still be a rich source of information.

VI. COMPUTER-ASSISTED INTERVIEWING AND USER DOCUMENTATION
Although it does not exclusively concern longitudinal surveys, the move to computer-assisted interviewing for
household surveys at Statistics Canada is raising some interesting documentation issues. We are finding that efforts to document the questionnaire are proving to be very labour-intensive and error-prone. We have been searching for tools and techniques to improve the process and trying to promote some measure of consistency across surveys.

A working group was set up recently in the household surveys area to address this issue. It looked at a number of options. One idea was to produce a print image of each screen. However, this would yield very bulky documents and, for surveys with complex branching (like SLID), it would be nightmarish to follow flows. Also, even with that level of detail, many special features such as hot keys and edits would not automatically be documented. Similarly, the idea of producing a diskette with the questionnaire is appealing at first blush but this would not be very meaningful as a "stand-alone" product. The user would need to learn the data collection software. Moreover, many survey applications — particularly longitudinal ones — do not start with a blank sheet. There are prefilled items that affect questionnaire flow. Without these prefills, one cannot get into various branches of the application.

After examining these and other options, the working group found that, at least for the time being, the best approach is to concentrate on producing a good survey codebook. Among other advantages, this is an approach where standards or guidelines across surveys are a reasonable goal and where the documentation reflects the data user’s perspective. This means that the user documentation of a questionnaire would begin with the output variables and work backwards, ending with the questions underlying the variables. Instead of expecting users to follow complex flows through hundreds of questions, each question or group of questions would have a "universe statement" describing the question’s target population.

The group also concluded that different surveys would require different supplementary tools, depending on audience, length, complexity and periodicity. In SLID's case, flow diagrams showing the organization of the survey content at increasingly detailed levels are being developed.

V. CONCLUSION

Once established, longitudinal surveys can be invaluable — but it can take time to become established. In the current fiscal and social policy climate, time is at a premium. New longitudinal surveys cannot afford many years to demonstrate their value. There is therefore a pressing need to support researchers in getting started. In this paper, some of the dissemination measures planned for SLID have been reviewed. Feedback on current plans will help us to get off to a good start. At the same time, this is a learning experience for survey staff as well as researchers. We fully expect to make adjustments to products and services and therefore hope to sustain a dialogue on enhancements.
APPENDIX: OVERVIEW OF SLID CONTENT

Partial List of Variables

I. Labour
*Nature and pattern of labour market activities*
- spells of employment and unemployment (start and end dates, durations)
- weekly labour force status
- total weeks of employment, unemployment and inactivity by year
- multiple jobholding spells
- work absence spells

*Work experience*
- years of full-time and part-time employment
- years of experience in full-time, full-year equivalent

*Characteristics of jobless spells*
- job search during spell
- dates of search spells
- desire for employment
- reason for not looking

*Job characteristics* (all characteristics updated each year and dates of changes recorded; collected for up to six jobs per year)
- wage
- work schedule (hours and type)
- benefits
- union membership
- occupation
- supervisory and managerial responsibilities
- class of worker
- tenure
- first date ever worker for this employer
- how job was obtained
- reason for job separation

*Characteristics of work absences lasting one or more weeks* (collected on first and last absence each year, for each employer)
- absence dates
- reason
- paid or unpaid

*Employer attributes*
- industry
- firm size

II. Income and wealth
*Personal income*
- annual information on about 25 income sources
- total income
- taxes paid
- after tax income

*Receipt of compensation* (whether benefits were received from each source and, if so, in which months)
- Unemployment Insurance
- Social Assistance
- Worker's Compensation

*Assets and debts*
Information might be collected once or twice in life of panel on roughly 20 asset and debt categories.

III. Education
*Educational activity*
- enrolled in a credit program, months attended
- type of institution
- full-time or part-time student
- certificates received

*Educational attainment* (updated annually)
- years of schooling
- degrees and diplomas
- major field of study

IV. Personal characteristics
*Demographics*
- year or birth
- sex
- current marital state and date it began
- year/age at first marriage
- number of children at home
- parents' schooling

Ethno-cultural
- ethnic origin
- member of an Employment Equity designated group
- mother tongue
- citizenship
- country of birth

Activity limitation
- annual information on activity limitations and their impact on working
- satisfaction with work

Information on person's children
- number of children born, raised
- year and person's age when first child born

Geography and geographic mobility
- economic region or CMA of current residence
- size of community
- moved during year
- move dates
- reason for move
- nature of move (full household/household split)

Household and economic family information (annual summary information at household level, e.g., size, type)
- key characteristics of other individuals in household (e.g., age, sex, relationship, income, annual hours worked)
- household/family size and type
- family income
- relevant low-income cutoff

- family events (separation, death, birth)

Main Features of SLID

Objectives
13 interviews over 6 years:
- preliminary
- 6 labour (Jan)
- 6 income (May)

First panel started Jan. 1993
31K persons 16 and over
Second panel starts Jan. 1996
Results of preliminary interview
released (publication)
Now processing first wave

1. Paper presented at IASSIST 21st Annual Conference May 9-12, 1995, Quebec City, Canada.

2. The first wave (including results from the preliminary interview) will, however, be released as a rectangular file. The content has not yet been finalized but our best estimate is that the record length will be about 3000 bytes for a total file length of roughly 90 Kb. Every year, the dataset grows, i.e., the second year's file will incorporate and replace the first.

3. The basic time unit used in dynamic variables differs depending on the state being measured. For example, spells of employment and unemployment are measured in weeks, as are work absences. Marital states, job tenure and receipt of UI are among the variables measured in months.


5. Abstracts appear in our quarterly newsletter, Dynamics. Also, an annual supplement to Dynamics presents abstracts for all research papers produced during the year. For major developments and issues, there is generally also a longer write-up in Dynamics.
Preliminary Announcement and Call for Papers

IAASSIST '96 brings together researchers, data producers, data archivists, data librarians and support staff to explore the changing roles and relationships among those who work with social science data.

This year's theme uses a metaphor to represent the interrelatedness of technology, social research, data, and services supporting these activities. "Weaving the Web" entails shaping the new technologies for the creation, storage, access, and analysis of social data. In this context, IASSIST '96 is an opportunity for the leaders in data provision and data support to discuss new opportunities, new solutions, and new problems in working with data.

Another aspect of this theme is the impact changes in the economic and political climate around the world have had on the overall fabric of social science research, data, and support. IASSIST '96 provides a forum for examining these influences and their possible outcomes.

A final focus for IASSIST '96 is the application of the new "data fabric" in the instructional setting. How are new tools and techniques for locating, sharing and analyzing data being put to effective use in the undergraduate and graduate classroom? Papers presenting innovations in this area are most welcome.

IAASSIST is an international organization of professionals who are engaged in the creation, acquisition, processing, maintenance, distribution, preservation and use of machine readable text and/or numeric social science data. IASSIST'96 will bring together researchers, data producers, archivists and data archivists, librarians and data librarians and other interested persons to explore our changing roles and relationships.

To facilitate this discussion the conference will provide a oneday overlap with the Computing in the Social Sciences 1996 conference. There will be three days of panels, paper sessions, poster/project sessions and special speakers followed by a day of workshops.

The IASSIST'96 theme draws on the history and experience of the textile industry in applying computer technology to the art of weaving. The Jacquard weaving process, one of the earliest uses of "machine-readable" punch cards, drew together the expertise of the artists, technicians, and engineers to create an aesthetically pleasing, functional and enduring product.

IAASSIST'96 seeks to attract a similar blend of expertise and perspectives to ensure the continuation of a viable data infrastructure which will support social science research and instruction into the future.

TOPICS OF INTEREST:

The program committee is inviting submissions for paper presentations, panel sessions, or poster sessions on any aspect of the conference theme; computing tools for social science research or instruction; technologies for creating, storing, accessing and analyzing data, and developing data needs and concerns. Possible presentation topics include:

- Data delivery mechanisms
- Documentation standards
- Instructional use of data and analysis tools
- New and improved data products
- Data client support services
- GIS/Mapping and data
- Data access and preservation
- Integration of scientific and social data
- Infrastructure and support for research
- Cross national comparisons and standards
- Community and collaboration
- Global and regional modeling
- Visualization and other graphic methodologies
- Impact of political and social change on data availability
Data management techniques and archives administration
- Intellectual and physical access to data
- Confidentiality and privacy
- Self publication and data archives
- Applications software

CONFERENCE DETAILS:
The conference will meet following the Conference on Computing for the Social Sciences (May 13-15) and will hold joint sessions on Wednesday, May 15 with CSS’96. A reduced rate will be available for individuals wishing to attend both conferences.

For further information on CSS’96 contact Ronald E. Anderson, Conference Chair, 909 Social Sciences, University of Minnesota, Minneapolis, MN U.S.A. 55455 (612-624-9554) <rea@soc.umn.edu>. CSS’96 Homepage: http://ag.arizona.edu/ssca/96anmeet.html

CONFERENCE LOCATION:
Radisson Hotel Metrodome
University of Minnesota
615 Washington Avenue S.E.
Minneapolis, Minnesota U.S.A. 55414

The Radisson Hotel Metrodome is conveniently located just three minutes from downtown Minneapolis and the Metrodome, and 15 minutes from downtown St. Paul via easy freeway access.

The Twin Cities (Minneapolis-St. Paul) area is home of the Charles Babbage Institute, Guthrie Theater, Walker Art Museum, the Mall of America, the American Swedish Institute, and the Museum of Questionable Medical Devices, as well as the Minnesota Historical Society. Come prepared for warm spring weather and walks along the Mississippi River. Lake Wobegon is nearby.

MINNESOTA: Celebrating 50 Years in Computing

DEADLINES:
December 15, 1995 - Proposals for papers, etc.
January 19, 1996 - Notification of proposal acceptance
May 15, 1996 - Papers due for publication

Notification of acceptance will be sent via e-mail. For those not providing e-mail addresses acceptance letters will be sent in lieu of e-mail.

CONFERENCE INTENTIONS FORM:

Name
Title:
Affiliation:
Mailing Address:
Electronic Mail Address:
Telephone: __________ Fax: __________

Check all that apply:
- I intend to submit a paper on the following topic/title: ____________________________
- I would like to hold a panel/seminar/roundtable discussion on the topic of: ____________________________
- I am interested in presenting the following poster session/display: ____________________________

(please include a 250-500 word abstract plus a 2 sentence biographical summary with any of the above)
- I will be willing to chair a session
- Please keep me on the mailing list
- Please add this person to the mailing list: ____________________________

Submit Intention Form before December 15, 1995 by e-mail, fax or mail to:
Wendy Treadwell, Program Co-Chair
Machine Readable Data Center
University of Minnesota
2 Wilson Library
Minneapolis, MN U.S.A. 55455
612-624-4389 FAX: 612-626-9353
E-MAIL: papers96@mrdc.lib.umn.edu

Summer 1995
The International Association for Social Science Information Services and Technology (IASSIST) is an international association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information system specialists, database librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hard copy as well as machine readable data.

Paid-up members enjoy voting rights and receive the IASSIST QUARTERLY. They also benefit from reduced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:
Regular Membership: $40.00 per calendar year.
Student Membership: $20.00 per calendar year.

Institutional subscriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subscription: $70.00 per calendar year (includes one volume of the Quarterly)

---

I would like to become a member of IASSIST. Please see my choice below:

- $40 Regular Membership
- $20 Student Membership
- $70 Institutional Membership

My primary Interests are:
- Archive Services/Administration
- Data Processing
- Data Management
- Research Applications
- Other (specify)

Please make checks payable to IASSIST and mail to:
Mr. Marty Pawlock
Treasurer, IASSIST
% 303 GSLIS Building,
Social Science Data
Archives, University of
California, 405 Hilgard
Avenue, Los Angeles, CA
90024-1484

---

Name / title

Institutional Affiliation

Mailing Address

City

Country / zip / postal code / phone