Sharing Information Begets Information

by Eric Tanenbaum[1]
ESRC Data Archive
University of Essex
Colchester
Essex C04 3SQ
England

Abstract

In common with many other national data libraries, the British Data Archive has a developed role as a broker of government and commercial data to the academic community. For the Data Archive, at least, it would not be excessive to claim that its *raison d'etre* lies in its fulfillment of this role.

Overcoming the technical and legal constraints on data accessibility revealed a knottier problem. Data archives' success in making vast volume of diverse data physically available to a heterogeneous research community heightened the need for knowledge about the data. Although recent developments in computer networking and software ease *data sharing*, they may inhibit *knowledge sharing.*

This paper describes several innovative programmes undertaken over the years to facilitate the exchange of data-orientated knowledge between (and among) depositors and analysts. Three are described: data-focused workshops, user groups and computer teleconferencing. Each approach has

[1]Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15–18, 1989

evolved through time, often based on the reaction of programme participants. The paper highlights noteworthy stages of the several iterative processes which led to the nourishment of new information-sharing networks.

Introduction

Throughout their history, social science data archives concentrated on a direct, unambiguous goal. They retained computerised information (data) files for distribution to secondary analysts. While archivists will not shrink from describing the difficulties still entailed in this aim, their success minimised general awareness of their efforts. Moreover, as I have suggested elsewhere[2], this task was necessary at a particular technological moment. Developments in information techniques[3] diminish the centrality of the traditional archive task. Put simply, archives have to reassess their role in a changing world.

Archives, like any other producer, can explore the characteristics of their market to be sure that they target their "product line" to market needs. The paper adopts a standard market analysis strategy to show that the British national archive's client community contains several segments. It finds that within the British user community there is a sizeable group of researchers who analyse large data suites. The paper then specifies the particular needs of this large cluster. They, more than others, require shared knowledge about the material they use. While informal networks are the best medium for exchange, existing professional networks cannt fill this function effectively. Here then, the paper suggests, is a market niche that archives are well-suited to fill.

Moreover, and at the risk of forcing the commercial analogy, this section of the paper also shows that in creating this marketable product, the archive is adopting the concerns of the currently popular ethical investment funds. Information-sharing networks are valuable not just because they bolster an archive's "profitability". They also offer an otherwise unavailable public good.

The second part of the paper describes three related projects undertaken by the Data Archive to foster data-centred networks. Each is described at some length to guide other archives who might be contemplating similar network-building activities.

[2]Tanenbaum, E. (1986) "Archive and Dinosaurs" IASSIST Quarterly

[3]e.g. local mass storage devices, inter-computer networking, database systems with common user interfaces

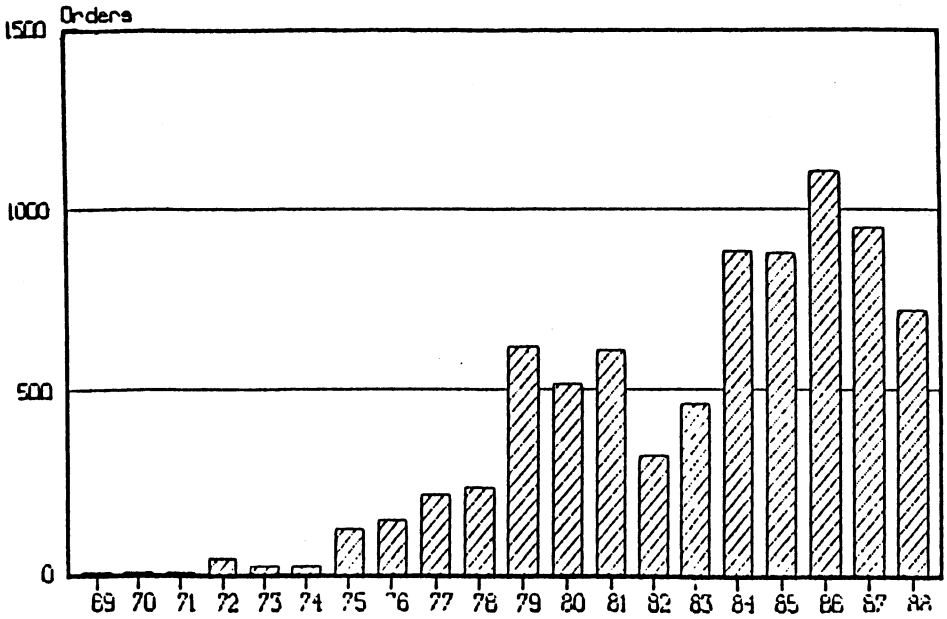
Section 1: Identifying the Market(s)

Success Has Its Perils

There is evidence that archives successfully meet their goal of recycling data to the research community. On the supply side, even in a period of economic restrictions, data archives continue to be funded with central funding. Moreover a growing number of multi-volume archive data catalogues suggests that data originators find depositing data to be both worthwhile and safe. On the demand side, less impressionistic (if more parochial) evidence can be drawn from one archive's experience to suggest that archives have also found acceptance among social scientists.

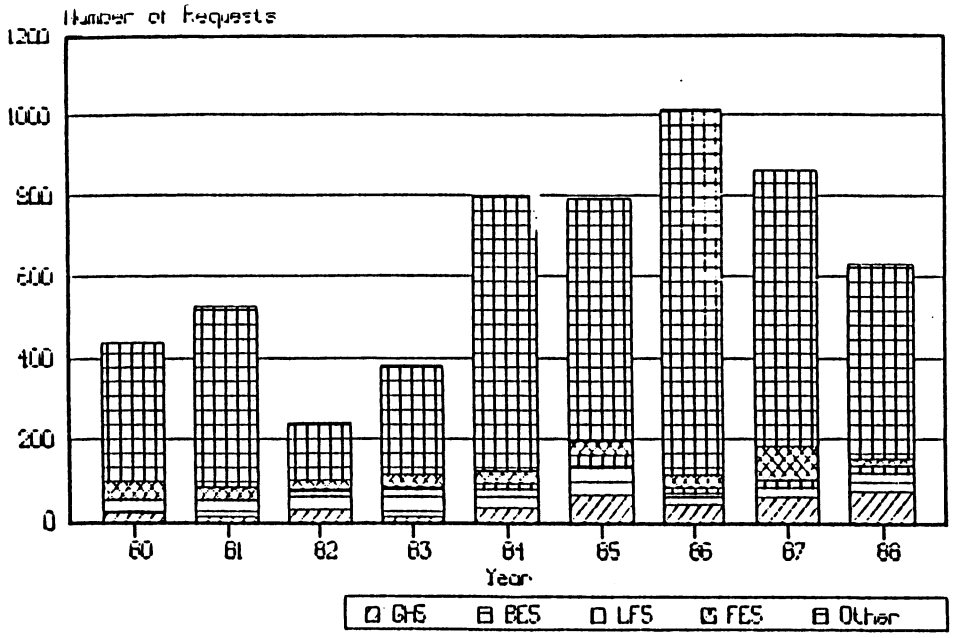
Figure 1 charts the development of British (Economic and Social Research Council) Data Archive academic user community since it "went public" in 1969. Of the several factors which contribute to the growth pattern[4], the increased range of rich data sets that have become available through the years stands out. Figure 2 demonstrates that a good portion of the growth comes from only a few datasets. To put the figure in perspective it is worth noting that the four series combined represent less than 3% of the Data Archive's data holdings. Thus while Archive users still address a varied body of data, some files are disproportionately popular. This has several implications for the Archive, the user community and, not least, this paper.

figure 1
Data orders



[4]e.g. growth in (a) social science comunity, (b) acceptance of empirical research, (c) access to computers, and (d) quantitative skills.

figure 2
Data orders for four series



At first glance the most remarkable feature of these data sets is the breadth of their substantive coverage. Each title referenced in Figure 1 is the generic name of a multi-member set. Table 1 describes the broad features of each set by time coverage (occasions), stimuli (variables, tests) and sample (population, people), the three facets of the data box described by Cattell[5].

Table 1 flags the potential technical and strategic proglem areas that anyone analysing these data will face. These are:

> *COMPUTING*: The sheer amount of data faced by the user of one of these data sets will almost certainly exceed anything in the analyst's experience. Success with these data will require familiarity with aspects of computer use which could happily be ignored when analysing smaller data sets (of the kind most researchers learn to use during their training)[6].

[5]R.B. Cattell (1978) The Scientific Use of Factor Analysis in Behavioral and Life Sciences (London: Plenum Press).

[6]For example, my first analysis encounter with the Family Expenditure Survey used a ten year slice

STATISTICAL: These studies rarely use simple random sampling. The elementary approaches to sampling error to which many social scientists cling tenaciously for decision rules may mislead. Equally troublesome, analysts quickly find that most coefficients are "statistically significant" with samples as large as those available from these files.

DESIGN: Each set supports cross-time trend analyses. Moreover, two have panel study characteristics and several have irregular time intervals. Exploiting the time dimension will often require learning new analytical strategies of the kind not normally covered in a social science curriculum.

DEFINITION: Although not unrelated to the design consideration just described, choosing the most appropriate operational definition is a sufficiently large stumbling block to justify separate mention. Three potential problems stand out.

First, the individual studies present many choices. Recurrent government studies[7] which serve several purposes (and sponsors) have multiple definitions of similar phenomena. For example, the Family Expenditure Survey offers its analysts at least a half dozen ways of measuring income. The would-be secondary analyst knows there must be a reason (or reasons) for the different approaches but may not be clear about their relative merits.

Second, definitions may change across years. The analyst has to decide whether the different question wordings are functionally equivalent[8] or are truly different measures.

Third, and although trivial it can be disastrous, the physical location of the same question/variable may change across years. In the uncommon situation faced by researchers using these data sets, a certain numbness sets in when writing the eighth format statement to extract the "same" data[9]. This is less likely a problem for the researcher analysing data from a single cross-section study.

STRUCTURAL: Although perhaps not immediately clear from Table 1, most of the studies referenced do not use the familiar simple data structures that most cross-sectional studies employ. Many researchers lack the software tools to unbundle complex data relations like those of the General Household Study. There, as the name implies, the household is the basic observational unit. However data appear separately for all individuals in the household. Moreover there can be variable amounts of information appearing for the separate individuals in a single house. Granted, the British Data Archive produces simplified data structures from

---

[6](cont'd) which meant a dataset with 250,000 observations. Although I had grown to love asterisks in psychology research reports, I was less enthusiastic when my standard computer package insisted in producing rows of asterisks in all the output I generated. The mass of data threw it into overflow fits as it tried to calculate simple variances across a quarter million observations.

[7]i.e. General Household Survey, Labour Force Survey, National Food Survey and Family Expenditure Survey.

[8]For example, the researcher interested in authoritarianism may feel foolish asking respondents to evaluate "zoot-suiters" but might still want to tap ethnocentrism with a more contemporary, but subjectively-equalivanet, stimulus.

[9]It may also be that many analysts of multi-year data sets are risk-takers. It is surprising how many request several years' data from a survey set based on reading the documentation for a single year.

these files. However an understanding of the particular logical view of the inter-entity relations that the Archive takes in making these simplications is desirable, or perhaps essential, to understanding the data themselves. The user of data from simpler sources can more easily ignore this difficulty because no choice is offered.

TABLE 1

Basic Characteristics of Four Series

STUDY SET	TIME PERIOD	STIMULI	SAMPLE SIZE
General Household Survey	1971– (annual)	housing/ employment/ education/ health/ leisure/ household structure/ income	11,000+ households (30,000+ individuals)
Family Expenditure Study	1961– (annual)	income/ expenditure/ household structure	11,000+ households (30,000+ individuals)
Labour Force Survey	1973– (biennial/ annual)	household composition/ economic activity	100,000 individuals annual) (panel component)
British Election Studies	1963– (election years)	demographic/ economic activity/ social attitudes/	2000+ individuals (panel component) voting

While the difficulties clustered in these five areas may plague any secondary analyst, they are particularly pernicious for the analyst of complex suites of data, such as those described in Table 1. The analyst who uses a single data file collected for a particular purpose may avoid the pitfalls charted in the preceding paragraphs unwittingly. These data files are smaller, temporarlly-specific and substantively-focusd at origin. With these data sets the analyst's tried and proven skills will be appropriate. The user of material from these larger series will be much more prone to error at every stage of the research project.

Whither the Archives?

This suggests several alternative strategies for data archives in their dual roles as purveyors of data and promoters of good social research. If they adopted a truly defensive posture, archives could dissuade researchers from undertaking the analyses of material derived from these complicated sources. However a little reflection will show that this is neither possible nor safe because of the requirements of data-based empirical research. An empirical social science *which says anything worthwile about social phenomena, requires rich data resources. Social reality is not simple.* There is no reason to suspect that it can be modeled with simple data.

If this defensive posture is not viable, neither is a *laissez-faire* attitude which allows the analysts to swim through the shoals if they can. The chances are too great that more would sink than survive. Moreover the analyst would not sink alone. Archives and social science would be trapped in a quagmire of bad social research. There are few countries in which the social sciences are strong enough to risk producing bad research. More parochially, there are probably even fewer publicly-funded archives which could survive in the face of being implicated in a long sequence of poor research projects.

Archives have to take a pro-active role to improve the secondary analyses of complex datasets. Besides providing well-documented data, they must begin to offer an effective "after sales" service. In the academic world this means working to foster, and then maintain, professional networks.

There is nothing novel in suggesting that networks are central features in the social researchers gambit. One has only to look at the activities of the many professional associations in the social sciences for evidence of their role. However while discipline-based networks are a model for this paper's proposal, they do not provide the vehicle. Figure 3 suggests why.

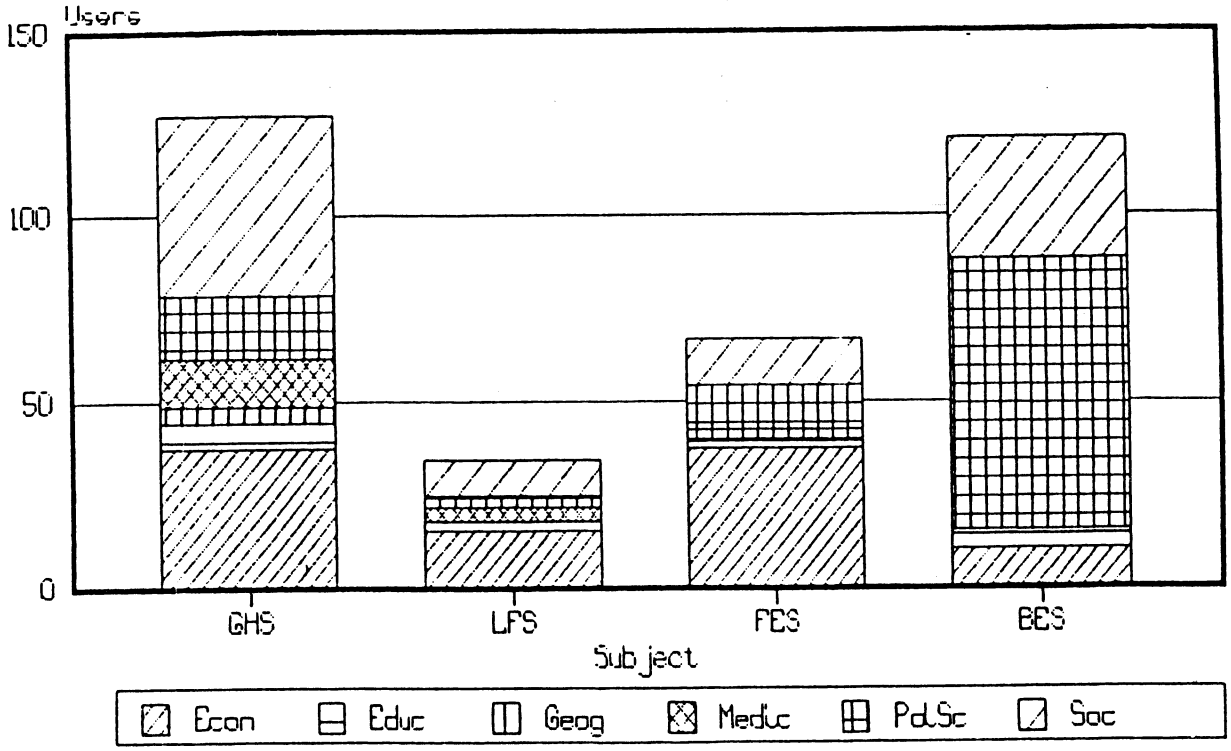
Again, this chart reports the Data Archive's experience. The users of the four data suites are described by discipline[10]. As might be expected from these datasets' diverse topic coverage, the graph shows that they attract a heterogeneous group of analysts. As they are "united" by a common resource rather than a common professional orientation, existing disciplinary networks will not provide an effective milieu for sharing experiences[11]. At the very least, archives themselves have to provide the catalyst for network creation. More likely, they will also have to provide continuing executive support.

---

[10]Because disciplines are identified by the user's departmental address, this is at best a rough approximation of what researcher affiliations are. For example, political scientists are defined as those users whose departmental address contained one of the following character strings: POL, INTER (for international relations), PEACE (peace studies), GOV (government) or PUB (public administration).

[11]Even application fields, which often transcend disciplinary boundaries, offer an inadequate spread of contacts. Appendix A lists a set of research projects undertaken in one year with one of these data series. It can be seen that the applications are almost as diverse as the researchers' professional orientations.

figure 3
Data users fo four series within discipline 1963 - 1988



Section 2: Fostering Data-Centred Networks

Three Ventures

The Data Archive has approached network creation in three ways: (a) data based workshops, (b) user groups and (c) computer teleconferencing. Each has been successful enough to justify its continuation. Nonetheless, they are resource intensive, albeit in different ways. Thus it is worth giving an account of their individual development, and the problems met, to guide other archives who might be considering similar activities.

Data Workshops

The data-based workshop series was the Archive's first formal attempt to encourage the interchange of knowledge. It has two targets. It tries to promote exchanges both among secondary analysts of particular data series and between the data series originators and potential secondary analysts.

The first workshop, which featured the Family Expenditure Survey, was held in 1976. Although the format of the workshops has changed through time[12], the basic principles remain unaltered.

Most important of these is the belief that *resource-centred* meetings are unique and necessary. As suggested earlier, social researchers have many opportunities to engage in substantive interchanges with their colleagues. Without these data-centred workshops they would have little chance to meet others using the same resource and so would have little opportunity to share problems and solutions. As many of these are technical, they are unrelated to substantive concerns.

The necessity of a forum for sharing fostered the second stable organisational principle. Informality rules. Unlike most academic workshops, Data Archive data workshops encourage audience participation. Through the years we used several ploys to help stimulate this ambience.

We use only a skeletal agenda — an example of one appears in Appendix B. Generally there are no more than four speakers. Each *introduces* a well-defined topic. Typically these topics are "the originator's perspective: data substance," "the originator's perspective: field work procedures," and, where possible, descriptions from two secondary analysts about their experience with the data. The invitation to speakers is careful to warn against a conventional paper. Experience, rather than findings, is what is required.

Although each segment is given at least an hour, speakers are urged to limit their prepared comments to 20–25 minutes. They are also asked to steel themselves against long, but we hope pregnant, silences as members of the audience wait for someone else to speak.

We adopted this approach because it is virtually impossible to predict what aspects of a dataset are not propoerly covered by its accompanying documentation. The important bits of knowledge are those identified by the users as being crucial. Creating a forum in which users can express and resolve concerns helps remove the "minor" stumbling blocks that inhibit analyses.

User interchanges are also encouraged. Because the Archive's main funding agency recognised that these workshops offered a valuable, but otherwise unavailable, service to social researchers they could be run free of charge[13]. A completed pre-registration questionnaire which asks about the participant's interest in and experience of the featured data set is the only admission price.

These questionnaires, which serve as confirmations of intention to attend[14], are collated before the

---

[12]Although it would be nice to think that the Archive is iterating toward the perfect solution for knowledge interchange one frequently suspects that we are simply fixed in a local minima.

[13]Indeed in the beginning it was also possible to offer travel support to all attenders; this is now only available to people on Research Council studentships.

[14]the venues are only announced to those who return them

meeting. Each participant gets a copy so that people with similar interests can locate one another during scheduled breaks. The collations also help the invited speakers direct their remarks to the audience's needs.

In general, the content of audience interventions meets our expectations. Questions about variable definition and sampling strategy stand out among the many topics that have arisen over the years. Furthermore, our follow-up questionnaires indicate that just meeting the originators of a data resource is reassuring. For many, this exceeds the value of responses to particular points. If nothing else, it helps defeat the gap between data collection and data analysis that is one of secondary analyses weaknesses.

When the series began, we expected that we would hold three a year. Each would correspond to a British academic term. We also intended to limit workshop coverage to the larger data sets (e.g. General Household Survey, Naitonal Child Development Survey, Family Expenditure Survey) so that they could be recycled biennially. The list of workshoips held in the last few years, which appears in Talbe 2, shows that this aim failed.

Table 2 also has information about the number of people attending the workshops. Attendance varies widely, although a ceiling is set at fifty. Occasionally we schedule a rapid repeat meeting to serve an overflow.

Another change has been that workshops now run for only half a day. Because most of the meetigs are held in London we must allow as many people as possible to take advantage of British Rail's single day return fare scheme. Thus meetings now begin at 12:30 and continue until 5:00.

The participants themselves requested this change in the Participant's Comments that we collect after each meeting. Here we ask people to reflect on the level and content of presentations, the day's organisation and to give ideas about workshops they would like us to hold. Certainaly any archive embarking on a similar series should be sure there is a way for participants to express their opinions. The Data Archive finds that these completed forms add to a waning collective imagination, something that was ineviable after ten years.

Not surprisingly the workshops are resource-intensive. However each only requires several concentrated spurts of activity. Except for recruiting speakers to introduce particular data sets, much of the work can be made routine.

Even getting speakers is not very difficult. The originators of the featured data often welcome the opportunity to meet people who will be analysing "their" data. While recognition of their contribution must be gratifying, we suspect that data originators feel relieved that contact with many naive users can be concentrated in one occasion.

The two biggest weaknesses of the workshops are their location and timing. Almost all occur in London. The London venue is almost mandatory because London is an "average" location. Like many averages, it misrepresents the extremes. In the British case, outliers are anyone living north of Newcastle or west of Liverpool. While it is possible to hold meetings away from London, it can be difficult getting London-located data originators to take the time necessary to travel to remote locations.

Timing is also a problem. The workshops offer the kind of information that is not readily-available elsewhere. However its usefulness depends partially on the analyst being at a stage in a project in which the material is pertinent. This is not likely to happen very often.

To overcome both these problems the Archive established two other initiatives: a computerised teleconferencing system and the data-oriented use group. Each will be described in turn.

TABLE 2

The Active Workshop Series

YEAR	TOPIC	ATTENDANCE
1982-83	Labour Force Survey	50
	1851 Census of Great Britain	15
	1981 Census for Sociolgists	15
	Cohort Study of the Unemployed	40
	JUVOS Unemployment Records	38
1983-84	CSO Macro-Economic Data Bank	24
	Women & Employment Survey	42
	" "	23
	Workplace Industrial Relations Survey	15
	JUVOS Unemployment Records	53
	Census "Public Use Files"	30
1984-85	British Crime Survey	36
	British Social Attitudes Survey	37
	British Election Studies	24
	JUVOS Unemployment Records	45
1985-86	On-Line Access to the Archive	34
	National Child Development Study	20
	Agricultural Census	23
	General Household Survey	18
1986-87	British Population Census	
	JUVOS Unemployment Records	
	Administrative Statistics	
	CSO Macro-Economic Data Bank	
	Family Expenditure Survey	

Computer Teleconference: The Archive Bulletin Board

Regardless of the type of data supplied, archives try to offer analysts ancillary support. The Data Archive, for instance, publishes a descriptive catalogue, a regular newsletter and, where necessary, specialised documentation about particular data sources. Archives also offer on–line telephone support, although as archival holdings grow, archives risk embarrassment by substantive ignorance. And, of course, for large sets of data, it sponsors the worshops described above.

Whatever their merits, these services inevitably suffer because of shortcomings in the media used. The newsletter, for example, is composed some weeks before distribution. Mail or telephone consultation minimises the delay between user need and archive response but here information which is potentially of general interest is broadcast to a narrowly defined audience — typically consisting only of the person who raised the question.

Bulletin Boards in General

Computer teleconferencing facilities can overcome these constraints. Although these facilities are familiar to teenagers, they are less so to older social scientists. A brief overview of their potential is warranted.

"Computer conferencing", a new venture in the social sciences, is attractive because it offers topic specialists an efficient, inexpensive communications facility. The December 1985 issue of BYTE (vol. X, 13) featured computer conferencing. Although dated, it remains a worthwhile reference for anyone interested in looking further into its use. The articles by Jacob Palme ("Conferencing Standards" and "Database Structure in PortaCom") are of particular interest in the context of the Archive's Bulletin Board as the Archive and the University of Essex have adopted Palme's COM for its conference system.

Teleconferences are a computerised version of the conference call[15]. The main difference is that interpersonal communication is via the computer keyboard, rather than by voice. Consequently, teleconferences also benefit from computerised storage systems which record messages and from the ability to transfer large messages efficiently with file transfers. Economy, gained through the use of packet switched stream networks is a less heralded virtue, perhaps because it is not always applicable. Nonetheless, even teleconferencing over the normal audio telephone system can still be cheaper than voice to voice contact because the message storage facility can benefit from time dependent telephone charge tariffs.

Economy and message storage aside, the point to stress about teleconference's contribution to the development of the networks described above is that network members can "interact" in pseudo–real time. That is a network member can read a message lodged hours before and, if he or she has a contribution, respond. The response, as well as the original message, can be broacast to as many members as the sender nominates. They in turn can react. Thus, although it is not normally viewed this way, an active teleconference can have many of the message linking features admired in

[15]Meeks, B. (1985) "An Overview of Conferencing Systems," in BYTE x,13:169–186.

hypertext system.

The Archive's System: Design

These characteristics, combined with the availability of a "free" inter-university/polytechic computing network, convinced the Archive that a teleconference system could support a data-oriented researcher network. The Archive set up a social-science oriented "Bulletin Board" on the University of Essex Computer System. Anyone who can contact the Essex Computer via JANET[16], PSS or a telephone modem can access the Bulletin Board. Once on the Essex Computer System, the teleconference user call the Bulletin Board directly. There is no need to use any other Essex computing facility. The Bulletin Board's facilities are free to registered users.

Formal registration itself is only required to make it easier to collect information about the user community. Initially people had to complete a two-page questionnaire as their admission price. However this went against the spirit of an easy interchange of information and so the Archive dropped the requirement. Now, a user can have free reign over the Bulletin Board by simply sending an electronic mail message asking for an identifier.

The Bulletin Board is organised as a suite of topical "conferences". These include:

The General Household Survey
The Family Expenditure Survey
The Labour Force Survey
The Population Census
The Agricultural Census
Election Studies
Central Statistical Office Time Series Data Bank
British Social Attitude Surveys

Any member of the Bulletin Board can join any (or all) of the conferences and thereby access their contents. There are two access forms: (a) passive access allows the member to read what others have written; (b) active access permits the user to lodge a general comment or question as well as send a private letter to another member (or set of members).

These topical conferences are a starting point. Others may be added while some may be combined. Since the Board began two years ago, the Archive has been its coordinator, thereby filling an editorial function. In time the Archive will produce a printed copy of each topic's contents which it will distribute to the conference's registered members.

the Archive's System: In Practice

It will come as no surprise to readers of this paper that the author is a teleconferencing enthusiast. After two years' experience, however, he feels that he may be alone among his British social scientists. On most counts, the Bulletin Board has not been a massive success.

---

[16]Joint Academic NETwork

To the idea's credit, it has attracted a disciplinary heterogeneous set of registered users. Thus it meets the first criterion for the data centred networks extolled earlier. However there have not been very many overall. At the moment, only 89 people have registered for the Bulletin Board — two of these live outside the UK and so cannot access it interactively.

Within the Board, most people join several, if not all, Conferences. This leads to the suspicion that people who joined are more interested in the computing aspect than they are in sharing data-specific knowledge. The logon pattern reinforces this view. People seem to sign on three times, with a long gap between the second and the third signons. Because few contribute information actively, it appears that people get bored at the static state of the messages. There is nothing to compare with the excitement evident from accounts of the pure computer-oriented teleconferencing systems. Several reasons for this suggest themselves.

Perhaps the most obvious possibility is that a newly found impulse on pin 22 of an 80286 chip is intrinsically more interesting than the ramifications of different definitions of income in the Labour Force Survey. However even if that is so, social scientists are stuck with their subject matter and so have to look elsewhere for the disappointing performance of the Bulletin Board.

The main reason why there are so few active users is that there are so few active users. Like any interactive seminar, a Bulletin Board requires a vibrant critical mass to keep it alive. Without this, the passive entertainment aspect simply kills it.

In the British social science community, the networked computer is not common enough to generate the required mass. The Archive's model for Bulletin Board implied (and expected) that quantitative analysts would connect to the system regularly, perhaps as part of their normal compute-bound routine. It will still be several years before the social scientist's working pattern incorporates daily access to a desktop, networked computing station.

In the interim, the Archive will continue to maintain the teleconferencing system as a background task. Other than a small license fee, the system does not consume many resources. However, if more resources become available the Archive will run an academic conference on teaching quantitative techniques. Although not the Bulletin Board as originally intended, such a conference may be enough to get people interested in the medium.

Data Centred User Groups: The GHS User Group

The data centred user group is perhaps the most ambitious, but certainly the most labour intensive, Archive project to foster a specialist network. While several of these groups have been planned, only one has so far been realised. However, the General Household Survey User Group has developed to such an extent that it is only possible to sketch its programme in this paper.

It should not be surprising that a study like the GHS can encourage the level of activity that it does. Conducted annually since 1971 by the Social Survey Division of the Office of Population Censuses and Surveys, it must count as the single richest source of social indicators in Great Britain. Every year the individuals in over 11,000 households are questioned about aspects of housing, education, employment, health and social services, transport, family life, leisure and social security. Information can be extracted by individual, household and/or by attribute. Multiple-deffined logical views are

possible, as are trend analyses on a vast range of social phenomena. Although the Survey steers clear of conventional attitudinal measures, it offers coverage of behavioural matters which are unequalled in any other data source held by the Archive.

The General Household Survey data analyst faces all of the problems described earlier in the paper. Indeed, the GHS is the archetypical case in the context of that list. Even so, it is one of the Archive's most popular datasets. The register of current GHS users maintained by the User Group lists over 75 active analysts.

The GHS user Group was established in 1986 to provide these users with a regular source of support, as well as a forum for the interchange of experiences with the datasets. The Group has a schedule of regular activities. These include a bi-annual meeting which focuses on a theme related to the analysis of the GHS data[17] and the maintenance of a research register of people interested in exploiting different aspects of the Survey. Equally important, it provides a convenient interface between the team in charge of the Survey and the community of users.

Although it is a mark of the level of interest in the Survey that many of these activities are undertaken by people outside the Archive, the Archive still has a central support role. Indeed its GHS specialist provides the Group with its executive function, to say nothing of serving as the *Newsletter*'s[18] editor.

As mentioned, the GHS User Group was initially established to test the viability of data specific user groups. It has been the Archive's intention to get similar associations started with several of its other larger holdings, but these have been slow to take root. Partially, the problem is one of limited resources. Ideally, each Group requires about 20% of an experienced staff member's efforts. This level of support has not been available so far. Perhaps, though, if the computing problems that took a significant share of archival resources are diminishing, as suggested in the beginning of the paper, that staff will be freed to embark on this new style of data provision.

Summary

The paper argued that archives could no longer survive simply as hewers of wood and drawers of data. Through an analysis of the British Data Archive's user community it showed that a significant share of that Archive's market would benefit from a specialised set of knowledge-sharing networks. Three ongoing projects were described which varied in resource utilisation and return.□

[17]For example, the next meeting of the group will focus on the topic "Using SIR in the Analysis of GHS."

[18] General Household Survey *Newsletter*

Appendix A

Research Projects Using a Single Data Source:
A Selective List

1. A Cross-Section Study of the Distribution of Earnings in the UK.
2. Sex Differences in Sickness Absence from Work - GHS 1975 and 1976.
3. Alternative Approaches to Classfying Women by Social Class.
4. Taxation, Incentives and the Distribution of Income.
5. The Economic Value of Life Saving: An Estimate from the British Labour Market.
6. Problems in Human Capital Analysis: The Case of Great Britain.
7. The Education, Occupations and Earnings of Men and Women.
8. An Anlysis of Variation in Job Satisfaction.
9. Determinants of Housing Tenure Choice in UK.
10. Patterns of Family Formation and Dissolution in Contemporary Britain.
11. Social and Economic Factors in Fertility Differences.
12. Circumstances of Families with Pre-School or Primary School Children.
13. The International Comparative Programme on Life Cycle Methodology to Integrate Social Indicators.
14. Economics of Discrimination.
15. An Anlysis of the Relationship Between Socio-Economic Factors, Self-Reported Morbidity and the Use of Health Services.
16. The Growth and Distribution of Fringe Benefits in British Industry.
17. An Analysis of Occupational Earnings.
18. Evaluation of the SIR Data Management Package.
19. Work, Household and Marriage in the Earlier Stages of the Life Cycle: 1850-Present.
20. Economic Analysis of Female Labour Force Participation Rates.

Appendix B

A Typical Workshop
Agenda

CJG/12th June 1987

FAMILY EXPENDITURE SURVEY SEMINAR

Room C119, London School of Economics, Houghton Street, London WC2A 2AE

Thursday, 25th June 1987

AGENDA

12:30–1.00	Registration and Coffe
1:00–2:00	FES: An Originator's Overview M. Janes (DE)
2:00–2:45	FES: Secondary Analysts' Encounters H. Sutherland (London School of Economics) P. Truscott University of Surrey)
2:45–3:15	Accessing the Data: SIR and Not–SIR A. Heath (SIR Inc)
3:15–4:00	FES: Data Collection B. Redpath (OPCS)
4:00–4:30	Open Forum

CJG/12.6.87.