Debates and directions in the future of opinion polling data
by Neil Guppy

The Swedish Election Studies: studies of 30 years of Swedish electoral behavior
by Iris Alfredson

CODA: a Concept Organization and Development Aid for the research environment
by James A. Dewar and James J. Gillogly

Tabulations on the DDA Study Description
by Karsten Rasmussen

Database directories
by Jim Jacobs

Contents of Current Journals
STEIN ROKKAN Prize
Editorial Information

The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine-readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine-readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors

The QUARTERLY is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press Manual of Style or Kate L. Turabian's Manual for Writers Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. "Bibliographic references for numeric social science data files: suggested guidelines". Journal of the American Society for Information Science 30(2):77-82, March 1979. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor: Walter Piøvesan, Research Data Library, W.A.C. Bennett Library, Simon Fraser University Burnaby, B.C., V5A 1S6 CANADA (01)604/291-4349 E-Mail: USERDLIB@SFU.BITNET

Book reviews should be submitted in duplicate to the Book Review Editor: Daniel Tsang, Main Library, University of California P.O. Box 19557, Irvine, California 92713 USA. (01)714/856-4978 E-Mail: DTSANG@ORION.UCL

Key Title: Newsletter - International Association for Social Science Information Service and Technology
ISSN - United States: 0739-1137 Copyright 1985 by IASSIST. All rights reserved.
Debates and directions in the future of opinion polling data

The following two papers were presented at the IASSIST '87 conference in a session entitled: The uses of socio-political data. The session focused on the comparability of electoral data, public opinion data and other comparative data projects, including technical and political factors affecting secondary analysis. (Ed. note).

by Neil Guppy
Department of Anthropology and Sociology
University of British Columbia

A social invention of this century, opinion surveys are now commonplace in liberal democratic societies. Major political parties cannot afford to ignore polling data, market researchers tap public opinion on a plethora of topical issues, and the mass media broadcast polling results in an incessant stream. Daily we see or hear new polling results which reveal how our contemporaries rate the politicians, or the postal service, or the latest soft drink.

Recently, pollsters have taken to asking the public how they feel about polls. Since pollsters are concerned with assessing the images and opinions of the population, it is hardly surprising that polls on polling, or surveys on surveys, have increasingly found their way into the polling literature (see e.g., Roper, 1986; Goyder, 1986). The irony of using polls to evaluate polls is not lost on the pollsters, and a good deal is revealed by these self-assessments.

This paper reviews this recent literature in an attempt to gain some leverage on the potential directions of public opinion research in the next decade or so. I begin with estimates of the sheer volume of polling data now being collected in different countries. This pervasiveness of polling, though, has generated substantial controversy and conflict in the practice of polling. In an attempt to understand the possible ramifications of current practices and techniques for the future of opinion polling data, I review general criticisms levelled against opinion polling. These criticisms are used to organize a discussion of future directions for both the industry, and by implication, for those who rely on poll-generated data.

The Prevalence of Polling

It is difficult, especially on an international scale, to ascertain exactly how much polling data is currently being collected. No central registry of polling data is available, so a variety of proxy estimates must be employed. 1 rely on

---

1This is a revised version of a paper originally presented at the Annual International Association for Social Science Information Service and Technology (IASSIST) Conference held in Vancouver, British Columbia, Canada on May 19-22, 1987.
three — the percentage of the population reporting participation in opinion research studies, the amount of money spent on polling activities, and the frequency of publication of polling results.

One method of assessing the prevalence of polling is to determine how many members of the general public have been involved as respondents in polling or survey research.

<table>
<thead>
<tr>
<th>Table 1: Trends in Respondent Involvement, 1978-1984</th>
</tr>
</thead>
<tbody>
<tr>
<td>%Ever</td>
</tr>
<tr>
<td>%Last Year</td>
</tr>
</tbody>
</table>

Source: Schleifer, 1986

Table 1 shows trend results from a series of U.S. polls where respondents were asked, first, to report whether they had "ever participated in a survey before", and second, whether they had previously been "interviewed for a survey in the past year" (Schleifer, 1986). Since 1980 the majority of Americans have been contacted at least once, and almost one-quarter have participated in at least one poll or survey in the past year.

Table 2 presents similar findings reported by Roper (1986). His findings parallel the results reported in Table 1, demonstrating that most Americans now have first-hand experience with polling and survey research. Furthermore, many Americans have multiple experiences as participants. Results from other countries are less systematic. Goyder (1986) reports that Canadians in a mid-sized city have experienced levels of participation roughly equal to the U.S. findings.

Table 2:
Respondent Involvement in Public Surveys, 1985

<table>
<thead>
<tr>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never</td>
</tr>
<tr>
<td>Once</td>
</tr>
<tr>
<td>Twice</td>
</tr>
<tr>
<td>3-5</td>
</tr>
<tr>
<td>6+</td>
</tr>
<tr>
<td>D.K.</td>
</tr>
</tbody>
</table>

Source: Roper, 1986

Estimates of the amount of money spent on polling are available only for the U.S., where Advertising Age annually reports financial data for polling firms. For the fiscal year ending in 1985, U.S. research firms in the marketing, advertising, and polling sector billed for $1,785.3 million, up 11.5% from the previous year, and more than three times the annual rate of inflation. Of this, approximately 78% was from U.S. based work (see Honomichl, 1986).

In the mid-1970s, Paley et al. (1980) reported that the New York Times ran news stories containing polling data on an average of one in every three days. A rough count of news, editorial, and feature stories in the 1985 New York Times Index reveals 278 items under the heading "public opinion polls" (a non-election year in the U.S.). Worcester (1980) reports that in the United Kingdom (as elsewhere), opinion polls dominate the front-page headlines during the build-up to national elections.

Polling is pervasive, and every indication suggests that growth has continued to this day. Recent advances in random digit dialing and computer assisted telephone interviewing have served to extend the pollsters' reach even
Criticisms of Polling

Several very general criticisms of polling are heard frequently. Often the charge is made that polls are, at best, only a superficial barometer of public beliefs. In its strongest guise, this argument suggests polling results are frequently wrong. A weaker version claims opinion polling gives only a perfunctory account of facile opinions. Others argue that polls are invasive, trampling public privacy by asking for personal information (e.g., political preferences). Yet others complain that polls have fundamentally altered the political process such that substantial policy matters are unduly influenced by popular and often ill-informed opinion rather than by thoughtful deliberation.

These are very basic arguments on which substantial ink has already been spilled. Rather than add to this area of the debate, I will attempt to look behind some of these general objections to more specific issues in the practice of polling.

1. Distortion — one concern is that people don't give true responses when answering the pollsters' questions (Lewis and Schneider, 1982). Individuals lie, or as the pollsters say, respondents "misreport". Sometimes this appears to be deliberate, as when people are asked whether they voted in the last election (evidence shows that more people claim to vote than actually do vote). On other occasions, it is less certain whether people actually lie or whether they are generally confused; a classic example here is a poll conducted by the German magazine Der Spiegel in which a fictitious cabinet member came sixth in popular rankings, ahead of ten real-life ministers of the crown.

2. Non-attitudes — if pollsters ask people questions about which they have no opinion, some people feel pressure to respond and instant opinions may be invented. Evidence suggests that the more remote an issue is from a respondent, the more random is the response. It is especially on this basis that critics claim polls are superficial.

3. Opinion change — individual attitudes are often not stable or deep-seated. Snap-shots from opinion polls may be as interesting as yesterdays news, but they are known to be poor launching pads for general social forecasts.

4. Issue complexity — few issues are so clear-cut that single-attitude questions can capture the essence of the matter. The black and white image of the world that one may acquire from reading opinion poll results does not do justice to the full array of public sentiment.

5. Words and deeds — the ease with which people may express an opinion on a topic is no guarantee of the direction their actions may take. The link between attitude and behaviour has been probed repeatedly, and we still have less than perfect knowledge of when any congruency between the two will hold.

6. Question wording — social scientists have known for some time that subtle changes in question wording can influence response patterns. Asking respondents whether they would "forbid" or "not allow" something leads to very different results, with a swing of some 20% in response frequencies. So too, the sequence of questions in an interview can influence responses.

7. Impersonality — just as students complain that multiple choice exams do not adequately assess the depth of their knowledge, so some
argue that polls similarly distort reality because people are forced to respond in fixed categories which rarely allow them any self-expression. The frame of reference for the entire polling exercise is determined *a priori* and this can easily disqualify certain questions and certain responses.

8. Sampling — the ability of samples of several hundred people (up to about 2,500) to accurately reflect the diversity of opinion in an entire nation has often been doubted. Polls seldom tap the rich or the poor in any society, thereby predominantly reflecting the views of the middle class.

The more general criticisms, and the eight more specific objections to polling listed immediately above, are likely to continue to surface in debates over polling in the foreseeable future. The veracity of these claims is often less compelling than the volume of their elucidation would suggest. Both polling experts and secondary users of polling data are conversant with the limitations involved. It does not follow from this, however, that these criticisms will gradually dissipate, or even more importantly, that they will have no consequences for the future of opinion polling. Public attitudes about opinion polling data are as likely to influence decisions about the future of polling as they are to influence the future of political parties.

What I turn to now is evidence pertaining to criticisms of opinion research in an attempt to develop some perspective on future directions in the polling marketplace.

---

**Directions and Tendencies**

**i) Assessments of Accuracy**

One recurrent question concerning polls is the frequency with which the pollsters accurately reflect public sentiment. William Buchanan (1986) has examined the results of election polling in several Western democracies in an attempt to examine the precision of election forecasts based on opinion surveys of voter intentions. In analyzing 155 polls from 68 national elections, he found that on 22 occasions the wrong party was predicted as being victorious. Beyond forecasting the wrong victor in 1 out of 7 attempts, there appears to be no trend of improvement since 1949. Similar findings are reported by Worcester (1980) for the U.K. In general, erroneous forecasts are made by a group of pollsters for particularly close elections. It is not true that the polls are always wrong, although it is the case that when the polls are wrong, they all tend to be wrong.

A second issue, linked to accuracy, is the actual reporting of polling results. Here the question is not so much whether the polls are correct or incorrect, but whether they are properly reported by the press. Reporting is crucial, because it is via press reports that most people form their perceptions of the practices of pollsters. Smith and Verrall (1985) undertook a critical evaluation of Australian television coverage of election opinion polls and they claim "poll coverage is extensive, superficial, and inaccurate". Typical errors included "temporal transposition" (incorrectly attributing past or present results to some future point), overgeneralization (extending claims to beyond the sample universe), overstatement (exaggerating the strength of findings), and making ambiguous contrasts (comparisons between poorly conceived groups or time periods).
Related to this is yet a third aspect, the completeness of press reports on opinion polls. Here the concern is with whether or not the press meets basic reporting standards so that consumers can make informed judgments regarding polling results. Table 3 contains a listing of the basic standards, showing the percentage of poll reports (either election or non-election polls) which comply with these basic levels of reporting adequacy. As the percentages reveal, the three papers under examination (L.A. Times, Chicago Tribune, Atlantic Constitution) do not do a particularly good job of providing basic information for informed judgments about polling results. These figures are from the 1970s, and there may have been improvements since this time, especially as the media assign specific people to do all their polling reports. No systematic evidence is currently available to assess possible improvements.

Table 3: Polling Standards versus Polling Practice 1972-79

<table>
<thead>
<tr>
<th>Standards</th>
<th>% Reported in Election</th>
<th>% Reported in Non-Election</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>89</td>
<td>81</td>
</tr>
<tr>
<td>Sponsor</td>
<td>80</td>
<td>87</td>
</tr>
<tr>
<td>Wording</td>
<td>71</td>
<td>34</td>
</tr>
<tr>
<td>Sampling Error</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td>Population</td>
<td>91</td>
<td>66</td>
</tr>
<tr>
<td>Method</td>
<td>62</td>
<td>38</td>
</tr>
<tr>
<td>Timing</td>
<td>76</td>
<td>48</td>
</tr>
<tr>
<td>N</td>
<td>61</td>
<td>55</td>
</tr>
</tbody>
</table>

Source: Miller and Hurd, 1982. [in Norwegian]

Finally, pollsters have asked the public about their perceptions of polling accuracy. Andrew Kohut (President of Gallup) reports that in 1985, 68% of respondents thought the pollsters got election forecasts correct most of the time (Kohut, 1986). This was an increase in public confidence from 57% in 1944. When asked about non-election polls, however, only a slim majority of people felt the polls were generally right in tapping the public mood (52% in 1944, 55% in 1985). Conversely, negative sentiments about non-election polling accuracy seem to have increased, with 21% saying they felt the polls were "not right at all" (up from 12% in 1944). The British appear to be a little more sceptical about polling results, with only 32% believing that "the opinion polls are normally right" (46% thought they were normally wrong, with 22% giving other responses — see Worcester, 1980: 561).

If public opinion is as influential as the practice of polling implies, then pollsters need to be conversant with the public images of polling. When gauged by specific measures of accuracy, polling does not have a massive majority of support.

**ii) Bogus Polls**

Polling got a bad name in the 1936 U.S. presidential election when the Literary Digest, a magazine for affluent Americans, asked readers to write in with their choice for president. On the basis of these responses, the Digest predicted that Alf Landon would win the election, opening itself and the prestige of polls to ridicule when Franklin Roosevelt won by a landslide. Pollsters have insisted that only surveys with "scientifically" selected random samples should be called polls, and for some time this seemed to be accepted practice. Recently, however, phoney or bogus polls have become more prevalent. For example, after the 1980 Carter-Reagan debate, ABC asked viewers to phone and report who they felt won the debate. On the basis of some 727,000 calls, ABC reported that their "poll" showed people felt Reagan had won 2-1. This phenomenon of
self-selected "samples" appears to be growing in the polling marketplace. QUBE is a more recent invention, allowing cable subscribers to send digital signals back through the video cable to record their "vote" on various issues. Political parties in several countries have taken to doing "surveys" of the public, asking people first to rate the current government, and then to donate money to help the party doing the survey.

The prevalence of these bogus surveys is hard to detect, but concern is mounting that sales people are using this technique to identify potential customers. Schleifer (1986) reports that in 1980 some 13% of U.S. respondents reported having been exposed to false surveys, a percentage that increased by 4 points to 17% in 1984. With almost 1 person in 5 being confronted with bogus surveys, the reputation of the industry could be tarnished quickly and decisively.

iii) Respondent burden

Knowing that 1 in 5 people are approached by phoney surveys, one wonders about the frequency with which people have been approached by legitimate pollsters or survey researchers. Estimates vary, as shown in Tables 1 and 2, but over one-half the population in the U.S. appears to have been involved in a survey or poll at some time. Schleifer (1986) reports that of the 23% of respondents who had been involved in a survey in the past year, almost 1 in 5 had participated in four or more polls. These latter individuals are known to survey researchers as "professional respondents" and the inclusion of the same people in multiple surveys has pollsters worried about the "freshness" of their samples.

The growth of polling raises the possibility of 'over-kill' — people will be 'turned-off' polls by too many requests for their help. One way of assessing respondent burden is to examine empirical evidence of possible overexposure.

Steeh's (1981) results, shown in Figure 1, chart refusal rates between 1952 and 1980 for two national U.S. samples, both conducted by the University of Michigan Survey Research Center. As the graph shows, refusal rates in both the election and consumer attitude series are rising, although whether or not this is due to respondent burden per se is difficult to determine. As Steeh notes it could be caused by one factor or a combination of factors, including overexposure, disillusionment with the use or accuracy of survey results (see above), or heightened concern about privacy and confidentiality. Goyder and Leiper (1986) report similar trends based on an analysis of polling and survey response rates in the U.K., the U.S., and Canada, and they point to rising criticism of Census practices, especially in Canada and the U.K.

iv) Exit polls

In 1980, Jimmy Carter conceded defeat before the polls had closed in the American west. One reason for this was that the television networks were using exit polls to predict the winner before everyone had had an opportunity to vote. Exit polls (or 'same-day polls' as they are called in Britain) are conducted by standing outside selected polling places and asking those leaving for whom they had voted. Based on these reports, the networks have been able to forecast with accuracy the eventual winner. The State of Washington was so upset with this practice that they banned exit polls, making it illegal for people to conduct surveys within 300 feet of a polling station. The media challenged the law in court, losing an initial verdict and then winning on appeal — the state is currently appealing the appeal.

Whatever the eventual outcome, the concern remains that the techniques and the process of polling have fundamentally altered the practice of politics. Whether exit polls actually alter the outcome of elections is debatable (see Sudman, 1986), although they do appear to have a
marginal impact on voter turnout when a landslide has occurred. The key point here, however, is not whether exit polls actually have any effect, but that people believe they have an effect. As pollsters themselves have shown, it is the image that is important.

v) Polling Initiatives

Pollsters have recently been expanding their craft at a rapid rate. The use of polls for marketing is an old and established pastime (Labatts Brewery in Canada has opinion data dating back to 1910 in Canada). More recently, polling has had an influence in the courtroom where survey results have been used in judgments over trademark protection (the NFL, Corning Glass Works), advertising claims (Pepsi vs Coke), and jury selection (Ford, IBM, MCI Communications). Furthermore, various departments of government charged with regulatory functions have begun to use polling data to assess the impact of certain initiatives. Listerine was required to engage in corrective advertising to dispel the myth they had created that the mouthwash would prevent people acquiring colds and sore throats. To assess the effectiveness of the correction, the U.S. regulatory agency that was responsible for enforcement used a poll to examine changes in opinion (see Crespi, 1987; Dutka, 1982).

As image and knowledge grow in importance, the pollster's craft is more in demand. But as the demand rises, the value of information escalates, and polling agencies in the private marketplace are less willing to freely relinquish their data. Beyond cost, the sheer volume of information frequently makes archiving data a burden to avoid — profit lies with the next project.

Conclusions

U.S. respondents continue to report that they feel polls are "a good thing" (73% in 1944 and 76% in 1985 — see Kohut, 1985), although the British are less sanguine, with a majority feeling that polls were "pointless" or "not very accurate" (Worcester, 1980: 560). Potentially, fatal dangers for the polling industry would appear to lurk in the areas of bogus polls and respondent burden. The very prevalence of polling may undermine the craft as individuals feel inundated with strangers asking dubious questions about issues which people increasingly define as nobody else's business (see Goyder and Leiper, 1986 for an analysis of increasing objections to the Census). Serious, but probably not fatal, dangers would appear to lie in the possibility of disastrous election predictions in several countries simultaneously, or the use of polls in a way so as to make people feel their personal freedoms or rights are subverted (as seems to be the case with exit polls).

Finally, several signals suggest that more and more public attitude data from polling firms will become off-limits. Currently the vast majority of opinion polling data is not publicly available. Increasingly, polling data will be kept secret as polling agencies realize the economic value of trend projections. As the value of information grows, pollsters will protect their investments and profitability. Several companies in North America now conduct omnibus surveys which they keep confidential. In addition, several countries now have governments collecting general social survey data. Thus there will be less pressure on the pollsters to serve the academic interest by releasing the poll data.
References


Honomichl, Jack. The nation’s top marketing/advertising research companies. Advertising age May 19, 1986.


The Swedish Election Studies: studies of 30 years of Swedish electoral behavior

by Iris Alfredson

Before using data for secondary analysis, the researcher must be aware of the principles used in compiling the material. The aim of this paper is to give a short description of the Swedish election studies, and discuss some of the problems which may arise if one is to study changes over time using these data.

Quite a few variables are common to all ten studies, but are they comparable? There may have been slight changes in question wording, or the coding of the variable may have changed over time. These are some of the changes that affect direct comparisons. Additional complications arise when making international comparisons, but such problems are outside the scope of this paper.

I hope that this presentation will act as an introduction to the extensive material available on Swedish elections. For those whose requirements are more comprehensive, a bibliography of books, reports and papers on the subject is provided.

Background

In conjunction with the local elections of 1954, Jorgen Westerstahl and Bo Sarlvik of the Department of Political Science, University of Gothenburg, conducted the first scientific field survey of Swedish voting behavior. The survey was a local election study conducted in Gothenburg and the countryside around Borås. The survey was inspired by that conducted by Lazarsfeld, Berelson and Gaudet in Eire County. The questions concerned mainly social background, political party sympathies and motives, previous political sympathies, stated and actual vote behavior, reasons for change in sympathies, public reaction to content of media coverage, political sympathies of friends, and knowledge of and political opinions on specific election campaign issues. The 1954 survey was a pilot test for a larger national study. Two years later, in conjunction with the parliamentary election of 1956, the first nationwide survey on party choices, participation and political opinions of the Swedish electorate was conducted. This was the real start of the Swedish election research program. Since then, similar surveys have been carried out at all elections. Further, studies have been conducted in conjunction with the two referenda that have taken place since then, the referendum on the general supplementary pension scheme (ATP) in 1957, and a referendum on nuclear power in 1980.

The Swedish election studies are, together with those carried out in the United States, Norway,
France and West Germany, the only academic election studies that extend as far back as the 1950s. The Swedish parliamentary election studies have been carried out at every election since 1956, making them one of the most comprehensive sources on voting behavior.

**Financing**

The local election study of 1954 and the national election study of 1956 were funded by the Swedish Social Science and Legal Research Council, the Foundation for Research in Sociology at the University of Gothenburg, the Swedish Broadcasting Cooperation and the political parties. Since 1960, the election studies have been financed through government grants and carried out as a part of the election statistics program by the Central Bureau of Statistics (Statistiska Centralbyrån).

**Survey Design**

Since the mid-1950s, there has been a political behavior research program at the Department of Political Science, University of Gothenburg. The election studies have, with the exception of the 1976 study, come into being through a close collaboration between the Department of Political Science at the University of Gothenburg and the Swedish Central Bureau of Statistics (SCB). The 1976 study was conducted by the Department of Political Science in Uppsala in collaboration with SCB.

The Survey Research Center of the Central Bureau of Statistics is responsible for the sampling for the election studies, their permanent interview organization performs the field work, and they also collect additional data from public registers. The research project at the Department of Political Science is responsible for the general planning of the studies, the construction of the questionnaires, and the analysis and presentation of data. The basic principles of the studies, the questionnaires, and the coding scheme were originally designed by Bo Sarlvik. The surveys were initiated by professor Jörgen Westerståhl and directed by Bo Sarlvik (1956, 1960, 1964, 1968, 1970, 1973), Olof Petersson (1973, 1976), and Sören Holmberg (1979, 1982, 1985).

The sample represents the resident, enfranchised population. The samples for the earlier election studies were drawn from the Survey Research Center's sampling framework which consisted of a nationwide set of primary sampling units which provide the framework for a 'general purpose' two-stage population sample. Since 1973, another method has been applied. The samples are now drawn directly from the SCB register over the total population (RTB), by means of the SCB standard program for random samples. Respondents not included in the target population because of ineligibility to vote are excluded from the sample. The target population even includes Swedes resident abroad, but these are not included in the sample. People above 80 years of age at the time of the study are excluded in order to avoid the difficulties encountered in interviewing very old people: the 1968 survey was an exception, the age limit was 84). The proportion excluded by this age limit is about 3% of the total sample.

In 1956, respondents were interviewed twice, once before and once after election day. From the 1960 study onwards, with the exception of the 1970 study, field work has been carried out in two stages. The total sample is split into two subsamples of equal size. One subsample is contacted for personal interviews during the field work stage preceding the election. Respondents in this subsample are contacted again after election day through a short mail
questionnaire. The primary purpose of this mail questionnaire is to obtain information about the final vote decision of these respondents. The second subsample is contacted for personal interview during the weeks immediately after the election. Most of the interviews are held between mid-August and mid-October.

The 1970 election study differs from the others because the entire survey was carried out after election day. Different techniques were used in 1970: about 1/3 of the sample were interviewed in their homes, and additional 1/3 through telephone interviews, and the remainder received a short mail questionnaire. In the other election surveys, respondents were interviewed in their homes. Some busy respondents were interviewed by telephone. Respondents interviewed before the election received, immediately after election day, a short mail questionnaire which mainly contained questions on final voting decision.

Panels

In order to facilitate the study of variation and constancy in voting behavior, a panel design has, with one exception, been used. The panel technique used has varied over the years. In 1956, respondents were interviewed twice, once before and once after election day. In 1960, no panel was used. In 1964, a three-stage panel was started, in which the same respondents were interviewed at the elections of 1964, 1965, and 1970. A major problem with a panel extending over such a long period of time is that the sample loss has a tendency to increase at every stage.

In order to facilitate the study of individual changes, but at the same time avoid too large a sample loss, a new type of panel was introduced in 1973. This was a kind of "rolling" two-stage panel, in which half of the 1973 sample was reinterviewed in 1976. The "new" respondents in 1976 were reinterviewed in 1979, and so on. In this way, all respondents are interviewed twice. At present, there are four panels: 1973-1976, 1976-1979, 1979-1982 and 1982-1985. In every survey, a supplementary sample of first-time voters, who have become entitled to vote since the last election, is also drawn.

In conjunction with the 1957 referendum, respondents were interviewed three times; this design was also applied to the 1980 referendum. The questionnaire which was sent out to respondents in the 1980 referendum sample was also sent out to respondents belonging to the 1979 election study sample. In this way, two long term panels were created, one for the years 1976-1979-1980 and one for the years 1979-1980-1982; this makes it possible to do analyses over a longer time span.

Sample Loss

During the first decades in which Swedish field surveys were conducted, sample loss rates were low, about 5% - 7%. In the 1960s and the beginning of the 1970s, there was a dramatic rise, the sample loss increasing to about 20% before stabilizing at that level. Sample loss in the Swedish election studies have also followed this pattern: the 1965 sample loss was 5%, during the 1960s it slowly increased (8%, 8% and 12%) until it reached 14% in the 1970 survey. Then there was a sudden increase, from 18% in the 1973 survey to 26% in the 1976 survey. In the latest election studies, sample loss has been reduced to less than 20%, mainly through the use of shorter interviews. The main portion of the sample loss was due to refusals to participate, and through shorter interviews with respondents who are unwilling
or are pressed for time, sample loss can be reduced. The shortened interviews take about half the time of a normal interview. Some respondents answered an extremely short interview, containing only those questions on final vote decision.

The framing of the questions

The Swedish election studies series now extends over a period of thirty years. At present, the series consists of ten surveys conducted in conjunction with parliamentary elections, and two conducted in conjunction with referenda. During this period, a great number of questions have been asked. Some questions are repeated in all surveys, which makes it possible to study changes over a period of thirty years, and there are also questions specific to one or two surveys. Some surveys have a large number of questions on media, while others (1970 and 1973) don’t touch on the subject at all. In the 1973 and 1976 surveys, there were many questions on international politics and events, an area not covered in the other surveys at all.

The 1968 election study was almost twice as large as the others, largely because of the goal of the survey which was to cover the system of representation. Comparable questions were also asked in an interview survey of Members of the Lower House of Parliament. Questions asked on current topics include: questions on public representation on bank boards (1968, 1970), possible Swedish membership in the European Economic Community (1968, 1970, 1973), questions on nuclear power (1976, 1979), and wage-earners’ investment funds (1979, 1982).

However, the central questions in an election study are always those on party preference. Therefore, all surveys contain questions on: respondent voting habits, party preferences, and voting in both current and previous elections. Voter participation is also checked using electoral registers. Information on father’s political sympathies is also available in most of the surveys. Other recurring questions cover political interest and party identification.

Social background factors such as information about the respondents’ date of birth, sex, marital status, education, occupation and trade union affiliation are also available. Information on occupation, education, trade union affiliation and political party membership in the 1970 survey (the third stage of the 1964–1968–1970 panel) was extracted from the 1968 survey, for all respondents except those lost from the 1968 sample and those added in the 1970 supplementary sample. For these respondents, information from 1970 was used.

The Swedish Social Science Data Service is at present compiling a continuity guide to all questions and variables used in the Swedish election studies.

Problems of comparability

A series of surveys extending over thirty years could be a gold mine for researchers wishing to study changes over time. Such comparisons are not, however, always problem-free: a number of factors make comparison difficult, such as changes in question wording between surveys. In addition, questions may not address the same groups, the variable coding may have changed, the source of background variables may change from registers to interviews, or vice versa. Changes in society, such as an increasing proportion of women in the work force, and

---

1. Representationsundersökningen, by Bo Sarivik et al.)
new education systems, also make direct comparisons difficult. In order to save time and trouble, it is necessary to study the construction of the variables carefully before starting analysis. The following are problems that have been identified in the Swedish election studies.

Marital status: Information on the respondent's marital status has alternately been gathered through questionnaires and from registers. In 1956, the question asked was: "Are you married, unmarried, widow/widower, divorced?". In 1960-1976, this information was extracted from a population register. In the most recent surveys, the information has again been collected in the interview. The question in 1979/1982 was worded: "Which of the following alternatives best describes your marital status (Married or co-habitating, unmarried or divorced, widow/widower)?" The advantage of this method is that the data collected are more up-to-date. The number of code categories used has also varied over time; from 1956 to 1968, four categories were used: married, widow/widower, divorced or unmarried. Since the end of the 1960s, it became more common that people live together without being married; this is reflected in the election studies. In the 1970 survey, the category 'unmarried but cohabiting' was introduced. Since the 1976 survey, the number of categories has been reduced to three: married/cohabiting couple, unmarried or divorced, widow/widower.

Education: Changes have been made to the variables on education over the years, because of changes in society. New school systems have been introduced, and the general level of education has risen dramatically. The coding scheme has also changed: the election studies conducted from 1968 to 1976 used a very detailed coding for education. Question wording has remained fairly constant over the years: Have you any education above 'folkskol' level? (IF YES:) What education do you have?" (1956-1964) "Do you have any practical or theoretical education above 'folkskol' level? What education? Have you any other practical or theoretic education?" (1968-1973) "What education do you have? Have you any other practical or theoretic education?" (1976-1982). The early election studies (1956-1964) have the same education code categories, except that category 2 (folkhögskola, yrkesskola) in 1956 was divided into two categories in the following two surveys. Categories in the earlier surveys remain in the more recent surveys (1979-1982), but changes in the school system are easy to track. Categories 3 and 4 in the 1960 and 1964 election surveys have been combined into a new category 3. This category includes 'grundskola', a level which did not exist earlier. There is a new category, '4', which includes education in 2-year 'gymnasie' courses; this is also a new level of education, as compared with earlier studies. During the period 1968-1976, a very detailed coding scheme was used for education. The question on education was split, with a variable for general basic education, followed by additional variables for other education. Education is coded with a three digit cod, of which the first digit denotes the main group, the second the level of education, and the third the type of education (degree).

Occupation: Major problems of comparability occur in the definitions of work and class. There are problems in the classification of married women and students, and a new classification system for occupation has been introduced. The variable 'occupation group' is present in all surveys. In the 1956 survey, it had 12 categories; since 1960, it has had over 30 categories. Married, non-working women have been included in different categories over the years: in the 1956 survey, all married women were classified according to husband's occupation. In 1960 and 1964, working married women were coded according to their own occupation, while non-working married women were coded according to husband's occupation. In 1966, 1970 and 1973, a special code was used for these women, and since 1976, they have been classified according to previous occupation.
Students have also been treated differently over the years. Since 1970 they have had a separate code category, but before that, they were coded according to the social status or occupation of the family head. Since 1968, a detailed, 3-digit code has been used for occupation. The first two digits denote area of occupation, the third digit status of occupation.

Place of residence: All surveys, with the exception of the 1970 and 1973 studies, include information on respondent’s place of residence. The categories have changed from survey to survey. With the exception of the 1982 survey, it is possible, by combining categories, to extract three comparable categories: large cities (Stockholm, Göteborg and Malmö), other towns, and rural areas and villages.

Party preference: The question "Which party did you vote for?" is asked in all surveys. Respondents interviewed before the election, with exception of the 1964 survey, received a questionnaire immediately after the election. Respondents in the 1964 pre-election sample were asked "Which party do you like best?" Information on who actually voted in the election is also available. In most surveys (1956 to 1964 and 1976 to 1982), one variable is used to summarize the contents of the two variables on voting and election participation. Respondents who stated that they voted for a certain party, but who, according to the election register, didn’t vote, are excluded. With the exception of the 1956 study, all surveys have contained a question on how the respondent voted in earlier elections: "Did you vote in the election 19…? (IF YES:) Which party did you vote for?" Similar data can, from the 1956 survey, be extracted by combining the answers to the following questions: "Did you vote for the same party at earlier elections?" and for those who answered ’no’: "Which party did you vote for?"

Party identification: All studies include questions on party identification. In the earlier surveys (1956–1964) there is only one question: "Some people are strongly convinced adherents of their party. Others are not so strongly convinced. Do you yourself belong to the strongly convinced adherents of your party?" In the later surveys, party identification is measured in three questions: "Many feel strongly for a particular party, whilst others do not feel the same allegiance towards any of the parties. How do you see yourself, as a liberal or social democrat or moderate or centrist or communist? Or don’t you have this attitude towards any of the parties?" The respondents who consider themselves party adherents are asked the following question: "Which party do you like best?" and finally, those who named a party were asked: "Some people are strongly convinced adherents …".

Newspapers: With the exception of the 1970 and 1973 surveys, questions on which newspapers respondents read have been asked. But one must be careful with these data. In the earlier studies (1956 to 1964), respondents were asked which newspapers they read daily, while in the remaining surveys (1968, 1976–1982), respondents were asked which papers they read regularly. In the 1976 survey, 'regularly' was defined as at least 4 times/week for daily newspapers and at least every second week for weekly papers, while in the 1979–1982 surveys, 'regularly' was defined as at least once/week for all papers. In 1968, only one-half the sample were asked which papers they read regularly. This question was asked in the mail questionnaire which was sent to pre-election respondents, and 'regularly' was not defined. The coding of the variable in the earlier surveys makes direct comparison with later surveys impossible. The code categories included a variety of combinations of information on subscriptions, political affiliation of the newspapers, type of newspaper, etc. Since 1966, the names of the newspapers have been coded.

A project to recode the oldest surveys, is currently ongoing at the Department of Political
Science, Gothenburg University. This will, hopefully, solve many of the problems of non-comparability among the surveys.

### Swedish election studies at the Swedish: Social Science Data Service (SSD)

All election studies through 1982 and the 1980 referendum survey have been deposited in the Swedish Social Science Data Service (SSD). With the exception of the referendum study, they have all been documented with the aid of the GIDO-system, which produces a data file and machine-readable codebook in OSIRIS format. These OSIRIS-format files can be converted to other formats. The 1980 referendum is currently being documented, and we hope that, in the near future, the 1985 election survey will also be available. It is also possible that the referendum study of 1957 will be documented.

The following is a short summary of the Swedish election studies currently available from the Swedish Social Science Data Service (SSD):

#### SWEDISH ELECTION STUDY, 1960 (SSD 0001)

Principal investigator: Bo Särvlvik, Department of Political Science, University of Gothenburg.
Total sample: 1,603
Number of respondents: 1,466
Sample loss: 137 (8.5%)
Number of variables: 215
Method: Interview in home. One half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.
Panel: None.
Format: OSIRIS, SPSS-x, machine-readable codebook.

#### SWEDISH ELECTION STUDY, 1964 (SSD 0007)

Principal investigator: Bo Särvlvik, Department of Political Science, University of Gothenburg.
Total sample: 3,109
Number of respondents: 2,849
Sample loss: 260 (8.4%)
Number of variables: 219
Method: Interview in home. One–half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.
Panel: The first stage of a three–stage panel study in which the sample was reinterviewed in conjunction with the parliamentary elections of 1966 and 1970.
Format: OSIRIS, SPSS-x, machine-readable codebook.

#### SWEDISH ELECTION STUDY, 1968 (SSD 0039)

Principal investigator: Bo Särvlvik, Department of Political Science, University of Gothenburg.
Total sample: 3,109
Number of respondents: 2,849
Sample loss: 260 (8.4%)
Number of variables: 219
Method: Interview in home. One–half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.
Panel: The first stage of a three–stage panel study in which the sample was reinterviewed in conjunction with the parliamentary elections of 1966 and 1970.
Format: OSIRIS, SPSS-x, machine-readable codebook.
Principal investigator: Bo Särkvik, Department of Political Science, University of Gothenburg.

**Total sample:** 3,356

**Number of respondents:** 2,943

**Sample loss:** 413 (12.3%)

**Number of variables:** 532

**Method:** Interview in home. One-half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.

**Panel:** The second stage of the three-stage 1964–1968–1970 panel study.

**Format:** OSIRIS, SPSS-x, machine-readable codebook.

---

**SWEDISH ELECTION STUDY, 1970 (SSD 0047)**

Principal investigator: Bo Särkvik, Department of Political Science, University of Gothenburg.

**Total sample:** 4,815

- Interview in home 1,602
- Telephone interview 1,580
- Mail questionnaire 1,633

**Number of respondents:** 4,130

- Interview in home 1,355
- Telephone interview 1,407
- Mail questionnaire 1,368

**Sample loss:** 685 (14.2%)

- Interview in home 247 (15.4%)
- Telephone interview 173 (10.9%)
- Mail questionnaire 265 (16.2%)

**Number of variables:** 223

**Method:** Three types of interviews: in-home interviews, a somewhat shorter telephone interview, and a mail questionnaire with only a small number of questions. All interviews were conducted after the election.

**Panel:** The study represents stage three in the 1964–1968–1970 panel.

**Format:** OSIRIS, SPSS-x, machine-readable codebook.

---

**SWEDISH ELECTION STUDY, 1973 (SSD 0040)**

Principal investigators: Bo Särkvik and Olof Petersson, Department of Political Science, University of Gothenburg.

**Total sample:** 3,179

**Number of respondents:** 2,596

**Sample loss:** 583 (18.3%)

**Number of variables:** 239

**Method:** Interview in home. One-half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.

**Panel:** The study represents stage one in the 1973–1976 panel.

**Format:** OSIRIS, SPSS-x, machine-readable codebook.

---

**SWEDISH ELECTION STUDY, 1976 (SSD 0008)**

Principal investigator: Olof Petersson, Department of Political Science, University of Uppsala.

**Total sample:** 3,580

**Number of respondents:** 2,652

**Sample loss:** 928 (25.9%)

**Number of variables:** 290

**Weight:** Respondents belonging to the 1973–1976 panel who did not respond in 1973 had sample probability halved in 1976; when processing the data, these were weighted by a factor of 2.

**Method:** Interview in home. One-half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.

**Panel:** The study represents stage two in the 1973–1976 panel and stage one in the 1976–1979 panel.

**Format:** OSIRIS, SPSS-x, machine-readable codebook.

---

**SWEDISH ELECTION STUDY, 1979 (SSD 0089)**
Principal investigator: Sören Holmberg, Department of Political Science, University of Gothenburg.

Total sample: 3,498
Number of respondents: 2,816
Sample loss: 682 (19.5%)
Number of variables: 303

Weight: Respondents belonging to the 1976–1979 panel who did not respond in 1976 had sample probability halved in 1979. These are represented by two records in the data file. The number of respondents is therefore 2,905 and the sample loss is 853.

Method: Interview in home. One-half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.

Panel: The study represents stage two in the 1976–1979 panel and stage one in the 1979–1982 panel and stage one in the 1982–1985 panel.

Format: OSIRIS, SPSS-x, machine-readable codebook.

SWEDISH ELECTION STUDY, 1982 (SSD 0157)

Principal investigator: Sören Holmberg, Department of Political Science, University of Gothenburg.

Total sample: 3,597
Number of respondents: 2,943
Sample loss: 654 (18.2%)
Number of variables: 303

Weight: Respondents belonging to the 1979–1982 panel who did not respond in 1979 had sample probability halved in 1982. These are represented by two records in the data file. The number of respondents is therefore 2,980 and the sample loss is 744.

Method: Interview in home. One-half of the sample was interviewed before election day, the other half after. Pre-election respondents also answered a short mail questionnaire after the election, which mainly contained questions on final vote decision.

Panel: The study represents stage two in the 1979–1982 panel and stage one in the 1982–1985 panel.

Format: OSIRIS, SPSS-x, machine-readable codebook.

Publications

A number of books, reports, and papers have been published based on the results of the election studies. The following list does not claim to be comprehensive.

Bibliography

Clausen, Aage, Sören Holmberg and L. Dehavens. Contextual factors in the accuracy of leader perception of constituents' views.


Westerståhl, Jörgen and Bo Särkvik. Svensk
Westerståhl, Jörgen and Bo Särövik. Svensk valrörelse 1956. Arbetsrapport I:
Intervjuundersökningen. [TS] Göteborg: Statsvetenskapliga institutionen, Göteborgs
universitet, 1957.
CODA: a Concept Organization and Development Aid for the research environment

by James A. Dewar and James J. Gillogly

Introduction

In 1982, the authors and colleague Morlie Graubard were led by their experiences to wonder what the role of computers might be in policy research (our primary occupations). The role of computers in analysis, particularly quantitative analysis, is extensive and well documented. Computerized text editing (as in the preparation of this manuscript) is quickly replacing the typewriter throughout the industrialized world. Artificial Intelligence researchers at Rand and elsewhere are exploring the possibility that computers might do some of a researcher’s thinking. Data retrieval systems are bringing vast stores of data within the reach of a researcher’s specialized interests. But it seems as if there ought to be other ways that the power of computers could help in the research process.

We focused on the part of the typical research project that involves reading vast amounts of information related to a research topic and remembering those parts which are pertinent to one’s current thesis. This part of research predates computers and is currently supported, to some extent, by a variety of file management and data base management systems. The idea of the computer acting as a long term memory for the researcher seemed a natural combination of the speed and memory of the computer with the more subtle, synthetic powers of the researcher.

In looking at several extant storage and retrieval systems, however, they seemed to be directed either toward very large data bases or at personal computer applications. As such, each seemed to have important limitations when applied specifically to the problems of the individual researcher. This led us back to take a look at the research process itself in an attempt to define those aspects of search and retrieval that most suited the policy researcher’s environment.

The key seemed to be that the research process of an individual was essentially an interactive process characterized by both a growing data base and a series of failed attempts at organizing the data into a coherent whole (followed ultimately by a successful attempt, of course). This led us to a hypothesis about the utility of computers “optimized” for the policy research process, and from there to a list of desirable characteristics for the associated computer aid.

HYPOTHESIS:

Computers can aid the policy research process by acting as a long term memory (storage and retrieval facility) for the researcher’s growing data base and changing concepts.

The realization of this hypothesis in the form of computer software specifications required constant referral back to the research process and an appreciation of the limitations of modern computers. The following list of desiderata reflects the results of that effort along with our Justifications for each.

Desiderata

1. Quick Boolean tag searches.

In retrieving data, we have found with other retrieval systems that even a 1 to 2 second delay between the request and the results was very distracting to a researcher’s thought processes. Boolean searches involve the use of the logical connectives AND, OR, and NOT; require longer than single search requests, in general; yet should respond just as quickly as single search requests if the researcher is not to be distracted from the problem at hand. It should be emphasized that the requirements here are placed on TAG searches—full-text searches are exempt from this specification because we thought that the primary search method would be by tags and that the researcher would not mind delays for the occasional full-text search.

2. Flexible tagging rules.

The process whereby a human retrieves data from his/her own memory is ill understood, except that it seems to involve a variety of mental processes. Retrieving data from a computer is more limited, but we wanted the researcher to be able to "mark" or tag data for retrieval with as much flexibility as possible. S/he shouldn't be restricted to words that appear in the text or single word tags or ...
3. Powerful tag changing capability.

This would seem to be the key for a policy researcher. As one’s concepts change, the relevance of one’s data to the new concept usually changes also. We wanted the researcher to have the ability to make wholesale changes to the markers or tags on data. This seemed to imply two capabilities: 1) The ability to do both full-text and tag searches and then to "prune" or shape the resulting list of data records, and 2) the ability to then make changes in the tags of those records all at once.

4. Recall by data capability.

This is a common capability in storage and retrieval systems and seemed a useful adjunct for a research environment that depends heavily on dated articles. This also includes the ability to do comparative operations on dates (such as, all dates from date 1 to date 2, etc.).

5. Data entry from keyboard or file.

This would seem to be useful for two reasons: 1) Many researchers already have an electronic data base that would benefit from this capability, and 2) Rand (among others) now has the capability of obtaining the results of large data base searches in electronic form, which could then be easily entered in that form.

These, then, formed the basic requirements that we thought defined a computer capability "optimized" for the needs of the policy researcher. With these specifications in mind, we again looked at available software for file and data base management. Without claiming to have done an exhaustive search, we did talk to a variety of data base specialists, search the on-line literature, and visit computer shows. In the end, we were sufficiently disappointed with lack of matching between our desiderata and the available software that we decided to build our own. As a point of interest, the common mismatches were either slow response times (typical of microprocessor based systems) or insufficient capability to do easy, wholesale changes to the tags in the system (typical of systems aimed at very large data bases).

The resulting system was called CODA (for Concept Organization and Development Aid) and that system is the topic of this paper. In the following sections we will describe the prototype system we built in order to test our hypothesis, the system’s capabilities and limitations, some of the details of its user interface, what we have learned both from the building and testing of the system, and, finally, some thoughts on further capabilities that appear to be amenable to computer implementation and that might aid the policy researcher.
The Prototype

The CODA program most properly qualifies as a file management system aimed at small data bases and a very limited number of users. The specifications followed in its development were the desiderata listed above, and the reason for developing it was to test the hypothesis that the specific task of policy research could be improved by an appropriate computer aid. It is a system designed and implemented by users (policy researchers) for testing some concepts about the users' environment. As such, there are some specific things that CODA is NOT. It is not a full data base management system for general use, it is not particularly suited for large data bases or numerical processing, and it is not a product that the Rand Corporation is trying to sell.

With those caveats in mind, we were interested in describing CODA to this conference both as a way of highlighting the data base needs of one special segment of the user community, and as a way of elicitng feedback from the professional community of data management specialists.

Before proceeding with a description of CODA, it is important to discuss briefly the major "buzz words" that we have come to use in our description of the system.

tag: This is a user-supplied word or phrase that is typically used in CODA for retrieving a data item. In other systems this is called a keyword. 'Tag' is used in CODA because it need not be something that actually appears in a record. A given record can have many tags.

record: This is another word for datum and refers to an individual recallable item of data in the system. In CODA, a record is any (unformatted) text that is of interest to the researcher. This is intended to include anything from a single word to several paragraphs, from a table of numbers to a collection of symbols, etc.

hit: This is any record that contains a specified search tag (or tag combination). If, for example, CODA were commanded to return all records having 'important' as a tag, CODA would return that it had found, for example, 34 'hits' or records that contained 'important' as a tag.

index: This is undoubtedly the ugliest of CODA's buzz words. There are two kinds of indices in CODA: date indices and others. Date indices are a way of grouping different kinds of dates for recall. In this way, the user is able to differentiate between, for example, the date on which the material in a record was published and the date on which it was entered into the system. The principle behind the other indices is much like that of the indices in a books. It is a way of grouping tags both for display and for search purposes. For display purposes, the intention was to give the user a tool akin to the Author index found in some books. It is much easier to look up a half-remembered author in a smaller author index than it is to do so in a large full index. In addition, one can retrieve, specifically, an author's works, for example, rather than retrieving his works as well as anything written about him.
Technically, the prototype CODA system has been written under UNIX (4.1 BSD) and is running on a VAX 11/780. It is written in C and uses the Curses screen management package. Data entry from the keyboard uses the Rand editor, a full-screen editor used by most of Rand's researchers and secretaries. CODA provides a menu-driven interface to users with a variety of terminal types, including Rand's standard Ann Arbor terminals (in several models) and personal computers connected to the VAX via modems. By servicing all of Rand's terminals, we were able to enlist a variety of Rand researchers to use CODA and feed back their comments as to its utility.

To achieve the recall speed specified in the desiderata, the tags are loaded into a hash table in memory with linked lists pointing to the data associated with each tag. The tags themselves are C strings, which allows for a wide variety of things the system will recognize as legal tags.

Capabilities and Limitations

CODA is a menu-driven system because it was our feeling that the typical non-computer-oriented researcher would find such a system the quickest and easiest to learn and the most comfortable to work with. The capabilities and limitations of CODA are best demonstrated directly through the menus that constitute the user interface. Below are the nine CODA menus arranged in a hierarchy based roughly on the connections between the menus.

**MAIN MENU**

- DATA ENTRY Options
- DATA RECALL Options
- HIT LIST Options
  - RECORD Options
  - CHANGE HIT TAG Options
  - REFINE HIT LIST Options
- TAG CHANGE Options
- TAG INDEX CHANGE Options

*Figure 1. Basic CODA Menu Structure*
Capabilities

In each menu, typing '?' will access a help file that describes the options in that menu, and typing '<ESC>' will get the user out of CODA entirely. The MAIN MENU has an introduction to CODA for the first time user, and is the main access path to the four menus in the first indentation in Figure 1. Of those four, two (TAG CHANGE and TAG INDEX CHANGE Options) relate to tag and index changes throughout the entire data base. The other two (DATA ENTRY and DATA RECALL) are the "guts" of the system, and, along with their submenus, they deserve more detailed mention in order to give the reader a good feel both for what CODA is and what it is not.

** DATA ENTRY Options **

a. Set up session tags  c. Transfer data (from file)  
b. Enter data (from keyboard)  d. Back to MAIN MENU (*)

Option?

Figure 2. Data Entry Menu

Figure 2 is the Data Entry menu, exactly as it appears in the bottom window of the user's CRT. Data entry from the keyboard is done in a full-screen two-window editor. An example of a record and its tags (as they have been entered in the two-window editor) is shown in Figure 3.

AGM-86 ALCM

The AGM-86 air-launched cruise missile is a small unmanned winged air vehicle capable of sustained subsonic flight following launch from a carrier aircraft. It has a turbofan engine and a nuclear warhead, and is programmed for precision attack on surface targets. When launched in large numbers, each of the missiles would have to be countered, making defense against them both costly and complicated. Additionally, by diluting defenses, the ability of manned aircraft to penetrate to major targets would be improved. Small radar signature and low-level flight capability enhance the missile's effectiveness. Production is expected to total 3,418 missiles between FY '80 and FY '87, with deliveries to be completed in FY '89. Initial funding for 225 AGM-86B ALCMs was provided in FY '80; 480 more were approved in FY '81, 440 in FY '82, and procurement of 330 is planned for FY '83. SAC's 416th Bombardment Wing at Griffiss AFB, N.Y., became the first Air Force unit to attain operational capability with ALCM in December 1982, with 12 missiles fitted externally to each of its 14 B-52Gs. It has been followed by the 379th Wing at Wurtsmith AFB, Mich. Other units to receive ALCMs are at Grand Forks AFB, Ark. Ultimately, each B-52G is intended to be modified to
have a bomb-bay rotary launcher

#General: weapon system; ACM-86; ALCM; nuclear; nonnuclear; guidance;
  inertial; TERCOM; speed; range; cruise missile; funding; air-to-surface; tech data
#Journal: Air Force Magazine
#Author: 
Date: 5/31/83

Figure 3. Data Entry Example in a Two-Window Editor

The first line of each record is printed out on the hit list in data recall, so it is commonly used as a summary line for the record. In the tag window, the indices are identified by '#' or ':' and the tags are separated by semi-colons. These are the only reserved symbols in CODA.

Session tags (option a. in Figure 2.) give the user the ability to set up tags to be put on all data items that are entered until the session tags are changed. This saves the user time in cases where several items to be entered will have tags in common. After being set up, the tags will automatically appear in the tag window in succeeding data entry calls. Session tags are also useful in data entry from non-CODA files. In order to enter data into CODA from a file, the file must have records separated by a single delimiter that doesn't appear elsewhere in the data: the records will be entered into the data base with the session tags set up for that purpose.

** DATA RECALL Options **

a. Enter tags for search  
b. Look at all system indices.  
c. Look at all tags (for an index)  
d. Back to MAIN MENU (*)

Figure 4. Data Recall Menu

The Data Recall menu, shown in Figure 4 as it appears in the bottom window of the user's screen, leads into the hypothetical heart of the system — data retrieval and manipulation. In any menu involving tags, the user has the option of looking at the current system indices and a glossary of the current tags. Searches can be done by tag, by full-text search, or by Boolean (AND, OR, and NOT) combinations of the two. After the search expression has been entered, CODA looks up the pieces.
of the expression in the tag hash table and offers to do a full-text search if it can't find them. It will also do a pure full-text search. A sample CODA response is shown in Figure 5 (in this case the special expression 'all' was entered and CODA returned all records).

There are 278 hits for this expression.
Expression: all

1. LGM-30F/G MINUTEMAN
2. MGM-118A PEACEKEEPER (MX)
3. AGM-69 SRAM
4. AGM-86 ALCM
5. BGM-109G GLCM
6. AGM-45A SHRIKE
7. AGM-65 MAVERICK
8. AGM-78 STANDARD ARM
9. AGM-88A HARM
10. GBU-15
11. AGM-109H TOMAHAWK (MRASM)
12. ALMV
13. AIM-120A (AMRAAM)
14. GBU-15s DESTROY SIMULATED MISSILE SITES
15. CRUISE MISSILE: WONDER WEAPON OR DUD?
16. BUSINESS OUTLOOK: CRUISE MISSILE PROGRAM
17. INGLORIOUS FAILURES PLAGUE PERSHING-2
18. MaRV: KEEPING A NEW NUCLEAR GENIE IN THE BOTTLE

*** HIT LIST Options ***
a. Look at specific record 
   e. Change tags on these hits
b. Output specific record(s) 
   f. Refine this hit list
c. Delete specific record(s) from hit list 
   g. Back to DATA RECALL (**)
d. Expunge specific record(s) 
   h. Back to MAIN MENU (*)
Option?

Figure 5. CODA Hit List Example

In some sense, the posited utility of CODA should be measured by the percentage of time the user spends in and around the Hit List menu. It is here that one's ability to rearrange and shape one's "long term memory" is most evident. It is here that we thought the researcher would spend time accessing, modifying and re-marking data in light of new information or insights. Three of the Hit List processing options (a, e, and f) lead to separate menus. The others lead to further prompts from CODA and will be described first.
The records in any subset of hit list records can be output (option b) in their entirety or to the first blank line (allowing for summary outputs), to a file or directly to the printer, with or without the associated tags. Any subset of the hit list can be deleted from the list (option c) as a way of shaping the list for output or tagging, or can be expunged from the data base entirely (option d). Paging through the hit list can be done with the +page and −page keys to be found on most terminal keyboards.

Any record can be seen in its entirety (option a) by entering its numerical identifier. This leads to a separate menu (Figure 6). In additional to mimicking the output, delete and expunge options of the Hit List menu, the user can edit the record and its tags (as in data entry), page through the record (if it is greater than one screen-full) with the +page and −page keys, or page through the full records on the hit list and its options.

<table>
<thead>
<tr>
<th>**** RECORD Options ****</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Edit this record (and tags)</td>
</tr>
<tr>
<td>b. Go forward to next hit list record</td>
</tr>
<tr>
<td>c. Go back to previous hit list record</td>
</tr>
<tr>
<td>d. Output this record</td>
</tr>
<tr>
<td>e. Delete this record from hit list</td>
</tr>
<tr>
<td>f. Expunge this record</td>
</tr>
<tr>
<td>g. Back to HIT LIST Options (***))</td>
</tr>
<tr>
<td>h. Back to DATA RECALL OPTIONS (***))</td>
</tr>
</tbody>
</table>

Figure 6. Record Options Menu

Option e on the Hit List menu (Figure 5) allows the user to make tag changes on all records in the hit list simultaneously. This leads to the menu shown in Figure 7. Adding, deleting, or renaming a tag takes place online and is reflected thereafter in any functions that involve the tags.

<table>
<thead>
<tr>
<th>**** CHANGE HIT TAG Options ****</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Add a tag to these hits</td>
</tr>
<tr>
<td>b. Delete a tag from these hits</td>
</tr>
<tr>
<td>c. Rename a tag on these hits</td>
</tr>
<tr>
<td>d. Look at all system indices</td>
</tr>
<tr>
<td>e. Look at all tags (for an index)</td>
</tr>
<tr>
<td>f. Back to HIT LIST Options (***))</td>
</tr>
</tbody>
</table>

Figure 7. Hit Tag Change Menu

Summer 1987
In addition to shaping the hit list by deleting individual hits, option f in the Hit List menu (Figure 5) allows the user to refine the hit list using the Boolean operators AND, OR and NOT. This leads to the menu shown in Figure 8.

**** REFINE HIT LIST Options ****

a. Current hits OR those with...
b. Current hits BUT ONLY those with...
c. Current hits BUT NOT those with...
Option?
d. Look at all system indices
e. Look at all tags (for an index)
f. Back to HIT LIST Options (***)

Figure 8. Hit List Refinement Menu

One of the major purposes of options a through c is to allow users unfamiliar with Boolean operations to use them nonetheless, by having them translated into "natural" English. Option a corresponds to the Boolean OR, option b to AND, and option c to AND NOT. With these three, the user familiar with logic can build up any desired Boolean refinement and the novice should, with successive refinements if necessary, be able to do the same thing. With each option, the user enters a search expression (as in data recall). CODA then performs the appropriate operation and gives the user back the refined hit list, the refined search expression and the Hit List menu.

While CODA has others, those major capabilities listed above are those most relevant to the hypothesis that we hoped to test with the system. In addition to the capabilities mentioned, there are limitations worth mentioning also. Some are a direct result of the design chosen to meet the desiderata and others have more subtle origins.

Limitations

Perhaps the most serious philosophical limitation on the system is that, while it is "optimized" for an individual policy researcher, its assets become increasing liabilities as the number of researchers using the same data base increases. This phenomenon is well known to large data base systems with a large number of users. The more users there are, the more important it becomes to limit both the number of people that can make changes to the system retrieval parameters as well as the frequency with which any changes can be made. CODA is specifically directed at, and can be used profitably by, only a very small number of users PER DATA BASE.
Although not necessarily a limitation, it is true that CODA is not a very "smart" system. Again, it was the intent of the design to leave the "smarts" in this man-machine system in the "man" since that is what "he" does best. The machine has been directed to do what it does well — store and retrieve data and make changes to the data on command.

The most noticeable limitation to the user is that the system can take a long time to load. As currently implemented, the hash table is built each time CODA is operated. Hence, the larger the data base and the heavier the machine processing load, the longer the system takes to load. With a 2000 record data base and several tags for each record, the loading can take several minutes on a heavily loaded machine. This is a "start-up" cost and is mainly due to the time required to link the data to their respective tags. While this could be done more efficiently, doing it in C will not be as efficient as it would be in a list processing language such as LISP.

In addition to loading torpidity, the hash table implementation leads to fairly large RAM requirements and to some inefficiencies in truncated searches. While a virtual machine is not overly sensitive to the RAM requirements of any program, the current CODA system, set to handle 10,000 records and a hash table set up of 10,000 tags, takes up about 700K in the VAX 11/780. Again, some storage requirements could be gained and efficiency sacrificed by putting the hash table on disk.

Another limitation from the hash table approach is that truncated searches (which aren't currently coded) can't be efficiently coded. An example of a truncated search is to look for all records that contain any work starting with 'bomb'. This would retrieve records with the words 'bomb', 'bomber', 'bombing', 'bombardment', 'bombast', 'bombazine', etc. In full-text search mode, CODA can do this quite well, but very slowly. In indexed tag searches, however, it requires searching through the entire hash table, which defeats the purpose of the hash table approach and takes decidedly longer than a single tag search. This is really only a limitation on the tag searches, and tends to be a problem with most other approaches that are geared for speed.

The above capabilities and limitations are basically the ones that were designed into the system. They are basically the things that can be said about the system AS SOFTWARE. It is important to reflect upon them in terms of trying to understand the tool that was developed to test our original hypothesis. In order to determine if this tool, CODA, was indeed of significant use of policy researchers, however, we went to the researchers, had them use it and asked them to give us their opinions on it. In the following section we tell...

What We've Learned

First, it is important to understand that CODA has basically been a "hobby" written on a shoestring budget. These constraints show up in the CODA code as a paucity of "gorilla proofing" and shortage of pre-release testing. They show up in testing as a lack of a rigorous hypothesis-testing methodology and a small number of "beta test sites." What we have learned about the system comes from the generous assistance of seven Rand researchers (including one of the authors [Dewar...
-- from whose work the examples in this paper have been drawn) and a total of nine data bases of various types. As examples, one data base contained data on long range non-nuclear weapons (about 300 records), another contained interview data from Russian emigres (about 2400 records), a third contained data on terrorist incidents world-wide (about 1800 records), yet another contained information on reviews for the Rand Journal of Economics (about 400 records), and one contained information on current data bases in the Rand data base library (about 100 records).

Perhaps the most interesting finding for us concerned the size of the data records. The system seemed most useful on records that were small. The two largest data bases had existed previously and were already organized as "bite-sized" data records with a small number of tags. In CODA, the number of tags on these data bases grew slowly as the data were retrieved, sorted and tagged in different ways, but there was no movement toward combining records. In contrast, in one of the smaller data bases that grew with time, there was also a slow move toward splitting large records into smaller, reasonably disjoint records or compacting large records by summarizing them. Records that were a screen-full or less in size were easiest to scan and absorb quickly.

CODA, as implemented, does indeed appear to be cumbersome for large data bases (which in this case should be taken to mean 1000 records or more). Loading time for the two largest data bases on a loaded VAX 11/780 could be several minutes; looking at the glossary of tags was cumbersome (particularly on the terrorist data where practically every incident had its own unique data identifier); and hit lists tended to be long and cumbersome to wade through while sitting at the terminal. Somewhat surprisingly, however, the users with large data bases have been happiest with CODA. Our guess is that this says much more about the utility of data base and file management systems in general than it does about CODA in particular.

The main "choke point" in the CODA process is definitely data entry and tagging. The ability to enter data from files was the only thing that made test with the largest data bases possible, but it is a slow and painful process in general to enter data into the system and to tag it. It was something of a surprise to find that it appeared to be more satisfactory to have a secretary put several records into the systems at a time (with a tag such as "untagged") and then to add tags to them by doing full-text searches (on the entire system), than it was to tag the records one at a time.

Originally, non-data indices were a way of grouping tags FOR DISPLAY PURPOSES ONLY. They were intended to function in much the same way that the Author index in the back of a book does. We found this to be of questionable use, and found increasingly that it would be nicer to have the ability to specify at times not only the tag, but also the index with which it was associated. That capability is now implemented, and, in our ongoing tests, is being evaluated.

Finally, we arrived at some very unscientific estimates of the point at which the CODA system stops being a corroboration of one's own memory and starts to function as a long term memory for things that have been forgotten. There seem to be two kinds of threshold -- one in terms of the number of data records in the system, and the other in terms of the time that has elapsed since the first record was entered into the system. According to our informal survey, it took one to two hundred records in the system before the first records had receded far enough in a researcher's memory that they appeared fresh when recalled. For data bases smaller than that, it took a few months of elapsed time before the computer's memory was clearly superior to the researcher's. These estimates do not claim to be scientific, but they do illustrate the "delay" a researcher can expect before a
system like this can be expected to begin paying noticeable dividends.

In using the prototype of the CODA system and getting feedback from others doing the same thing with different kinds of data bases, one not only learns things about the system as it is, but one also starts to get a feel for the general class of improvements that would enhance or establish its utility. It is to this set of reasonable (and not so reasonable) future system possibilities that we now turn.

Wish List

There were three general capability enhancements that occurred to us during the testing and, in decreasing order of practicability, they can be identified as 1) a bibliographic formatting capability, 2) thesaurus capability, and 3) optical data entry.

In at least two of the data bases there were a large number of direct quotes. While it is possible to put enough tags on these records to enable one to construct a bibliographic reference, there are computer programs available that will construct an appropriately formatted bibliographic reference. Data entry in these systems is of the "fill-in-the-template" form, and CODA allow one to set up user templates. While significant work would be required to build a bibliographic formatting capability for CODA, the effort would appear to be worthwhile in future systems of this type.

The second area of improvement that seemed reasonable from our test experience was to build some type of thesaurus capability into CODA. The desirability of a thesaurus that could provide "synonyms" for a given tag seems to increase as the size of the tag glossary increases. This would be a slight departure from the philosophical notion that it is best to rely on the researcher for the intelligence in this kind of man-machine system. Nonetheless, giving a CODA-like system some capability to remember or to look for similarities among tags might be a useful "advisory" capability. The details of such a thesaurus capability must remain vague at this point, but some notions of such a capability can be described. One possibility would be to "wire in" a thesaurus, in which case the researcher would be responsible for creating and maintaining the thesaurus and CODA would only respond with synonyms upon request. Another possibility would be to "teach" CODA concepts of similarity and have it constantly review the tag glossary and, on request, suggest synonyms to the user.

The most desirable improvement comes directly from numerous confrontations with the data entry bottleneck. The most obvious data entry mechanism would be an optical character reader about the size of a light pen that one could use much the way one uses a highlighting marker to mark for recall passages in a text. Entering them directly into CODA in this manner would be much more satisfactory than current data entry methods. While such a capability is an easier technological problem to solve than the more obvious voice entry capability, the feasibility of such a hand-held mechanism is still, sadly, beyond the current state-of-the-art. In fact, ANY better mechanism for entering data would measurable improve the utility of systems like CODA.
But back to reality. There are some things to be said for the utility of CODA as it is currently constituted and they are the topic of the final section of this manuscript.

Conclusions

Our original ponderings about computer-aided policy research led us down a somewhat tortuous path to the development of CODA. The prototype system was designed to test our hypothesis that computer-aided policy research could be improved, and to determine if our specific set of desiderata for such a system was a path to such improvement. CODA was built roughly to specifications implied by our desiderata, we enlisted Rand researchers to use and comment on it, and we have learned from the process.

As to whether a CODA-like system serves a useful purpose in the community of data base users, our work only leads us to suggestive conclusions. The eagerness of our volunteers to use CODA for more traditional file management purposes and for larger-than-we-envisioned data bases suggested the ever-growing recognition of the utility of computer-aided data management. This, in turn, suggests that specialized communities such as the policy research community are still awakening to the possibilities of computer-aided data management in a variety of forms.

Among the researchers who used CODA as we intended, there was a growing appreciation for and dependence on CODA's capabilities. This growing appreciation corroborates the earlier note that there is an inevitable time lag between the beginnings of a CODA data base and the appreciation of its long term memory utility.

Several of the desiderata built into CODA appeared indeed to have recognizable utility in the policy research world. The most controversial of these was the on-line ability to change tags simultaneously on large subsets of data. As mentioned earlier, in the general data management world this capability has serious drawbacks. In the community of data bases that have a small number of users, however, this becomes a very powerful tool for reforging the long term memory to conform to the current concerns and theses of the users.

While it is a very subjective judgement at best, search retrieval under one second are a definite improvement over systems with turnaround times only slightly longer. This appears to be much akin to satellite telephone conversations in which a one-second delay in conversational responses is distractingly noticeable.

The ability to enter data into CODA electronically from a file was very useful in transferring extant data bases into the CODA system. In addition, this ability led to some serious musings on the integration of CODA-like capabilities with larger data base management systems that have on-line capability for retrieval from very large data bases.

By way of improving the policy research process, the one currently feasible desideraum the CODA prototype seemed to lack was a bibliographic formatting capability. With the addition of this item to
the list of specifications, the general design goals of a useful policy research computer tool would seem to be complete.

In summary, our work with CODA leads us to believe that there is room for improvement in the area of computerized data management aids specifically designed for the policy research and related communities of users. While CODA may not be the optimal realization of that goal, the desiderata that led to its creation, along with the addition of a bibliographic formatting capability, form an excellent foundation upon which to build such a system. □
Tabulations on the DDA study description

by Karsten Boye Rasmussen
Danish Data Archives
Odense University

The object of this paper is to give a brief introduction to the Standard Study Description and to add a few remarks on its recent history and development. For those already familiar with it, the first section will also answer the question: 'Whatever became of the ACCESS project?' The main part of the paper concentrates on a presentation of the holdings at the Danish Data Archives (DDA) in the form of cross-tabulations based on a data file compiled from the contents of the study descriptions at the DDA.

The Study Description

For those not familiar with the term 'Standard Study Description', allow me to clarify: the phrase is meant to be read backwards.

First of all, the Standard Study Description is a 'description'. It is a machine-readable document written in a specific format. It is like a library catalogue card, except that the object is not a book but a machine-readable data file, or, to take another step backwards, a study. The data file or study is, for the purposes of this paper, within the broad area of the social sciences.

The format of the Standard Study Description (SSD) is somewhat complex. The SSD contains a large number of items, which can be compared to variables. Each item consists of a numeric identification code and an entry containing specific information. What makes the format complex is that the different entries may contain different types of information. At the IFDO/IASSIST conference in Grenoble in 1981, I presented a working paper\(^1\) which extensively expounded the format of the SSD. In this paper, a few examples should suffice to illustrate the complexity of the SSD. For example:

1. Item 101 contains the title of the study. It is an unstructured text item.

2. Item 212:04 contains the unweighted number of cases in the data file. A numeric subitem (:04) inside a structured item.

3. Item 222 describes the target population using predefined codes. A precoded item.

---


Thus the study description contains text items, numeric items and precoded items; of these, the precoded item is the most complex.

In precoded item 222, an entry "01" will signify that the target population is restricted by "age limits". There may be further restrictions specified by further codes. When a precoded item is viewed as a variable, what we have is indeed complicated (like a multipunch column on an old punch card).

To complicate things further, the SSD format allows a text explanation of unlimited size to follow any type of item, whether it be a text, numeric, or precoded item.

The SSD is like a data file: it is incomprehensible without the proper documentation describing the significance of the item numbers as well as a 'codebook' describing the codes used in the SSD. This documentation is, of course, machine-readable as well. With a computer program, one can merge the SSD data file and the codebook information to produce a human-readable printout describing the study.

Working backwards we have now reached the last word in the term 'Standard Study Description'. The SSD has not yet been accepted as an international standard. Given that this year is the 12th anniversary of the SSD, I shall not insist on the word "standard". Rather, in the remainder of this paper, I shall refer to the 'Study Description' or the SD. On the other hand, it is a standard of sorts, in so far that some of the European social science data archives use the SD extensively, with only minor differences.

Those interested in the true story of the Standard Description are advised to read the paper "Standard Study Description as a meta research data base" given by Per Nielsen at the 1983 IASSIST conference. That paper outlines the development of the SSD and includes references to historical papers on the subject. This paper is an updated version of Per Nielsen's paper.

The Function of the SD

From the beginning, the object of the SD has been to fulfil several functions as a tool for the data archives and the social science community. Some of the functions have changed, primarily because new technology has facilitated the achievement of new goals.

1. Data abstracting and catalogue production

The main function of the SD is (as described above) to produce human-readable printouts of the study descriptions. This function is what was originally termed "data abstracting and catalogue production". Since 1978, the DDA catalogues have been produced using computer programs to generate prototype setting instructions from the SDs and a "skeleton file". At the same time, a considerable amount of indexing of the descriptions is done automatically. The same procedure has been used at the Zentralarchiv in Cologne, at the Steinmetzarchief in Amsterdam, and is presently being used at the ESRC Data Archive in Essex. It should be noted that other data organizations use similar techniques. They are

---

1 Per Nielsen: "Standard Study Description as a meta research data base". Paper presented at
not mentioned here because they do not use the SD as the basis for their catalogue production.) At the DDA we are now completing the 1985 catalogue of holdings. Only a subset of items from the study descriptions is being printed, but a number of COMfile containing the complete descriptions will be supplied with each copy of the catalogue.

The production of catalogues has many drawbacks. First of all, it is very expensive. Of course all the SDs are available, as they are produced as part of the documentation process when data are deposited in the archive. Nonetheless, a lot of proof reading is necessary before the catalogue is ready for printing. Another drawback is the problem that the catalogue is outdated before it even reaches the market. Thus, online access to a computerized catalogue is necessary for up-to-date information.

2. Mapping and Methodological Research Base

The collection of SDs has long been regarded as the perfect object for methodological research or presentation of holdings. But the perfection resides in the very detailed information they contain, and not the process of compiling the information from the SD format to a rectangular data matrix ready for analysis. Because of the complex format of the SDs, such compilation requires some computing, to produce a rectangular numeric file without text information. This paper includes a chapter presenting the holdings at DDA with tables based on such a rectangular file. All the programming, including the extraction of information, was done with the software package SAS.

3. Data (Re)Analysis Prerequisite

At the DDA, the documentation of studies deposited in the archive includes a machine-readable codebook with complete questionnaire text as well as coding instructions plus one-way distributions of each coded variable. The codebook describes the data at the variable level. It also includes the total Study Description describing the background, objectives and outcome (publications) of the data collection. Thus the SD is a prerequisite for the process of secondary analysis.

4. Intra-Archival Loggin

This function, which was to supply the archives with a tool for keeping track of the processing of their data, has completely lost significance in the technological race of the last decade. The development of interactive data bases with immediate updating facilities, as opposed to a sequential and both time- and cost-consuming method, has, at least at the DDA, led the archive to implement data base applications which have the power to keep the most important information ready at hand ("just a PF-key away").

5. Inter-Archival Exchange

However the SD is still a standard for inter-archival exchange. It is an exchange format which is simple enough to be read into any machine (including microcomputers). Special software is needed, however, to process the SDs.
It is therefore my impression that the SD will remain a standard format for exchange purposes, and that its list of items will be a check-list of the kinds of information which should be provided for each study. The actual storage mode of the SDs at individual archives may differ, depending on what data base facilities are available, but the data base application should be able to both 'export' and 'import' SDs in the standard format.

Information Retrieval Data Base

The archives in Cologne, Amsterdam and Odense have developed retrieval systems for searching their own SDs. The most comprehensive system is the ZAR system at ZA in Cologne, which has been described earlier in IASSIST surroundings. The data archive at Essex has recently announced that they are setting up an online information retrieval system. Other archives have information retrieval systems as well, but the archives mentioned above are all using the SDs.

The ACCESS Project

The idea of using the SD-format as an exchange format as well as the possibility of setting up retrieval systems on the basis of the SDs, led to a project amongst the European archives within CESSDA (Committee of European Social Science Data Archives). Under the project heading "ACCESS: Integrated European Archive Inventory" a catalogue was to be published, SDs to be exchanged, and a data base retrieval system to be set up with common access (available through EURONET). The EEC was to finance the project, but the demands of the bureaucrats in Brussels made the project much less attractive, and it was finally abandoned.

The project has instead developed into an ongoing effort at the four archives (mainly the DDA). But due to a lack of funding, this project competes with regular activities of the archives and has therefore often been postponed. The tables in this paper are based on the collections of a single archive, the DDA. It is my hope that within a reasonable time period I shall be able to present similar tables comparing the holdings of the four European archives. At present the DDA has received SDs from the Steinmetzarchief; with the completion of the ESRC Data Archive catalogue, the DDA will receive a new batch of SDs, and finally the SDs from ZA.

Setting up a retrieval data base as described in the ACCESS project will demand a considerable amount of work. At present, network facilities are still not sufficiently effective to supply online access to other computers. When these techniques have been improved, the online Integrated European Archive Inventory will become a reality.

As mentioned above, the four archives use slightly different formats for the SD. As long as the differences are fully documented, this presents only minor problems. The ZA and DDA formats are very similar, although the ZA does not use as many items as the DDA. At the Steinmetzarchief, the numbering of the items is different, but the mapping of the formats is the same. At the ESRC-DA, depositors, etc. are identified by a number from a special file. At the DDA, the latest change in the SD format has introduced an item (220) pertaining to historical data materials, which shows the time period covered by the data, and an item (225) to show to what regional area or country the data describe.

---

Tabulations on the DDA SDs

The tables in the remainder of this paper describe 962 studies. These tables will not be extensively commented upon nor compared with the findings of Per Nielsen in 1983\(^7\) as the purpose of this section is not primarily to comment on the development of the data holdings of the DDA since the 645 datasets were investigated in 1983. The tables refer to the datasets described in the forthcoming DDA catalogue and give an overview of the contents of that catalogue.

There has been no exclusion of datasets with missing data. For some items, missing data should not exist, but for others the existence of missing data is perfectly alright.

Some of the variables are coded with multiple codes. To illustrate this, some of the tables have information on the number of codes. If, in a precoded item, a study has more than one code, the entries are weighted accordingly (e.g. 2 codes, weight = 0.5, 3 codes, weight = 0.33 etc.). Therefore even in these multiple response items, all the tables still total 962 studies (apart from rounding error).

The mean of the 962 studies is approximately 160 lines of information. The comparison between an SD and a library catalogue card is therefore not a very good one. The SDs at the DDA are more like a library catalogue card plus a very elaborate abstract of the study. The magnitude of information shows the potential for making a retrieval data base with the SDs as input data.

In an ongoing archival process, not all studies have optimal documentation; many studies are not yet fully processed. A few remarks about the DDA level of documentation may be useful here. At the DDA, two categories are of special interest to users. Studies in class "D" contain a complete machine-readable codebook. From this documentation, we generate setups in SPSS- or SAS-format for the user as well as deliver published documentation on the study. Studies placed in class "C" do have some machine-readable documentation, but are not as "polished" as studies in class "D", and do not contain a machine-readable codebook.

The Study Description item '001' contains the status of the study. However, at the DDA, this item is updated by extracting information from our data base which keeps track of the processing of studies. As this process had not taken place at the time that I computed these statistics, I have computed table 2 directly from the processing data base for the purpose of showing the status of the studies.

This total differs from the number of studies drawn from the SDs. This difference is due to the fact that all the other statistics are made on the basis of the SDs to be published in the catalogue of holdings. Since the cut-off of new additions to the catalogue, 141 studies have come to our attention; these studies are typically placed in class L.

\(^7\)op.cit.

\(^1\) (tables have been collected together at the end of the article. Ed's note)
Description of the Studies

This section will illustrate the kinds of studies available from the DDA.

Table 3 shows the subject headings or areas covered by the DDA holdings.

The table shows a heavy bias towards the traditional areas of the social sciences. Election studies, general sociology and political science together constitute more than 60 percent of the studies.

A similar picture is displayed when looking at the kind of data on which studies are based. As shown in Table 4, 83 percent of the studies consist of survey data.

Earlier, approximately 76 percent of all studies had been generated by the old-fashioned oral interview. Since then, there have been an increasing number of mail surveys. It is possible that the rising cost of conducting the traditional personal interview has led investigators to use mail surveys. Furthermore, it is my impression that many surveys are now carried out by telephone, but this is not yet reflected in the distribution of these data.

Typically, the time span between data collection and the deposition of data in the archive is between 3 and 4 years. (Table 5).

The studies stored in the DDA also appear to be conservative with respect to the cases on which the data are based (the "units of observation"). Close to 90 percent of all studies are based on individuals. (Table 6).

In Table 7 the definition of the universe is shown. A central variable is age. Most election surveys concentrate on those over 18 (which is the age at which one obtains the right to vote in Denmark). Please note, that this table indicated a frequency total of 1303. Of the 831 studies described in this item, many have more than one code specified. Of interest are also those categories which are lacking in the table below. In the SD 'skeleton' file, provision is made for definition of the universe in terms of 'race' or 'religion'. These do not occur in any of the DDA studies.

When we consider the applied sampling procedures, we find that most surveys fall into one of two categories. Those coded 'no sampling' are typically studies carried out at a distinct location (e.g. a working environment). The other major category is the many studies (25 percent) which are based on some kind of multi-stage sampling. (Table 8).

Without performing detailed cross-tabulations, it is nonetheless easy to outline the characteristics of a typical study in the DDA collection: it would seem to be an election study, carried out as a multi-stage sample survey, with individuals as participants and as units of analysis.

The DDA Data Bank As Potential For Analysis

It is interesting to note that, according to the frequencies on Item 211, approximately 23 percent of the studies are panel studies. This high percentage is due to an agreement between DDA and the public opinion and marketing bureau Observa A/S to the effect that all their political panels from 1967 to the present are being stored in the DDA6. Apart from the Observa studies, the remainder of the panel studies are typically election studies also.

It is also worth noting, in Table 9, that the largest number of studies are in the category of cross-sectional sectional studies with replication.

6The OBSERVA project was described by Karsten Boye Rasmussen and Lone Borgersen in DDA-nyt nr. 34, 1985.
of one form or another. These, together with the large number of panel studies, comprise a majority of the studies at the DDA which are connected as part of a panel study or having variables which are very similar, and which therefore present great potential for secondary analysis.

The European archives have often tried to identify studies carried out within the same period of time in a number of the European countries. At the DDA we have introduced a new Item (225) to indicate area of coverage.

Table 10 shows that the vast majority of the studies at the DDA are national and therefore cover all Denmark. About 40 of the studies, however, are cross–national. These studies are typically the Euro–barometers conducted for the European Economic Commission, but within the last few years the DDA has identified some other cross–national studies. It should be noted here that the DDA does not publish information about studies also stored in other data archives, unless they contain information on Danish matters.

To obtain access to a dataset deposited in the DDA, the user is asked to fill out a requisition form and send a one–page description of how the data are to be analyzed. This information is then sent to the depositor or other person authorized to permit access. Most of the studies (65 percent) are without any access restrictions for the typical user working in the social sciences. Surprisingly, a large number of studies are categorized as being available only by special arrangements with the access–granting authority. At the DDA, we prefer not to store studies that are not available to users. The studies in this category may possibly be those of which the principal investigators have not yet finished analysis, but it might be worth rechecking the access conditions of studies in this category. (Table 11).

---

**Time of Study and the Data Matrix**

In this section we investigate the hard facts of some numeric items from the studies placed at the DDA. A discussion of whether this material may in any way be regarded as being representative of social science data in Denmark can be found at the end of this paper. Tables 12–14 are believed to show nothing but the distribution of the studies placed at the DDA.

One of the key variables is the starting year of the time period covered by each study (Item 220:01). This new variable is not be confused with Item 231:01, the data collection date. For surveys, these two dates will of course be identical, but for historical studies based on old documents (e.g. parish registers or census lists) Item 220 is indispensible. Subtracting Item 220:01 from Item 220:02 (end year) shows that 75 percent of the studies start and end in the same year. On the other hand, 15 studies cover a time period of more than 100 years.

In Table 12 the start year is cross–tabulated with a grouped variable containing the number of cases in each data file.

Although the DDA was founded in 1973, more than 20 percent of the studies in the archive deal with a time period previous to that year. Of the 62 studies covering the period before 1950, 42 studies are concerned with a time period before 1900. These historical studies have typically a large number of cases, but are problematical in that they are not easily sampled. The cases are inter–related (i.e. family reconstitution data) and therefore all cases must appear in the data material so that the relationships can be determined by computer.

Item 212:01, the number of cases, is missing in approximately 28 percent of the studies. Most of these studies are new. The number of cases is missing because for many of these studies...
only a limited description is as yet available; DDA has as yet received neither the data file nor precise information about the file dimensions yet. The typical data file has between 800 and 2500 cases.

These newer, partially documented studies are also a major portion of the great number of missing cases in table 13 below showing the number of variables in the datasets.

The studies dealing with the time period before 1950 (the 'historical' studies) contain, as expected, a low number of variables. The information concerning the historical cases is not very full, but the number of cases is - as shown in the previous table - often huge. In the course of the last 25 years there seems to have been a rise in the number of variables in a single study. The average number of variables is approximately 160 variables per dataset.

Table 13 shows the cross-tabulation of number of cases by number of variables. As has already been mentioned, there seems to be a weak reverse relationship. On the one hand, the datasets with few variables have a large number of cases. On the other hand, the datasets with more than 40 variables are concentrated around the 801 to 2500 case size.

The collection of descriptions of Danish social science studies provides an opportunity to examine this collection as a sample of Danish social science empirical research. But is it possible to draw valid conclusions from this sample? I do not intend in this paper to present a solution to this problem. But I should like to discuss some of the problems of bias that must be discussed before any conclusions can be drawn from the collection of SDs. It is my intention, in raising these problems, to stimulate a discussion which will be of benefit to the future comparative analysis of the characteristics of the data holdings of the other European archives.

It is indeed questionable if the collection of studies in the DDA is representative of Danish social science research.

First of all, the studies represented consist of data collections or empirical studies. Secondly the studies must be machine-readable, which will normally mean that computer analysis has been performed on the data. The target of the analysis will therefore at least be limited to Danish machine-readable empirical social science studies.

The most serious threat to the validity of the analysis is whether or not there has been a change in the DDA's criteria for incorporating a study into the data archive. Given the limited resources available at the data archive, it is to be expected that over the years some changes in the basic criteria may have taken place. Furthermore, it is to be expected that such a 'drift' in criteria may have happened unobserved and without being part of explicit archival policy.

One way to prove or disprove this hypothesis would be to compare the holdings of the DDA with a complete inventory of Danish social science research. But the DDA catalogue of
holdings is the only available source which is close to being a complete inventory. Such a comparison would result in tautological nonsense. Instead, a few limited areas could be compared. The DDA has deposition agreements with some research institutions as well as with the Danish Social Science Research Council. Thus the DDA could check whether these agreements are being fulfilled, or if some studies are for any unknown reason not brought to its attention.

Until there has been a further, more thorough investigation into the representativeness of the studies placed in the data archive, the tables above cannot be construed to be representative of social science research. On the other hand, even if the representativeness of the sample is not tested, we can argue as follows:

Bias is inherent in the selection of social science data files for archival storage. One major source of bias is technical. A simple 'rectangular' survey file is the archetypic data file in data archives. These files are practically ready for storage on receipt by the archive, while a hierarchical study demands more data processing by the archive. Both types of studies, of course, need the production of the proper machine-readable documentation.

The other major source of bias in the selection of social science data for archiving lies in the nationalistic characteristics of the social sciences. Because of the similarity in standards and technical capabilities of the four European archives (in Germany, Great Britain, the Netherlands and Denmark), the technical bias can be isolated.

Based on these assumptions, a table showing the differences in the holdings at the European archives may serve as a guideline for the actual differences in social science research being carried out in the four respective countries.

### Tables

#### "Number of lines in SDs"

<table>
<thead>
<tr>
<th>100000</th>
<th>frequency</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-50</td>
<td>56</td>
<td>5.8</td>
</tr>
<tr>
<td>51-100</td>
<td>395</td>
<td>41.1</td>
</tr>
<tr>
<td>101-200</td>
<td>458</td>
<td>47.6</td>
</tr>
<tr>
<td>201-300</td>
<td>44</td>
<td>4.6</td>
</tr>
<tr>
<td>301+</td>
<td>9</td>
<td>0.9</td>
</tr>
<tr>
<td>Total</td>
<td>962</td>
<td></td>
</tr>
</tbody>
</table>

**Table 1**

#### Class/Status frequency

| D: fully machine readable documentation | 285 |
| C: no codebook                          | 56  |
| B: available from primary investigator  | 72  |
| O: being processed/ongoing acquisition | 456 |
| L: only preliminary donor contracts    | 234 |
| Total                                  | 1103|

**Table 2**

#### "AREA"

<table>
<thead>
<tr>
<th>ITEM 002</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>2.0</td>
<td>0.2</td>
</tr>
<tr>
<td>organizational</td>
<td>61.5</td>
<td>6.4</td>
</tr>
<tr>
<td>general sociology</td>
<td>134.8</td>
<td>14.0</td>
</tr>
<tr>
<td>history, demography</td>
<td>53.1</td>
<td>5.5</td>
</tr>
<tr>
<td>law &amp; criminology</td>
<td>25.0</td>
<td>2.6</td>
</tr>
<tr>
<td>political science</td>
<td>91.1</td>
<td>9.5</td>
</tr>
<tr>
<td>social physics</td>
<td>33.7</td>
<td>3.5</td>
</tr>
<tr>
<td>social medicine</td>
<td>75.8</td>
<td>7.9</td>
</tr>
<tr>
<td>welfare &amp; leisure</td>
<td>48.7</td>
<td>5.1</td>
</tr>
<tr>
<td>socialization</td>
<td>41.7</td>
<td>4.3</td>
</tr>
<tr>
<td>election studies</td>
<td>363.2</td>
<td>37.8</td>
</tr>
<tr>
<td>macroeconomics</td>
<td>17.0</td>
<td>1.8</td>
</tr>
<tr>
<td>microeconomics</td>
<td>13.8</td>
<td>1.3</td>
</tr>
<tr>
<td>(Total # of codes = 1358)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3**
**Table 4**

### Kind of Data

<table>
<thead>
<tr>
<th>ITEM 202</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>9.0</td>
<td>0.9</td>
</tr>
<tr>
<td>survey</td>
<td>797.5</td>
<td>82.9</td>
</tr>
<tr>
<td>census data</td>
<td>9.1</td>
<td>1.0</td>
</tr>
<tr>
<td>statistics</td>
<td>52.0</td>
<td>5.4</td>
</tr>
<tr>
<td>legislative roll</td>
<td>4.0</td>
<td>0.4</td>
</tr>
<tr>
<td>clinical data</td>
<td>8.7</td>
<td>0.9</td>
</tr>
<tr>
<td>textual data</td>
<td>3.2</td>
<td>0.3</td>
</tr>
<tr>
<td>coded textual data</td>
<td>23.2</td>
<td>2.4</td>
</tr>
<tr>
<td>coded documents</td>
<td>55.3</td>
<td>5.8</td>
</tr>
</tbody>
</table>

(Total # of codes = 1042)

**Table 5**

### Method of Data Collection

<table>
<thead>
<tr>
<th>ITEM 232</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>100.0</td>
<td>10.4</td>
</tr>
<tr>
<td>oral interview</td>
<td>452.7</td>
<td>47.1</td>
</tr>
<tr>
<td>telephone survey</td>
<td>13.8</td>
<td>1.4</td>
</tr>
<tr>
<td>mail survey</td>
<td>317.5</td>
<td>33.0</td>
</tr>
<tr>
<td>pencil &amp; paper</td>
<td>73.5</td>
<td>7.6</td>
</tr>
<tr>
<td>psychological test</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>other</td>
<td>0.9</td>
<td>0.4</td>
</tr>
</tbody>
</table>

(Total # of codes = 1019)

**Table 6**

### Units of Observation

<table>
<thead>
<tr>
<th>ITEM 211</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>80.0</td>
<td>8.3</td>
</tr>
<tr>
<td>individuals</td>
<td>858.0</td>
<td>89.2</td>
</tr>
<tr>
<td>families/household</td>
<td>16.0</td>
<td>1.7</td>
</tr>
<tr>
<td>groups</td>
<td>5.5</td>
<td>0.6</td>
</tr>
<tr>
<td>other</td>
<td>2.5</td>
<td>0.3</td>
</tr>
</tbody>
</table>

(Total # of codes = 972)

**Table 7**

### Definition of Target Population

<table>
<thead>
<tr>
<th>ITEM 222</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>131.0</td>
<td>13.6</td>
</tr>
<tr>
<td>age limits</td>
<td>535.5</td>
<td>55.7</td>
</tr>
<tr>
<td>sex</td>
<td>5.7</td>
<td>0.6</td>
</tr>
<tr>
<td>marital status</td>
<td>1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>ethnic group, nationality 1.8</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>language characteristics</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>location of unit</td>
<td>69.7</td>
<td>7.2</td>
</tr>
<tr>
<td>housing conditions</td>
<td>5.7</td>
<td>0.6</td>
</tr>
<tr>
<td>position in family</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>occupation</td>
<td>47.1</td>
<td>4.9</td>
</tr>
<tr>
<td>education</td>
<td>17.0</td>
<td>1.8</td>
</tr>
<tr>
<td>physical conditions</td>
<td>8.2</td>
<td>0.8</td>
</tr>
<tr>
<td>mental conditions</td>
<td>3.2</td>
<td>0.3</td>
</tr>
<tr>
<td>time limits</td>
<td>116.4</td>
<td>12.1</td>
</tr>
<tr>
<td>other</td>
<td>19.0</td>
<td>2.0</td>
</tr>
</tbody>
</table>

(Total # of codes = 1303)

**Table 8**

### Sampling procedures

<table>
<thead>
<tr>
<th>ITEM 223</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>329.0</td>
<td>34.2</td>
</tr>
<tr>
<td>no sampling</td>
<td>178.4</td>
<td>18.3</td>
</tr>
<tr>
<td>quota sample</td>
<td>2.4</td>
<td>0.2</td>
</tr>
<tr>
<td>simple random number 91.4</td>
<td>36.2</td>
<td></td>
</tr>
<tr>
<td>stratified random sample</td>
<td>36.2</td>
<td></td>
</tr>
<tr>
<td>area-cluster sample</td>
<td>13.0</td>
<td>1.4</td>
</tr>
<tr>
<td>multi-stage sample</td>
<td>245.7</td>
<td>25.5</td>
</tr>
<tr>
<td>other</td>
<td>66.2</td>
<td>6.9</td>
</tr>
</tbody>
</table>

(Total # of codes = 1019)

**Table 9**

### Time Dimensions

<table>
<thead>
<tr>
<th>ITEM 221</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>59.0</td>
<td>6.1</td>
</tr>
<tr>
<td>cross-sectional</td>
<td>332.6</td>
<td>33.5</td>
</tr>
<tr>
<td>as above - with partial replication</td>
<td>350.0</td>
<td></td>
</tr>
<tr>
<td>panel study</td>
<td>218.6</td>
<td>22.7</td>
</tr>
<tr>
<td>trend study</td>
<td>10.0</td>
<td>1.0</td>
</tr>
<tr>
<td>other</td>
<td>1.0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

(Total # of codes = 974)
### Geographical Code

<table>
<thead>
<tr>
<th>ITEM</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>2</td>
<td>0.2</td>
</tr>
<tr>
<td>local</td>
<td>133.5</td>
<td>13.9</td>
</tr>
<tr>
<td>regional</td>
<td>34.5</td>
<td>3.6</td>
</tr>
<tr>
<td>national</td>
<td>750.5</td>
<td>78.0</td>
</tr>
<tr>
<td>cross-national</td>
<td>41.5</td>
<td>4.3</td>
</tr>
</tbody>
</table>

(Total # of codes = 985)

Table 10

### Accessibility

<table>
<thead>
<tr>
<th>ITEM</th>
<th>FREQUENCY</th>
<th>PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>38</td>
<td>4.0</td>
</tr>
<tr>
<td>no access restrictions</td>
<td>169</td>
<td>17.6</td>
</tr>
<tr>
<td>no restrictions to scientific usage</td>
<td>454</td>
<td>47.2</td>
</tr>
<tr>
<td>no publication without permission</td>
<td>13</td>
<td>1.4</td>
</tr>
<tr>
<td>no use of data without permission</td>
<td>149</td>
<td>15.4</td>
</tr>
<tr>
<td>available after special arrangement</td>
<td>136</td>
<td>14.1</td>
</tr>
<tr>
<td>other access conditions</td>
<td>3</td>
<td>0.4</td>
</tr>
</tbody>
</table>

(Total # of codes = 965)

Table 11

### Dimensions of Data - number of variables

<table>
<thead>
<tr>
<th>D212:02</th>
<th>(DIMENSIONS OF DATA number of variables)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D220:01</td>
<td>(TIME PERIOD begin year)</td>
</tr>
</tbody>
</table>

Frqncy missing before 1950-1960-1970-1980-

<table>
<thead>
<tr>
<th>sing</th>
<th>1950</th>
<th>1959</th>
<th>1969</th>
<th>1979</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>43</td>
<td>120</td>
</tr>
<tr>
<td>1-40</td>
<td>34</td>
<td>25</td>
<td>12</td>
<td>12</td>
<td>45</td>
</tr>
<tr>
<td>41-100</td>
<td>11</td>
<td>17</td>
<td>16</td>
<td>33</td>
<td>56</td>
</tr>
<tr>
<td>101-250</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>154</td>
</tr>
<tr>
<td>250+</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>69</td>
</tr>
</tbody>
</table>

Total 51 62 32 99 495 223 962

Table 13

### Dimensions of Data - number of cases

<table>
<thead>
<tr>
<th>D212:04</th>
<th>(DIMENSIONS OF DATA number of cases)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D213:02</td>
<td>(DIMENSIONS OF DATA number of variables)</td>
</tr>
</tbody>
</table>

Frqncy missing 1-40 41-100 101-250 250+

<table>
<thead>
<tr>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>missing</td>
</tr>
<tr>
<td>1-40</td>
</tr>
<tr>
<td>41-100</td>
</tr>
<tr>
<td>101-250</td>
</tr>
<tr>
<td>20000</td>
</tr>
<tr>
<td>20001+</td>
</tr>
</tbody>
</table>

Total 344 134 159 215 110 962

Table 14
Database directories

by Jim Jacobs

Introduction
The following is a highly selective list of directories of [American] machine-readable data files. Emphasis is on those directories which are most current, most complete, or are unique in some useful way. This list is current as of May 1986.

DIRECTORIES OF ONLINE DATABASES


Describes over 2800 publicly available databases, most of which are available online. Gives rather detailed descriptions of each database. A subject index lists databases in 550 categories. This directory can also be searched, full-text, on Dialog (file 230). Volume one covers databases in science, technology, and medicine; volume two covers business, law, the humanities, and social sciences; multi-disciplinary databases are listed in both volumes.


Identifies and describes machine-readable database, both bibliographic and non-bibliographic, which are available for public access online in North America. Lists fewer databases (about 1700 versus more than 2700) than Directory of Online Databases or Computer Readable Databases. Well indexed by subject, producer and vendor. Also available for searching full-text on BRS (database label: KIPD).


Lists and describes machine-readable databases available online to the public. The spring 1985 issue lists 2760 databases, only slightly fewer than Computer Readable Databases, which includes a few which are not available online. Published quarterly and available online on Westlaw. Good subject and other indexes.

1Editor's note

This directory is useful if you have the name of a particular periodical and you want to know if it is indexed or abstracted online, or available for full text searching online. Will cover 25,000 periodicals when all 3 volumes are published.

SPECIALIZED DIRECTORIES


Lists and provides profiles of members in this organization of data users, producers, and distributors. All members are organizations.


Describes holdings of machine-readable data in the National Archives. Arranged by the same record groups as the National Archives Guide. These files are not available online, but most can be purchased on tape. A new edition is due in 1986.


The Roper Center is an archive of sample survey data from over seventy countries. Data Acquisitions lists and describes new surveys in the archive. Currently there are over 900 studies available through the center in machine readable form. A newsletter, Data Set News Roper, announces new acquisitions.


Lists and describes over 1000 federal databases which are for sale from NTIS on computer tape. This catalog does not indicate online availability although some files may be available through vendors. Indexed by subject and agency.


Although dated, this directory is unique and is still valuable for identifying organizations that collect data files, and the types of files they collect. Many entries are for local data centers collecting locally produced data. Some examples: "Historical Data on the Social Welfare Policies in Europe", (1850-1965), "Polish Immigration in the U.S." (1776- ), "Urban Transportation Study, Amarillo Texas", (1940- ). Few, if any, of the data files listed here are available online.

Describes the Bureau’s holdings of machine-readable data and how they may be obtained. Also see: Census Catalog and Guide, an annual publication of the Bureau.


The intention of this directory was to list and describe all federal statistical data files which were either available to the public or available for processing by federal agencies at a user’s request. This was a "preliminary" edition and was not complete. No other edition has been prepared by the government, but Oryx Press has announced the intention of publishing an updated and expanded directory, Federal Statistical Data Bases: A Comprehensive Catalog of Current Machine-Readable and Online Files, in September 1986.


For each of the thirty-eight organizations of the United Nations, this directory lists and describes "information systems" (which includes over 600 machine-readable data files, as well as libraries, clearinghouses, information centers, etc.). Indexed by subject.


This directory of organizations that produce, access, and service machine-readable data is particularly useful for identifying otherwise elusive datafiles including files which are available only off-line. About 3300 organizations and 3600 databases are covered. Particularly useful indexes.


Identifies, describes and tells how to get access to over 3000 databases. Some are available directly online through the government, some are available only on computer tape and must be purchased from the government. Almost 300 of the databases listed are commercial databases which contain government produced information. Although there is no subject indexing, one section lists and describes government data files by broad subject.

Describes federal sources and information systems which are maintained by executive agencies and which contain fiscal, budgetary, and program-related data and information.


Guide to Resources and Services. Inter-University Consortium for Political and Social Research. (annual). Ann Arbor, MI: ICPSR.

This is the catalog of over one thousand data files maintained by the ICPSR. Data files maintained by ICPSR include surveys, censuses, election returns, legislative records, and other data on social and political attitudes and phenomena in over 130 countries. Data files are available for purchase. The catalog describes each file and gives ordering and codebook information. It is arranged by subject categories and includes a separate subject index. A newsletter supplements the catalog between editions. Also available online through ICPSR.

[Rutgers Inventory of Machine-Readable Texts in the Humanities.]

This inventory is not yet published but should be available by summer or fall of 1986. It will be based on an inventory project conducted at the Archibald Stevens Alexander Library at Rutgers. The inventory may also be available on RLIN and OCLC. (See: "Rutgers inventory of machine-readable texts in the humanities" by Marianne I. Gaunt in: International Conference on databases in the Humanities and Social Sciences, ed. by Robert F. Allen. Osprey, FL: Paradigm Press, 1985. pp. 283-290).

JOURNALS AND NEWSLETTERS

The following is a selected list of academic journals and other serials which feature articles on machine-readable information fairly regularly.

ACH Newsletter. Association for Computers and the Humanities.


Computers and the Humanities.

Computers and the Social Sciences.

International Conference on Computing and the Humanities.

Scholarly Communication Online Publishing and Education (SCOPE).

Social Science Microcomputer Review.
Contents of Current Journals

SSD kontakt
(Swedish Social Science Data Service)
3-4/1986

p. 1 Workshop on a Swedish database of regional time-series data. [in Swedish]
p. 4 Regional codes and individual-level data (microdata)./ Lennart Brantgarde [in Swedish]
p. 8 Swedish electoral data from 1911. [in Swedish]
p. 9 A cartographic program at SSD. [in Swedish]
p. 10 The meeting of the SSD user's group March 26th 1987. [in Swedish]
p. 12 Codebooks in the university libraries. [in Swedish]
p. 18 SSD disseminates data from all over the world. [in Swedish]
p. 21 News from other data archives. [in Swedish]
p. 23 New foreign data files in SSD. [in Swedish]
p. 24 Newly acquired and processed Swedish data files. [in Swedish]
p. 32 SSD class 1 data files. [in Swedish]
p. 33 English summary.

ESRC data archive bulletin
N.37 May 1987

News:

p. 1 "Looking back" by D.E. Allen
p. 2 Data for schools service
   Forthcoming workshop on FES

Summer 1987
The Domesday Project and beyond
p. 3 The IRON system
    Call for papers: Teaching quantitative data analysis a teleconference
    The 1991 census
    Rural areas database
p. 4 Regional research laboratories The general households survey
p. 5 The FES information pack

New acquisitions:

p. 5 A selection of new acquisitions
p. 5 Handicapped and impaired in Great Britain, 1968–1969 (OPCS)
p. 6 Disarmament negotiations between USA and USSR, 1986 (Steinmetzarchief)
p. 7 Domestic energy use data archive (Building research energy conservation support unit)
p. 7 The new earnings survey
p. 8 Health and lifestyle survey
p. 8 Revisions and updates to existing holdings
p. 9 Updates to serial holdings

Research organisations, data institutions and foreign archives:

p. 9 Danish data guide 1986
p. 9 Index to the Eurobarometers
p. 9 General social survey bibliography
p. 9 Oxford text archive
p. 9 The social research unit (Cardiff) working papers
p. 10 Advice on cross-national research

Software bulletin:

Notes:

p. 10 The measurement of social class
p. 10 Social research association
p. 11 "Searching and extracting data from the CSO macro economic time series"
p. 11 Statistical news
p. 11 Employment gazette
p. 11 Recent Department of Employment research papers
p. 12 ESRC newsletter
p. 12 Population trends
p. 12 Area
p. 12 The telephone and the voter
p. 12 Historical social research
p. 12 International social science journal
p. 13 Yuppies vs New Deal Democrats - ISR reports
p. 13 Computers and human interaction

Summer 1987
p. 13 European political data newsletter
p. 13 Social research association seminars
p. 13 Forthcoming events

**Books:**

p. 20 Human behaviour in geographic space, ed. by J. Paelinde (review by Roy Drewett)
p. 20 Social action and artificial intelligence, by Gilbert and Heath (review by R. Banks)

**Book notes:**

p. 21 Information systems education, by Buckingham et al.
p. 21 New methods and techniques for information managers, ed. by Mary Feeney
p. 21 Introduction to information science, by Flynn
p. 21 Analysis of panel data, by Hsiao
p. 22 Introduction to expert systems, by Jackson

**Appendices:**

p. 22 A: Recent additions to the ICPSR archive at Michigan
p. 23 B: Datasets acquired since the publication of Bulletin 36

---

**Bits & bytes review**

**vol. I(4), March 1987**

p. 1 The Apple Macintosh SE: the penultimate Macintosh./ John J. Hughes
p. 9 The Apple Macintosh II: the ultimate Macintosh./ John J. Hughes

---

**Government publications review**

**vol.14(3) 1987**

p.273 Dialling for documents: the distribution of electronic publications by the U.S. Department of Commerce./ Ann Bregent and Rushton Brandis
p.287 Information policy in an era of illiteracy: the U.S. Pension Bureau before the Civil War./ James W. Oberly
p.295 The availability and use of international documentation in Yugoslav libraries./ Zaneta Barsic
p.311 On site indexing of 1980 U.S. Census of Population and Housing./ Edward Herman
p.341 Diskette data from federal agencies./ Susan Anderson
p.347 The distribution of Food and Agriculture Organization publications to United States Land Grant Institution libraries: a research note./ Philip van de Voorde
p.351 Nicaraguan government document update./ Thomas Bloch
p.353 United States Census Bureau CD-ROM evaluation project: a news note./ LeRoy C. Schwarzkopf
p.355 Reviews./ Robert A. Walter
Books and other materials received./ Robert A. Walter
News from Washington./ LeRoy C. Schwarzkopf
International organizations news./ Robert W. Schaaf

Database
V. 10(2) April 1987

Feature articles:

p. 15 Database interviews Marydee Ojala, an information professional./ Jeffery K. Pemberton and Nancy Garman
p. 23 Searching the engineering databases./ Virginia N. Anderson
p. 29 Company rankings: whose top 20?/ Ruth A. Pagell and Michael Halperin
p. 36 The world of demographic data./ Diane Crispell
p. 44 Demographics on a laserdisk: Infomark./ John E. Rogozenski, Jr.
p. 47 Downloading and post-processing online numeric data./ Philip M. Clark
p. 56 Computer databases: a survey. Part 3: product databases./ Mick O'Leary
p. 66 Makedb: a low-cost alternative to Sci-Mate, Pro-Cite, et al. ...usable with free "generic" retrieval systems./ Clyde W. Grotophorst
p. 73 NewsNet in a corporate library./ Patricia Meyer
p. 83 First look: Supertech./ Jane Stolarz

Columns and special features:

p. 4 Letters to the editor./ a column by our readers
p. 5 The linear file - How microcomputing writers and editors are dBASEing the language of information professionals...a wordsmith's journey among the trendies and vendies./ Jeffery K. Pemberton
p. 11 SDI - the database news section./ June Thompson
p.100 The dollar sign - searching for general business/management information./ Marydee Ojala
p.107 Caduceus - NurseSearch: a nursing database on floppy disk./ Bonnie Snow
p.114 Source code - code translations: dBASE II, R:BASE 5000, dBASE III PLUS./ Paul W. Kittle
p.116 Database design - australia and america and database design./ Betty Eddison
p.119 The friendly user - Soviet information sources online./ Lucinda D. Conger
p.122 The silver disk - between a rock and a hard place: preservation and optical media./ Nancy Herther
Literary and linguistic computing
vol. 2(2), 1987

p. 61 Word-patterns and story-shapes: the statistical analysis of narrative style./ J.F. Burrows
p. 71 Greek syntax: a new approach./ R. Wonneberger
p. 80 Dialect analysis and automatic cartography by means of a microcomputer./ E.W. Schneider
p. 86 The computer and Sophocles Trachiniae./ E.M. Craik and D.H.A. Kaferly
p. 98 Uforanderlige and uforanderlighed: more about their differences./ S. Hogue and A. McKinnon
p.106 Sound and sense in the Divine Comedy./ D. Robey
p.116 Exploration of foreign languages speech synthesis./ M. Struit, G. Weston and D. Burkhardt
p.120 The value of the confidence interval of the consonant-vowel ratio as an indicator of the type
of linguistic material./ Y.A. Tambovtsev
p.125 The Oxford concordance program version 2./ S. Hockey and J. Martin
specialises./ R. Pellen and J. Pradines
p.141 Diary
p.142 News and notes
p.144 Documents received

European political data newsletter
nr. 63, June 1987

p. 4 Editorial note.

Data section:

Research in progress:

p. 5 Historical election-research at the University of Passau./ Stefan Immerfall
p. 8 The British election campaign study 1987./ David Broughton et al.
缇. 10 The international social survey program.

Archive news:

缇. 14 Austrian data archive.
缇. 15 Data development for international research funded.
缇. 16 Euro-barometer key-word documentation.

Books:
p. 19 Centre–periphery structures in Europe.

Other publications:
p. 21 Eurostat index.
p. 22 General social survey bibliography

Forthcoming events:
p. 23 VIII Nordic political science congress.
p. 24 Sixth international conference of europeanists.
p. 25 European consortium for political research: directory of services.

Computer section:
p. 30 Exploring aggregate data; conceptual and practical aspects.
p. 48 Clustar–PC
p. 49 Notes on software
p. 52 Mapics

Scope: humanities computing update
vol. 5(3), May–June 1987

p. 26 Software
p. 27 Grants
p. 27 Hardware
p. 28 Publications
p. 29 Databases
p. 30 How the Writer's Workbench came to Colorado State University./ Charles R. Smith
p. 31 Why can't we use standard software to teach our students?/ Bill Kemp
p. 31 Quotes
p. 32 Courseware
p. 33 Software reviews
p. 34 Calendar

Summer 1987
Data user news from the Bureau of the Census
vol 22(6), June 1987

p. 1 Child care costs $11 billion for households with working mothers.
p. 3 Economic data centers proposed as part of SDC program.
p. 4 Get ready for the '87 agriculture census.
p. 5 No microfiche from the 1987 economic censuses? Keeping track of trucking.
p. 6 Who's minding the kids?
p. 8 1990 census planning: packaging '90 census data.
p. 11 Speed through the trends.

Data user news from the Bureau of the Census
vol 22(7), July 1987

p. 1 After-tax income up for fourth consecutive year. Birth rates still down.
p. 2 Age, sex, and race estimates for your county. New economic census history.
p. 3 Rental vacancy rates still high.
p. 5 Census staff study marketing applications.
p. 6 Children and their participation in government assistance programs.
p. 8 1990 census planning.
p. 9 Homebuilding permits total $98 billion.
p. 10 Studying wages and salaries in hospitals, banks, and other industries.
   Occupational outlook shows where the workers are.

Data user news from the Bureau of the Census
vol 22(8), August 1987

p. 1 People over 100 - a future 'boom'?
p. 2 Dallas-Fort Worth tops Houston as eighth largest metro area.
p. 3 Our aging world.
p. 4 Conference focuses on data use, '90 census and more. [APDU] Manufacturers invest $2.8 billion to reduce pollution.
p. 5 CPS computer files - you create the statistics! Conference on families and households.
p. 6 State Data Centers can perform many services.
p. 7 Federal data highlights.
p. 8 U.S. statistics at a glance.
DDA-nyt
(Danish Data Archives)
nr. 41, Spring 1987

p. 3 Editorial. [in Danish]
p. 6 The history and computing conference in Westfield 1987./ Lou Burnard & Hans Jorgen Marker
p. 11 Articles in DDA-NYT – an overview./ A. B. Lauritzen [in Danish]
p. 21 Processed surveys: [in Danish]
  DDA-0048 Danish leisure survey 1975.
p. 29 DDA-0188 Students at the University of Copenhagen 1760-1967.
p. 35 DDA-0189 Living conditions and hospitalization of children in Copenhagen.
p. 36 DDA-0620 Social reform survey: social workers in county boards of appeal.
  DDA-0621 Social reform survey: social service advisors. p. 37 DDA-0678 Political participation and political attitudes, 1979
p. 38 DDA-0680 Danish electoral and census statistics since 1970, II.
p. 55 English summary.

NSD brukermelding
1987:6, September 1987

p. 2 Hordaland in the elections. [in Norwegian]
p. 4 Social science research 1985-1986. [in Norwegian]
p. 5 International social survey program. [in Norwegian]
p. 7 What you should know about research and confidentiality. [in Norwegian]
p. 11 Survey data. [in Norwegian]
p. 12 Pension data.
p. 12 SPSS in the regional colleges. [in Norwegian]
p. 13 Citation of data files. [in Norwegian].

Social science microcomputer review
V.5(3) Fall 1987

p. vii Editor's announcement./ G. David Carson
p.289 Social science and the new generation of microcomputer hardware./ Marc A. Triebwasser
p.304 LOGLIN: a microcomputer log linear analysis program./ Donald R. Ploch
p.313 Automating social science examinations: an application of spreadsheets and word processing to mathematically oriented questions./ Thomas Palm
p.325 A simplified approach to creating software for computer-assisted instruction./ Jay R. Alperson

Summer 1987
and Dennis H. O'Neil
p.331 Database management with dBASE and compilers: a review and tutorial./ Stanley P. Littlefield

Communications and reports:

p.341 Requiring computer and information systems courses in MPA curricula: a report of faculty opinion./ Lowell Douglas Kiel
p.346 C + I + G: an introduction to macroeconomic modeling./ Frank Vorhies
p.347 Analysis with P/G% in public policy./ Len Faulk
p.350 Which statistic?/ O. Zeller Robertson, Jr.
p.354 News and notes.

Software reviews:

p.377 New software for psychologists and social scientists.
p.378 ROOTS II
p.383 Clinical interview: the mental health series.
p.386 Test plus.
p.388 A mind forever voyaging.
p.389 A Congressional bill simulator.
p.390 American history achievement series.
p.392 Law.
p.392 Statistix.
p.395 Statpac gold.
p.399 Visual statistics.
p.400 Insight 2+.
p.404 Expert system reference engine (ESIE).
p.408 Individual training for project management.
p.410 Showcase.
p.411 Simulation construction kit.
p.412 Reflex: the database manager.
p.414 R & R report writer.
p.415 SQZ.
p.416 Lotus freelance plus and Lotus freelance maps.
p.417 Lotus HAL.
p.418 Lotus manuscript.
p.420 Goal solutions.
p.422 Rightwriter.

Book reviews:

p.424 Technology and the character of contemporary life./ Albert Borgman
p.425 People and computers: the impacts of computing on end users in organizations./ James N. Danziger and Kenneth L. Kraemer
p.426 Learning with personal computers./ Alfred Bork
p.428 Computing in psychology: an introduction to programming methods and concepts./ James H. Reynolds
p.430 Computers in criminal justicis administration and management./ William G. Archambeault and Betty J. Archambeault
p.432 Computers in the language classroom./ Robert M. Hertz
p.434 In search of the most amazing things./ Tom Snyder and Jane Palmer
p.436 dBASE III plus power tools./ Rob Krumm
p.437 Essential guide to bulletin board systems./ Patrick R. Dewey
p.437 Computing information directory 1987./ Darlene Myers Hildebrandt

ACM computing surveys
V.19(1) March 1987

p. 1 About this issue./ Salvatore March
p. 3 About the authors.

Communications of the ACM
V.30(9) September 1987

Articles:

p.770 Corrigenda: An empirical validation of software cost estimation models./ Chris F. Kemerer
Laws of programming./ C.A.R. Hoare et al

Computing practices:

p.772 An object-oriented programming discipline for standard Pascal./ Jonathan P. Jacky and Ira J. Kalet

Research contributions:

p.777 Processing encrypted data./ Niv Ahituv, Yehekel Lapid and Seev Neumann
p.781 A metamodel of information flow: a tool to support information systems theory./ Niv Ahituv

Departments:

p.749 Authors.
p.750 ACM president's letter.

Summer 1987
The American Statistician
V. 41(3) August 1987

p.167 In memoriam: Theodore Alfonso Bancroft, 1907-1986./ Wayne A. Fuller and Oscar Kempthorne
p.169 Analysis of data from the Places Rated Almanac./ Richard A. Becker, Lorraine Denby, Robert McGill and Allan R. Wilks
p.187 Quality/value relationship for imperfect information in the umbrella problem./ Richard W. Katz and Allan H. Murphy
p.190 In search of the optimum Harvey Wallbanger recipe via mixture experiment techniques./ Herman F. Sahrmann, Gregory F. Piepel and John A. Cornell

Teacher's corner:

p.195 Goodness of fit statistics for general linear regression equations in the presence of replicated responses./ Potter C. Chang and A.A. Afifi
p.199 Nonestimable parametric functions for some discrete distributions./ Shaul K. Bar-Lev
p.200 A new look at quartiles of ungrouped data./ John E. Freund and Benjamin M. Perles
p.204 Partial round-robin comparisons with perfect rankings./ Robyn M. Dawes and Joseph B. Kadane
p.205 A unique unbiased estimator with an interesting property./ Colin L. Mallows and Vijayan N. Nair
p.206 The fixed X assumption in econometrics: can the textbooks be trusted?/ James K. Binkley and Philip C. Abbott
p.214 A sample-size problem in simple linear regression./ Charles C. Thigpen
p.215 A note on confidence intervals for proportions in finite populations./ John P. Buonaccorsi
p.218 Characterization of risk sets for simple versus simple hypothesis testing./ Daniel Q. Naiman
p.220 Accent on teaching materials.

Statistical computing:

p.222 Effective microcomputer statistical software./ Kenneth N. Berk
p.229 Statistical computing software reviews.

Summer 1987
Letters to the editor:

p.242 Letter.
p.248 Corrections.

(National Archives of Canada)
vol. 5 (1&2), 1987

   CUIO files – an update. / Chrisiuna Lloyd

p. 3 1987 IASSIST Conference.
   Update on CULDAT.
   "Archives in the Information Age." / John McDonald & Sue Gavrel

IFDO news
(International Federation of Data Organizations for the Social Sciences)
nr. 7, 1987

p. 2 IFDO assembly Ann Arbor.
   IFDO board meeting London.
   Moscow data archive.

p. 3 All-Union Sociological Data-Bank./ Vladimir Andreenkov

p. 4 IFDO seminar Budapest.

p. 6 UK data catalogue project.

ZUMA Nachrichten
nr. 20, Mai 1987

p. iii My own concern./ Max Kaase [in German]

Essays:

p. 1 The Mannheim regional centre as a part of GESIS. [in German]
p. 8 How stable are survey data? Description and preliminary results from the ALLBUS 1984

Summer 1987
Test-retest study. [in German]
p. 32 General Inquirer. [in German]

Project reports:

p. 37 Personal networks in social surveys 1: on the design of a research methodology. [in German]
p. 44 Personal networks in social surveys 2: computer-aided fieldwork management. [in German]
p. 51 Personal networks in social surveys 3: data management in an SIR-database. [in German]

Publications:

p. 57 ZUMA publications in process. [in German]
p. 58 ZUMA research reports [abstracts] [in German]
p. 63 ZUMA Handbook of social science scales – 3rd edition. [in German] p. 65 Publication on the principles of the ZUMA organization. [in German]

Reports on ZUMA:

p. 67 ZUMA research: 'Data analysis and qualitative data collection methods' 7-9/11/1986. [in German]
p. 68 ZUMA research: 'Causal models with latent variables in comparative analysis of multiple groups' 10-13/11/1886. [in German]
p. 69 ZUMA workshop: 'Content analysis' 17-21/2/1987 [in German]
p. 70 ZUMA research: 'Modelling social processes' 10-12/3/1987. [in German]
p. 71 4th conference on the scientific use of statistical software. 23-26/3/1987. [in German]
p. 73 ZUMA workshop: 'Multidimensional scaling' 30/3-3/4/1987. [in German]

Coming ZUMA activities

p. 74 ZUMA workshop: 'GAUSS' 19-23/10/1987 [in German]
p. 75 ZUMA research: 'Practical usage of theoretical panel research' 9-13/11/1987. [in German]
p. 76 Visitors to ZUMA [in German]

News

p. 77 SOCIAL-SCIENCE-BUS: a quarterly, multi-thematic survey. [in German]
p. 78 News from the computing centre. [in German]
Humanistiske data
nr. 1-87, 1987

p. 4 Computer entry for interdisciplinary analysis of archaeological finds of human skeletons./ B.J. Sellevold & J.-R. Naess [in Norwegian]
p. 12 Use of computing at the archaeological museums./ E. Mikkelsen [in Norwegian]
p. 17 Excavations in towns from the Middle Ages./ P.B. Molaug [in Norwegian]
p. 21 Natural language in knowledge-based systems./ I. Utne [in Norwegian]
p. 39 Basic material for information retrieval at the Norwegian terminology bank./ I. Utne [in Norwegian]
p. 54 A Norwegian 'Writer's workbench' - intelligent word processing./ P. Vestbostad [in Norwegian]
p. 60 Computers in language teaching./ P.M. Mathisen [in Norwegian]
p. 65 Experiences with writing interactive line-oriented language teaching programs./ S.M. Sanne
p. 72 The Centre's videodisc project./ R. Erlandsen, C. Huitfeldt & O. Reigem [in Norwegian]
p. 85 Bibliographic databases in the humanities./ A.H. Langballe [in Norwegian]

Reports

p. 88 Computing for humanities students. Experiences with a new 'grunnfag' at the University of Oslo./ A. Braendeland & J. Lanestedt [in Norwegian]
p. 95 Communication theory and semantics – a progress report./ A.J.I. Jones
p. 99 Three resident programs for MS-DOS computers./ E.S. Ore [in Norwegian]
p. 106 Nordic conference on text comprehension and information retrieval./ J.H. Hauge [in Norwegian]
p. 109 Information meeting on the Centre’s videodisc project./ E.A. Drivenes [in Norwegian]
p. 116 Notices [in Norwegian]
p. 139 Summary

Reports

Summer 1987
p. 74 Five CAL programs from the Computer Secretariate./ E.S. Ore [in Norwegian]
p. 80 The MLS Bibliography Generator./ R. Jewell
p. 83 Political attitudes./ E.-A. Drivenes [in Norwegian]
p. 86 CALICO '87./ J.H. Hauge [in Norwegian]
p. 90 Computers and teaching in the humanities./ K. Natvig [in Norwegian]
p. 99 Optica '87./ O. Reigem [in Norwegian]
p. 104 'The use of computers in the teaching of language and languages.'/ P. Vestbostad [in Norwegian]
p. 107 ICAME 8th./ K. Hofland [in Norwegian]
p. 111 XIV ALLC conference./ E.S. Ore [in Norwegian]
p. 112 The 1987 developmental seminar 1987./ K. Natvig, O. Reigem & P. Vestbostad [in Norwegian]
p. 116 Notices [in Norwegian]
p. 126 Summary

ACSPRI newsletter
(Australian Consortium for Social and Political Research Inc.)
nr. 16, September, 1987

p. 1 New ACSPRI members.
   New SSDA catalogue.
p. 2 Fourth ACSPRI summer program.
p. 4 Microcomputer – software info.
p. 7 A.B.S. news.
p. 9 Information sources.
p. 11 Australian research news.
p. 13 Conferences.
p. 14 Australian data available.
p. 19 ICPSR additions to holdings.
p. 20 Contributions to the newsletter.

ZA–Information
(Zentralarchiv für Empirische Sozialforschung)
nr. 20, Mai, 1987

p. 4 Editorial. [in German]
p. 5 Infrastructure of the social sciences institutionally secured through GESIS./ E. Mochmann [in German]
p. 8 Infrastructure as constraint on and opportunity for social science research./ E.K. Scheuch [in German]
p. 18 'Improbable – but not happenstance.' On the establishment of GESIS./ F. Neidhardt [in German]
p. 21 The social sciences in social change./ A. Brunn [in German]
p. 27 Sociology and method. Reflections on the politics of sociological research present and future./ J. Rembser [in German]
p. 34 New data collections of the Zentralarchiv. [in German]
p. 36 Microcomputer diskettes and computer networks for the ZA data service. [in German]
p. 39 A new workbook in the series: International Social Science Council Workbooks in Comparative Analysis. [in German]
p. 40 The current theme: 30 years of European partnership.

Research notes

p. 44 Black & white, response uncertainty, mover-stayer, third force, or...? Further reflections on Jagodzinski's analysis of the post-materialism panels./ R. Langeheine [in German]
p. 56 Some use and interpretation problems in the responsible exploitation of data./ W. Jagodzinski [in German]
p. 64 How should agreement be measured?/ V. Thiessen [in German]

Notices

p. 71 From Glasnost to 'Dschojnt Wentschurs' in social research. Report on a 'Comparative research' seminar in Moscow. [in German]
p. 74 The position of the Methodology Division of the German Sociological Association on the 1987 census. [in German]
p. 75 Methodological problems of the U.S. census. [in German]
p. 77 New methodology literature: [reviews] [in German]
p. 79 Visiting fellowship at the Zentralarchiv.
p. 80 Announcing ISSC STEIN ROKKAN PRIZE in comparative research.

Summer 1987
STEVIN ROKKAN PRIZE
in comparative research

Announcing ISSC STEIN ROKKAN PRIZE in Comparative Research

The International Social Science Council, in conjunction with the Conjunto Universitario Candido Mendes (Rio de Janeiro) announces that the next STEIN ROKKAN PRIZE will be awarded on November 1988.

The prize is intended to reward an original contribution in comparative social science research by a scholar under forty years of age on 31st December 1988. It can be either an unpublished manuscript of book length or a printed book or collected works published after December 1985.

Four copies of manuscripts typed double space or of printed works should be delivered to the International Social Science Council before 15 March 1988, together with a formal letter of application with evidence of the candidate's age attached. Work submitted will be evaluated by the International Social Science Council with the assistance of appropriate referee or referees.

The AWARD will be made at the ISSC General Assembly meeting in November 1988. Its decision is final and not subject to appeal or revision.

The Prize is US dollars 2,000. It may be divided between two or more applicants, should it be found difficult to adjudicate between equally valuable works submitted.

For further enquiries, please write to:

The Secretary General
International Social Science Council
UNESCO - 1 rue Miollis
75015 Paris, France

Summer 1987
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, data base 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: $20.00 per calendar year.
- Student Membership: $10.00 per calendar year.
- Institutional subscriptions to the quarterly are available, but do not confer voting rights or other membership benefits.

Institutional Subscription: $35.00 per calendar year (includes one volume of the Quarterly)

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

- $20 Regular Membership
- $10 Student Membership
- $35 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:

Ms Jackie McGee
Treasurer, IASSIST
% Rand Corporation
1700 Main Street
Santa Monica
California USA 90406

Name/phone

Institutional Affiliation

Mailing Address

City

Country/zip/postal code