
Provider Sophistication Versus User Simplicity: European Servicing Through Bridging the Gap

by Per Nielsen¹
Danish Data Archives
Odense, Denmark

Introduction: Working on the right problem

When I first entered the “data archive movement” (February 1, 1974), everybody seemed to be very preoccupied with the creation of advanced software for the mainframe: Report generators (even though there was little to report on), search systems (even though there were few surveys to search among), and data base systems. Later on, we realized that the *de facto* standards were developed at larger organizations - either within (OSIRIS from the ICPSR, SPSS from NORC) or outside (SPSS Inc., SAS in Ralceigh, and all the other business firms all over the marketplace) “our world” of social science research institutes and data archives. I am sure that we spent quite a lot of time during the early years working on the wrong problems given the needs of the time; however, the work initiated the habit of engaging in cooperative projects among the European archives. This good working habit has survived ever since; as institutions and as individuals alike, the European archives have a close and pleasant collaboration program, ranging from responding to incoming servicing requests over staff-relevant Expert Seminars² to Business Meetings once or twice a year.

One topic that was hardly ever discussed during these early years (even internally among data archivists) was the actual level of servicing provided by the data archives during a given year; it was the tacit understanding that the actual (quantitative) level of servicing was not a topic that would underline the *raison d'être* of the new data organizations, data archives, and data libraries.

Now, almost twenty years later, we have vast amounts of data sets in custody, and the demand for data for secondary analysis has increased dramatically, not least with the advent of the PC. Unfortunately, we are not so well-prepared to meet this demand as one would expect, given the advanced techniques developed and applied in the take-off phase. In Babel-like Europe, there are still huge obstacles to a free data flow from data providers to data users. We are working on these, but there is a long way to go. At a recent seminar, two American scholars³ told us (the data archive professionals) that we had been overtaken by the “ordinary” library people in terms of computer mediated communication. Shame on us!

The real problem during the sixties and seventies, looking in the rear-view mirror, was to localize and collect data, to develop standards for documentation, to teach data collectors “sound methodological/technical practices” during the data generating process, and to store the data safely - with a long term archival perspective in mind. We did perform all of these tasks once we found out that we could import most of the software tools from the outside; however, then we worked little on the search systems and the other advanced tools that successively became relevant as we had thousands of data sets in our holdings; in many archives, the then advanced retrieval systems of the seventies were maintained and slightly developed; some of them are still in use in the early nineties.

The “right problem” is simple: To find relevant data sets that meet the specifications of a user and transfer the data to that user - with the shortest possible elapse time and with the least possible input of human and machine resources in both ends of the communication line.

The topic: Remote Access and New User Services

My impression is that American scholars find it very difficult to get an overview of the European data marketplace; honestly, it is sometimes difficult even to Europeans working in the market!

In this paper, *Remote Access* is interpreted as “Access from North America to European Data” - rather than looking at the technical notion of remote access (i.e. running jobs on a distant machine). The underlying philosophy is that the user is substantively oriented rather than technically fascinated; the user would rather have data available in a known environment than shuffle around in dozens of differently functioning systems to dig out what (s)he needs.

Given this interpretation, *New User Services* will, to a certain extent, become equivalent with present user services. Some 80% or more of the DDA-servicing is domestic (i.e. national), so the new services will be developed for the national market first. Consequently, I can claim to be knowledgeable about the Danish situation, only; and, being a native of a small country with a peculiar language, I realize that this situation may be of

minor interest in North America.

Finally, my personal feeling is that a paper on *Remote Access and New User Services in Europe* would be much better in 1995 (third year of the Open Internal Market, cf. below); after a couple of decades with consolidation on the (national) *archival* side of the data organizations, we shall now move into an era where the (national and international) *servicing* aspects gain more weight. This gradual shift in emphasis is (among many other indicators) reflected by the fact that the ECPR Council has accepted the theme of *Integrating the European Data Base* for the ECPR Joint Sessions of Workshops 1992 (Limerick, Ireland)⁴.

The scene: The integrated Europe (United States of Europe?)

Being in North America, a notion about the EEC-generated phantom of European Integration is perhaps necessary. With the introduction of the Open Internal Market (end of 1992) and the current plans regarding an economic and monetary union (three stages during the nineties), with or without a political (and maybe even foreign policy and defence) union, the Brussels establishment (especially the Commission) has succeeded in moving some frontiers of thinking, especially in business - and maybe even more so in North America and Japan than in Europe!

Even though the United States of Europe is being discussed, by supporters and adversaries alike, this vision will remain a phantom to most Europeans for another couple of decades. In the perspective of *Integrating the European Data Base*, the differences of language, ethnicity, culture, religion, wealth and political and social science tradition represent obstacles to the free data flow; the same is true of the differences in economic development and technological sophistication - a gap which is more evident in the East-West dimension than on the North-South continuum.

Furthermore, it is important to notice that the cooperation between European research institutions (and hence social science data archives) has never been limited to European Community members; it has been open to any institution with the interest and the capacity to participate. Generally (and this is especially true with respect to the data archives) the nation-state has been the represented unit. It has been difficult in some countries to find the relevant (i.e. nationally representative) institution; and this problem area will return to the scene as new states emerge (East) and as some countries develop specific institutions to deal with data archiving (e.g. specialized historical data archives in the larger countries).

Finally, with a landscape of "peer partners" among European archives, it has not (yet?) been possible to establish the *European Data Archive*⁵ that might be the gross dealing agent in Europe, comparable with the ICPSR in North America. Needless to say, it would be easier from the outside to address one central agent that collected and disseminated all major European data sets of broader interest. Our response so far to this demand is: Contact any one of the archives, and they will (ideally!) let the message pass to everybody else. As a matter of fact, this procedure has proven to be efficient on a number of occasions; but please be accurate and specific when elaborating the request!

Which are the heavy resource demanding servicing tasks?

In the American context, the European archives should be understood as an amalgamation of the gross dealing agents (e.g. the ICPSR) and the local retail servicing facility (university data libraries, state data archives, etc.) Covering the whole set of archival as well as servicing procedures, we have a good overview of the costs involved in different parts of the whole process.

At the archiving end, it is the cumbersome data processing to a *standard archival format* that digests heavy resources. Even though standards may vary from place to place, most of us want to produce standard codebooks (in a format derived from the OSIRIS dictionary-codebook format - type 3 to have a clean ASCII character set.) Most of us want to have a good study description (sub-structured in a more detailed way than the ICPSR free text study description), and most of us want the data to be immediately accessible for the major analysis packages like SAS and SPSS.

At the servicing end, the gross-dealer functions take little time: Users requesting specific data sets that have already been processed to the standard archival format can be serviced from one day to the other - or even within hours; they will receive easily accessible data and can start off with their analyses immediately, cf. below. The resource-consuming customers are those that want to obtain data that meet certain search criteria (often too vaguely defined!) - and this is especially time-consuming if the user wants to perform cross-national comparisons and/or if non-standard data sets are involved.

Obstacles facing the user of European data

Below is a list of existing obstacles that the data user may face when trying to get hold of data relevant for (cross-national) analysis. Imagine that the user asks the local national archive in one country to facilitate access to data from several European countries, relevant to a certain topic; what is the process ahead?

1. The local archive sends out a "search warrant" to all other European archives. This can be done quite quickly via E-mail. However, most European countries still do not have a data archive; and some of the existing national archives are heavily under-staffed, so that the requesting archive will get a late reply or even no response at all. Result: With good luck, relevant data will be found in 3-5 countries, only.

2. Some of the data sets localized may have been produced by central statistical offices (CSOs) or administrative agents; in these cases, data may not be available at all - or the user will have to go to the country in question because data export is not allowed. Another possible obstacle is an embargo period for secondary analysis, introduced by the primary investigator.

3. Some of the relevant data sets may have an inappropriate format, i.e. they are not immediately available for analysis with SAS or SPSS. It may take months or even years until the data archive has improved the technical availability of the data set.

4. The data user may have to sign undertakings with each of the primary investigators before the data can be delivered for secondary analyses.

5. The documentation of the relevant data sets may be available in the local language, only; and this, of course, is the rule rather than the exception.

6. Remedying obstacles 1-5 may cost real money that the user may not have available.

7. The (few) data sets that actually pass the obstacles 1- 6 will now be sent to the requesting archive; they in turn pass the data sets on to the user.

8. If, during the secondary analysis process, problems arise with the data or the documentation, the trouble-shooting will be quite difficult also.

The above picture of the obstacles is quite pessimistic, some European data archive people might say; unfortunately, my feeling is that it is a realistic one. There is a long way to go for the archives united in CESSDA* (Committee of European Social Science Data Archives) before we have an Integrated European Data Base.

With 15 years of fieldwork, we have not yet fully implemented the visions presented in a paper at the CESSDA founding meeting by one of the fathers of the Data Archive Movement, the late professor Stein Rokkan: "Our basic philosophy is very simple: we do not believe the archival movement in Europe will get anywhere unless there is a real break with the tradition that archives are there simply to store, clean and reformat separate data sets. The future lies with active reorganization of data: linkage across files, build-up of time series sets, preparation of handy packages for use in the classroom, integration of packages with better computer routines for graphic display, cartography, visual model-to-data fitting." (Rokkan's underlinings).

Remedying the obstacles: Remedying the obstacles: State-of-the-art and planned activities

We are doing our best to try to smoothe the facilitation of European data to the user. Let's reiterate on the obstacles mentioned above and see what is being done and what can be done within each of the "obstacle fields" identified in that section. Doing so, we shall look at the European level first, and I shall add a few comments about the Danish situation - which I know best!

Localization of relevant materials

Most archives do have printed *catalogues of holdings*, with *multiple indexes*, from which you can figure out whether the other archives do have relevant data sets in a given field. However, the printing is expensive, and the paper-bound inventories tend to become outdated quite quickly - both with the acquisition of new data and in terms of "processing classification," access restrictions, etc. It is possible, of course, to acquire the machine readable text from the catalogues of each archive and search these in your own retrieval environment; but this does not solve the updating problem. Consequently, the most evident solution is either to integrate the primary cataloguing at one central location or to search in the catalogues of the other archives, via telecommunication, at their own computer installation.

The first path, *integrating the catalogues*, has been worked on with some energy. The Commission of the European Communities (CEC) had actually granted money that would allow catalogue integration in a project with DDA, ESRC-DA, STAR and ZA as the major project partners. This project stranded because the CEC demanded that the resulting data base had to be commercially viable after the 2-year project period. (It probably would not have been after 10 years; but the CEC bureaucrats, preoccupied with "commercialization," are not at all sensible to the special problems in the academic sector!)

The second path, *searching via telecommunications*, is

probably a more realistic one. Given that most archives now have TCP/IP and FTP facilities available at the installation where the catalogue information is stored, the searching as well as the actual exchange of data may take place using these communications and file transfer facilities. This procedure assures the user that the most updated version of the catalogues is searched and the most processed data set is transferred.

Even though many of the archives have retrieval facilities that are open to the user, most searching is still done by the staff of the archives on behalf of the user; this is probably going to be the case in the foreseeable future for all other than very heavy users of data.

In the case of the DDA, we have one central retrieval system, DDAGUIDE, based on the study descriptions. It is available to most potential Danish customers, located at UNI*C (a national computing center for research and education). However, it is not yet open to users without an account number at UNI*C. In addition to DDAGUIDE, we have several in-house search systems (some mainframe-based, others PC-based) searching the contents of the machine readable codebooks. In order to integrate retrieval at the study description and codebook levels, we have designed an integrated system and applied for money from the SSRC to have the UNI*C-people implement the system; hopefully, there will be a remote user access to this integrated retrieval system⁷. (On the other hand, the language stored will be Danish, cf. below, where the DDAGUIDE stores English language texts!)

Handling access restrictions on the data

In most European countries, the access to process-produced data (administrative and statistical data) is hampered - mainly due to three factors: (1) The bureaucratic traditions of government and a lack of Freedom-of-Information-tradition; (2) the privacy legislation; and (3) the wave of cutting in public spending - indicating that the statistical bureaux (CSOs) and other data owners want to sell their data rather than offer the data for free (or at a quite low price) via the social science data archives. Let us take a look at a few examples:

In Norway, where the relations between the central statistical bureau (CSO) and the NSD has been better than in most countries, the NSD can disseminate a lot of statistical data for research and educational uses; this has mainly been done with regional data, where the NSD probably has the largest collection of commune-based data in Europe. On the other hand, there are severe restrictions to the access to survey-data (data on individuals); for instance, the later election studies in Norway have been collected by the CSO; consequently, the data may not be taken out of the country, so that the user will

have to go to Norway to use such data sets.

In Sweden, also, the election study data sets and other CSO data on individuals may not leave the country; the SSD has tried to apply the LIS-model (Luxembourg Income Studies) to gain indirect user access to such data⁸.

In the United Kingdom, the ESRC-DA is "re-selling" selected data series from statistical authorities to the academic community. Also in Hungary, Tárkai has quite close relations with the statistical office in Budapest.

Data from the academic community are usually more easily available than data from the CSOs. Even so, it is in some countries (for some studies) necessary to ask the primary investigator's permission. It seems to be generally accepted that the primary investigator may impose up to a 2-year embargo on the data.

In Denmark, the CSO (Danmarks Statistik) tries to make a lot of time series data banks available on a commercial basis. Four rather large "data banks" are available, covering national economic time series (DSTB), commune statistics on the 275 local administration units (KSDB), labor market statistics (ABBA), and business related statistics (ESDB). However, the Danish CSO is very reluctant to release survey data of any kind. The big public survey organisation, The Danish National Institute of Social Research, on the other hand, generously puts all its surveys (in a *de facto* anonymous form) at the disposal of the DDA and her users - free of cost. This fact demonstrates that it is the interpretation of the data legislation rather than the Acts themselves that render access impossible.

Handling technical problems with the data

There is little standardization across Europe with respect to the processing classes (cf. ICPSP's Classes I-IV - a data class structure which is presently being revised). Some archives (e.g. DDA, SSD, ZA) follow a strategy pretty much like that of the ICPSP, having machine readable codebooks for their top-class studies. Other archives (e.g. NSD, STAR) try to make as much as possible available as SPSS Export files - not necessarily having all information from the questionnaire (or other instrument) in the machine readable codebook. The ESRC-DA has realized that they do not have the resources to process their several thousand studies to the "ICPSP Class I"-level; instead, they have developed a thesaurus and apply that to append relevant search entries to the study descriptions in order to have a search base without having fully-fledged codebooks for all studies.

In general, most European archives aim at making the data sets available for analysis with SPSS and SAS.

However, most archives hold many data sets that have not (yet) reached this level of processing (ICPSR Classes II-IV); for some of these data sets, the user will have to produce the setup on his own, based on a card-image data set and a paper documentation thereof.

At the DDA, we stick to the traditional "ICPSR-type" of documentation with respect to the codebook, adding a (structured) standard study description to this codebook. Streamlining the data documentation and processing work (mainly based on programs such as SAS, KEDIT and REXX in an OS/2 networking environment) we can more than keep pace with acquisitions, so that the number of "non-Class I" data sets is diminishing.

Data ownership as a restriction to remote access

Whereas many archives have *searching facilities* available for remote access from the users, few archives have the data files as such available for immediate analysis. This is usually not due to technical restrictions; rather, it is based on proprietary considerations: The principal investigator is the official owner of the data, and sometimes her or his written consent is required in each individual case before the archive can offer access to the data set itself.

This type of restriction should probably be removed in the future; given the fast technological development (with communications protocols like TCP/IP and file transfer protocols like FTP on the one hand, and distribution on CD-ROM or other mass storage media on the other hand), the provision of free distribution should be granted to the data archives by primary investigators.

Vis-à-vis the researchers, the DDA has not yet found the formula that will allow access to all stored materials without prior written consent from the depositor. However, access is never denied, so we consider it feasible to get an agreement about unrestricted access with most donors, once we really need that - either in order to allow remote access for analysis on our computing facilities or to distribute selected data sets on mass storage devices.

The tremendous language problem in Europe

Within each country, it is considered "normal" or even indispensable that the documentation be produced in the national language; the only general exception to this rule seems to be The Netherlands, where Steinmetzarchief produces SPSS-setups for all files in English rather than in Dutch. Most archives do have catalogue information available in the English language, but the codebooks and/or the questionnaires (or other instruments of data collection) are available in the national language, only. This is an obstacle to remote (*in casu* foreign) access - to which there is no readily available solution.

Everybody who has been engaged in cross-national research projects will know that it is a tremendous problem to produce cross-culturally comparable data, in part due to the language problem. This is true even in the culturally relatively homogeneous European Community (reflected in the Euro-Barometer surveys); but the difficulties are even greater if one goes to Second or even Third World nations (which, for instance, the ISSP-program is doing).

Definitely, there is not enough resources within the European archives to produce all documentation in the national language *and* in a world language (e.g. English). All users of European data should be aware, consequently, that they will have to be able to read the language of the nation under investigation - with the aforementioned exception of The Netherlands. (Some scholars would argue that you would have to know some language and culture prior to engaging in quantitative (or qualitative) investigations of a specific nation *anyway*; we shall not engage ourselves in that discussion here.)

At the DDA, we keep study descriptions in both Danish and English; we can, therefore, inform about our data (in catalogues and *ad hoc* listings of selected topics) in either language. But with the codebooks it is different; even though we have produced some English language codebooks (in addition to the Danish ones) for a few frequently exported data (e.g. the *Continuity Guide to Election Studies*, which is also disseminated through ICPSR), the bulk of the codebooks are available in the Danish language, only.

It may be a future project among European archives to produce English language codebooks within areas where cross-nationally comparable data sets can be "constructed" by the archives. Such projects have been successfully carried out in the past; for instance, at lot of National Election Studies (and Continuity Guides based on these) have been produced in cooperation with the ICPSR and are now distributed from Ann Arbor to the membership.

Fee schedules among the European archives

In some countries, the servicing of users is free of cost (apart from "media" such as paper-codebooks, diskettes, etc.); other archives have a fee schedule - the size of the fee depending on factors such as size and complexity of the data sets delivered, staff time involved, etc. Sometimes there is a discriminatory pay-schedule, where students are at the cheap end and business applications in the expensive end - sometimes with researchers in a middle position. It is beyond the scope of this paper to try to spell out all the fee schedules; they change now and again, so the user has to ask in each specific case.

At the DDA, servicing of archived files is free (media cost recovery is demanded if the media are not returned). Also, the staff and machine time consumed by performing searches for users is free. However, special services such as "super-quick processing" of DDA-studies, processing of requestor's own data sets, or translation of codebooks will have to be paid for by the requestor.

Actual transmission of data to the user

As a consequence of an agreement between all CESSDA archives, there are certain rules that apply to international data transfers; the major contents of this agreement can be summed up as follows:

- * Each national archive is the primary repository regarding data from that nation. ("Fishing-zone agreement").
- * All requests from within one of the "CESSDA-countries" should be directed through the home archive. (This seemingly bureaucratic rule is administered liberally: If a foreigner requests Danish data, for instance, we would always inform the relevant "home archive" and, if they so wished, send the data via that archive). The idea is, of course, that the users should not be able to circumvent pay schedules by going abroad to the free or low-cost archives.
- * Cross-national data sets are processed and disseminated according to mutual agreements among the relevant archives.

As mentioned before, the actual transfer of an already processed data set is not very time consuming. However, the actual procedures may differ from place to place; needless to say, in inter-archival transfers the technicalities would be agreed upon beforehand (if they are not already known from earlier transfers).

At the DDA, the transfer is done according to the specifications wanted by the actual user. If the user works on a mainframe, we would normally send the data to that mainframe from our central archive (magnetic tapes at a central UNI*C mainframe). If the user wants to work on a PC - which some 90% of users do, we will send the data on a diskette, containing a DOS BAT-file that will do all the work necessary before the analysis: Make the necessary directories on the user's hard disk, copy the files onto the hard disk, UNZIP the packed files, and, if necessary because of multi-volume delivery, put split files back together.

Let us assume that the study number DDA-9999 was requested and sent on one or several diskettes; after the

user has run a DDA-COPY.BAT job, (s)he now has the following files in a directory with the name C:\DDA-9999:

DICB9999.OSI	(OSIRIS-like dictionary-codebook, ASCII)
DATA9999.OSI	(OSIRIS-like data file, ASCII)
OSI-SPC.EXE	(DOS-program for system file, cf. below)
LIST9999.PRT	(ASCII-listing file with SD and codebook)

All the user has to do now is to run the OSI-SPC program (a DDA-utility) which will ask (1) whether the user wants an SPSS or a SAS file; (2) which variables the user wants to include in the systems file; and (3) the names of the dictionary-codebook input file and the setup output file. When OSI-SPC has finished (in seconds, even with very large files), a setup is ready to build the specified SPSS or SAS systems file - with all necessary variable and value labels in place. Again, the user needs to specify only the data set names before running the systems file generating job. The LIST9999.PRT file is just a stream of lines that can be printed on any type of printer; the idea is that the user can save the money for the printed documentation if (s)he prefers to run off a printed copy instead.

The supply of data on diskettes takes place from a copy of the original archive which contains the ZIP-files; this "archive copy" is kept on a 1 Gb traditional disk on a net server, so there is very quick access. Needless to say, such files can be sent over the external networks to the user instead of using diskettes; however, the DDA is waiting for the OS/2 version of TCP/IP and FTP - which is to be in the market very soon.

Our experience is that users (and we ourselves, when receiving foreign data like that) are very satisfied with the present procedure.

Problems with data or documentation during secondary analyses

It is important to know which archive is responsible for the data sets that "drift around" in the international social science community; otherwise, when errors or omissions occur during the secondary analyses, it is close to impossible to remedy or clarify such problems. A couple of examples will demonstrate this problem:

parties (Venstre, i.e. the (Conservative) Liberals) are absent from data as well as from documentation. Who "produced" the error: The principal investigators (Jacques-René Rabier, Helene Riffault, Ronald Inglehart); the Danish data collector (Danish Gallup); the international coordinator (Faits et Opinions); the Zentralarchiv (where the file is first available); or the ICPSR (where the final documentation is produced)? - Needless to say, the error is critical to a political scientist who will use this variable extensively during the analyses.

A Danish user participating in the repeated International Value Project finds out that there are problems with the "oversampling" of young people in the 1981 Value Project file for Denmark. There were no Danish researchers involved in the 1981 Value Project; all you can do is to ask the Danish collector (Observa, which has changed name and staff in the meantime); the Coordinator (British Gallup); or the involved archives (*in casu* the ESRC-DA, but we got the file from NSD). It is more than likely that you can never solve the problem! (Which, in this case, derives from the fact that archives entered into the process of preserving the data long after the primary investigation.)

In conclusion: What is achieved, and where are we heading?

In the European countries with a national data archive, a huge data resource is immediately available for analysis to everybody who knows the human language of the country; the range of users has been augmented laterally (data in schools⁹ and easily operated analysis packages¹⁰) and horizontally (the concept of "social sciences" is broadening with "new" disciplines all the time.)

The users want the data "packaged" to their needs, on their own computer and for their own analysis package; it is my feeling that we have accomplished this task on most levels of user sophistication: The number of users is constantly rising, and the servicing is becoming very efficient in the archives. To move from diskettes to FTP transfer of data is a technical detail of minor importance to the user; however, it may save resources among the data suppliers and make the transfer across long distances quicker.

The broadening of the topical area covered by the archives is a process that may take different paths in different countries. Take the historical data as an example: In Germany, the Zentrum für historische Sozialforschung started its career as an independent institute; later, the ZHSF was moved to form a department at the ZA. In The Netherlands, a historical archive is on the steps; right now, however, funding is lacking. In the UK, the situation is under review.

In smaller countries, it seems likely that the existing archives will cover the "new" disciplines. In Denmark, data from history as well data from social medicine are archived at the DDA - and have been for years. The DDA hosts the up-coming AHC Conference 1991 to demonstrate that fact to everybody in Europe¹¹.

Obviously, the immediate future in Europe will be devoted to the project of *Integrating the European Data Base*. CESSDA is right now "incorporating" (with a formal Constitution and membership fees) as one step in that direction. Projects are underway that will facilitate searching across archives. Data exchange will take even less time in the future using telephone lines for transmission rather than snail-mail with data media.

Let the remote users come; we shall give them access and demonstrate that our services are better and quicker than ever before!

Footnotes:

¹ Presented at the IASSIST 91 Conference held in Edmonton, Alberta, Canada. May 14 - 17, 1991.

The Danish Data Archives (DDA) is a national social science data archive established in 1973. Since 1978 the DDA has been located at Odense University; some time in 1991, the DDA is likely to be relocated, most likely to be a department of the Danish National Archives. Danish Data Archives, Munkebjergvej nget 48, DK-5230 Odense, Denmark, E-mail: DDAPN@NEUVM1.BITNET or ddapn@vm.uni-c.dk, Telephone: (+45) 66 15 79 20 x 2810, Fax: (+45) 66 15 83 20

² The European archives arrange so-called Expert Seminars for staff-members, hosted by one of the archives, once or twice a year. It should be noticed that the data archive "milieu" is more institution-based than person-based compared to the situation in North America, cp. the large IASSIST constituency in North America compared to Europe.

³ Professors Harold Clarke and Mark Franklin, Texas, made these comments during the ECPR planning session *Integrating the European Data Base* (Essex, UK, March 22-26, 1991).

⁴ This Workshop, co-chaired by Ekkehard Mochmann (ZA) and Eric Tanenbaum (ESRC-DA) was being prepared in March of this year, cf. note 3 above.

⁵ Unfortunately, a local research institute in Mannheim (FRG) has adopted the name EDA (European Data Archive); this may confuse some users inside and outside Europe. The name is misleading also in the sense that the holdings of EDA are mostly "second hand data" (i.e. data from other archives).

⁶ CESSDA was founded in 1976 in Amsterdam (May 31 - June 1) as an informal cooperation between existing social science data archives. CESSDA is the European branch of IFDO (International Federation of Data Organizations). Active European archives are - with an asterisk (*) in front of the CESSDA founding members:

- ADB (All-Union Data Bank), Moscow, USSR
- * ADPSS (Archivio Dati e Programmi per la Scienze Sociali), Milan, Italy
- * BASS (Belgian Archives for the Social Sciences), Louvain-la-Neuve, Belgium
- BDSP (Banque de Données Socio-Politiques), Grenoble, France
- * DDA (Dansk Data Arkiv), Odense, Denmark
- * ESRC-DA (Economic and Social Research Council Data Archive), Essex, UK
- * NSD (Norsk Samfunnsvitenskapelig Datatjeneste), Bergen, Norway
- SSD (Svensk Samhällsvetenskaplig Datatjänst), Gothenburg, Sweden
- * STAR (Steinmetzarchief), Amsterdam, The Netherlands
- TARKI (Social Science Information Center), Budapest, Hungary
- * ZA (Zentralarchiv fuer empirische Sozialforschung), Cologne, FRG
- WISDOM (Wiener Institut fuer sozialwissenschaftliche Dokumentation und Methodik, Vienna, Austria)

Other countries, e.g. Czekoslovakia, Ireland, and Switzerland, are expected to set up similar archives in the near future.

⁷ This system represents the first stages of a larger project presented by Karsten Boye Rasmussen at the IASSIST Annual Conference 1990. The remaining parts, having searched study descriptions and codebooks, include an automatic downloading of the relevant data sets for analyses.

⁸ The LIS-model aims at securing access to confidential data in the following manner: Instead of distributing the "real" data, a constructed "model data set" with the same distributional characteristics is disseminated. Once the user has produced the setup that generates the right analyses, (s)he sends that setup to the data base administrator - who then runs it against the real data. The administrator sends the output to the user after checking that no confidential information can be disclosed in the output. (Setup and output can be sent via electronic mail.)

⁹ Most "CESSDA countries" have made teaching packages based on some of their more "popular" data sets for

undergraduate or even school students. In Denmark, some 25% of all schools acquired a teaching package during the second half of the eighties.

¹⁰ NSDstat from the NSD is presently being distributed in several of the other "CESSDA countries". NSDstat was presented and demonstrated during one of the workshops prior to the IASSIST Edmonton Conference.

¹¹ Association for History and Computing 6th International Conference will take place in Odense, Denmark, August 28-30, 1991. Usually, several hundred historians from all over Europe (and some overseas guests) participate in the AHC Annual Conferences.