Facilitating Access to Comparative Data

by Ekkehard Mochmann & Lorenz Gräf[1],
Central Archive for Empirical Social Research (ZA);
at the University of Cologne

Mandate of the ZA as part of a Social Science Infrastructure
The Central Archive for Empirical Social Research (ZA) at the University of Cologne serves as a research, training and resource center for social research. Founded in 1960 by the Faculty of Economics and Social Sciences of the University of Cologne, it soon developed into a data service with a supraregional and international clientele. As a central node in the international data service network it became the starting point for a more comprehensive social science infrastructure, the German Social Science Infrastructure Services (GESIS e. V.). This association was created as a response to needs formulated by the social science profession in 1986 to provide infrastructural services in all fields of social research with particular emphasis on:

- collecting data and making it available for further research.
- informing about social science literature and research projects.
- development of research methods, teaching instruments and methods consulting for research projects.

The core of the ZA mandate is to facilitate access to already existing data, especially survey data, which can be used for secondary analysis. The holdings cover all fields of empirical social research. Beyond survey data there are collections of statistical data, regional data, various types of quantitative historical data and machine readable texts for computer assisted content analysis, as well as party manifestos and other text collections.

ZA provides services in the area of acquisition, processing, documenting and making available data for social research, especially survey data. ZA offers consulting services for secondary analysis. Training in complex analysis methods takes place twice a year in the ZA spring seminar for empirical social research and the autumn seminar for quantitative historical research. Beyond this, ZA creates ex post statistical time series and supports comparative international studies for the analysis of long term social developments .

ZA holdings of empirical social research data include European time series and comparative studies. The ZA department ZHSF (Center for Historical Social Research) develops data bases, in some cases going back to earlier centuries. The ZA holds nearly 4000 data sets and data collections. Even though there is no particular topical restriction, emphasis is on topics such as political attitudes, election studies, education, unemployment, leisure and occupation, media and the environment.

Among the data sets intensively used are the EUROBAROMETERs (a data pool of comparative surveys from European countries taken for more than 15 years), the German General Social Survey ALLBUS, which is conducted every two years, the International Social Survey Program (ISSP) for 25 countries from Australia, America, Europe to Japan. Similar attention is paid to the monthly POLITBAROMETER series provided by the Research Group Elections (FGW: Forschungsgruppe Wahlen) which is also presented on the second public TV station (ZDF) every month and the collection of surveys to the national parliament (Bundestag) since 1949.

A GESIS branch in Berlin is now focusing on data and information transfer from and to Eastern Europe. Recently more than 400 data sets from surveys conducted in the former German Democratic Republic (GDR) since 1975, were included in the ZA holdings and were processed for secondary analyses. Currently emphasis is on supporting initiatives to create infrastructure institutes in Eastern Europe and to develop a service network for European wide data transfer.

The ZA has access to data held in the social science data archives world wide. International data transfer is coordinated with the Council of European Social Science Data Archives (CESSDA) and the International Federation of Data Organizations for the Social Sciences (IFDO). Access to internationally distributed data bases is supported by making use of modern telematic services like WAIS, WWW, FTP on the INTERNET and other computer networks.

Selecting relevant data and solving methodological problems relating to secondary analysis is an essential part of individual consulting. The newsletter ZA INFORMATION and the journal Historical Social Research (HSR) inform about new data

sets, methodological developments, research findings and conferences. A documentation of more than 1000 empirical research projects conducted in Germany, Austria and Switzerland is published annually.

Organizational priorities in collecting and distributing data
From the very beginning the ZA philosophy was to develop services in close interaction with the scientific community. As a consequence of this philosophy ZA also supports a small research and training department which focuses on new methodological developments in data collection and analysis. Under a guest professor scheme scholars from abroad are supported in their research from planning new surveys to secondary analysis of available data. Experts in data management and analysis offer advice from the selection of appropriate data to advanced statistical analysis.

Already in the 60s Erwin K. Scheuch, one of the ZA founding fathers, created a climate for comparative research which was inspired by the Standing Committee for Comparative Research of the International Social Science Council, in which he cooperated with Stein Rokkan and Warren Miller. This orientation was enforced by the emerging European Unification and the globalization of social research.

Over the years several international research projects have chosen the ZA as their resource center for creating an integrated data bases. Integrating national data sets into internationally comparative data sets includes comprehensive documentation of methodological, technical and historical background of a study and additional interpretation knowledge to facilitate further comparative analysis. Currently ZA serves in this function for the EUROBAROMETERS (jointly with ICPSR and Swedish Data Services (SSD), the International Social Survey Program (ISSP), and the major election studies to national parliaments in Europe (ICORE)[2]. Bringing together researchers working on the data and the data management experience of ZA provides a unique working environment for creating an integrated fund of knowledge on core topics of European social development. In cooperation with the principal investigators and other European data services ZA coordinates and creates European data bases, which could otherwise not be made available to the scientific community, relying just on national resources.

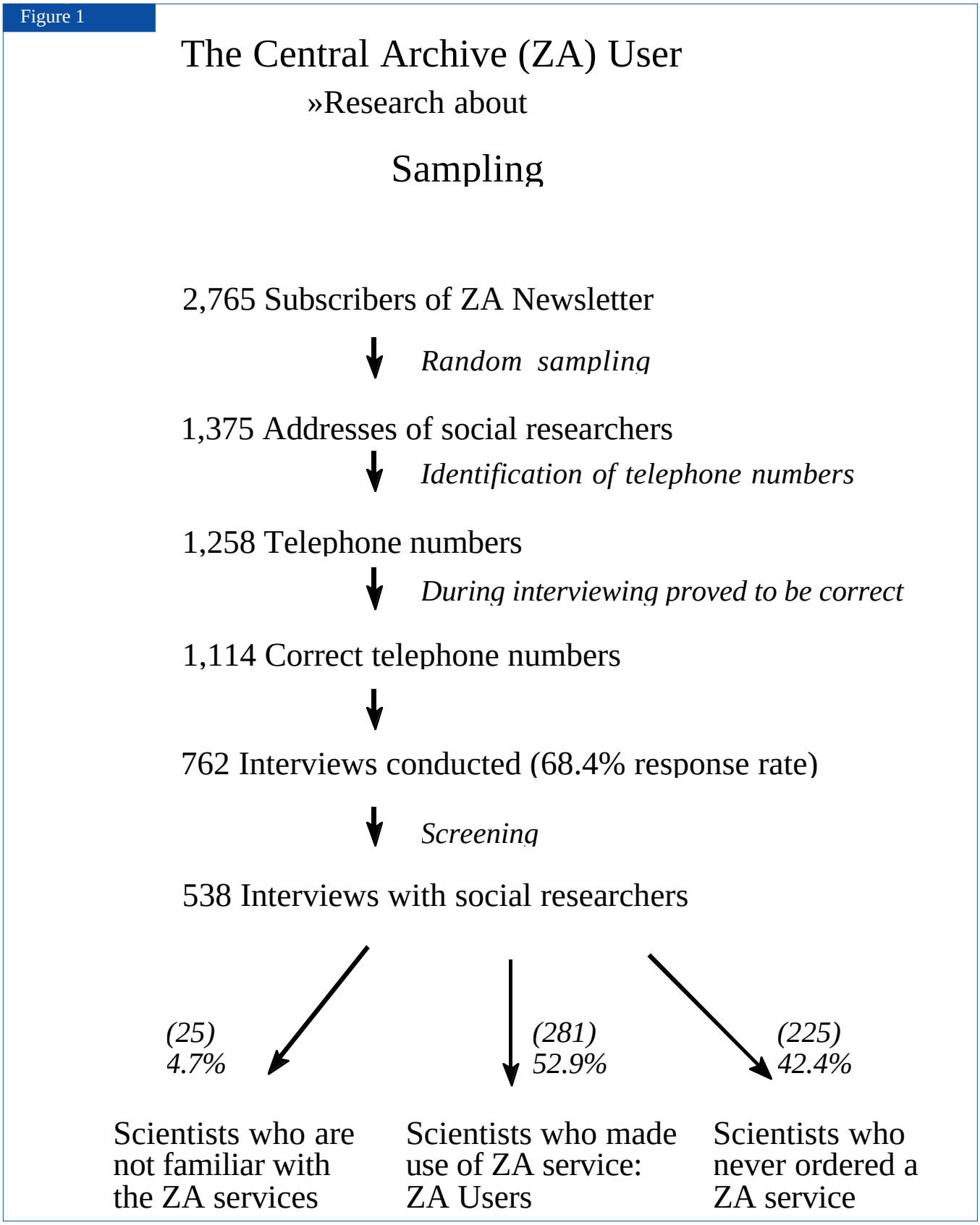
ZA strongly supports a policy of labor division between European archives according to topically focused European data collections. Under tightening resources this is a must for integrating the European data bases. Inspite of intellectual and political efforts there is an ongoing demand for additional European resources to achieve what cannot be covered by the subsidiarity principle: the data service capacities are by and large absorbed by the national demands and there is little leverage to cope with additional international workloads.

Direct access to the expertise and information banks on social science literature and research projects, as well as to the methodological expertise of its GESIS partner institutes complement this infrastructural support for the production and analysis of comparative data bases on Europe.

The ZA User Survey
Although the central archive has always made efforts to communicate with its clientele we have found it necessary to get more information about our clientele to face the rapid technological changes which are taking place. In the past, information concerning needs and demands of the clientele were mostly gathered by mail surveys. This procedure involves three major drawbacks. First, only the users of the institution are surveyed. So we would miss the comments of those researchers who did not make use of the data services. Second, the findings are often biased because only the most motivated people contribute in these surveys. Third, the response rates in mail surveys are low. The installation of a laboratory for telephone surveys at the University of Cologne last autumn gave us the opportunity to avoid these drawbacks. In a pilot study to test this telephone facility we could interview social researchers about their research environment and about their impression of the central archive.

Description of the sampling procedure
The target population of this study were all social scientists engaged in empirical research. For this purpose we defined empirical social research as a quantitative approach which is done with the methods of empirical social research, mainly interviewing, observation and content or document analysis (cf. Obershall 1972) Since there is no list of scientists using this very approach in their research, we could have started our project with a list of institutions known to us as informants for our documentation. But this procedure would have led to some sort of snowball sample resulting in an unpredictable sample structure. Furthermore, we wanted to interview even those people who do social research but do not want to appear in our documentation so they do not inform us about their work. Eventually, we came to the conclusion that a sample drawn out of the subscribers of the ZA Newsletter would suit our needs best, for they may be assumed to be highly interested in the application of the methods of social research. It was equally important for our purpose that fifty percent of the ZA Users were also subscribers to the Newsletter. Sampling under the subscribers of the ZA Newsletter gave us the opportunity to get the

Figure 1


The Central Archive (ZA) User
»Research about
Sampling
2,765 Subscribers of ZA Newsletter
Random sampling
1,375 Addresses of social researchers
Identification of telephone numbers
1,258 Telephone numbers
During interviewing proved to be correct
1,114 Correct telephone numbers
762 Interviews conducted (68.4% response rate)
Screening
538 Interviews with social researchers
(25)
4.7%
(281)
52.9%
(225)
42.4%
Scientists who are not familiar with the ZA services
Scientists who made use of ZA service: ZA Users
Scientists who never ordered a ZA service

feedback of those who had already made use of the services of the Central Archive, and of potential users as well.

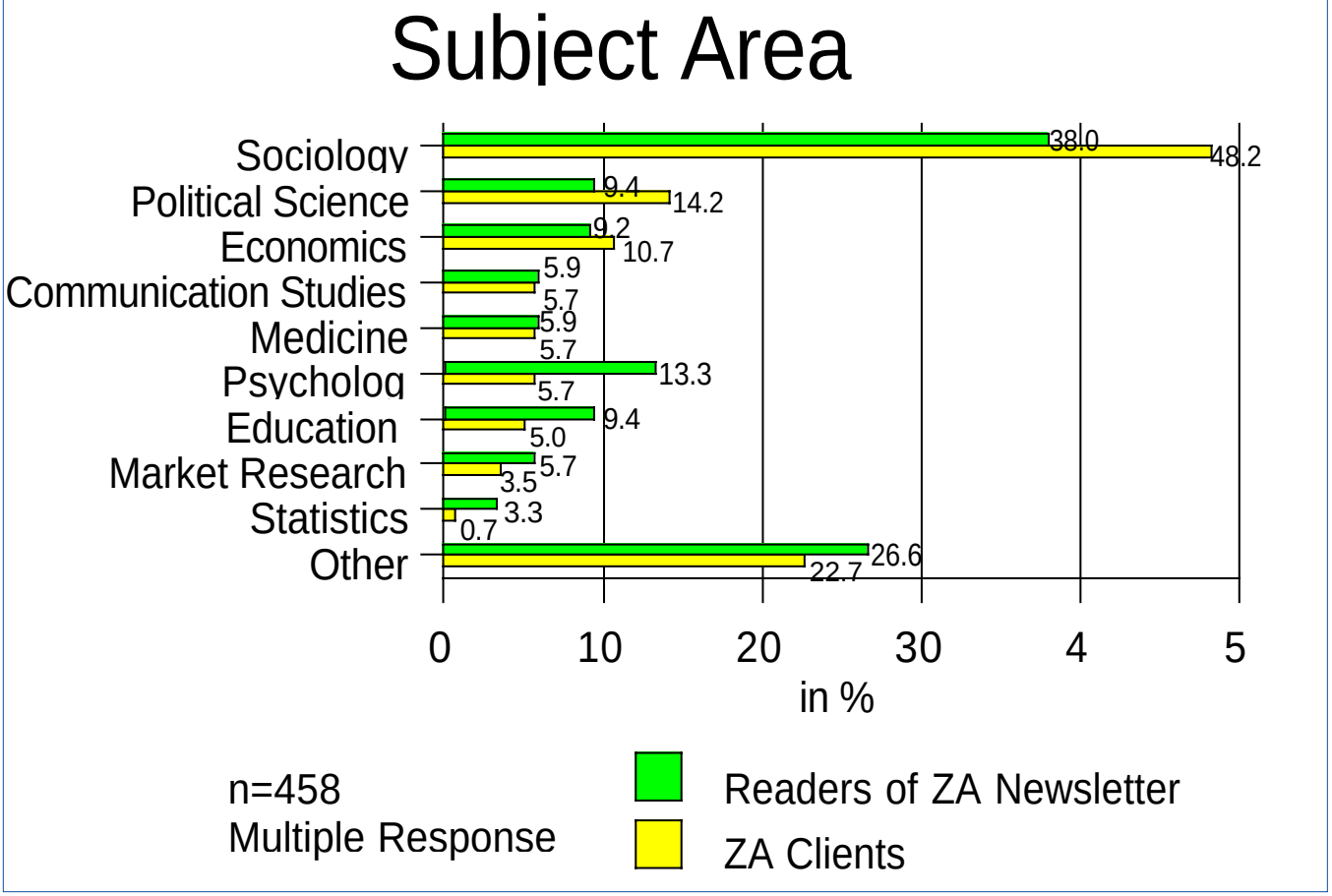
We planned to collect about 500 interviews. In advance, we estimated that about 25% of the subscribers were not directly involved in empirical research, e.g. librarians or local staff of university computer centers. From over 2,500 subscribers of the ZA Newsletter we drew a sample of 1,375 people. Since we only had their addresses we had to find out their phone numbers. Using a telephone directory CD-ROM and directory inquiries we managed to locate more than 1,100 potential respondents. The survey took place between Nov. 28 and Dec. 5, 1995. We conducted over 700 interviews. More than 200 respondents were not engaged in empirical social research so we finished with a sample of 538 social researchers. Figure 1 shows the details of the sampling procedure.

The interviews consisted of three parts. The first part dealt with the institutional affiliation of the researcher. The description of the actual empirical work formed the content of the second part and finally the respondents were asked questions concerning the performance of the Central Archive. In this paper we will focus on the description of the research community and the ZA clientele.

Characteristics of the ZA Clientele

Empirical social research is done in a variety of disciplines. The readers of the ZA Newsletter are heavily inclined to sociology as shown in figure 2. Two in five researchers (38.0%) belong to an institute which is situated in the field of sociology. The relevance of sociology is outstanding. It is mentioned nearly three times as often than is psychology (13.3%), which ranges second. Next follows a group of three fields with a proportion of ten percent each: Economics, Political Science and Education. These five subjects together form the core of the Social Sciences. Medicine, Communication Studies and

Figure 2

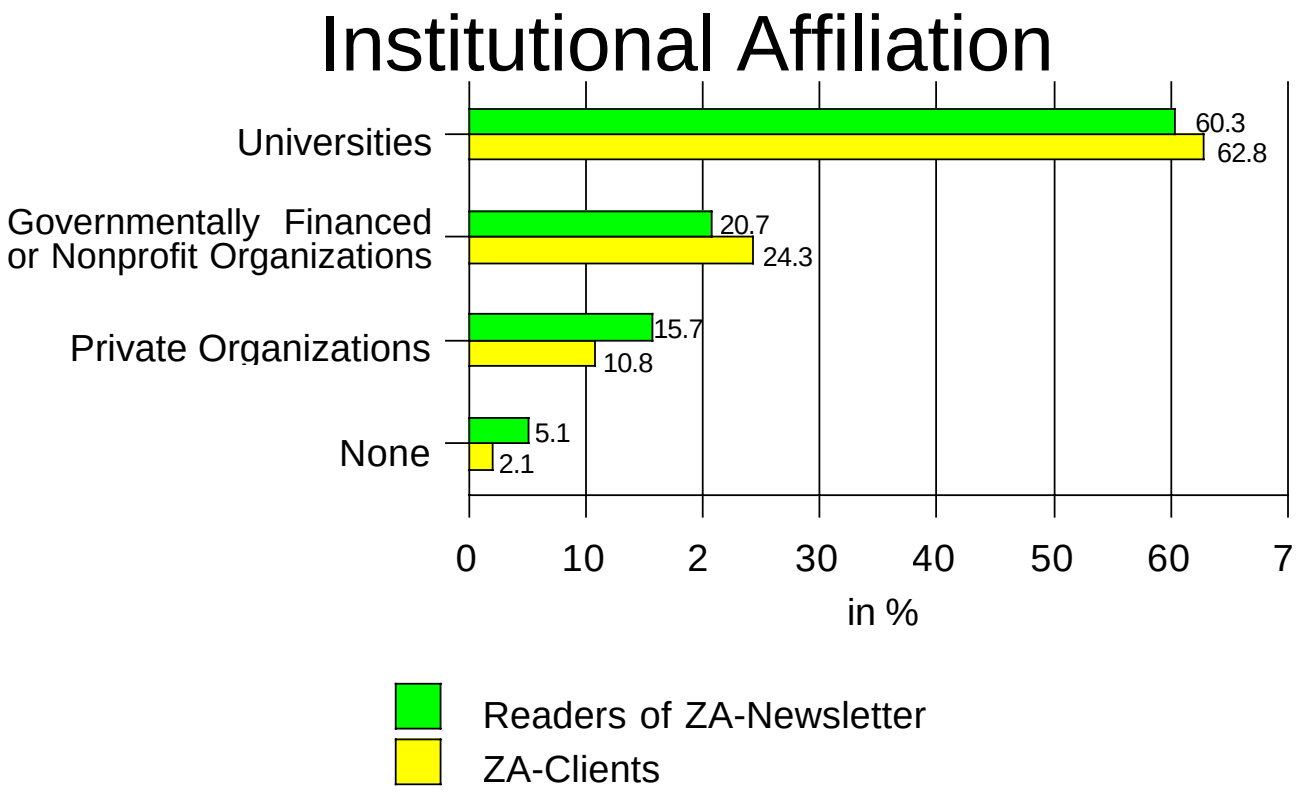


Market Research gain nearly 5% each. Statistics ranges last in this list of disciplines. There were 27 other subject fields named by the respondents, like Geography, History, Criminology, Social Psychology etc. But none gained more than two percent.

Let us now focus on those interviewees who have formerly received data sets from the Central Archive. Among those people nearly 50% belong to an institute of the research field of Sociology. The second most important discipline is Political Science. Psychology which was second among the readership of the ZA Newsletter now follows in the fourth position. This is due to the fact that psychologists do not deal that much with survey data. They prefer experimental data mostly collected from college graduates. They adopt the methods of empirical research but they normally do not need nationwide survey data. Political scientists on the other hand very often look for election data or data concerning the nationwide electorate. So it is not surprising that they range second as users of the ZA data service. Ranging third among the users of the ZA Data service are researchers belonging to institutions in the field of Economics. They gain a proportion of nearly ten percent. Psychology, Education, Communication Science and Medicine gain 5% each. Obtaining data from the Central Archive is of less importance for people belonging to Market Research Institutes and to the Statistics Branch. The former do not care much about surveys carried out by other scientists and the statisticians do not seem to be in particular demand of survey data.

As shown in figure 3 two thirds of the users of the Central Archive work in an academic institutional background. 20.7% of the respondents are employed in publicly financed research institutes. They consist mainly of federal research agencies, like the Bundesinstitut für Bildungsforschung, or governmentally financed large scale institutes, like the Max-Planck-Institute or the Wissenschaftszentrum Berlin für Sozialforschung. Only a small number of them are private nonprofit organizations like the Konrad-Adenauer-Stiftung. 13.9% of the readers of the ZA Newsletter work in private organisations in the commercial sector, mainly within the field of market research. There are only slight differences in the percentage between the researchers who have made use of the ZA data service and those who have not. While emphasis is on providing services for the academic community, the clientele also includes researchers from public administration and the media. In the ZA User Survey people

Figure 3

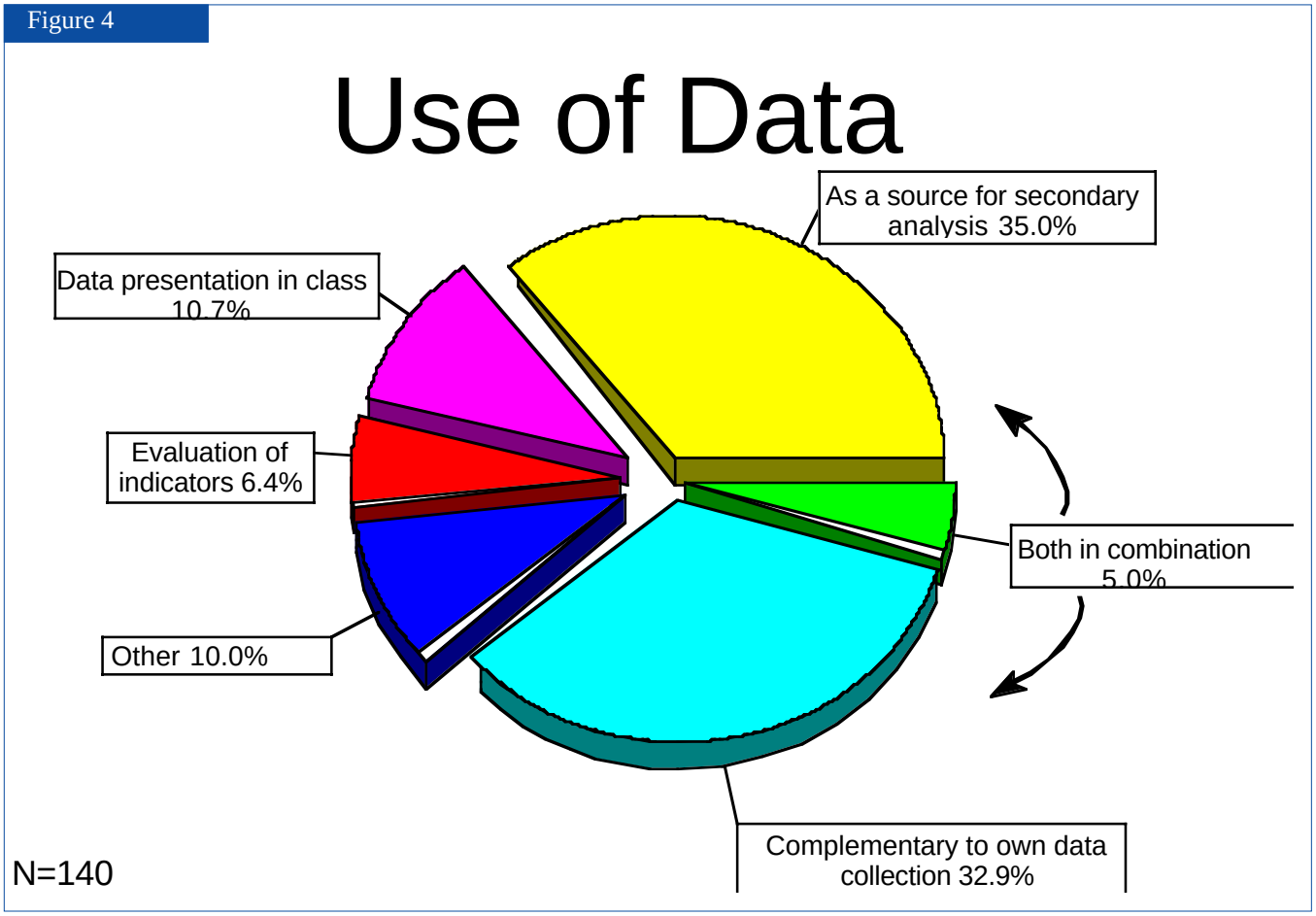


who work in the media are underrepresented to some extent. They normally do not subscribe to the ZA Newsletter because they are less interested in methological issues. The survey focused on people who are actually doing social research. But a remarkable part of the ZA clientele is mainly interested in getting information about the distribution of attitudes in the population.

For what purpose do the clients use the data?
We asked those persons who had at least once received data from the Cologne Archive what purpose they followed in examining the data. The question was posed as open ended question and respondents could give multiple responses. Nevertheless we got a clear-cut picture. There are two main intentions behind the ordering of data. One third of the population uses the data as a source for secondary analysis under a new research question. Another third employs the data as a supplement to own data sets. This completion was mostly sought in time dimension. In most cases this means that researchers who have already got data at the present stage want to make comparisons concerning the same population at some former point in time. The intention to conduct an international or intercultural comparison as a supplement to their own data was mentioned by a smaller fraction. 10.7% of the researchers used the data for teaching purposes in class and 6.4% used the data in order to evaluate indicators used by other researchers. Another 10% named other intentions, like information about the distribution of certain opinions in society, compilation of dissertation thesis etc.

Internet as a Research Tool
In the near future the already heavy use of the internet by the research community will drastically change and hopefully improve the conditions for doing scientific research. Computer mediated communication via email will expand the possibility to collaborate and interact with distant colleagues. The flow of information will accelerate and increase when scientists begin to use virtual arenas (multi user dungeons, news groups, mailing lists, virtual conferences, electronic journals etc.) to discuss and distribute new ideas. With more scientists using the internet the demand for quick and easy access to socio-economic data will increase. Therefore it is vital for data archives to know how many of their clients have access to the internet and for

Figure 4



how many of them it has already become an ordinary research tool.

In November 1995 when we asked German social scientists 56.7% of them had direct access to the internet from their working place and 35.2% made frequent use of it. This low percentage indicates that the internet-revolution has still to gain ground in Germany. Only one third of the subscribers of the ZA Newsletter have adopted the internet as a research tool. Broken down into institutional affiliation we find a big difference between researchers working in the academic context and those who work outside university. Already 68.4% of the respondents belonging to university institutes have access to the internet. This figure is nearly twice as high as in non-academic institutes. In private organisations only 32.2% of the employees dispose of a direct connection to the internet. In governmentally financed and nonprofit institutes the adoption rate is higher and amounts to 42.7%.

	Access	**Frequent use**	**Ratio** (use/access)	**N**
University	68.4	44.1	64.5	(320)
Government / Nonprofit	42.7	26.1	61.1	(119)
Industry	32.2	15.1	46.9	(73)
No institutional affiliation	33.3	26.7	80.2	(15)
	56.7	35.2	62.1	(527)

Table 1: Access and use of the internet by institutional affiliation

Access to the internet does not imply that scientists make use of the internet in their daily work. Only two thirds of the researchers with access to the internet adopt an internet based service as an ordinary research tool. There is still a lot of hesitation in exploring the usefulness of the internet. The ratio of use to access of the internet is nearly the same in university institutes and in governmentally financed institutes. But in the industrial context only one half of the people with direct access to the internet make frequent use of email, WWW, FTP or some other internet service. The situation seems even worse if we look at the percentage of scientists in the industrial context using the internet. Only 15.1% of them mention the internet as a useful research tool. In Germany, at the time of our survey, the internet was still an academic challenge. But we suppose that the low adoption-rates in the industry sector are only due to the fact that the internet is basically an academic invention. With a time lag of a few months we expect that researchers in non-academic institutes will use the internet with the same frequency as their colleagues in university institutes.

It is often assumed that one of the major impacts of the use of internet-services will be a gap between young and skilled persons who adopt the new technologies quickly and older people who will be excluded from the new information technologies (cf. Negroponte 1995). In our study we do not find support for the thesis of a widening gap between the generations. We do find differences between young and older scientists in the access-rates but there are no differences in adoption-rates. Since the internet has not yet arrived in non-academic institutes we confine the analysis of this thesis to institutes in universities. As shown in table 2 nearly three quarters (73.8%) of the young scientists (age < 40) have access to the internet. Among the older scientists (age ≥ 40) 65.6% dispose of a direct access. If we focus on those people in universities who dispose of a direct access to the internet the percentage of frequent users among young scientists is nearly the same as among older scientists. 65.9% of the young scientists make frequent use of internet-services compared to 64.0% of older scientists. Thus the adoption-rates in the two age-groups are almost identical. If there was an effect of age on adoption we would expect a much higher adoption-rate among the younger scientists. We can conclude from these findings that differences in the percentage of internet-users between age-groups are only the result of differences in access-rates. Presumably older people get access to the internet at a later stage of the innovation-process than younger people. But if there is a direct access to the internet the same fraction of researchers will use the internet in the older and in the younger generation. This indicates that the use of internet-services is already a valuable research tool and that it depends mostly on the institutional context in which the scientist works whether he adopts internet-services or does not. But we can expect that in the near future the use of internet-services will be as natural as that of personal computers is now. Therefore archives have started to prepare themselves for the coming internet age.

	Access	Frequent use	Ratio (use/access)	N
under 40 years	73.8	48.6	65.9	(107)
40 years and older	65.6	42.0	64.0	(212)
only university institues				

Table 2: Access and use of the internet by age

Desiderata and Recommendations of the ZA Clientele

At the end of the interview we asked the respondents if there was anything that the Central Archive should improve or which services should be introduced. A large fraction of the respondents commented on the information policy. They wanted information to come more frequently and more directly to their working place. Another group recommended to give more detailed information. Some researchers gave the advice to foster the effort of addressing people outside the core-disciplines of Social Sciences. The second main topic was the dissemination of data. Many of the users wanted to have quick and easy access to the data via FTP and to have more data sets made available on CD-ROM. Some mentioned the present pricing policy and expressed their wish for reduced charges for data access. As the third main topic some users pointed to topically focused data collections and to a better and easier access to international data. Some researchers would be glad if we could offer more surveys from the field of commercial market research and if we could offer more recent data.

Facilitating Access to Comparative Data

Using the Internet and Publishing on CD-ROM

The central archive has always made the effort to expand its services and to use new technology to disseminate data and to communicate with its clientele as shown in the first chapter of this paper. In response to the answers the researchers gave in the user survey the central archive will strengthen these efforts. We will spread information about new data sets and other relevant news through a mailing list. More detailed and always up-to-date information can be found on our web pages (http://www.za.uni-koeln.de/) just now and will be developed further in the near future.The question text, codebook information and marginal distributions of the International Social Survey Program (ISSP) e.g. are searchable in the Internet under WAIS. Soon data will be accessible by FTP-Transfer. Furthermore we will enlarge our collections of data sets available on CD-ROM. Third we are engaged in a multinational project, named ILSES which aims at the development of an integrated library- and dataservice. Finally, we have installed a scientific laboratory equipped with all the infrastructure needed for comparative research. Also, the European Data Archives are creating a virtually integrated catalogue of their holdings, accessible via Internet.

Social Research Labs / Large Scale Facilities

As we start aging in the virtual scientific community we learn that the dream of information and data traveling to any place in the world is becoming true, yet it does not provide the ideal research environment for comparative research. Researchers may be well informed about major events in their societies that might have had an impact on attitudes and behavior of respondents. The further we progress in time, the more interpretation knowledge must be transferred to the collective memory of researchers in order to provide the context that was decisive in the phase of data collection. This is particularly relevant for information about other societies which are not part of the daily information routine of the researcher.

Contextual information, cultural background and historic knowledge which may be necessary for sound interpretation of empirical evidence do not automatically travel with the collection of data sets from different societies . Bringing together relevant data is still an exercise in systematic selection of comparable variables, data recoding and overcoming transboarder data flow hurdles emanating from data protection and data access regulations.

A response to the needs of comparative research may be social science data labs, in which all relevant data and information for a particular research field is at the fingertips.Over the past two years ZA has created a EUROLAB which provides access to major comparative studies and related background material (e.g. party manifestos, media-reports, event data bases, fact books etc.). The study collections include among others the International Social Survey Program, the Eurobarometers and major election studies on national parliaments in Europe.

The Standing Committee for the Social Sciences of the European Science Foundation had pointed to the need for better integration of the European data base and brought to the attention of the European Union that social science data bases are the equivalent to large scale research instruments of the natural and technical sciences. A study panel proposed to acknowledge

the need of social research for Large Scale facilities where researchers not normally having access, could come to profit from available resources. The Institute for Social Sciences in Essex and the Zentralarchiv in Cologne received recognition as first Large Scale Facilities in Europe under the Training and Mobility Program of the EU. This will allow to cover travel and subsistence costs for scholars from EU member and associated states who want to make use of these resources subject to approval of their applications.

Over the next three years this will allow the ZA to have scholars and research teams not only making use of the resources, but at the same time enjoying truely comparative research by bringing together their specific knowledge about different countries. Thus they can help to validate data and background material. Ultimately this will improve the research resources for the scientific community at large, since validated data and background information may be compiled in knowledge basis for general distribution.

References:

Lazarsfeld, Paul F., 1962: The Sociology of Empirical Social Research; in: American Sociological Review, Volume 27, No. 6

Lazarsfeld, Paul F., 1972: Foreword; in: Anthony Oberschall (Ed.): The Establishment of Empirical Sociology: Studies in Continuity, Discontinuity, and Institutionalization; New York

Marcson, Simon, 1972: Research Settings; in: Saad Z. Nagi and Ronald G. Corwin (Ed.): The Social Contexts of Research; London

Negroponte, Nicholas, 1995: *Being digital*, New York

Oberschall Anthony, 1972: Introduction: The Sociological Study of the History of Social Research; in: Anthony Oberschall (Ed.): The Establishment of Empirical Sociology: Studies in Continuity, Discontinuity, and Institutionalization; New York

Zuckerman, Harriet, 1988: The Sociology of Science; in: Neil J. Smelser (Ed.): Handbook of Sociology; Newbury Park, London, New Delhi

1.Paper presented at the annual meetings of IASSIST, May 15, 1996, Minneapolis, Minnesota.

2. International Community for Research into Elections and Representative Democrarcy