# Data Services in a Non-Academic Environment

*by Ineke A.L. Stoop[1] ,*
*Social and Cultural Planning Office,*
*The Netherland*

*The Social and Cultural Planning Office (SCP) is a Dutch government research agency. It contributes to policy making and appraisal in The Netherlands by producing a coherent picture of the state of social and cultural affairs. The production of such an overview requires an ever increasing amount of empirical data. During the 23 years of SCP's existence several hundred data files have been edited, stored and made available to the researchers. Data come from a large number of providers, such as Statistics Netherlands, commercial public opinion and market research firms, ministries and the Dutch data archive. Secondary use of data is strongly advocated Data services in a small non-academic organization differ by scale alone from data services in large academic institutions. Some problems, however, are universal.*

## The Netherland's Social and Cultural Planning Office

In 1973 the Dutch government founded the Social and Cultural Planning Office (SCP). SCP is an inter-ministerial scientific institute which conducts independent research on social and cultural aspects of government policy. Research findings are mainly intended for the Government, the Upper and Lower Houses of Parliament, senior officials of ministries, and for professional and academic staff in the public sector and academic circles.

SCP publishes the *Social and Cultural Report* every two years in September. An English version is available. The Report describes the social situation of the Dutch population and their attitudes to government policies on social and cultural matters. It covers health care, the social services, employment, social security, justice and criminal procedures, housing, education, leisure, the media and cultural affairs. Other publications deal with one of these subjects or with issues like the distribution of income and wealth, the economics of the public sector, the position of target groups or local policy matters.

SCP employs about 50 researchers, each an expert in one or several subject areas. In addition everyone has to be aware of current policy issues and to be versed in empirical research and data analysis. The department of Information Services supplies them with data, software, hardware and statistical consults.

## Information Services at SCP

The department information services at SCP (Information and Automation, I&A) comprises a relatively extensive group of people (10) for a relatively small number of researchers (50). Information services covers a wide area, though: information technology, communication, statistical consulting and data services.

Information technology at SCP stands for providing researchers and administrative staff with PC's software and information systems. All PC's are linked in a Novell network. Most users work in a Windows95 environment. Most statistical analyses are processed on dedicated PCs for running under OS/2. I&A personnel take turns in staffing a help desk and are also responsible for internal and external electronic communication (E-mail) and Internet facilities.

Statistical consulting encompasses a wide variety of activities, such as assisting researchers using SPSS, creating scales, and developing models, suggesting statistical methods, arranging courses and giving in-house demonstrations. When mounting surveys, methodological issues have to be taken care of, related to the determination of adequate sample sizes, weighting data, inputting missing data and wordy questions. The area of survey methodology might also be considered as part of the data services provided by I&A.

## Data Services at SCP

Data services are part of the information services at SCP. Researchers need PC's to operate, statistical programs to manipulate, hypotheses and above all ... data to contemplate. As Sherlock Holmes said, in **The Adventure of the Copper Beeches:**

> *"Data! Data! Data! I can't make bricks without clay."*

The first task of data services at SCP is determining what kind of data is required in a particular study. Researchers are urged to use existing data. For this purpose SCP has a wide range of data files available on-line:

- longitudinal series of population surveys from Statistics Netherlands on subjects like well-being and health, employment, education and expenditures;
- longitudinal series of surveys (partially) funded by SCP on facility use, time budgets, cultural changes and opinions;
- international surveys (Eurobarometer waves, International Social Survey Programme (ISSP) modules);
- institutional data (hospitals schools);

- one-time data (troubled youth, opinions on social security,
  environmental behavior);
- demographic data.

Data from Statistics Netherlands are obtained by way of the Scientific Statistical Agency, an intermediary institution acting as a go-between for (academic) researchers and Statistics Netherlands. Data from Statistics Netherlands are secured from disclosure by leaving out regional variables, minority group indicators and other identifying characteristics. In some cases only outdated survey waves are to be had, from other surveys recent data files are available. All things considered, Statistics Netherlands takes a co-operative line in providing research institutions with data. Other providers can be much harder to cope with. Sometimes data are withheld for strategic reasons or privacy considerations. Other times they are very expensive.

Mounting surveys funded by SCP start out;by selecting the research institute. In some cases a particular research institute has carried out a survey for years and is thus an obvious choice.

The preparation of a new survey requires a lot of time. Every detail requires attention. With respect to existing surveys, however, where comparability is paramount, routine set in as sampling procedures and questionnaires content are changes as little as possible.

Once data files have been obtained, they have to be edited and standardized. Weighting and imputation of missing data may have to be performed. Data files are all described in the Meta Information System SCP (MISS). MISS contains information on study, survey and variable level. Inclusion of the questionnaire and keywords may be possible in the future.

When data files are 'ready' they are all on-line available for researchers to analyze. In most cases these analyses are run on dedicated PC's using SPSS-X. SPSS for Windows, however, is gaining popularity. I&A comes into action again only for advanced data usage, e.g. when data files have to be matched synthetically or are molded into models. Of course, researchers often require assistance, either when they have questions about the survey or statistical or methodological questions.

During the years SCP has acquired a mountain of data. Sometimes data files are small, other times large with respect to number of cases, number of variables or complexity (multi-level data). Data files contain information on health, well-being, time budget, employment, social security, education, housing, media, expenditures, and so on. Data objects range from individuals and families by way of organizations and communities to postcode regions and EC-countries. Variables indicate opinions, describe behaviors or contain measurable quantities like expenditures, time

budgets, possessions (persons) or personnel, funding, housing (organizations). And finally, they are supplied by a large group of providers, as outlined above.

**Quality Control**
Sound government policy requires high standard research and high standard research requires first-class data. Unfortunately, social research in the Netherlands suffers from some major drawbacks, summed up by SCP as *the hole in the data layer* .

The first flaw in data collection is the poor response to surveys in the Netherlands, even in face to face research. Mind that there is no census in The Netherlands and that participation in official surveys is never mandatory. A response rate of 40% is no exception. Response to surveys is selective in a way not always to be remedied by weighting. The elderly, people from minority groups and inhabitants of inner cities are often underrepresented. Weighting by using postulation characteristics may partly compensate for this kind of nonresponse. Nonresponse may be related to the topic of the survey. In this case weighting will not always be possible. And finally, whereas SCP publications allegedly cover the entire population, actually, a lot of unsuitable cases are excluded from general social surveys. Inmates of prisons, hospitals, homes for the elderly and mental institutions are not interviewed. People who cannot see, hear, speak, write or read are frequently not able to participate and thus registered as nonresponse. Nonnatives may be under represented because of language problems. People either too young or too old are excluded as a standard procedure. Yuppies are never at home, the homeless cannot be reached, the mentally retarded may not understand every question and drug addicts live in another world. Though some of these problems may be solved by interviewing by proxy, a lot of people quietly drop out of the sample frame.

Nonresponse and noncoverage are tackled in three ways:

- firstly, take pains to boost response,
- secondly, find out in what manner respondents differ from nonrespondents,
- finally, compare results with data from other sources and, possibly, use data from public records.

Besides having an acceptable response rate, data ideally should come up to the following standards:

- have national coverage, and thus allow for regional comparisons;
- cover the entire Dutch population, and thus allow for comparison between groups;
- come from recurrent surveys (containing identical questions), and thus allow for comparisons over time;
- be fully documented.

*Strengths and weaknesses*

SCP publications are based on a firm empirical tradition. They nearly always contain empirical analyses. This gives SCP a clear advantage as compared to other institutions which mainly rely on material published elsewhere. Another reason why SCP publications tend to get a wide press coverage is that they generally bear upon current policy issues.

The main strength on SCP thus may be summarized as a tradition of empirical, policy relevant research. A major weakness originates from the size of the institute (rather small) and the range of subjects SCP covers (fairly wide). Researchers sometimes extend deadlines in order to create a better product. In other cases they desperately try to finish on time, putting up with data files not being analyzed thoroughly and only presenting cross tabulations, where as publications may be delayed, or data files may be analyzed insufficiently. Finally, analyses are not always documented in full. Reproducing analyses of former waves of a survey may thus require a laborious process of reconstruction.

Data services at SCP have their own strengths and weaknesses. After almost 25 years SCP has obtained a hoard of data on almost any social and cultural subject. Old files can be used to outline developments. New files can be used to illustrate current dilemmas. As said earlier, secondary use of data is advocated. Central storage of a single copy and on-line availability guarantee data integrity. Editing and standard format prevent errors. The main strength of data services at SCP, however, is that data services, information technology and statistical consulting are thoroughly integrated within one department . Conflicts of interest can be solved at the root, data collection can profit from methodological expertise and computational facilities are geared to a research environment.

Still all is not well. Firstly, much of the information about data files only exists in the head of the data people. Documentation is often incomplete. Questionnaires get lost or exist as a computer program only. And secondly, the tradition of secondary analysis does not come naturally to researchers, especially when they come straight from university. Most of them need time to get rid of the academic attitude: *get new data to test new hypotheses*.

**Threats and opportunities**

In its more complacent moods SCP might consider itself a victim of its own success. Competitors seem to be attracted to SCP research methods. Universities start proclaiming the use of secondary data and building large data bases to be used and reused by students instead of collecting data time and time again. Statistics Netherlands wants to get rid of a large series of unattractive publications filled with huge tedious tables containing detailed statistical data. Instead it wants to publish semi-glossy compilations of statistical data, including pictures and sexy graphics. These new publications may appeal to the general public. SCP, on the other hand, was and still is extremely fond of huge tales and extremely detailed statistical data to be stored in SCP data bases and use for further analysis.

Another threat to SCP and its data services is the aforementioned hole in the data layer. Data deficiency is, among others, caused by a variety of restrictions of data use. Embargoes, privacy regulations and disclosure avoidance procedures limit the availability of data. The progressive use of public records by Statistics Netherlands, a development to be applauded in research itself, also threatens SCP data services, as administrative data will not be available for research outside the statistical office. As commercial data providers or associations of nonprofit organizations start to appreciate the strategic importance of information, the availability of data will fall off and the price will go up. And, as said earlier, the growing nonresponse rate in the Netherlands is a serious threat to data quality.

The final threat mentioned here will be considered an opportunity by a lot of people. In the old days, statistical programs, like SPSS required a set up which was generally stored for later use. Data files had to be stored centrally as local PC disks were much too small. SPSS for Windows does not require a set up; operations are performed directly on the data file. Analyses results can be cut and pasted into the final text. Recorded files for individual use can analyzing data, much more discipline is required to control quality. The development of explicit quality control procedures and tools at SCP is meant to help researchers to maintain this discipline.

Not all projects are bleak, though they require renewed efforts to keep data services at SCP up to par. Luckily, some new phenomena may help. The electronic super highway may be seen as an overload of use information. Still, when used prudently, it may offer useful tools for communication by electronic mail and web sites, for data collection and for data acquisition. Though SCP words for the Dutch market, international data can put this market into perspective. The Internet may help to obtain international data. And finally, the Internet may help in developing meta data at SCP, as it opens us international efforts and international standards for meta data and documentation. In its small way, SCP is starting to Weave the Web.