Providing Local Data Services

by R. de Vries [1]
Steinmetz Archives/SWIDOC.
Amsterdam, The Netherlands

The Steinmetz Archive, as the Dutch national data archive for the social sciences, is a somewhat special case in the context of this workshop. We function both as a data clearinghouse and as a data library per se. As data library, the archive operates in an area that, in other countries, would be seen as a "regional" (as opposed to "national"); the following are some reasons why the Steinmetz can serve as an example of a data library providing local data services.

- dependency on someone else's computer center. I.e., decisions on installation of software packages, policy on how to handle mass storage problems and the safekeeping of magnetic tapes, participation in networks, etc., are all beyond our direct control; we can argue, but have no real influence on such decisions, let alone the means of implementing them on our own.

- dependency on the willingness of research organizations and their funding bodies to deposit data (survey or otherwise) in the Archive's holdings. Again, we have no financial means with which to buy large datasets, nor the manpower, in the case of published statistical data for example, to generate new data from published sources.

- a small staff (5).

- a strong emphasis on documentation and reference service, aided by a reference database containing study descriptions of every stored dataset. "Documentation" here refers to the original questionnaire, research report, print-outs of frequencies, etc.

- given the national role of the Archive a less than desirable situation, but in the context of "local data servcies" quite reasonable: we cannot give access to our holdings through a network, nor is the reference database available online. Data exchange is via magnetic tape, and available datasets are brought to the attention of potential users via regular newsletters and a published catalogue. A service that is in my opinion typical of local data services, data exchange on floppy discs, is possible but we have hardly any experience with it as yet.

If one sees a data library as a "local" service backed by a central data clearinghouse or central acquisition and processing centre, then indeed one expects an organisation with a small staff, using external computers and software, concentrating on reference as well as actual dissemination in as friendly a manner as possible, generating subsets, documentation and other special requests. Also, seen geographically, the data library should be within one day's travelling distance for its users, for consultation purposes. Given this definition, the

[1]Paper prepared for presentation at the IFDO/IASSIST Conference, Workshop on Data Services, Amsterdam, May 20 – 23, 1985.

Steinmetz archive does serve as an example of a data library.

Are there any other local data services in Holland for the social sciences? For survey data: no. For statistical information at the level of cities or regions: yes. There are several specialised databases owned by government organizations for planning and policy making, and by university departments, for research and training. These are, on the whole, "local" in the sense of being accessible only to their own people, not to outside researchers, either academic or otherwise.

In the second part of this account, I will briefly outline our approach in the following areas: data acquisition, data dissemination, storage and maintenance of data, documentation and reference services.

Data acquisition is accomplished by routinely checking registers of ongoing research, social science periodicals, and reports of finished research. This is facilitated by the Steinmetz Archive's participation in the Social Science Information and Documentation Centre. There is no human network of researchers or fund raisers in the field who could report to the Archive interesting projects or data. (Nor would I would expect the kind of data library that provides local data service to rely on such a network for data acquisition.)

Dissemination. Users in Amsterdam, where the Steinmetz Archive is located, have direct access to the data through the local university-owned computer centre (SARA). From a local terminal, a user can get a copy of a dataset by simply starting a job, that has only one variable: the Steinmetz number given to the particular dataset. Central logging of these jobs and who has started them, is automatically reported to the Archive, thus enabling a monthly overview of this type of usage. Other users receive the data, and often an SPSS setup, on tape. This arrangement makes, of course, no provision for users without at least access to a minicomputer with a tape drive and statistical package, such as SPSS. For example, we are unable by these means to provide service to schools. Data transfer to such users should be through floppy discs, a service that we have not really started yet.

As an archive, with an obligation to disseminate 10 to 15 year old datasets, it requires that we have strong **Data storage and maintenance** systems. This we achieve with a system of multiple tape backups and a tape refreshing scheme to guarantee that no tape is physically more than three or four years old. All tapes are stored in the computer centre. One might expect a data library per se to be more relaxed in these matters; whatever gets lost can be replaced upon request from a central data organisation, but this is not so in our case. Developments in "laser disc" technology could ease local mass storage problems, and at the same time ensure long term reliability.

How is the user introduced to these masses of carefully preserved data? Through **Documentation and reference services.** The Steinmetz Archive, as mentioned previously, helps users find data suitable to their needs through a catalogue, which is easily, and regularly, produced from a reference database, various indices on microfiche and paper, and through the original documentation produced by the principal investigator. Introductions to the principles of empirical research and the data available for secondary analysis by means of making "teaching packages" available and giving lectures at schools and colleges, are other means to assist users. The Steinmetz does not give lectures but does offer a teaching package together with the relevant data; the package was developed by an outside institution. Data libraries, which should need to put less effort into such activities as acquisitions, processing and maintenance, might profitably put more effort into actively getting users acquainted with computer-assisted analysis, data sources, etc. ¤