
Social Science Data and the Usage of Data in Hungary

by Prof. Tamás Kolosi¹
Social Science Informatics Center
H-1027 Budapest
Frankel Leó u. 11

The collection of data for social sciences has a rich tradition in Hungary. Hungarian sociology – apart from short periods after the turn of the century and after the Second World War, has undergone rapid changes of late (most notably from the sixties on). Hungarian social statistics undertook extensive and very profound data collection operations as early as the period between the two world wars. A striking example of this state of affairs, is that it was in Hungary, for the first time anywhere, that inter-generation mobility data were collected on a nationwide representative sample (1930).

There has been a very close connection between social statistics and sociological research. The mobility surveys repeated at ten-year intervals (1964, 1974, 1984), the income assessments conducted every five years, and data from the time balance surveys in 1964 and 1977 frequently become data sources for not only Hungarian, but also for international comparative sociological research. A particularly rich collection of data is available to us as regards social stratification and inequality. The stratification survey work performed by KSH (Central Statistical Office) in 1963 for the middle-eastern-European region was the first one to provide an empirical base, on a large nationwide sample, for criticising the ideologically enforced Stalinian class model.

¹Presented at the IFDO/IASSIST 89 Conference held in Jerusalem, Israel, May 15–18, 1989

The largest-scale empirical data collection undertaken thus far was in 1982–83. Through the combined effort of several research institutes, this data collection went on for one and a half years, during which nine different questionnaires were completed on the same sample and in which the data (2000 variables or so) almost comprised the entire matrix of measurable social inequalities.

Currently, there are three large-scale data collection systems in Hungary. The Unified Population Data Collection System of KSH which dates back almost twenty years. Beyond regular household and manpower surveys, this system comprises collection of such large-scale social statistical data on mobility, time use and prestige of professions. However, these large sample (about 50,000 persons) data collections are in part very expensive and, on the other hand, due to their statistical character, are not fully suitable for the purposes of sociological research inquiring into, or wishing to inquire into finer social structures.

These considerations motivated us to establish the so called Joint Survey System of TARKI (Social Science Information Centre) from 1982. This is a nationally representative sample involving 6000 persons in which there are omnibus-type survey periods twice a year. The omnibus character means that the different research units may either turn in their questions or their questionnaire blocks, or they may even occupy an entire questionnaire. The data collections are then coded after arranging the questions from different research aspects into a unified questionnaire. The researcher then gets back a labelled SPSS system-file within six months. The sample is replaced after five data surveys and the questions are then collated with the data which originated from within a two and a half year period. Proceeding like this makes data collection not only substantially cheaper (it is sufficient to call for standard information only once) but provides an immensely greater scope of possibilities for analysis. In the framework of the system, data were collected for a number of widely different purposes, from the inequality system to the economy of households, to the study of law-consciousness and to the research of cultural stratification. Inquiries are made every year on the basis of this system in connection with the questionnaire block of the Social Survey Program which now extends to nine countries.

The third system is operated by the Hungarian Public Opinion Research Institute and is beyond just political public opinion research, this institute is also concerned with investigations relating to the acceptance of media communication, which is primarily a regular monitoring of TV watching and radio listening.

In addition to these large-scale research activities, there have been many smaller, special purpose sociological data surveys over the recent thirty years. But for a long time, the usage and utilization of data was essentially limited to analysis and publication by principal investigators. As a result of this, the paradoxical situation has arisen in Hungary that apart from some analyses which aroused international interest as well, today we possess much richer data sources than what are indicated by the combined achievement of Hungarian sociology. With a bit of malice one may say that it was not the original character of analyses and the scientific level that, primarily made somebody a well known researcher, but the personal ability and cleverness to acquire the financial means required for collecting the data.

With a view to altering this situation, in 1985 we started to organise a Hungarian social science data bank in the framework of TARKI which is a joint venture of the five major research units in the country. (Currently there are fourteen proprietor-institutes, and almost all the Hungarian social

science research bodies are associated with us as users). Our data bank was developed from drawing on the experiences of the data banks in Cologne, Colchester and Amsterdam. Furthermore we established the connections by becoming members in International Federation of Data Organizations (IFDO) and CESSDA. We have data from 250 research surveys available in our archives, in addition to the raw data, we provide access to the files as an SPSS system file. Although a majority of the data from earlier research is not available in an archivable format, we do possess a partially completed materials originating from work from the end of the seventies. In addition to the sociological data collections, the social statistical data surveys by KSH (including data from the latest two population censuses) are also available in our data bank.

The data bank is entirely public and access to data conform to international standards. The SPSS-format data tapes or original data are made available to university students free of charge and to professional researchers for a small fee. Users may also arrange to conduct their analyses on our computer facilities if they have no access to computers.

As a new service, we undertake to carry out secondary analyses of the available data for customers inexperienced in sociology such as the press and government agencies. We will also provide customers with material in tabulated or graphic form or in the form of short analyses.

The development of factographic data bases has also been ongoing during the past two years. At present two such data bases are widely available.

The data base called TARKA is usable on a IBM-compatible PC if the user has an SPSS PC program. Essentially, it is a handling system whereby the SPSS system files of different subjects stored on HD floppies are integrated into a unified data base.

The TARKADAT data base is another venture by us but on a grander scale. In this system, the data base is accessible on-line through a science based, packet-switched Hungarian data network (currently about 70 institutions share the system, – research institutes, universities and libraries and some industrial companies engaged in major development jobs as well) and it contains all the information from all the data surveys performed within the TARKI Joint Data Survey System (this today amounts to almost 8000 information with a combined extent of 8 megabytes).

The system is composed of two parts and its use requires no special information or technical knowledge whatsoever. The first part makes it possible for the user to select the required information from a tree-structure thesaurus in an interactive way and to determine the necessary sample. Data processing is performed in the second part. There are two possibilities here: simpler statistical operations (such as frequencies, descriptive statistics, crosstabulation, correlation, grouping means and standard deviations) can be performed also on-line. (Since the data are labelled at both variable and value levels, this kind of processing leads to output of final tables). There is, however, a possibility for the user to have the output from the first step printed into his own file (either as a raw data file, SPSS system or export file) or to arrange to have the data written to magnetic tape or floppy disks. In this latter case, the data is available for further processing in an unrestricted way. (The system has an automatic billing for data use charges). Access to the Hungarian packet-switched data network, via internet, has been scheduled to be made available by mid-summer of 1989. This will allow the international social research community relatively easy access to our TARKADAT data base.

This is important not only for the expansion of international relationships but also because the use of data banks and data bases in Hungary is spreading rather slowly. This is strikingly illustrated by the fact that in 1988 TARKI had a higher file trade abroad than in Hungary.

International experiences indicate that in the first years subsequent to organising data banks the initial number of users is rather small. Then there is a period of moderate usage. Studies indicate that about 10–15 years are necessary for the system to reach full operational status. The twenty-three data files circulated by us in 1988 corresponds to the expected useage statistics of a young data bank. Nonetheless we are eager to expand our services and useage. Attempts are therefore made to develop as user-friendly systems as possible. To this end, intensive training activity is also carried on.

There are very rich data sources currently available in Hungary. In our view, access to these data is quite easy both technically and legally. The lack of expertise by users, the low level of computational and secondary analytical cultures and the shortage of financial funds available to potential users, which is due to economic crisis, all combine to result in a still relatively poor utilization of our data holdings and the research possibilities available therein.□