

# Building Regional Data Files in the Federal Republic of Germany

---

by Edwin Ferger <sup>1</sup>  
Zentralarchiv für empirische Sozialforschung  
Universität zu Köln

---

The Zentralarchiv für empirische Sozialforschung, Universität zu Köln, has engaged itself in the creation of the German part of the 'Joint European Time-Series Data Base'. The development of this data base for comparative social science research is a project by the IFDO (International Federation of Data Organizations for the Social Sciences).

In the following paper, I will describe some of our experiences building regional data files, I will point out where we stand now and some problems we face.

---

<sup>1</sup>Presented at IASSIST/IFDO International Conference May 1985, Amsterdam

---

## Experiences in the Past

Although the Zentralarchiv's primary tradition is to file surveys, there are several election studies in its holdings containing aggregate data; for the most part these are variables of basic demography and occupational structure, as well as political variables.

Many talks with social scientists have shed light on the fact that research interests are diverse. However, there is a strong desire for a common dataset on the lowest possible regional unit. Some regional data files exist, but access is sometimes restricted; data in different files are seldom comparable because they are valid for different demarcations or time points. Data comparability also suffers from boundary or definition changes. Some other problems apply to special circumstances in the Federal Republic of Germany.

During the past years many data files were created in Germany. Unfortunately the data are as heterogeneous as the projects and the underlying research interests from which they result. Although the data files contain precious data, they only have a few variables in common. Knowledge of the existence of data files seems not to be widespread, as seen in the duplication of effort that is usual because of the many authorities and researchers that use large amounts of manpower to create machine readable data sets with more or less overlapping variables.

Usually, existing data files are not easily available. Data sets under the jurisdiction of authorities may theoretically be accessible to the public, however, to get the data is another matter. The transmission of data to a third researcher or research institute almost always involves prolong negotiations or severe restrictions.

One of our aims is to build a *data pool* that can be enlarged by the addition of data from different sources. There are some studies already in the Zentralarchiv holdings which contain aggregate data. These sources should be usable for the pool, together with the data we transcribe from printed material to a machine readable format.

At the present time efforts are made to complete the IFDO list of variables as far as possible for the Regierungsbezirk level and later on for the Kreis level.

We try to collect data from other researchers' projects and studies not only for these high levels of aggregation but also for smaller units. The intended combination of data for different regional levels makes data available for multilevel and contextual analyses.

It seems to be reasonable to use data from these sources as the nucleus of the data pool and to complete the gaps that remain in the IFDO list of variables. Most of the variables to be added concern settlement structure, education, income, production, health and household conditions.

When the data are merged, they should be subjected to reciprocal control routines. Regional data are hierarchical in structure across the different levels of aggregation and thus can be subjected to control procedures. These can use sum, line or row percentages. It is impossible, however, to execute checks "totally automatically". Security restrictions and unavailability at lower levels also hinder one from coming in contact with the data. Other data, for example relational data, cannot be aggregated.

From our point of view, control routines are necessary in any case; whether data from existing files are merged or transcribed from printed sources. If machine readable data are merged, one should not start with the premise

that existing files are faultless.

Considering the diversity of data files in the university sector, especially their regional and temporal coverage, efforts should be strengthened for the better use of the holdings of the statistical offices of the Länder and the Federal Statistical Office. The development of data banks or statistical information systems in most of these institutions during the last decade seems to open new possibilities for such improvement. Experience will show if these data banks can be used to construct regional data files covering the whole territory of the Federal Republic. There have been some disappointing experiences in the past - the fact that the majority of files have been constructed by transcribing data from printed sources, demonstrates this. On the other hand, the willingness of the authorities to cooperate with the scientific community has increased recently. The data banks and information systems have been established with enormous financial and personal expenditures, and users are now made welcome; a large number of users will mean that their efforts have not been in vain.

---

#### Where we stand and the problems we face

In our efforts to transcribe data from printed sources to a machine-readable format, we have built four SAS-files, following the 'List of IFDO variables' (LOIV). We have created one file for each of the three census years: 1950, 1961, and 1970, and another file containing annual data for the variables, area, population, births and deaths. These files have been sent to the NSD. (*editor's note: Norwegian Social Science Data Service*).

Accordingly, three questions arise:

1. Why didn't we obtain a data file with the full set of IFDO variables for 1980?

Unfortunately the latest census in the Federal Republic of Germany was conducted in 1970. About 1980 and for several years thereafter a broad public discussion on data protection took place, which prevented a new census. Only recently the federal government decided to postpone the next census until 1990.

The statistical offices in our country have annual data, of course, but it has merely been updated by different authorities for various reasons.

Moreover, there is no single, definite time point for updating all variables. Another point to be noted is that many data are registered in connection with special events only (with the exception of a census).

Let us take for example a student who has lived for the rest of his life in the same house in the city in which he attended university. According to the registration office, he will have died at the age of 75 still as a student, assuming no census was carried out in the intervening years and he does not move to another house.

Examples like this highlight the problems of reliability of data based on information updated by administrative authorities only on the occurrence of vital events.

In the future, we will try to procure data of this type for 1980, as well as secure information about reliability limitations.

2. Why has the data been stored in separate files?

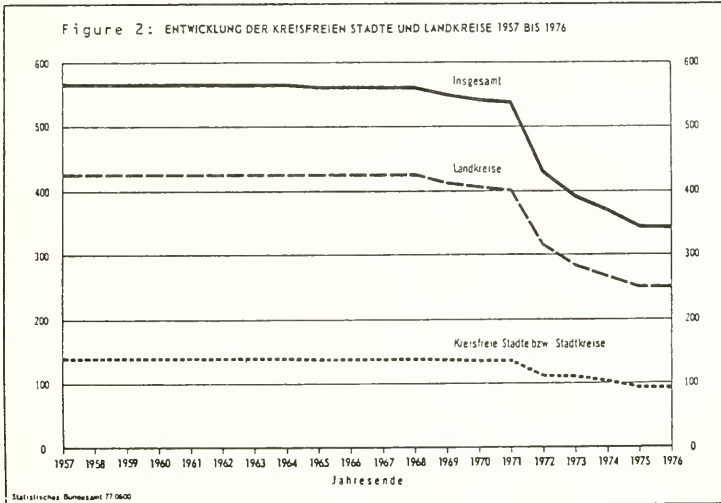
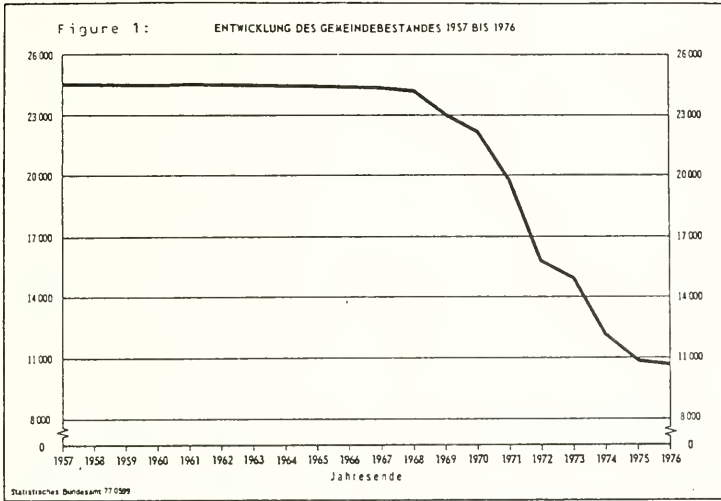
The answer is succinct and very important: great changes occurred in the numbers, boundaries and names of the regional units as a consequence of territorial reforms in Germany.

The main problem for the development of a regional data file consists in the instability of regional boundaries over time. From 1968 until the end of the seventies, a far reaching territorial reform took place, which affected not only most communes, but caused changes in regional units at higher aggregation levels as well. The territorial reforms were brought about to reorganize the communal self-administration at the local level in order to create an efficient system of common welfare and social security. Walter Christaller's central-place theory is basic to this concept.

For IFDO purposes, the greatest difficulty is the comparison of variables over time. Territorial reform did not only affect communes as a whole, for example the incorporation of an entire commune into another, but many others were divided, and their parts merged with yet other communes.

Figures 1 and 2 give a sense of the amount of change that has taken place in the last decade. Indeed, it's difficult to find any commune with stable boundaries over this period. The statistical offices of the Länder are able to 'reconstruct' the old commune limits by a "backward recalculation" ("*Rückwärtsaktualisierung*").

When the data on the newly delimited communes are converted into data of communes under the old demarcation, two different systems must be considered if the new commune was an amalgamation of old communes or parts thereof. Under the old system, all data on the commune to be divided in the course of territorial reform, was transmitted to the commune that incorporated the largest portion of the population within its enlarged limits. The new system focuses on distributing area and



Source: Statistisches Bundesamt Wiesbaden (1977): Devölkerung der Gemeinden 1976. Fachserie 1, Bevölkerung und Erwerbstätigkeit, Reihe 1.2.2. Stuttgart: Verlag Kohlhammer. Page 6.

population data exactly. Other data are appropriated to the new enlarged communes according to the percentage of population they received from the old divided commune. A well-known conclusion is as valid here as elsewhere in regional research. The errors, resulting from boundary changes, that occur in comparisons over time tend to vanish at higher aggregation levels.

3. Do we need data which allow comparisons for regional units over time if the regional units have changed?

Changes in the boundaries of regional units are modifications of the units of analysis. If one supports the idea that regions like Gemeinde, Kreis or Regierungsbezirk influence people's lives, one must consider their geographic reality in present as well as earlier times.

Given the many far-reaching changes in regional units by the territorial reform, the aggregation or disaggregation of the old regions within the present-day boundaries means nothing apart from getting a stable scanning pattern. It does not, however, have any influence on people's orientations in their geographic context.

From this point of view, we suggest that the several thousand boundary changes can be tolerated in Germany during the territorial reforms. Regional analysis at different time points are thus rendered possible. Comparisons over time, however, are not only a problem of data availability in Germany but also of adequate research design.■