Flexible regional references for administrative data
Andreas Gleich
Bureau for Statistics of the City of Augsburg (Germany)

In many European countries the next census will be a register-based census. Registers (and also sample surveys) are becoming more and more a main source of statistical information. The previous censuses (irregular including housing, buildings, occupation, work place) in Germany had been implemented by interviewing citizens at their homes. For the next census in 2011, the administrative registers will be used. It will be organised by the Federal Statistical Office in cooperation with the statistical offices of the 16 federal states. The primary objectives are to meet the requirements of the European Union and to get comparable results for the federal states and cities. Local administrations will not get the intraregional raw data. Register data may not be corrected during the census data collection or with the results of the survey.

The data of previous census is the basis for general statements in a large area until the results of the following census are published. Normally census data will not be updated between the surveys and the data does not change in time. Their regional references can be neatly constructed as a fixed structure valid for this set of data and refer only to the due date of the census. There are some approaches to update census data by the changes in the registers. But these extrapolations differ with the register more and more each year.

Most of the urban policy has an important geographical aspect and must be based on a thorough understanding of the territorial structures, geographical relations and developments. Politicians, planners and researchers are usually interested in geographically differentiating a local situation or development. Regional and urban planning is based on spatial analyses and this is based on geographical units. Statistical data must refer to a well-defined territory with a unique code, valid for the same point of time. Statistical agencies should be able to provide the user with a digital representation of the boundaries of the territory to which the statistical data refer.

The main problem in performing spatial and periodical analyses is to collect and to store the content data for these purposes. On the geographical micro-level, the situation changes quickly and ad-hoc answers are requested. They cannot be based on new surveys whenever a new question arises. But also the used data should be available for long-term planning. Municipal registers help to provide up-to-date information and data on the changes. Periodical excerpts from these registers can be used to analyse previous progress and to pre-estimate trends. Different questions require different geographical allocations and aggregates. It is the aim of the regional reference system to make any kind of geographical analysis possible with any kind of geographically referenced content data taking into account their date of validity and the validity of the geographical references requested.

Administrative registers are updated whenever the information on a change reaches the administrator. Not only content data, such as births, deaths or residence, are changing but also the regional reference. In contrast to the stable, abstract and usually extended regional references of census data, local administrative registers use postal addresses applied in everyday life as their basic regional reference. These addresses are in practical use by everybody, as the descriptions of locations for postal and similar services identifying localities for delivering post, goods and services. It is essential that they are up-to-date. However, addresses are changing often: renaming, splitting or merging of streets, building or demolition of houses, renumbering of addresses by different reasons and so on. Thus, at different moments there exist different addresses in the same area. Addresses in registers
have to be continually updated so that they are valid when needed. Excerpts from administrative registers for statistical purposes contain addresses which were true at the certain point of time, when these excerpts were made. Each element of the set of content data must have an unambiguous code for its location. For the use of administrative sources of different dates, the date of validity of the addresses is an important property. Content data and geographical codes must have a date of validity.

Many administrative and planning purposes require regional groupings of addresses, e.g., for school enrollment, for organisation of elections, delimitation of catchments areas of kindergartens, planning of bus routes, etc. If data has not been collected for the particular purpose of such an analysis, the analyst is usually confronted with the problem that the territorial references of the content data differ from the territorial structure for which he/she is aiming. It must be possible to geographically group and aggregate statistical content data according to the specific geographical questions to be answered (Fig. 1). It is useful to have small territorial units that are clearly identified by the territorial codes and can be composed like a mosaic according to the different needs of the users. Their numerical code is therefore organised in hierarchies and numerically linked to higher territorial units by a territorial reference system. One strategy for flexibility in geographically grouping the data according to user needs is to store them with their references on the lowest possible territorial level. As long as the available content data are referenced by this code (for example data about inhabitants, dwellings or other "content" of buildings), it is not a great problem to group them according to any of the different area types, i.e. zip code areas or school districts, if for each of these areas there is a complete list of territorial units located within its boundaries at a given point in time. This translation of the available reference data into the references of the target areas is an everyday task in database management.

To facilitate this task, municipal statisticians have defined so-called building blocks delineated by the surrounding streets or "natural" boundaries, like waterways or railway tracks. The lowest territorial level of this system is the "side" of a building block from one street corner to the next. These units are arranged in a standard hierarchy so that the planning or administrative districts can be composed and defined by the largest possible territorial units in the hierarchy completely lying within the area of the target unit.

The problem with spatial data is their instability in time. As long as a change occurs within an existing territorial unit in the hierarchy, only the lowest possible unit affected must be adjusted, thus adjusting automatically all units above it. Whenever an address changes, the different references for regional groupings of addresses must be updated as well. This is greatly facilitated if the references are organised hierarchically so that only the smallest territorial “building block” as an element of all higher levels of regional aggregates must be updated to be effective for all higher levels. The system of building blocks in a territorial hierarchy makes the territorial references quite stable, because all hierarchical and compounded elements must not be changed if e.g. a new address or a new street is added.

Each territorial unit can be composed like a mosaic to form any desired new structure. The composed target area (and with it all the associated elements) can be described just by listing the largest possible territorial unit of which its area is composed, without referring to each individual address. Automatically, any address in the city is assigned by the hierarchical code of the geographical element it belongs to. These references to the target areas have to be updated only if the target areas themselves change their boundaries or if the units describing the target areas change. The boundaries of such areas are change over time, just like the addresses they contain e.g. new housing areas or a new school requires new school districts and so on. However, they change independently of each other. The date or period of
validity is therefore just as important a property of the regional references as of the basic addresses and must be observed in any application. If, however, each of the possible target areas had been described by the individual postal addresses belonging to it, the amount of time for updating the references would be enormous. As a rule, the total area of the city is structured this way, so that a hierarchical code can be assigned to all data with a location anywhere in the city.

Geographical names, for example postal addresses, are used for different content data. They also change in time. They must be continually linked with the content data to guarantee consistency between the geographical and the statistical database. As this kind of change affects all administrative registers, it needs to be processed in all of them - and because of the links between these registers – in as uniform a way as possible. Because of their universal importance, these particular changes should be managed with special care and should be organised in such a way, that all registers can rely on a uniform and well coordinated system of valid addresses. They must be updated in a way that guarantees their permanent consistency, if more than historical evaluations are requested. In consequence, the combination of addresses with the hierarchical code reference for regional groupings is a task that requires a well-organised management system. The geographical reference systems should therefore be managed separately in an accepted organisational unit for universal use. The statistical information system, or at least the part of it to be linked with the geographical system, must satisfy the conditions of territorial plausibility and precise definition of its validity in time. The spatial dimension should include the time dimension in order to allow for evaluations of historic data. That is to say, the territorial reference of the content data must be coherent, also with respect to its validity, with the territorial references in the geographical system. This way, the territorial codes of the target areas can be retrieved in the geographical system and linked to the content data.

It is no surprise that many municipalities, at least in the decentralised form of organisation of Germany, have separated the management of their address system and of the regional references from the individual administrative registers and have concentrated it in a central regional reference system for all parts of the administration. In order to serve many different applications, this management system of addresses and of regional references must be complete, reliable, up-to-date and flexible.

For this purpose German municipal statistical offices have developed and established a standard geographical reference system under the name of AGK, separated from the statistical information system to facilitate handling and providing the geographical references to all its data. Management of updating the geographical data, interrelations and applications can thus be organised with a high degree of reliability, comparability in time, flexibility, and it is open to be applied in different technical environments. The data model of this standard geographical reference system comprises the following types of elements: postal address, street (name and code), sections of streets, sides of building blocks and building blocks. Aggregates are composed of these elements, some hierarchical up to the statistical sub-city district and some administrative, like constituencies, postal code areas, etc. (mostly more than 20 different administrative and planning areas per city). The alphanumerical elements are stored without geometry in a database management system (DBMS) but include their “history”, where each element knows its relationship to the other elements and also its former or next state at every point in time (Fig. 2). People without specific skills can use this system.

This regional reference system on the sub city level in Germany is based on recommendations of the Statistics Committee of the Union of German Municipalities of 1976 and - after reunification with East Germany – of 1991. The concept and technical tools for
this geographical management system have been developed over the last decades and are now coordinated and further developed in a kind of co-operative of about 40 municipalities lead by the Statistical Office of Augsburg. It is possible to transpose the urban examples presented and the solutions proposed into similar tasks and solutions on the regional level.

In a number of cities this alphanumeric hierarchical coding system has already been enlarged, following the Geo-code principle, by a topological net composed of segments, knots and mesh as the areas between them. They form a geometric pattern of the city, the “lines” (streets, transport lines, rivers, etc.) as well as the areas. Any information identified by an address, the code of any of the elements of the system or by coordinates can automatically be mapped or evaluated by Geographical Information Systems (GIS). For each alphanumeric element, the system links its codes to the codes of all other relevant elements and to their geometry in the Geographical Information System (Fig. 3). The information on the beginning and end of validity of each element supports historical comparisons maintaining the full flexibility of geographical referencing according to the user needs. In a standardised geo-database (for ArcView GIS 9.x), any geometry can be created on the basis of postal addresses and sides of building blocks from which, in principle, all other geographical aggregates can be derived. Thus, all data of the statistical information system are integrated in the geometry and vice versa. In contrast to geodesy, this combination of the alphanumeric coding system and geometry enables the user to visualize any content information directly, without having to digitize it beforehand and is independent of the map scale. The “geometry” for mapping is derived directly from the geo-coded reference system that is, in combination with software and the statistical information system, a powerful statistical information system. Of course, there is still much to be done to fulfil all user needs, e.g. the translation of geographical aggregates into the grid system and improvements in the geographical visualisation of maintenance activities (Fig. 4).

It is a common goal of all urban and regional researchers to increase flexibility and reliability in this field. However, according to their educational background or institutional position, they either try to integrate the interfaces for this task in their statistical information system of the content data (usually relying on numerical references) or they try to integrate the content data together with the management of the geographical references in their geographical information system (usually relying primarily on coordinates and geometric algorithms). The state of affairs and the perspectives presented try to avoid overloading either of the systems with tasks for which they have not been designed, combining the strengths of both systems by linking (instead of integrating) the geographical reference and information system (managing the geographical data and tools) to the statistical information system (managing the content data and statistical tools).

Comparability and compatibility are the central requirements of statistics and must be based on standards. These standards can only become effective, if they are verified and stable. Equally important is simplicity. If standards are too complicated, they will not be applied. The underlined statements in this paper shall force the current discussion on recommendations for the management of regional reference systems by referring to the address and reference system of German municipalities.

Key words: regional references, administrative registers, municipal statistics
Fig. 1: Simplified process chain from raw data to product

Fig. 2: Hierarchy and simplified data model for the elements of the regional reference system
Fig. 3: Thematic map (right) produced based on the alphanumeric code (left)

Fig. 4: School districts with school locations (left) “compared” with the amount of students in grid cells (right)