Adding Value to Statistical Information by Georeferencing

Norbert RAINER
Head of Registers, Classifications and Methods Division

Ingrid KAMINGER; Gernot KATZLBERGER
Cartography and Geographic Information Systems

All: STATISTICS AUSTRIA

Georeferencing of the base data held in linked registers has multiple advantages. Besides the obvious one of knowing the exact location for each object, data can be aggregated to individual or functional areas independent of administrative units. With the help of GIS methods, georeferenced data allow the establishment of new and enhanced statistical information and analysis which go far beyond traditional regional or spatial statistics. This contribution intends to give an overview of the conditions for georeferenced data and the possibilities that such data permits for the users.

1. Introduction

For many years the only spatial resolution for which statistical data was available were administrative areas; however, administrative areas have several disadvantages: they change over time, are heterogeneous within and between countries and are usually too large so that they comprise different kinds of subareas inside them. With the introduction of georeferenced data, the scope of possible methods and tools has widened and the creation of small area statistics became possible.

The first steps towards assigning statistical data to grids at Statistics Austria go back to the 1980s. However, this was only at an experimental stage. It has taken another two decades for comprehensive georeferenced base data to become available. This was an outcome of the preparatory works for the establishment of the Buildings- and Dwellings Register in 2003. Since then the use of GIS-based methods and the provision of georeferenced data for our users have become standard.

This paper provides a short overview of the experience of Statistics Austria in providing georeferenced data; it focuses especially on the principles under which the system was developed and maintained and how the georeferenced data is being made available to the public. Furthermore, it gives a short outline of the various uses of georeferenced data by our users.

2. Principles of Statistics Austria concerning the provision of georeferenced data

In developing and providing georeferenced statistical data, Statistics Austria has elaborated the following principles. These principles served as guidelines and have been further developed according to the experience made.

2.1. Register based approach

Like in other National Statistical Institutes, the long term strategic goal of Statistics Austria was to develop integrated statistical registers and to maintain these registers mainly by using administrative data sources. A register-based statistical system has a lot of advantages.

It was therefore obvious that also the georeferencing should be integrated into the register system. This was done in the Buildings- and Dwellings Register (BDR), where a software application is accessing online the digital cadastre map of the National Mapping Agency and allows allocating the coordinates to each estate and to each building by simple clicks. Two types of coordinates are applied: coordinates for the estate (usually placed at the estate entrance) and coordinates for the building (also placed at the entrance of the building).

As the BDR is linked with the other statistical registers the georeference can be utilised in various other statistical domains. Integrating georeference in a register-based system supports the availability of up-to-date information by continuously updating the register information. Coordinates for each building provide the most precise and smallest location reference.

2.2. Grids as spatial unit

To deal with the challenge of providing regionally detailed data but keeping individual data as well as the exact location confidential, Statistics Austria has adopted the system of statistical grids as the smallest statistical area unit for which data can be passed on. Statistical grids are defined as a region-wide spatial reference system for geodata and are described by a

coordinate system. A hierarchy of different grid sizes starting from 100m x 100m up to 10km x 10km ensures that the appropriate data base is available for the purposes of the customers.

There are various advantages when using grids. Grids are independent of administrative boundaries and therefore allow an adequate delineation of areas. Because of their large scale, spatial distributions can be more easily identified. Technically, grids are easy to handle with today's geoinformation systems. Using the georeferenced Buildings- and Dwelling Register as basis, Statistics Austria is able to provide grid data compiled in a "bottom-up" approach (aggregated grid data). This method is very efficient and automatic mapping between various grid systems can also be done easily. From an analytical point of view a comparison between different regions is easy to achieve because of the equal area of the grid cells. Also temporal comparison is possible because the grids are stable over time unlike administrative areas. In addition, grid data are the adequate approach for cross-border analysis.

2.3. Confidentiality concept

Concerning disclosure control a solution had to be found between the user needs of detailed spatial data and the necessary compliance of confidentiality. Statistics Austria is using a combination of methods to prevent users from deriving individual data from small area aggregates. Some of the base data is perturbed (target record swapping), data below a certain threshold is not published and the published data only include few predefined variables and characteristics with reduced level of detail (e.g. age groups rather than single years). The general principle is: the smaller the spatial area, the fewer data are made available.

2.4. Data packages as main form of data provision

Compliance with data confidentiality was also the main reason for providing ready-made data packages. This also supports more efficient ways of data provision to the customers. Where possible, the same predefined variable structure is maintained for updates. Furthermore, users are requested to sign legal terms of use before the data is provided.

Data packages provided on grids come from different sources, mainly from censuses or from registers. Datasets for the smallest grid sizes (100m/125m and 250m) have been greatly demanded, as these show the distribution of spatial phenomena best.

2.5. European standards

When Statistics Austria first introduced its national grid system in 2003, it was designed based on the Austrian national projection (Lambert Conformal Conic projection). This ensured that it suited other Austrian national geodata. The costumers accepted the grid system of Statistics Austria as the de facto standard and used it in their analysis. With the launching of INSPIRE a harmonised European grid system was introduced. It was originally meant for environmental data sets, but later on statistical data got integrated. Since it is going to be the future European standard also for statistical data Statistics Austria has switched to that system.

The European grid system based on the ETRS89-Lambert Azimuthal Equal Area Projection is parallel to the 52° latitude and 10° Meridian and hence no longer square when reprojected to suit the Austrian national geodata. Both the national and the European grid systems were being run in parallel during a transition period, but since 2011 all grid data packages of Statistics Austria are only provided in the European grid system. Also the users of the former grid had to change to the new system.

2.6. Role of GIS

Having georeferenced data available also allowed the use of Geographic Information Systems (GIS) at Statistics Austria. GIS is understood as an instrument supporting the production, presentation and analysis of statistical data with respect to their spatial dimension. It is mainly used as an instrument to create added value based on the georeferencing of the base data. GIS is also used for cartographic purposes and the delineation of functional areas. Thus, for Statistics Austria, GIS is used like other methods which support the creation and presentation of statistical data. Statistics Austria is not providing specific GIS based studies and analysis addressing specific regional planning questions. However, Statistics Austria provides a lot of georeferenced data for such analysis undertaken by universities, research institutes and other institutions providing regional and spatial analysis and planning.

3. Implementation of georeferenced data

Together with the National Mapping Agency (www.bev.gv.at) Statistics Austria acts as a service provider to the municipalities making available an online application called the "Address-Buildings and Dwellings Register Online". This joint application ensures a consistent data base of estate and building addresses, which also serves for various other

statistical and administrative registers, e.g. the central register of residents. The application includes a geo-client for the georeferencing of both estate addresses as well as building addresses (see Figure 1). Each building therefore has x and y coordinates. Of course, the buildings and dwellings register includes a lot more characteristics than just the coordinates.

Figure 1: Geocoordinates per building; register of buildings and dwellings

Via the unique building identifier, other registers, census results and other data sets can be linked and the georeference of the building can be applied to these data sets as well (see Figure 2).

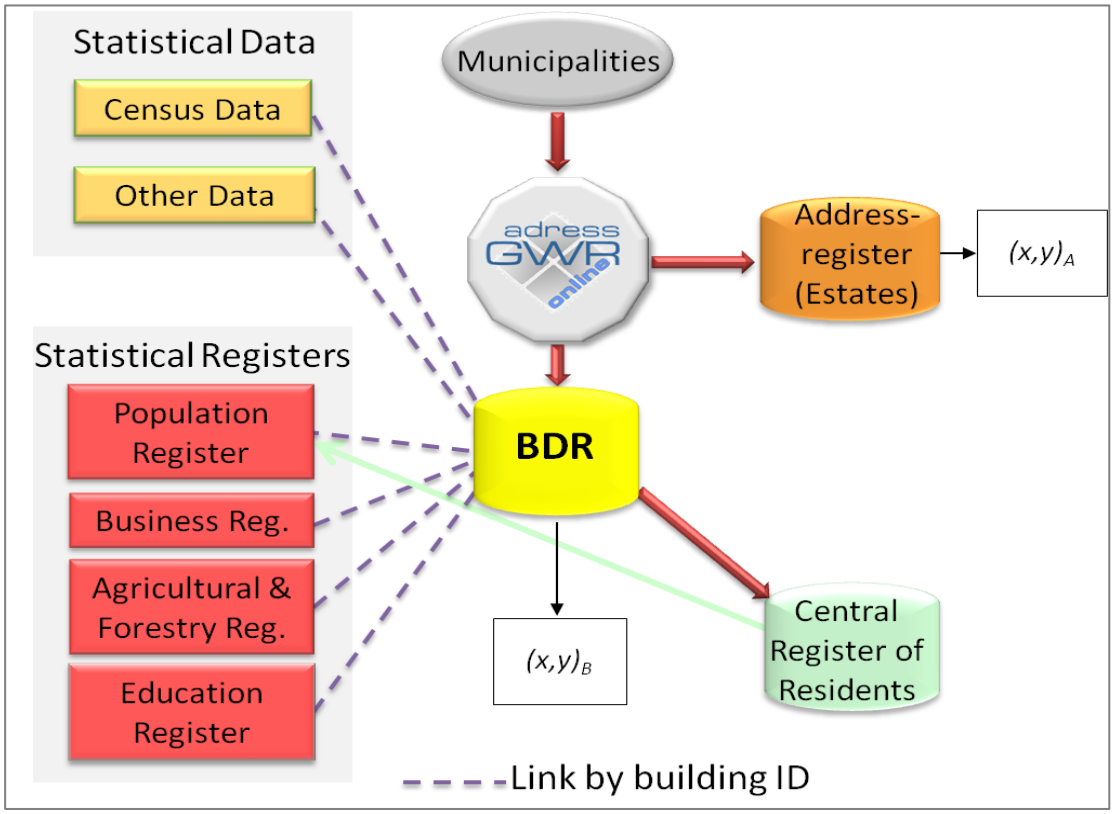

Figure 2: Link with the BDR (x, y - Coordinates) by building ID, Statistical data and Statistical Registers

The system of registers

The Building- and Dwellings Register and the Central Register of Residents represent the backbone of the register-system. The other base registers are the statistical Business Register, the Agricultural and Forestry Register and the Register of Educational Attainment, all of which are maintained by Statistics Austria. All of them use the address basis of the BDR and are therefore implicitly a georeferenced database.

Geodatabase

To manage the statistical data at the different regional levels, at the different points in time as well as and the connected geodata, a geo-enabled database was established (IBM-DB2). It includes the geometries and territorial status for all regional levels (e.g. municipalities, census districts, grids, geocoded buildings,...) as well as selected information of the georeferenced base data. When the statistical data is imported it is automatically aggregated to all regional levels. Pre-defined SQL-queries are integrated into the geodatabase, which can be triggered to create the required data packages for grids or other regional levels and the required metadata for any combination of territorial status and reference date. In particular this makes it easier to provide time series taking into account changes of administrative units.

For confidentiality reasons there are two database views of the data. One view is only for internal use providing all information and one for exporting data to external costumers where the data protection rules are applied. This geodatabase is the backbone to serve the costumers of small area statistics and is an effective way to provide the data packages. It also serves as the basis for the delivery of regional statistical information for costumers and for INSPIRE.

4. Main Uses of georeferenced data

Within Statistics Austria georeferenced data is mainly used as basis for delineating functional/thematic areas independently from administrative boundaries. Examples are the delineation of urban regions (metropolitan areas), rural areas, localities (settlement areas), etc. (http://www.statistik.at/web_en/classifications/regional_breakdown). The precise location is also used to calculate commuting distances along traffic routes as well as population weighted centroids for grids and enumeration districts.

Additionally to the grid data packages new forms of presentation of statistical information are provided. One example are the interactive maps

(http://www.statistik.at/web_en/publications_services/interactive_maps/). This tool enables interactive thematic and topographic maps on a variety of subject areas like population, agriculture or tourism to be made available on the website. In addition to municipal and district maps, data is also visualised on the basis of the statistical grids (see Figure 3). Also geodata itself, without the statistical information, is made available to our customers (http://www.statistik.at/web_en/publications_services/geodata/).

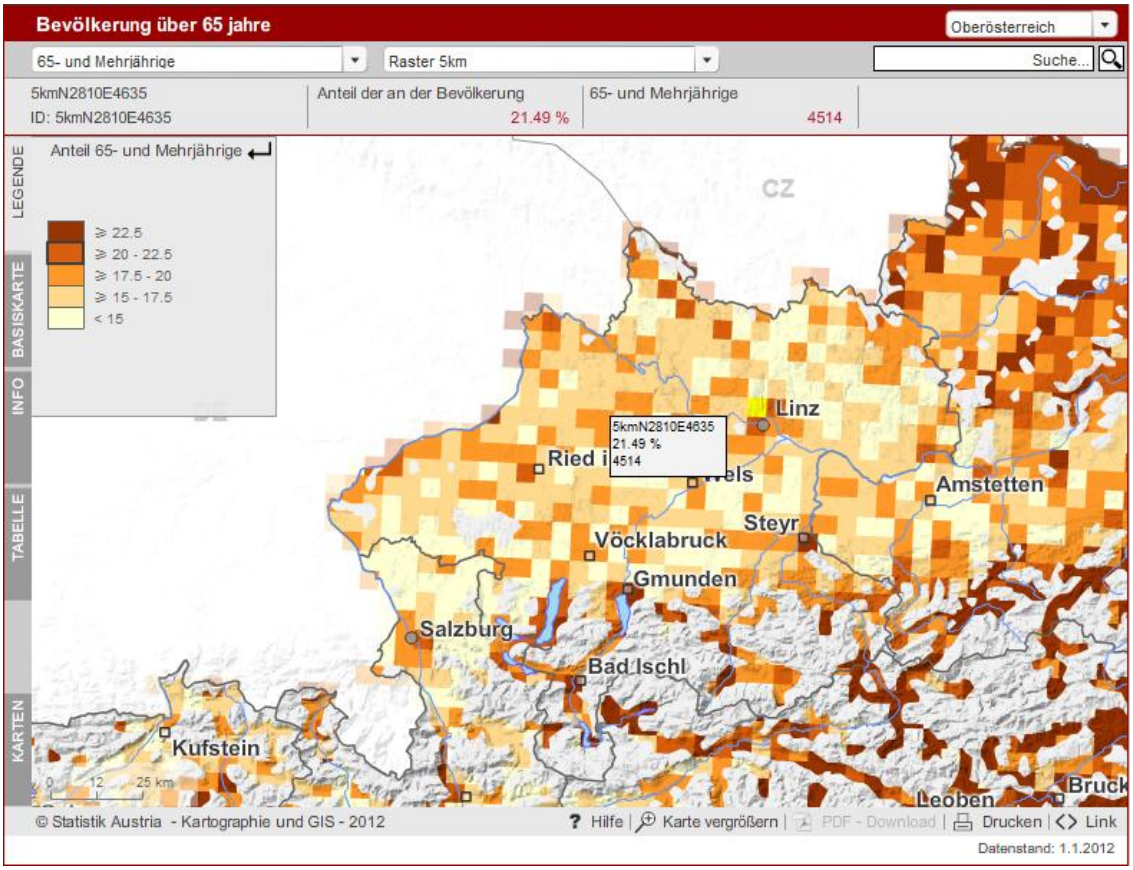

Figure 3: Interactive map visualising the population over 65 years in Upper Austria (5km grid).

4.1. Use of data on the basis of statistical grids

Grid data packages are provided to customers for spatial analysis, regional planning or geomarketing purposes. Therefore, the main users are governmental agencies, private companies, geomarketing companies and universities/research institutes. Over the past few years Statistics Austria has provided data based on georeferenced information (grid data, customer defined polygons and geodata) to approximately 150 customers, mostly from Austria, but also from other European countries and from Overseas.

The majorities of customers use the grid data in the fields of regional planning, the planning of mobile phone base stations, shop location analysis, sales territory optimisation, market analysis and research on energy supply.

An advantage of using standardised grid systems is that data users can combine data sets from Statistics Austria with data sets from other data sources. Taking an example for agriculture, statistical data on agricultural farms can be combined with data of soil types, water conditions, permeability and soil depth, or data on damage by hail, frost or drought or other climate data, such as precipitation, humidity or sunshine duration.

In Austria some federal provinces are using grids for *regional planning* and to create concepts for the local development. For this task of planning, the smallest possible statistical grid cells such as the 100m/125m size are an adequate data basis.

4.2. Customer defined polygons

Besides providing data on the basis of grids, the geodatabase system of Statistics Austria is also suited to provide aggregated data of any customer defined polygons. These data can refer to the number of buildings and their characteristics, the number of residents, etc. Of course, the confidentiality criteria have to be met. This is a service, which is either used by customers, who do not use GIS themselves or by customers needing very precise data, for which the exact location of objects is needed and hence not even the smallest grids can answer their requests.

Analysis based on customer defined polygons is used in various projects. One example is the number of buildings and residents in high risk flooding zones (see Figure 4). Which measures should be taken beforehand to prevent big damages in case of flooding? And in the case of an emergency it is important to know which buildings are concerned, whether the buildings are used for living and whether there are elderly people, children or other people with special needs. Also the information on the type of heating used in the buildings concerned is of importance, as experience shows that oil tanks in the basements can burst and cause further environmental damage.

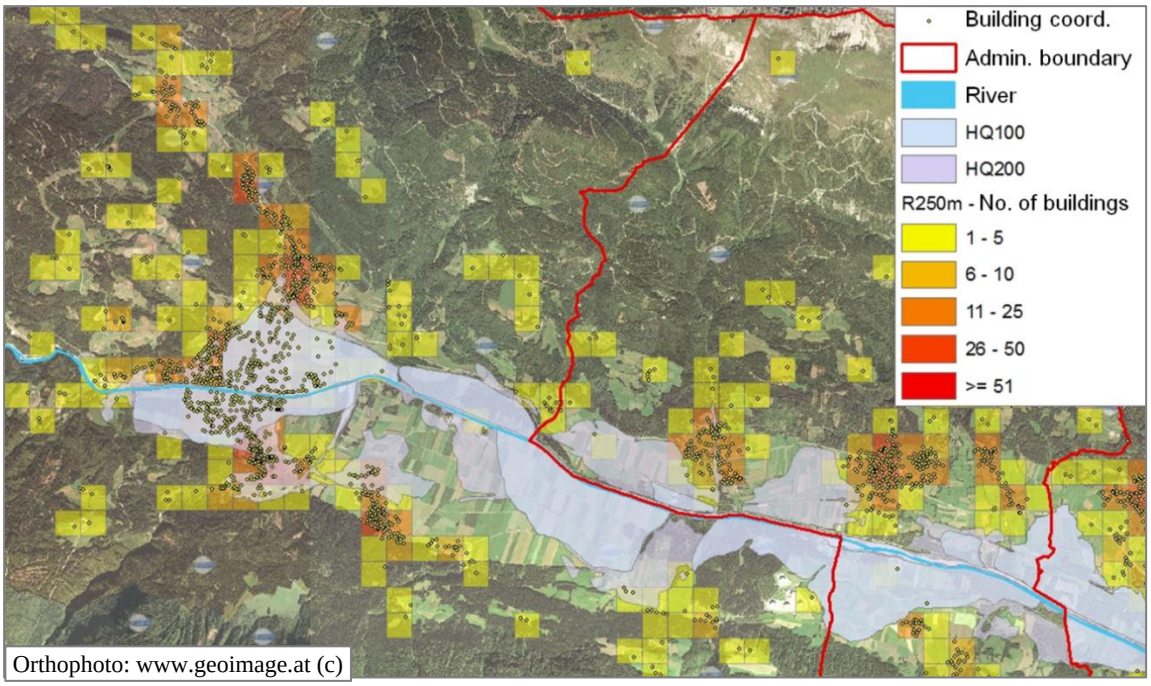

Figure 4: Number of buildings in a high risk flooding zone (250m grids versus building coordinates)

Another customer is an Austrian airport providing noise-level-polygons for approaching and departing corridors. Statistics Austria calculates the number of residents per noise-level for each of the affected municipality. Similar examples are the noise level polygons along high frequency roads and high-ways. In these calculations the distance to the road/high-way as well as the digital elevation model is used to provide exact data on the number of residents affected.

5. References

[1] GEOSTAT Project Team (2012), Final report of the GEOSTAT 1A project on the development of a European population grid dataset.
http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/documents/ESSnet%20project%20GEOSTAT1A%20final%20report_0.pdf
http://www.efgs.info/geostat/geostat-1a

[2] INSPIRE Reference Systems Thematic Working Group (2008), D2.8.I.2 Specifications on Geographical Grid Systems. Draft Guidelines.
http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_Specification_GGS_v3.0.pdf.

[3] Kaminger I. (2010) Auf dem Weg zu einer rasterbasierten Regionalstatistik in Europa. In Flächennutzungsmonitoring II, *Konzepte – Indikatoren – Statistik, Leibniz-Institut für ökologische Raumentwicklung (Hrsg. Meinel, Schumacher);* IÖR Schriften Band 52, Rhombos-Verlag Berlin.

[4] Kaminger I., Katzlberger G. and Rainer N. (2010), Statistical grids as a tool for small area analysis. *27th SCORUS Conference. Latvia*.
http://scorus.csb.gov.lv/images/stories/papers/4_5_kaminger_katzlberger_rainer.pdf

[5] Kaminger, I. and Wonka, E. (2005), Regionalstatistische Gebietsgliederungen nach geographischen Rastern, *Statistische Nachrichten, Statistik Austria*, Wien.

[6] Statistics Finland (2010), Production and dissemination of grid data since the 1970 Census in Finland, *Conference of European Statisticians, Fifty-eighth plenary session*, Paris.

[7] Statistics Norway (2009), Bloch, V. and Strand, G. Statistical grids for Norway
http://www.ssb.no/emner/01/90/doc_200909/doc_200909.pdf

[8] Statistical Office of the Republic of Slovenia (2010), Establishing a national hierarchical grid in Slovenia – lessons learned and future challenges, *Conference of European Statisticians, Fifty-eighth plenary session*, Paris.

[9] Wonka, E., Kaminger, I. and Katzlberger, G. (2009), Regionalstatistische Auswertungen mit geographischen Rastern in der Raumplanung, *Informationen zur Raumentwicklung, Bundesamt für Bauwesen und Raumordnung*, Heft 10.