



## **Annex 1. Use cases**

Final report from the GEOSTAT 2 project

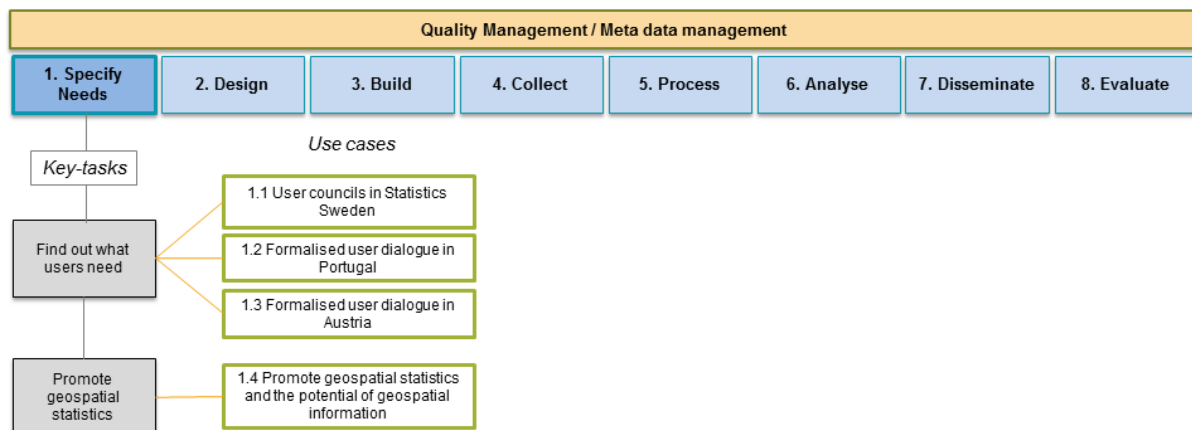
## Content

1. Specify needs.....	5
1.1 User councils in Statistics Sweden.....	5
1.2 Formalised user dialogue in Portugal.....	5
1.3 Formalised user dialogue in Austria .....	7
1.4 Active promotion of geospatial data and statistics in Norway .....	7
2. Design .....	9
2.1 The Swedish National Geodata Cooperation .....	9
2.2 Data access through Norway Digital .....	10
2.3 Legal basis for data and geospatial data access in Portugal .....	11
2.4 Legal basis for geospatial data access for Statistics Austria.....	12
2.5 Acquisition and processing of geospatial reference data in Poland .....	12
2.6 GIS Resource Centre in Statistics Norway .....	13
2.7 Resource setup in Statistics Finland .....	14
2.8 Resource setup in Statistics Poland.....	15
2.9 Resource setup in Statistics Portugal .....	16
3. Build.....	18
3.1 The Geography Database - production set-up for point-based geocoding in Statistics Sweden	18
3.2 The GeoDatabase - production set-up for point-based geocoding in Statistics Austria .....	19
3.3 The production set-up for point-based geocoding in Statistics Portugal.....	21
3.4 The Labour Force Survey (LFS) sample in INSEE.....	23
3.5 Sampling Frame - National Dwellings Register - Portugal.....	24
3.6 The sampling frame for social surveys (SFFS) in Poland.....	26
4. Collect.....	27
4.1 Managing temporality – The Regional database and the Address- Buildings & Dwellings register at Statistics Austria .....	27
4.2 Address point acquisition in Statistics Poland.....	32
5. Process.....	34
5.1 Cross-check and geospatial data validation in Statistics Portugal .....	34
5.2 Geospatial data assessment in Statistics Norway .....	35
5.3 Non-spatial data assessment in Statistics Norway .....	35
5.4 Maintaining the quality of a point-based Buildings and Dwellings Register in Statistics Austria	36
5.5 Quality indicators for non-spatial data.....	37
5.6 Geocoding population data in Statistics Sweden .....	41

5.7 Geocoding workplaces in Statistics Sweden .....	43
5.8 Geocoding workplaces in Statistics Portugal.....	44
5.9 Geocoding practise in Statistics Finland.....	46
5.10 Geocoding practise in Statistics Austria .....	46
5.11 Geospatial statistics portfolio in Statistics Poland .....	48
5.12 Geospatial statistics portfolio in Statistics Austria .....	49
5.13 Metadata and INSPIRE compliance – Finland .....	49
6. Analyse .....	51
6.1 Data dissemination in Statistics Portugal .....	51
6.2 Dissemination of grid data for tax information in France: issues and solutions.....	52
6.3 Constrains on data dissemination in Austria.....	53

This annex is a part of the final report from the GEOSTAT 2 project. The use cases presented in this annex are linked to Chapter 6 of the main report and represent national practices found in the project countries.

## 1. Specify needs



### 1.1 User councils in Statistics Sweden

Statistics Sweden currently has nine user councils<sup>1</sup>. User councils are established by a decision of the director-general, who also decides on the instructions, chairpersons and other members of the user councils. The user councils have an external chairman and consist of approximately ten members.

The purpose of the user councils is to maintain a system of organised user interaction to continually provide Statistics Sweden with input about the new and changing needs for statistics and to obtain the views of key users regarding changes in the statistics. User councils form a kind of network of information, whereby Statistics Sweden can spread and get feedback on ideas and plans. They have an active role in prioritising and follow-up as well as an advisory capacity to SCB. The user council members also receive comprehensive information regarding Statistics Sweden, such as the annual report, business plan, budget, etc.

The user councils have an advisory role and shall represent the users of the government funded statistics that Statistics Sweden is responsible for within each subject area. This refers to both the general needs for statistical information and the more specific needs of different users.

The user councils have 2–4 meetings each per year. The discussions are documented and an annual activity report is submitted. The activity reports provide the basis for the reporting of user councils' activities in Statistics Sweden's annual report.

There are no user councils specifically addressing geospatial statistics but issues related to small area statistics or geospatial data production are mainly handled by the User council for regional statistics and the User council for land use and Real estate statistics.

### 1.2 Formalised user dialogue in Portugal

The Statistical Council (SC) is the state entity that presides over, guides and coordinates the National Statistical System (Statistical Law, Article 3(2)). It was established to provide room for the debate between producers and users of official statistics and especially to ensure the reliability of the statistical system. The SC is chaired by the member of government responsible for Statistics Portugal,

<sup>1</sup> For further information: <http://www.scb.se/en/About-us/Main-Activity/Councils-and-boards/User-councils/>

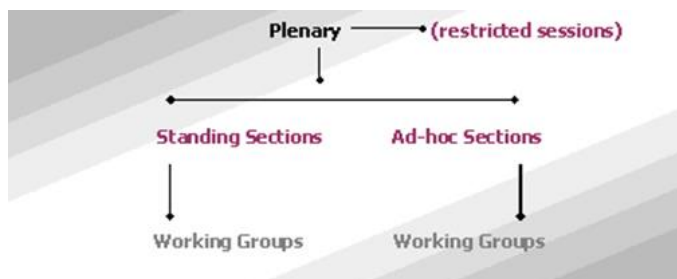
at present the Secretary of State for Administrative Modernisation. The Director General of Statistics Portugal acts as the Vice-Chairperson of the SC.

According to the Law of the National Statistical System, the SC has the powers to:

- (...) define and approve the general guidelines of official statistical activity and relevant priorities;
- define, on an annual basis, nationwide official statistical operations and those solely in the interest of the Autonomous Regions, upon proposal of statistical authorities (...);

The 28 members of the Statistical Council are representatives of Statistics Portugal, all other national authorities producing European and national statistics, Banco de Portugal, public services (being major users), employers' confederations, union confederations, association of municipalities, the Consumer Protection Association, universities, the Data Protection Authority, and five persons of recognised scientific merit and independence. The members are appointed for three years by a decision of the Prime Minister, upon proposal by the ministers of the represented ministries, the Director General of Statistics Portugal and the Council of Portuguese University Rectors.

The SC has five standing sections on overarching topics and on specific statistical areas. Ad-hoc sections can be established. Currently there is one ad-hoc section dedicated to the revision of the Statistical Law. These sections may also set up their own task forces, working groups and technical groups.



**Figure 1: Chart illustrating the organisational setup of the Statistical Council of Portugal.**

For regional statistics, the activities are carried out by the Standing Section on Territorial Base Statistics, which have the followings powers:

- Cooperating with the Standing Section of Statistical Coordination in the preparation of the document Linhas Gerais da Atividade Estatística Oficial (General Guidelines of Official Statistical Activity);
- Monitoring the production of territorial base statistics, namely by assessing its adequacy to user needs;
- Monitoring, in strict cooperation with the appropriate Sections, the production of territorial base statistics through the analysis of statistical projects with relevant implications for statistical data at regional and local level;
- Developing activities that foster maximum use of administrative data for statistical purposes, in cooperation with the appropriate Section;
- Fostering the employment of existing statistical operations targeted at a maximum use of their potentialities for the enhancement of territorial base statistics;

- Monitoring, through national institutional participants, the activities of Committees or Working Groups operating within the European Union and the relevant international bodies regarding its scope of action;
- To assess Reports and supervise Working Group Monitoring Plans in place within the Section.

Statistics Portugal has some feedback mostly from the users, including the municipalities (one of the most important users and cooperation partners) and from enterprises, researchers or student's that consume statistical products regularly.

From a purely internal perspective, the Department of Methodology and Information Systems receives annually - from the all other departments at Statistical Portugal - their needs for applications, data warehouse, methodology, IT and geographic information, related to new, revised or current surveys.

### 1.3 Formalised user dialogue in Austria

In the Austrian Statistics Act the establishment of the Statistics Council ("Fachbeirat") is regulated as well as the frequency of meetings, the process of deciding on resolutions and a definition of its duties.<sup>2</sup>

The Statistics Council comprises 16 members representing government, various ministries, the national bank, the Austrian association of municipalities/towns and various federal chambers. The Statistics Council shall meet as and when required and at least quarterly and a written record of the proceedings and resolutions of the Statistics Council shall be prepared. Among the duties of the statistics council is also the preparation of an annual report and the issue of recommendations and statements.

In addition it is possible to establish national expert working groups on specific topics (e.g. working group on territorial matters; Building and Dwelling Register forum), which have the task to advise the statistical office on specific thematic matters.

Statistics Austria conducts user surveys to learn about user requests and additional user requests also come in by email/telephone, generally through contact of the staff with the customers. There is no limit to what users would need – they usually wish for data as small scaled as possible (building level) and as timely as possible (data as of today).

### 1.4 Promote geospatial statistics and the potential of geospatial information

Statistics Norway participates actively and presents their geospatial statistics products in different forums, ranging from presentations for GIS-experts in county administrations, to regional stakeholders in the "Norway digital cooperation" and national meetings for municipalities, public and private institutions/companies in the geospatial sector. Statistics Norway also regularly conducts guest lectures at a university which educate land use planners.

To identify "hidden" needs for geospatial information internally, Statistics Norway have, after years of recognising the potential, established a project within the statistical institute, to promote the potential of geospatial information ("GIS resource centre", see case 2.6 *GIS Resource Centre in*

---

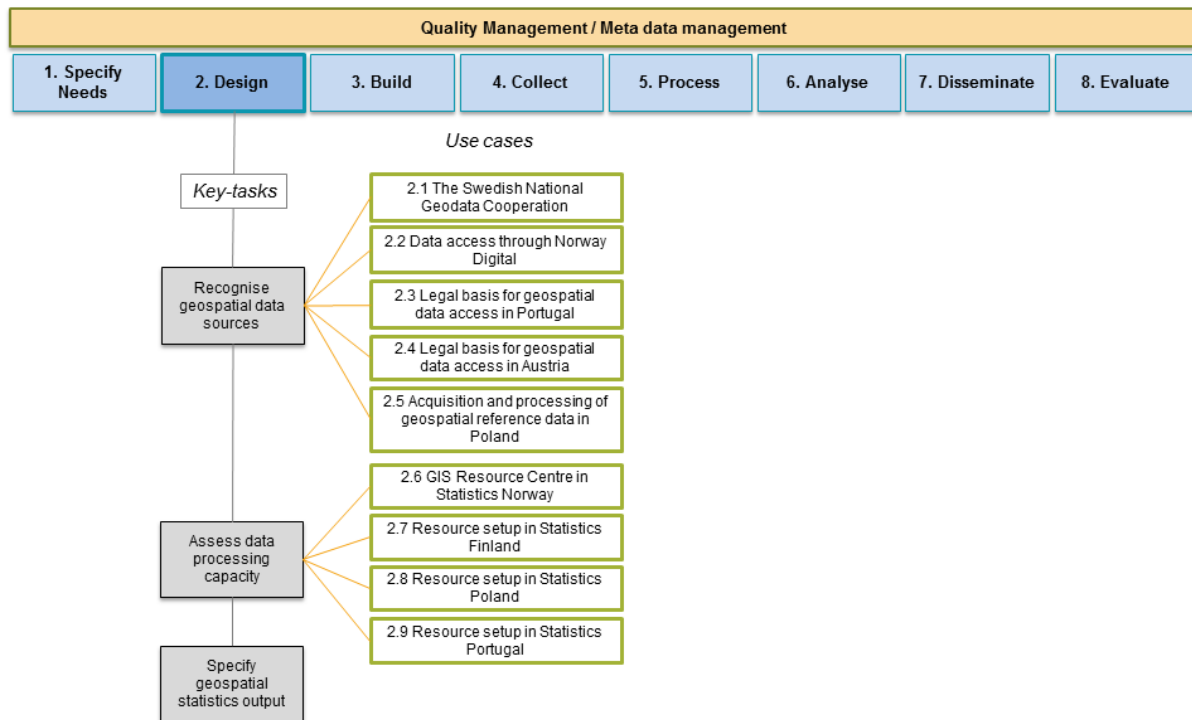
<sup>2</sup> English version of the law available under:

[www.statistik.at/wcm/idc/idcplg?IdcService=GET\\_PDF\\_FILE&dDocName=027192](http://www.statistik.at/wcm/idc/idcplg?IdcService=GET_PDF_FILE&dDocName=027192)

*Statistics Norway*). The goal is to ultimately make geospatial information management an integrated part of the statistics production.

Statistics Norway concluded there was a need to establish infrastructure for management of geospatial data, inform the statisticians by means of information seminars/workshops and internal courses. To identify concrete projects, a survey was sent to all heads of unit. The resulting projects, big and small are being followed up by the resource centre. The projects lead to various outcomes, but even more important is the resulting competence building. The resource centre is just to support all aspects of the production when it comes to geospatial information (from education, information and advising to development of production lines involving geospatial data).

## 2. Design



### 2.1 The Swedish National Geodata Cooperation

The Swedish Geodata Cooperation Agreement is the national foundation for a sustainable cooperation within the SDI. The cooperation was launched in 2011 in order to fulfil the obligations of the INSPIRE directive, yet it goes beyond the aspirations of INSPIRE as it reaches further than the scope of environmental information. As of today, the Geodata Cooperation offers a pool of data exceeding 400 geospatial data products from 19 different providers. The very basic concept of the cooperation is that parties sign *one* agreement and pay *one* annual fee but get access to geospatial data products from all data providers. This concept stands in sharp contrast to the situation before 2011, where data acquisition entailed complex business models and expensive agreements.

The cooperation is managed by the Swedish NMCA (Lantmäteriet). Parties to the cooperation are authorities with a data management responsibility according to the Swedish Act and Ordinance on Spatial Information, based on the INSPIRE directive, and municipalities, government agencies and other organisations with official duties.

The Geodata Cooperation Agreement includes how to handle organisation, steering, coordination and responsibilities as well as technical prerequisites, forms of supply and terms of use of spatial data. The parties in the Geodata Cooperation offer each other spatial data for official use at an annual fee. Available data are presented and described in a Product Catalogue. Municipalities, government agencies and other public organisations which conduct official duties can also join the Geodata Cooperation, and thereby get access to all data in the Product Catalogue, for official use.

The contents of the Product Catalogue will change over time, with the aim to include as much spatial information as possible from all authorities in the cooperation. The INSPIRE regulation gives a minimum requirement, but in order to fulfil also the goals in the National Geodata Strategy, the

cooperation has a broader scope: by making available as much spatial information as possible the benefits from sharing information will increase within the public sector.

In general terms, data sharing is a cost-effective way of enabling the entire public sector the use of high quality data for a wide variety of tasks. By making spatial data available as services on the web, it is also easier for the private sector to benefit from this infrastructure, as it gives easier access with known conditions and licenses.

One of the most obvious results of the cooperation is reduced administration regarding licensing and data acquisition. The simple and straight forward business model makes it easy to budget future (next years) costs for Spatial Information. The situation before the Geodata Cooperation suffered from multiple agreements between data providers and users and complex terms of use obstructing flexible use of data and creation of services.

## 2.2 Data access through Norway Digital

Access to data (including geospatial data) is regulated by the Norwegian Statistics Act. The act provides access to all relevant registers free of charge. However, data providers may charge for the work consumed by data extraction.

Regarding geospatial data, there is an agreement for exchange of data with all public providers. “Norway Digital” is the name of the national spatial data infrastructure (SDI). The Norway Digital collaboration is a formalised agreement based on cooperation between municipal, regional and national organisations responsible for producing geospatial data or is big users of geospatial data. The partnership was established to build and operate a national spatial data infrastructure to ensure the whole community good access to geospatial data. The law on infrastructure for geographical information (Geodata Act) was adopted in 2010 and is also an implementation of the INSPIRE directive establishing an Infrastructure for Spatial Information in the European Community. The law aims to promote good and effective access to geospatial data for public and private purposes, and assumes that this will be achieved by strengthening and continuing the voluntary collaboration in Norway Digital. The agreement grants access to geospatial data as well as participation in different specialist committees. There is an annual fee for being part of the cooperation.

Besides the Norway digital agreement, there are additional “agreements of data quality” with some data providers. These agreements secure regular delivery and provide a formalised way of exchanging information about quality. Regarding geospatial data, the agreement with the National Mapping Authority (NMCA) is most important. A massive amount of geospatial data are transferred from the NMCA to Statistics Norway each year. Data are provided both as download services and as web services. Given the amount of data, and the need to access “historical” data, relying solely on web services, has not been an option so far.

Regarding cadastre data on buildings and addresses, a statistical copy of the cadastre is updated daily. From this data base, “situation extracts” are created regularly. From some of these “situation extracts” ready-to-use geospatial data are prepared in the central geodata base. (Address points with population, building points with core information, enterprise points with core information).

## 2.3 Legal basis for data and geospatial data access in Portugal

Under the National Statistical System's Law<sup>3</sup> Statistics Portugal is mandated to use administrative records for the purposes of compiling statistics. Administrative data owners have to provide the data free of charge.

Since 2013 and under the legal act that regulates the Indicator System for Construction and Housing Statistics (SIOU) the municipalities are obliged to provide to Statistics Portugal, on a monthly basis, all the completed buildings and buildings permits including X,Y location and addresses of each building. SIOU is, in this way, the reference data source to update dwellings in the National Dwelling Register and buildings in the Buildings Geographical Database.

However, although the Statistical Law apparently covers sufficiently the accessibility of administrative data, Statistics Portugal is not always successful in accessing administrative data.

To solve this problem in the Peer Review report of March 2015, on compliance with the code of practice and the coordination role of the national statistical institute, is mentioned that Statistics Portugal (i) will make proposals to change the National Statistical Law in order to ensure timely and effective access to administrative data for statistical appropriation (ii) will make proposal to change the statistical law in order to ensure Statistics Portugal right to be involved in the design and modification of any administrative data system in order to improve its suitability and quality for statistical purposes.

Beyond these mandatory terms, Statistics Portugal has established collaborative processes, signing memorandum agreements and protocols with several institutions, to provide and share data and geospatial data. The most relevant examples are the Memorandum Of Understanding between Statistics Portugal and the Directorate-General for Territory– the Portuguese National Mapping and Cadastral Agency (NMCA), to access national official geospatial reference data (ex. orthoimagery and official administrative boundaries map - and the Protocol Agreements celebrated with all the Portuguese Municipalities for the development of a common base geography for census.

Most recently, Statistics Portugal has as well established an agreement with ADENE, the national authority responsible for managing the Energy Certification System of Buildings, to share data and location regarding Statistics Portugal's National Buildings and Dwellings register.

In addition, Statistics Portugal is a full member of the Advisory Council of Geographic Information National System (CO-SNIG) for INSPIRE, which is the strategic body responsible for the implementation of the Directive in Portugal.

At a technical level Statistics Portugal is also strongly engaged in the INSPIRE Cross Thematic Working Group, mainly focused on metadata and services, as well as in five out of nine INSPIRE Working Thematic Groups, which were created similarly to the INSPIRE Thematic Clusters.

Within SNIG (National Portuguese Spatial Data Infrastructure) and the implementation of INSPIRE (mainly under the Law "DL nº 180/2009, from 7 of august" that constitutes the legal basis for SNIG constitution and activity and for the application of INSPIRE in Portugal), Statistics Portugal also contributes to the national repository of geospatial data. For the INSPIRE themes under its

---

<sup>3</sup> [Law Nº 22/2008 of 13 May 2008](#)

responsibility Statistics Portugal provides metadata for INSPIRE compliant datasets and services to the national geospatial data metadata catalogue and also provide geospatial information through services.

## 2.4 Legal basis for geospatial data access for Statistics Austria

The legal basis for access to the geocoded Buildings and Dwellings Register, which is used for geo-enabling other administrative data sources used for statistics, is regulated both in the Census Act<sup>4</sup> and the Buildings and Dwellings Register Act<sup>5</sup>.

The Census Act regulates the access to administrative data sources for statistical purposes and hence the Buildings and Dwellings Register can be used for statistical purposes as it is an administrative data source. The contents and structure of the data in the Buildings and Dwellings Register and its access of use are regulated in the Buildings and Dwellings Register Act.

Besides the unique Id's and structural data on buildings and dwellings this includes a pair of geocodes as spatial reference for the address and building respectively, and can therefore be considered to be the backbone of geocoded registers and geospatial data at Statistics Austria.

## 2.5 Acquisition and processing of geospatial reference data in Poland

TERYT is the register of Poland's territorial division. It is maintained by the Central Statistical Office (CSO) and contains descriptive information about territorial geographies, localities, streets, buildings and dwellings. In order to use geographic data in Polish censuses, this data needed to be obtained for statistical units and dwellings.

Maintaining address locations lies in the scope of work of national NMCAs. Unfortunately, at the time such data was needed for the census, the Polish NMCA did not possess a complete set of address points for the whole country. Therefore address points had to be prepared by official statistics with the use of various reference materials.

Two sources of geographic information have been maintained by official statistics: on topographic and cadastral maps at various scales and coordinate systems, boundaries of statistical division of the country were drawn and updated throughout the years. Also, since 1978, situation sketches of household locations have been maintained and updated regularly in regional statistical offices and by enumerators in the field during censuses. All of these paper materials had to be transformed into digital maps. The first step of the process was scanning and georeferencing all statistical maps. Boundaries of statistical geographies and census areas were then digitized, maintaining compliance with boundaries of municipalities.

Carrying out both censuses gave CSO the legislation tools needed to acquire spatial data from any state agency free of charge. Following datasets have been acquired: National Register of Boundaries and Areas of the Country's Administrative Division, cadastral data, Land Parcel Identification System (LPIS), Topographic Data Base and ortophotomaps.

---

<sup>4</sup> Census Act (German version):

<https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20004583>

<sup>5</sup> Buildings and Dwellings Register Act (English and German versions:

[http://www.statistik.at/web\\_en/publications\\_services/online\\_address\\_buildings\\_and\\_dwellings\\_register/legal\\_basis/index.html](http://www.statistik.at/web_en/publications_services/online_address_buildings_and_dwellings_register/legal_basis/index.html)

[http://www.statistik.at/web\\_de/services/adress\\_gwr\\_online/allgemeines/gesetzliche\\_grundlagen/index.html](http://www.statistik.at/web_de/services/adress_gwr_online/allgemeines/gesetzliche_grundlagen/index.html)

National Register of Boundaries and Areas of the Country's Administrative Division contains boundaries of municipalities, counties and voivodeships. The municipality area boundary is superior to statistical region and census area boundaries and hence has been adopted as a spatial reference – boundaries of statistical geographies have to be collinear with municipality boundaries where applicable.

Cadastral data in Poland is maintained at a county level. A unified standard exists for exchanging this data, however mapping authorities in each of the 379 counties have their own way to interpret this standard. Furthermore, the data is maintained in various coordinate reference systems, some of them being local coordinate systems. A huge effort had to be put to process this data to make it useful for address point acquisition purposes.

Land Parcel Identification System (LPIS) contains cadastral parcel boundaries for whole of the country excluding big cities. This data served as a secondary reference material where cadastral data was not present in numeric form or could not be processed.

The Topographic Database is a system for dissemination of high quality spatial data with address points as one of its feature classes. Unfortunately at the time of address point acquisition in Polish official statistics, The Topographic Database covered only approximately 10 percent of Poland.

Ortophotomap for the whole of Poland has also been acquired from the National Mapping Agency to support visual identification of buildings in the process of spatial address database creation.

Some of the above mentioned reference materials – those covering the whole of Poland (National Register of Boundaries and Areas of the Country's Administrative Division, Land Parcel Identification System, and ortophotomap) were directly after acquisition by CSO delivered to the regional units along with a set of scanned and georeferenced statistical maps. The Topographic Database and cadastral data needed some initial processing before they could be taken advantage of.<sup>6</sup>

## 2.6 GIS Resource Centre in Statistics Norway

Up until 2016, use of geospatial data has been decentralised in Statistics Norway. Geospatial data processing was limited only to a few divisions.

No one in Statistics Norway had the overall coordination or responsibility for an overview of the infrastructure, technical needs, GIS-expertise or user needs for GIS and geodata. There was a weak understanding of the informal – and formal responsibilities between the departments in Statistics Norway concerning: Obtaining, quality control and storage of spatial data, operation and maintenance of the geodata base, facilitating and improvements of the infrastructure for GIS tools, and dissemination of the result as geospatial data. Not enough funding and competence was allocated to these tasks.

In order to extend the use of geospatial information to all relevant divisions and to meet needs for new and better statistics, Statistics Norway decided to establish a “GIS resource centre”.

---

<sup>6</sup> Mirosław Tadeusz Migacz “Geostatistics Portal – a platform for statistical data geovisualization”, Statistical Journal of the IAOS 31 (2015) 463–470, DOI 10.3233/SJI-150920, IOS Press 2015, paper written by Mirosław Tadeusz Migacz CSO Chief GIS Specialist.

The resource centre was initiated as a project, but after the project phase the goal is to streamline use of geospatial data and integration with statistical and administrative data into the work of the sectorial statistics divisions. The resource centre will still be conducting some central tasks, but first and foremost, provide operational support for the organisation along the GSBPM.

The resource centre now coordinates and make sure that the infrastructure and technical needs regarding GIS and geospatial data are met.

It can be assumed that some divisions will embrace the use of geospatial data to a large extent and thus maintain the know-how and competence necessary, yet other divisions may need more help and support. Some tasks may even be done solely by the resource centre in support of the division/statistics at hand. The vision is to provide easy access to tools and software and strengthen the competence in the organisation in such a way as to make it as natural for the statistician to include geographical analysis as it is to use conventional statistical processing tools.

In the process of setting up the resource centre, Statistics Norway made the following strategic considerations:

- Tasks close to data and the subject at hand could preferably be decentralised.
- Tasks regarding maintenance of competence and effective use of geospatial technology should preferably stay centralised.

The resource centre provides internal GIS courses as well as organising special courses with external lecturers. A regular internal “GIS user forum” has been set up to discuss matters of common interest as well as to inform about new features and present best practice cases.

In connection with the annual updating of the operational planning, the resource centre makes an inquiry to all heads of division asking them to forecast the need for assistance regarding GIS in any part of the GSBPM.

Currently, the resource centre is located within the Division for natural resource- and environmental statistics. This is because this division is the biggest single user of GIS and it ensures competence and knowhow regarding production of statistics based on geospatial data.

## 2.7 Resource setup in Statistics Finland

Statistics Finland’s *GIS strategy*, as well as GIS related notations in the *ICT strategy for 2015 to 2018* promoted and created a legitimate status for geospatial data and techniques at Statistics Finland.

Both strategies recommended that a *GIS technology review* should be carried out. The GIS strategy also affected the management of geospatial infrastructure, relocation of personnel (to a *GIS core team*) and founding of an in-house *GIS expert group*.

The position of the GIS core team, responsible for geospatial core data and techniques, was strengthened in the ICT Management department. The GIS technology review was carried out in 2015 to 2016. The report offered three alternative technological solutions on how to handle geospatial data and its requirements in data collection, production and dissemination. A so-called hybrid solution was chosen. The solution combines Open Source technologies with licensed GIS technologies. E.g., Open Source products were mainly the chosen solutions for server-side

functionalities or for light thematic mapping purposes (Geoserver, QGIS). By contrast, licensed products were chosen (ESRI ArcGIS, SAS, SQL Server spatial data type) for professional use, e.g., for spatial analysis or spatial data editing. Planned renewals of geospatial production and new techniques were combined with the GIS Technological Renewals project (2016 to 2017).

In addition, main geospatial codes of practices were made for future actions. They included seven principles, from strengthening of the recognition of the geospatial dimension in statistical production processes to quality aspects of spatial data use and acquisition, as well as the importance of in-house support and instructions for everyone.

Based on Statistics Finland's experiences, the commitment of the organisation's management is a prerequisite for a successfully implement common GIS infrastructure and technological choices. A GIS Strategy is cross-statistical and it is needed through the statistical production process. In order to implement the strategy, adequate resources with capabilities need also to be allocated. Planning the use of chosen technologies and an organised way of acquiring new technologies, gives perseverance and structure for operation and development.

Well-organised spatial data, technology and resource management provide awareness to the organisation of its own know-how and sufficiency of the chosen technology in relation to the objectives of the organisation's strategy.

## 2.8 Resource setup in Statistics Poland

### Human resources

For the purpose of trial censuses, statistical address point databases for approximately 20 municipalities have been prepared by geoinformation staff of the Central Statistical Office (6 persons). They were the starting point for employing GIS technology in the censuses and nationwide address point creation.

The Central Statistical Office (CSO) has 16 regional offices (one per voivodeship), each has few branch offices. Overall there are 64 regional units in Polish official statistics – each with a department responsible for maintaining the descriptive part of the register of Poland's territorial division (TERYT). Each unit has a certain number of counties in its scope. 210 employees from these units have been assigned to spatial address database creation works and trained in basics of ArcGIS. A CSO unit comprised of 5 GIS specialists had been assigned to offer them constant support. Also, several illustrated tutorials have been created in CSO to support the works.

With geospatial data on statistical geographies present since October 2009, CSO could focus on address point acquisition. Maintaining address locations lies in the scope of work of National Mapping Agencies. Unfortunately, at the time such data was needed for the census, the Polish Mapping Agency did not possess a complete set of address points for the whole country.

In order to use spatial data in the Agricultural Census 2010, the spatial address databases with nationwide address point coverage had to be ready by June 30th 2010. For that to happen, reference materials had to be collected, processed and delivered to the regional units<sup>7</sup>.

Nationwide address point acquisition took place before the Agricultural Census 2010 and lasted 6 months. The statistical address point database is maintained and quarterly updated by GIS operators in the regional statistical offices (currently approx. 160 people).

### **IT resources**

The Geostatistics Portal infrastructure ensuring the functioning of the website and the map application and contains 11 identical servers with different roles: 2 web servers, 3 application servers, 2 database servers and 4 mapping servers.

Backups of all servers are performed cyclically according to the defined schedule. Applications, which are part of the Geostatistics Portal System, are configured to ensure uninterrupted operation of the Portal. In the event of failure of a server, application or service, the system automatically switches to the backup environment and works on it until the failure is dealt with. From the point of view of the user failures are not noticeable.

The Spatial Address Point Database System relies on 17 servers: a central server located in the CSO and 16 voivodship servers - one in each regional statistical office. The central server holds reference materials for GIS operators. The data is automatically distributed through regional servers to operators' workstations. Operators at the end of each update of databases export the results of their work to regional servers, where the data is automatically merged and sent to the central server to create a continuous database for the whole of Poland.

## **2.9 Resource setup in Statistics Portugal**

Statistics Portugal has a Cartography/GIS Unit since the final of 1990 decade. Nowadays the Geoinformation Unit, which is integrated in the Department of Methodology and Information Systems, has 25 technicians with GIS expertise distributed by the central Lisbon Office and by the 4 regional Delegation Offices of Statistics Portugal.

The use of cartography has supported census data collection at Statistics Portugal since 1981. In 1995, Statistics Portugal started the preparation of the 2001 census cartography, which was named "Geographic Information Referencing Base" (BGRI 2001) and was based on Geographic Information Systems. Since 2006, with the production of the BGRI 2011 to support the 2011 census, Statistics Portugal has been developing a Spatial Data Infrastructure (SDI) and carrying out other statistical activities in a permanent effort to introduce the spatial perspective across the different phases of statistical production.

The SDI is currently being used, in a transversal way, at Statistics Portugal activities, promoting the integration of the spatial component in the statistical production process, in order to achieve efficiency and accuracy, within several domains such as the sampling process, the data collection or the dissemination of statistical information.

---

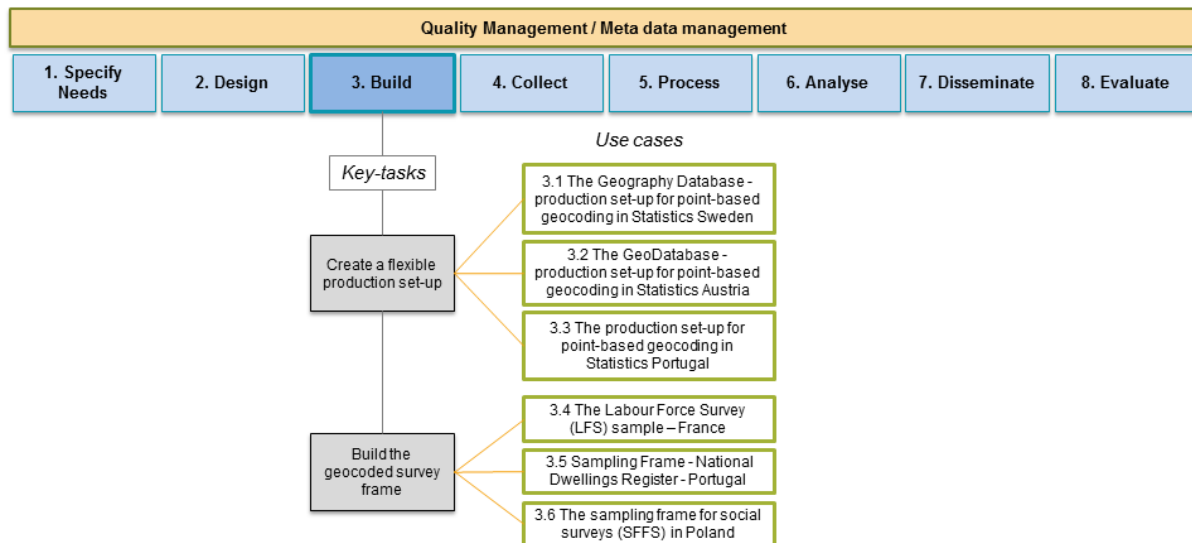
<sup>7</sup> Mirosław Tadeusz Migacz "Geostatistics Portal – a platform for statistical data geovisualization", Statistical Journal of the IAOS 31 (2015) 463–470, DOI 10.3233/SJI-150920, IOS Press 2015, paper written by Mirosław Tadeusz Migacz CSO Chief GIS Specialist.

Regarding the technological components, Statistics Portugal has a hybrid solution. Commercial software for desktop and server solutions (ESRI ArcGIS, ORACLE, SQL Server spatial data type) are combined with open source software (GeoServer, QGIS, Pmapper).

All the geospatial dataset are filed in an ArcSDE/Oracle geodatabase that is managed by GIS and ABD administrators. The process requires a close work relationship between the Statistics Portugal Geoinformation Unit and IT team. The process to update the spatial data is conducted by Statistics Portugal within the Geoinformation Unit.

Despite the work that has already been developed regarding spatial information within Statistics Portugal, it's still lacking an internal and national global spatial data strategy for the next years. The INSPIRE Implementation in Statistics Portugal and in several other national institutions will be one of the drivers to achieve it.

### 3. Build

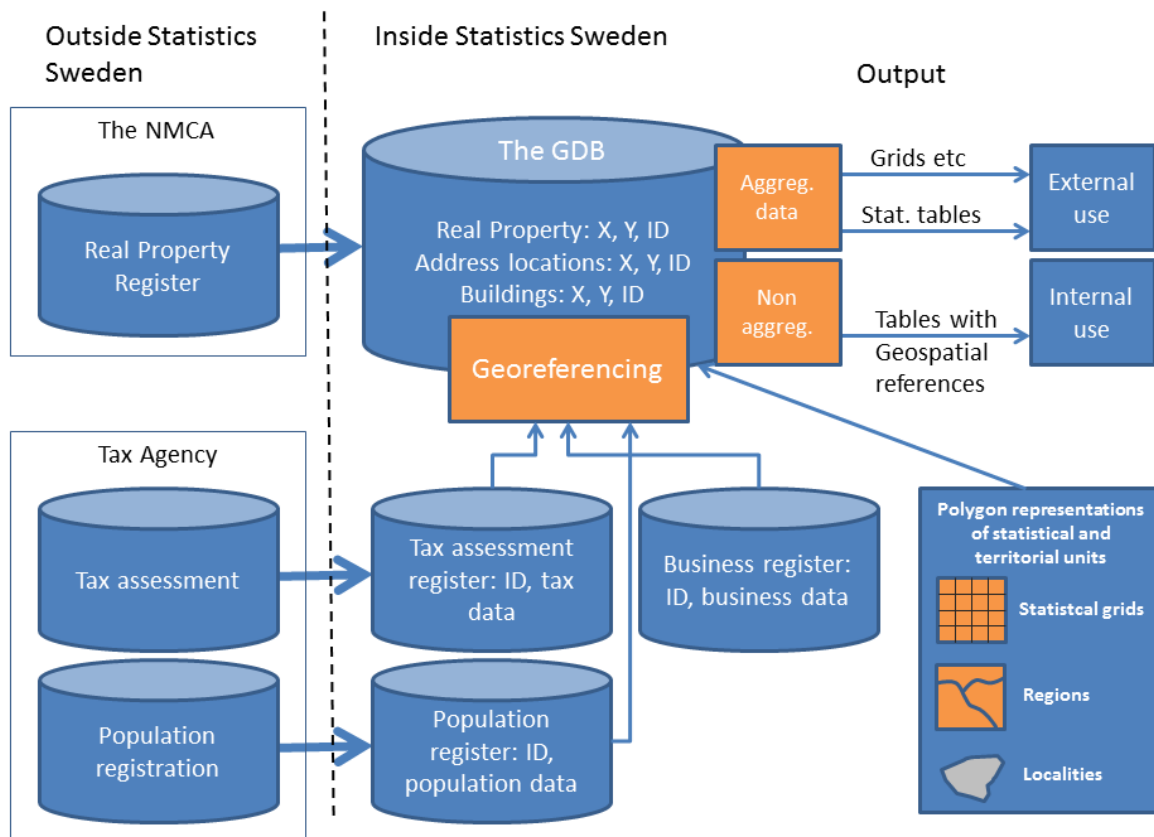


#### 3.1 The Geography Database - production set-up for point-based geocoding in Statistics Sweden

The “Geography database” (GDB) is the main environment for geocoding in Statistics Sweden. In essence, the GDB is a copy of the National Real Property Register maintained by the NMCA stored in an MS SQL environment. The Real Property Register comprises a national authoritative location data framework. Data is transferred automatically by means of weekly notifications from the Real Property Register to Statistics Sweden and loaded into the GDB. The GDB is basically a location reference database used to link various spatial locations with data from different administrative records using a set of identifiers. As such, the GDB does not contain statistical information.

All cadastral parcels, buildings and address locations have unique identifiers which correspond to identifiers used in unit record data. For example, a cadastral parcel is identified by its municipal belonging (municipality code) + the name of the real estate unit and its block code (Fastighet 1:1). In a similar way, an address location is identified by municipality code, street name and number + code of dwelling. Since 2012, each object also has a UUID.

Figure 2 below illustrates the role of the GDB and its relations to other data sources outside and inside Statistics Sweden.



**Figure 2: The Geography Database (GDB) and its relations to other data sources**

Each object in the Real Property Register (coordinate points of cadastral parcels, address locations and buildings) is tagged with references to official administrative geographies, such as the municipality or county to which it belongs. To add additional references to territorial or statistical geographies (not found in the Real Property Register), in the GDB, each object will have additional information on the most common output geographies, such as postal code area, electoral district, locality, small statistical area, NUTS regions, grid cells, etc. assigned to it. Polygon representations of desired geographies are loaded into the GDB. Attributes from polygon features are transferred to point data by use of simple point-in-polygon arguments.

As the GDB contains a variety of references to common geographies, processing of data can mostly be conducted without using GIS software. By means of SQL statements data from the GDB can be instantly linked with unit record data from the Business Register or the Population Register etc. and aggregated on the basis of the spatial references stored in the GDB. In case aggregations are needed for output areas not present as geographies in the GDB, data needs to be exported to GIS environment for further processing.

### 3.2 The GeoDatabase - production set-up for point-based geocoding in Statistics Austria

The GeoDatabase (GDB), an IBM DB2 database including spatial extender, is the central data base environment for geocoding in Statistics Austria. The central part of the GDB is a copy of the Buildings and Dwellings Register maintained by the municipalities with support from Statistics Austria.

The Address Register and the Buildings and Dwellings Register have a common recording pathway but for reasons of access and usage rights, they are managed separately; the Address Register at the national mapping agency (BEV) and the Buildings and Dwellings Register at Statistics Austria. The establishment of a common recording pathway guarantees consistent management of the address and building data in both registers. It also ensures that the municipalities, responsible to maintain the two registers, do not have to enter data twice. Data processing can be done either via an Internet application (Web client) or via an interface (XML client).

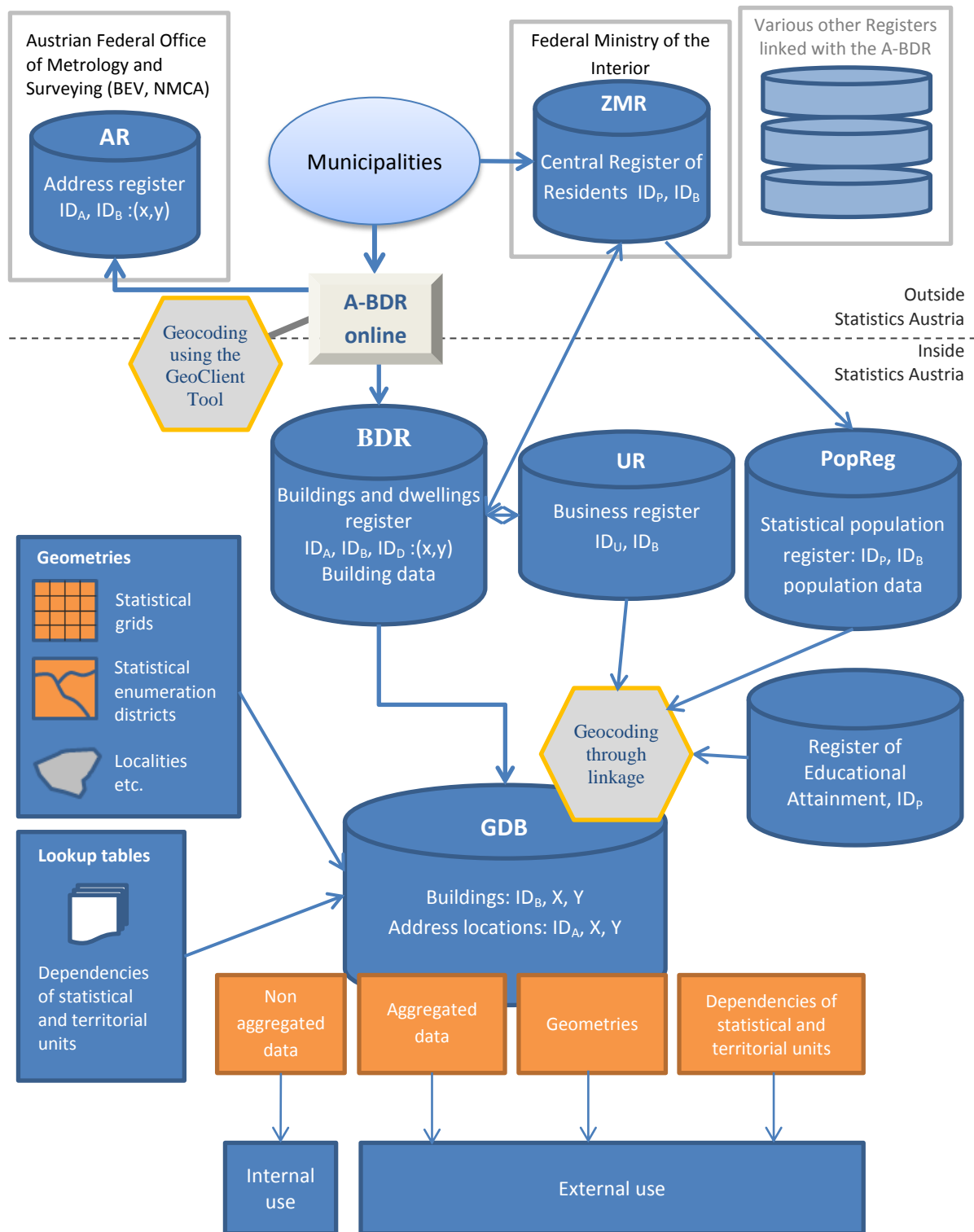
The Buildings and Dwelling Register is the connection to other national registers, for example the Central Register of Residents, and to local registers in the municipalities.

As Statistics Austria is the service provider for the maintenance of the Buildings and Dwellings Register, data is available at any time. However copies are loaded into the GDB only when needed and kept for a certain period to ensure consistency (e.g. until all tables for the same “as-of-date” have been aggregated). The GDB is basically a location reference database used to link data from different administrative records with various spatial locations using a set of identifiers. All addresses and buildings have unique identifiers which correspond to identifiers used in unit record data. As such the GDB serves as basis for the aggregation of statistical data to all available geometries.

Each object in the Buildings and Dwellings Register is tagged with the reference to the municipality to which it belongs. In addition the local reference to the postal code area and the Statistical Enumeration Districts is included. To add additional references to territorial or statistical units (not found in the Buildings and Dwellings Register), the GDB contains historicised Lookup-Tables showing the dependencies for each object to the most common output areas, such as grid cells, NUTS, electoral district, etc. assigned to it. Polygon representations (geometries) of territorial units are also part of the GDB.

As the GDB contains a variety of references to common geographies, processing of data can mostly be conducted without using GIS software. E.g. unit record data from the statistical population register can be linked with data from the GDB by SQL statements and aggregated on the basis of the Lookup tables stored in the GDB. In case aggregations are needed for output geographies not present as spatial references in the GDB, data is exported to the GIS environment for further processing.

Figure 3 below illustrates the source and role of the GDB and its relations to other data sources outside and inside Statistics Austria.



**Figure 3: The GeoDatabase (GDB) and its relations to other data sources**

### 3.3 The production set-up for point-based geocoding in Statistics Portugal

The process to maintain and update the SDI spatial data is conducted by Statistics Portugal within the Geoinformation Unit. Statistics Portugal Geoinformation Unit has 25 technicians with GIS expertise that manage and edit the several features of the GDB. The Geoinformation Unit has 17 members at

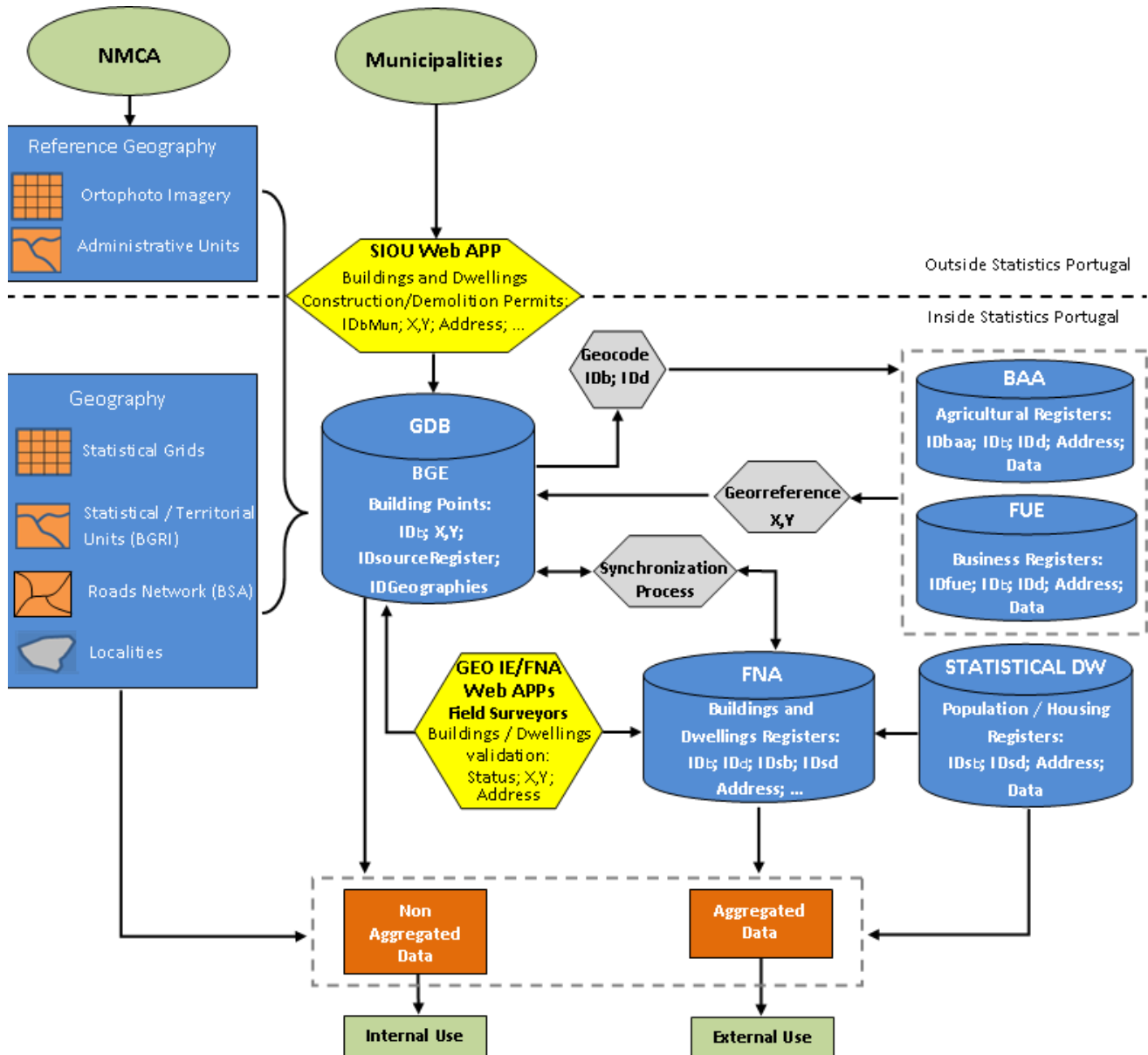
the central Lisbon Office and 8 technicians distributed by the 5 Statistics Portugal Regional Offices (the “regional staff” is an extremely important link with municipalities).

One of the main focuses of statistics Portugal Geoinformation unit is set on editing and updating the Buildings Geographical Database buildings points and addresses (including the buildings dwellings addresses).

The Buildings Geographical Database (BGE) is being continuously edited in a quality control process in order to increase the accuracy of X,Y geometric position and its consistency with the addresses of the buildings. The Statistics Portugal Department of Data Collection, in straight relation with the Geoinformation Unit, has a strong contribution in this process. Following the National Dwelling Register constitution, Statistics Portugal has developed internal applications (GEOIE/FNA Web APPs) for the interviewers that conduct the field surveys to report errors and updates in the location or the addresses of the buildings and dwellings of the Buildings Geographical Database.

The update process regards the inclusion of new residential buildings and also other types of buildings, with no residential capabilities (such as public facilities), at the Buildings Geographical Database. The work is developed through, (1) the Indicators System of Urban Operations (SIUO) dynamically maintained by the Municipalities, using an electronic web based platform, concerning the urban operations related with building permits and completed construction work and based on (2) the Business Register georeferencing process.

All the editing work is done with GIS software, desktop and WEB applications developed in house or acquired abroad. The reference data (orthoimagery) and the official administrative boundaries for this activity are being provided by Portuguese National Mapping Agency under the existing Memorandum of Understanding.



**Figure 4: The GeoDatabase (GDB) and its relations to other data sources**

### 3.4 The Labour Force Survey (LFS) sample in INSEE

The French Labour Force Survey (LFS) has been an area survey for decades. Each area consists in 20 contiguous dwellings. The sampling frame has not been geocoded until 2009. Since such a geographical level does not exist among the various official or statistical geographical French classifications, Insee has had to construct these areas using only paper maps. The work load was very heavy (140 000 hours of working time over 2 years every ten years). In 2009, the reshaping of the sampling design relied on a geocoded sampling frame which enabled an automatic construction of the areas by a small team within a few months.

The Labour Force Survey (LFS) sample is drawn from the tax files that have the cadastral parcel ID among their variables. Cadastral information is widely used to automatically construct small specific areas of 20 main dwellings to facilitate fieldwork.

The new sampling frame for all the French household surveys will be geocoded. The location of the statistical units will be used to improve the design of the sampling, for example, to ensure that the sample is well spatially distributed.

Three chapters of the future Handbook on spatial statistics, being written at INSEE, will be devoted to highlighting the links between survey sampling and spatial information. The first will deal with the improvement of sampling design. The second will focus on the use of statistical spatial tools with non-exhaustive data. The latter will address the issue of small area estimation.

### **3.5 Sampling Frame - National Dwellings Register - Portugal**

Between 2006 and 2010 Statistics Portugal conducted a fieldwork to guarantee that the Master Sample 2001 could be used until 2013, the year of the transition to the new sampling frame, the National Dwelling Register (FNA).

The National Dwelling Register, created from the 2011 census micro data (buildings and dwellings), constitutes information of statistical production support infrastructure, from which sampling frames and samples are drawn in order to support Statistics Portugal social surveys.

The National Dwelling Register is updated with information mainly from surveys and from the Urban Operations System – SIOU. This last source represents administrative data from the Portuguese Municipalities, concerning urban operations which are related with building permits and completed construction works. In the future it is foreseen that the National Dwelling Register will be updated also with external administrative sources. This process is a major challenge given its size and complexity.

The lack of a common identifier between the different sources requires the analysis and comparison of non-harmonised addresses, which constitutes one of the main constraints to the National Dwelling Register update process.

#### **New Geographical Referential**

The Spatial Data Infrastructure (SDI) of Statistics Portugal provides the foundation for the geography of the National Dwelling Register. Regarding the National Dwelling Register, the SDI main components are the Buildings Geographical Database and the GEOSTAT European 1x1 square kilometer GRID (Grid\_ETRS89\_LAEA\_1K). The National Dwelling Register comprises all the residential dwellings of the country which are integrated in the Buildings Geographical Database.

Each building has assigned an address, coordinate of its location, census variables and its relation to the GEOSTAT European GRID.

The main aspects underlying the creation of the new Portuguese geocoded sampling frame were (i) relating the Census Buildings to the GEOSTAT European GRID; (ii) order each 1Km<sup>2</sup> grid cell in each NUTS 3 area; (iii) define the geographical primary units (PU) that consist on a subset of the 1 Km<sup>2</sup> grid cells.



### 3.6 The sampling frame for social surveys (SFFS) in Poland

The sampling frame for social surveys (SFFS) includes information on persons and dwellings. Since 2015 the SFFS has constituted a basis for the creation of sampling frames for the social surveys carried out by the Central Statistical Office, as well as aided the conduction of statistical analyses. The first supply of data to the Sampling frame for social surveys included the address and dwelling sampling frame created for the purposes of the 2011 National Census on the basis of the register of Poland's territorial division (TERYT) and the PESEL population register, which included four types of structures connected to one another with unique IDs:

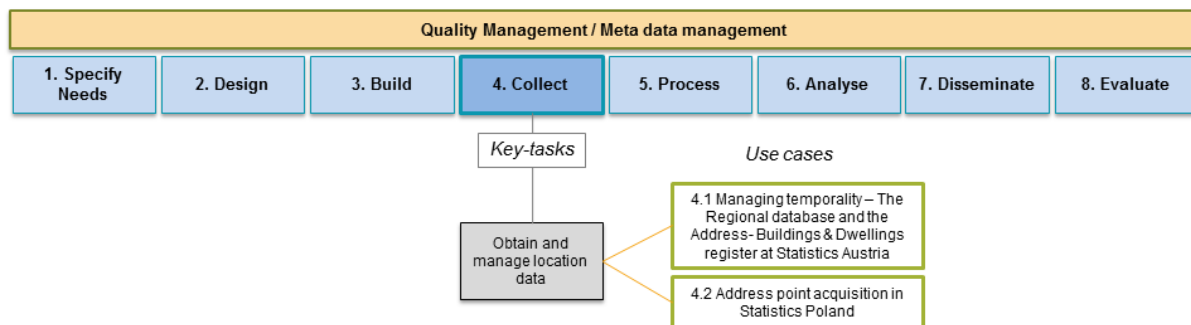
- buildings,
- dwellings,
- persons and
- collective accommodation establishments.

According to the Act on official statistics, the present structure of sampling frame includes data on persons in respect of their:

- name, date and place of birth, sex,
- PESEL (population register) No.,
- citizenship, marital status, date of marriage, date of marriage termination,
- education, profession, type of workplace or place of education,
- tax identification number, title of insurance, excluding the part of the code, which is subject to confidentiality requirements,
- address of permanent and temporary residence and/or mailing address including the x,y coordinates of address points,
- e-mail address and telephone number.

In general, geospatial data are included in sampling frames but are not directly used in the process of creation of survey frames.

## 4. Collect



### 4.1 Managing temporality – The Regional database and the Address-Buildings & Dwellings register at Statistics Austria

The administrative or regional unit to which a building belongs to at a set point of time has traditionally always been saved hard coded with the statistical base data. This has made it difficult in the past to react to administrative changes and aggregating statistical data for time lines required special attention. Changes such as the merging or splits of municipalities or changes in the boundary could not be followed both ways, data from the past to current geometries or current data to former geometries.

To allow for a full historicisation of associated data and its regional aggregates two aspects of temporality of a building need to be historicised. The first has to do with the life cycle of the building itself, describing the period during which data can be associated with the building. This aspect is directly handled in the building and dwelling register.

The second aspects deals with the regional belonging of a building, which is saved together with the address details of a building directly in the Buildings and Dwellings Register, but only for the current point of time. However the regional belonging might change from time to time, when affected by administrative changes. To overcome the problem of this temporal aspect Statistics Austria is currently working on optimising the look-up system for the Buildings and Dwellings register by creating a regional database which should be used during production as well as for the subsequent data aggregation.

#### The life cycle of a building and its building address

To manage temporality it is important to work with date fields (start date and end date of validity) and status fields to describe the status during the validity phase of a record (Address, building, dwelling,...). Historic records are still necessary to be kept for later data joins, so records in registers should never get deleted, but their status should be set to 'not active' and an end date given.

It is equally important to record status changes with date of validity in all linked registers and associated geometries and dependencies in order to allow for correct linkages for any date of interest. Relevant for this at Statistics Austria is the Buildings and Dwellings Register as well as the new project of the regional database, described below.

The Address- Buildings & Dwellings register is a register system consisting of the address register run by the mapping agency and the buildings and dwellings register run by Statistics Austria. The address

register is also part of the Buildings and Dwellings Register and contain parcel addresses which are the bases for the building addresses.

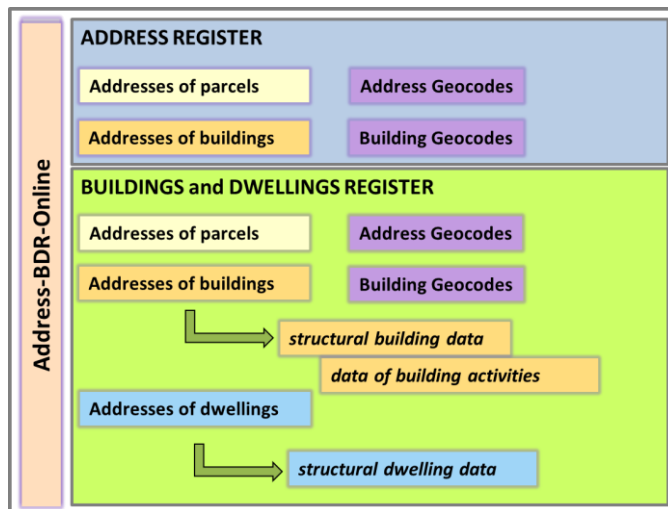


Figure 8: The content of the Buildings and Dwellings Register

These parcel addresses have a unique address code, address information including geocodes, a status field and validity from and to dates.

ADR CODE	MUNIC. CODE	ENUM DISTRICT-ID	MERID	RW	HW	STATUS	FROM	TO
7209214	32415	004	M34	23841,3	330631,1	active	2003-10-17	2999-12-31

Table 1: Address register: example of some fields with regional and status information

The life cycle of a building starts with its building permit and construction phase and the entry as construction activity in the building activity table, which is a separate table in the Buildings & Dwellings Register. The building-code (unique lifelong identifier) is assigned automatically and the building address defined. The building address also includes the corresponding address code from the active parcel address; the address code has to be selected in the application.

This record is saved with a starting date “FROM” and the end date “TO” is set to 2999-12-31 (meaning open) and further construction details added to the record, including a type field “CONSTR. ACTIV.” describing the type of the construction activity as ‘New building’.

ADR CODE	BUILDING CODE	MUNIC. CODE	CONSTR. ACTIV.	MERID.	RW	HW	STATUS	FROM	TO
7209214	2665547	32415	N	M34	23850,4	330625,9	open	2010-05-12	2999-12-31

Table 2: BDR: building activity table: example of some fields with regional and status information

When the construction of the building is finished, the type of the construction activity is changed to ‘finished’ and the date saved as change date in the building record. In fact the record with construction type ‘New building’ is saved as historic record with an end date.

At the same time this building is integrated in the building stock and a building record saved in the main table of the Buildings and Dwellings Register. The building address details are copied from the record of building activity table, further structural data of the building is entered and status (“active”)

and date fields added. From that moment on it is considered a building and can be used by other integrated register systems, e.g. the central register of residents to register residents.

ADR CODE	BUILDING CODE	MUNIC. CODE	MERID.	RW	HW	STATUS	FROM	TO
7209214	2665547	32415	M34	23850,4	330625,9	active	2012-08-22	2999-12-31

**Table 3: Buildings and Dwellings Register, example of some fields with regional and status information**

Processes that trigger historicised entries in the tables of the Buildings and Dwellings register are changes of status information and changes of geocodes. Currently the recorded dates (from, to) often represent the date of change in the Buildings and Dwellings register and not the date of validity of a building object or an address. Also as mentioned above, the regional attributes are only saved according to the current situation, so the change information on regional matters is dealt with in separate processes. In the future this will be covered by a new and central application for the Buildings and Dwellings register and Geodatabase (the regional database) currently being developed. So dates of validity will be included and the assignment to regional entities will match the date of validity.

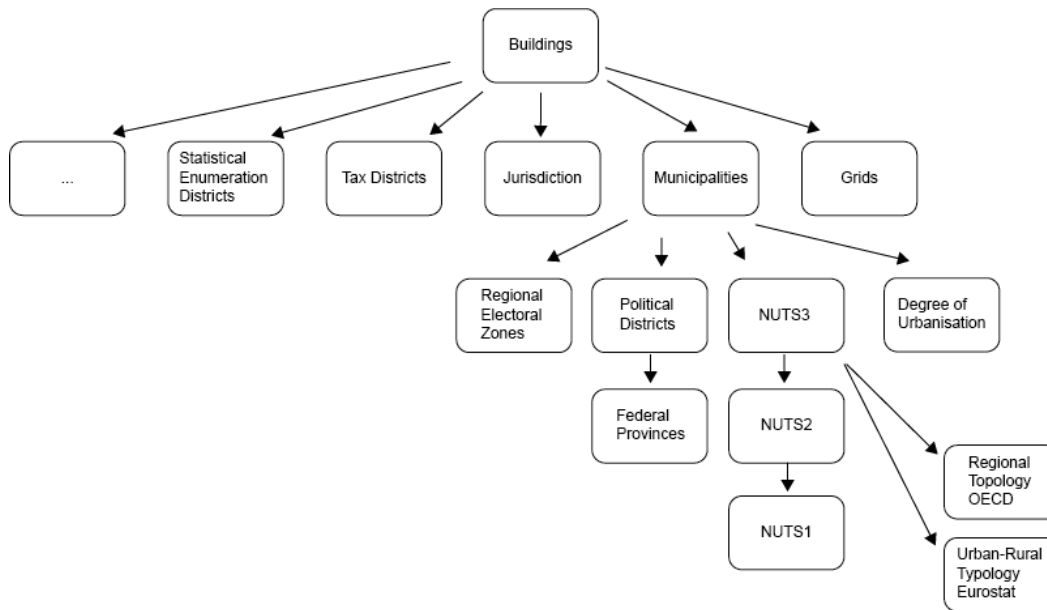
### The regional database

As mentioned above the regional belonging of a building might change from time to time, when affected by administrative changes or when changes of typologies occur affecting the dependencies. The basic regional information is stored in the Buildings and Dwellings register for the current date with the building data. It needs to be kept up-to-date and the information has to get provided to the Buildings and Dwellings register. The changes should also be stored as historic entries to allow data aggregations for different points of time.

To overcome the problems of this temporal aspect Statistics Austria has started an internal project to set up a new optimised look-up system for the Buildings and Dwellings register during production as well as for the subsequent data aggregation. The look-up system is based on the regional database and will be included in the geodatabase (see case 3.2 *The GeoDatabase - production set-up for point-based geocoding in Statistics Austria*) providing historicised look up tables for the dependencies. The accompanying application has the aim to administer all regional breakdowns within Austria in the regional database with a validity time stamp. All different typologies (nearly 20) like administrative geographies (e.g. federal provinces, municipalities), breakdown of areas by function (e.g. regional electoral zones) or geometric area divisions (grids) with different spatial resolutions (e.g. 9 federal districts vs. 8825 statistical enumeration districts) will be included.

The main focus of the regional database is on:

- Classification of all regional breakdowns
- Dependencies between regional breakdowns (e.g. NUTS are dependent on municipalities, e.g. Grids are dependent on the geocodes of the buildings) for different points of time
- The storage of historicised regional entities
- And the simple expandability for new regional classifications.



**Figure 9: Typology of regional breakdowns**

### Dependency of regional breakdowns

There are different hierarchy levels of the regional breakdowns, most of which are built up from municipality level, but all are connected to the buildings of the Buildings and Dwellings Register. So with the storage of the location as geocodes and the municipality code all regional breakdowns should be derivable. The regional database and the Look-up-Tables of the Buildings and Dwellings Register are linked. A split or merge of municipalities is registered in the regional database and is synced automatically to the Buildings and Dwellings Register.

The creation of the geometries for all these historicised breakdowns is a separate process that will use the data stored within the regional database.

### Historicisation

To achieve a full historicisation of the data, all changes must be logged and stored with a “from-” and “to-” date for its validity. This applies to modifications of the typologies which usually take place as of 1<sup>st</sup> of January, to administrative changes which can also be effective by any date of a year (e.g. as of 1<sup>st</sup> of May) and also to buildings of the Buildings and Dwellings Register where daily changes are made.

The changes of the typologies can be

- Split, merge or delete of a regional entity
- Change of the name
- Renumeration of entity identifier

Some changes (e.g. split, merge or delete of municipalities) automatically trigger the change of the dependency tables, so both, the list of regional breakdowns as well as the corresponding dependency tables, have to be stored historically.

**Example: Merging of municipalities**

4 Municipalities with codes 61702, 61704, 61707 and 61732 are merged to be one on 1.1.2015. The new municipality got the new code 61756 as of 1.1.2015. This information is stored in the list of regional breakdowns.

MUNIC. CODE	FROM	TO
61702	1.1.1900	1.1.2015
61704	1.1.1900	1.1.2015
61707	1.1.1900	1.1.2015
61732	1.1.1900	1.1.2015
61756	1.1.2015	31.12.2999


**Table 4: List of regional breakdowns**

The corresponding dependency tables are still being developed, but will contain the information to which higher regional entity each municipality belongs to before and after the merge and will also include possible changes of the dependencies previous to or following the merge. Yet it is not clear at this stage how far back the dependencies will be integrated.

This change of the municipality code affected all buildings of the four former municipalities and their records in the Buildings and Dwellings Register were updated as of 1.1.2015. The change is also saved as historic records for all of the buildings in the regional database.

As in this case the former municipality areas were incorporated as separate enumeration districts of the new municipality, a backward and forward historicisation is also possible on the bases of enumeration districts as long as the predecessor / successor information can be derived. The predecessor / successor information will be incorporated in the regional database.

MUNIC. CODE	ENUM DISTRICT-ID	FROM	TO
61702	61702000	1.1.1900	1.1.2015
61704	61704000	1.1.1900	1.1.2015
61707	61707000	1.1.1900	1.1.2015
61732	61732000	1.1.1900	1.1.2015
61756	61756000	1.1.2015	31.12.2999
61756	61756001	1.1.2015	31.12.2999
61756	61756002	1.1.2015	31.12.2999
61756	61756003	1.1.2015	31.12.2999



**Table 5: Table illustrating the predecessor / successor information to be incorporated in the regional database.**

**Example: Change of boundary between two municipalities**

This example deals with the change of boundary between the two municipalities 61710 and 61756 affecting only a few buildings. This information is stored as historicised building information in the regional database, which for building-ID 1301211 looks like this (see table 6 below). It was once in 61707, after the merge (see example above) in 61756 and belongs to municipality 61710 from 1.1.2016.

BUILDING-CODE	MUNIC. CODE	ENUM DISTRICT -ID	FROM	TO
1301211	61707	61707000	1.1.1900	1.1.2015
1301211	61756	61756002	1.1.2015	1.1.2016
1301211	61710	61710000	1.1.2016	31.12.2999

**Table 5: Illustration of historicised building information in the regional database**

The enumeration district for these four buildings changed too. In fact all regional entities concerned by this change have to be updated, which is done either by the look-up tables or by a point in polygon analysis.

With this system it will be possible to get a log file with all changes of certain regional breakdowns. When fully implemented some questions like these can be answered in a quick way

- Get a list of all municipalities in 2012
- Get all changed municipalities since the census 31.10.2011
- List all changes of a specific municipality starting from a specific timestamp
- List the NUTS3 classification (based on municipalities) from 2011
- ...

In a final phase this system will be accessible via web services. So other registers in Austria (e.g. Business Register) will be able to retrieve the needed information over a standardised interface.

## 4.2 Address point acquisition in Statistics Poland

Address points in the Territorial Identification Registry (TERYT) are described by a set of unique identifiers. Based on that, statistical address databases created in the census preparatory phase are maintained and quarterly updated. Locations of new address points are determined by GIS operators with use of the ortophotomap and descriptive information from municipalities (e.g. number of cadastral parcel the new building is on).

Address points related to buildings and dwellings are described as follows: XXXXXX X + YYYYYY Y + RRRRRR O + UUUUU + B, where:

- XXXXXX X – administrative unit ID (voivodship + county + municipality),
- YYYYYY Y – locality ID,
- RRRRRR O – statistical region ID (6 digits) and census enumeration area ID (last digit),
- UUUUU – street ID,
- B – address number.

If the address register of official statistics is kept with respect to unique identifiers of all statistical and administrative units, external sources such as registers or other spatial data can be used to ensure quality of data, provided they use the same identifiers.

Development of the statistical address point database allowed the Central Statistical Office (CSO) to assess the quality of reference materials acquired from the Mapping Agency and other authorities. The main focus was on data coverage (% of the country area) and availability of descriptive data that would allow determining geometries for addresses. Address points as a layer were available for approx. 10 percent of the country from the Topographic Objects Database. Cadastral data in numeric

form was available for approx. 60 percent of the country but not all datasets possessed descriptive data of a quality that could support address point creation. Address data extracted from the Topographic Database and cadastral data were compiled into one dataset for each county with the following set of attributes: name of locality, street name, address number. Those datasets were then distributed to regional units, which assigned appropriate identifiers (locality, street) and then imported the address point to the statistical address database for the county.

Address points that could not be extracted automatically from the reference materials were created manually by geoinformation staff of CSO. In the first step the address points imported from external sources were corrected and complemented with the missing dwelling locations. This is where situation sketches have proven to be a worthy source of address locations. All sketches have been updated in the field by enumerators during the Population and Housing Census 2002. After that all new dwellings were drawn on sketches based on information submitted by municipalities and those that could be gathered by workers responsible for maintaining the register of Poland's territorial division (TERYT). Information on each new building with dwellings needs to be submitted by municipalities to the regional statistical offices in order to register it in the Dwelling Register (NOBC). However, there is no obligation to submit information on geographic location of the building. Bearing that in mind and the fact that the last field check took place in 2002, the situation sketches were not a 100 percent correct source of information. Still, at the time of creating the spatial address database, it was the only address location data source covering the whole of the country.

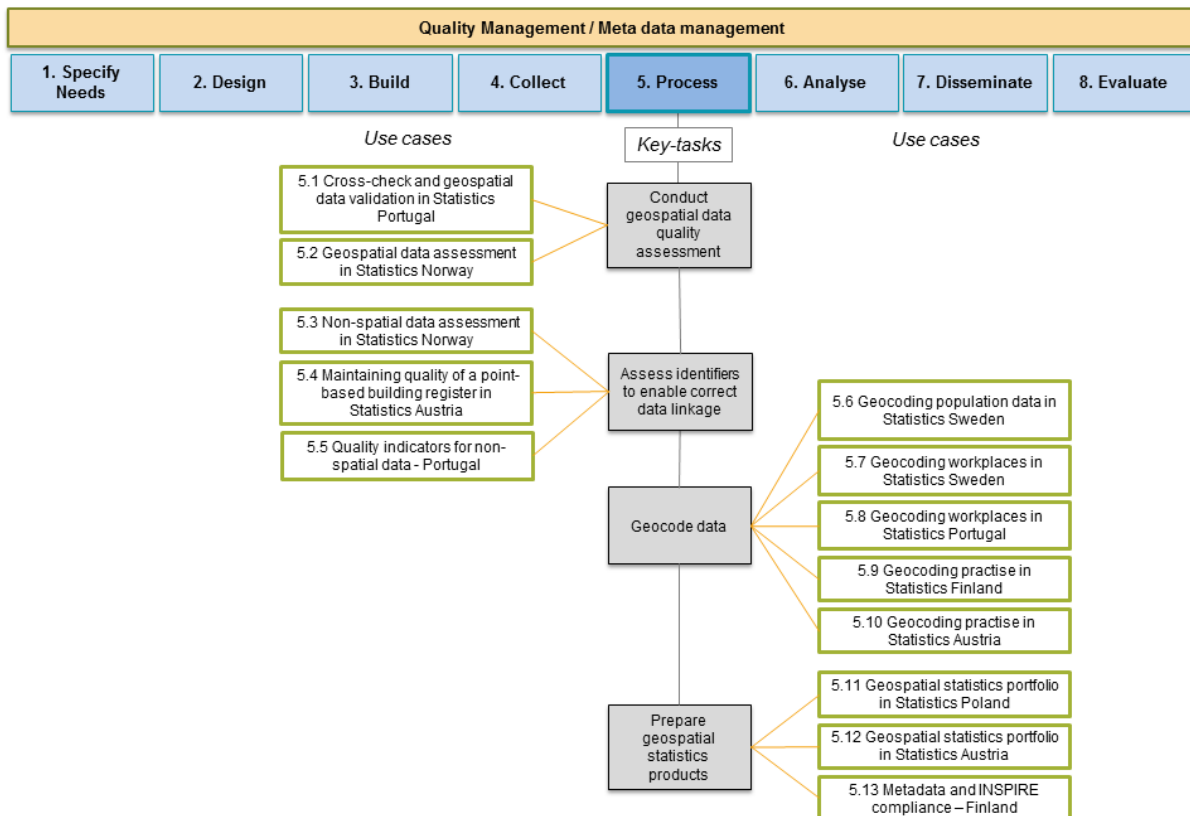
The first datasets with extracted address points were delivered to regional offices in January 2010. While the spatial address database for the first counties was being created in the regions, CSO was successively extracting address points for other counties. 210 GIS operators in CSO have put an enormous effort to create a complete set of address points representing dwelling locations for the whole country. On June 30th 2010 the dataset was ready to be used in the Agricultural Census 2010. 5,7 million address points were acquired in half a year. The spatial address database is to date the most complete and accurate source of information on dwelling locations in Polish public administration.<sup>8</sup>

The statistical address point database was used in both nationwide censuses and its quality has been improved by municipality authorities, which were obliged to verify and correct address point locations. The locations were also verified and corrected in the field by enumerators in a pre-enumeration phase of the census. Enumerators used hand-held devices with GPS, which allowed them to verify and correct address point locations on the spot. Their work was supervised by dispatchers, who used desktop GIS tools to assign tasks, monitor enumerators' daily path and work progress. All address points were verified in the field during that stage of the census.

---

<sup>8</sup> Mirosław Tadeusz Migacz "Geostatistics Portal – a platform for statistical data geovisualization", Statistical Journal of the IAOS 31 (2015) 463–470, DOI 10.3233/SJI-150920, IOS Press 2015, paper written by Mirosław Tadeusz Migacz CSO Chief GIS Specialist.

## 5. Process



### 5.1 Cross-check and geospatial data validation in Statistics Portugal

Within SDI management, Statistics Portugal has a set of quality control routines for its own data during the editing process. Those routines are mostly GIS based processes developed in-house that allow identification of topological and attribute errors for the:

- Enumeration Areas - blocks with a three level structure – sections, subsections and localities – integrated with the official administrative boundaries;
- the Road Segments Network – street line coverage at national and local level edited with geometric and alphanumeric data from the municipalities and used for the delineation of the Geographic Information Referencing Base;
- the Buildings Geographic Database, Geographic Information Referencing Base and Road Segments datasets.

Additionally for the data provided by the municipalities, X,Y coordinates of the Indicators System of Urban Operations (SIUO), dynamically maintained by the Municipalities, concerning the building and dwelling permits and completed construction work, there is also a spatial cross check routine to evaluate the location data quality and accuracy prior to the insertion of the building point in the Buildings Geographical Database. The addresses of these permits are also harmonized and cross checked with existing addresses in the National Dwelling Register and validated before they are loaded into the National Dwelling Register.

The SDI is also used by Statistics Portugal to manage the INSPIRE national thematic datasets under its responsibility. Statistics Portugal is a full member to the Advisory Council for INSPIRE which is the strategic body responsible for the implementation of the Directive in Portugal.

At a technical level, Statistics Portugal is strongly engaged in the Cross Thematic Working Group, mainly focused on metadata and services, as well as in five out of nine Thematic Working Groups which were created similarly to the INSPIRE Thematic Clusters (Annex 1: List of the INSPIRE Thematic Clusters and related Data Themes). Due to the assistance of this thematic working group, Statistics Portugal has developed important skills concerning validating GML INSPIRE harmonized data according to the directive data specifications.

## **5.2 Geospatial data assessment in Statistics Norway**

All national geospatial data is documented and should follow national standards (geonorge.no). All geospatial data which is produced by Statistics Norway and delivered as external products is also documented accordingly on the national geospatial portal (run by the National Mapping Authority/ National inspire geo coordinator/ Norway digital secretariat). However, internal ready processed geospatial data which forms the basis for statistics shall according to archive laws and regulations be long term stored and documented. For all non-geospatial data forming the basis for statistics, there is a central system in Statistics Norway for long term storage and documentation. Regarding point-based geospatial data, this can be handled just as all other non-geospatial data (but not area-based data, which will be subject to further assessments).

Before geospatial data arrives in Statistics Norway from the NMCA, it has already been subject to quality checks by the NMCA and the data has been put together to national data bases from municipal data bases. However, Statistics Norway is building routines for automatic checking and uploading to the central geodatabase. One key aspect in these routines is automatic procedures for comparing against last year's data. If there are big differences, they are recorded to an output list for further inspection and quality assessments.

Regular meetings between the NMCA and Statistics Norway are conducted on issues related to the address and building register. These meetings concern all aspects of quality, not only the geospatial quality. Issues are discussed and the NMCA discusses and informs about the issues to the municipalities, which are actually responsible for maintaining the source data. A generic observation and experience from the cooperation in Norway; it is better to establish routines and checks when data is being put into the register, rather than trying to fix errors afterwards.

## **5.3 Non-spatial data assessment in Statistics Norway**

There are a number of generic checks (non-geospatial) put in force in the statistical version of the Cadastre (addresses, buildings), but routines specific to statistical products are also applied. One simple test is checking whether the address or building point is within the right municipality or county (or country!). Concerning building area, figures from the building (polygon) map have been used for imputation.

There is a specific division responsible for statistical population registers in Statistics Norway which maintains statistical versions of registers covering population (residents), addresses, buildings and enterprises and monitors the quality therein, including accurate linkages between unit record data and geospatial data.

## 5.4 Maintaining the quality of a point-based Buildings and Dwellings Register in Statistics Austria

As the Buildings & Dwellings Register (BDR) and its address and building geocodes respectively form the basis for further point based statistics the quality of the register is of outmost importance. The following aspects are particularly crucial and need special attendance when linking other registers and data:

- Address-ID and building-ID should be unique and unambiguous.
- Address-ID and building-ID should be “active” for the required time-stamp
- Address-ID and building-ID should correspond; it should be clear which parcel address a building address lies on.
- Address-ID and building-ID should have valid geocodes

To safeguard correct linkage between geospatial data and registers usually the building-ID is the common field. In the case where the address information is only available on parcel level, the parcel address and its geocode is used. So as long as data is provided with the correct and valid building-ID or at least address-ID the linkage with the geocodes should not be a problem.

Due to possible changes in administrative boundaries between the timestamp of the data to be linked and the timestamp of the Buildings and Dwellings register the dependency between building-IDs and municipality codes in the data has to be checked and possibly updated.

Experience also shows that sometimes records cannot be linked due to mistakes in the Buildings and Dwellings register or due to inactive or out-dated building-IDs in some registers. There is no definite procedure to overcome this, but miss-matches between statistical registers/records and geospatial address data/building data are checked both by the staff responsible for the statistics as well as the geoinformation staff. If it turns out to be a mistake in the Buildings and Dwellings Register it is reported to the Buildings and Dwellings Register hotline, which then contact the municipality and if possible, mistakes are corrected right away. If the mistake is found to be due to inactive or out-dated building-IDs in the data to be linked currently the solution is to accept inactive or out-dated building-IDs and use their geocodes or that of neighbouring addresses/buildings for further analysis.

If records are missing a correct building-ID they cannot be linked with the Geodatabase and hence have no geocode to be used for point-based statistics. This is the reason why statistical aggregates from points (grids and point based analysis) do not amount to 100 percent of the population. Being now aware of the problems this causes to the world of spatial analysis the departments try to reduce the number of missing building-IDs by looking for the addresses also in the building tables for other dates of validity.

By using the geocodes for spatial analysis or the aggregation of grids and other derived products, internally or by customers, some mistakes in the geocodes have been discovered in the past. They were mainly due to historic entries into the system, which were still done by hand (miss typing), so before the Geoclient (mentioned in Case 5.5 *Geocoding practise in Statistics Austria*) was installed. These were reported to the Buildings and Dwellings Register hotline, which then contacted the municipality and the mistakes were corrected.

Examples:

- Geocodes from the Buildings and Dwellings Register are also used for traffic navigation by emergency suppliers (ambulance, fire brigade...), so it is a living register. Errors can be fatal and get reported, so the quality gets constantly improved.
- When calculating the commuting distance based on the geocodes from the Buildings and Dwellings Register the result for the commuting distance for all pupils of a certain primary school in Vienna was more than 15km. This was clearly implausible and it was soon clear that the geocode of that school was wrong! The geocode was corrected in the Buildings and Dwellings Register and the routes recalculated.
- Built-up grid cells in the middle of a lake (no, there is no island there) made incorrect building geocodes visible. The problem was: some buildings at the shore of the lake only had a geocode for the parcel address and this was placed in the centre of the parcel (piece of "land"), which reached far into the lake. Again these building geocodes were corrected.

## 5.5 Quality indicators for non-spatial data

The lack of a common identifier link between different sources requires analysis and comparison of non-harmonized addresses, which constitutes a constraint factor to the update process of the National Dwelling Register (FNA) in Statistics Portugal.

In this context an analysis system was created in Statistics Portugal Data Warehouse, where a set of indicators for buildings and households has been defined in order to measure and monitor the quality of the National Dwelling Register and some of the attributes, and their completeness, of the family surveys samples. Several indicators have been developed, however this case focusses on the **Address Quality Degree Indicator (AQD)**. The indicator is explained in detail below.

### Indicator "Address Quality Degree" (AQD)

The indicator was created to assess the syntactic quality of an address. The AQD variable is based on the features described below, in which the following concepts were adopted:

- Mono-household building - Building with a single household;
- Multi- household building - Building with two or more households.

The AQD includes the following features:

- Structure of syntactic particles of address;
  - the absence of street name (R);
  - the existence of street name but the absence of the building door, number or name (name = pair of attributes: building type prefix and building name) locator (P);
  - the lack of households or the existence of duplicated households in multi-households buildings (A);
  - Non unique Address, till the buildings door number or name descriptor, between household and building in mono-household buildings (U);
- Absence of the designated family representative;
- The geographical area of the household location:
  - The 2011 Census Place. Characteristics of three different types of geography were studied: the Parish (2014 TIPAU), the 1Km<sup>2</sup> GRID (household density); the 2011

Census place (buildings). Given the more detailed level of spatial analysis that it allows, it was decided to feature in the AQD the 2011 Census Place.

The AQD variable is intended for use in the following situations:

- Monitor and measure the update/quality status of the households addresses for mailing purposes;
- Possible use, in the sample selection process, in order to prevent the selection of households whose addresses do not have a minimum quality that will allow addressing the circular (postal notification);
- Contact the municipalities whose geographic places have a large number of buildings with no street names in order to validate this information and to try to raise awareness to this reality.

### Type of 2011 Census Place

Statistics Portugal has proceeded to create a classification of 2011 Census Places regarding the existence of street names in the addresses. The 2011 Census Places typology has been defined on the basis of the following actions:

- Identify, for each building, its 2011 Census Place;
- Count the total number of buildings of each 2011 Census Place (N);
- Count, for each 2011 Census Place, the number of buildings with street names in the address (M);
- Calculate the percentage of buildings with street names in the address (PE) regarding the total number of buildings of the 2011 Census Place ( $PE = M * 100/N$ );
- Set a percentage scale with intervals of 5% of PE (PE);
- Represent, in a table, the 2011 Census Places on the PE scale;
- Identify, on a map, the 2011 Census Places on the PE scale;
- Define a threshold on the PE scale from which it is considered that the existence of addresses without street names is a real characteristic of the 2011 Census Place and not the result of poor quality addresses. In other words we intent to know if the lack of street name in the address of a building results from the poor quality of its address or if it's a real characteristic of the 2011 Census Place were the building is located ("Insufficient Address - **Real**")

The results of this indicator application are presented in the table below, which represents the number and percentage of 2011 Census Places at the PE Scale:

PE Scale	Nº of 2011 Census Places	% of 2011 Census Places	Accumulated % of 2011 Census Places
A-0%	3083	11,58%	11,58%
B-0% a <5%	446	1,67%	13,25%
C-5% a <10%	376	1,41%	14,67%
D-10% a <15%	208	0,78%	15,45%
E-15% a <20%	203	0,76%	16,21%
F-20% a <25%	149	0,56%	16,77%
G-25% a <30%	173	0,65%	17,42%
H-30% a <35%	137	0,51%	17,93%
I-35% a <40%	126	0,47%	18,41%

PE Scale	Nº of 2011 Census Places	% of 2011 Census Places	Accumulated % of 2011 Census Places
J-40% a <45%	135	0,51%	18,91%
K-45% a <50%	100	0,38%	19,29%
L-50% a <55%	141	0,53%	19,82%
M-55% a <60%	123	0,46%	20,28%
N-60% a <65%	163	0,61%	20,89%
O-65% a <70%	135	0,51%	21,40%
P-70% a <75%	173	0,65%	22,05%
Q-75% a <80%	186	0,70%	22,75%
R-80% a <85%	217	0,81%	23,56%
S-85% a <90%	326	1,22%	24,79%
T-90% a <95%	593	2,23%	27,01%
U-95% a <100%	2699	10,14%	37,15%
V-100%	16736	62,85%	100,00%
<b>Total</b>	<b>26628</b>	<b>100,00%</b>	

**Table 6: 2011 Census Places distribution at the PE Scale**

Definition of a two value typology, with a threshold at 75 percent, to classify the 2011 Census Places regarding the existence or absence of street names:

- **T** – 2011 Census Places with street names: 2011 Census Places with PE > 75% - it means that more than 75% of the 2011 Census Place buildings have street names and it is acceptable to assume that addresses of buildings without street names are the result of errors that must be corrected.
- **S** – 2011 Census Places with no street names: 2011 Census Places with PE or <= 75% PE – it means that 75% or less of the 2011 Census Place buildings have no street name and it is acceptable to assume that buildings addresses without street names are considered as real, as a characteristic of the 2011 Census Place itself, and not as the result of an error at the buildings address.

#### Household's Addresses Classification Algorithm (Algorithm and AQD categories)

For the AQD categories the following algorithm was set:

- **(R) "Address without street name"** – Address just with Place or Locality attributes filled (Note: Validation Rule – If Place is filled the Locality must also be filled - there are no address registers with the Place filled and without the Locality filled):
  - **T** (2011 Census Place with street names) -> **C.R** (Insufficient address - without street name - Correctable)
  - **S** (2011 Census Place without street names) without the designated family representative -> **R.R.** (Insufficient address – Real - without Street name)
  - **S** (2011 Census Place without street names) with the designated family representative -> **S.R.** (Sufficient address without street name)

- **(P) "Address with street name and without building door number/name"** – Address with street name filled but without building door number or name (pair of attributes: building type prefix and building name):
  - **T** (2011 Census Place with street names) -> **C. P.** (Insufficient address – with street name but without building door number or name - Correctable)
  - **S** (2011 Census Place without street names) unnamed family representative -> **R. P.** (Insufficient address – Real - without building door number/name and family representative name)
  - **S** (2011 Census Place without street names) with the name of the family representative -> **S.P.** (Sufficient Address - Real - without building door number/name but with the family representative name)
- **(U) Building with 1 single household but with different addresses between building and household:**
  - **C. U.** (Insufficient Address - Correctable - without unique address between building and household)
- **(A) Building with more than one household without indication of floor number and/or side (right, left, etc.) or with inconsistent floor or side information** (duplicated floor and side, etc.):
  - **C.A.** (Insufficient Address - Correctable - without floor number)
- **(B) Remaining addresses, good:**
  - **B** (good address)

Notes:

- Without family representative name – Field "Name" = "FAMILY REPRESENTATIVE"
- With family representative name – Field "Name" <> "FAMILY REPRESENTATIVE"
- To aid the correction of all the potentially correctable Addresses it have been created specific excel files
- In the reports documents it only be presented the classification at one level. For example:
  - If IR.R it will only be reported the IS.

**AQD Households Addresses Classification:**

B – Good -> Selectable and Addressable

S – Sufficient -> Selectable and Addressable (despite not all of the address attributes - locators - are filled)

**S.R.** (Sufficient address without street name)

**S.P.** (Sufficient Address - Real - without building door number/name but with the family representative name)

C – Insufficient correctable -> non selectable and not addressable (Address with errors that must be corrected in order to be selectable and addressable)

**C.R.** (Insufficient address - without street name - Correctable)

**C. U.** (Insufficient Address - Correctable - without unique address between building and household)

**C.A.** (Insufficient Address - Correctable - without floor number)

**C. P.** (Insufficient address – with street name but without building door number or name - Correctable)

R – Insufficient real -> selectable and non-addressable

**R.R.** (Insufficient address – Real - without Street name)

**R. P.** (Insufficient address – Real - without building door number/name and family representative name)

## 5.6 Geocoding population data in Statistics Sweden

Due to the large number of references to statistical and administrative geographies in the Geography database (GDB), standard aggregation of statistics to various geographies is possible without using spatial operations (for further information about the Geography database, see case 3.1 *The Geography Database - production set-up for point-based geocoding in Statistics Sweden*). Even aggregation of data to statistical grids can be based on regular SQL statements. This case illustrates a standard production chain for gridded population data comprising geocoding of micro data and aggregation of statistics. The result is a grid map covering the entire country with statistics on population by age. The grid map will have a 250 x250m grids in urban areas (localities) and 1x1km grid in rural areas.

In this case, two data sources are used:

- 1) The Population Register which contains information about all inhabitants in Sweden as of December 31 each year. Each individual comprises one record in the register. Central to this operation is the unique identifier, in this case the name of the real estate unit (Cadastral parcel) and the municipality code. Other unique identifiers possible to use are physical address and dwelling-ID.
- 2) The GDB, in this example with cadastral parcel coordinates along with the Real Estate identifier found in the Population Register. Each object in the GDB is aware of its location with regards to localities, NUTS areas, statistical grids etc.

Step one is to link grid references to each unit record in the Population Register and, in the same step, to aggregate population into age-groups and to 250 meter grids in urban areas and 1km grids in rural areas. This operation is conducted by a script in Microsoft SQL, which is the environment where all registers at Statistics Sweden is stored. The grid reference is created by means of

truncation of X and Y coordinate values to retrieve the coordinates of the upper left corner of the grid cell in which the spatial objects are situated.

Population register						
Municipality	Real Estate Unit	Personal id	Sex	Land of origin	Age	AgeID
011401	ANTUNA 2:1	192613015070	1	SWEDEN	16	30772
011804	HEDEN 2:1	193203356002	2	SWEDEN	78	28782
188001	MELONEN 4	193204001029	2	SWEDEN	32	28735
188001	MELONEN 6	193718041555	2	SWEDEN	22	27024
051301	KAJAN 7	196508568733	1	SWEDEN	45	16562
051401	SKEDVI 2:11	194607746453	1	SWEDEN	64	23649
051401	SKEDVI 2:11	195013299361	2	SPAIN	43	21988
012201	ANTUNA 2:1	194805126695	1	SWEDEN	8	22939

Geography Database						
Municipality	Real Estate Unit	Xcoord	Ycoord	Locality code	NUTS3	ETRS89LAEA
011401	ANTUNA 2:1	6918811	619720	7688	SE232	1kmN3867E4509
228301	TEGELBRUKET 6	7005877	616941	7660	SE232	1kmN3862E4494
106002	HOLJE 2:246	6236100	471816	2720	SE232	1kmN3854E4573
011804	HEDEN 2:1	6394488	698073	0000	SE232	1kmN3854E4573
051401	SKEDVI 2:11	6259331	444381	2956	SE232	1kmN3884E4586
252301	HAKKAS 35:6	7436106	787874	0000	SE232	1kmN3884E4586
012006	BJÖRNÖ 1:481	6570581	700694	0106	SE232	1kmN3887E4576

**Figure 10:** The figure illustrates the content of the Geography Database (GDB) and the Population Register and how the linkage between location data and unit record data is conducted.

As shown in Figure 10 above, two individuals are registered at the real estate unit “SKEDEVI 2:1” in municipality 051401. By using the municipality code + real estate unit as a unique identifier, both individuals can be linked to the corresponding point-location in the GDB. In Figure 11 below the SQL statement used to link and aggregate data in the same step is shown.

```
SELECT
GridSize= CASE WHEN locality = '0000' THEN 1 ELSE 2 END,
GridCode = CASE WHEN locality = '0000'
THEN CONVERT(char(6), (Ycoord/1000)*1000) + CONVERT(char(7), (Xcoord/1000)*1000)
ELSE CONVERT(char(6), (Ycoord/250)*250) + CONVERT(char(7), (Xcoord/250)*250) END,
Pop0_19 = SUM(CASE THEN age between 0 and 19 THEN 1 ELSE 0 END),
Pop20_39 = SUM (CASE THEN age between 20 and 39 THEN 1 ELSE 0 END),
Pop40_64 = SUM (CASE THEN age between 40 and 64 THEN 1 ELSE 0 END),
Pop65_w = SUM (CASE THEN age >= 65 THEN 1 ELSE 0 END),
TotPop = COUNT(*)
INTO GridStat
FROM [...].rtb2015..rtb a JOIN [...].GDB2016..RealEstateCoordinates b
ON a.Municipality = b. Municipality and a.RealEstateUnit = b.RealEstateUnit
GROUP BY
CASE WHEN locality = '0000' THEN 1 ELSE 2 END,
CASE WHEN locality = '0000'
THEN CONVERT(char(6), (Ycoord/1000)*1000) +
CONVERT (char(7), (Xcoord/1000)*1000)
ELSE CONVERT(char(6), (Ycoord/250)*250) + CONVERT(char(7), (Xcoord/250)*250)
END
```

**Figure 11:** SQL statement used to geocode and aggregate population data in the same step

Table 7 below shows the result of the combined geocoding and aggregation operation. Disclosure control has been conducted, in this case for cells with figures below 3 which is the standard threshold. The column **GridSize** identifies whether the grid is located in a rural area (1) or in an urban area (2). **GridCode** is the grid identifier built composed by the coordinates of the lower left corner of

the grid combined. The columns **Pop0\_19...** contain aggregated unit record data, in this case population by age groups and total population.

GridSize	GridCode	Pop0_19	Pop20_39	Pop40_64	Pop64_w	TotPop
1	3170006585000	3	0	3	0	6
1	3170006599000	0	3	6	3	12
1	3170006600000	5	9	4	4	22
1	3180006588000	0	0	4	3	7
1	3180006589000	0	0	3	3	6
2	3180006600000	27	17	30	12	86
2	3180006601500	4	6	6	3	19
2	3187506602000	0	3	3	3	9
2	3485006641250	45	24	32	18	119
2	3490006641250	19	51	51	92	213

**Table 7: Table showing the result from the combined geocoding and aggregation operation.**

The result is typically delivered along with a corresponding grid net GIS database, but it could also be disseminated as a stand-alone table, as some users already have a grid net at disposal and wishes only to update the population data.

## 5.7 Geocoding workplaces in Statistics Sweden

Geocoding of workplaces (business premises), schools or other public facilities in Statistics Sweden is conducted on the basis of authoritative address records comprising the physical address location. Physical address data are obtained from the NMCA and stored as a full address table in the Geography database (GDB) which is the main environment for geocoding in Statistics Sweden (See Use Case 3.1 *The Geography Database - production set-up for point-based geocoding in Statistics Sweden*).

As opposed to geocoding of population data, described in the previous Use Case (5.1 *Geocoding population data in Statistics Sweden*), geocoding workplaces is a more demanding process. Whereas people need to register by a full physical address validated against an address index, companies are not subjected to the same rigorous point-of-entry address validation processes. As a consequence, there is a certain degree of mismatch between the address information reported by companies to the business register and the authoritative address location data. Possible causes of mismatch are;

- The address reference in unit record data is incomplete, as of missing street name or missing street number
- The street number in unit record data is not valid and the closest valid number in the authoritative address table is too high or too low to accurately locate the workplace
- Information regarding city or municipality is incorrect in unit record data
- Street name is misspelled in unit record data
- The address in unit record data is correct but new and has not yet arrived in the GDB due to lagging of data transfer between Statistics Sweden and the NMCA

To undertake high precision address geocoding of a workplace the following information needs to be present in the unit record data:

**[Code of municipality + name of the city/postal area + street name + street number]**

Prior to linking unit record data with location data, the address information needs to be examined and undergo a number of correction procedures. The first step is to standardise and “repair” the reported address in unit record data. For example, in case the reported address of the unit record to be geocoded is “Queen str. 14 LGH1202”, the result after standardise and repair, will be “Queen street” (name) and “14” (number) stored in separate columns. The last part of the original address string, “LGH1202”, is a reference to the dwelling number, used mainly to assign population to a specific dwelling for census purposes. This information is redundant when geocoding workplaces and can be dropped.

The second step is to correct misspelled addresses using so-called alias tables. A typical example is “2:nd Cross Road” which is an incorrect textual representation of the address. The address will be translated to the correct spelling, in this case “Second Cross Road”.

Linking of unit record data with the physical address record in GDB holding the coordinates is conducted entirely within a MS SQL server environment. Matching is an iterative procedure conducted in the following steps:

1. Direct match on correct original address
2. Match on original address against table of popular name in GDB. Popular names comprise alternative references to a place, e.g. “Old farm” instead of street name and number.
3. Match on standardised address
4. Match on original and standardised address against previous year’s address record in GDB
5. Match on original address without using information about city. This can only be conducted if street address is unique.
6. Match on closest valid street number (if the street number reported in unit record data is invalid). The closest valid street number can be both odd and even.

Each geocoded object will be tagged with a code (1-6) declaring how the coordinates were retrieved. This is important as to know the level of spatial accuracy to be expected.

## 5.8 Geocoding workplaces in Statistics Portugal

The National Dwellings Register is the master file to which any address-based system must be matched for the purpose of geocoding existing databases at Statistics Portugal, specifically through its related *Buildings Address Database* subset, which is the key for pinpointing the Business Register records.

The methodology for building the Geographic Database of the Business Register (BRGD) considers the address as the key element to directly or indirectly match the records with the existing Buildings Geographical Database, by following a step-by-step approach based on different locators capable of sequentially pinpoint the Business Register records.



**Figure 12: Illustration of the step-by-step approach to sequentially pinpoint the Business Register records by means of a set of locators progressively less precise in terms of spatial accuracy from Locator #1 to Locator n.**

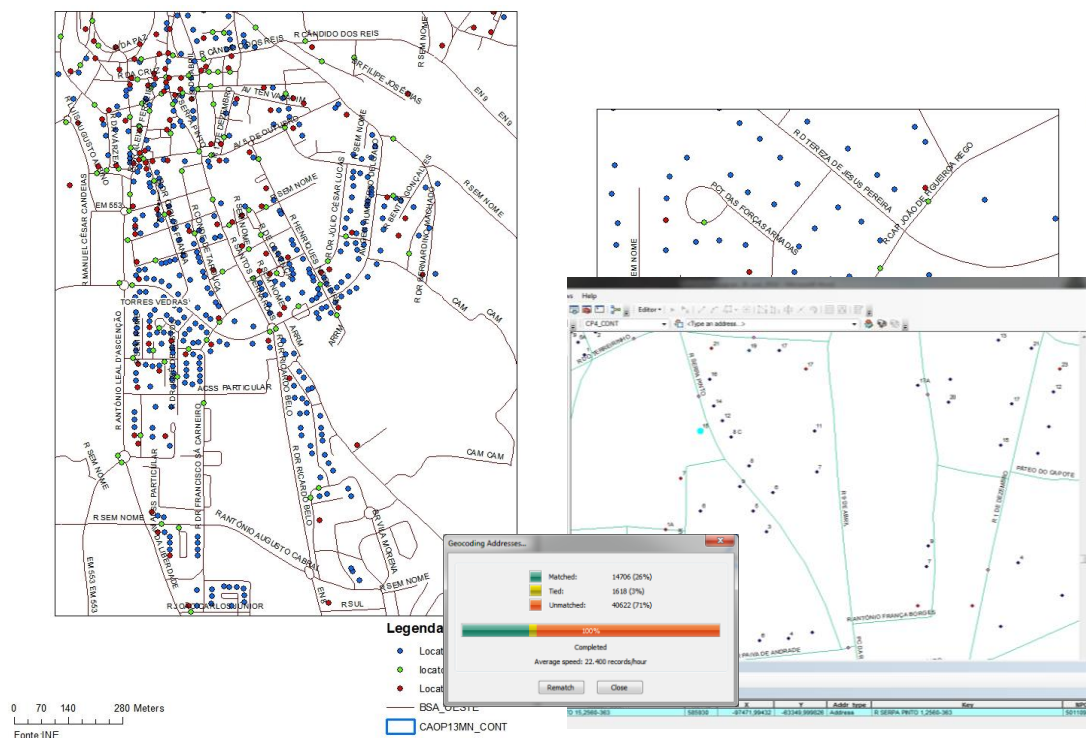
The generic methodological approach relies on performing address-matching routines for the Business Register directly over the Buildings Geographical Database, which is synchronised with the National Dwellings Register, being the locators progressively less precise in terms spatial accuracy from Locator #1 to Locator n.

Due to specific aspects revealed by the analysis of the Business Register, some adjustments are considered, namely the use of locators retrieved from the Road Segments Database, which is indirectly related with the Buildings Geographical Database through spatial attributes, or the edition of specific fields in order to improve the results obtained by a given locator.

The locators used to pinpoint the Business Register records are described in table 9 below.

Locator	Description
<b>Locator #1</b>	Complete address composed by type of road, name, number, 7 digit postal code
<b>Locator #2</b>	Used over the Road Segments Database in order to overcome discrepancies in the address
<b>Locator #...</b>	Uses the locality name and the 4 digit postal code
<b>Locator n-1</b>	Based on the 7 digit postal code, which is a linear structure used to code each block façade composed by the CP4 and 3 additional digits
<b>Locator n</b>	Based on the 4 digit postal code, which is a polygonal structure used to code each postal distribution area

**Table 9: List of locators sequentially used to pinpoint the Business Register records**



**Figure 13: A different mix of locators has been used for the cases processed.**

## 5.9 Geocoding practise in Statistics Finland

At Statistics Finland, location information of statistical single units comes initially from registers. Location information can be linked by using the same IDs of single units in separate statistical systems.

Statistics Finland makes quality checks but usually not corrections. If the quality of register-based location information is proved to be poor, the location information of that particular single unit is suppressed.

Needs for geocoding varies. The main cases are:

- identifying the right building for an address (point) e.g. for a business premise
- identifying the correct part of a street for a point e.g. road traffic accidents
- reverse geocoding, identifying the correct address for a building (point)

Known location information, as well as generation of location information from auxiliary data may be part of the statistical production system. In this case, the aim is to automate the production phases. A service-based geocoding system for addresses has been created. It is based on location information of buildings but in case of missing coordinates, location information is conducted from street data. The system was created by using OpenLayers, SAS, SQL and service-based interfaces. An idea is to have the system widely in use in different statistical production procedures. This kind of service may later represent a GIS related CSPA-service.

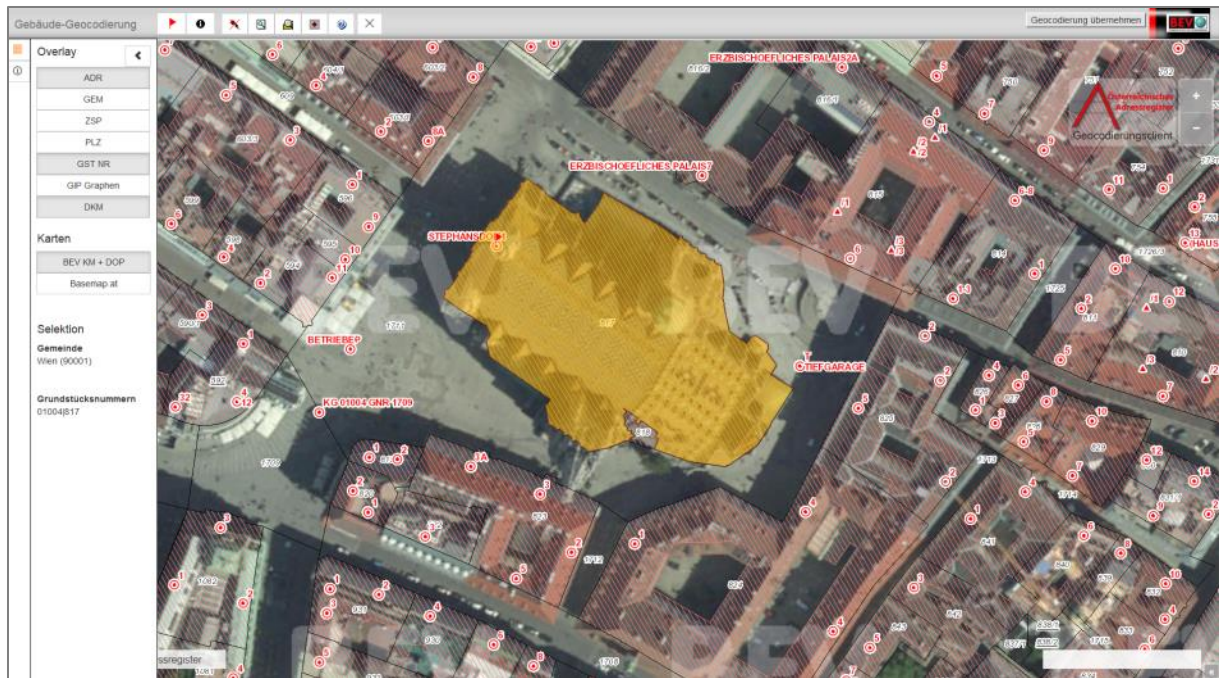
Road traffic accidents renewed their statistical system in 2015 to 2016. In addition of a transition to use OS GIS tools in GIS related production phases, the geocoding of road traffic accidents were made more integrated with the statistical production system. SQL Server spatial data type is utilised for geocoding, as well as for further descriptive data search by location.

## 5.10 Geocoding practise in Statistics Austria

The Buildings & Dwellings Register demands the geocodes of each building to be entered into the system. The tool to do this is the so called “Geoclient”, which is included in the application of the Online Address- Buildings and Dwellings Register. On click it opens up a separate window showing an aerial map, parcel boundaries and its building areas, zoomed in to the selected parcel address or building and its surroundings. The toolbar (flag symbol) provides the function to set the geocodes of the selected building. As a rule the ideal position for the geocode is within the building polygon and near the entrance.

When clicking the save-button (“Geocodierung übernehmen”) a check is run in the background that verifies that the geocode is set within the selected building area.

The geocodes are then saved in the local reference system of the municipality, which is one of the three Gauss Krüger stripes (EPSG 31254, 31255 and 31256).



**Figure 14: Screenshot of the GeoClient with Saint Stephan's Cathedral in the centre and a flag for the representation of its coordinate point (its geocode).**

Through this system there is a geocode for each building (in fact for each building part corresponding to each entrance). Other datasets containing the building-ID hence can be linked via the building-ID and get the geocode from the Buildings and Dwellings Register. These datasets can therefore be considered to be point-based too.

Pop-ID	Building-ID	Male	Female	Building-ID	x-coord (3035)	y-coord (3035)
Pop1	1234567	1	0	1234567	4426955	2681683
Pop2	1234567	1	0	1234568	4427245	2681356
Pop3	1234567	0	1	1234569	4427552	2690218
Pop4	1234567	0	1			
Pop5	1234567	1	0			
Pop6	1234568	1	0			
Pop7	1234568	0	1			
Pop8	1234568	1	0			
Pop9	1234568	0	1			
Pop10	1234569	0	1			
Pop11	1234569	1	0			
Pop12	1234569	1	0			
Pop13	1234569	0	1			
Pop14	1234569	0	1			
Pop15	1234569	1	0			
Pop16	1234569	1	0			

Building-ID	Total Population	Male	Female	x-coord (3035)	y-coord (3035)
1234567	5	3	2	4426955	2681683
1234568	4	2	2	4427245	2681356
1234569	7	4	3	4427552	2690218

**Figure 15: Linking population to building by means of building-ID**

## 5.11 Geospatial statistics portfolio in Statistics Poland

The introduction of x, y coordinates and address points in statistical data collection enabled significant change of the previous spatial identification system and resulted in shifting from area assignment (statistical and administrative units) to point assignment. It had a fundamental consequence for the application of geoinformation in the Central Statistical Office (CSO). The change of the assignment mode allowed for a more flexible aggregation of data collected in CSO even for the smallest areas. It also facilitated the creation of a spatially-oriented micro database, enabling the conduction of geo-statistical analyses. Building a point-based foundation of statistics was included in the preparation phase for 2010/2011 census round, when it quickly became apparent that address point reference (x, y coordinates) is needed in all stages of the census. Collection of data with reference to x, y coordinates also allows producing spatial statistics, which are independent from boundary changes. Changes in the territorial division of the country usually influence changes in geometries of statistical geographies and census enumeration areas. This facilitates efficient analysis of time series, regardless of the changes place in statistical and administrative geographies. An additional advantage is the possibility of data aggregation to territorial geographies as well as grids.

In order to publish census results CSO prepared the Geostatistics Portal (<http://geo.stat.gov.pl>). The portal provides tools for creating all sorts of thematic maps.

Statistical data available for the portal are absolute values – these can be directly visualised on various kinds of diagram maps. Users also have at their disposal a set of normalisation attributes that are used for on-the-fly calculation of relative values (statistical indicators) to be presented on choropleth maps.

For all types of presentations, users can select thematic phenomena from a predefined list to define the aggregation level (territorial unit) of the output as well as the following (if applicable): symbol, colour range, number of classes and classification methods.

Spatial reference of statistical data kept in the portal database goes as low as LAU2 level (municipality). However, for the data about population all users have access to more detailed data. They are able to create spatial queries for an aggregated value of a phenomenon within a user defined polygon as well as perform advanced spatial analysis on micro data. Results of such analysis may then be published directly in the Portal or become part of a publication, provided they meet requirements of statistical data confidentiality.

Apart from thematic maps and spatial analyses in the Geostatistics Portal, users can find spatial data maintained within CSO (such as boundaries of statistical geographies). The portal serves as a publishing platform for spatial data services that the Central Statistical Office is obliged to provide according to the INSPIRE directive. A discovery service has been established as well as view and downloads services for two spatial data themes: statistical units and population distribution (demography). Access to all services (discovery, view, downloads) is public and free of charge.<sup>9</sup>

Currently CSO presents also various population grids. Data presentation in grid cells is accurate and allows easy comparison as all cells have the same size and are stable over time. Moreover grids

---

<sup>9</sup> Mirosław Tadeusz Migacz "Geostatistics Portal – a platform for statistical data geovisualization", Statistical Journal of the IAOS 31 (2015) 463–470, DOI 10.3233/SJI-150920, IOS Press 2015, paper written by Mirosław Tadeusz Migacz CSO Chief GIS Specialist.

integrate easily with other scientific data (e.g. meteorological information) and grid systems can be constructed hierarchically in terms of cell size, thus matching the study area. Grid cells can also be assembled to form areas reflecting a specific purpose and covering the study area (mountain region, water catchment).

### 5.12 Geospatial statistics portfolio in Statistics Austria

Statistics Austria provides a fixed portfolio of geospatial statistics on the basis of grids from smaller to larger grid sizes. 100m is the smallest size for absolute values; further characteristics subject to confidentiality are available from 250m. Grid statistics are available from various data sources (<http://www.statistik.at/reg-datenkatalog/>). As this is a commercial product a licence agreement has to be signed and the data has to be paid for.

Some analysis demand precise results (e.g. only buildings within the flooding zones etc.) which cannot be obtained using grids. Statistics Austria can use non-aggregated point-data for internal production or analyses, but for confidentiality reasons data can only be published in an aggregated way. Therefore Statistics Austria also offers customised analysis, where customers provide the polygons of interest and get the results of a point in polygon analysis in return. Again this service has to get paid for and is subject to confidentiality.<sup>10</sup>

Exceptions as to using non-aggregated data are in the mapping environment, where road accidents are mapped on the precise location.<sup>11</sup> Similar applications are currently being developed on “buildings under construction” and “schools” where the precise point-locations will be used.

### 5.13 Metadata and INSPIRE compliance – Finland

Statistics Finland publishes geospatial data through Geoserver, both as view services (WMS, Web Map Service) and download services (WFS, Web Feature Service). Data are published in the National Geoportal according to INSPIRE specifications for services and INSPIRE themes. At the moment, there are more than 100 datasets provided by Statistics Finland in the National Geoportal and the data are published using the WMS and WFS services that Statistics Finland provides.

Statistics Finland has been asked by customers to open geospatial data also on smaller scale than municipalities. Statistics Finland has embraced Open data policies and after receiving governmental funding for opening data, Statistics Finland could respond to these wishes. Since January 2015, Statistics Finland has provided postal code data and geospatial data as open data through geospatial interface services, both view services (WMS) and download services (WFS).

INSPIRE has been a driving force for implementing new technologies at Statistics Finland, especially open source tools. We’ve adapted open source technologies, for example Geoserver, that can provide data through OGC-standard download services (WFS) and view services (WMS). The database used for open data is also open source technology, PostgreSQL.

Metadata of the geospatial data that is provided through interface services have been published accordant to INSPIRE requirements in the National Metadata Portal for Geospatial Data. The National Metadata Portal for Geospatial Data is linked to the National Geoportal, so Metadata of the

---

<sup>10</sup> For further information, please see:

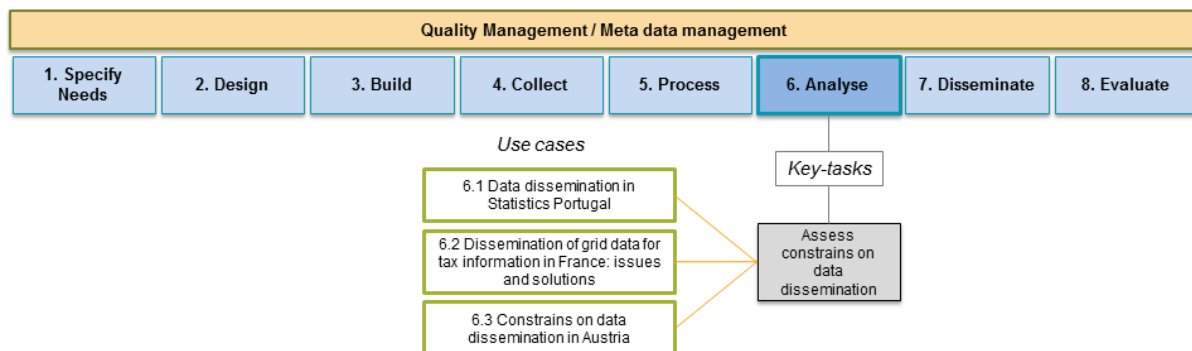
[http://www.statistik.at/web\\_en/publications\\_services/customer\\_defined\\_polygons/index.html](http://www.statistik.at/web_en/publications_services/customer_defined_polygons/index.html)

<sup>11</sup> For further information, please see: <http://www.statistik.at/verkehrsunfallkarte/>

geospatial data can be accessed also that way. Metadata provided in the National Metadata Portal for Geospatial Data, is also linked to the geospatial data provided through WMS and WFS services.

INSPIRE has brought a lot of new prospects to Statistics Finland, but the work to reach INSPIRE compliance in regards of the data and services still continues. Availability of WMS and WFS services provided through the Geoserver is on good level, but the data itself that should be provided in accordance with INSPIRE requirements, are not yet provided as INSPIRE compliant data products.

## 6. Analyse



### 6.1 Data dissemination in Statistics Portugal

Statistics Portugal Dissemination policy<sup>12</sup> follows a clear customer-oriented policy, assigning the greatest importance to meeting customers' needs and expectations. Wide and easy access to Statistics Portugal's information is of priority, as well as improving the quality of the service provided. Transparency, which must underlie the production and dissemination of official statistics, thus requires a detailed clarification of the Revisions policy.

The Dissemination Policy of Statistics Portugal lays down the fundamental principles governing the dissemination of official statistics, directly or indirectly produced under its responsibility.

#### Dissemination service

Concerning geospatial data, on the website, it's possible to:

- browse and download free of charge a significant volume of information (including geospatial data ex. Geographic Information Referencing Base shapefile);
- have access to other electronic services, which includes subscribing contents or sending requests on information not available on the website;
- have access to the Digital Library of Official Statistics, that contains the images of all publications issued by Statistics Portugal since 1864, totalling over 1.5 million pages.

#### Sale service

The customers can also acquire both published and "tailor-made" information.

- Published information can be acquired at Statistics Portugal premises or ordered on-line, by fax or email.
- "Tailor-made" information involves research and/or treatment of information. As a result a value-added product is obtained, whose costs are borne by the customer.

The deadline for its delivery depends on the complexity of the research required. Its sale is subject to statistical confidentiality and data representativity. This information can be obtained through the same channels as the published information.

<sup>12</sup> For further information, please see:

[https://www.ine.pt/ngt\\_server/attachfileu.jsp?look\\_parentBoui=55229723&att\\_display=n&att\\_download=y](https://www.ine.pt/ngt_server/attachfileu.jsp?look_parentBoui=55229723&att_display=n&att_download=y)

## 6.2 Dissemination of grid data for tax information in France: issues and solutions

In 2013, INSEE has released for the first time grid data for sensitive variables, namely tax information resulting from Localised Tax Revenues (“Revenus fiscaux localisés”, RFL) relative to year 2010. This information is disseminated in a grid of squares of 200m.

Squares constitute a precise and regular partition of national territory. They are also stable over time and independent of any other administrative partition. Therefore, grid square data make more relevant diagnoses on territories.

However, releasing grid data requires a precise geolocation process that can be cumbersome and costly, and it also requires some disclosure control methods to preserve statistical confidentiality. In the case of tax data dissemination, stakes are important because tax secrecy is particularly strict, since no figure can be disseminated relating to a population less than a threshold of 11 households. In Metropolitan France, among the 2.3 billion of inhabited squares (of 200m), 36% have only one household, and 80% have less than 11 households.

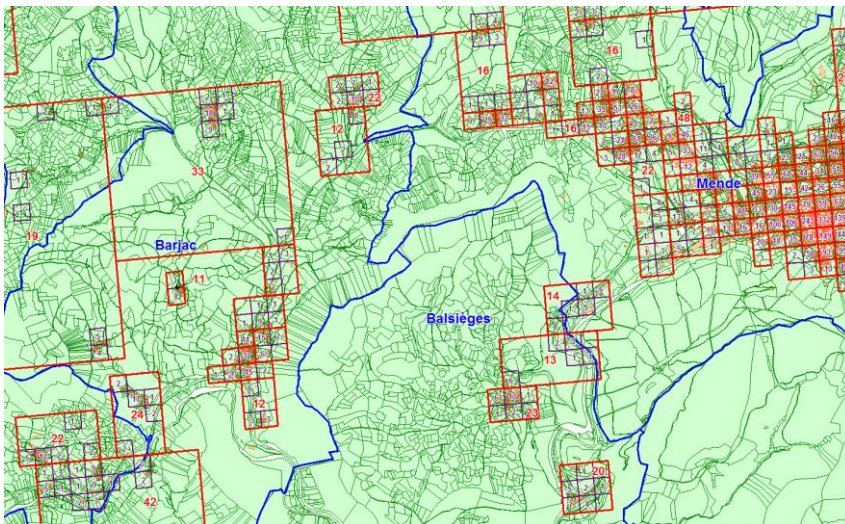
Grid data dissemination had to be associated to disclosure control methods. Initially, several methods were considered, including record swapping, imputations etc. But finally, INSEE decided not to disseminate data at the level of the square if the square counts fewer households than the threshold, but to aggregate squares into rectangles of at least 11 households.

Additionally, issues related to geographic differencing have to be treated, when sensitive variables are disseminated in different non-hierarchical partitions of the territory. In these cases, INSEE chose not to disseminate anything about observations within the overlap areas.

The algorithm of constitution of rectangles is the following one:

- Metropolitan France is divided into 36 large squares of identical size (for reasons of optimization of processing time)
- At first stage, every large square is cut in 2 either horizontally or vertically by its centre of gravity (weighted by the population) to form 2 rectangles. The algorithm arbitrates between a horizontal or vertical split by choosing the one that minimizes dispersion of inhabited squares within the 2 resulting sub-rectangles.
- The algorithm goes on until further cutting would result in non-compliance with the diffusion threshold rule.

As a result, the rectangles obtained can have variable sizes (see image below).



**Figure 16: Illustration of the rectangles created by the algorithm.**

For users, the file of rectangles has to be seen as an intermediate file: it should not be used as such, in particular for mapping representations. Indeed, since the rectangles are of varying sizes, counting maps would be erroneous, or internal spatial disparities could be masked within largest rectangles.

To go further, INSEE also disseminates the file of squares, but containing only numbers of household per square, without any variables. The user is then invited to create a new file by allocating, for each variable, the total number of the rectangle in proportion to the total population in each of its inhabited squares. INSEE disseminates a methodological note indicating the procedure to implement such a procedure with MapInfo or QGIS.

### 6.3 Constrains on data dissemination in Austria

Constrains on data dissemination mostly occur where data may be confidential or extra manpower is needed to prepare data according to customer wishes. Generally, as stressed in its mission statement, Statistic Austria's tries to provide data information in an easily accessible and comprehensible manner for further processing, following scientific principles and using various communication channels.

Having customer wishes in mind, who need tailor-made information or to whom place matters the standard regional units for dissemination are not always detailed enough. Yet, the highest responsibility must be taken to prevent the identification of individuals (persons, business enterprises etc.). The solution was found in offering the service of customer defined polygons and data on the basis of grids.

Grid statistics and the service of customer defined polygons require manpower; therefore they are commercial products and have to be paid for. As these products contain a lot of information on the location of the statistical data, the data portfolio is restricted and customers of grid statistics and customer defined polygons have to sign specific terms of use. With signing the terms of use customers agree to avoid misuse, updates are regulated and the usage rights defined for the different uses, internal use and commercial use.

In addition these small area products have to be evaluated for quality and confidentiality to prevent the identification of individuals. This is taken care for by applying the method of target record

swapping for some of the sensitive variables in the base data. In addition confidentiality thresholds apply based on the absolute count (e.g. of population, buildings, dwellings etc.) in each grid cell and customer defined polygon respectively.