

NEW WAYS FOR THE DISSEMINATION OF OFFICIAL STATISTICS: ACCESSING DATA THROUGH GIS

Enrique Morán Aláez¹, Mariano González Izquierdo², Jaime Garrido Vázquez³

¹ *Basque Statistics Institute - Eustat, Vitoria-Gasteiz, Spain; Enrique-Moran@eustat.eus*

² *Basque Statistics Institute - Eustat, Vitoria-Gasteiz, Spain; Mariano_Gonzalez@eustat.eus*

³ *Basque Statistics Institute - Eustat, Vitoria-Gasteiz, Spain; Jaime_Garrido@eustat.eus*

Abstract

Eustat has prepared a new way of accessing statistical data by means of a GIS. We think this new channel will provide users with a powerful tool to create and access to very detailed territorial data, not only for administrative or statistical areas but also to user-defined areas.

Eustat has created a new statistical database which includes data for each building; that is data about persons, dwellings, companies and establishments and their main characteristics (gender, age, facilities, economic activity, level of education, number of employees, etc.) tabulated at this level.

This web tool to access the data allows selecting a geographical area through a GIS; the users will be given the choice of selecting an area from an existing or of drawing an own layer, by polygons designed freely by themselves, by distances to a point or an axe or others.

Then they will be able to choose a subject from those offered by the database, a table, and some or all the cells in the table, which are aggregated to that particular geographical area.

An important question is to check for disclosure control. Due to the free selection, the area can be very small or it can include very few units. In both cases there is a minimum that must be attained to be allowed to download the information. Otherwise the user will be informed that no information is delivered due to statistical disclosure control.

There is also access to the statistical units from economic directories, with special restrictions to comply with the Basque statistical act. Some of these units are also elaborated as lists of equipment and facilities and presented as geographical layers.

Keywords: Geographical information, disclosure control, layer, free selection, territorial data.

1. Background: The history of the statistical production in Eustat

1.1. Territorial terminology.

From its very beginnings, Eustat had special interest in territorially detailed data. This was in part due to its limited scope of competence, but primarily due to the characteristics of its main users, amongst which Provincial and Town Councils immediately stand out, as do the Departments of the Basque Government.

As such, the 1981 Population and Housing Census - which was conducted in conjunction with the renewal of the Municipal Register of Inhabitants and in collaboration with the Spanish National Statistics Institute - began the coding of territorial units in terms of streets and numbers. In other words, what we now refer to as the “portal”.

The streets were coded by municipality in accordance with practices of Basque Town Councils and Provincial Councils since municipal data was first computerised - especially the Municipal Register of Inhabitants.

However, despite high expectations, tables of statistical data could only be disseminated on a municipal level, and only a very limited amount of data could be disseminated across districts and census sections.

1.2. The introduction of the “segment list”.

In 1991, the boundaries of districts and census sections and the boundaries of population entities came into contact with the municipal street map and the list of existing numbers and portals for each street. This led to the formation of the first “segment list”, which is the minimum set of streets plus a range of numbers belonging to a single village and a single census section.

This “segment list” was predominantly used so that Town Councils were able to classify their population and their housing correctly and hence avoid discrepancies between the most common territorial units at that time.

1.3. The creation of the territorial Database and the GIS.

In 1995, a Database that was based on the “segment list” was created, and it defined a new territorial unit known as a “space unit”. This unit was representative of housing and properties and was characterised by a horizontal and vertical postal address associated with the different existing territorial zones: municipalities, villages, districts and census sections.

The main source of data was to be the Municipal Register of Inhabitants and a thorough review of Basque territory carried out during the same year in preparation for the first release of Population and Housing Statistics, which was conducted independently by Eustat in 1996. For this reason, the data in the space unit s set aside for housing was poor quality.

Alongside this, and based on developments made by the Provincial Councils, we managed to come up with the first standardised Geographical Information System, which contained buildings and the various territorial layers mentioned.

As a result of this, Eustat was able to publish statistical data in 1996 for all residential buildings (ERAIKIN96) in the form of an Access Database in which all the buildings were identified with a coded postal address.

1.4. The integrated data system and NORA.

Once the Integrated Data System was launched, in 2000, the territorial Database was transformed into a corporate Database, and in 2006, Nora was set up as a corporate territorial referencing system for the Basque Government as a whole. Eustat’s role in this involved managing data related to the street numbers and buildings, so dealing with not only the alphanumeric aspect – i.e. the verification of addresses, management of registrations/de-registrations – but also the geographic aspect as all the new buildings are georeferenced using co-ordinates from the building’s main access point.

This made using administrative data generated by the Departments of the Basque Government easier and more reliable, and this data is in fact one of the main sources for producing statistics at the moment.

2. Main characteristics of LurData

2.1. Demand for territorially detailed data.

Since 1996, we have been able to deal with requests for provincially detailed data. However, it has only been possible to produce standard data on an infra-municipal level for neighbourhoods in municipalities with more than 10,000 inhabitants.

Data requests, on the other hand, have become more demanding: population by portals, employment by industrial zone, population and jobs in the areas surrounding to train stations, population by post code, ratio of bars by inhabitant and by street, commercial establishments just around the corner and so on.

LurData, which is the trade name of the new application, seeks to respond to these requests simply and online.

2.2. The statistics database.

As a result, Eustat has created a new database containing data from each building currently in the Basque Country, a figure that comes to about 200.000, in relation to the population and businesses inside them.

This database has been set up using data from the 2011 Population and Housing Census and the 2014 Directory of Economic Activities (DIRAE). Data from Municipal Statistics on Inhabitants, Housing and Education from 2012 to 2015 were then integrated, and the DIRAE update from 2015 was subsequently added as well.

The aim is to carry on integrating all exhaustive Eustat statistics such as statistics on births, deaths, marriages, migration and so on in addition to other existing directories. However, the main aim is to keep all the data that makes up the database up to date, and to ensure the respective statistical operations achieve this on a consistent, regular basis.

2.3. The graphic interface.

Eustat has developed a geographical interface based on the GeoEuskadi interface used by the Basque Government's Department for the Environment and Territorial Policy. This interface

allows you to access all the geographical layers developed by each dependent organism and Department - including those created and maintained by Eustat.

The creation of adequate graphic services enables you to use the different territorial units to request statistical data, or you can sketch other units freely using the tools available (polygons, lines and dots). If the sketch consists of one line, you are asked for the length of the area surrounding it, and if it consists of one dot, you are asked for the radius of the sphere of influence.

Once you have chosen or created your area of interest, the system saves it as a geographical selection criterion, and you are granted access to the statistical databases. Registered users will also have the chance to save these selections in a suitable workspace so they can use them at a later date. They can also download their own territorial zones.

2.4. Data searches.

The geographical selection you make then becomes a building selection, and the geographical coordinates for these buildings are included in the area/areas selected. You then select your subject of interest (population, economy, society and so on), the table/tables that contain the data requested and, in each of those tables, the cells required –it is indeed possible to select all the cells.

The application will then add the data from the buildings for the area/areas in question as a whole, and it then goes on to present them in an Excel spreadsheet instantly. This does not occur if the area is too small - in terms of the buildings it includes - or if the buildings do not include enough population, housing or establishments. If this happens, you are informed that the data has been deleted due to disclosure control.

2.5. Facilities and services layers.

Facilities and services layers have been developed that use existing data from the Directory of Economic Activities and from the Territorial Database. The units sketched move away in essence from DIRAE classifications of economic activity in order to become categories that bear more relation to non-statistical users.

The benefit of these layers, aside from showing the location of various services and facilities, is their ability to actually connect these services and facilities to the population using them, or to the economic activities that supply them. It is thus possible to determine the population living in the sphere of influence of a certain facility, such as a hospital, as well as the population that has a certain facility in its proximity.

To this end, there is a plan to create a “building sheet” what will record the presence or absence of a set of facilities and services that are defining features of that population's living conditions in an area located 10-20 minutes away from the building's entrance.

3. Pending issues and upcoming developments

3.1. New territorial layers

Eustat intends to create and issue new territorial zones for general interest, such as Areas of Economic Activity. This will include industrial estates, business parks, transport and logistic parks, technology parks, shopping centres, leisure centres and service centres.

It will be possible to use these zones to disseminate statistical data in tables in operations such as the Directory of Economic Activities. However, they are primarily intended for use by Town Councils, Provincial Councils and Autonomous Administration.

Subsequent to this, a zone related to university campuses will then be added.

3.2. The incorporation of municipal data

Eustat has a wide range of municipal data at its disposal. This data includes historical series that go back to 1981 and earlier, tackling more topics than the statistical database for buildings. The aim is to make this data accessible via GIS, an idea that may be beneficial seeing as this would make it easier to locate the municipality on the map, and seeing as a municipal/regional environment would be added to that municipality's location.

3.3. Other issues

The future development of LurData requires a huge effort on the part of Eustat in order to ensure that statistical units contain accurate territorial referencing, and that they can be captured on a map. There is also a need to make NORA the standard tool of use in administrative tasks that involve locating persons and activities in such a way that the resulting data can be integrated into ordinary statistical production.