

STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 20  
English only

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Geographical Information Systems  
(Brighton, United Kingdom, 22-25 September 1997)

Item (iii) of the provisional agenda

# **GISCO**

**STATE OF THE ART  
BASIC BUILDING BLOCKS IN THE MEMBER STATES  
A REPORT ON STATISTICAL SYSTEM IN THE EU**

Submitted by Eurostat

**Spring 1997**

## Management Summary

The focus of this study has been to investigate the different statistical systems in the European Union with respect to issues of input, processing and output. It has been considered important to understand the methods whereby the spatial components of small area statistics may be realised. The overall findings are summarised in this the "State of the Art Report". The main body of the report is presented in the following chapters in which each National Statistics Institute surveyed is discussed. This work is a continuation of two earlier reports<sup>1</sup> and is a part of the larger study "GISCO and Geographic Information supply, management and the implementation of GIS applications at NUTS 5 level".

Statistical systems are classed as being either register-based, area-based or some mix of the two. A register-based system is described as a system into which data are gathered, managed and manipulated via administrative records. Comparatively, area-based statistical systems generally use data gathering techniques that are highly dependant upon field work. For example, questionnaires and sample surveys are delivered to individuals or households whose answers are then recorded and allocated to a specific territorial unit (e.g. post code area, census tract). These classifications of the statistical systems have then been refined and presented as sub-categories.

Although originally the statistical data might have been gathered from either register-based or area-based inputs, the final structure of the small area statistics will be area-based. Perhaps some of the most important information for Eurostat lies in understanding these statistical products, known as output areas and functional zones. In fact, comparability of statistics will depend not only upon the way in which the data is collected, but also upon the methodology used to generate these output areas and functional zones. Likewise, it is suggested that processing methods and basic algorithms are different from Member State to Member State, but the exact extent of the differences is not yet known. Finally, the present inventory of the outputs of the different Member States suggests the short-term expectations for functional zones that can be provided to Eurostat. On the other hand, the medium- to long-term opportunities for functional zones will depend on a more complete understanding as to the processing methods on-going within each Member State.

### 1. STATISTICAL SYSTEMS

This work examines the state of the art with respect to the statistical systems of the different Member States of the European Union. And it is an attempt to summarise the most important aspects regarding the basic units of input, the different methods of processing and the potentials for spatially referenced statistical output.

The different statistical systems can be termed as being either area-based or register-based, or some mix of the two. Of the seventeen organisations studied, three were register-based, three were mixed systems, and eleven were area-based. And almost all of the seventeen organisations are operated in a centralised manner. Exceptionally, France, Germany and Spain have statistical systems that are decentralised both operationally and physically.

#### 1.1 Register-based

A register-based system is described as a system into which data are gathered, managed and manipulated via administrative records. The register-based system is dependent upon a number of factors, for example:

- the quality and dependability of the registers themselves
- uniquely identifiable records and codes

At present, systems that are entirely register-based include: Denmark, Finland and Sweden. Member States whose statistical systems are partially register-based include Austria, Belgium and the Netherlands.

**Denmark** has been a leader in developing register-based census, and ninety per cent of all data gathered is found in over thirty-seven different administrative/government registers.

Likewise, in **Finland** all data is being gathered through administrative registers. And as a matter of policy all

---

<sup>1</sup> *Geographical base unit: inventory of the situation in the Member States*, Document E/LOC/64. Presented at the Meeting of the Working Party on Economic accounts and statistical indicators at regional level - European infra-regional information system. 24/25 March 1997. *Census Geography as applied by the Member States*, Document GISCO/96/5. Presented at the Joint Meeting with National Statistical Offices and National Mapping Agencies. 25 November 1996.

administrative information is considered to be potential raw data useful for statistics. Finnish data are being spatially referenced via coordinates to precise locations; this geocoded information is found in the different government registers.

**Sweden** maintains a register-based input where the smallest units of collection are of two kinds: the individual person and the exact coordinates of the real estates (buildings). Until very recently, the Swedish system used a mix of registers and questionnaires. But in the future, and especially with regard to population censuses, information will be derived completely from administrative registers. To ensure the availability of data for the production of statistics, SCB has sought cooperation with other administrative authorities.

## 1.2 Area-based

Area-based statistical systems generally use data gathering techniques that are highly dependant upon field work. For example, questionnaires and sample surveys are delivered to individuals or households whose answers are then recorded and allocated to a territorial unit (e.g. post code area, census tract). Systems that are fully area-based: Germany, Spain, France, Luxembourg, Greece, Ireland, Italy, Portugal, and the United Kingdom for England/Wales, Northern Ireland and Scotland. To illustrate, in **Greece** the statistical system operates from two tiers: all data are finally compiled and published in a centralised way, but primary gathering of statistical information is regional process and relies on face-to-face interviews with the heads of households. Registers have not been used in the past because they were too limited in coverage and accuracy.

In fact, statistical systems that are primarily area-based often have and maintain different administrative registers. The degree to which these registers are (or might be) useful in preparing statistics depends on how the records are spatially referenced, and the level of accuracy that is maintained. In some cases the role of registers is diminished not due to accuracy, but for reasons of confidentiality. Information has shown that due to a strong reaction to concerns for privacy, a few Member States are moving away from, or are restricting the use of data found in their registers. This is the case with Germany, Luxembourg and France.

It is interesting to note that **Germany**, up until 1987, had been more open to the use of registers. And the results of the census used to serve administrative purposes, such as updating local population registers, but a ruling by the Constitutional Court has stopped this practice. **Luxembourg**, following the German example, no longer updates its administrative registers based on its census returns.

As of 1990, a number of new restrictions were placed onto the use of census information in **France**. For example, the town hall was no longer allowed to keep census information, nor could it use census data to update administrative registers. Conversely, the French statistical system provides "anonymous" individual data (microdata) to users in a form that refers to samples of the population (25, 3, 1, 0.1 per cent) or to the whole population. These data are made anonymous by deleting, restricting or omitting identifying information.

On the other hand, information shows that a number of so-called traditional area-based statistical systems use their administrative registers as supporting statistical resources. In some situations data gathering and management methods make equal use of register-based and area-based techniques; those Member States whose statistical systems are only partially area-based include Austria, Belgium and the Netherlands. Notably, a few Member States are beginning to move more towards having operational register-based systems. This is the situation for Belgium (which is already partly register-based) and Spain.

Presently, the **Spanish** statistical system draws information from the general censuses and from different administrative registers. For census information, the Census Section is the smallest unit of input (BTU).

Whereas, data from registers have not used because of their unreliability. Notably, each of the 8 080 municipalities maintains its own register of the population, and until recently there was no coordination between them - so omissions, inaccuracies and inconsistencies ("inflation") were inevitable. But due to recent legislative reforms, the situation with regards to the population censuses will lead to the creation of a centrally maintained register.

And, although Spanish law does not permit the communication of individual statistics (even to other Ministries), nor can the INE supply samples from a census or a survey (to scientific or market research purposes), it is newly allowed (as of 1996) for municipal registers to be updated from population censuses.

On a similar line, **Belgian** law, as does **Italian** law, requires that population registers are updated and amended in accordance with the general census. The Belgian statistical system is being oriented to move towards a register-based solution, on the condition that there is not a loss in the quality of the data that can be gathered.

### 1.3 Mixed

Perhaps what can be termed as a transition phase is typified by a mix of the two directions. As already indicated, mixed systems distinctly include: Austria, Belgium and the Netherlands. Specifically the Netherlands is the most strongly oriented towards the use of registers. But within this classification of "mixed", the individual statistical systems are moving in different ways.

**Austria**, like Ireland and the Netherlands, uses more than one BTU for input, depending on specific criteria.

Register-based information is then used to describe these BTU areas. For example, the Register of Buildings contains the main addresses of all buildings, and each building is considered as a separate entity and is assigned an object number (OBJNR). The Register of Buildings is in fact made possible by linking together other administrative registers. Eventhough, the underlying principles for gathering, processing and disseminating statistics remain area-based, and data records are summarised only down to the smallest geographic area, the building.

The **Belgian** statistical system is presently area-based, but is moving cautiously towards becoming a register-based system. The problem is that the available registers do not cover a number of statistical sectors that the traditional population census covers.

In **the Netherlands** the statistical system uses a specific mix of area-based and register-based statistical gathering techniques. Various administrative registers are used to collect information about small areas. For example, because every individual record in the municipal population register also contains the address, individual data can be aggregated to household information which can then be aggregated to sub-commune areas like census tracts.

### 1.4 Overview of the systems

The following summary table shows the diversity of the situations amongst the different Member States with respect to basic units of input that are available in digital form. The table reconsiders the classifications of the statistical systems, and presents what are very general groupings of sub-categories.

The category "Area-based" is a simple grouping of the different small areas for which digital boundaries exist. For each Member State the smallest boundary set covering the entire country is considered. Names such as enumeration area, census section, enumeration district, quarter, etc. are not used for two reasons: first - they are too confusing, and often have different spatial meanings from country to country; second - some countries use interchangeably two or more names to describe the same small area / boundary set.

The category "Register-based" differentiates between the manner in which the information that is derived from registers is linked and manipulated. All Member States that reported using registers also indicated that much of the information could be directly referenced to an exact address. In several cases the exactness of knowing a record's location is a matter of X-Y coordinates. In other situations, records are related to a boundary set by address matching techniques.

Table: Over-all comparison of the Member States.

Area-based			Register-based	
BTU	Country	Description	Country	Description
Very small size: sub-commune	Germany	30-50 households	Denmark	exact address + point coordinates
	UK - Scotland	15 households	Finland	
Small size: sub-commune	Austria	900 inhabitants ±	Sweden	exact address / no coordinates
	Spain	500-2500 electors	Belgium	
	Greece	40-50 housing units	Austria	
	Ireland	250-400 households	Netherlands*	
	Italy	400 survey units - max.		
	Portugal	300 dwellings - max.		
	UK - E/W	50-250 households		
NUTS - 5 "commune"	Belgium	commune		
	France	commune		
	Luxembourg	commune		
	Netherlands	municipality		
	UK - NI	electoral ward		

\* the Netherlands only uses coordinates with respect to grids.

## 2. INPUTS INTO THE STATISTICAL SYSTEMS

Each statistical system has its basic building blocks, whether these are area-based or register-based units of input. The criteria for defining the basic units of input will influence the entire method for gathering statistics, and thus influencing the eventual statistical and geographic output types.

Inputs are the means whereby data is identified and brought into the statistical system for eventual processing, manipulation, storage and use. In some cases the input unit is spatially defined with a unique area reference (e.g. housing block or census tract). In other cases the input unit is derived from a record in a register that is linked with some type of spatial definition (e.g. an exact address with XY-coordinates).

### 2.1 Register-based inputs

Member State	national level	sub-national level
Denmark	the smallest unit of input is the single unique	post codes are regulated by the
Finland	the smallest unit of input is the record of the-	
Sweden	the smallest units of input are the register of	the local tax office is responsible for

### 2.2 Mixed inputs

Member State	national level	sub-national level
Austria	the smallest unit is the SED (Statistical	each commune is free to choose
Belgium	the commune is the smallest "official" unit of input	each commune is free to choose
the Netherlands	the smallest units of input include: municipal	municipalities are responsible to

### 2.3 Area-based inputs

Member State	national level	sub-national level
Germany	enumeration districts (ED) are the	each Länder is responsible for creating
Spain	census sections are the smallest units	municipalities are responsible to establish
France	the commune is the smallest unit of "local"	branches of the INSEE define the
Greece	the census district (or enumeration	communes do not participate in conducting
Ireland	there are several BTU for input:	there is no specified involvement from the
Italy	the census section is the smallest unit	of communes are responsible to carry out the
Luxembourg	the only digital boundary data set	-
Portugal	the smallest area units for input include	census sections are proposed by the INE
England and Wales	enumeration districts (ED) are the	-
Northern Ireland	enumeration districts (ED) are a	-
Scotland	the postcode is the smallest unit of input	the Royal Mail creates the postcode

### 3. OUTPUTS FROM THE STATISTICAL SYSTEMS

Outputs are the ways in which small area statistics are presented. The original data might have been gathered from register-based or area-based inputs, but the final structure of small area statistics made available is area-based. Comparability of the statistics depends not only upon the way in which the data is collected and processed, but also upon the methodologies used to generate output areas. Here we suggest that processing methods and basic algorithms play a very important role. What remains to be seen is the extent to which these are different from Member State to Member State.

Additionally, not all statistical systems have as yet integrated GIS into their day-to-day work environment: Belgium, Greece, Ireland, Portugal, Northern Ireland, and until recently Spain. Therefore, not all output can be fully expressed with regards to its spatial dimension, and statistical output for different geographic levels (smaller than NUTS 5) is sometimes limited to a data base of statistical tables.

Table: Inventory of standard area-based (digital) output units from the Member States.

Standard output type	Description	Member State
<b>Postcodes</b>		
postcode areas	based on the postal rounds of the post office	DK, FI, AT, NL
postcode area	boundaries are derived from Thiessen	SE
enumeration district (1981); Output	these are based on postcodes (the BTU)	SE
<b>NUTS 5 level</b>		
parish, municipalities, communes, <del>and</del>	NUTS 5	DK, FI, SE, AT, BE, NL,
<b>Below NUTS 5</b>		
municipal sub-area	hierarchical areas, to 3-tiers below the	NUTS
SAMS - Small Area Market Statistics	boundaries are derived from Thiessen	SE
electoral district	boundaries are derived from Thiessen	SE
buildings (groups of)	buildings may be grouped together to form	AT
census district	former (old) municipalities/communes; <del>fi</del>	AT
census tract / census district	harmonised to match NUTS 5 boundaries	SE
census section	defined using the Administrative hierarchy	FR, NL
quarter	regroupings of the districts upto 5 000	FR
district	Originally used by the INEE for the	FR
census section (EA - enumeration area)	cover all the country, and are available	IT
BGRE: census sections	these boundaries are being made available	BE
enumeration district	same as the input BTU; 50-250 households	DK, NL
Standard output type	Description	Member State
<b>Grids</b>		
grid squares	1X1 km	FI, SE, NI
grid squares	500X500 metre	FI, SE
grid squares	100X100 metre : used only in cities	SE
<b>Urban Areas</b>		

urban areas / localities	UN definition: clusters of buildings with	at, SE, IE
population centre	defined using the Nomenclature hierarchy	FR
îlots	housing blocks that are to be available	FR
BGRE: census sub-sections	these boundaries are being made available	FR
<b>Rural Areas</b>		
rural agglomeration	population is 50 -199	SE
rural parts of NUTS 5	specific areas of NUTS 5 that are not urban	SE

#### 4. PROCESSING THE INPUTS AND OUTPUTS

One of the objectives of Eurostat is to achieve a harmonised approach to small area statistics. Thus it is very interesting to consider the results (output) of the different statistical systems. Outputs from the Member States, especially functional zones, provide the range of possibilities for infra-regional statistics.

The purpose of this section is to investigate the actual presence and use of output areas that are or will lead to the creation of functional zones - per Member State.

##### 4.1 Links between input, output and functional zones

What are termed functional zones are a type of special output created by Member States, but which are dependant upon certain user-defined criteria, often expressed in the form of algorithms (for example Urban Agglomerations).

The creation of user-defined areas is not only demand driven, but is dependant upon specific developments: technological, political, financial. Technical issues include: digital boundary management; GIS; telecommunications networks; other hard- and software developments. On the political front there are issues such as: confidentiality; copyright and IPR; trans-border and other trans-national issues; policy development and follow-up. Linked closely to the previous topics, financial resourcing will underwrite the extent to which a technical and political agenda can be achieved.

The following is a review of the situation with regards to the potential for functional zones in some of the different Member States. This information is illustrative in that it presents a sketch of the state of the art and a suggestion as to what is possible in the future.

##### 4.1.1 Functional zones in register-based systems

In **Finland** the basic unit of input, the georeferenced record, can be aggregated into different spatial patterns of output, depending on the users' needs. For example, from the different administrative registers it is possible to derive household-dwelling units, family data, place of residence - the latter by reference to commune, village property or block, building and dwelling and full street address. Already, there are a number of existing "standard" output areas (sub-municipal level) to which individual records can be aggregated, namely: municipal sub-areas, localities, postcode areas, grid squares. And because it is possible to identify the location of statistical data by map coordinates, a wide variety of other geographical areas can be produced.

Functional zones are possible due to the geocoding/georeferencing of individual records. Units are linked together via personal ID codes and domicile codes, additionally these units are linked to a system of coordinates (the centroids of buildings). Information about where people work is linked to other data on establishments via the company trade registration codes and the exact addresses.

The situation in **Sweden** is one similar to Finland in that data from different registers are brought together by real estate point coordinates and personal identification numbers - every person is connected to the Real Estate where they live, and every Real Estate has X-Y point coordinates. The actual production of small area statistics requires the use of unique, stable identifiers that are easy to use. For example, any analysis of urban and regional planning will have three data objectives: the individual (PIN), the firm (business code), the real property unit (property code).

Many requests are for output customised small area statistics. To manage these demands, the "Key-Code" System was developed by Statistics Sweden specifically for communes to use when delineating functional planning

areas. Local authorities draft a subdivision of their territory and list the property units in each of these sub-areas, by doing so the configuration of the sub-area is a product of the local conditions within each commune. This information is presented as a list containing the important codes that link different statistics to these specially defined areas.

Functional areas can be achieved in different ways, either by aggregating small areas to create new larger (hierarchical) areas or by using "artificial" (non-hierarchical) areas created with Thiessen polygons. For example, a number of artificial digital boundaries are created using the technique of Thiessen polygons: boundaries are placed half way between points having different codes (such as postcodes, electoral district codes, Small Area Markets). On the other hand, NYKO's are sub-commune areas that fit within the hierarchical scheme, but only a few communes have these boundaries in digital form. Nevertheless, it should be possible to create the NYKO boundaries in the following manner:

- a.) Both the Real Estate Register and the Register of Total Population contain codes for counties, municipalities, parishes, and the NYKO. As all Real Estates have point coordinates, it may be possible to allocate the individual Real Estates to create Thiessen polygons;
- b.) In the event that all boundaries of the Real Estates are digitised, then these can be aggregated to create NYKO boundaries that respect other hierarchical boundaries.

#### 4.1.2 Functional zones in area-based systems

The **Spanish** statistical system operates under a dual system of codes and data organisation. Data referencing and management can be either (or both) the "Administrative" or the "Nomenclature" hierarchies. The extent to which these hierarchies overlap or contradict each other is yet to be clarified. Eventually the expectations for the management of a centralised population register will have to be detailed in advance of planning for the provision of functional zones.

What is known is that data are collected and organised with reference to the Census Section. Each section can only fall within a single municipality, and the minimum size criteria is set by the municipality itself, but the size of sections are determined by the number of electors and not on surface area.

In **Greece** there is - at present - no direct link made between the statistical data and its geographic representation. This situation is seen as being primarily due to the lack of technological solutions such as an operational GIS. Any eventual link will depend on a number of issues being finalised. In fact, each Census District already has a unique code number and a delimited boundary (in analogue form), and individual data records are associated to these unique codes. But information shows that the final reliability of census data has been called into question; specifically, the number of persons indicated in the census as being employed are much less than results from the Labour Force Survey.

In principal Census Districts can be integrated to form functional areas such as urban, semi-urban or rural areas for national purposes. Other functional areas can also be provided upon user request. The degree to which this is possible must be determined.

As per the definition used in **England and Wales**, functional entities exist to fulfil some need for a specific type of information, but for which there is not an administrative body created: administrative entities are not the same as functional entities. The Enumeration District (ED) is linked to attribute data, and is used to create administrative and functional entities.

In fact, input and output ED are the same thing. "Input ED" are planned to respect administrative boundaries (e.g. civil parishes, electoral wards, local authority districts, counties) and major physical obstacles (e.g. motorways, rivers, railway lines). Local requirements (e.g. to reflect homogenous housing) are also taken into account where they do not conflict with other planning constraints. Furthermore, ED are classified as being either urban or rural - by first determining the urban boundaries and then allocating individual ED to urban areas through the use of their centroids. "Output ED" provided the basic building blocks from which statistical information were created - based on the 1991 population census. In England and Wales the term "output area" refers to any type of geographical area for which statistics are produced. This is in contrast to Scotland in which Output Area (OA) are a specific type of geographic area. The smallest units of output for non Census statistics are wards or local authority districts - the final type of output area will depend on the type of statistics being used.

Recent efforts have been made to develop an address-based geography using postcodes. Here individual addresses are foreseen as the basic planning unit and would allow for the grouping of postcodes, thus providing a more flexible output geography. This change will impact both aspects of Input and Output, thus more information about the details of this initiative would be useful to Eurostat's planning.

In **Northern Ireland** with respect to Censuses, Enumeration Districts (ED) are used as input to the statistical system, and originally were designed to respect rural townland and urban ward boundaries and other functional boundaries. However, an electoral ward boundary revision occurred in 1992, and now the ED do not respect current electoral ward boundaries (the current NUTS 5 areas). At present, the ED are only used to process census forms, and are not yet in digital form. Each returned census form contains the ED code, and nearly all summary statistics produced by the Census Office for Northern Ireland are based on aggregations of the ED codes. Other statistics generated use the postcode or grid coordinate reference as the identifier for aggregation purposes. The extent to which ED might serve for functional areas ought to be considered.

Statistical connectivity is managed via the ED code, the postcode and grid coordinate. The hierarchical geographical structure of data is in three lines: the Census Enumeration District, the Postcode Unit, the 100m Grid Square. The first two can be aggregated upwards through a hierarchical system that fully covers all of Northern Ireland and respect statutory boundaries. Whereas the latter, Grid Square, does not coincide with the NI boundaries.

Links between the postcode and the **Scottish** statistical system are provided via the postcode index. Information contained in the index links the postcode with various geographical areas, namely:

Administrative areas - Local Government Region, Local Government District, Health board area, New Council area.

Electoral areas - Electoral division/district ward, Electoral ward, Parliamentary constituency, European parliamentary constituency.

Other census areas - Output Area, Locality, Urban/rural type, Civil parish, Inhabited island

Position and size codes - National Grid reference, Resident count, Household count, Delivery point count, Non-residential delivery point count

Other indicators - Split postcode indicator (delivery points in more than one local authority), Linked small user postcode, Date of introduction, Date of deletion

Each postcode has a number of delivery points, these are decided by Royal Mail (RM is the UK postal agency). RM does not consider ward boundaries when assigning these delivery points. The allocation of postcodes to the various areas held in the index is achieved by the process of "point-in-polygon". This GIS process requires digital boundaries for each output area type held in index. For these area types, the General Register Office for Scotland - GRO(S) stays informed of any changes, and based on this information the digital boundary files are regularly updated.

In fact, the standard output from the Scottish statistical system is presently the "Output Areas - OA", this has replaced the previous output of Enumeration Districts - ED. Due to the problems of comparability, etc. GRO(S) has proposed to Eurostat that the NUTS 5 area for Scotland should become the postcode sector. These actual changes and proposals should be considered with respect to the needs for the provision of functional zones.

#### 4.1.3 Functional zones in systems that are "mixed"

In **Austria** the basic inputs are both the individual buildings (real estate) and the administrative municipalities. Like Ireland and the Netherlands, Austria uses different BTUs depending on different criteria. Approximately 1.81 million buildings were counted in the 1991 Combined Census of population and housing (GZ). As of 1997, the automated property register contains details of approximately 1.9 million buildings. Buildings can be grouped together to form spatial entities of any kind, but always fall within a single NUTS 5 area (municipality). For most other statistics, the BTU are municipalities or hierarchically superior administrative units such as political districts or federal states.

By means of the units of the administrative breakdown (e.g. communes) other optional units (functional zones) can be created. Examples include: judicial districts, financial districts, labour office districts, or any of the NUTS levels. Units for which all subject-related data are presented include Länder, administrative districts, communes, SEDs, MEDs. Units for which only certain data are presented, but where all subject-related data may be presented by means of aggregation are: NUTS 1, NUTS 2, NUTS 3, juridical districts, urban regions, and certain data in the data base ISIS. All data is available by the aggregation of buildings. (ÖSTAT-c, Desoye).

Units for which only a restricted number of data is presented are the break down of settlements (e.g. Ortschaften with 17 200 and Ortschaftbestandteile with 34000), localities (a continuously built-up area as per the UN definition), and localities are determined expressly for the purposes of the census. (Desoye).

Some territorial breakdowns in Austria were designed to address special administrative responsibilities, and produce data only for specific tasks (e.g. financial and juridical authorities). (Greul).

Urban Regions are functional regions that integrate the major urban centres with the countryside. By 1981 the

continuously built-up SEDs were combined into localities, yet in 1991 the localities were no longer based on SED's but were defined on the basis of Ortschaftsbestandteile ("structurally coherent parts of the settlement"). (Greul).

Functional zones (tailor made areas) are areas defined by the clients and follow certain procedures of aggregation, and these functional zones are different to those previously mentioned. The clients are obliged to pay the costs, and they are supplied with non-confidential data matching the characteristics as defined. (Desoye). For example, by using the results of the combined census, buildings can be grouped together to form any type of area (also called project areas), and data for these areas can be obtained from the GZ, but only if at least 4 buildings or 30 inhabitants are concerned, and data protection and statistical confidentiality must be assured.

Some of the functional area that are created in Austria are beyond the normally anticipated needs. For example, in Austria it is not considered very important to devise an urban/rural classification, but the definition is carried out based on international recommendations and to answer international questionnaires.

In the example of **the Netherlands** we find two very important types of geographic support systems: administrative and geometric. In the processing phase these support systems are linked with automated data files as well as digital map data.

To georeference data administratively the following codes can be linked:

postcodes, Census Tract codes, national coordinates of grids of 500X500m and municipal codes. Every link between these various codes is possible using the address, which is the postcode and house number.

municipal codes are linked with the codes for the various regional subdivisions of the country, of which there are more than 20.

All administrative areas *as well as nearly all functional areas* in the Netherlands share the municipality as the basic spatial unit in their delimitation. Furthermore, some of these areas are hierarchical by design, and respect the boundaries of higher administrative levels. And the relationship between most of the administrative and functional areas, as well as the relationships amongst the different administrative areas are complex and overlapping. (Glickman).

The relationship between subdivisions on the sub-municipal level are complex and overlapping as well, and many intersections exist between the various areas. To cope with these intersections, relational systems have been developed between such areas. The most important being the GBF: a relational system between postcode areas, Census Tracts and districts as well as grids of 500 X 500 metre, using the address key.

Statistical (or functional) subdivisions are defined mainly by statistical criteria and have the specific purpose to give insight into spatial distributions of different phenomena. These statistical subdivisions have importance for policy and research purposes. (Vliegen).

Two approaches are recognised for the purpose of delimiting regional subdivisions in the Netherlands: zonal and nodal. In the first case, zonal identifies regions that can be considered internally homogeneous while being different with respect to the relevant characteristics used for delimitation (e.g. agriculture or tourist regions). The second situation is nodal, and emphasises functional relationships (e.g. home-to-work) between the different parts of a region - usually a hub and its satellites.

## 4.2 Processing the inputs

Processing is addressed to the methods used to manage and manipulate the data gathered from the input. Processing is driven by a number of forces, namely: organisationally perceived needs, technical solutions and operational constraints (e.g. confidentiality laws). These forces will influence the types of statistical information to be generated and disseminated. Therefore, processing methodologies need to be examined further with respect to:

- present situations
- future alternatives

Within the framework of processing, a clear understanding is needed as to the use and potential of GI and GIS by the different statistical systems. Within the scope of this present work, it is known that GIS have been operationalised in a number of Member States, and for the remaining statistical systems, the integration of GIS is planned. The following offer some examples as to the motivations (organisational objectives) and uses (special units) of GIS in the statistical systems.

In 1991 GIS came to Statistics Finland with the objective to design new statistical products, to improve the handling of large bodies of place-referenced data and to maintain the organisation's own boundary information.

To better use the GIS resources, Statistics Finland set up a special unit to provide geographic information by

grid square, block, town quarter, postcode area, commune sub-area, locality, commune, economic area, work area, etc. In other words, the GIS unit is fulfilling the needs of users for functional zones.

Geography Area Planning System (GAPS) in England and Wales applies a GIS to define EDs for the 2001 Census using digital maps and boundaries and the OS Address-Point gazetteer. The systems at the test stage.

The Geographic Support Service (OGSS) has been set up to provide a corporate geographic service for OPCS (now the ONS) in England and Wales. Part of this service includes the development of a data base of geographic areas and unique names and codes for every geographic area.

The census office for Northern Ireland does not have a GIS, but plans to acquire one for the 2001 census. GIS is not used during the collection or processing of statistics. It is planned to develop a Geography Area Planning System (GAPS) that is similar to that of England and Wales. The Census Office (NI) also envisages creating new areas by using 1991 ED digital boundaries together with other Ordnance Survey (NI) products. (OPCS-c).

In the past there was automated cartography but now ÖSTAT is just being develop its GIS office, and plans are being laid out for the new possibilities in local statistics analysis that will be made possible. The GIS is being established using ARC/INFO. ÖSTAT has already digitised a number of features, including enumeration districts and settlement gravity centres. Furthermore a hierarchical grid (2.5, 5 and 10 metres) was developed, and by means of the gravity centres of the enumeration districts, it is possible to present census data to the grid. (ÖSTAT-a).

### 4.3 Feasibility to Eurostat

A geography for the European Union is an ambitious undertaking. As a discipline geography requires the use of geographic space and location to bring together in a meaningful way information about where people live and work, and the links between time and space. What is needed is a geography that can be used at the European level to describe the environment, the physical factors of the Earth and the different ways in which European citizens through their policies and preferences, go about the task of day-to-day living.

Therefore, it would be useful to consider how to arrive at a method whereby Eurostat receives output (functional zones) from the Member States based on well defined needs and objectives which have been translated into algorithms useful for processing spatially referenced data.

The inventory of the outputs of the different Member States describes the short-term expectations for functional zones that can be provided to Eurostat. On the other hand, the medium- to long-term opportunities for functional zones will depend on a more complete understanding as to the processing methods on-going within each Member State. Already, based upon the known situations with respect to "Input", an appreciation exists as to the constraints that will influence subsequent processing and output. There are a number of different types of output that can be considered, namely: Grid-based, Area-based with coordinates, Register-based with coordinates.

#### 4.3.1 Example of the Urban Agglomeration

One phenomena that is apparent in a number of cases is the detailed definition of urban areas and locality. There is a certain amount of information describing the extent of the urban area at sub-commune levels. This situation should be explored with respect to the potentials of algorithm definition, the ability to apply algorithms uniformly, the flexibility of the Member States to manage and update this information, the difference between register and non-register approaches, etc.

The following is a review of the urban and rural information that is presently available in the different Member States.

In **the Netherlands** municipalities are classified as rural or urban. Rural municipalities have the distinction made between built-up areas and sparsely populated areas. Urban municipalities make the distinction between the degree of built-up areas, with the differences between the urban areas being derived from socio-economic information.

Urban/rural classification in **Sweden** is evident in the distinction of the rural parts of the parish (NUTS 5) and in the rural agglomerations.

In **France**, "urban units" are the communes defined as being urban, and having built-up zones with more than 2 000 inhabitants.

The Census Districts in **Greece** can be integrated to form functional zones such as urban, semi-urban and rural areas. And as well, the smaller unit of the "building block" can also be grouped to form these same area types. But at this time these are only available in analogue form. It remains to be seen if these will come into digital form and when.

In **Ireland** the situation is quite unclear. The output areas for which digital boundaries exist include all the DEDs and

only the urban EAs. These boundaries are available from the NMA, but are not used by the Central Statistics Office - CSO(NI). The CSO does make Small Area Population Statistics available for the DED level on a country-wide basis, while EA data is provided only for the larger urban centres. In Ireland, urban areas play an important role, and these areas are only defined after the census and following UN guidelines. Once defined, the urban areas are used to aggregate various statistics. The extent to which these boundaries exist in digital form, or how they might be managed is yet to be determined.

For **Belgium**, the sectors have information about urban and rural classification. Sectors are based on the old commune boundaries, and have characteristics on social order, economics, urban areas. Each sector has a code that links it to its historical commune, the category of urban or rural and change information from one census to another. Sector boundaries are not available digitally at this time.

Regarding the situation in **Germany**, it is not known if, or to what extent the ED boundaries from the Länder are made available in digital form. Nor is there information about the type of urban-rural information that might be available at a level below the NUTS 5.

In **Italy** the communes identify what are termed "inhabited localities", which are areas that (following the UN definition) have a place name and upon which houses are located. The distinction for "built-up area" is based upon the type of human settlement found. In the early 1990's, via the CENSUS project, urban areas (built-up areas) were defined with the help of Remote Sensing technology.

Further possibilities for creating functional zones are provided by the Census Sections (also known as EAs). It is possible to aggregate census data into new areas that are built with the EAs. What remains is to determine the "how" of aggregating; this is a matter of understanding the criteria and defining the algorithms.

For **Luxembourg** there is statistical information (only from the population census) made available at the level of the quarter and locality. But, this is constrained by the fact that the boundaries for these areas are not yet digital.

In **Portugal**, below the NUTS 5 level are the sections and sub-sections (referred to as the BGRE). Specifically, the statistical sub-section is an urban area that corresponds to a continuous built-up area (following the UN definition): in urban areas these are blocks of buildings, and in rural areas these are locations.

Statistical sub-sections also have three levels of classifications: urban, semi-urban, rural. These are used to identify the NUTS 5 unit (the freguesias) as being urban, semi-urban or rural.

The Census Office of **Northern Ireland** plans to put the 1991 ED boundaries in digital form, and from this other "new" areas (functional zones) will be created. This is interesting in so far as the 1991 EDs respect the 1991 Electoral Ward boundaries (1991 NUTS 5). The EDs do not respect current Electoral Ward boundaries.

The CO (NI) considers the basic area unit of the census to be the grid square. All information is gathered at household level, and is then referenced to various area units - including the grid coordinates. Until recently the only digital / spatial representation for statistics, besides the NUTS 5 unit, has been the grids.

The General Register Office (**Scotland**) has used postcodes as input areas to define localities (urban areas) by using a mixture of criteria, including the density of population per postcode. Once localities had been defined, OAs were assigned to the localities by selecting one postcode (usually the one with the most households) belonging to the OA, and then assigning the entire OA to areas containing that postcode. To map the boundaries of localities, the aggregation of postcodes was used rather than the OAs.

The postcode index contains the electoral ward (1996 vintage). These wards are themselves assigned into approximate travel-to-work areas (labour market areas). In the past, these functional areas were originally defined as aggregations of wards (1983 vintage).