ESS Coordination of Geospatial Information and Statistics

Prague September 2012

Mr. Pedro DIAZ MUNOS
Mr. Danielle RIZZI
Mr. Gunter SCHAEFER

All statistics have a spatial dimension: a vision for production efficiency, cost reduction and generation of more accurate and value added information. Traditionally, geospatial information and statistics have been attributed to different organisational entities with very little need for cooperation. Technological progress and new policy demands have shown that both fields an beneficially be applied in their combination. For the European Statistical System this creates new organisational challenges and the need to cooperate both within the Member States with other organisations, such as the National Mapping Agencies, as well as on a European scale within the ESS. The 2012 DIGINS will discuss the various implications and sketch an action plan to approach the issue in a well-coordinated way.

1. BACKGROUND

1.1. Growing importance of GIS for statistics

Europe 2020 Europe's growth strategy for the decade, requires easily accessible, harmonised, timely, and appropriate information for its implementation and for measuring the achievements of its targets. In the current socio-economic context, marked by globalization and environmental issues, and affected by a financial crisis requiring structural changes in the European marketplace, an even deeper improvement of the EU information capacity has become essential. In particular, for efficiency and cost benefit considerations, the territorial dimension of socio-economic phenomena needs to be better taken into account in order to define tailored, location based policies. The EU's new high-level strategy, Europe 2020, explicitly introduces a third dimension on cohesion approach - territorial cohesion – which calls for greater territorial

disaggregation of traditional statistics and for a territorial perspective on cross cutting issues.

Citizens are directly affected by decisions having an influence on their immediate neighbourhood. Governments and local authorities need detailed information to illustrate the local effects of programmes, policies or infrastructure projects and must be able to demonstrate the benefits but also threats of their actions on neighbourhoods.

It is essential that the spatial dimension is taken into account and embedded into the statistical production chain from its very beginning, beyond the traditional approach often related to territorially large-scale administrative units. The spatial dimension, in fact, is already implicitly present in statistics, as a dimension affecting data collection, storage, analysis and dissemination, and as a means for creating different territorial aggregations. What is needed to achieve the objective of improving the EU information capacity is a more in-depth and integrated approach, where spatial dimension, at the most appropriate level of accuracy, is associated at all stages of the statistical data production cycle.

1.2. Mutual benefits for both areas - cross fertilisation

Beyond the improvement of the support to policy making, achievable through more accurate and better exploitable georeferenced statistical information, the further integration of the spatial dimension in statistics would generate added value at several levels:

- At the level of data collection, National Statistical Institutes costs could be reduced e.g. through the use of registers, with the support of GIS techniques for the optimisation of samples, and the simultaneous collection of statistical and geographical data, made possible by the availability of cheap and easy to use location enabled devices;

- Improvement of output and reduction of costs could be obtained due to better exploitation of the same reference information, whose georeferentiation would make possible aggregations to multiple, different geographical units (e.g. demographic or environmental statistics aggregated by statistical units, functional

areas, river basin districts, and at no extra cost), cross-theme linkages, and multidimensional spatial analysis;

- New requirements deriving from additional reporting obligations at the EU level could also be more easily fulfilled without the systematic necessity of implementing additional surveys.

The georeferentiation of statistics would also greatly improve the synergies between National Statistical Institutes, National Mapping Agencies and other GI authorities, improving the exchange and integration of information, avoiding possible duplications in data collections and facilitating the fulfilment of national and European reporting obligations (INSPIRE, environmental legislation) with the availability of harmonised and integrated spatial and statistical information.

This integration, therefore, in order to be as effective and resource efficient as possible, rather than be seen as a progressive convergence of two separated streams of information, made compatible and exploitable towards the end of the production chain, should be focussed on setting harmonisation principles at the start of the entire process. This means that the infrastructure, the georeferenced data themes, data models and tools should be harmonised as much as possible at Member State level, naturally leading to the harmonisation of the information collected at the EU level, thus contributing to the achievement of the ESS vision for the implementation of a comprehensive production following the "data warehouse approach".

The greatest challenge is therefore to implement a wide interoperability of the current statistical infrastructure with the standard functions of a Spatial Data Infrastructure so that they can become integrated parts of a comprehensive statistical production process.

2. CHALLENGES FOR THE ESS

In coping with the need to incorporate the spatial dimension into the statistical production chain, Eurostat and National Statistical Offices face a number of important challenges.

2.1. Bridging the geographical – statistical divide

Traditionally, geospatial information and statistics have been attributed to different organisational entities with very little need for cooperation. Thus, the division into Statistical Institutes and Mapping Agencies is common place in Europe. Cooperation between these types of agencies was fairly limited and is still today rather exceptional even for subjects for which this cooperation could be very useful. Beside the allocation of responsibilities, general practices, such as access by third parties to the information, were very differently organised and the condition of access differed greatly between the statistical and the GI world.

With the growing importance of geospatial information for statistics and the emergence of entirely new groups of users for Mapping Agencies, the need to organise this cooperation in a more effective way has arisen. In countries where location strategies have been defined, and as a consequence of the requirements generated by the implementation of the INSPIRE directive, a closer cooperation between NSIs and NMAs is taking place. Nevertheless, the two types of organisations have to continue and increase this cooperation in order to better exploit the information potential and make the information itself accessible for further use and in a user friendly way. Metadata practices with different content and structure are an example of this need for cooperation across the organisations. Often within statistical offices competences in geospatial information are still concentrated in specialized teams of GI experts; in these cases, optimising the exploitation of these competences throughout the whole organization would be very important.

2.2. Projects combining geospatial and statistical methods

The setting-up of successful flagship projects in which geospatial information is used as well as the successful dissemination of their results is the key to demonstrate the benefits of fully integrating geospatial information into the whole statistical process. Through good examples, for which the preconditions, resources, and other factors are well documented and communicated, the utility of georeferenced statistics can be concretely demonstrated. Such projects can show the advantages of the exceptional use

of geospatial information and can thus promote their evolution to a normal working tool for statistics.

For Eurostat and NSIs, it will be a challenge to select and organise such projects at a time of very limited resources. Therefore, the projects have to be well chosen. The cooperation within the ESS can be very helpful in identifying the most promising projects and the best possible way to organise them; pitfalls and exaggerated expectations could thus be avoided.

2.3. Cooperation in the ESS

The ESS has already a very good track record of successful cooperation on complex subjects. Working Groups and Task Forces have been effective means of organising common actions and are effective platforms for the exchange of information. In recent years, the ESS-Net projects have opened up new possibilities. The GEOSTAT project has already shown the usefulness of this instrument in the area of geospatial information. It has created already in its first phase a European wide dataset on population distribution on a grid based level. This detailed information has already been very useful for a joint project between the OECD and the Commission on a harmonised classification of rural and urban regions. This common concept of the degree of urbanisation introduced into EU wide surveys (LFS, SILC) will generate new statistical indicators that take into account the territorial diversity of Europe.

This shows to what extent close cooperation in the ESS can be very beneficial for future projects. On the one hand, successful projects of individual NSIs need to be communicated as best practices and, on the other hand, common projects between NSIs in some cases with the support of the Commission can be identified. Furthermore, this cooperation will be very useful to identify promising technologies or spatial analysis techniques.

One of the most important components of the cooperation within the ESS will be the identification of the issues to be better addressed at European level; handling confidentiality is one of the most critical ones. For these issues, Eurostat could take the lead in defining, in close collaboration with NSIs, recommendations, guidelines and

where appropriate legislative instruments for a harmonised approach in view of the definition of a seamless information infrastructure. Furthermore, a key factor for creating an effective information infrastructure is to have national legislation allowing the use of administrative registers, at the same time considering the methodological challenges necessary for making administrative data fit for purpose.

Cooperation should also take place for sharing knowledge and, whenever possible, developing common tools; solutions which are platform and vendor independent, based as much as possible on an open source software approach and on standardised exchange formats as XML should be privileged.

2.4. Creation of an effective information infrastructure

Geospatial information is already widely available but statistical techniques offer many more opportunities to increase the potential of this information. The key concept is georeferencing. Knowing the exact location of statistical indicators at the highest possible level of detail allows using this information for analyses well beyond the original purpose for which the data has been collected. Point based georeference is a robust and universal location concept which ensures stability over time and represents the best harmonisation base across all countries. One example is the analysis of social or economic characteristics for geographical entities different from administrative regions, such as mountain regions. Address coding is therefore a major issue on implementing an effective information infrastructure.

Furthermore, very different types of information can be combined in new ways. Examples could be the analysis of commuting pattern and for the optimisation of public transport from the location of employees to their workplace where both types of locations are known and can be combined.

The census 2011 exercise already triggered in several countries a strong move towards georeferenced data collections and to the spatially enabled statistics paradigm in general. The exploitation of census 2011 information provides a good opportunity for the setting up of a geostatistical information infrastructure whose benefits could be used to support its extension to other statistical domains. Besides this output oriented

advantages, the focus on geospatial data through the whole statistical production chain can improve the efficiency of this production cycle. Examples are GIS techniques used during the census 2011 to widely improve the efficiency of field work, or for the optimisation of samples.

An effective georeferenced statistical information infrastructure will of course need to be consistent and interoperable with the spatial data infrastructure under implementation according to the INSPIRE directive. With this consistency in place, geospatial information will be seamlessly exploitable at any level of the statistical production process, constituting in fact geospatial and statistical information, a whole set of data measuring physical or human phenomena, to be used together to organize the collection of new information (e.g. for planning and optimizing surveys), to aggregate information in different ways for different purposes (e.g. on grids, statistical units, functional areas, always using the same basic information), to derive new information (spatial analysis – proximity, service areas) and finally to disseminate in the most appropriate way, including the guarantee of the required degree of disclosure control.

2.5. Methodological issues

Working with geospatial information in statistics requires reconsidering a number of established statistical conventions and rules. The most prominent one is certainly the question of confidentiality. Traditionally, statistics put all necessary efforts into making the identification of statistical entities impossible or at least very difficult. This is done by eliminating information on statistical entities that would allow identification or aggregation to a level that prevents disclosure of individual records. Geospatial information is, however, largely dependent on reliable geographic coding. This will require specific considerations and possible different types of safeguards compared to traditional statistics, to be taken into account at all stages, from data collection and access to registries to data aggregation and dissemination. The stronger use of registries for statistics has indeed already raised this question and the work with georeferenced statistical data poses new requirements.

Disaggregation techniques are also needed to estimate information for small geographical areas. Methods such as modelling on small areas are relevant.

2.6. Accessibility and interoperability issues

Seamless access to information is the key aspect of a successful information infrastructure. To achieve this, two components need to be addressed: technical interoperability and the access/data sharing agreements.

A long lasting and solid experience on cooperating for improving whenever possible the interoperability of statistical information has already been developed, e.g. statistical metadata for exchange, discovery and use (SDMX, etc.). For implementing an efficient and effective geostatistical information infrastructure, further harmonisation issues, such as the integration/ alignment of the statistical and spatial (INSPIRE) standards for metadata and web services, need to be addressed.

Moving towards a richer information, including the spatial component of statistical information and potentially making available more specific and detailed information, raises issues not only at the level of confidentiality, but also regarding business models and commercially relevant data. The definition of harmonised conditions for easy or even free access to spatial and geostatistical information and, where needed, common licencing schemes will also have to be addressed.

2.7. Gaining competence in geospatial information

All NSIs have already gained experiences with geospatial information and Geographical Information Systems (GIS). The preparation of statistical maps is an established practice to display and illustrate statistics. Larger Member States also use statistical maps for showing statistics on regional and other smaller scale geographic levels. In many cases, NSIs have gained experience with spatial analysis projects. Thus, GIS is by no means a new technology for the members of the ESS.

However, in general, the competence in GIS aspects has been limited to a few GIS specialists while the work of the large majority of statisticians has hardly been impacted. These specialists use standard tools for creating statistical maps and/or consult a specialised service to create them for their specific purpose, in particular if

more complex representations or the combination of different display tools are aimed at. So far, the spatial dimension in the statistical context has mainly been employed in user friendly illustrations of statistical indicators at the end of the production process; statistics are traditionally associated to administrative areas which can be graphically represented together with other appropriate topographic layers (roads, rivers, etc.).

The full use of the potential of geospatial information and techniques requires a more intensive contact of statisticians with the subject. It would certainly be exaggerated to expect that the specific competence of GIS experts could be acquired by the average statistician. However, the level of competence needs to be raised, in particular, the awareness of the existing geospatial datasets that can be useful for the work of statisticians, and the recognition of the potential of spatial analysis and tools as well as projects in which geospatial information has already proved useful for statistics. This applies also to managers who should encourage staff to extend their competences to spatial matters.

2.8. Georeferencing at the core of the production process

It is of particular importance that the integration of the spatial dimension takes place at the data collection step of the production process, e.g. in surveys. In this way, the overall information (that is, fully georeferenced statistics at microdata level) can later on be fully exploited, allowing its use in more flexible ways and for a larger range of analyses, which would not be possible with the traditional indicators aggregated on the basis of administrative units. This broadened flexibility would imply a twofold improvement: the use of the same indicator for multiple purposes; the analysis of issues not necessarily already known at the moment a survey was designed.

3. HOW TO ORGANISE COOPERATION ON GEOSPATIAL INFORMATION WITHIN THE ESS?

Already today an established form of cooperation is in place:

1. The so-called GISCO Working Group meets once a year and reports to the DIMESA Group of Directors for environmental statistics. This GISCO working group has concentrated on providing information to Member States about GIS activities in the Commission as well as in related institutions; it also provides information on initiatives,

such as the Commission projects GMES and SEIS, as well as projects related to geographical analysis in the context of regional policy by the ESPON Programme. To a limited extent, delegates also come from National Mapping Agencies.

2. The cooperation with the European Association of Mapping Agencies (EuroGeographics) is a further regular subject.

3. The GEOSTAT initiative, now made concrete with two ESSnet projects and the support to the European Forum for Geostatistics (EFGS) initiative, was begun in the context of the GISCO Working Group. This initiative aims at developing common guidelines for the collection and production of spatial statistics and grid statistics within the ESS. The first GEOSTAT project has already delivered a prototype European population grid dataset for the reference year 2006 at 1km² resolution; the on-going new phase of GEOSTAT aims at the creation of a European population grid dataset of the 2011 census.

Eurostat proposes to upgrade the GISCO Working Group shifting its focus from information to cooperation. This may require a more frequent meeting schedule, and a higher level reporting channel as well as supplementary Task Forces to prepare or accompany common projects or to prepare recommendations on important issues. Evidently, it will be important to have the right levels of competence in the delegations of Member States in these meetings, combining knowledge on spatial data infrastructure, thematic data production cycle and data needs for territorial development. A closer link between GISCO, the Working Group on Regional and Urban Statistics and the EC interservice group working on INSPIRE (the so called INSPIRE Team) is therefore needed. The definition of additional Community supported projects is a further option for discussion in the Working Group.

Registers and other administrative data sources are often managed by other public authorities outside the ESS, e.g. population registers managed by municipalities. These authorities will have to be directly or indirectly involved in the cooperation, being part of the key stakeholders necessary for a successful European data infrastructure for geospatial information.

Given the cross cutting nature of georeferentiation in statistics, the GISCO working group requires a mandate from the ESSC also? with regard to coordination with the other statistical working groups to avoid that georeferencing is dealt with in isolation.

Since cooperation on geospatial information is a rather new field of statistical work, Eurostat proposes not only to continue reporting to the DIMESA Group of Directors, but also to report annually to the ESSC.

4. SUBJECTS TO BE TREATED IN THE COOPERATION ON GEOSPATIAL INFORMATION

The strengthened cooperation within the GISCO Working Group will need to tackle the following subjects:

- Definition of the priorities, of the issues to be better addressed at the EU level, of the steps for the integration of the spatial dimension in the statistical production chain, of the appropriate level of harmonisation to be achieved and of the optimal instrument for their implementation (guidelines, recommendations, legislation)

- Concentration on cooperation in geospatial information via common projects, on specific priority themes, and the exchanges of best practices in spatial analysis projects

- Analysis of the possible development of a EU Location Strategy, based on the on-going experiences taking place in Member States and at EU level

- Facilitation of the relationships between NSIs and National Mapping Agencies and other Geographic Information Authorities/Organisations, including EuroGeographics on the European level.

- Awareness raising and training of statistical officers on geospatial information principles, techniques and tools, addressing in particular the association and use of the spatial dimension in statistics.

- Support on INSPIRE implementation, especially in the area of statistical information; preparation of recommendations for the harmonisation or coexistence of a number of different standards, i.e. on metadata: INSPIRE vs. SDMX/ESMS

– Definition of guidelines for georeferentiation of the statistical units; definition of guidelines on georeferentiation for different types of statistical information and in different contexts, such as surveys or administrative information, e.g. registries, this last option to be privileged whenever possible

– Evolution towards a shared data and services spatial and statistical infrastructure concept by the development of web services and linked data for the dissemination of statistical information and the association between statistics and spatial data

– Recommendations on confidentiality rules and guidelines for internal and public access

– Contribution to the UN-GGIM initiative; definition of European contributions and priorities to be addressed at the global level

5. CONCLUSIONS

The ESSC recognised during its meeting of February 2012 the need for stronger cooperation on the use of geospatial information and its integration into the statistical production cycle.

The September 2012 DGINS will be an important step for endorsing a vision for a coordinated approach on geospatial statistics in the ESS and for launching important initiatives for its implementation.