

Findings from the 2020 GISCO / GEOSTAT 4 survey

Geostat 4

Title: Findings from the 2020 GISCO / GEOSTAT 4 survey

Project: Eurostat ESSnet grant project GEOSTAT 4

Grant agreement number: 945503 - 2019-FI-GEOSTAT4

Author: GEOSTAT 4

It is permitted to copy and reproduce the content in this report. When quoting, please state the source.

© GEOSTAT 4 and Eurostat 2021

Acknowledgments

On behalf of the GEOSTAT 4 project and the European Forum for Geography & Statistics (EFGS) we would like to express our sincerest gratitude to those of you who kindly responded to the GEOSTAT 4 survey. We understand that some questions were ambiguous and difficult to answer. Some of you also patiently faced additional questions and requests for clarifications.

With your committed support and valuable input, we have been able to achieve a high level of response for this second round of survey (first round by the GEOSTAT 2 project in 2015). This allows us not only to make a comprehensive overview of the current situation in Europe, but also to measure progress over time. We believe this is of value to the statistical-geospatial community far beyond the GEOSTAT 4 project team.

The GEOSTAT 4 consortium

Content

Acknowledgments.....	2
List of Abbreviations	4
1. Introduction.....	5
2. Result	6
3. Conclusions	18
4. References.....	19

List of Abbreviations

API – Application Programming Interface

EFGS – European Forum for Geography and Statistics

ESS – European Statistical System

EU – European Union

GIS – Geographic Information System

GISCO – ESS working group on Integration of Statistical and Geospatial Information

GSGF – Global Statistical Geospatial Framework

NGIA – National Geospatial Information Agency

NSI – National Statistical Institute

1. Introduction

As part of the GEOSTAT 4 project (February 2020 - February 2022), a questionnaire was sent to all the national representatives of the GISCO Working Group, prior to the working group meeting in March 2020. The aim of the survey was to collect information about the current state-of-play in Europe regarding infrastructure for data integration and expectations from the community on the GEOSTAT 4 project in terms of guidance, practice and training. Some questions were designed to allow a follow up on surveys conducted by previous GEOSTAT projects and to measure progress made in countries.

Respondents were urged to coordinate their reply between National Statistical Institutes (NSIs) and National Geospatial Agencies (NGAs) and if possible, leave one reply per country. A few countries left multiple replies. In these cases, the responses were merged into one reply.

The initial plan was to conduct an interactive workshop in the GISCO Working Group meeting 2020 where the result was going to be presented, combined with discussions and collection of additional input using an interactive platform. However, as the meeting was cancelled due to the global Covid-19 pandemic, the survey result was the only way to systematically collect input for the project. The result from the survey has been fed into the GEOSTAT 4 project, for the consortium to be able to provide the best possible guidance in line with the expectations of the statistical-geospatial community.

The survey target group comprised the ESS countries (EU and EFTA), together with EU candidate countries, potential candidates and the former EU Member State United Kingdom. In total 40 countries were included. The overall response rate was good, in total 34 countries responded. The non-responding countries in the target group were Bosnia and Herzegovina, Czech Republic, Liechtenstein, Moldova, Montenegro and United Kingdom.

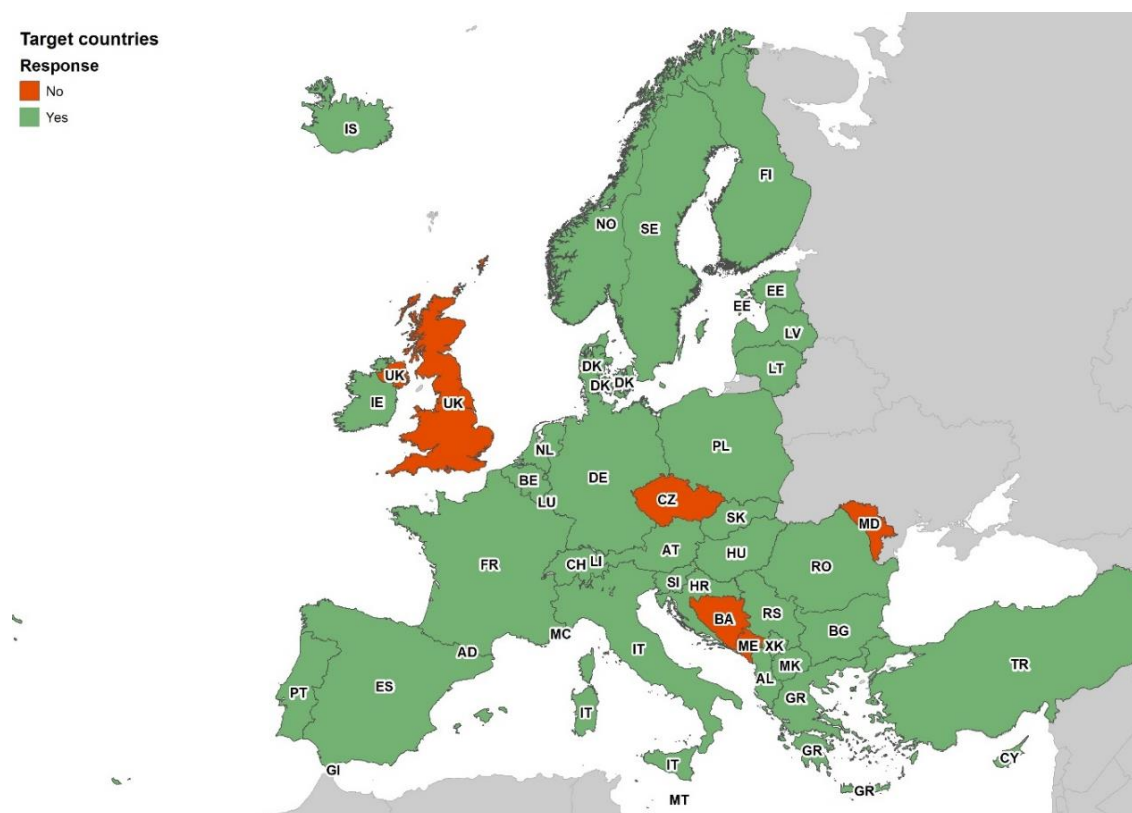


Figure 1. Map of target countries and responding countries as of April 9, 2020.

Among the EU Member States (EU27), the response rate was 96 percent. Czech Republic was the only non-responding country in this group. Among the 31 ESS countries (EU27 + Iceland, Lichtenstein, Norway and Switzerland) the response rate was 94 percent. In this group, Czech Republic and Lichtenstein were the only non-responding countries.

Compared to the GEOSTAT 2 survey, conducted in 2015 (GEOSTAT 2, 2016), the response rate of the 2020 survey was slightly lower. On the other hand, the 2020 survey included responses from Malta and North Macedonia, which did not participate in the 2015 survey.

2. Result

Question 2.1 - What is the lowest possible geographical level to which your country will be able geocode population data for census 2021?

Question 2.1 was designed to follow-up on the progress made among countries in terms of data infrastructure for point-based geocoding of census data. The same question, with similar response options, was included in the 2015 GEOSTAT 2 survey and a related question (not identical though) was included in an earlier assessment made by the GEOSTAT 1 project already in 2010 (Rademacher 2012).

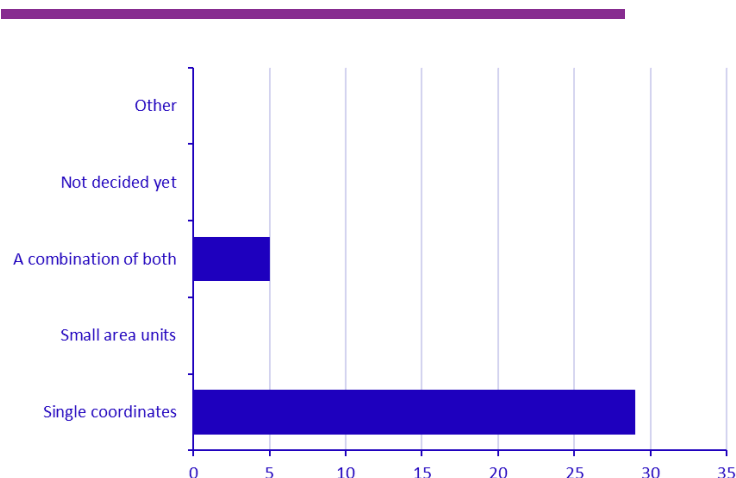


Figure 2. What is the lowest possible geographical level to which your country will be able geocode population data for census 2021 (number of countries)?

Full comparability between surveys is not possible due to different response rates (and not the same countries covered in all surveys) but the overall result shows a substantial progress made during the last 10-year period. In the 2020 survey, 29 of the responding countries report that they are able to geocode population data to single coordinate location.

In the GEOSTAT 2 survey, 21 countries reported that the lowest possible geographical level to which they could geocode population data was single-point coordinates (e.g. address locations or buildings). If we put together the results from 2015 and 2020, it is reasonable to believe that at least 80 percent of the 40 target countries now have the ability to fully geocode population data to the level of single point-coordinates.

Another striking observation is that in the GEOSTAT 2 survey, 10 countries reported that the lowest geographical level to which they could geocode population data was small areas (such as enumeration districts, blocks or small administrative units). In 2020, none of the responding countries has reported this option. However, five countries report a combination of single coordinates and small area units as the lowest level.

Question 2.2 - Sustainability of the data infrastructure needed for geocoding and integration of statistical and geospatial data

Question 2.2 was designed to gain a better understanding of the sustainability of the underlying data infrastructure used for geocoding and data integration. Whereas question 2.1 only gives a momentary view of the situation, question 2.2 seeks to investigate if the data infrastructure is concurrently and effectively maintained or if big efforts have to be undertaken to put data together and harmonise it for the purpose of Census operations etc.

This question partially follows up on previous surveys but the question was not asked exactly in the same way as in previous questionnaires. Consequently, full comparability with the previous survey is not possible though some general conclusions can be drawn on the progress over time.

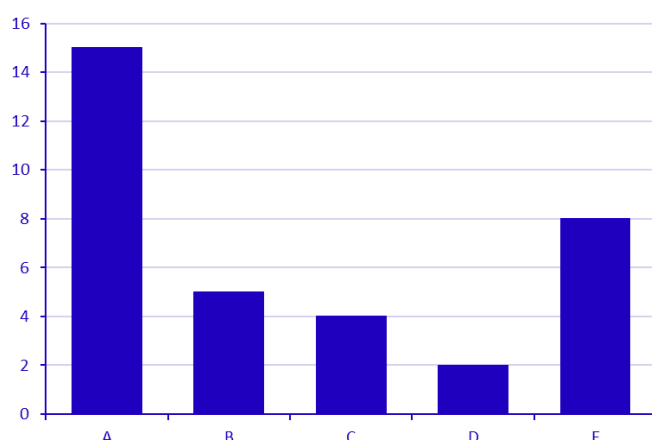


Figure 3. Sustainability of the data infrastructure needed for geocoding and integration of statistical and geospatial data (number of countries).

A - High quality standardised and continuously maintained data on address locations and/or buildings suitable for geocoding purposes exist in our country. Data can easily be obtained via national access points. A number of public institutions use the same data sources.

B - High quality and continuously maintained data on address locations and/or buildings exist in our country, but cannot be easily obtained via national access points. Data has to be retrieved from a number of regions and/or institutions and brought together and harmonised before use. Besides lack of national access points, data is more or less fit for purpose.

C - Data on address locations and/or buildings exist in our country, but is geographically scattered and with uneven quality. The lack of conformity and standards prevents us from using this data in Census operations (e.g. we have to create our own census address or building files).

D - Data on address locations and/or buildings does not exist or has only partial coverage in our country.

E - Other

The majority of the respondents (15 countries) reported option A, which can be considered the highest level of sustainability. This indicates that high quality standardised and continuously maintained data on address locations and/or buildings suitable for geocoding purposes exists and that data can easily be obtained via national access points and is widely spread within public sector. If put together with option B, which is the second most sustainable level, roughly half of the target countries have a very high, or relatively high, level of sustainability.

In total six countries have reported C or D, which indicate a low level of sustainability due to lack of harmonised national data or problems with coverage or quality. A fairly large number of countries reported option E, which indicates that the respondent wants to give an answer different from the four predefined options. Typically, the reason is that there is development under way to improve or resolve the data situation, potentially moving the country to option A or B within a few years' time.

Question 2.3 - Who is responsible for creating and maintaining the point-based reference data that are used in your country to geocode statistical unit record data?

Question 2.3 was designed to obtain a better understanding about the responsibility to produce the geospatial reference data needed to geocode statistical unit record data.

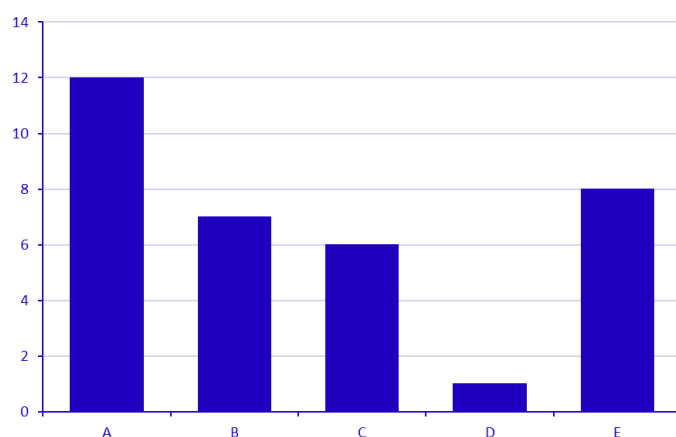


Figure 4. Who is responsible for creating and maintaining the point-based reference data that are used in your country to geocode statistical unit record data (number of countries)?

A – National Geospatial Agency (alone or in cooperation with regional agencies and/or local authorities)
B – NSI (alone or in cooperation with regional agencies and/or local authorities)
C – Both NSI and National Geospatial Agency (including cooperation with regional agencies and/or local authorities)
D – No one/very unclear responsibility
E – Other

From the GEOSTAT 2 survey, it is known that the situation can be quite complex, involving several producing organisations on different level of government (national, regional and local). The 2020 survey confirms this complexity and the diversity of arrangement found in different countries.

Obviously, the most common situation is that the National Geospatial Agencies (typically National Cadastral and Mapping Agencies) are responsible for collection and distribution of this data in collaboration with regional and/or local authorities.

Answer E reflects the fact that in a number of countries, there is no national agency involved at all or the role of the national agency is only to provide a common access point for data, e.g. they are not involved in the production/collection of the data itself. Comments provided for option E also indicate that there can be other national institutions (neither NSIs nor NGIAs) responsible for national repositories of address or building data.

Question 2.4 - Sustainability of the data management environment for geocoding and integration of statistical and geospatial data

Question 2.4 was designed to obtain a better understanding about data management environments for geocoding and integration of statistical and geospatial data. This question is related to question 2.2 but addresses the technical set up and tooling for data integration rather than the data infrastructure.

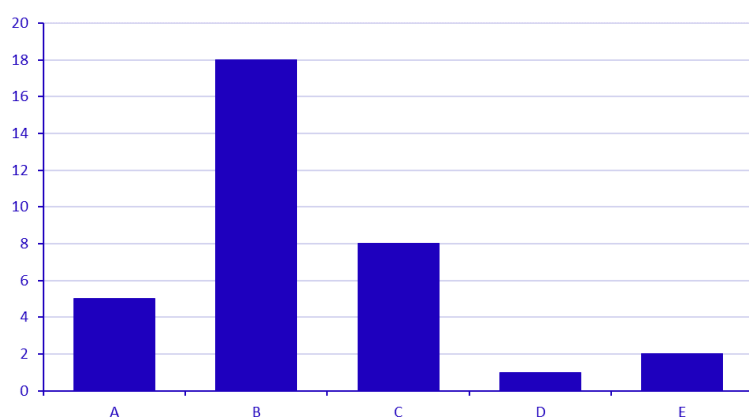


Figure 5. Sustainability of the data management environment for geocoding and integration of statistical and geospatial data (number of countries).

A – We have a well-structured and well-documented data management environment supporting systematic geocoding and automation in production of geospatial statistics data without particular needs for improvements.

B – We have a well-structured and well-documented data management environment supporting systematic geocoding and automation in production of geospatial statistics data BUT we see a need for improvement or modernisation.

C – We do not have a well-structured and well-documented data management environment supporting systematic geocoding and automation in production of geospatial statistics data. Our way to organise production may be non-efficient but it does not affect the quality of output in a negative way.

D – We do not have a well-structured and well-documented data management environment supporting systematic geocoding and automation in production of geospatial statistics data. Our production suffers from lack of efficiency and unfortunately restricts the content and quality of output.

E – Other

The answers reveal that most responding countries are highly or quite content with sustainability of their current situation. The majority of countries (18) reported option B, indicating that they have a well-structured and well-documented data management environment supporting systematic geocoding and automation in production of geospatial statistics. However, there seems to be room for improvement and modernisation. In total five countries have reported option A, which is the highest level of sustainability. If adding up option A and B more than half of the target countries have a sustainable data management environment for geocoding and integration of statistical and geospatial data.

In total eight countries have explicitly replied that they do not have a sustainable data management environment. Most of these countries believe that the lack of sustainability mostly have an impact on efficiency rather than the quality of the final output.

Question 2.5 - Threats and obstacles to statistical geospatial integration

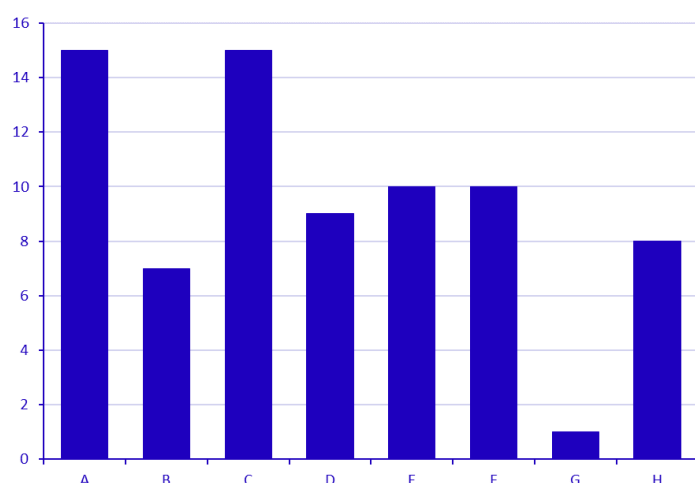


Figure 6. Threats and obstacles to statistical geospatial integration (number of replies, multiple options possible)

A - National standardised data (address records, building registers etc) does not exist, is incomplete or poorly maintained

B - Access to data is restricted (by legal or financial reasons)

C - Poor semantic or technical interoperability between different data sources or cross data domains (e.g. lack of consistent identifiers to link data or inconsistent data models etc)

D - Lack of coordination between data custodians and unclear responsibilities

E - Lack of know-how and/or human resources

F - A bit of all or some of the above mentioned, but no major obstacles

G - No particular obstacle at all, things work quite smoothly!

H - Other

The most significant obstacles for statistical-geospatial integration seems to be a combination of lack of standardised data (A) and poor semantic or technical interoperability between different data sources or cross data domains (C). These two options seems to be very much inter-related.

Very few countries describe a situation without any obstacles, though a quite large number of countries indicate that they encounter minor challenges but no major obstacles.

The questions in the GEOSTAT 2 survey were stated partially in a different way, making comparisons over time difficult. However, the general picture seems quite similar. In 2015, the most scored option was “inconsistencies in geospatial data needed for geocoding”. Largely, this question captures the essence of option A and C in the 2020 survey.

Question 2.6 - Use of administrative data sources for geospatial statistics

Question 2.6 was designed to assess to what extent countries have started to use administrative data sources as a basis for geospatial statistics. Increased use of administrative data sources is one of the strategic goals of the ESS vision.

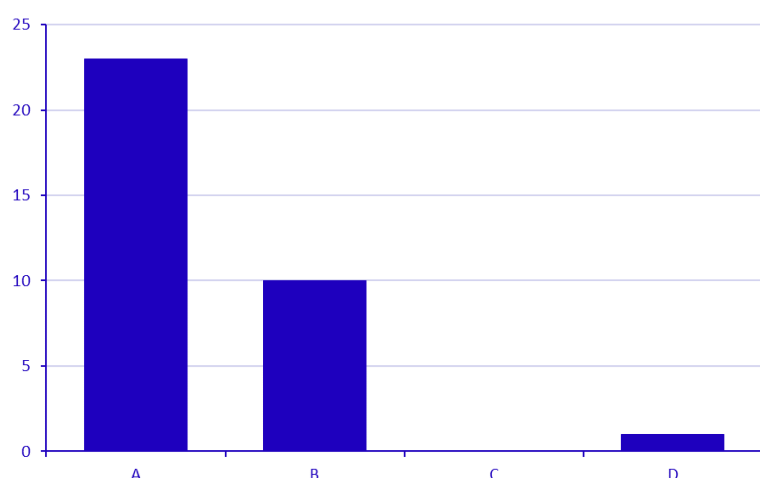


Figure 7. Use of administrative data sources for geospatial statistics (number of countries).

A - Administrative data sources have already been implemented in regular production of one or more of our geospatial statistical products

B - NOT yet implemented in our regular production, but we are currently looking into it or have plans to do it in the near future.

C - NOT implemented in our regular production. We have no plans to do it and we do not expect to be able to use administrative data sources in the near future.

D - Other

The overwhelming majority of the responding countries have already implemented administrative data sources in regular production of geospatial statistical products. Eight countries replied that administrative data has not yet been implemented in regular production, but they are currently looking into it or have plans to do it in the near future. All responding countries seem to foresee that administrative data will come into use in the near future.

The question does not reveal details about what kind of administrative data are used, nor does it describe how or for what purpose data is used.

Question 2.7 - Quality aspects of geospatial statistics

Question 2.7 was asked to provide input for the GEOSTAT 4 project as regards to its plan to develop a quality framework for geospatial statistics.

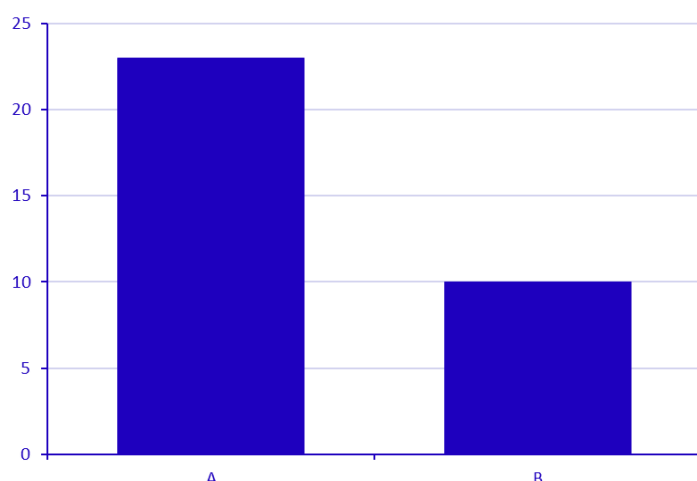


Figure 8. Quality aspects of geospatial statistics (number of countries).

A - No, our quality framework does not cover the geospatial aspect of the statistical production process and/or we do not have quality indicators in place to assess the quality

B - Yes, our quality framework cover the geospatial aspect of the statistical production process and/or we have quality indicators in place to assess the quality

Prior to this survey, little has been known about the existence of national quality frameworks particularly addressing quality aspects related to geospatial statistics.

Obviously, only a minority of countries have their own quality frameworks covering the geospatial aspect of the statistical production process and/or quality indicators in place to assess the quality. This fact strongly confirms the relevance of the decision by the GEOSTAT 4 project to develop such a framework.

Question 3.1 - Need for guidance

One of the main objectives of the GEOSTAT projects has been to provide guidance for countries to improve their capabilities to produce geospatial statistics. Question 3.1 aims to serve as a direction for the GEOSTAT 4 project in terms of development of further guidance. The options A-F loosely relates to the different principles of the Global Statistical Geospatial Framework (GSGF).

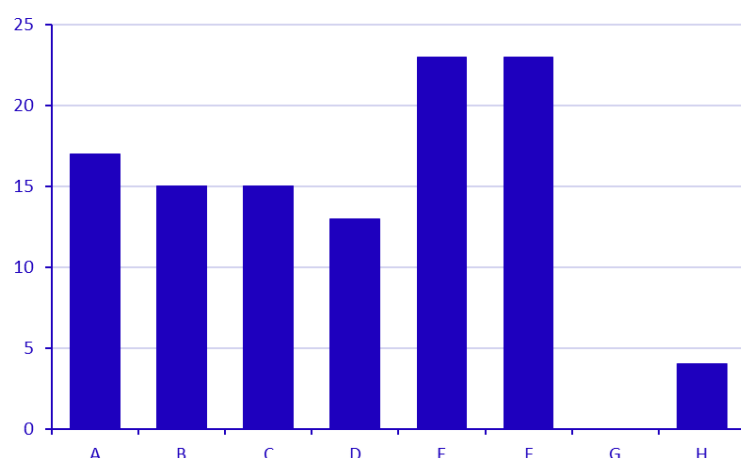


Figure 9. Need for guidance (number of replies, multiple options possible).

A - Data sources and data quality assessment

B - Geocoding and other methods and tools for data integration

C - Data management issues and architecture

D - Frameworks for, and management of, common geographies

E - Interoperability issues and standards data

F - Web services and tools for data dissemination

G - No guidance needed

H - Other

The responses clearly indicates that there is a significant need for guidance around interoperability issues and standards (E) and web services and tools for data dissemination (F). These two options score the same numbers. Judging from the comments provided, option F is also closely related to demand for guidance concerning dissemination of INSPIRE services.

An interesting observation is that the need for guidance is rather evenly occurring across the spectrum of target countries. There is no or little systematic difference between countries with a long history in the ESS and new EU Member States and candidate countries in terms of need for guidance.

Question 3.2 - Type of guidance

Question 3.2 relates to question 3.1 and aims to provide a direction for the GEOSTAT 4 project in terms of the type of guidance that is considered most useful by the statistical-geospatial community.

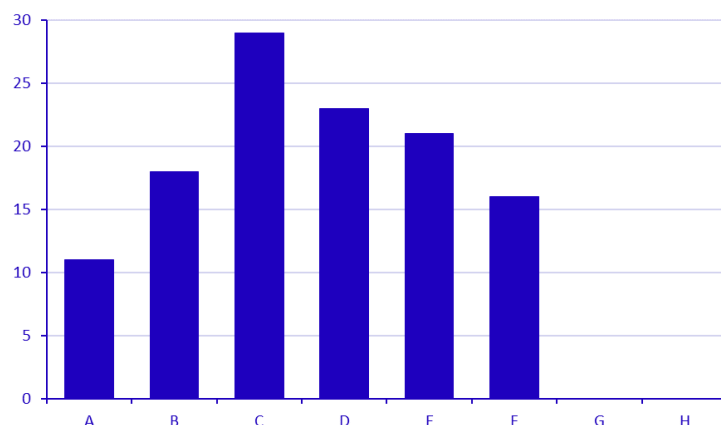


Figure 10. Type of guidance (number of replies, multiple options possible).

- A – Better or more elaborate interpretations of the different elements of the GSGF
- B – Common reference architecture models to support production of geospatially enabled statistics
- C – National good practice cases to benchmark with, or get inspired by, other countries
- D – Technical guidelines and manuals
- E – Concrete business cases to promote the potential of statistical-geospatial integration
- F – Proof-of-concepts for tools or services that can be tested and evaluated
- G – No guidance needed
- H – Other

The answers show a strong support/request in particular for guidance in the form of national good practice cases to benchmark with, or get inspired by, other countries (C).

Also technical guidelines and manuals and concrete business cases to promote the potential of statistical-geospatial integration is requested (D and E). However, in principle the general conclusion is that there is a demand for most types of guidance.

Question 3.3 - Business cases to promote statistical-geospatial integration

As a follow up question to 3.2, question 3.3 concerns suggestions on content for business cases to promote statistical geospatial integration.

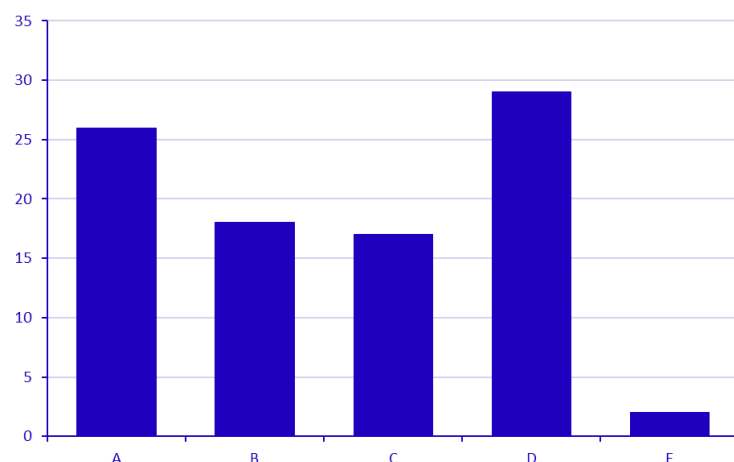


Figure 11. Business cases to promote statistical-geospatial integration (number of replies, multiple options possible)

A – The benefits of, and new products that can be retrieved from, a fully geocoded business register

B – New applications and products based on geospatially enabled health data

C – Accessibility studies involving a range of geospatially enabled data sources

D – New applications and products based on a combination of big data, Earth Observation data and “traditional” geospatial and statistical data sources

E – Other

The most popular suggestions are new applications and products based on a combination of big data, Earth Observation data and “traditional” geospatial and statistical data sources (D). The second most popular option is A, the benefits of, and new products that can be retrieved from, a fully geocoded business register.

Question 3.4 - Need for training and capacity building

One of the tasks of the GEOSTAT 4 project is to draft a concept for training and training material in accordance with the needs expressed by the statistical-geospatial community. Question 3.4 aims to collect input concerning content of future training and capacity building.

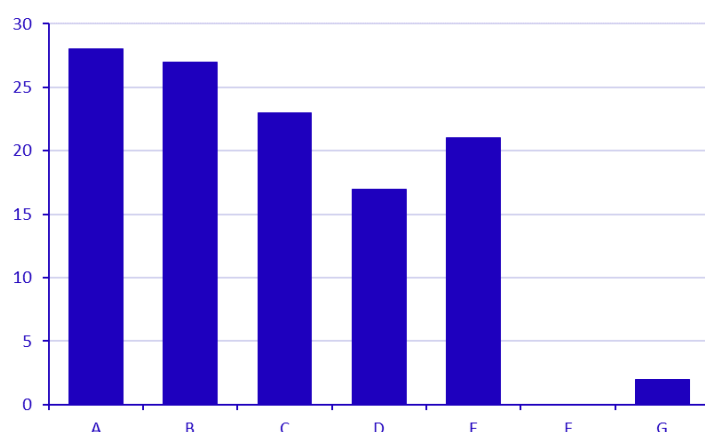


Figure 12. Need for training and capacity building (number of replies, multiple options possible).

A – Advanced scripting and programming in GIS for increased automation in data production

B – Web mapping tools and services, APIs, linked data etc.

C – Interoperability and standards

D – Use of Earth Observation data

E – Advanced spatial analysis

F – No particular training needed

G – Other

The conclusion from the responses to this question is that there is a high demand for training across a number of themes. None of the options provided in the survey has been scored low.

However, among the options provided the most popular options are A-C, comprising advanced scripting and programming, web mapping tools and services, including APIs and linked open data, and interoperability and standards.

As for question 3.1, there is none or little systematic difference between countries with a long history in the ESS and new EU Member States and candidate countries in terms of need for training.

3. Conclusions

The replies from the responding countries give the impression that progress has been made in the field of statistical-geospatial integration over the last five years. The data access situation seems to have improved in a number of countries. In 2015, according to the GEOSTAT 2 survey, lack of data /restricted access to data was a recurrent theme in multiple questions. The goal that the entire ESS (and candidate countries) should be able to adopt a point-based foundation for statistics no longer seem as a distant vision.

However, despite an ever-increasing access to data, lack of interoperability remains an issue. Many countries report that lack of interoperability and conformity of data poses obstacles to a successful data integration. Also, interoperability and standards (probably in a broad sense, including both semantic and technical aspects), seems to be an issue that many countries are struggling with.

Another positive indicator is that many countries rate the sustainability of their data management environments for geocoding and integration of statistical and geospatial data quite high. The vast majority of countries find their data management environments well-structured and well-documented, supporting systematic geocoding and automation in production of geospatial statistics. Remaining issues concern improvement rather than total remake of systems.

In terms of input for the GEOSTAT 4 project, the responses confirm several of the considerations already made in the start-up phase of the project. The benefit of including a quality component in the project has been clearly demonstrated by the lack of national frameworks for quality assurance for geospatial statistics reported by the countries. Another aspect is the support for the approach of using national good practice cases as to inspire and to facilitate knowledge exchange between countries.

The questions in the survey were not designed to obtain very detailed information. The aim was rather to obtain a quick and brief overview without putting a lot of burden on the respondents. The initial idea was to drill deeper in some of the most interesting findings from the survey in the GISCO Working Group meeting. Unfortunately, this was not possible due to the cancellation of the meeting.

4. References

GEOSTAT 2 (2016). Spatialising Statistics in the ESS. Results from the 2015 GEOSTAT 2 survey on geocoding practices in European NSIs. EFGS and Eurostat.

(<https://www.efgs.info/wp-content/uploads/2017/03/GEOSTAT2-Spatialising-statistics-in-the-ESS-2015.pdf>)

Rademacher, W (2012). Population Grid Statistics from Hybrid Sources. In: REALITY, DATA AND SPACE INTERNATIONAL JOURNAL OF STATISTICS AND GEOGRAPHY. Vol. 3 Núm. 3 septiembre-diciembre 2012.