

Results from the GEOSTAT 4 /GISCO survey – what can we learn for the project?

Jerker Moström



Aim

- Go through the result and assess if and how the result affects the priorities in the project
- What kind of conclusions can we make for different WPs and actions?

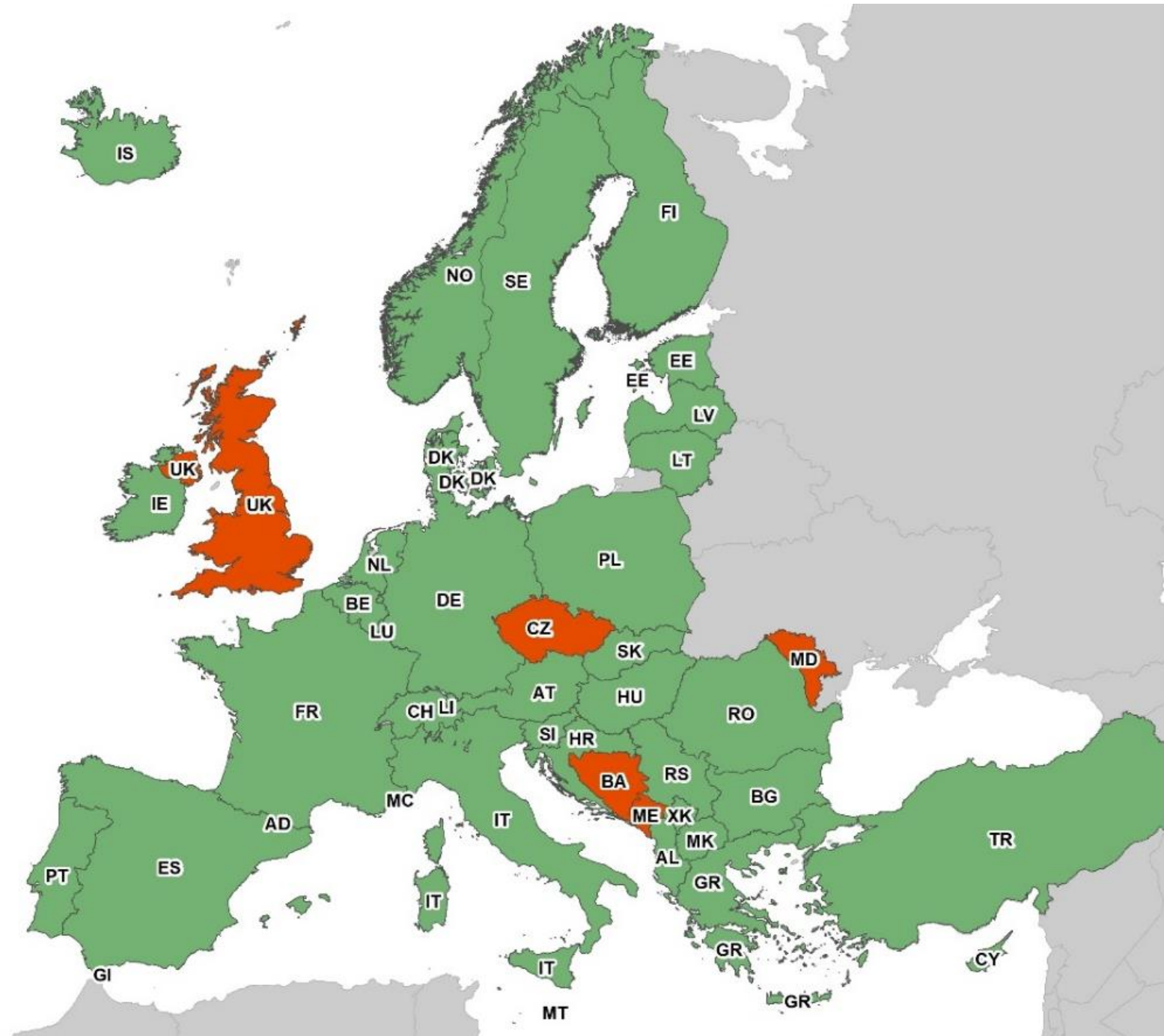
Target countries

Response

No

Yes

Responses 2020



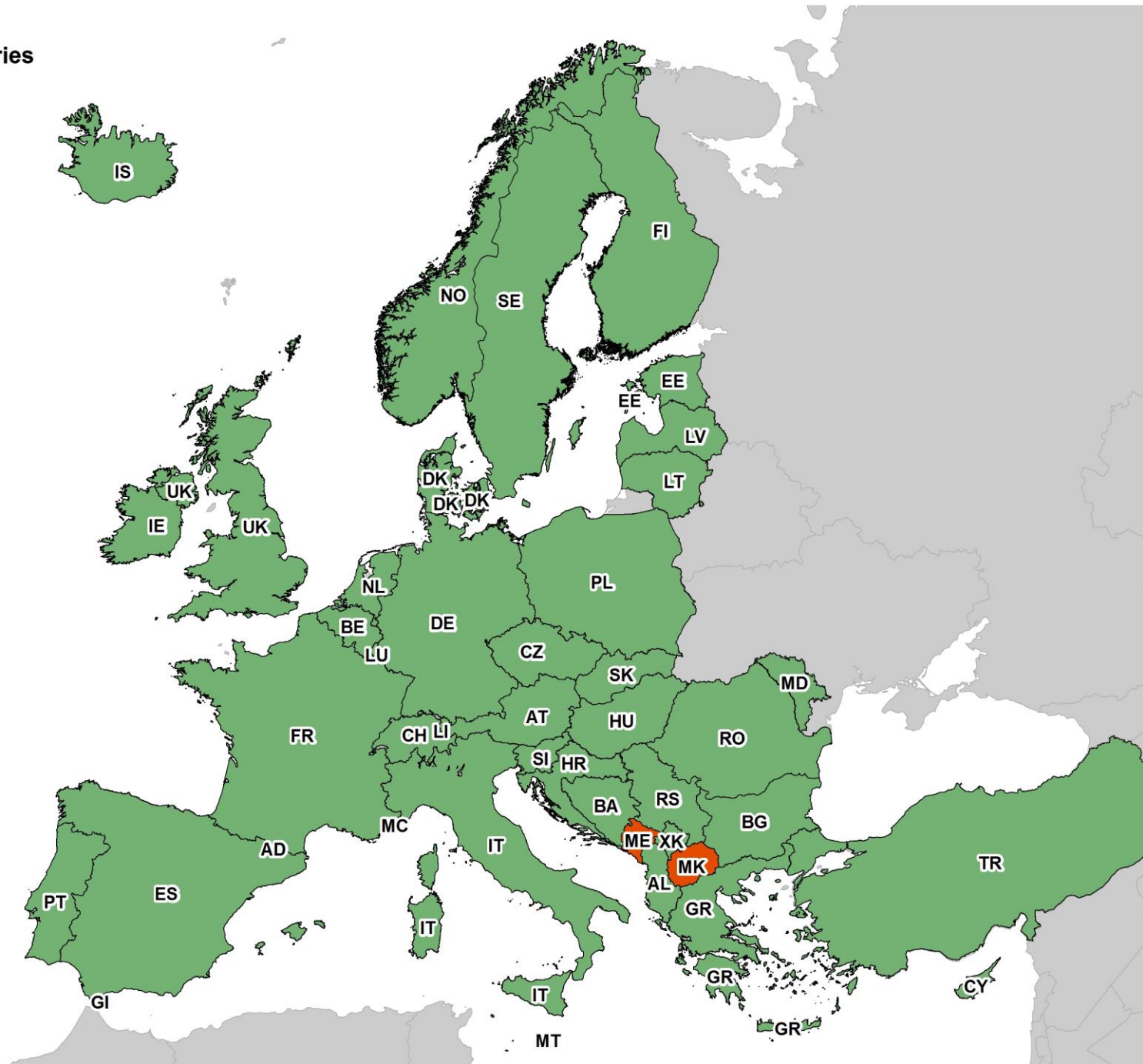
Target countries

Response

No

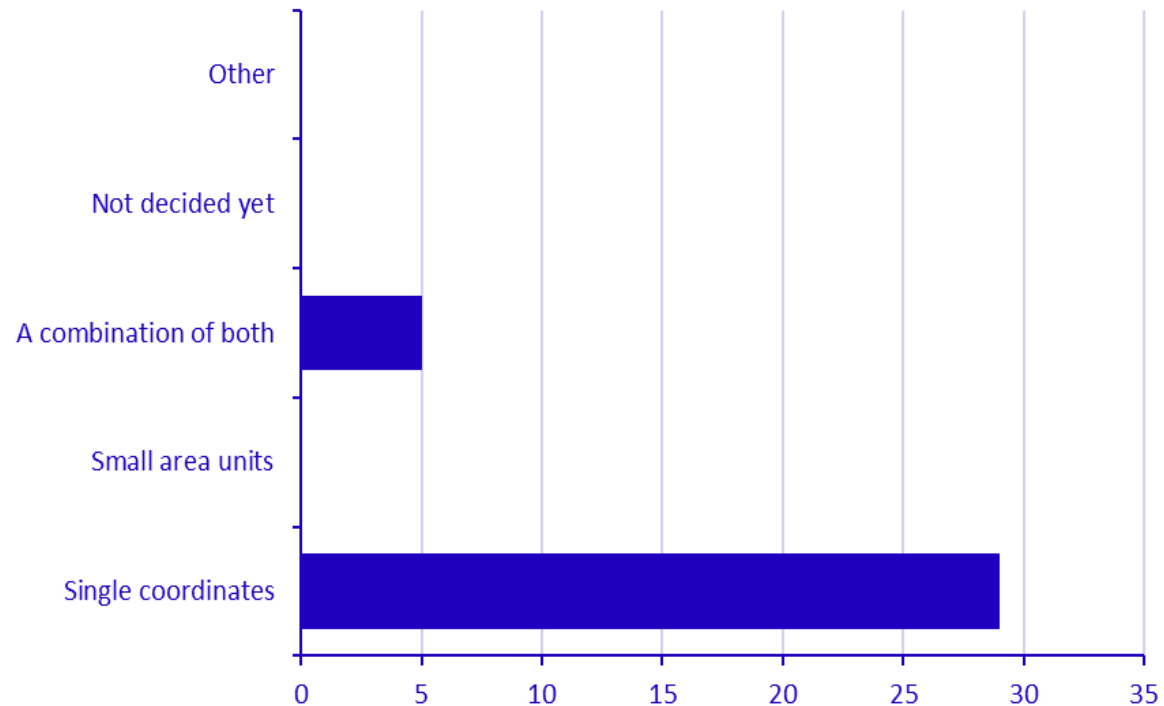
Yes

Responses 2015



Question 2.1

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

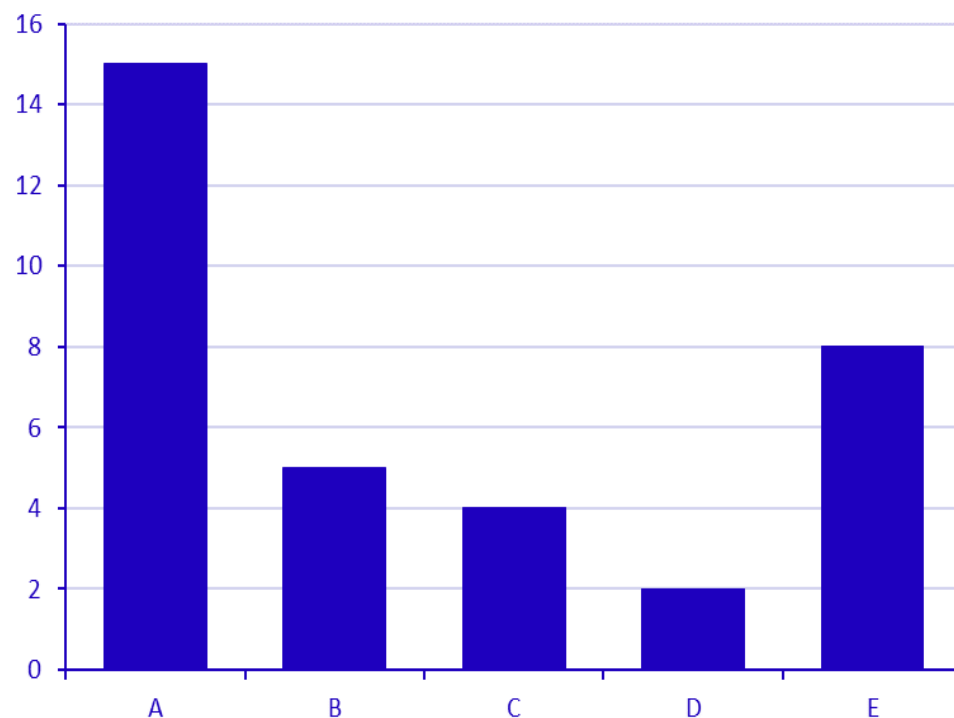


Question 2.1 - Conclusions

- Great progress since 2010 and 2015
- Assumption: *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
- Point-based geocoding starting to establish as a European “standard”

Question 2.2

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



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.

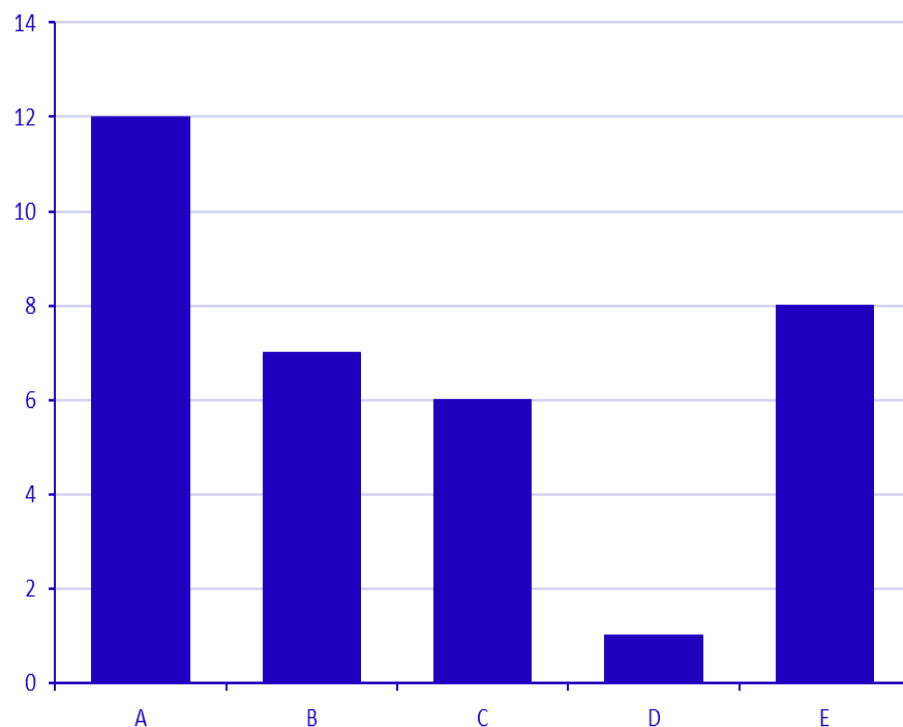
E - Other

Question 2.2 - Conclusions

- The majority of the respondents (15 countries) report option A, which can be considered the highest level of sustainability
- If A put together with option B, which is the second most sustainable level, roughly 50 % of the target countries have a very high, or relatively high, level of sustainability
- In total 6 countries have reported C or D which indicates a low level of sustainability due to lack of harmonised national data or problems with coverage or quality

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?



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

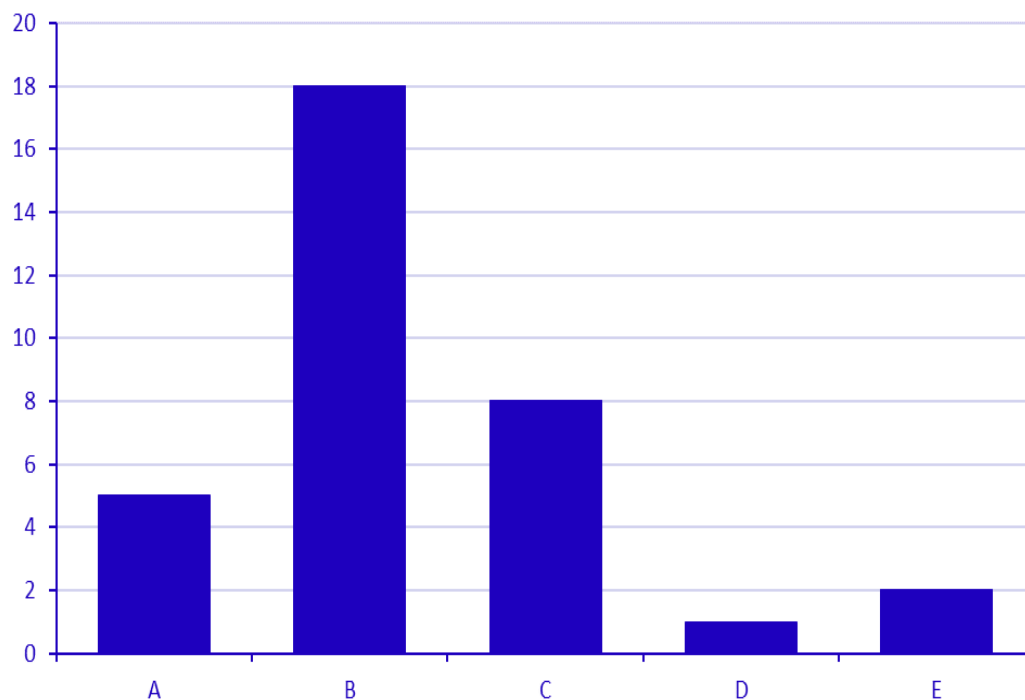
E - Other

Question 2.3 - Conclusions

- Complex and diverse situation in countries
- 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
- More or less similar situation compared to 2015

Question 2.4

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



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.

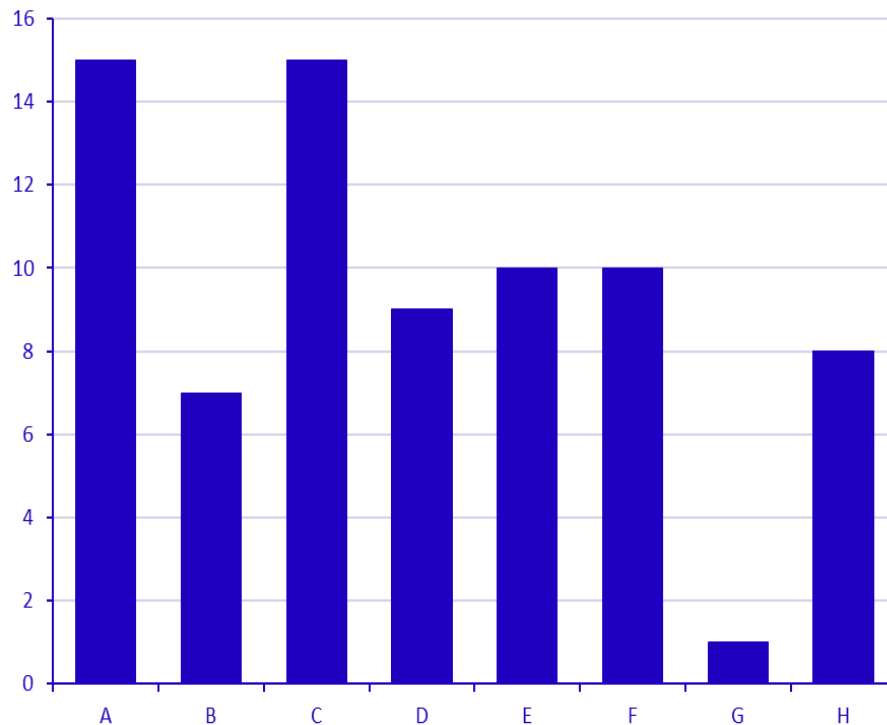
E – Other

Question 2.4 - Conclusions

- The majority of countries have reported option B, indicating that they have a well-structured and well-documented data management environment
- Room for some improvement and modernization
- In total eight countries have explicitly replied that they do not have a sustainable data management environment
- 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



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!

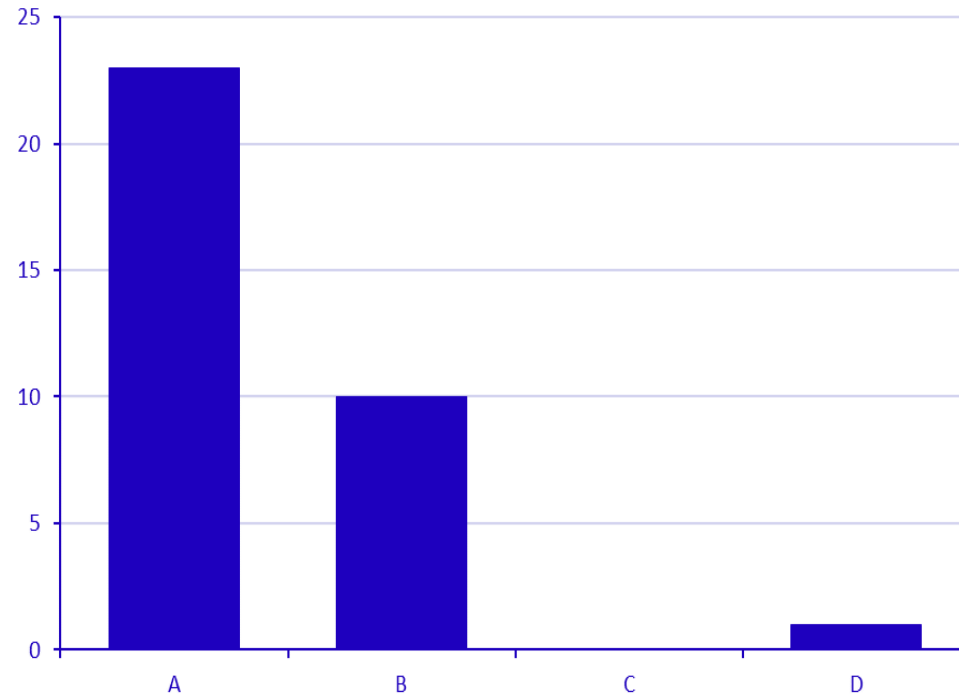
H - Other

Question 2.5 - Conclusions

- The most significant obstacles for statistical geospatial integration is 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 related.
- Very few countries describe a situation without any obstacles, though a quite large number of countries indicate that they can see minor challenges but no major obstacles.

Question 2.6

Use of administrative data sources for geospatial statistics



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.

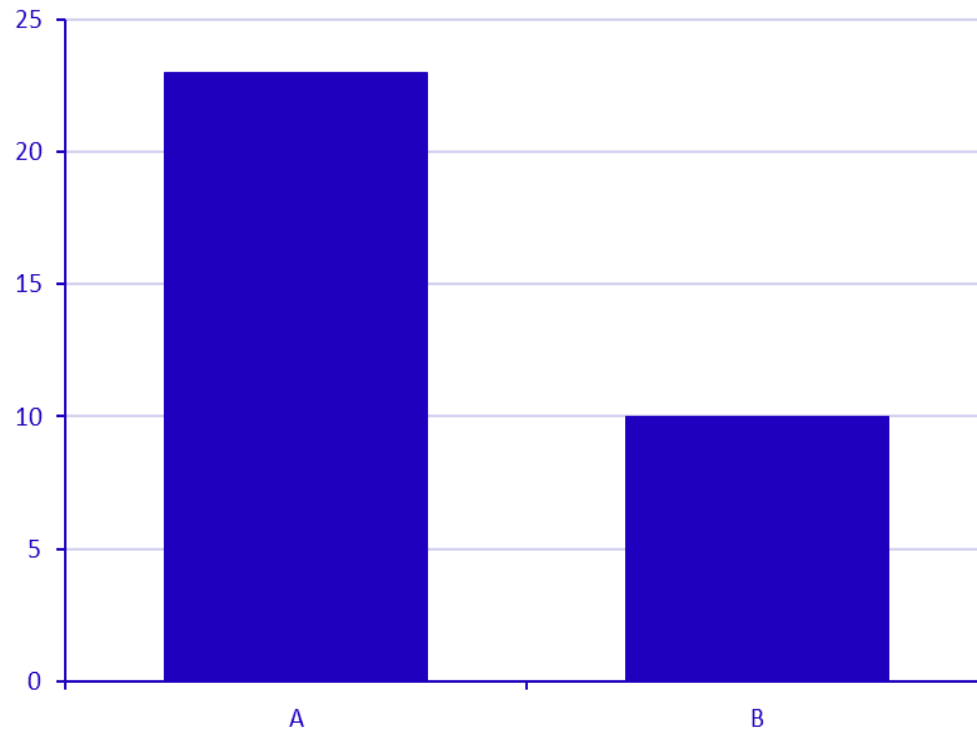
D - Other

Question 2.6 - Conclusions

- The overwhelming majority of the responding countries have already implemented administrative data sources in regular production of geospatial statistical products
- All responding countries seem to foresee that administrative data will come into use in the near future
- No deeper knowledge on what kind of administrative data are used, nor do we know how or for what purpose data is used.

Question 2.7

Quality aspects of geospatial statistics



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

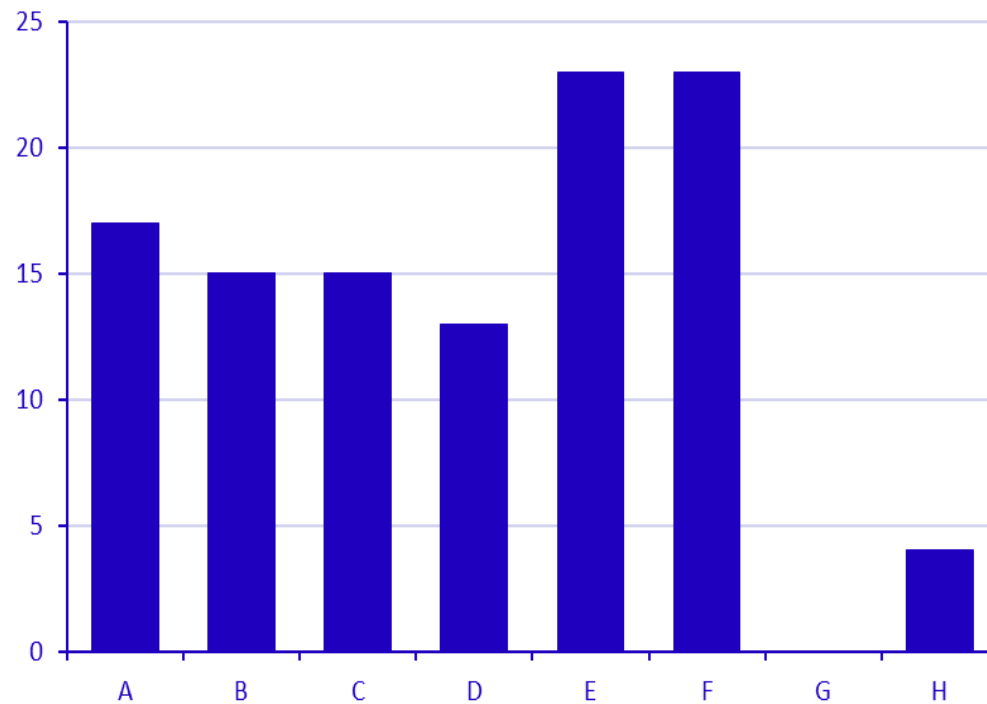
B - Yes, our quality framework cover the geospatial aspect of the statistical production process and/or we have quality indicators in place to assess the quality

Question 2.7 - Conclusions

- First time we have obtained information about quality frameworks (geospatial)
- 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



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

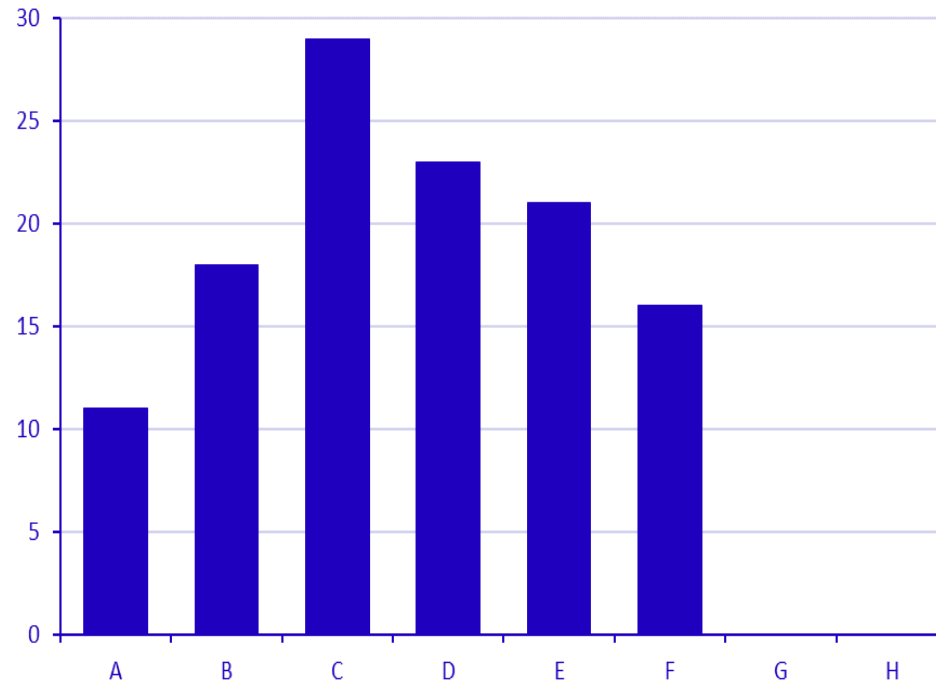


Question 3.1 - Conclusions

- Significant need for guidance around interoperability issues and standards (E) and web services and tools for data dissemination (F)
- F is also closely related to demand for guidance concerning dissemination of INSPIRE services
- Need for guidance is rather evenly expressed 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



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

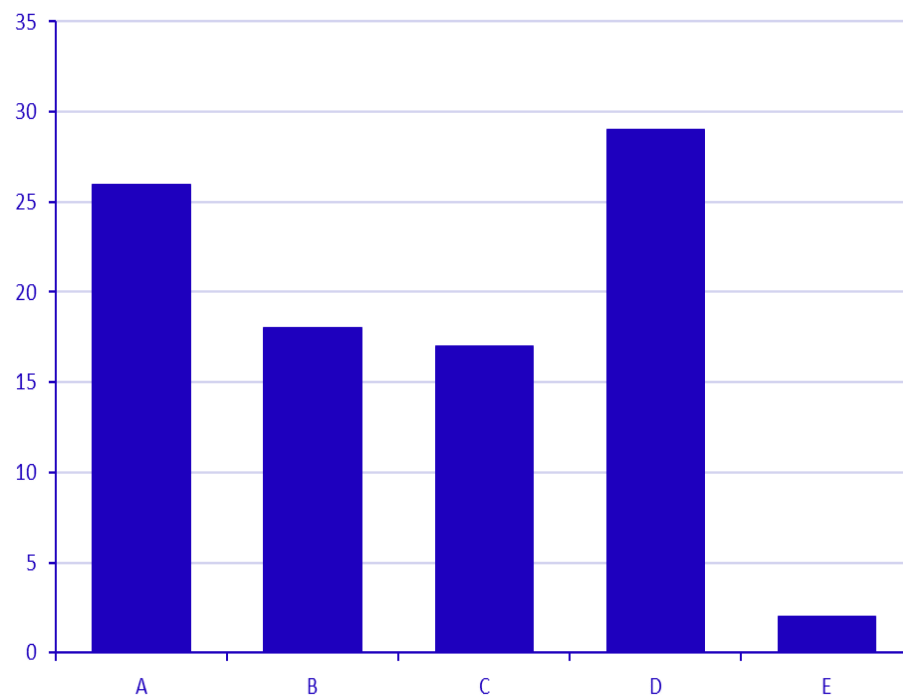
H - Other

Question 3.2 - Conclusions

- Strong support particularly 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 asked for (D and E)
- 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



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

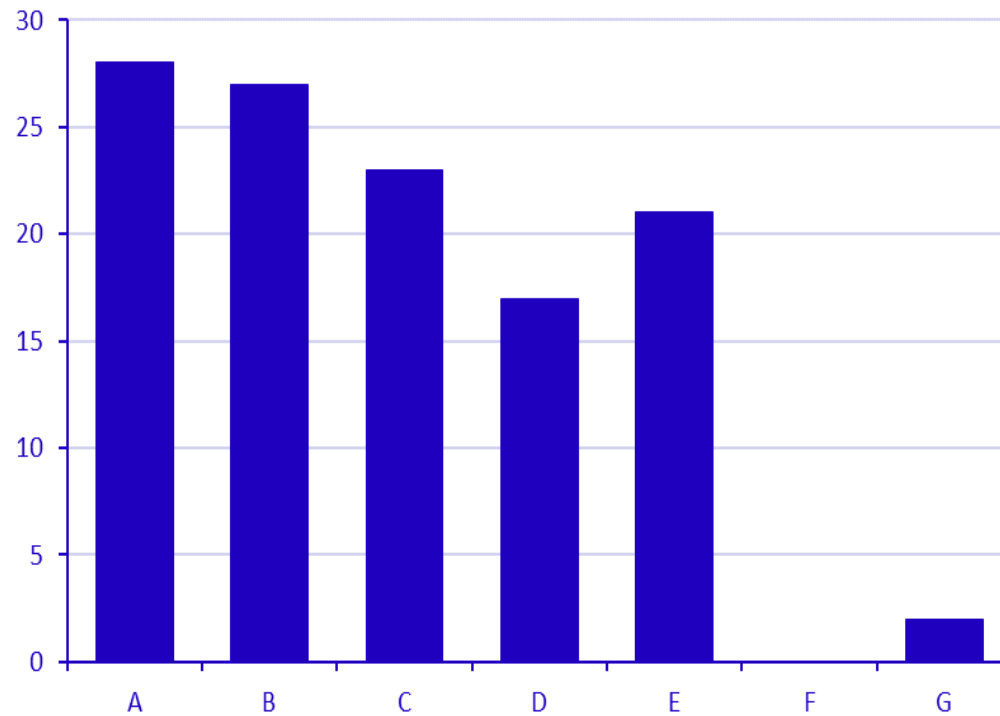
E – Other

Question 3.3 - Conclusions

- 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



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

Question 3.4 - Conclusions

- high demand for training across a number of themes. There is not really any of the options provided that has been scored low
- A-C, comprising advanced scripting and programming, web mapping tools and services, including APIs and linked open data, and interoperability and standards
- 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 training

General conclusions

- Progress has been made in the field of statistical-geospatial integration over the last five years
- The data access situation seems to have improved a bit in a number of countries
- Lack of interoperability remains an issue
- Many countries rate the sustainability of their data management environments for geocoding and integration of statistical and geospatial data quite high
- In general terms, the GEOSTAT 4 seems to be positioned with relevance

Questions

- Considering that the biggest threats to data integration is identified as lack of, incomplete or poorly maintained national standardised data along with poor semantic or technical interoperability between different data sources or cross data domains – what can be done within the GEOSTAT 4 project to mitigate this? What is within the scope of the project and what is not?
- Obviously the need for guidance is rated high. The most asked for issues of guidance are interoperability issues and standards along with web services and tools for data dissemination? How can we respond to this kind of requests?
- Also the need for training is strongly expressed. Especially training related to advanced scripting and programming in GIS for increased automation in data production, web mapping tools and services, APIs, linked data etc and interoperability and standards. What paths can you see for the project in order to meet this demand?

