



Spatialising Statistics in the ESS

Results from the 2015 GEOSTAT 2 survey on geocoding practices in European NSIs

Jerker Moström & Amelia Wardzińska-Sharif

Acknowledgments

On behalf of the GEOSTAT 2 project and the European Forum for Geography & Statistics (EFGS) we would like to express our sincerest gratitude to those of you who kindly contributed to the GEOSTAT 2 survey. We know that some questions were difficult to answer, and that the timing of the survey, in the middle of the summer, wasn't really perfect. Some of you also patiently faced additional questions and requests for clarifications.

With your committed support and valuable input, we have now achieved a high level of response, allowing us to make a comprehensive overview of the situation in Europe which we believe is of value to the community much bigger than the GEOSTAT 2 project team.

Jerker Moström & Amelia Wardzińska-Sharif

12.01.2016

Table of Content

1 Introduction.....	2
1.1 Background.....	2
1.2 The scope and aim of the survey.....	2
1.3 The content and structure of the survey.....	2
2 Results of the survey	4
2.1 The existence of and possible access to data used to spatialise statistics.....	4
2.2 A point-based foundation for statistics.....	7
2.3 The state of geocoding practices.....	12
2.4 The responsibility for and maintenance of geospatial information.....	17
2.5 The conditions for and obstacles to successful geocoding practice	18
2.6 The current use of and needs for spatial information in statistical production	21
2.7 The role and benefits of geospatial information in statistical production.....	25
2.8 The geocoding capacity of Census 2021.....	26
3 Discussion and conclusions	32
4 List of terms.....	35
Appendix 1- Questionnaire	36

1 Introduction

1.1 Background

GEOSTAT 2 is a two year ESSnet¹ grant project, launched in February 2015. The project is coordinated by Statistics Sweden and has participants from the National Statistical Institutes (NSIs) of Austria, Finland, France, Poland, Portugal and Norway. The overall aim of the project is to foster a better integration of statistics and geospatial information so that the statistical community could provide more qualified descriptions and analyses of the society, the economy and the environment.

One of the main project objectives is to propose a generic model for a national, point-based geospatial reference framework for statistics, building on nationwide addresses, and building and/or dwelling registers. The reference framework should be appropriate for statistics in the widest possible sense, which explicitly includes its use for enumerations and sampling for surveys. The model must build on, and take into consideration, the diverse situation in Member States regarding access to data, responsibilities, organisational setup and resources.

1.2 The scope and aim of the survey

To gain a better overview of all these aspects, and in order for the project to make recommendations relevant and useful for the current and future ESS Member States, the project principals agreed on the necessity to undertake an inventory of existing national and sub-national spatial reference frameworks in Member States, candidate countries and potential candidates, including statistics and administrative data. The inventory was carried out as a web-based questionnaire addressing National Statistical Institutes (NSIs). However, as the questionnaire contained a set of questions regarding the existence of geospatial data, respondents were urged to liaise with their National Mapping and Cadastral Authority (NMCA), or other authorities providing spatial data in their country, before answering the questionnaire.

The survey was announced through the Eurostat GISCO and DIMESA groups in March and June 2015 respectively, and it was closed in October 2015. The results obtained in the survey are analysed and presented in this report.

At a late stage of the work, a request from Eurostat to support the current work of the Census 2021 Task Force with some background information resulted in an amendment to the survey. A small set of follow-up questions were addressed regarding the prospect of geocoding the prospective 2021 Census data to the geographical level of point coordinates. Only those countries indicating in the original survey that they had not developed a point-based infrastructure in place to date were involved in the follow-up questions.

1.3 The content and structure of the survey

The content of the survey can be roughly divided in the following sections, each addressing a particular set of topics relevant to the further work of the project.

- The existence of and possible access to registers or other datasets used to spatialise statistics
- A point-based foundation for statistics
- The state of geocoding practices
- The responsibility for and maintenance of geospatial information
- The conditions for and obstacles to a successful geocoding practice
- The current use of and needs for spatial information in statistical production (including quality aspects)

¹ <http://www.essnet-portal.eu/>

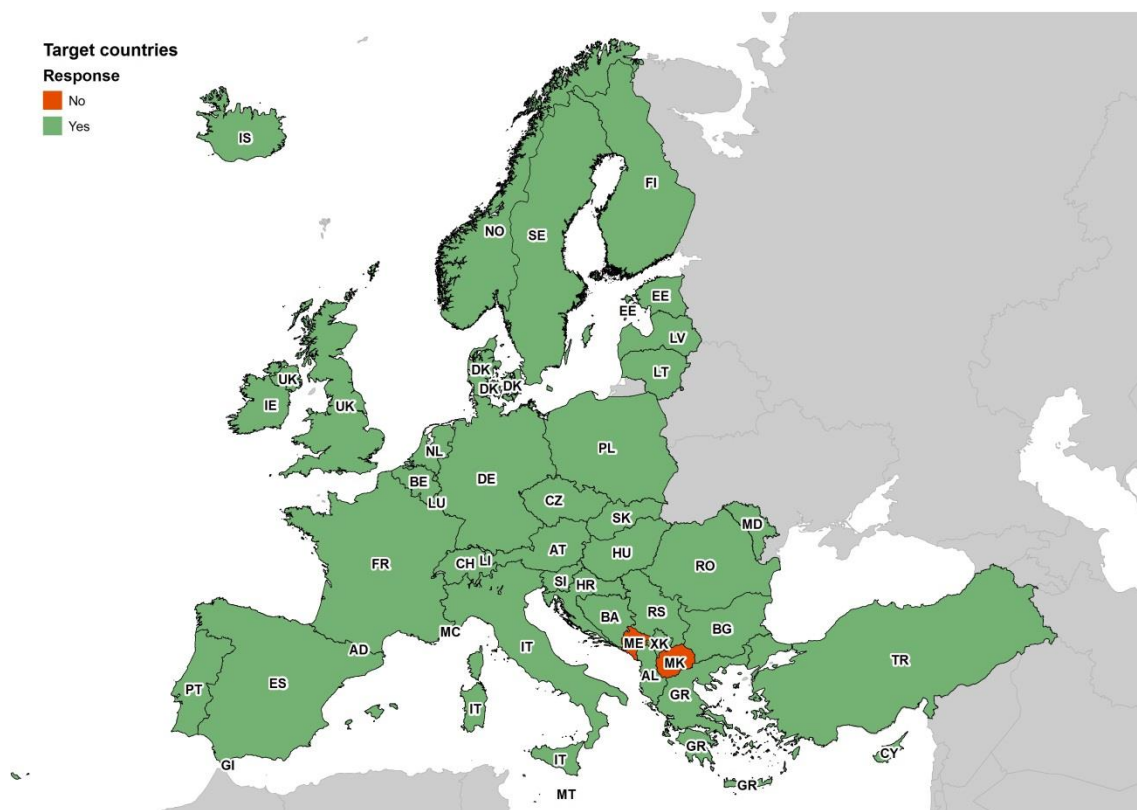
- The role and benefits of geospatial information in statistical production
- The geocoding capacity of Census 2021

The questions were prepared by the GEOSTAT 2 project group in collaboration with Eurostat. The web-based questionnaire was designed and implemented by Statistics Poland. In total, the survey comprised a set of 38 questions. The majority of the questions were based on predefined options (single-choice or multiple-choice questions) in order to avoid the ambiguous interpretation of the answers and to speed up the submission of responses. In most cases, however, the option to provide “other” input was available for respondents. Only 6 questions were entirely based on free-text input.

The survey target group comprised the ESS countries, together with EU-candidate countries and potential candidates. In total 40 countries were included. The overall response rate was very good; 37 countries responded, but some of them did not provide answers to all questions in the survey. The non-responding countries in the target group were Malta, Montenegro and The Former Yugoslav Republic of Macedonia.

Among the EU Member States (EU28), the response rate was 96 percent. Only Malta did not respond. Among the 32 ESS countries (EU28 + Iceland, Lichtenstein, Norway and Switzerland) the response rate was 97 percent. Also in this group Malta was the only non-responding country.

Map 1. Map of target countries and responding countries



Source: GEOSTAT 2 project/EFGS

The high response rate, in combination with the geographical distribution of the responding countries, provided good coverage of a variety of practices in geocoding and traditions regarding the use of registers and administrative data.

2 Results of the survey

2.1 The existence of and possible access to data used to spatialise statistics

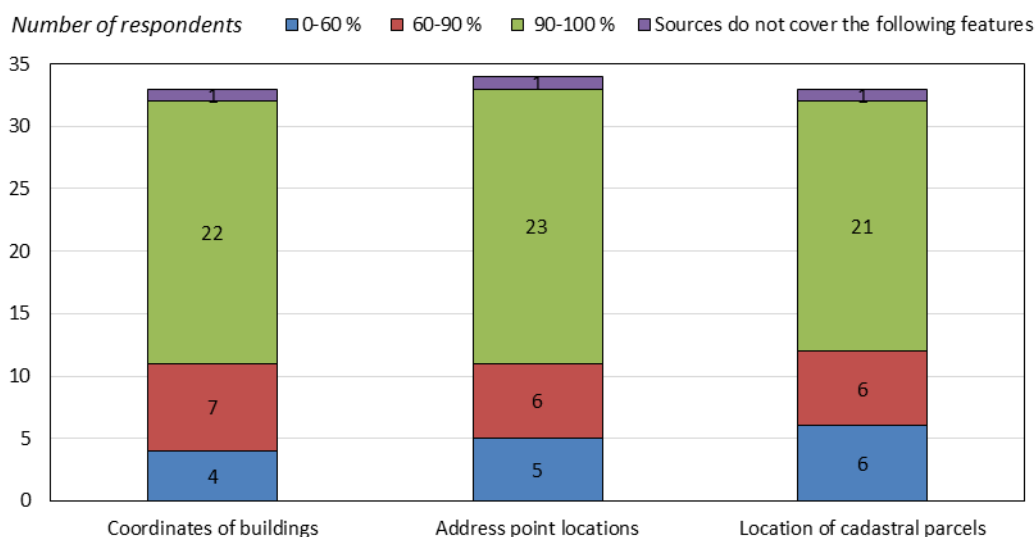
In this section, the questions aim to gain insights in the general ability to spatialise statistics among the Member States. When asking about “any sources of data for which it is possible to retrieve the spatial location of its units with coordinate accuracy” (Q6), the vast majority of the countries (97%) gave a positive answer. Only Greece reported that no such sources were at hand.

The option to specify “if it is possible to transform the existing spatial registers to acquire spatial data of point accuracy” (Q7) was given for those countries which provided a negative response to Q6. Greece gave a specification of the transformation of existing spatial registers:

“An effort was made in the context of providing grid-based data for GEOSTAT 1A & 1B using the EFGS/EUROSTAT Grid. This was performed under the EUROSTAT grant on “Merging Statistics with Geospatial Information in Member States” for the 2001 and 2011 Census data. There was a hybrid procedure combining population data at the census-block level, along with data at the settlement level (area data) that were rendered to the EUROSTAT Greek grid portion.”

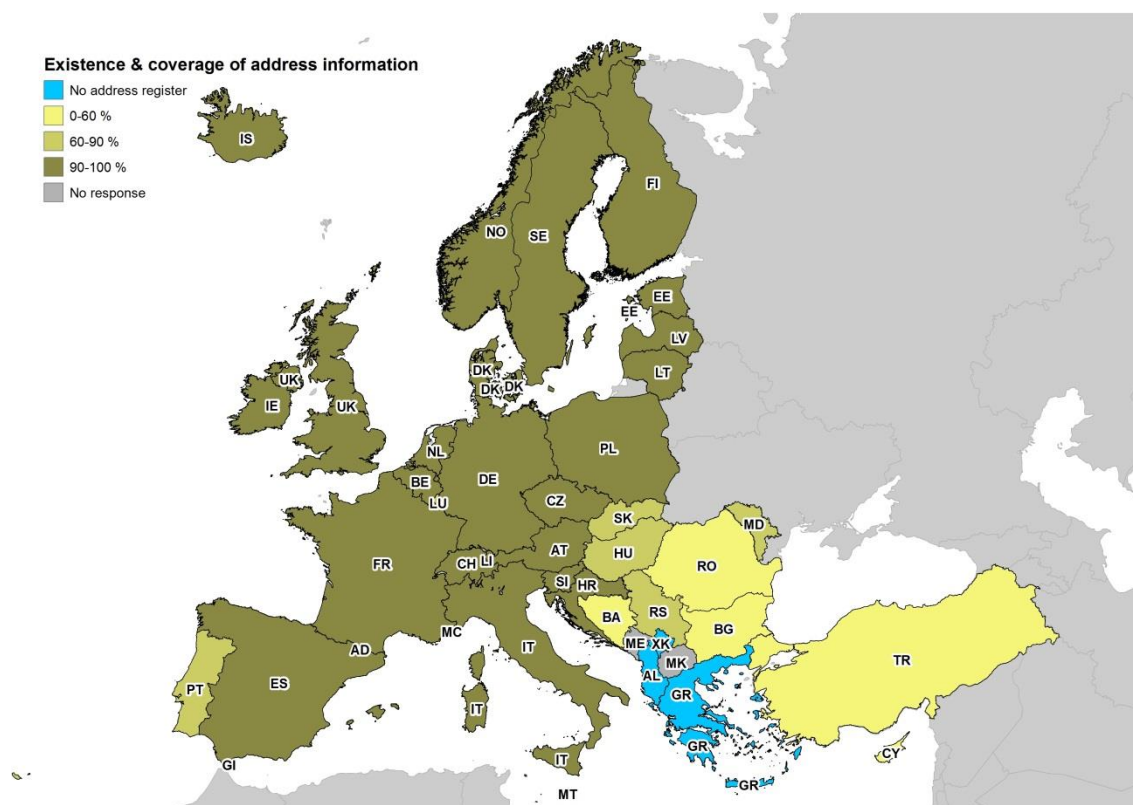
The countries which provided a negative response to Q6 were asked to further specify whether the data sources (referred to in Q6) encompassed the coordinates of buildings, address point locations and/or the location of cadastral parcels, and to what extent those data sources covered the territory of the country. The question allowed for multiple choices. In total 36 countries responded, of which 33 countries reported the use of the coordinates of buildings, 34 reported the use of address point locations, and 33 reported the use of the location of cadastral parcels. In total, 31 countries reported the use of all three data sources.

Results obtained for Q8. Do the sources of data mentioned in Q6 encompass the following features (the coordinates of buildings, address locations and/or location of cadastral parcels) and to what percentage do they cover the territory of your country (fixed intervals: 0-60%, 60-90% or 90-100%)?



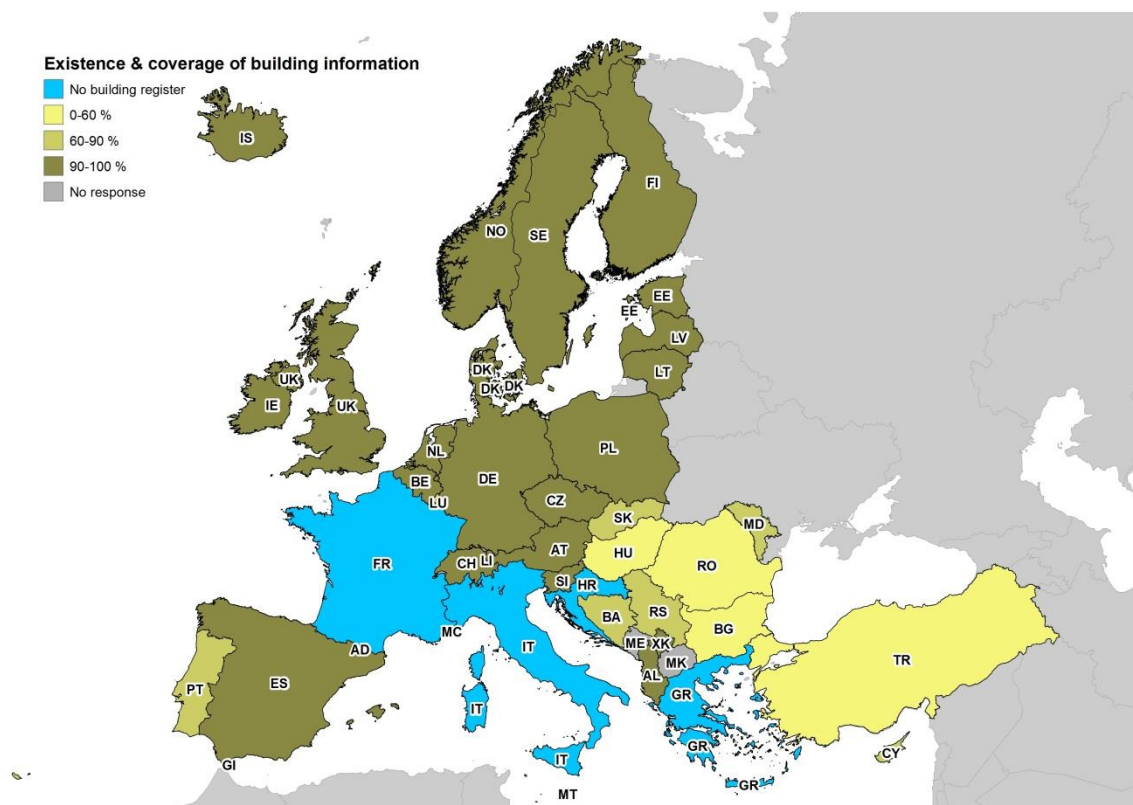
The above chart shows that the degree of coverage of the respective features (the coordinates of buildings, address point locations and the location of cadastral parcels) is very similar in many countries, the differences being insignificant.

Map 2. The existence and coverage of address information by country



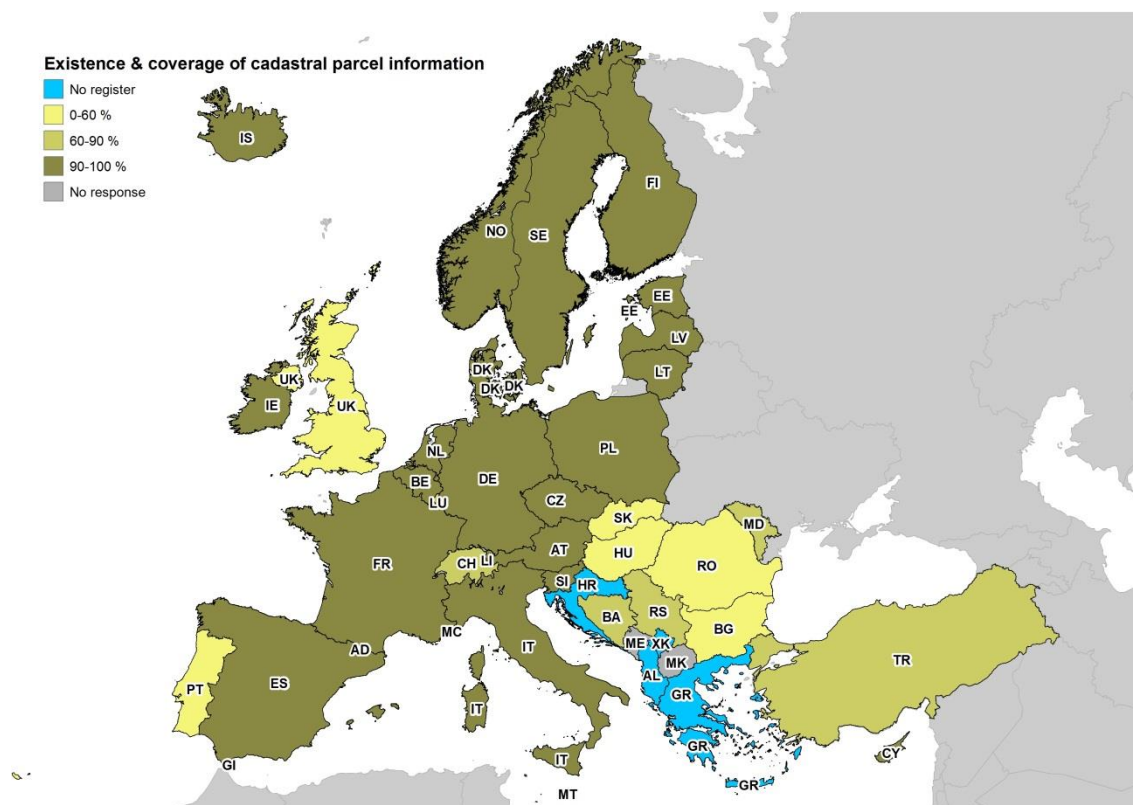
Source: GEOSTAT 2 project/EFGS

Map 3. The existence and coverage of building information by country



Source: GEOSTAT 2 project/EFGS

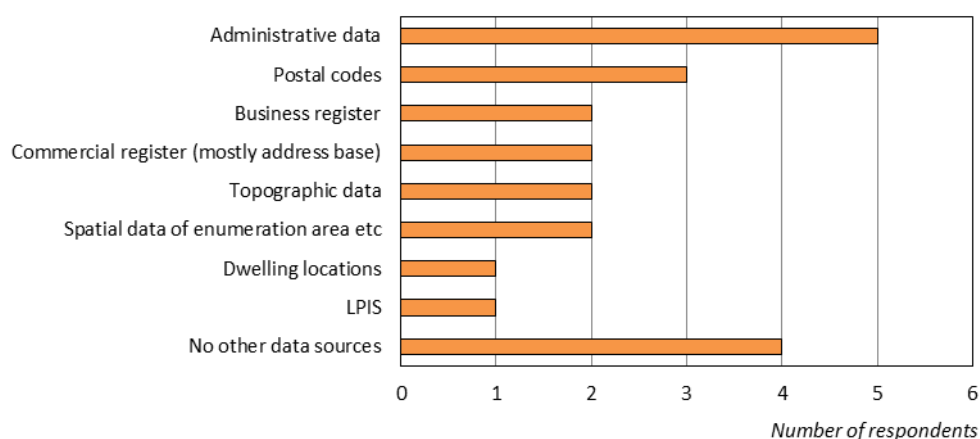
Map 4. The existence and coverage of cadastral parcel information by country



Source: GEOSTAT 2 project/EFGS

In Q9 respondents were given the opportunity to indicate whether they could think of other data sources (other than the coordinates of buildings, address locations and cadastral parcels) which could be useful for geocoding statistical information. 14 respondents did not indicate any sources, but the rest reported a variety of data categories. As there were no predefined options in this question, it is difficult to make precise aggregations of data sources, but some rough categories have been identified as illustrated in the bar chart below.

Results obtained for Q9. Are there any other sources which you think could be used/useful for geocoding statistical information?



Based on the responses, it can be inferred that:

- The vast majority of the countries in the ESS already have point-based location data in place with a reasonable degree of coverage. Most of the countries indicating a lower degree of coverage are candidate countries or possible candidates. It can be assumed that the main reason for the lower degree of coverage in those countries is that the location data frameworks are under construction and have not yet reached completion.
- Moreover, it seems as if the location data frameworks in the Member States usually include not only an address register, but also very likely a building register and/or a register of cadastral parcels.
- Even though there are very small differences between the different categories, address locations seem to be the most-mature category of location data, at least in terms of coverage.

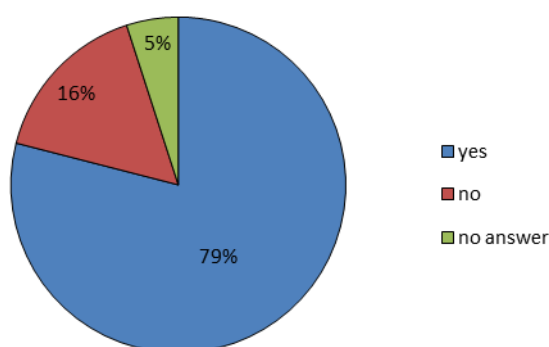
2.2 A point-based foundation for statistics

To verify the working hypothesis of the GEOSTAT 2 project, which is that coordinates of address locations and/or building points and coordinates of cadastral parcels *more or less* form a complete basis for a point-based production system, respondents were asked to agree or disagree with such a statement (Q10).

Roughly 80 percent of the countries agreed with this statement with no further comments. Respondents that did not agree with the statement were asked to specify additional data to add to the core data for a more-complete basis for point-based statistics (Q11). The following additional data were reported

- geo-referenced building register maintained by official statistics,
- precise location of agricultural holdings,
- road network with road names,
- mobile household (boats, caravans) locations,
- environmental reports and permits including coordinates, such as location for emissions.

Results obtained for Q10. According to your experience do the following two sets of point data represent the complete basis for point-based statistics: 1) coordinates of address points* and/or building points and 2) coordinates of cadastral parcels?



Q10 and Q11 were followed by a set of more detailed questions about the existence and accessibility of registers of population, dwellings and buildings. The aim of these questions was three fold;

- to learn more about the existence of administrative registers facilitating yearly or even more frequent updates of geospatial statistics (as opposed to survey or census data),
- to learn more about the issue of access to such registers, and finally,
- to learn more about the use of keys (standardised identifiers) linking records of statistical or administrative data to geographical location.

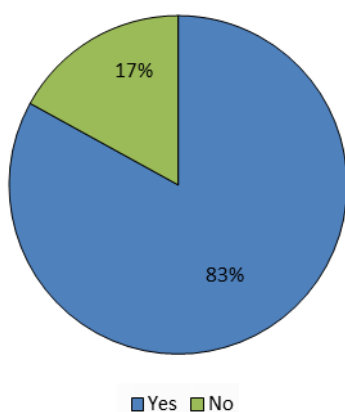
The results (Q12 A, B and C) indicate that approximately three out of four of the responding countries possess both population (individual persons) and dwelling registers. The question did not explicitly imply that these registers must actually be used to produce official population or housing statistics. Accordingly, the answers do not necessarily reflect that three out of four of the responding countries actually use these registers for the regular production of statistics.

Nearly 60 percent of the responding countries reported that they could link population to dwelling location by means of standardised identifiers. This result indicates that some of the countries that reported that they had population registers and geo-enabled dwelling registers were not able to connect population with dwellings.

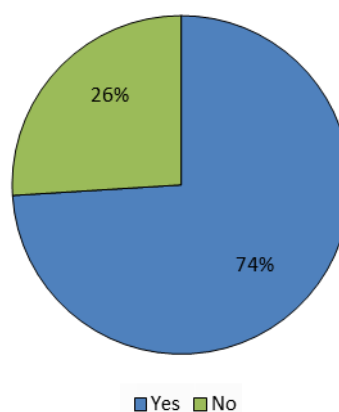
In Q13 the question about the existence of population and dwelling data was followed by a question about the accessibility and terms of access. Two thirds of the responding countries reported that they had access to these data without having to pay for it, or at least incurred very low expenses. Only one country indicated that access was on a payment basis and 4 countries had no access at all. The rest provided special comments, saying that they had access only to a building/dwelling register, whereas an address register was currently being developed, or that they had restricted access to population data, or could use only census data.

Results obtained for Q12. Population data and dwelling data

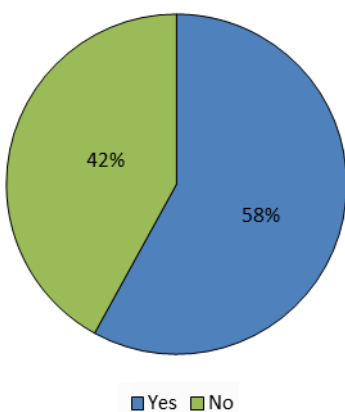
A. Does a register of population (individual persons) exist in your country?



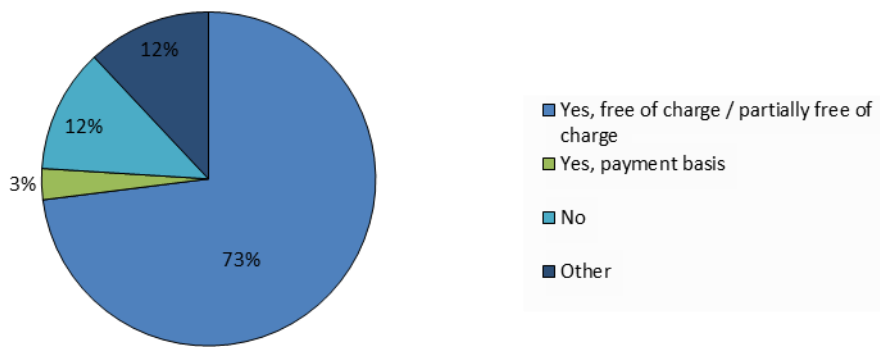
B. Are there registers from which you can obtain information on dwelling locations (in the form of address points)?



C. Can population be linked to dwelling location (geocoded) by means of a standardised identifier?



Results obtained for Q13. Does your NSI have, and on what terms, access to the data mentioned in Q12?



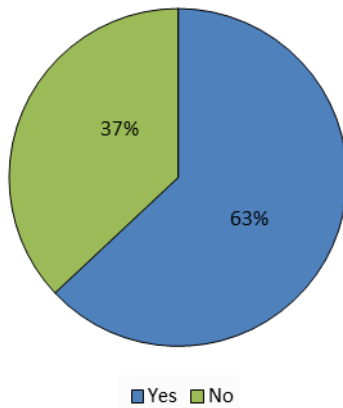
Q14 is equal in structure to Q13 though it focuses on the existence of and access to building data. Only around 60 percent reported the existence of building registers which was less than the 80 percent of countries which reported that they could obtain the locations of buildings from other sources. The most plausible interpretation is that they can use address information as proxy data to derive building locations.

Only two out of three countries can link building to location by means of standardised identifiers. This is not so surprising, considering that one out of three countries does not have a building register at all. However, among the countries which have a building register, 3 countries reported that they could link building to location by means of identifiers.

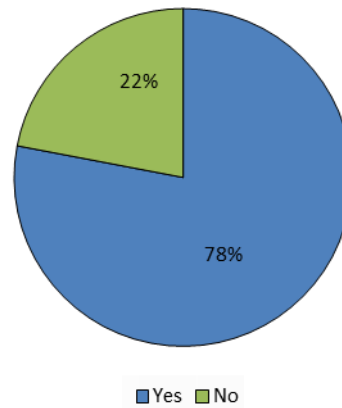
Regarding the terms of access, building information reporting has a similar profile to population and dwelling data. Around 70 percent of the responding countries can access building data free of charge or at low cost, which is similar to population and dwelling information. However, 3 countries have access on a payment basis. There are also special conditions reported, such as access only to new buildings, the fact that a building register is under construction, or that access accounts for the locations of dwellings/functional units inside the buildings rather than the building itself.

Results obtained for Q14. Building data

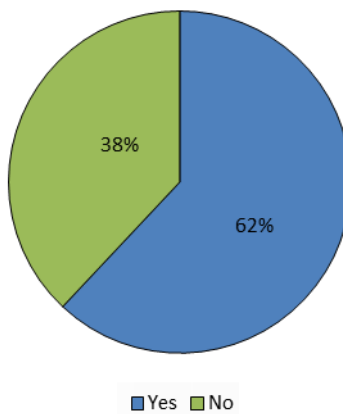
A. Does a register of buildings exist in your country?



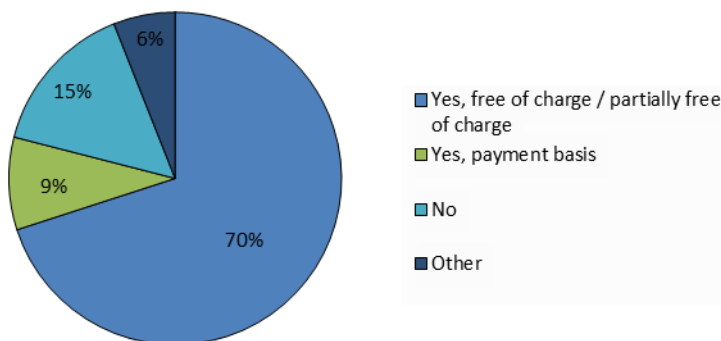
B. Are there registers from which you can obtain information on building locations?



C. Can buildings be linked to locations (geocoded) by means of a standardised identifier?



Results obtained for Q15. Does your NSI have, and on what terms, access to data mentioned in Q14?



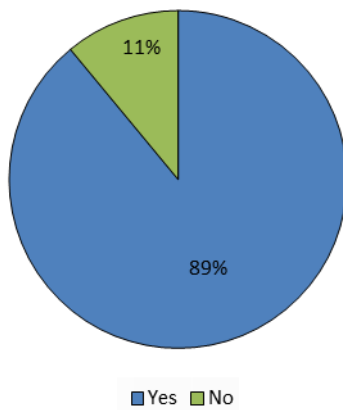
The last battery (Q16 and Q17) of detailed questions about the existence of and access to data sources, concern cadastral registers. The majority of the responding countries, almost 90 percent, have a record of individual cadastral parcels, but a slightly lower share of countries reported that they could obtain the actual locations of the individual parcels.

Roughly three-fourths of the countries can link cadastral information to location by means of standardised identifiers. As revealed by the answers, cadastral information seems to be the least accessible of the location data sources investigated. Roughly every second country reported that cadastral information could be obtained free of charge or at low cost. Around 10 percent could obtain data on a payment basis and nearly 25 percent of respondents indicated that the information was not available for them at all.

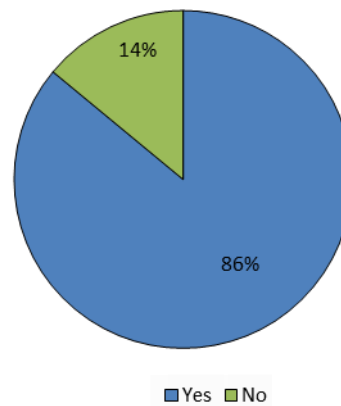
Of the types of location data that was investigated, cadastral data seems to be the least utilised category for geocoding purposes. Several countries commented on the fact that cadastral parcel registers were not needed, since they used other sources (e.g. buildings). This might also partly explain the fact that many NSIs do not have access to cadastral information. They simply don't need it. There are also countries that reported that special agreements were needed to access the information.

Results obtained for Q16. Cadastral parcels

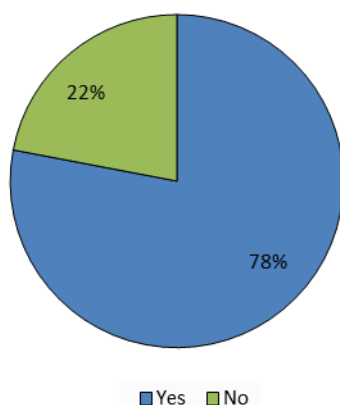
A. Does a record of individual cadastral parcels exist in your country?



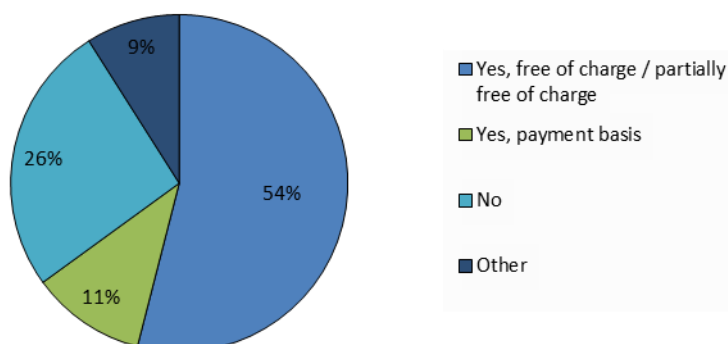
B. Are there any registers from which you can obtain information on cadastral parcel locations?



C. Can cadastral parcels be linked to locations (geocoded) by means of a standardised identifier?



Results obtained for Q17. Does your NSI have, and on what terms, access to data mentioned in Q16?



Based on the responses, the following main conclusions can be drawn:

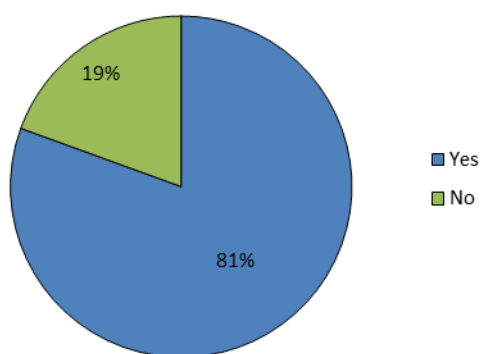
- Despite the fact that some of respondents indicated a need for additional data sources to make up a complete basis for point-based statistics, there is strong support for the working hypothesis that the core set of location data on which a point-based production model needs to be based is address and/or building information.
- Indeed the comments reflect that there is information that cannot be geocoded by means of building, address or even cadastral parcel coordinates; yet it is obvious that the majority of the information can. Nevertheless, the GEOSTAT 2 project needs to take into consideration the cases mentioned by respondents.
- Some of the additional data sources mentioned can be considered as complementary to the core set of location data, e.g. aiming to improve the geocoding of the population where address information fails (such as mobile households, road network etc.), whereas other datasets provide the ability to geocode information with weak or non-existing links to core-set location data (such as the location of agricultural holdings and environmental reports and permits).

2.3 The state of geocoding practices

For this section, the questions were designed to obtain a better understanding of the terms and conditions for geocoding information in the Member States. The aim was to find out to what degree geocoding is actually being conducted, and, if geocoding is not being conducted, what are the reasons behind this? Who is responsible for the geocoding process and what kind of information is being geocoded?

In this context, *geocoding* of statistics should be understood as the process of *geospatially enabling* statistical unit record data (i.e. data relating to individual persons, households, dwellings, businesses or buildings) by assigning a location descriptor (x, y, z coordinates, address, spatial ID, spatial feature). The geocodes (location coordinates, address IDs, or geographic areas codes), obtained from this process can be stored directly on the statistical-unit record or linked in some way to the record. The term *data pairing* is also used in the survey, referring to the process of merging data from two different sources - in this case linking data from different registers with its spatial reference (x, y coordinates).

The survey shows that more than 80 percent of the responding countries conducted some geocoding of statistical data. Only 7 countries reported that they had not conducted any geocoding at all. One country did not provide an answer to this question.

Results obtained for Q18. Has any statistical data geocoding process been conducted in your country?

Note: 1 of the countries participating in the survey did not provide an answer to this specific question.

Countries which answered “No” were asked to comment on the main obstacles preventing them from introducing procedures for geocoding statistical information or administrative data (Q19). Multiple choices were possible.

The main reasons, according to the countries which have not conducted geocoding, are the lack of resources and knowledge. The second most prominent reason is the lack of geospatial data or the fact that geospatial data have poor quality. A few countries have too many legal restrictions or bureaucratic reasons (for example, public institutions do not cooperate well), or indicate the lack of identifiers to connect data with geographical locations. A few countries also reported problems with statistical information or administrative data which were not collected and stored in a way that made geocoding possible or meaningful. There are two countries in which the absence of legal support for spatial statistics prevents them from conducting geocoding. The lack of quality in data location was indicated only by one country, and so were high costs of geospatial data. The latter was a bit surprising, as high costs were expected to pose a more significant problem.

It is also interesting to note that some of the options given in the questionnaire were not reported as a problem at all. No country reported “No domestic demand for spatial statistics” or “the lack of a uniform reference system between different administrative data sources”.

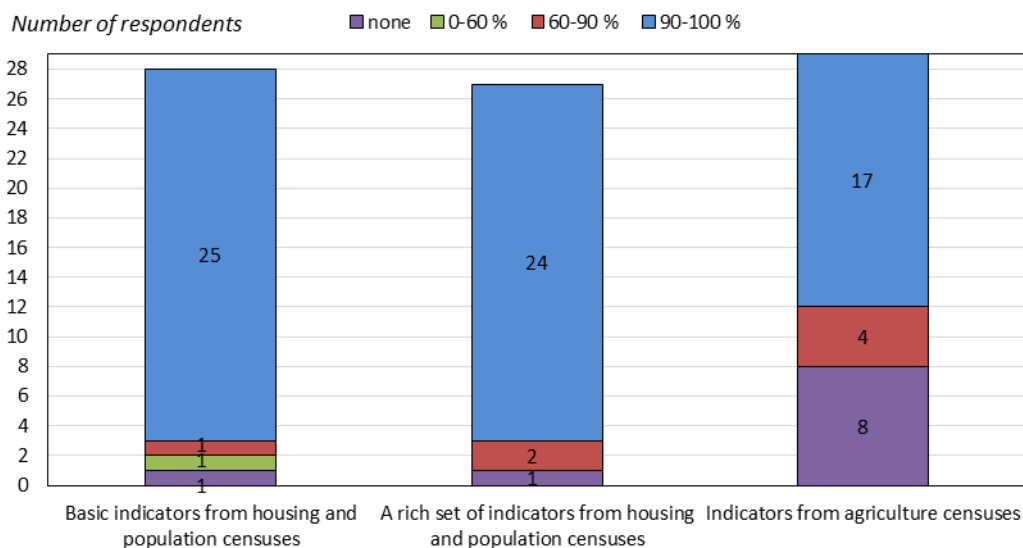
Q19 was followed by an enquiry into the kind of information which was geocoded (Q20). The complete phrasing of the question was as follows: “Please provide the scope of data that were geocoded and the percentage of the successful combination on the national level”. Three predefined options were given, along with a fixed interval for coverage.

The responses reveal that 25 countries have geocoded basic indicators from housing and population censuses (total population, number of dwellings etc.) to the level of 90-100% coverage. Two countries reported that they had only partially been able to geocode basic indicators, and only one country reported not having been able to geocode basic indicators from housing and population censuses at all. A number of countries did not provide an answer to this question.

Furthermore, 24 countries have been able to geocode a wide set of indicators from housing and population censuses (e.g. building use/type, age/sex, nationality, educational attainment of population etc.) to the level of 90-100% coverage. Two countries reported that they had only partially been able to geocode basic indicators and only 3 countries reported *not* having been able to geocode a rich set of indicators from housing and population censuses. Another 7 countries did not respond.

Around 20 countries have geocoded indicators from agriculture censuses with total or partial coverage. The remaining countries did not respond or reported that they had not been able to geocode any indicators from agriculture censuses.

Results obtained for Q20. Please provide the scope of data that were geocoded and the percentage of the successful combination on the national level (territory)



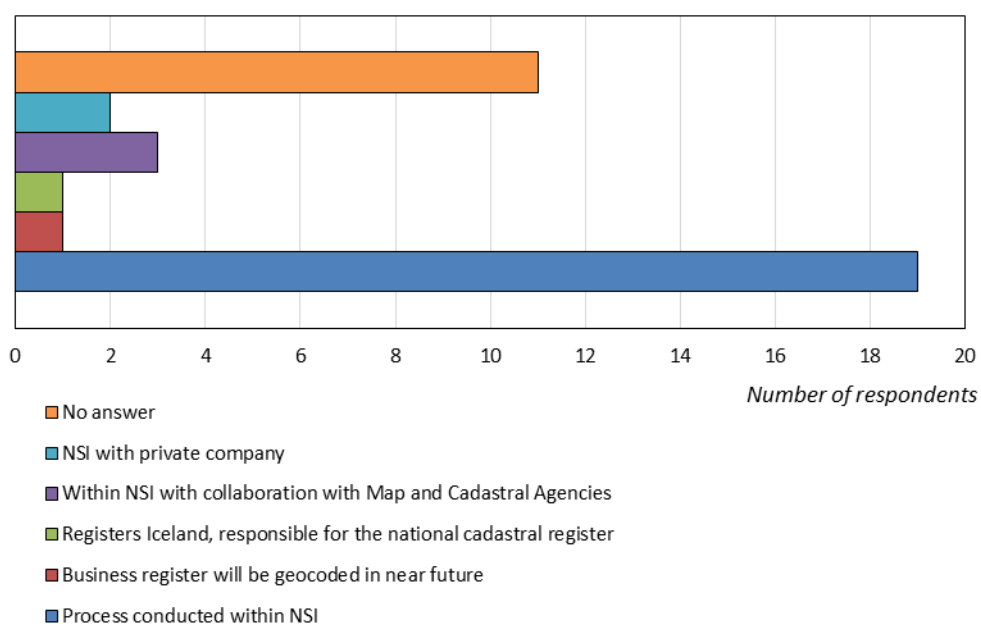
Q20 related only to a limited set of indicators from population and agriculture censuses. Following Q20 was an open question (Q21) concerning the scope of other registers or administrative data (business register, population register, tax files etc.) that were geocoded and to what percentage a successful combination could be achieved at national scale. This question allowed for free text input.

The most significant input to Q21 concerned business registers. Roughly 40 percent of the countries mentioned business registers explicitly. Some countries did not mention it explicitly, but it could be assumed from their responses that business registers were part of their portfolio of geocoded data.

In order to find out about the capacity of the NSIs to geocode information, Q22 asked about who conducted the process of data pairing².

In 19 countries NSIs conducted the process of pairing statistics and geospatial data without any external help. Three countries reported that they conducted the process for data pairing by collaboration between the NSI and the NMCA. Furthermore, two countries conducted the work through cooperation between the NSI and private companies.

² Data pairing – the process of merging data from two different sources - in this case linking data from different registers with its spatial reference (x, y coordinates). In other words – geocoding.

Results obtained for Q22. Who conducted the process of data pairing?

The aim of Q10³ was to establish whether the coordinates of address points and/or building points, along with the coordinates of cadastral parcels, were considered to form the complete basis for point-based statistics. In order to learn more about the preferences regarding these datasets, in Q24 respondents were asked about the key point-based geospatial data used by NSIs for geocoding. Multiple choices were allowed.

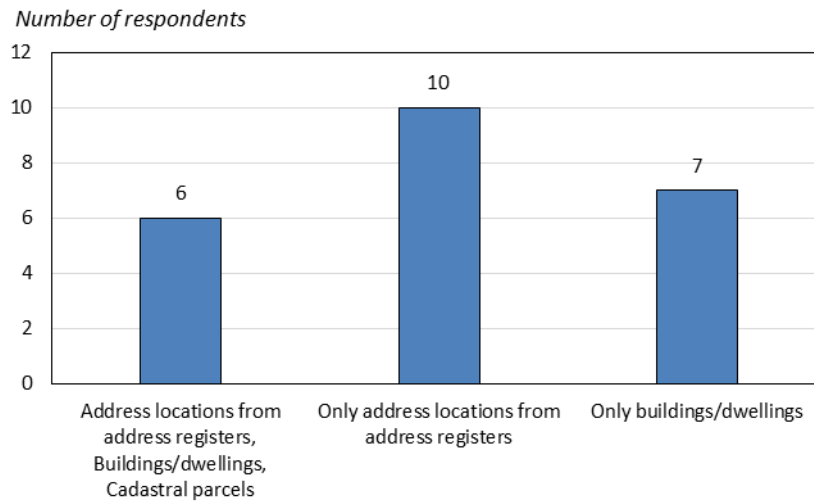
The majority of the countries (23 countries) indicated address registers as the key point-based geospatial data used for geocoding statistics in their NSIs. Roughly half of the countries indicated that records of buildings/dwellings were key point-based geospatial data used for geocoding statistics. Only 9 of the countries mentioned cadastral parcels as key point-based location data. 4 countries lack access to any point-based spatial data for geocoding.

As the question allowed for multiple answers, it is worth noting that in 6 countries both address registers, building/dwelling registers *and* registers of cadastral parcels are considered key point-based geospatial data used for geocoding statistics.

In 10 countries, only address registers are considered to be the key point-based geospatial data used for geocoding statistics and in 7 countries only buildings/dwellings registers were mentioned. None of the countries consider *only* records of cadastral parcels to be key point-based location data.

³ Phrasing of the question: According to your experience does the following two set of point data represent the complete basis for points based statistics: 1) coordinates of address points* and/or building points and 2) coordinates of cadastral parcels?

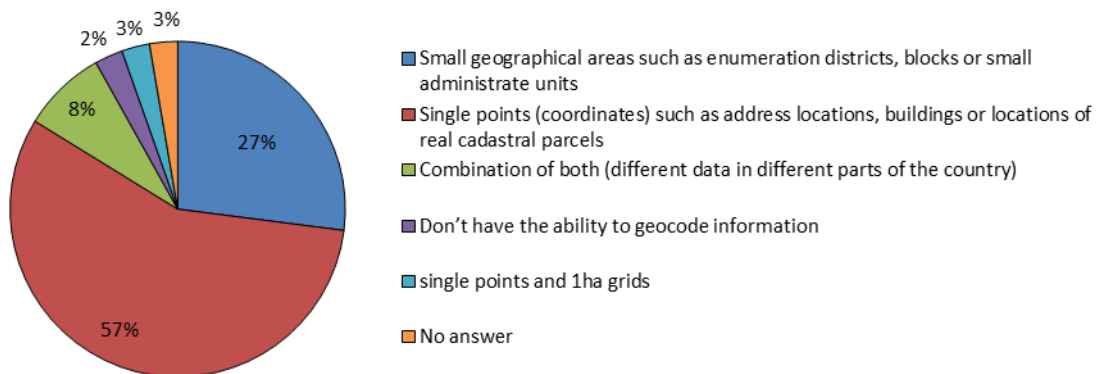
Results obtained for Q24. What is the key point-based geospatial data used for geocoding statistics in your NSI (multiple choices possible)?



The majority of countries (22) reported that single points (coordinates), such as address locations, buildings or locations of cadastral parcels, were the lowest possible geographical level to which they could geocode population distribution information. Around 10 countries answered that small geographical areas, such as enumeration districts, blocks or small administrative units, constituted the lowest possible geographical level to which they could geocode population data. Hence, one third of the countries did not report having a fully point-based production model for population statistics. A couple of countries use a combination of both approaches (due to different data in different parts of the country).

One country does not have the ability to geocode information at all and one country did not provide an answer.

Results obtained for Q26. What is the lowest possible geographical level to which you can geocode population distribution information?



Based on the responses, the following main conclusions can be drawn:

- There is a relatively high maturity among NSIs regarding the geocoding of statistical information in general. At least three-fourths of the countries have performed some kind of geocoding, and in a majority of the countries, this has been done by the NSIs alone or with support from other expert authorities, such as the NMCA.
- However, it is also clear that the level of complexity in data pairing varies between countries. The geocoding of basic census data indicators, such as population distribution, seems to be a reasonably established activity, whereas procedures for geocoding other information are less developed.

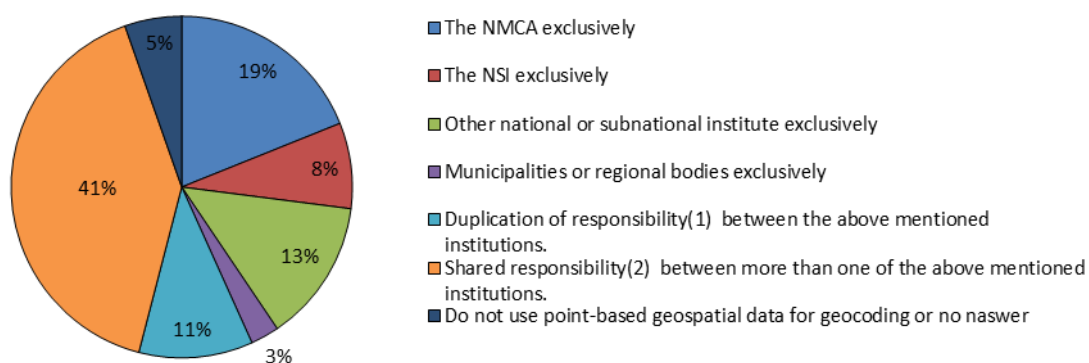
2.4 The responsibility for and maintenance of geospatial information

The question of responsibility for the production and maintenance of geospatial information is of vital importance for long-term work towards a better integration of geography and statistics. Already from the start of the project it was quite well known that different models would apply throughout Europe. However, no systematic knowledge on these different models existed prior to the survey.

In Q23 respondents were asked to describe who was responsible for creating and maintaining the key point-based geospatial data used for geocoding statistics. Not surprisingly, the most common model seems to be the shared responsibility of more than one institution. Mainly, the institutions involved in the model of shared responsibility are NMCA's and municipalities. But also other combinations exist, such as NSIs together with NMCA's, or NSIs and municipalities. The second most frequent model is that NMCAs are exclusively responsible for the production and maintenance of the key point-based geospatial data. "*Other national or sub-national institutions*" accounts for some 14 percent of the countries, and in some 8 percent of the countries, the NSIs are exclusively responsible for the production and maintenance of data.

One of the possible options for this question was "*the duplication of responsibility*" as opposed to "*shared responsibility*". Four countries reported such circumstances, potentially indicating a conflict in terms of authoritative data.

Results obtained for Q23. Who is responsible for creating and maintaining the key point-based geospatial data which you use for geocoding statistics in your NSI?

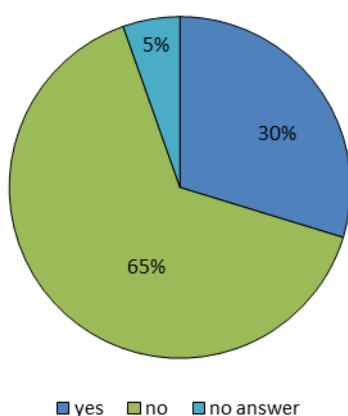


2.5 The conditions for and obstacles to successful geocoding practice

This section aims to understand the conditions for successful geocoding practice and possible obstacles preventing countries from geocoding information and/or producing spatial statistics. The questions were designed to shed light not only on technical aspects and data access, but also on legal issues and the organisational setup.

Respondents were asked to define whether there was any legislation which allowed *or* prevented their NSIs from collecting data with the accuracy of address points. The majority of countries responded that there were no such regulations. Among those who responded positively, there was no country indicating legislation preventing such data collection. One of respondents concluded that there was no law mentioning geocoding explicitly, but in general that there were no major obstacles in this matter. Two countries did not provide answers to this question.

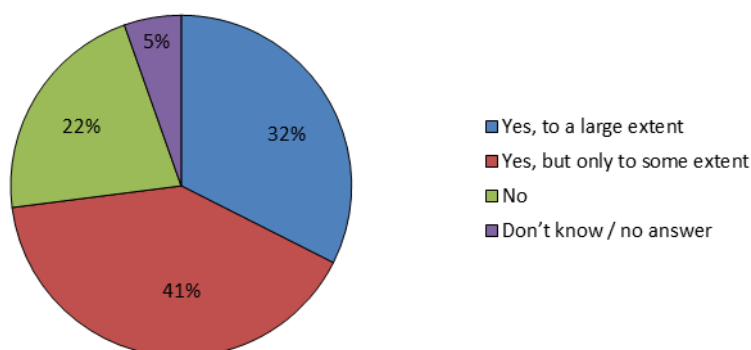
Results obtained for Q27. Is there any legislation which allows OR prevents NSI data collection with the accuracy of the address point?



A point-based production system cannot run on the presence of coordinates alone. There need to be unique identifiers and consistent keys linking statistical information or administrative data to specific locations. In terms of the traditional collection and structuring of statistical information, e.g. census data, it is within the scope of the statistical institutes themselves to support the implementation of keys compliant with those used by authoritative location data (e.g. standardised address codes). The increased use of administrative sources poses a greater challenge in terms of the implementation of unique identifiers as those data sources to a large extent are external to NSIs.

The responses provided to Q28 indicate that in less than one third of the countries there is clearly a formal policy or custom among public institutions to use standardised identifiers in registers or administrative records (address information, personal IDs, real-estate codes, building IDs etc.). Not surprisingly, this category is dominated by countries with a long-standing tradition of using centralised registers, such as the Nordic countries, Austria and the Netherlands. In around 40 percent of the countries, such policies exist, though only to a limited extent.

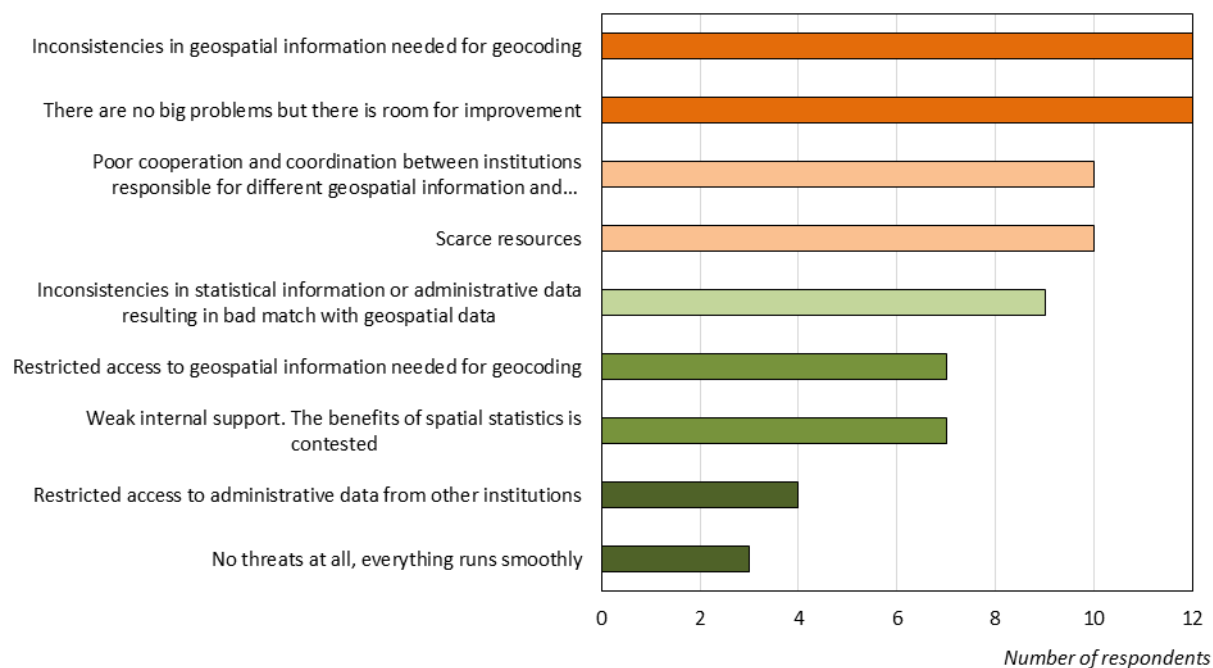
Results obtained for Q28. Is there is a formal policy or informal custom among (public) institutions to use standardised identifiers in registers (address information, personal IDs, real-estate codes, building IDs etc.?)



In the survey, respondents were asked to define what they believed to be the main threats related to current geocoding practice. Multiple choices were possible for this question (Q32). A significantly large number of respondents seem to be quite content with the current situation in their countries. Some 40 percent of the countries reported that either they saw no threats at all or that they saw no big problems; however, there was room for improvement.

On the down-side, *“inconsistencies in geospatial information needed for geocoding”* seem to be the most prominent threat, closely followed by *“scarce resources”* and *“poor cooperation and coordination between institutions responsible for different geospatial information and administrative data”*. It is also worth noting that *“restricted access to administrative data from other institutions”* is the least frequent category of threats. Only 4 countries assessed this as a threat.

Results obtained for Q32. What are the main threats in your country related to your current geocoding practice?

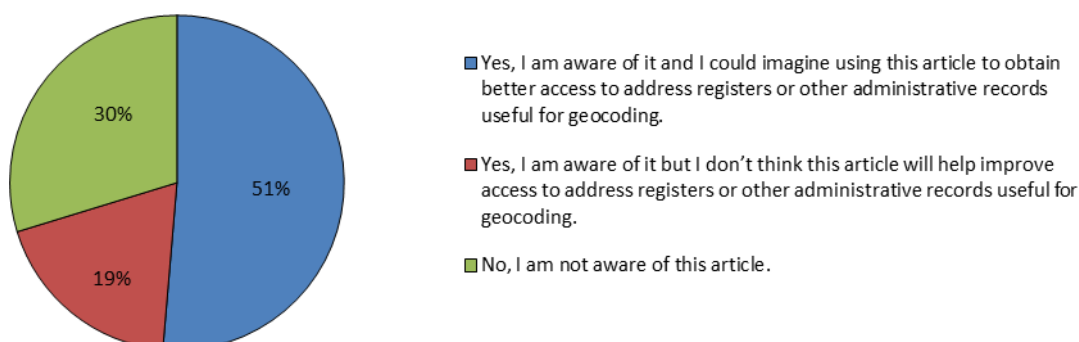


Prior to the survey, the Regulation (EU) 2015/759 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 29 April 2015, amending Regulation (EC) No. 223/2009 on European statistics, was passed, giving NSIs stronger rights to access and use administrative records. The regulation does not explicitly address the geocoding of

administrative records, but the GEOSTAT 2 project wanted to use the opportunity to ask respondents about their views on this regulation as a possible instrument to increase access to administrative records for geocoding.

Around 70 percent of countries indicated that they were aware of this regulation. Among the countries that were familiar with its articles, the majority made the assessment that it could be helpful in improving access to address registers or other administrative records for the purpose of geocoding. Around 20 percent of respondents were aware of the regulation but did not appreciate it as a means to improve access to data.

Results obtained for Q37. Are you aware of the REGULATION (EU) 2015/759 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 29 April 2015, amending Regulation (EC) No. 223/2009 on European statistics, which gives NSIs stronger rights to access and use administrative records?



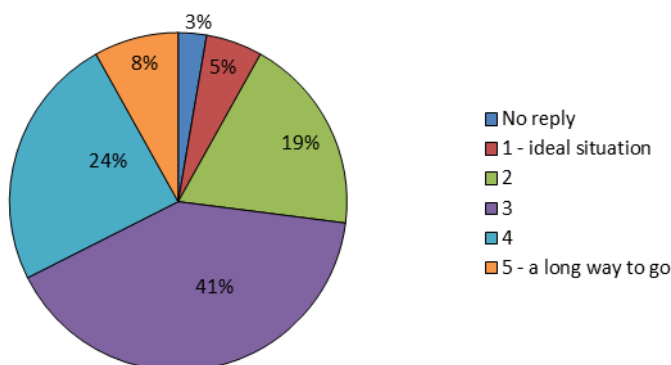
The next question, associated with the section dealing with the conditions for and obstacles to the successful geocoding practice, concerned the kind of support the NSIs need in order to strengthen their infrastructure for the production of spatial statistics. The question was open for free-text input. Despite a broad range of comments, it is possible to distinguish a few recurrent issues. Many respondents made references to resources/financial support/labour force etc., revealing that activities related to geospatial information, geocoding or production of spatial statistics were not at the core of priority in many NSIs. The need for more resources, in order to establish an infrastructure for the production of spatial statistics, is more commonly mentioned by new EU members and candidate countries. The exchange of best practices related to the problems with geocoding, geocoding historical data (methods), and a common approach towards INSPIRE recommendations, were also mentioned.

Another recurrent theme of support is cooperation. Cooperation was usually mentioned in relation to the NMCA, from which the NSIs need more support. Cooperation can involve access to data and also to national spatial data infrastructures. Finally, legal issues seem to be a common concern. Several countries expressed the need for EU regulations as well as national laws.

In the final question in the reference section (Q36), respondents were asked to make a self-assessment regarding the performance of the infrastructure for the production of spatial statistics in their countries (in respect of technical solutions, the legal situation, financial and human resources, access to data, quality of data, cooperation with other institutions, etc.). Respondents were asked to first imagine an ideal situation and then indicate how close or how far from this situation they were, using a five-point scale, where 1 was ideal and 5 far from ideal.

Roughly 40 percent of the countries placed themselves on 3, indicating that the situation fairly good, yet not ideal. Almost 25 percent placed themselves in the upper range, indicating that the situation was either ideal or close to ideal. At least one-third of the countries placed themselves in the lower range of the scale, indicating that they had a long or very long way to go to achieve an ideal situation.

Results obtained for Q36. Imagine an ideal situation regarding the infrastructure for the production of spatial statistics in your country and NSI (in respect of technical solutions, the legal situation, financial and human resources, access to data, quality of data, cooperation with other institutions etc). How close or far from this ideal situation are you today? (1 indicates that you have an ideal situation and 5 indicates that you have a long way to go).



Based on the responses, the following main conclusions can be drawn:

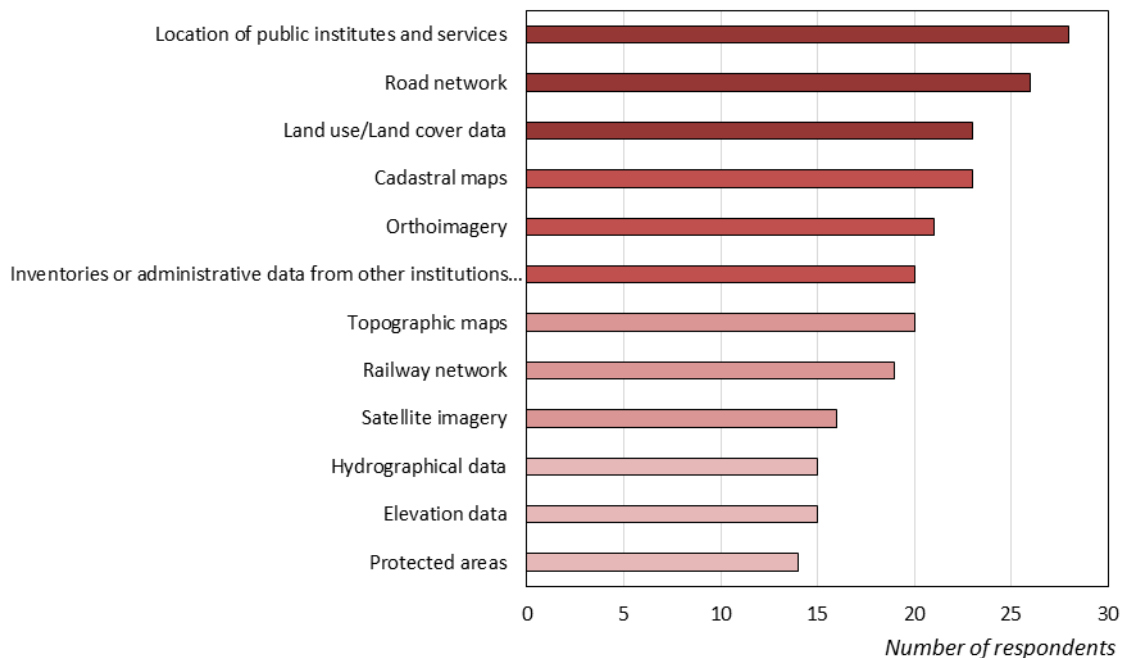
- Legal restrictions do not seem to be a significant reason preventing the geocoding of statistical information in the Member States. The regulations are rather considered a tool to facilitate geocoding by placing obligations on public institutions to share data with the NSIs.
- A big challenge to the long-term goal of establishing point-based production models in the ESS seems to refer to sufficient resources and knowledge. Activities related to geospatial information, geocoding or the production of spatial statistics are not at the core of priority in many NSIs.
- Poor cooperation between institutions is a major obstacle in many countries. The lack of a collaborative climate is most likely also related to the problems faced with inconsistencies in data (both spatial and administrative).

2.6 The current use of and needs for spatial information in statistical production

One aim of the survey was to shed some light on the use of geospatial information in statistical production besides point-based location data used for geocoding statistics (investigated in previous sections).

The first question in this section (Q29) investigated what datasets, according to respondents, should *“supplement address and/or building locations to form a fully fledged spatial reference framework”*. Multiple choices were allowed. The question is unfortunately a little vague in its definition of the spatial reference framework; hence it is difficult to interpret the results. It can be discussed whether, for example, orthophoto imagery could be considered part of a spatial-reference framework *per se*, or whether it should be considered complementary to such a framework. Regardless of the interpretation of “the spatial reference framework”, the answers provide valuable insights in the actual needs of the NSIs.

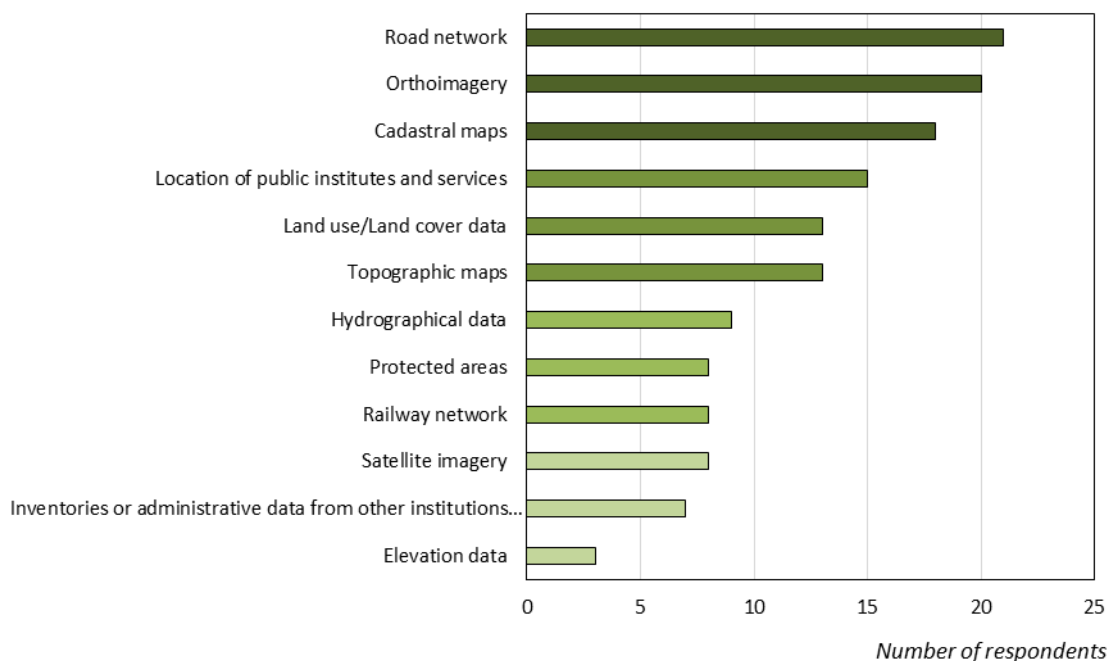
Results obtained for Q29. The ultimate purpose of the GEOSTAT 2 project will be to develop a spatial reference framework for statistics. Hence, although the GEOSTAT 2 project mainly deals with address or building point data as part of this framework, we would like to know more about the use and needs regarding other point-based geospatial data or data with other topology that are needed for a fully-fledged framework. Which of the following datasets should supplement address and building points to form this spatial reference framework (multiple choices possible)?



The most useful supplementary datasets indicated by respondents were “*location of public institutes and services (schools, hospitals, universities etc.)*” closely followed by “*road networks*”. In the next position equal importance was given to “*cadastral maps*” and “*land use/land cover data*” and “*orthophoto imagery*”. As the least important data “*protected areas*” were indicated, along with “*hydrographical data*” and “*elevation data*” (see the bar chart for the complete ranking of data).

In the next question (Q30), respondents were asked to rank the same datasets as in Q29, but this time they were asked if they were actually using the data in their current production of statistics. 6 countries did not give an answer to this question.

Results obtained for Q30. Please indicate which of the data from Q29 you already use in the production of statistics within your NSI (multiple choices possible)?



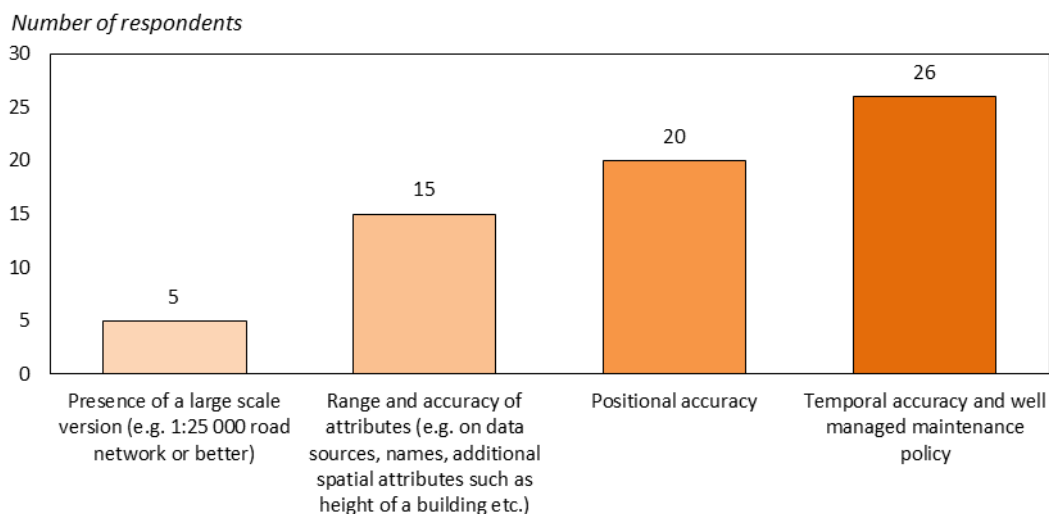
Of the datasets to choose from in the survey, “*orthophoto imagery*” and “*road network*” are the most widely used in the current production of statistics. Other quite commonly used categories of data include “*cadastral maps*”, “*location of public institutes and services (schools, hospitals, universities etc.)*” and “*topographic maps*” and “*land use/land cover data*”. The least frequently used category of data is “*elevation data*” (see the bar chart for the complete ranking of data).

When comparing the results obtained for Q29 and Q30, it is interesting to note the discrepancy between the demand and supply (or actual use) for some categories of data. An overwhelming majority of countries indicated that data on “*location of public institutes and services*” (76 percent) was crucial to supplement address and/or building points to form a spatial reference framework, but 40 percent of the countries are actually currently using this kind of data. The same pattern goes for “*inventories or administrative data from other institutions captured and stored with coordinates (environmental permits, Pollutant Release and Transfer Registers, Urban waste water treatment plants)*”. This category of data was selected by nearly 55 percent of the countries for Q29, but according to Q30 less than 20 percent are currently using it. Presumably the discrepancy can be explained by the lack of data.

The situation regarding “*orthophoto imagery*” is the reverse. According to Q30, this is the most widely used category of data; yet it is only ranked as number 5 for Q29. Hence, it can be assumed that the coverage of, and access to, “*orthophoto imagery*” is reasonably satisfactory as compared to other categories of data. However, it should be stressed that not all countries provided input for Q29 and Q30, and some countries provided input for Q29 but not or Q30, or the other way around. This fact prevents any viable comparison between the two questions.

The final part of the section concerns quality. Respondents were asked to define what quality characteristics of the above-mentioned datasets (categories from Q29 and Q30) were the most important. Respondents were allowed to choose a maximum of two of the options given. 33 countries provided answers to this question.

Results obtained for Q31: If you were to choose, which of the following quality characteristics of the above datasets are the most important?

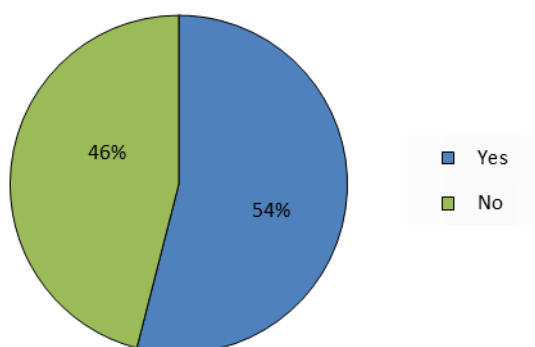


According to respondents, by far the most important quality characteristic of the datasets is “*temporal accuracy and well-managed maintenance policy*”. The answers imply that temporal accuracy and well managed maintenance policy is even more important than “*positional accuracy*” of the data.

“*Range and accuracy of attributes (e.g. on data sources, names, additional spatial attributes such as height of a building etc.)*” is of less importance, but the least important is “*presence of a large scale version (e.g. 1:25 000 road network or better)*”. Only five countries chose this option.

The second question on quality concerns the existence of geospatial data correctness verification processes (Q33). Around 46 percent of the responding countries gave a negative answer to this question, implying that no such procedures existed. Accordingly, the remaining respondents indicated that they conducted some form of geospatial data correctness verification and they also specified how this was done. The main method, indicated through comments from respondents, is associated with pre-census works and post-census works. Other forms of verification are connected with periodic controls and the accuracy of location information. Some countries reported systematic feedback-loops to NMCA’s regarding errors found in address registers during preparatory census work.

Results obtained for Q33. Is there any form of geospatial data correctness verification conducted in your NSI (e.g. post-enumeration surveys, pre-enumeration surveys [census preparatory works])



Based on these responses, the following main conclusions can be drawn:

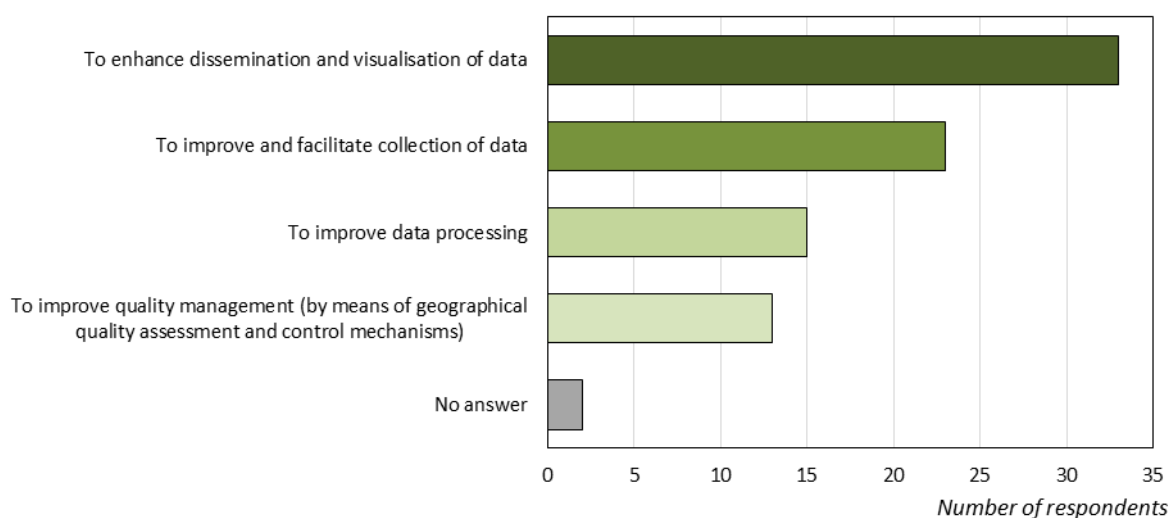
- There is undoubtedly a huge potential for the increased use of geospatial data in statistical production. A general observation is that the demand for geospatial information by NSIs generally exceeds the supply.
- A wide range of geospatial information is considered extremely useful in statistical production; yet the top priority of NSIs is any data that can contribute to facilitate the geocoding of statistical information.
- Even though geospatial data might exist, the survey indicated that the NSIs were not fully satisfied with the temporal accuracy of the information.
- The various uses of geospatial data and the varying intensity at which data is used, reflected by respondents, is most likely due to the different tasks and mandates of the NSIs. NSIs entrusted with the production of environmental statistics, land use statistics, and transport and agriculture statistics, are likely to be larger consumers of geospatial information.

2.7 The role and benefits of geospatial information in statistical production

The final section of the survey aimed to gather information on the role and benefits of geospatial information in statistical production. It is well known that geospatial information is used by many NSIs for the dissemination and visualisation of statistics. In several countries NSIs provide more or less advanced atlases or web services for regional statistics. But considering the potential use of geospatial information in the entire chain of production, less is known about the situation across Europe. Basically, two similar questions were asked. In the first question (Q34) respondents were asked to reflect on the role of geospatial information in their NSI, and in the second question they were asked to give their opinion on the main benefits of using spatial data in their NSI. In both questions multiple choices were allowed.

Regarding the role of geospatial data, it was expected that many countries would choose *“to enhance dissemination and visualisation of data”*. A large proportion of countries also chose *“To improve and facilitate the collection of data”*, which was also expected, due to the fact that geospatial information is involved in the census work in many countries. The least-recognised options were *“To improve data processing”* and *“To improve quality management (by means of geographical quality assessment and control mechanisms)”*. Only 15 and 13 countries, respectively, chose these options. 8 countries chose all four available options.

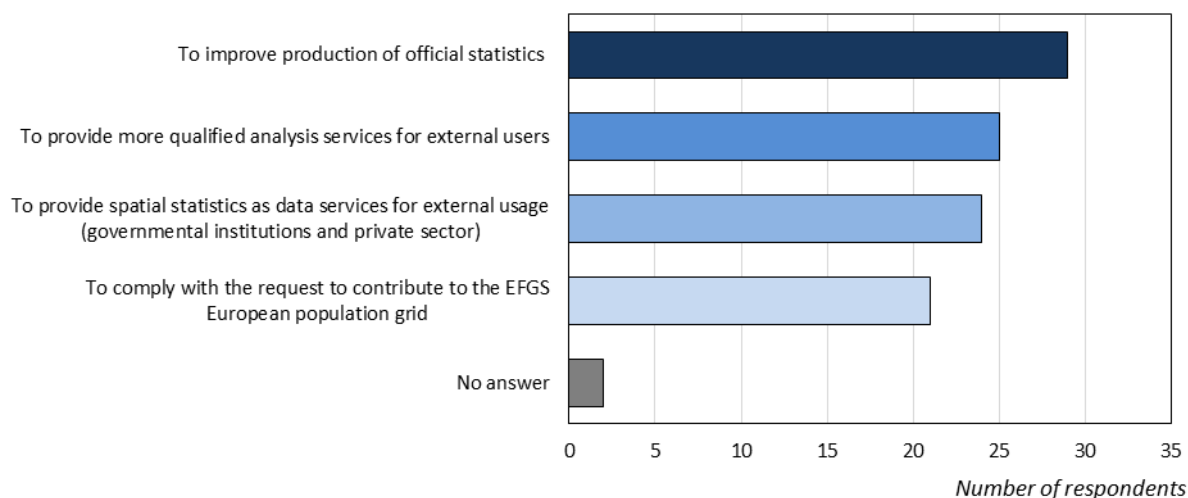
Results obtained for Q34. What is the role of spatial data in your NSI?



In Q35 respondents were asked about the main benefits of using spatial data in their NSI. They were given four pre-defined options to choose from (multiple choices allowed). The most common answer, indicated by 29 countries, was *“To improve the production of official statistics”*. The least common answer, indicated by 21

countries, was *“To comply with the request to contribute to the EFGS European population grid”*. In total, 14 countries indicated all four options and 3 countries did not choose any of the options.

Results obtained for Q35. What are the main benefits of using spatial data in your NSI?



Based on the responses, the following main conclusions can be drawn:

- The survey indicates the broad application of geospatial information in statistical production. Yet, in many countries the use of geospatial data is still mainly connected with the tasks of collecting data, and disseminating and visualising the output. Improvements in quality management and data processing (the throughput) constitute the domain which can be expected to unfold in the near future.
- A review of the use of geospatial information in line with the GSBPM model would contribute to further explore the benefit of the information.

2.8 The geocoding capacity of Census 2021

As mentioned in the introductory chapter, the GEOSTAT 2 project is working in parallel with the preparations for the 2021 round of the population census. However, the GEOSTAT 2 project will not make any proposals directly addressed to the undertaking of the census. A special task force has been set up by Eurostat to deal with the initial coordination of the upcoming census.

A strategic proposal put forward by the census 2021 task force is to make population distribution on 1 km² grids the standard output of the census. Such a proposal would in principle require the implementation of a point-based production model in the Member States, with the capacity to aggregate population and associated information to grid cells. In order to assess the basis for such a proposal, Eurostat and the census 2021 task force turned to the GEOSTAT 2 project for advice from the survey results.

Eurostat would like to propose the option to introduce implementing legislation as a temporary statistical action under the Regulation on European Statistics (Regulation (EC) 223/2009). This option would assist those NSIs which would be unable to participate in the collection on 1km² grid census data unless there was an obligation under EU statistical legislation.

Regulation (EC) 223/2009 clarifies the arrangements under which an implementing statistical regulation may be introduced to provide a legal basis for temporary direct statistical action. Article 14 of the Regulation states the main conditions which need to be applied to such an act:

...

(a) the action does not provide for data collection covering more than three reference years;

(b) the data are already available or accessible within the NSIs and other responsible national authorities, or can be obtained directly, using the appropriate samples for the observation of the statistical population at the EU level with the appropriate coordination with the NSIs and other national authorities; and

(c) the EU makes financial contributions to the NSIs and other national authorities to cover the incremental costs incurred by them, in accordance with Regulation (EU, Euratom) No 966/2012 of the European Parliament and of the Council (1).

With regard to condition b), Eurostat has requested relevant information as part of the GEOSTAT 2 survey, addressed to the geographical units of the NSIs. The key concern is whether the NSIs can already provide (or expect to be able to provide for the 2021 census) population data at the level of the 1 km² grid. To ensure high quality, the production of such a grid should, as far as possible, be based on the aggregation of population data geocoded to single point (coordinate) locations (bottom-up method).

As the GEOSTAT 2 survey was designed to describe the current state of play, hence made no references to future prospects, it was decided that some additional questions had to be amended to the original survey to be able to meet the needs of the census 2021 task force.

In this respect two additional questions were prepared and sent to all countries which, in question 26 of the original survey, stated that they currently lacked the capacity or the infrastructure to fully geocoded population distribution to single-point location.

Countries which in the original survey indicated that their production systems already supported a full point-based geocoding of population distribution needed take no further action. On the basis of their response, it was assumed that the aggregation of 2021 census data to 1km² grids would be feasible for them.

The first follow-up question was:

“Do you have or do you foresee that you will have the necessary data and infrastructure to geocode the next population census 2021 on single points (coordinates) such as address locations, buildings or locations of real cadastral parcels?”

Respondents were asked to use some of the following predefined answers:

- *“YES”*
- *“NO”*
- *“NO, but I foresee having access to these data for 2021 census”*
- *“Probably, we are already working towards this goal”*
- *“Possibly, but no fixed plans yet”*

The second follow-up question was

“Do you foresee that you will be able to generate 1km² population grids with census 2021 data?”

Respondents were asked to use some of the following predefined answers

- *“YES”*
- *“NO”*
- *“Probably, we are already working towards this goal”*
- *“Possibly, but no fixed plans yet”*

From the answers it can be inferred that the vast majority of the EU/EFTA countries already have, or foresee having, the necessary data and infrastructure in place to enable the geocoding of the next population census

2021 to the level of single-point (coordinate) location. Out of the 32 ESS countries, 23 already have the necessary data and infrastructure in place, one country foresees having it and three countries responded that they would *probably* or *possibly* have it in place. Five countries remain in the group which does not foresee having the necessary data and infrastructure in place for the point-based geocoding of the 2021 census.

Despite the fact that five of the ESS countries will not be able to set up a point-based environment for geocoding, most of them are positive about being able to generate a population grid with 2021 census data. In total 28 of the 32 ESS countries foresee that they will be able to provide gridded population data, one country will *probably* be able to do so and three countries believe it is *possible*. No country responded negatively to this question. Countries without the capacity for point-based geocoding will have to use disaggregation methods to produce proxy estimates of single-point coordinates.

On the basis of the results obtained for Questions 1 and 2 it appears that condition b) in Article 14; Regulation (EC) 223/2009 will be met.

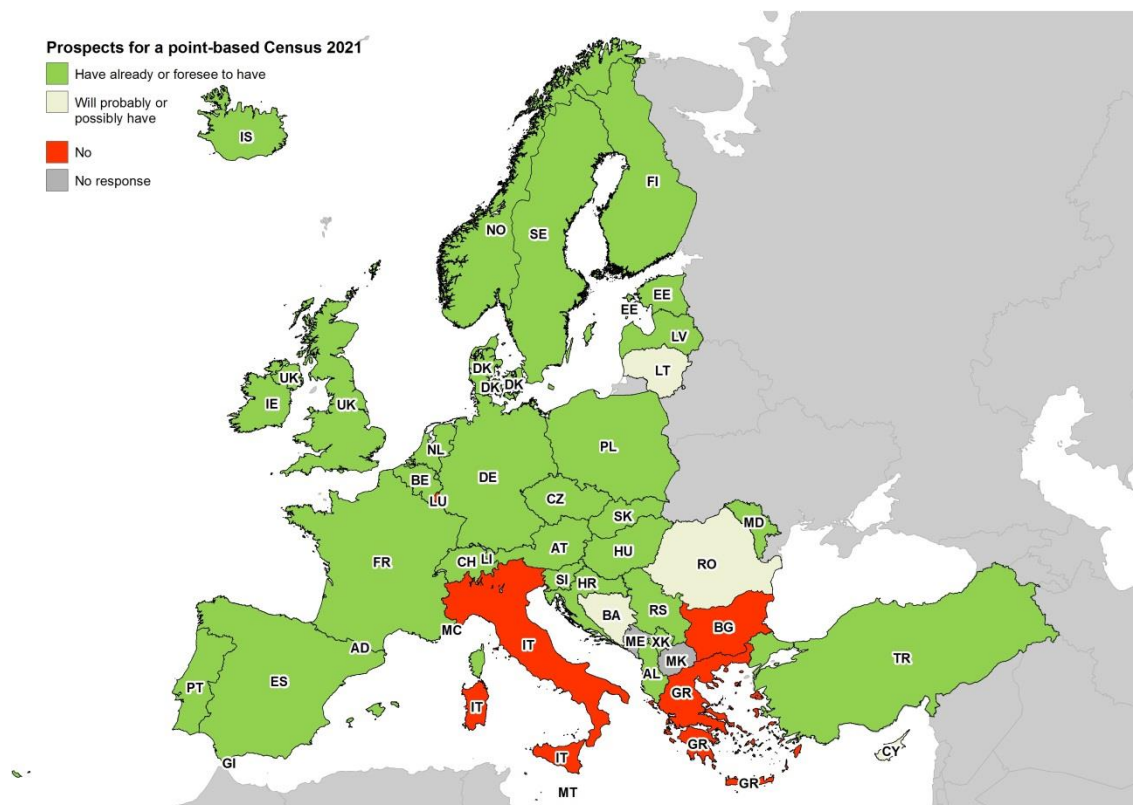
The table below shows the results for question 1 and 2 country by country. The inquiry from Eurostat concerns only the situation in the EU/EFTA countries; however in the lower part of the table are also included results from non-ESS countries.

Table 1. Results obtained for question 1. Do you have or do you foresee that you will have the necessary data and infrastructure to geocode the next population census 2021 on single points (coordinates) such as address locations, buildings or locations of real cadastral parcels?

	Countries	Have the necessary data and infrastructure already in place to geocode population data to the level of single point (coordinate) location	Foresee having the necessary data and infrastructure in place to geocode the next population census 2021 to the level of single point (coordinate) location	Will <i>probably</i> have the necessary data and infrastructure in place to geocode the next population census 2021 to the level of single point (coordinate) location	Will <i>possibly</i> have the necessary data and infrastructure in place to geocode the next population census 2021 to the level of single point (coordinate) location	NO, but I foresee having access to these data for 2021 census	NO
1	Austria	●					
2	Belgium	●					
3	Croatia	●					
4	Czech Republic	●					
5	Denmark	●					
6	Estonia	●					
7	Finland	●					
8	France	●					
9	Germany	●					
10	Hungary	●					
11	Iceland	●					
12	Ireland	●					
13	Latvia	●					
14	Liechtenstein	●					
15	Netherlands	●					
16	Norway	●					
17	Poland	●					
18	Portugal	●					
19	Slovenia	●					
20	Spain	●					
21	Sweden	●					
22	Switzerland	●					
23	United Kingdom	●					
24	Slovak Republic		●				
25	Romania			●			
26	Cyprus				●		
27	Lithuania				●		
28	Bulgaria					●	
29	Greece					●	
30	Malta					●	
31	Italy						●
32	Luxembourg						●
<i>Responses from non ESS countries</i>							
1	<i>Republic of Moldova</i>	●					
2	<i>Albania</i>		●				
3	<i>Kosovo*</i>		●				
4	<i>Serbia</i>		●				
5	<i>Turkey</i>		●				
6	<i>Bosnia and Herzegovina</i>			●			

* This designation is without prejudice to positions on status, and is in line with UNSCR 1244 and the ICJ Opinion on the Kosovo declaration of independence.

Map 6. Results obtained for question 1. Do you have or do you foresee that you will have the necessary data and infrastructure to geocode the next population census 2021 on single points (coordinates) such as address locations, buildings or locations of real cadastral parcels?



Source: GEOSTAT 2 project/EFGS

Table 2. Results obtained for question 2. Do you foresee that you will be able to generate 1km² population grids with census 2021 data?

	Countries	Yes	Probably, we are already working towards this goal	Possibly, but no fixed plans yet
1	Austria	●		
2	Belgium	●		
3	Croatia	●		
4	Czech Republic	●		
5	Denmark	●		
6	Estonia	●		
7	Finland	●		
8	France	●		
9	Germany	●		
10	Greece	●		
11	Hungary	●		
12	Iceland	●		
13	Ireland	●		
14	Italy	●		
15	Latvia	●		
16	Liechtenstein	●		
17	Luxembourg	●		
18	Netherlands	●		
19	Norway	●		
20	Poland	●		
21	Portugal	●		
22	Slovenia	●		
23	Spain	●		
24	Sweden	●		
25	Switzerland	●		
26	United Kingdom	●		
27	Slovak Republic	●		
28	Romania	●		
29	Malta		●	
30	Cyprus			●
31	Lithuania			●
32	Bulgaria			●
<i>Responses from non ESS countries</i>				
1	<i>Republic of Moldova</i>	●		
2	<i>Albania</i>	●		
3	<i>Kosovo*</i>	●		
4	<i>Serbia</i>	●		
5	<i>Turkey</i>	●		
6	<i>Bosnia and Herzegovina</i>		●	

* This designation is without prejudice to positions on status, and is in line with UNSCR 1244 and the ICJ Opinion on the Kosovo declaration of independence.

3 Discussion and conclusions

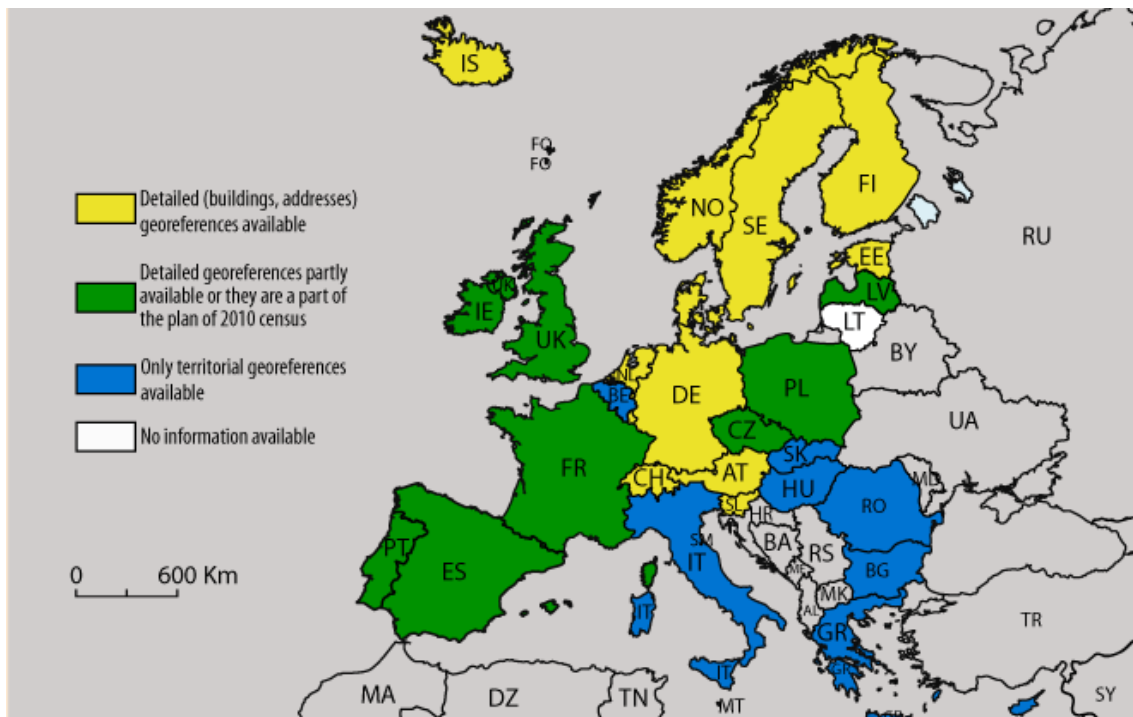
Prior to the survey, little systematic knowledge on the current situation regarding practices for geocoding statistics and the use of geospatial information in European NSIs, was at hand. However, in the GEOSTAT 1 project, an assessment within the ESS of the current state of affairs as regards the geocoding of population statistics and of the 2011 Census was undertaken in 2010. Even though the two surveys are not equal in terms of the scope, content and questions, it is possible to make some comparisons. A striking observation is the strong progress being made in the ESS during the last few years.

As of today, the majority of the countries have, or foresee having, a point-based production model for spatial statistics in the near future. Roughly half of the countries reported that point-based production models are already in place. Looking ahead, most of the remaining countries anticipate being able to geocode the next population census (scheduled for 2021) to the level of a single coordinate points.

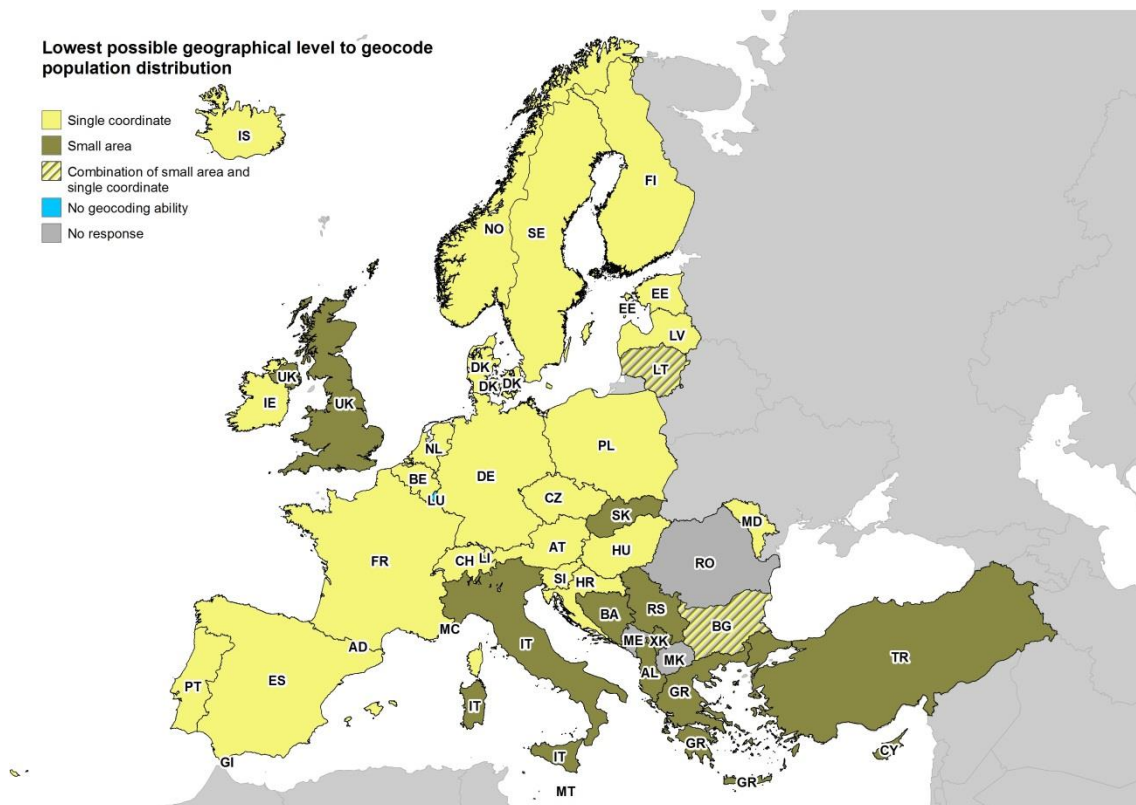
Most of the countries indicating a lower degree of coverage for point-based location data are candidate countries or possible candidates. It can be assumed that the main reason for the lower degree of coverage in those countries is that the location data frameworks are under construction and have not yet reached completion.

As shown in Map 6 below, the situation was quite different only five years ago. At that point, location frameworks to geocode census information were dominated by area-based data (territorial geo-references) or a combination of area-based data and point-based data (detailed geo-references). A typical trend is that the majority of the countries which previously used a hybrid area-based/point-based location framework have since successfully implemented a point-based framework.

Map 7. The available geo-references on the 2011 Census data



Source: 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.

Map 8. The lowest possible geographical level to which population can be geocoded 2015

Source: GEOSTAT 2 project/EFGS

The situation described above concerns the specific geocoding of population statistics but the trend is generic in the sense that geospatial information management as a whole is an expanding activity in European NSIs. The survey shows that almost every country possesses some sources of data for which it is possible to retrieve the spatial location of its units with coordinate accuracy – these are mostly coordinates of buildings, locations of address points or location of cadastral parcels. Furthermore, the absolute majority of the countries have already conducted some geocoding of statistical data, and most of those who have done so were able to geocode a set of indicators from population or agricultural censuses with full or partial coverage. In most cases, the NSIs conduct the process of pairing statistics and geospatial data without outside help. From the point of view of the project outcome, this is promising, as there seems to be fertile ground for the receipt and implementation of the project results.

Despite the fact that some respondents indicated the need for additional data sources to make up a complete basis for point-based statistics, there is a strong support for the GEOSTAT 2 working hypothesis, e.g. that the core set of location data on which a point-based production model needs to rest is address and/or building information. With some few exceptions, a well-managed authoritative address and / or building register is fit for the purpose of geocoding most of the information found in statistical institutes or other public bodies. The exceptions mentioned by respondents can be divided into two categories:

- Cases where additional data are needed to improve the geocoding of information where address-match fails (such as population in mobile households etc.)
- Cases where the use of address locations will produce a non-appropriate spatial representation of the phenomena. Such examples can be location of agricultural holding or discharge points for water and pollution from industries.

In many countries, the location data frameworks contain not only an address register, but also very likely a building register and/or a register of cadastral parcels. Even though there are very small differences between specific categories, address locations seem to be the most mature category of location data, at least in terms of coverage. From the European perspective it is of less importance if a Member State builds its point-based production model on address data, building data or cadastral parcel information. The choice of location data should rather be guided by principles of authoritativeness and potential for temporary maintenance. According to the survey, temporary accuracy and well-managed maintenance policies are valued even higher than positional accuracy.

Despite promising developments in recent years, the survey indicates several important obstacles preventing NSIs from the successful implementation of geocoding procedures. Prior to the survey, legal restrictions were expected to pose some obstacles. However, according to respondents the legal restrictions do not seem to be a significant reason *preventing* the geocoding of statistical information in the Member States. Regulations are rather considered a tool to *facilitate* geocoding by placing obligations on public institutions to share data with the NSIs and to use authoritative geospatial data to enable the geocoding of administrative data records at the unit record level.

A big challenge for the long-term goal to establish point-based production models in the ESS seems to be sufficient resources and knowledge. Activities related to geospatial information, geocoding or production of spatial statistics is not at the core of priority in many NSIs. The need for capacity building remains to transfer knowledge both between countries and between institutions in countries.

Poor cooperation between institutions is reported to be a major obstacle in many countries. Lack of a collaborative climate is most likely also related to the problems reported with inconsistencies in data (both spatial data and administrative data).

4 List of terms

Data pairing

Data pairing refers to the process of merging data from two different sources - in this case linking data from different registers with its spatial reference (x, y coordinates). Data pairing is used as a synonym for *geocoding*.

Geocoding

The geocoding of statistics describes the process of *geospatially facilitating* statistical unit record data (i.e. data relating to individual persons, households, dwellings, businesses or buildings) by assigning a location descriptor (x, y, z coordinates, address, spatial id, spatial feature). The geocodes (location coordinates, address IDs, or geographic area codes), obtained from this process can be stored directly on the statistical-unit record or linked in some way to the record.

Geo-enabling

Geo-enabling, or geospatially enabling, is used as a general reference to the process of assigning spatial location to statistical information. Statistics can be geo-enabled by means of *geocoding*. Geo-enabling is used as a synonym to *spatialise*.

Location data

Location is a general term to describe a place on, or near, the surface of the Earth. Location data is information which has any location component and is often used when referring to geospatial information. In this context, location data is typically the location of address points, building points or the location of cadastral parcels.

Spatial statistics

Location or extent are the main characteristics of spatial statistics. Spatial statistics are geocoded to small administrative or non-administrative geographies. Spatial statistics can also result from the integration of statistical and geospatial information during the statistical-production process, although the product might be regional statistics.

Spatialising

Spatialising is used as a general reference to the process of assigning a spatial location to statistical information. Statistics can be spatialised by means of *geocoding*. Spatialising is used as a synonym for *geo-enabling*.

Appendix 1- Questionnaire

GEOSTAT 2 - Survey about geocoded administrative data in terms of their use for statistical purposes

Welcome to the GEOSTAT 2 questionnaire!

You are going to fill in the questionnaire which is a survey within ESS (European Statistical System) countries about administrative data in member states in terms of their use for statistical purposes (collecting information about data maintenance, access conditions, privacy, organizational responsibilities, and possible need for data transformation).

The survey is focussing on the data situation, but it also looks into obstacles preventing statistical institutions from integrating geospatial data and statistics. Special attention will be paid to the challenges found in those member states that are least advanced in terms of using point based reference frameworks. It is therefore essential that all NSIs participate in this survey. The results from the survey will be analysed and presented in a state of progress report.

The questionnaire contains a set of questions regarding existence of geospatial data in member states. You may therefore liaise with your National Mapping Authority or other authorities providing spatial data in your country before answering the questionnaire.

Questions marked with a red asterisk require answers.

IMPORTANT:

After completion of the survey you should be redirected to the page with acknowledgement. It means that the survey has been successfully forwarded.

In order to secure the process of transferring answers from all surveys we kindly ask you before clicking "SEND" (at the end of questionnaire) to print the completed form to PDF and to send it at: a.wardzinska@stat.gov.pl.

If you have any questions about the questionnaire, please do not hesitate to contact Ms. Amelia Wardzińska-Sharif at Central Statistical Office of Poland: a.wardzinska@stat.gov.pl phone: +48-22-608-3833

The survey will be open until 8th of September 2015.

Your contribution to the survey is crucial in order for the project to make an accurate assessment of the situation in the ESS. The project team and Eurostat therefore highly appreciate your commitment!

1. Country*

Enter your answer...

2. Name*

Enter your answer...

3. Institution*

Enter your answer...

4. e-mail address*

Enter your answer...

5. full phone number*

Will be used only in case we will need to clarify your answers. Telephone contacts will always be agreed upon in advance via e-mail.

Enter your answer...

6. Are there any sources of data for which it is possible to retrieve spatial location of its units with coordinate accuracy (points) in your country?

☐ Yes ☐ No

7. If in q6 the answer is no - please specify if it is possible to transform existing spatial registers to acquire spatial data of point accuracy and with what precision such transformation can be carried out:

Enter your answer...

/

8. If answer for q6 is yes, please indicate if the sources of data encompass the following features and to what percentage do they cover your country?

	0-60 %	60-90 %	90-100 %	sources do not cover the following features
coordinates of buildings	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
address point locations	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
location of cadastral parcels	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

9. Are there any other sources, that you think could be used/useful for geocoding statistical information? please specify:

Enter your answer...

/

10. According to your experience does the following two set of point data represent the complete basis for points based statistics: 1) coordinates of address points* and/or building points and 2) coordinates of cadastral parcels?

*in particular all address points connected with buildings, in case of population statistics - address points linked to buildings that contain at least one dwelling.

☐ Yes ☐ No

11. If in q10 the answer is no, please specify other data sources which in your opinion represent the complete basis for points based statistics:

Enter your answer...

/

12. Population data and dwelling data

	Yes	No
Does a register of population (individual persons) exist in your country?	<input type="radio"/>	<input type="radio"/>
Are there registers from which you can obtain information on dwelling locations (in the form of e.g. address points)	<input type="radio"/>	<input type="radio"/>
Can population be linked to dwelling location (geocoded) by means of a standardised identifier?	<input type="radio"/>	<input type="radio"/>

13. Does your NSI have, and on what terms, access to data mentioned in q12?

☐ Yes - entirely or partially free of charge ☐ Yes - on payment basis ☐ No
☐ Other:

14. Building data

	Yes	No
Does a register of buildings exist in your country?	<input type="radio"/>	<input type="radio"/>
Are there registers from which you can obtain information on building locations	<input type="radio"/>	<input type="radio"/>
Can buildings be linked to locations (geocoded) by means of a standardised identifier?	<input type="radio"/>	<input type="radio"/>

15. Does NSI have, and on what terms, access to data mentioned in q14?

- ☐ Yes – entirely or partially free of charge
 ☐ Yes – on payment basis
 ☐ No

☐ Other:

16. Cadastral parcels

	Yes	No
Does a record of individual cadastral parcels exist in your country?	<input type="radio"/>	<input type="radio"/>
Are there registers from which you can obtain information on cadastral parcel locations?	<input type="radio"/>	<input type="radio"/>
Can cadastral parcels be linked to location (geocoded) by means of a standardised identifier?	<input type="radio"/>	<input type="radio"/>

17. Does NSI have, and on what terms, access to data mentioned in q16?

- ☐ Yes – entirely or partially free of charge
 ☐ Yes – on payment basis
 ☐ No

☐ Other:

18. Has any statistical data geocoding process been conducted in your country?

- ☐ Yes (if yes - go to questions 20 and 21)
 ☐ No (if no - go to question 19 and skip questions 20, 21 and 22)

19. what are the main obstacles in your country that prevents introduction of procedures for geocoding of statistical information or administrative data? (If you already have procedures for geocoding, skip this question)

- ☐ No domestic demand for spatial statistics
 ☐ No legal support for spatial statistics
 ☐ Lack of knowledge
☐ Lack of resources
 ☐ Lack of geospatial data or geospatial data have poor quality
☐ Geospatial data are available but too expensive for my NSI
☐ Geospatial data are not available to my NSI because of legal restrictions or bureaucratic reasons (public institutions don't cooperate well).
☐ Administrative data are not available to my NSI because of legal restrictions or bureaucratic reasons (public institutions don't cooperate well)
☐ Administrative data are stored in a way that makes geocoding impossible (lack of identifiers to connect data to geographical locations)
☐ Statistical information is not collected in a way that makes geocoding possible or meaningful
☐ The lack of uniform reference system between different administrative data sources

Other:

20. Please provide scope of data that were geocoded and what was the percentage of the successful combination at national scale

	none	0-60 %	60-90 %	90-100 %
Basic indicators from housing and population censuses (total population, number of dwellings etc):	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
A rich set of indicators from housing and population censuses (E.g. building use/type, age/sex, nationality, educational attainment of population etc):	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Indicators from agriculture censuses	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

21. Please provide scope of other registers or administrative data (business register, population register, tax files etc) that were geocoded and what was the percentage of the successful combination at national scale

(this question is an extension for q20)

Enter your answer...

↗

22. Who conducted the process for data pairing?

- ☐ Process conducted within NSI
 ☐ Process conducted within other institute with responsibility of official statistics
 ☐ Other:

23. Who is responsible for creating and maintaining the key point-based geospatial data that you use for geocoding statistics in your NSI?

(1) Duplication of responsibility may encompass creation and maintenance of several competing data sets due to lack of data coordination or lack of data sharing procedures between institutions (2) Shared responsibility may encompass a shared responsibility for data collection but it can also mean that one institution is responsible for data collection and another institution responsible for centralized data storage and services.

- ☐ 1. The NMCA exclusively
 ☐ 2. The NSI exclusively
 ☐ 3. Other national or subnational institute exclusively
 ☐ 4. Municipalities or regional bodies exclusively
 ☐ 5. Duplication of responsibility(1) between the above mentioned institutions.
 ☐ 6. Shared responsibility(2) between more than one of the above mentioned institutions.
 ☐ 7. No one responsible
 ☐ 8. Do not know
 ☐ 9. Do not use point-based geospatial data for geocoding

24. What is the key point-based geospatial data used for geocoding statistics in your NSI (multiple choices possible)?

- ☐ Address locations from address registers
 ☐ Buildings/dwellings
 ☐ Cadastral parcels
 ☐ Do not have access to any point-based spatial data for geocoding
 Other:

25. If in q23 answers 5 or 6 were chosen please indicate which institutions:

Enter your answer...

11

26. What is the lowest possible geographical level to which you can geocode population distribution information?

- ☐ Small geographical areas such as enumeration districts, blocks or small administrative units
☐ Single points (coordinates) such as address locations, buildings or locations of real cadastral parcels
☐ Combination of both (different data in different parts of the country) ☐ Don't have the ability to geocode information
☐ Other:

27. Is there any legislation which allows OR prevents the NSI data collection with accuracy of the address point?

- ☐ Yes (if yes please specify in 'other') ☐ No Other:

28. Is there a formal policy or informal custom among (public) institutions in your country to use standardized identifiers in registers (address information, personal IDs, real estate codes, building IDs etc)?

- ☐ Yes, to a large extent ☐ Yes, but only to some extent ☐ No ☐ Don't know

29. The ultimate purpose of GEOSTAT 2 project will be to develop a spatial reference framework for statistics. Hence, although the GEOSTAT 2 project mainly deals with address or building point data as part of this framework, we would like to know more about the use and needs as regards other point-based geospatial data or data with other topology that are needed for a full-fledged framework. Which of the following datasets should supplement address and building points to form this spatial reference framework (multiple choices possible)?

This question is particularly relevant for the UN-GGIM: Europe process and the definition of geospatial core data.

- ☐ Cadastral maps ☐ Topographic maps ☐ Orthoimagery ☐ Satellite imagery ☐ Elevation data
☐ Hydrographical data ☐ Land use/Land cover data ☐ Road network ☐ Railway network
☐ Protected areas ☐ Location of public institutes and services (schools, hospitals, universities etc.)
☐ Inventories or administrative data from other institutions captured and stored with coordinates (environmental permits,

Pollutant Release and Transfer Registers, Urban waste water treatment plants)

30. Please indicate which of the data from q29 you already use in production of statistics within your NSI (multiple choices possible)?

- ☐ Cadastral maps ☐ Topographic maps ☐ Orthoimagery ☐ Satellite imagery ☐ Elevation data
☐ Hydrographical data ☐ Land use/Land cover data ☐ Road network ☐ Railway network
☐ Protected areas ☐ Location of public institutes and services (schools, hospitals, universities etc.)
☐ Inventories or administrative data from other institutions captured and stored with coordinates (environmental permits, Pollutant Release and Transfer Registers, Urban waste water treatment plants)

31. If you were to choose, which of the following quality characteristics of the above datasets are the most important ones (max 2):

- ☐ Positional accuracy ☐ Presence of a large scale version (e.g. 1:25 000 road network or better)
☐ Range and accuracy of attributes (e.g. on data sources, names, additional spatial attributes such as height of a building etc.)
☐ Temporal accuracy and well managed maintenance policy

32. What are the main threats in your country related to your current geocoding practice?

- ☐ No threats at all, everything runs smoothly ☐ There are no big problems but there is room for improvement
☐ Weak internal support. The benefits of spatial statistics is contested ☐ Scarce resources
☐ Restricted access to geospatial information needed for geocoding
☐ Inconsistencies in geospatial information needed for geocoding
☐ Poor cooperation and coordination between institutions responsible for different geospatial information and administrative data
☐ Inconsistencies in statistical information or administrative data resulting in bad match with geospatial data
☐ Restricted access to administrative data from other institutions Other:

33. Is there any form of geospatial data correctness verification conducted in your NSI (e.g. post-enumeration survey, pre-enumeration survey [census preparatory works])

- ☐ Yes (if yes please specify in 'other') ☐ No ☐ Other:

34. What is the role of spatial data in your NSI?

- ☐ To improve and facilitate collection of data
☐ To improve quality management (by means of geographical quality assessment and control mechanisms)
☐ To improve data processing ☐ To enhance dissemination and visualisation of data

Other:

35. What are the main benefits of using spatial data in your NSI?

- ☐ To comply with the request to contribute to the EFGS European population grid
☐ To improve production of official statistics
☐ To provide spatial statistics as data services for external usage (governmental institutions and private sector)
☐ To provide more qualified analysis services for external users

Other:

36. Imagine an ideal situation regarding infrastructure for production of spatial statistics in your country and NSI (with respect to technical solutions, legal situation, financial and human resources, access to data, quality of data, cooperation with other institutions etc). How close or far from this ideal situation are you today?

1 indicates that you have an ideal situation and 5 indicate that you have a long way to go.

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

37. Are you aware of the REGULATION (EU) 2015/759 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 29 April 2015 amending Regulation (EC) No 223/2009 on European statistics that gives NSI stronger rights to access and use administrative records?

http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=uriserv:OJ.L_.2015.123.01.0090.01.ENG
(http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=uriserv:OJ.L_.2015.123.01.0090.01.ENG)

- ☐ Yes, I am aware of it and I could imagine using this article to obtain better access to address registers or other administrative records useful for geocoding.
- ☐ Yes, I am aware of it but I don't think this article will help improve access to address registers or other administrative records useful for geocoding.
- ☐ No, I am not aware of this article.

38. What kind of support does your NSI need in order to strengthen infrastructure for production of spatial statistics?

Enter your answer...

⌵

39. IMPORTANT: After completion of the survey you should be redirected to the page with acknowledgement. It means that the survey has been successfully forwarded. In order to secure the process of transferring answers from all surveys we kindly ask you before clicking "SEND" (at the end of questionnaire) to print the completed form to PDF and to send it at:
a.wardzinska@stat.gov.pl

REMINDER

Enter your answer...