



GEOSTAT 4, Project Meeting, 6 October

Venue/time:

On-line meeting, 6 October, 12-14 CET.

Participants

Rina Tammisto, Amelia Wardzińska-Sharif, Ana M. Santos, Anna Slawinska, Antti Santaharju, Arnaud Degorre, Hanna Brenzel, Igor Kuzma, Jennika Leino, Jerker Moström, Julien Gaffuri, Jørn Kristian Undelstvedt, Kathrin Gebers, Marianne Vik Dysterud, Marie-Hélène Kerouanton, Marina Backer Skaar Mervi Haakana, Nikolaos Roubanis, Karin Hedeklint, Outi Ahti-Miettinen, Pekka Latvala, Rossano Figueiredo, Simen Sommer, Tapio Kytö, Thomas Burg, Timo Aarnio, Tuuli Pihlajamaa, Vilni Verner Holst Bloch, Vincent Loonis

The purpose of the meeting

A monthly GEOSTAT4 project group meeting to review the situation of the work packages and to inform members about ongoing matters.

Agenda and notes

Announcements

Nikolaos Roubanis announced that he will be transferred to new duties in a month and his journey with us will come to an end. Nikolaos introduced his successor, Julien Gaffuri, who will start working with us in the GEOSTAT4 project. Nikolaos promised that he will not forget us entirely and can be contacted if needed. The GEOSTAT4 group thanked Nikolaos wistfully for all the help he has given us and congratulated of the new opportunities. The group also welcomed warmly Julien to work with us in GEOSTAT4.

Project management (budget, interim report, outputs)

In the last meeting we discussed about opening the budget to transfer the saved travel budgets to be used in some other ways. It seems that this is not relevant at the moment and we will consider this again in next autumn if needed.

In the halfway of the project we should deliver an interim report. This report should be ready by the end of February 2021. For the report all WPs are asked to gather the working days from the beginning of the project to the 31 December 2020. This information is needed by the end of January 2021. The WP leaders are also asked to look at how the Covid-19 has influenced the work from three perspectives:

- The work of the work package
- For the WP leader country and contacts to other members inside the WP
- For the contacts to other members in GEOSTAT4 project and other stakeholders

Two outputs of the project have been already delivered, from WP2 the Benchmark study and from WP6 the Communication plan. Rina apologised that these were delivered late, she first sent them only to Nikolaos, but they should have been uploaded into the portal as well.

Thomas asked if there is a common template for the GEOSTAT4 documents, but we do not have this yet. In the WP6 it is possible to have the needed templates derived by the graphic designer.

Other announcements

No other announcements

Notes of the previous meeting

Notes of the previous meeting have been sent and members are asked to comment by 9 October. After that the notes will be published.

WP output reviews

WP1, Completed GSGF Europe Specific requirements – Big Data sources, Arnaud Degorre

Arnaud Degorre presented his study about the specific requirements considering Big Data in the GSGF. The starting point is that Big Data is only indirectly handled in the GSGF, it is mentioned in some requirements but seen usually as an alternative data source alongside administrative data. The differences between Big data and administrative data can be reflected on the 3Vs model, that describe the dimensions that characterise the challenges and opportunities of large data: Volume, Velocity and Variety.

The volume is not an issue here, both administrative and big data sources contain large data sets, but the velocity and variety may concern the work with GSGF. The administrative data is structured data, but Big Data brings the challenge with its unstructured nature. Both will need to be mapped to statistical standards. The velocity may also be a key aspect, it is specific to Big data processes based on continuous stream of data.

In the presentation Arnaud also observed the Big Data from the GSGF principles point of view and listed what kind of new opportunities as well as challenges the Big Data would bring for geocoded data, consistent semantics, data quality and the workflows of statistical production. For example, ready geocoded and timestamped information could be gotten from IoT (internet of Things) based Big Data and the data collection could turn more data rich and continuous. But statistical community may have to review its core concepts and definitions used so far to describe statistical unit in a survey-based paradigm, as continuous measurement of the change in time and space may lead to measure differently, and even adopt new definitions, of spatial phenomenon: for instance, measurement and even definition of home-to-work shuttles are quite different in a traditional survey and continuous GPS-based data flow. Other topics are also to be covered regarding GSGF, as new scale or frequency for output statistics and looped cycle of production improvement. The challenges relate Big Data being as it is and that it cannot be easily harmonised or defined in used data models. The quality can vary considerably and sustainability of the systems can be out of hands.

Arnaud raised questions that would be relevant to discuss together:



- Is it relevant to include Big Data as non-conventional sources to be addressed in GSGF Europe requirements?
- Is it OK to focus on Big Data velocity in general, and more specifically to look at continuous streaming of data / real-time statistics?
- Are requirements we have pointed out so far are adequate? Are there other Big Data aspects to be analysed?

The discussions confirmed the interest of investigating the case of Big Data. While aiming to limit the number of new recommendations to keep the GSGF Europe at a general level, WP1 will establish a selection of items covering the challenges for geospatial big data statistical pipelines. A summary note will be proposed for the consortium's proofreading in early November.

WP1, Specific requirements for geospatial statistics based on survey data or non-traditional data, Ana Santos

Ana Santos related in her presentation how the Statistics Portugal has utilised the National Dwellings Register as a primary source for the Social Surveys since 2012. The register consists of continuously updated list of buildings and dwellings that were collected in the 2011 Census. The information on this register comes from fieldwork as well as from the Indicators System of Urban Operations, which is the main source of information on new buildings/dwellings and demolishes.

Based on the experiences in Portugal, Ana presented that the specific requirements for survey data or non-traditional data based geospatial statistics should focus on the geocoding process. The requirements for geocoding are classified in three thematic area: methodological, practice and guidelines and institutional. The methodological requirements consist suggestions for example, using standards for the location references, the use of unique identifiers and development of tools, repositories and platforms to enhance the processing of geospatial data.

In the practice and guidelines section, which should be defined in the European level, is described for example the need for technical data and metadata management practices, privacy and confidentiality practices and the development of common solutions for utilising the data and tools.

The institutional requirements describe the need of collaboration between different organizations and exchange of statistical and geospatial data in the scope of cooperation models, references and taking the laws and policies into account.

WP3 Draft output of the proposal for changes in the Quality Assurance framework, Thomas Burg

Thomas Burg presented the ESS Quality Assurance Framework (QAF) document and what kind of changes or enhancements could be done from the geospatial point of view. The QAF include various types of data, for example, survey and administrative data. The geospatial data is not mentioned specially and therefor is not acknowledged appropriately. In this task the aim is to bring the geospatial aspects to be considered on the quality framework.

The task is challenging because the existing principles are quite general, and the geospatial additions will go in more detailed level. There are some suggestions on

the wordings, like including the terms “geospatial data” or “and for geospatial data” on needed items and larger changes suggesting on new methods to cover the geospatial elements in the indicators. Another challenge is the terminology, the quality and geospatial experts each have their own standards where the terms may vary.

The document can be found from Teams and the project members are welcomed to comment the suggestions before they will be presented to the quality group.

Current situation in work packages (+Planner tasks)

WP1, Mervi Haakana

The schedules and roles of WP1 have been reviewed and cleared. In the discussions it was noted that a workshop would be a good idea to drive the work forward. A summary of the outputs so far and the connections to other work packages is also worth to produce.

WP2, Jerker Moström

At the moment the good practices cases are under way. The purpose is to find a good method to get the most out of them and how to publish the findings in a relevant way. The work with the Table Joining Service is also on-going and could be demoed some time. The task about the business cases is a bit late but Jerker will contact to arrange the start-up meeting.

WP3, Thomas Burg

The quality report has been started and suggestions for quality indicators are under way. The SIMS (Single Integrated Metadata Structure) is now studied as well.

WP4, Anna Slawinska

In the EFGS web page the webinar site was updated with latest information and the preliminary agenda. In the news there will be published the reminder to register for the webinar.

WP5, Amelia Wardzińska-Sharif

For the EFGS webinar a good number of abstracts were delivered and of them the presentations have been chosen. The agenda has been published, although some presenters have not confirmed yet. This is fine, because there are still left some extra presentations that can be taken, if needed. For the registrations there a still a couple of days left, so more participants can still be expected. The tests with the webinar platform will start next week.

Rina pointed out that she and Antti will present this project in the webinar too.

WP6, Igor Kuzma

The GEOSTAT logo plans have been reviewed in a small group and the designer will continue from there. Igor will upload them to Teams, so everyone can check them.

Other issues

Anna reminded that the EFGS conference 2021 will be in 7 and 8 September 2021.

Next project meeting

The next meeting will be on 30 October.