Understanding the challenges in creating pan-European datasets

Carol Agius, EuroGeographics
Fourth International Workshop on Spatial Data Quality
October 2023

eurogeographics

eurogeographics

Eurostat service contract (2023 – 2026)
Task 2

- Deliver effective topics groups to support the progress of NMCAs in addressing existing obstacles in the availability of pan-European datasets

eurogeographics

Knowledge Exchange

- Barriers to seamless pan-European datasets
- Understand members experience in topic areas
- Organise knowledge exchange activities:
  - Learn from members experiences
  - Members' capacity development
  - Feed into OME2
  - Raise awareness of OME2 production process and tools

Generalisation

Edge-matching

Quality

Life-cycle management

Harmonisation

The objectives of the TDKEN are:

- Establish a network of experts and expertise in the topic areas
- Increase the understanding and capacity in the topic areas in member organisations
- Establish current situation for each topic group in various countries, as well as at the pan European level
- Identify and note key challenges in the topic areas
- Share knowledge, national experiences and good practices on topics areas

eurogeographics

TDKEN Outputs

Output description
State of play document; which will detail the current situation for each topic group in various countries, as well as at the pan European level. It will highlight current good practices and note key challenges, both at the national and Pan-European level
A proposed plan / terms of reference for each topic group, detailing the proposed methodology and approach to progress the topic area

Item	Description	Outputs	Date
State of play	Draft baseline state of play questionnaire	Baseline questionnaire	May – June 2023
	Launch questionnaire		July – August 2023
	Initial analysis of baseline questionnaire	Initial identification of current state of play and gaps	September 2023
	Present responses at SDQ Workshop		Oct 2023
	Develop proposed plan and more detailed terms of reference, detailing the proposed methodology and approach to progress in each of the topic areas will be developed.	Detailed TORs	December 2023
Webinars	Launch call to members for presentations on topic areas, to share experiences and current practices	List of potential presentations	June 2023
	Schedule and present webinars	Webinars	Sept 2023 – March 2024

eurogeographics

Baseline survey - methodology

- Aim to establish a baseline of the state of play in order to asses and establish current situation for each topic group in various countries, as well as at the pan European level.

- Drafted questionnaire in SurveyMonkey

- Asked questions on all five topic areas, mainly open-ended questions

- Sent to all EuroGeographics PCs in July with September deadline

- Received 20 responses

www.eurogeographics.org

eurogeographics

Generalisation

- Investigate generalisation tools (tools for 10k to 50k,100k, 250k)

- Consider the feasibility of generalizing EBM and ERM from the large-scale dataset.

- Investigate methods to maintain and improve the generalization process to produce EGM.

Highlights

Do you use generalisation tools in your production processes?
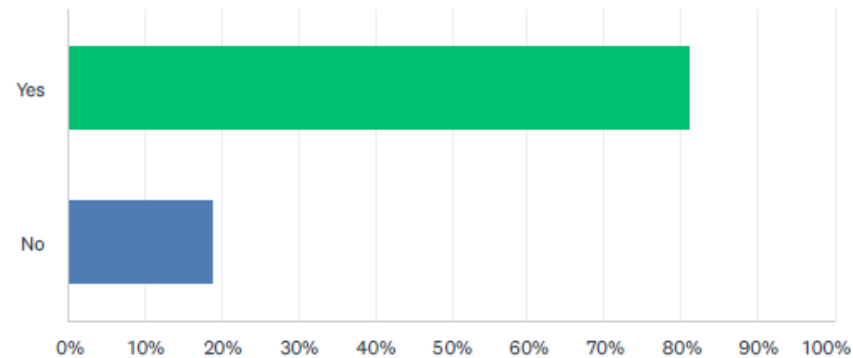
Answered: 17   Skipped: 3
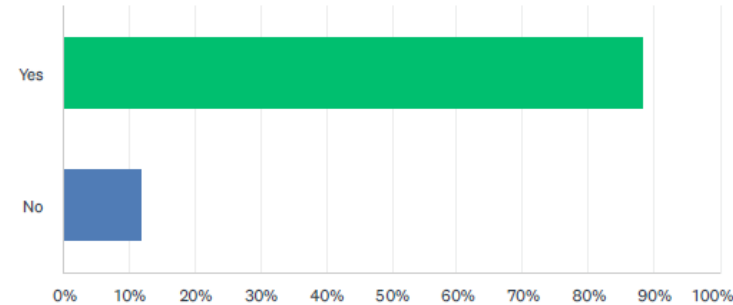
- Half of the responses indicate that the most common source scale is 1:10K to produce products at scales from 1:25K – 1:3M while about 25% of responses use 1:5K

- By far is the generalisation of vector data to produce topographic data sets (buildings, roads, water, infrastructure, terrain, and admin boundaries)

- ESRI the most common tools used for generalization with some use of FME and open source and a few developed in-house tools

- Most do not use raster generalization, and those that do use it for vegetation, rock relief and hydrography

- Concerns include generalization is not easy (complex) requires manual intervention and quality loss. One response mentioned looking into database of rules for generalization as a base for AI.

eurogeographics

Edge-matching

- Investigate and develop the tool required to edge-match large-scale data along the international boundaries according to a specified connecting features dataset. The aim is to ensure topological consistency of the data across borders.
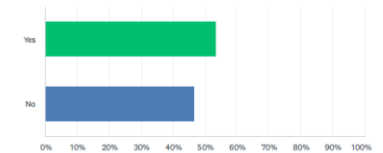

Do you use edge-matching in your production processes?
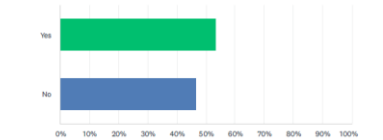
Answered: 16   Skipped: 4

Do you perform any edge-matching along international boundaries except within the production of pan-European datasets for EuroGeographics?

Answered: 15   Skipped: 5

Are there discussions or processes in place with your neighbouring countries at master level for edge-matching cross-border data?

Answered: 15   Skipped: 5

- Most edge-matching is internal (divided by sheets or municipality) while just over half responses mentioned edge-matching along national boarders with neighbouring countries – mainly with agreements on common points

- Almost 50/50 use of ready to use automatic or semi-automatic tools and manual methods

- Not many issues, noted that the work is done with care

Highlights

Harmonisation

- Investigate and develop a model conversion tool based on mapping tables between national and large-scale specifications

- Create an inventory of geometrical homogenisation cases to process

- In cooperation with users set up a tool to homogenise geometrical representations to fit users' needs

- Includes cross country comparability

Highlights

Do you encounter similar situations in your national production processes where you need to gather data coming from different producers or sources?
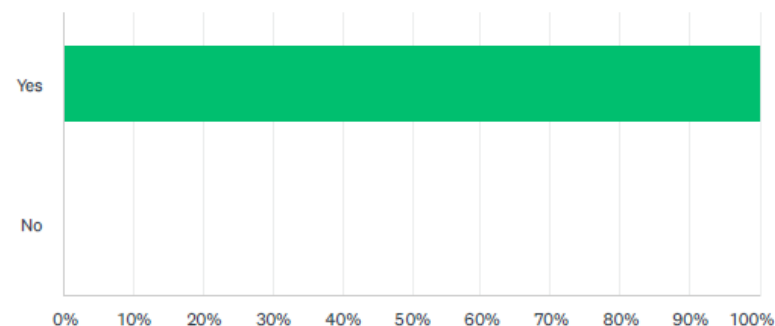
Answered: 15   Skipped: 5

Yes
No

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
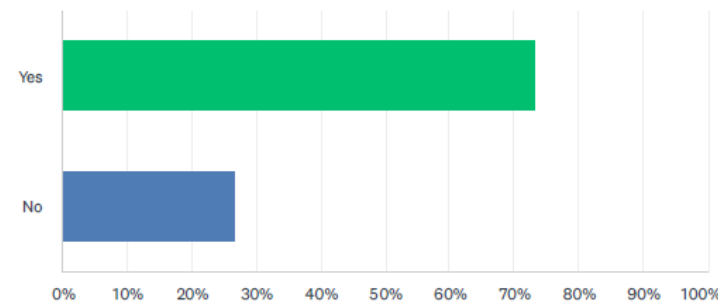
- Different sources internal (different departments), different orgainsations, municipalities and data processes

- Mainly manual processes using ArcGIS, FME

- All sorts of different datasets – admin boundaries, transportation, buildings, protected areas, utility, water and governmental services

- Not many issues highlighted – but those that did said it is laborious

eurogeographics

Quality

- Investigate and develop a tool that checks the quality of the datasets based on the specifications

- Consider the reporting and management of quality checking based on the stakeholders involved in the process

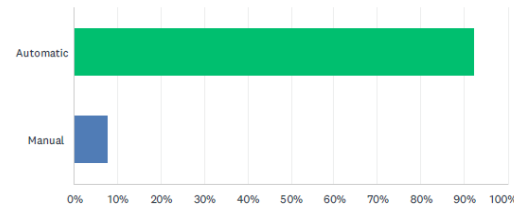- Includes quality control and correction

Highlights

Do you implement quality control methods in your production processes?
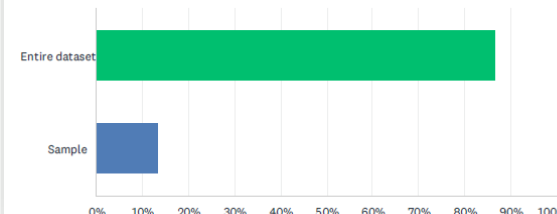
Answered: 15   Skipped: 5

Q27 Are the quality control methods automatic or manual?

Answered: 13   Skipped: 7

Q28 Are they carried on the entire dataset or a sample?

Answered: 15   Skipped: 5

- Quality checks mostly on topology, attributes and content
- Built into the process and also in some cases at the end of the editing process
- Almost half used standard commercial tools and just over a third used self-developed tools
- Mainly automatic checks on the entire datasets not a sample

eurogeographics

Life-cycle management

- Investigate and develop the implementation plan for an integrated life-cycle management system based on the experience of IGN-F within its own national activities, taking into account developments and experience from other NMCAs.

- Includes Incremental updating and continuous updating. *As stated in Eurostat's call for tenders, incremental updating concerns the capacity to disseminate incremental updates, that is only the changes between two consecutive versions of a dataset. Continuous updating concerns the capacity to improve the update frequency (how often new updates are published) and update delay/timeliness (the delay between the update reference date and the publication date, which should be as small as possible so that the dataset describes the current situation)*

Highlights

Do you implement object life-cycle management in your production process?

Answered: 14   Skipped: 6
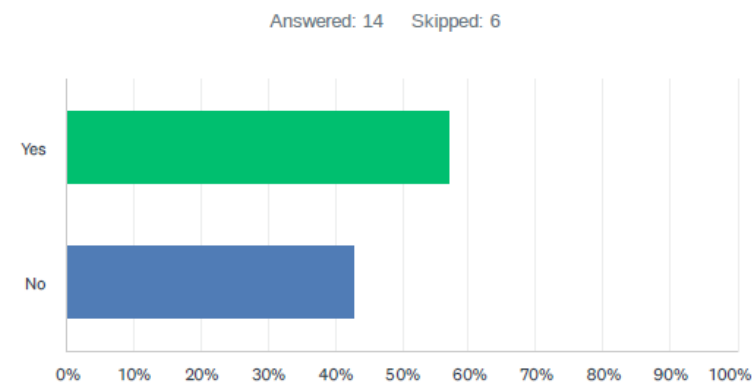
- Update cycle depends on the type of data set – topographic data anything between 1 – 5 years (and even 10 years)
- Cadastral data less – continuously to 6 months
- Those who implement it objects have life-cycle attributes

Next steps

- Anonymise the survey responses
- Identify the gaps and issues for each of the topic areas
- Organise webinars for members and stakeholders to agree how to progress the knowledge and learning of the topic areas
- All invited to contribute