

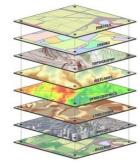
# Some activities on geographical data quality management at Eurostat GISCO

Julien Gaffuri - Eurostat

4th International Workshop on Spatial Data Quality, 11-12 October 2023

### GISCO – GIS at the Commission





GISCO is a permanent service of Eurostat that answers the common needs of Eurostat and the European Commission for geographical information at the level of the European Union (EU), its Member States and regions.

- Provision of GIS (reference) data, services and software,
- Support cartographic and spatial analysis activities,
- Stimulate the use of GIS to support commission activities,
- Support Eurostat activities on the integration of statistical and geospatial information.
- <u>https://ec.europa.eu/eurostat/web/gisco</u>





#### Outline

- 1. Quality requirements
- 2. Quality control
- 3. Quality influence on spatial analyse



## Quality requirements



### Quality requirements

 Thematic coverage: Base topographic datasets on buildings, ground infrastructure, transport networks, land use and cover, hydrography, orography, administrative boundaries, geographical names, POIs, etc.



BD-L-TC, ACT Luxembou

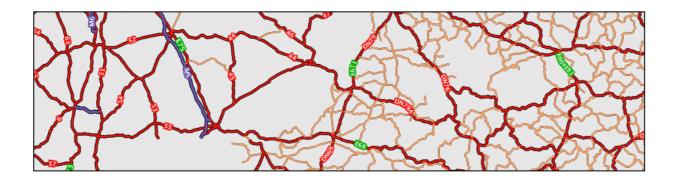
- Requirements for completeness, positional accuracy, thematic accuracy, temporal quality (timeliness and update frequency, versioning with persistent identifiers).
- Call for tenders ESTAT/2022/NP/0010-GISCO acquisition of topographic data layers.
- EU context: Specific quality requirements



### Quality requirements – Specific



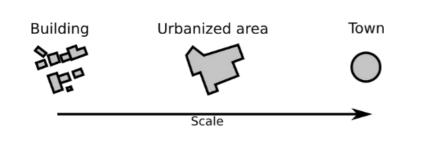
- Geographical extent, comparability
  - Several countries, cross-border regions or the entire European territories
  - Homogeneity of the quality across space
  - Topological consistency at border (with edge-matching)

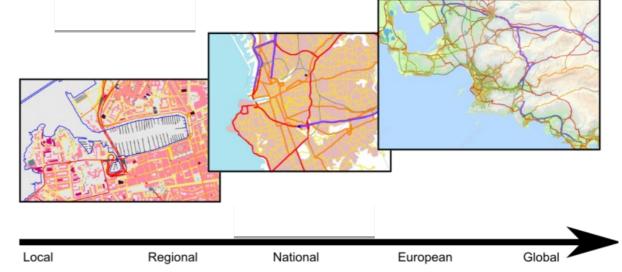




#### Quality requirements – Specific

- Need for detailed data (1:10k) <u>and</u> generalised data (1:50k, 1:100k, 1:250)
- Need for multi-scale data derived with AI-based automated generalisation.





#### Quality requirements – Specific

- Meta-quality
  - Quality must be known, measured, documented.
  - Stability of quality across versions.
- Sustainability need for governance and resources.



# Quality control



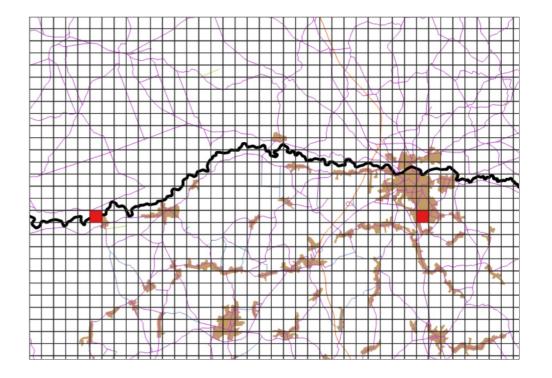
#### Quality control at Eurostat-GISCO

- Structure and geometry validity
- Completeness
- Topological consistency noding and edge-matching
- Timeliness
- Generalisation



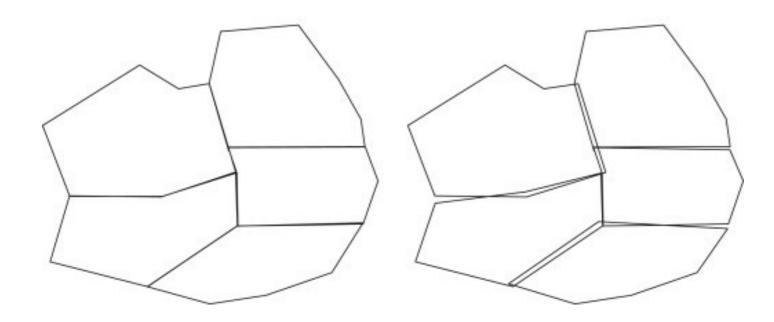
#### Quality control – Completeness

- Assess the number of omissions and commissions for some feature classes.
- Pseudo-random sampling of 1km grid cells:
  - 30 grid cells (one per country)
  - 17 cross-country grid cells
- For each grid cell:
  - Cross-source comparison



### Quality control – Topology – Tessellations

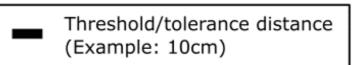
• Topology validation: No gap – no overlap



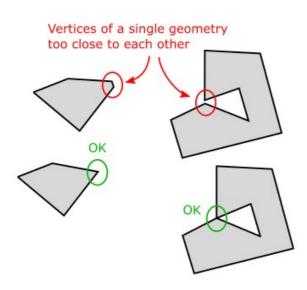


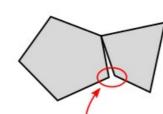
## Quality control – Topology – Tessellations

• Topology validation: Strict noding

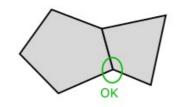


#### PointPoint issue

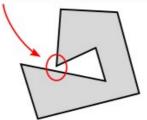


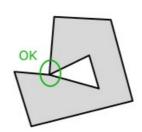


Vertices of a different geometry too close to each other (but not exactly at the same position)

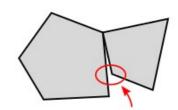


A vertex of a single geometry too close to a segment/line of its outline

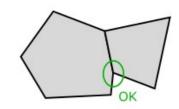




#### PointLine issue

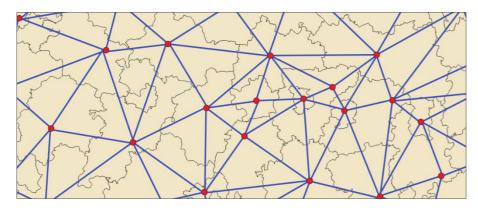


A vertex of a geometry is too close to the segment/line of another geometry outline



## Quality control – Topology – Road network

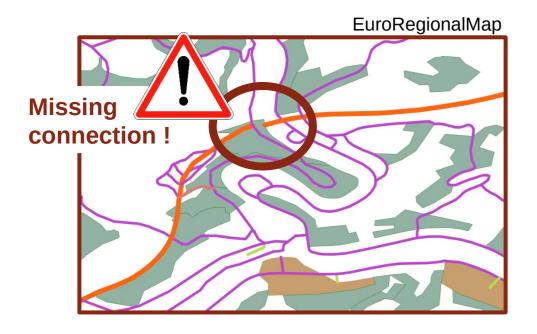
- Origin-distance matrix on a sample of 2053 locations.
- Driving time for Delaunay triangulation segments.



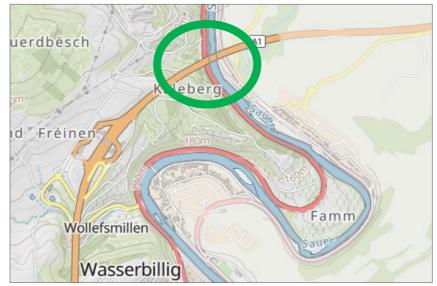
- Comparison of results based on different datasets and versions of them.
- Outlier detection
- Give insight on: Edge-matching issue, thematic accuracy, completeness.



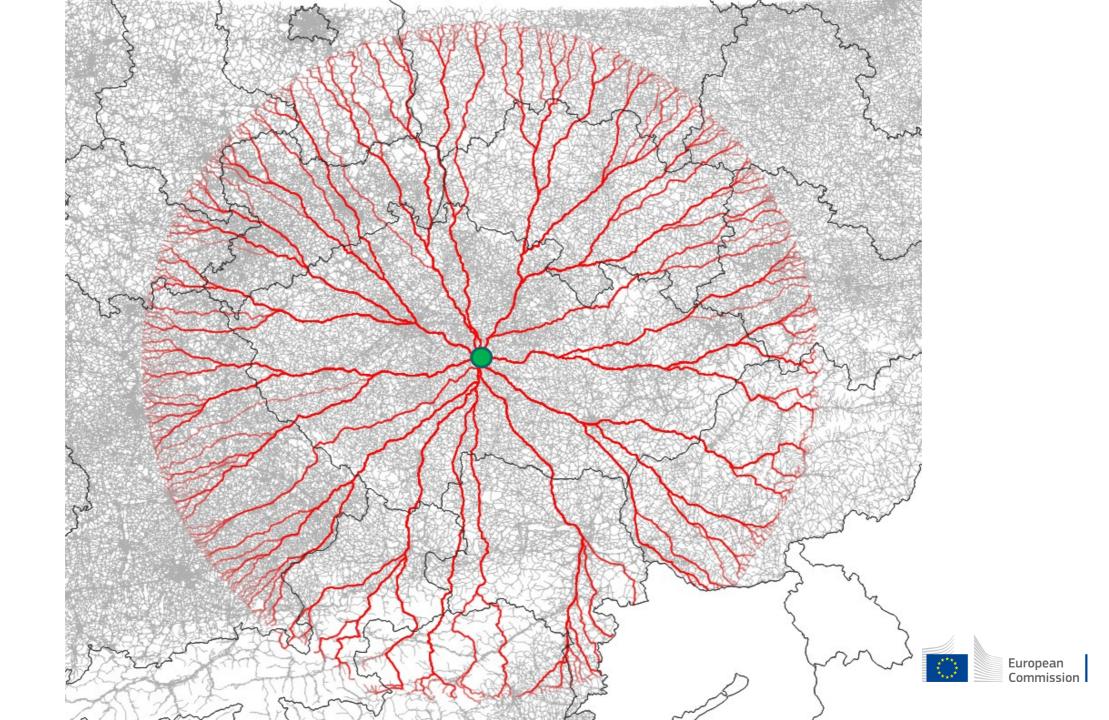
### Quality control – topology – Road network



BD-L-TC, ACT Luxembourg





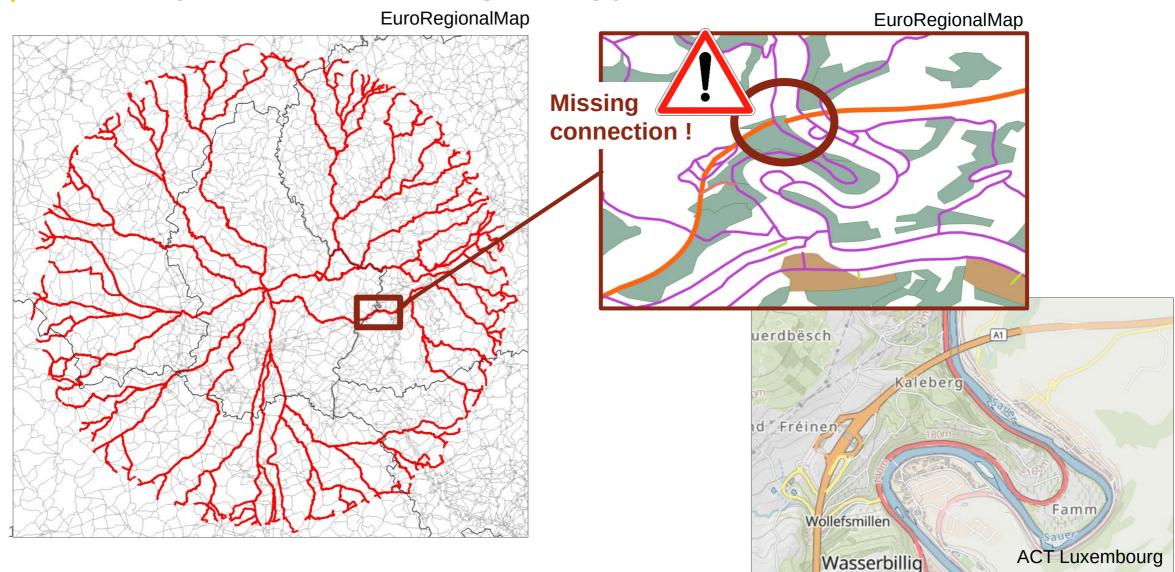


#### Quality control – topology – Road network

EuroRegionalMap **Missing cross-border connections !** 

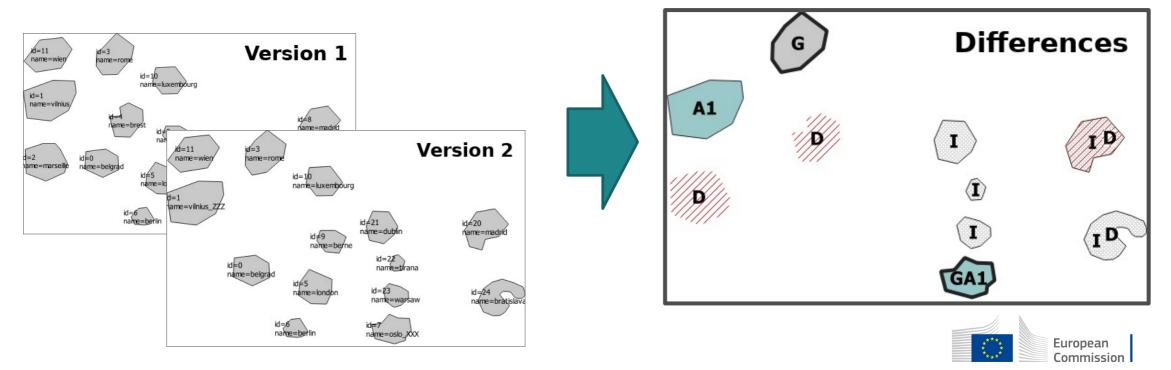


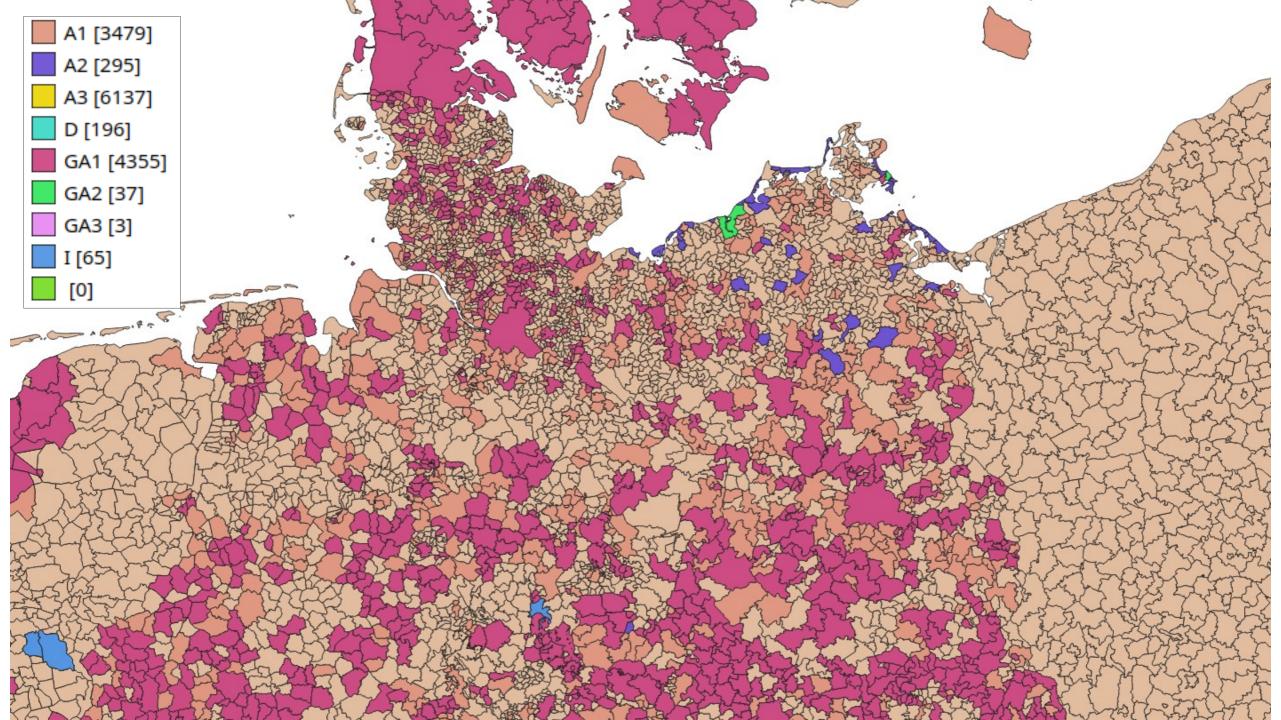
#### Quality control – topology – Road network



#### Quality control – Timeliness

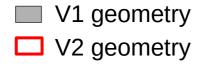
- Compare two consecutive versions of a dataset analyse the changes.
- Based on GeoDiff tool (<u>https://github.com/eurostat/GeoDiff</u>)
- Check identifier stability

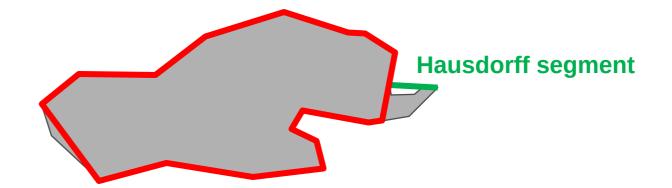




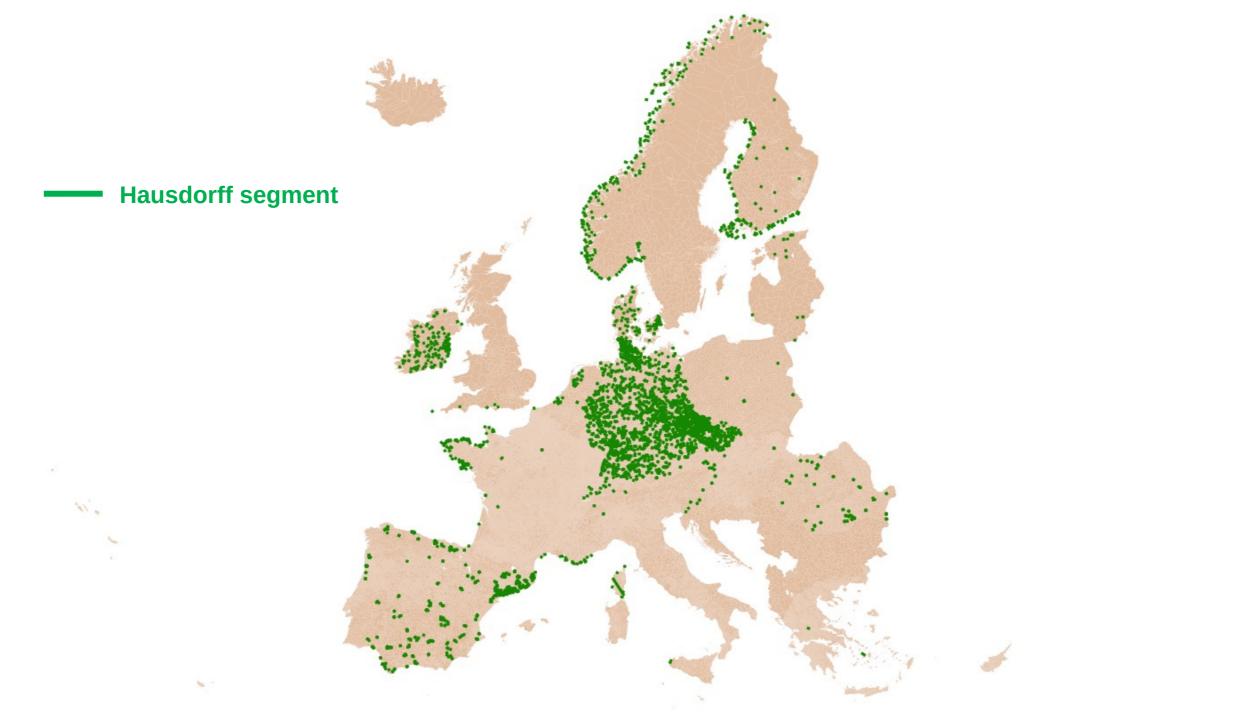
### Quality control – Timeliness

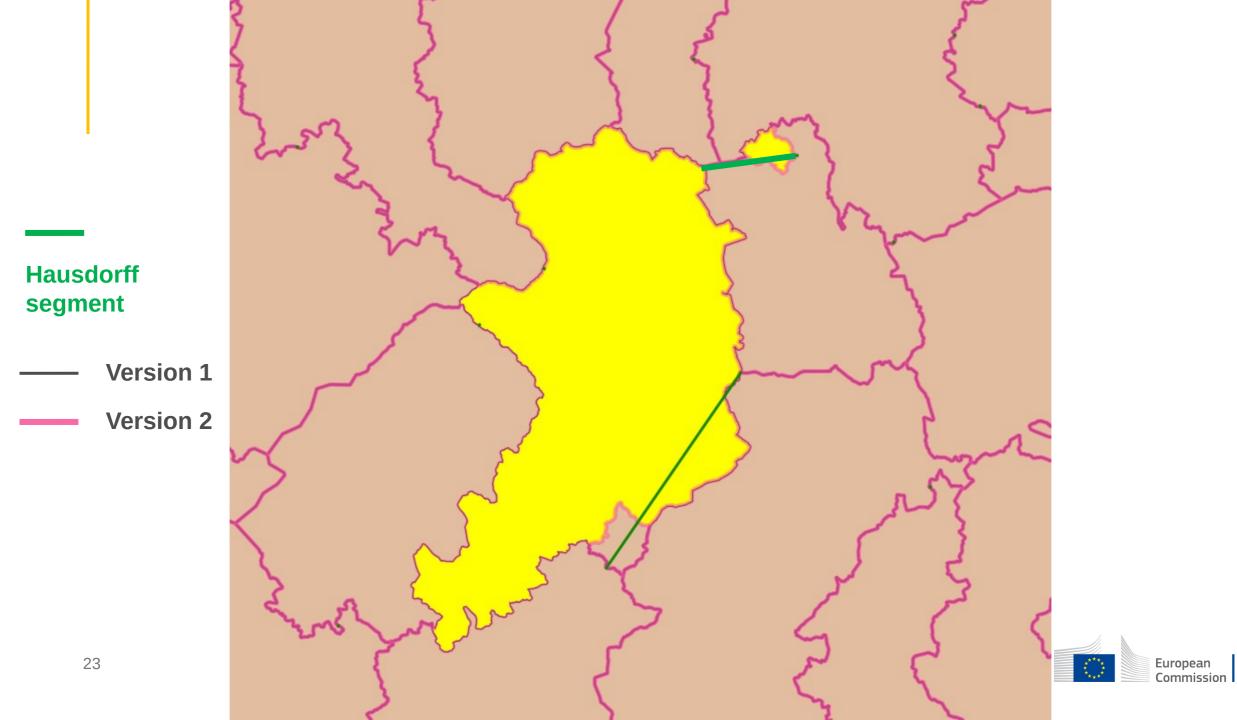
• Analyse geometrical changes, based on Hausdorff distance.

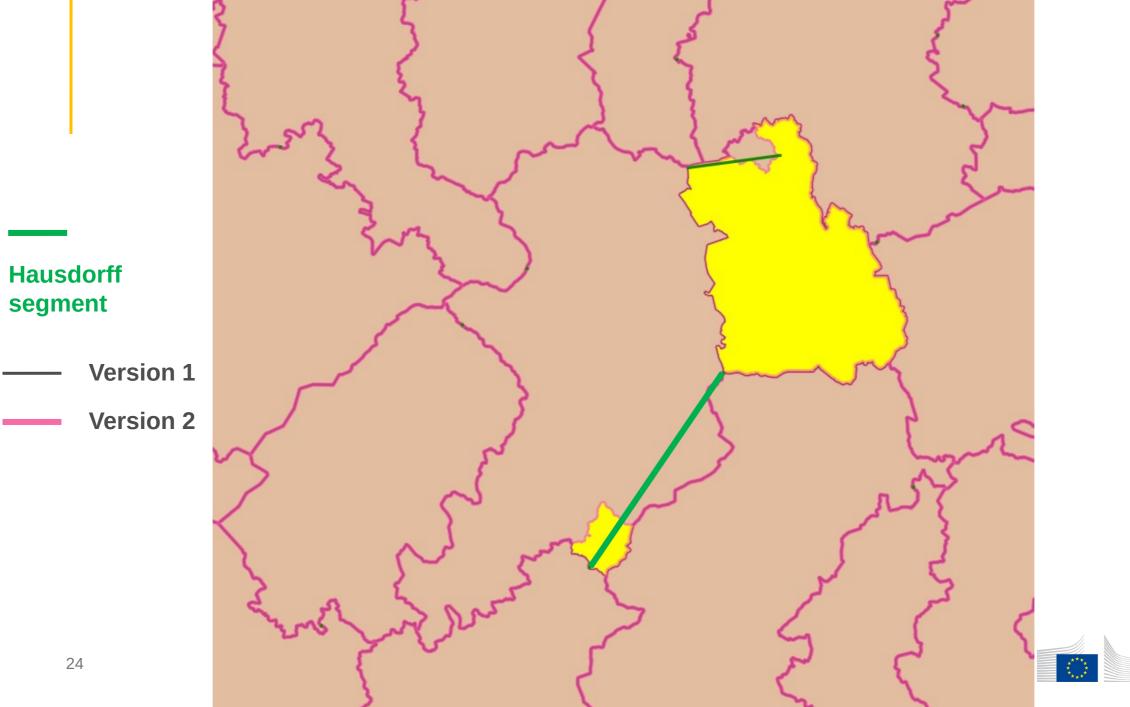




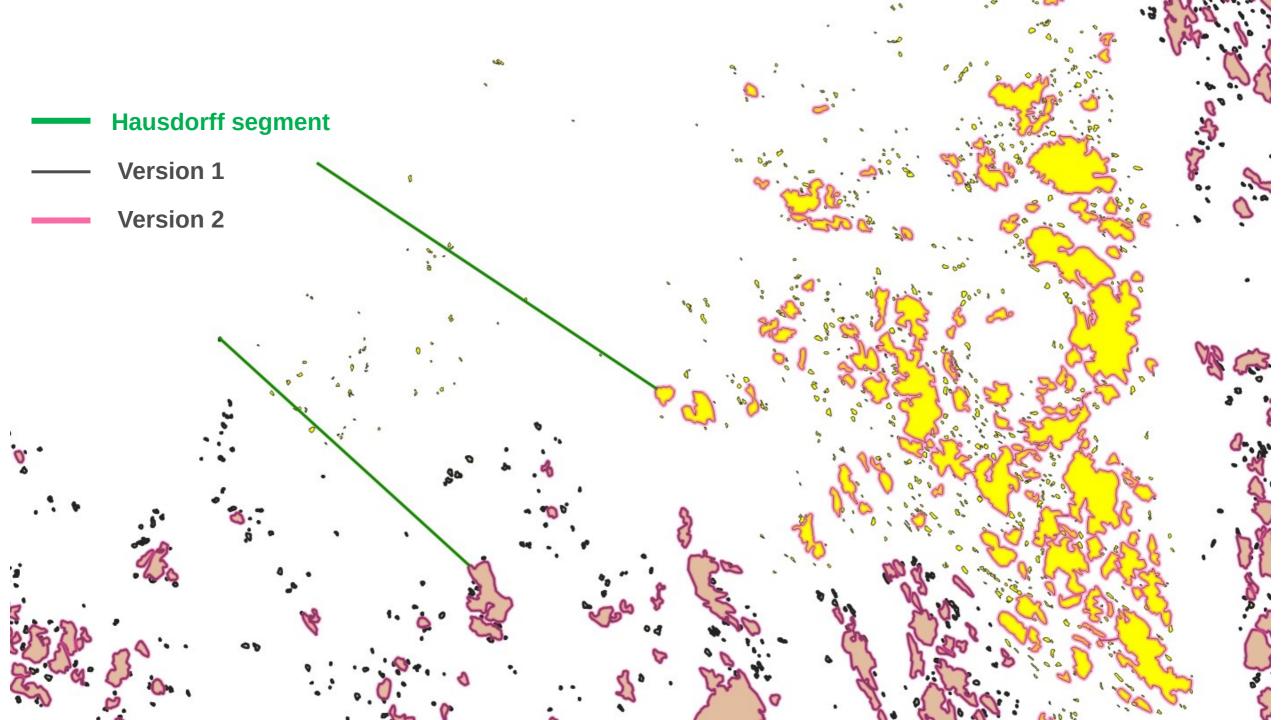


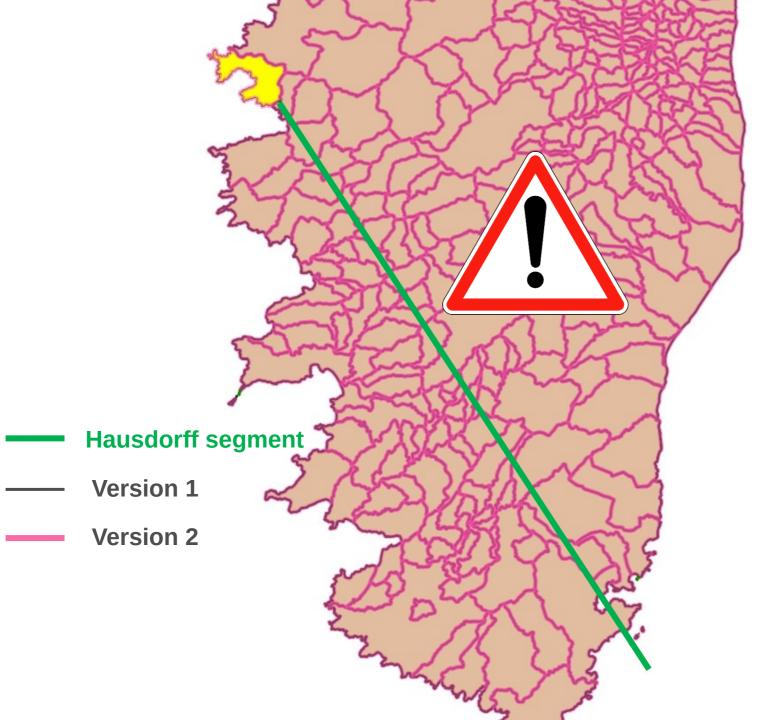






European Commission





#### Identifier stability issues

5

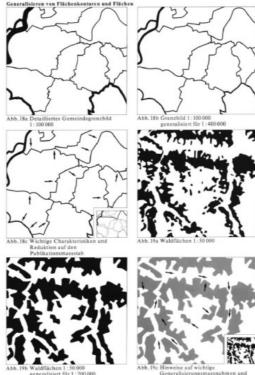
÷

Features which did not change (significantly) from a version to another must keep the same identifier.

#### Quality control – Generalisation

- Geometric level of detail: Validation for administrative units.
- Minimum mapping unit / distance

Scale	Resolution in map mm	Resolution in ground meter
1:1M	0.2mm	200m
1:3M	0.2mm	600m
1:10M	0.2mm	2km
1:20M	0.2mm	4km
1:60M	0.2mm	12km









Oberbegriff Getreide zusammengefasst



Abb. 14b Legende zu den Abbildungen 14





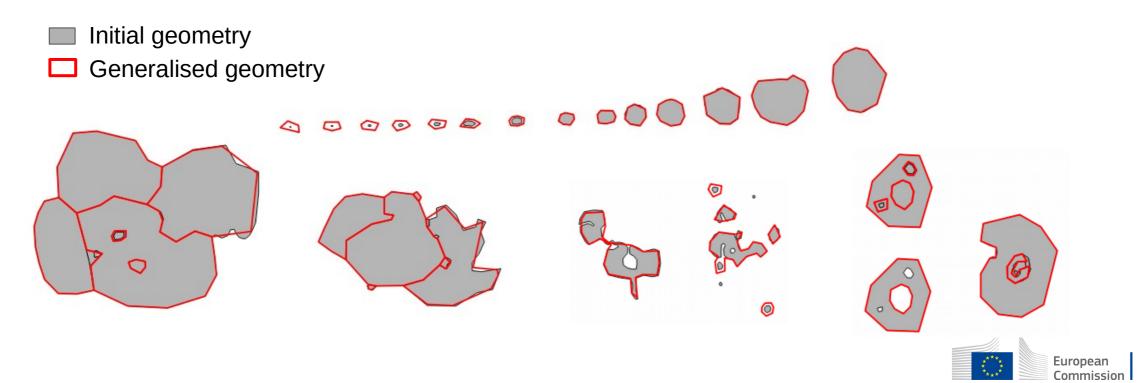
Abb. 15 Reduktion der Abbildungen 14 auf den Publikationsmassstab



Abb. 14d Zusammengefasstes Getreide geometrisch generalisiert

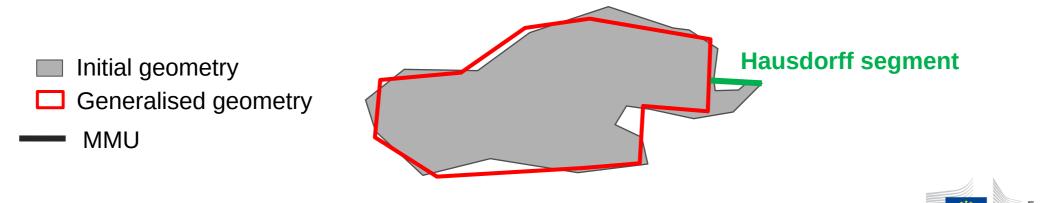
#### Quality control – Generalisation

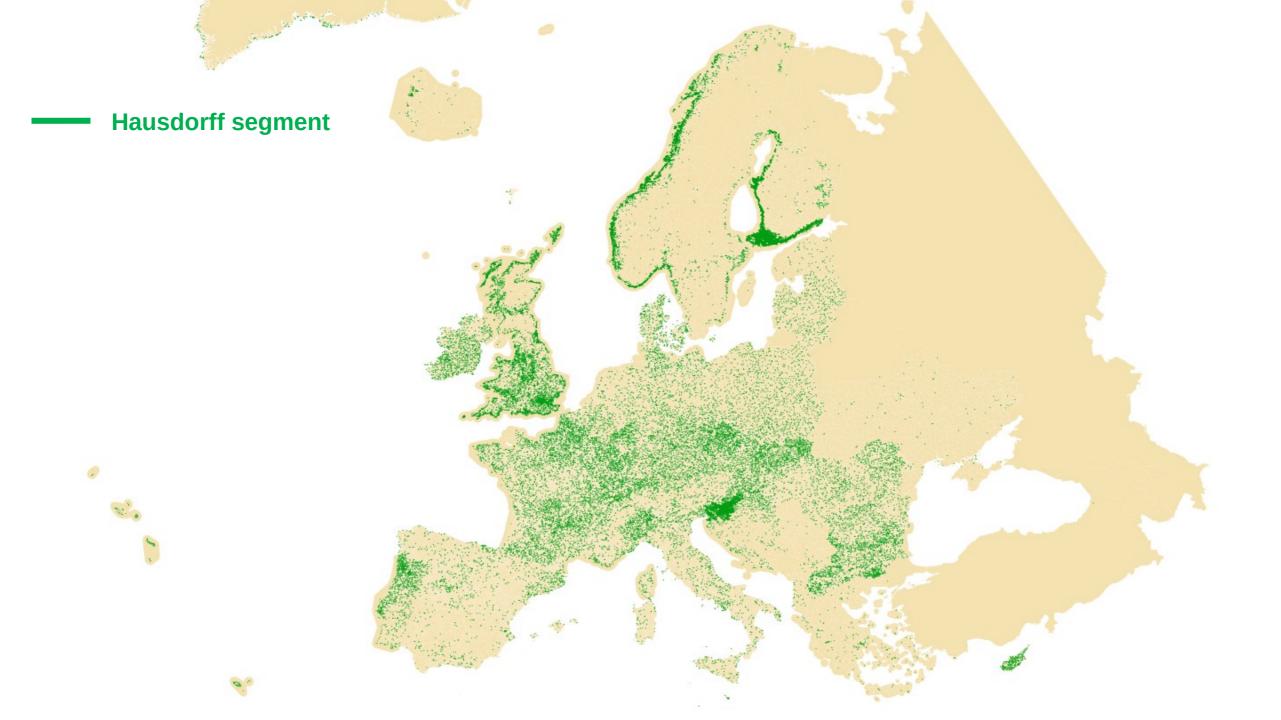
- Geometric level of detail: Validation for administrative units.
- Based on RegionSimplify tool (<u>https://github.com/eurostat/RegionSimplify/</u>)

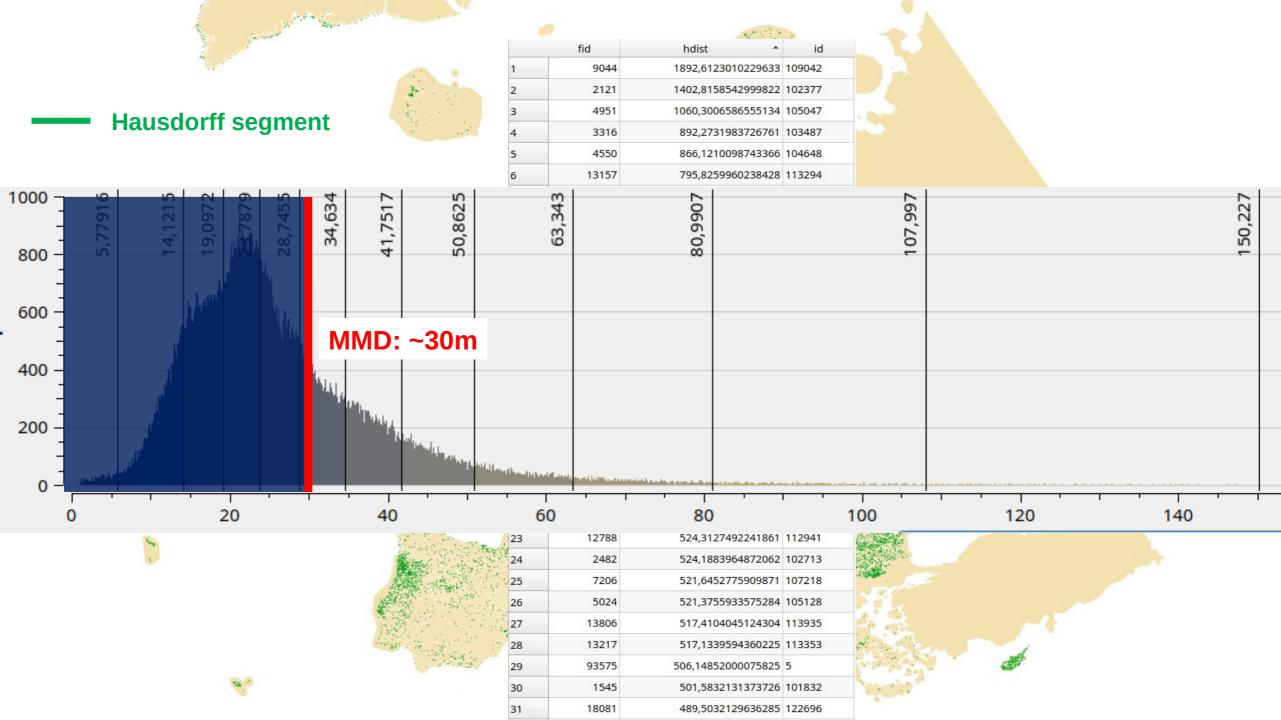


#### Quality control – Generalisation

- Geometric level of detail: Validation for administrative units.
- Process:
  - Run **RegionSimplify** on the dataset to be validated.
  - Compare the outcome with the dataset to be validated, with **GeoDiff**, based on Hausdorff distance.

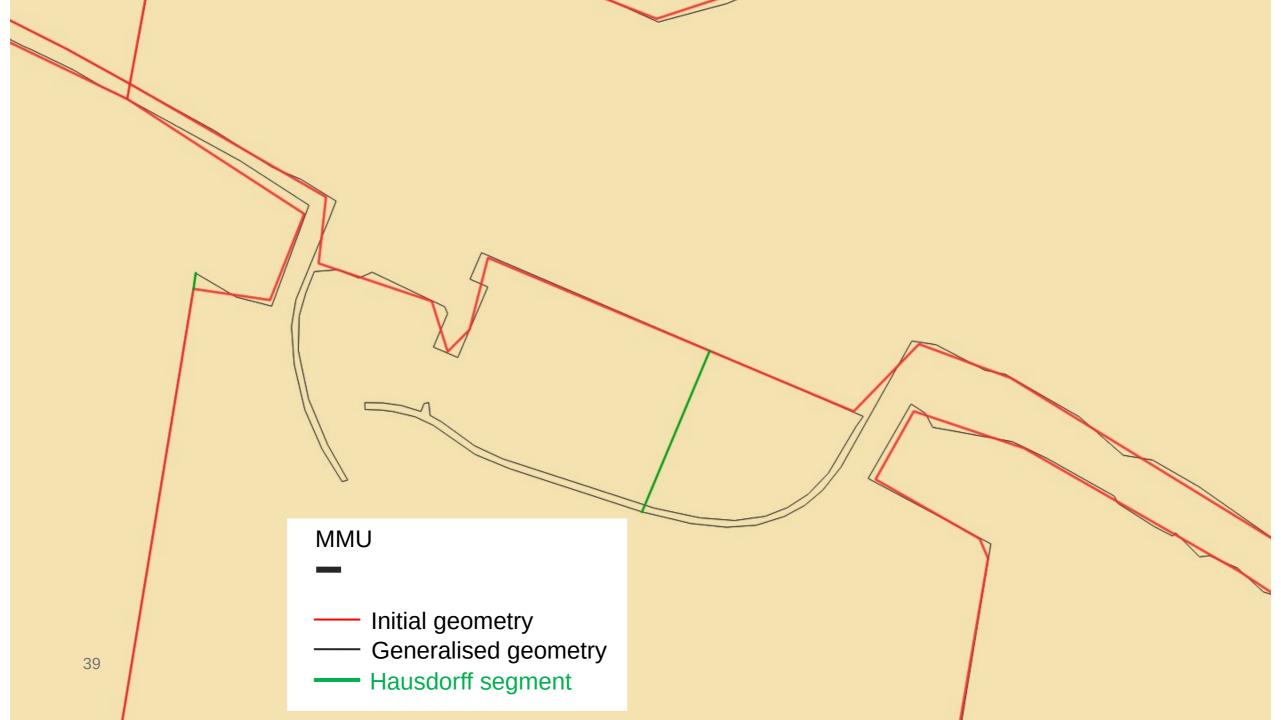


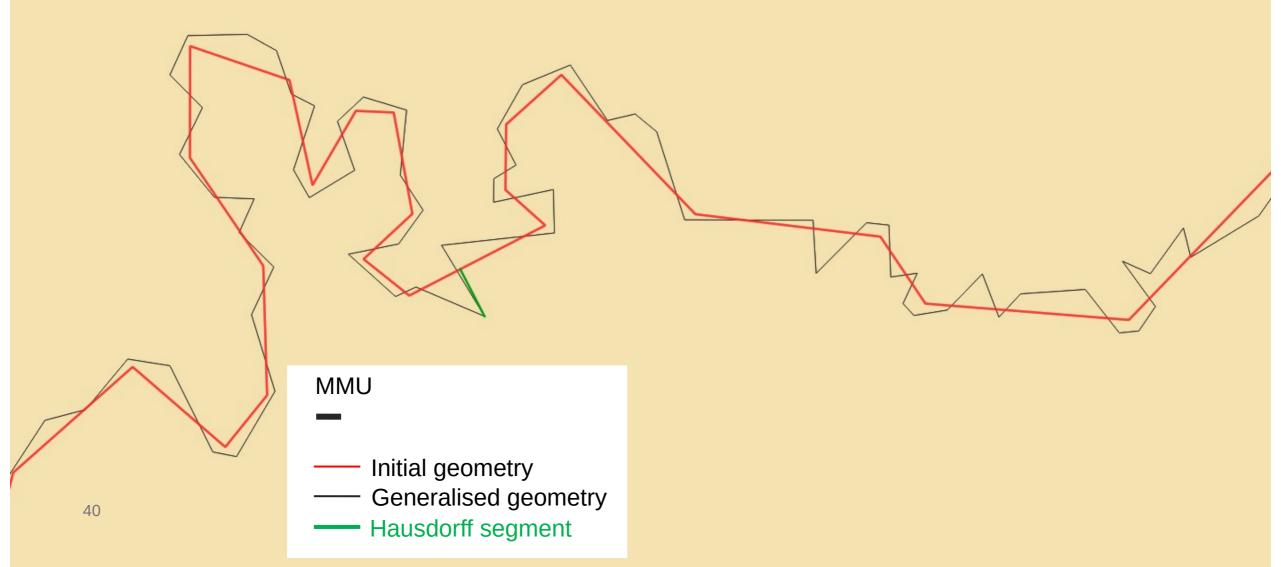


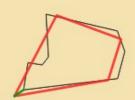


#### MMU

Initial geometry
Generalised geometry
Hausdorff segment







#### MMU

41

- Initial geometry
- Generalised geometry
- Hausdorff segment

# Quality influence on spatial analyses



#### Data quality and spatial analysis

- Accessibility analysis to healthcare and education services.
- 1km resolution population grid.
- Travel time to nearest service by road transport.



		Dresden	and frances and		
A STATE OF A	Erfurt	A Real W	A AREA A		ravel time to nearest health service, in minutes
Bad Hersfeld	Gera			Wałbrzych	5 10 15 20 30 45 60 90
A C P Land M G	CONTRACTOR AND	Market My	A CALL AND A CALL		
	nenau 👘 🔒	Ústí nad	Labem		No data
Gießen Fulda	A TOWN AND A PLAN	W ALL IN I		Strates	CHAT SHE HE COM
	Plauen		C		The second se
	APRIL	· · · · · · · · · · · · · · · · · · ·	and and a second second	radec Králové	治世人 大学 不是 "
「「「「「「「」」「「」」「「」」「「」」」「「」」」「「」」」」	Coburg, Karlo	ovy Vary		****	and the second
Frankfurt am Main		A Carl Carl	Praha	1. 1. 1	Ostrava
					the should be a start of the sec
	新しまして来たい後した		the second second	All States	
	amberg		Vlašim 🚺		Olomouc
Würzburg		Plzeň 👘		12	
<b>主义和你这些公司</b> 。""这个人并不	Parts I and Mary United	The section of the	A CONTRACT OF	1 A 36	
annheim/Ludwigshafen	And the store		Tábor Jihla	val	Zlin
国。2月11日(1990年)。199	Nürnberg	Klatovy		and as he	Brno 👘 👘
		11 11 11 11 11 11 11 11 11 11 11 11 11		44 - C - 4	and the second second
A HOLL AND CARE THE	A T I A A A A A A A A A A A A A A A A A	「「「「「「」」	No. In Mary and a	Real Providence	Trenčín
		and a second sec	České Budějovice.	147 - SAML	Hodonín
	Regensburg	Regen		Znojmo	and the second second
	人名格兰 法法律	a la harris		1. E	
Stuttgart 🖉 🖓				Star L. Star	
TRANSFER TO A TRANSFER	建市行业部门 二十年年	Passau	经不关于 美国 生代	the stand	E CALL & SHOW MENTER
Stand Bridge Barrier	and the second	Fassau	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1.1.1.1.1.1.1.1	A CONTRACTOR OF THE
A CARLEN CARLES AND CA		A COLOR			ionit
			📲 Linz	VV	ien Bratislava
KINE AND A LARGE MARKET AT		Braunau am Inn	「「「「「」」」、「」」、「」」、「」」、「」」、「」」、「」」、「」」、「」		A State of the second sec
	München München	10 Mar 10 1	and the second second		
gau 📚				6 9 C	
And the second sec	APPE LA APPE A		二十二 一十二	544 6 16	Győr
	· 李小王 · · · · · · · · · · · · · · · · · · ·	Salzburg	a start to		ooption -
https://ec.europa.eu/eurostat/	statistics-explained/index n	hn?title=Llrhan-ri	ural Europe - a 🕌	Y STA	
https://ec.europa.eu/eurostat/ uality_of_life_in_rural_areas#					A start a start
durity or no intractination	THE COURSE IN THE	and the second of the		1 P. A.	SzomPathali

and the state

1 ST 1 - 3

#### Data quality and spatial analysis

 Tests based on 3 data sources: NMCA data, TomTom multinet, OpenStreetMap.

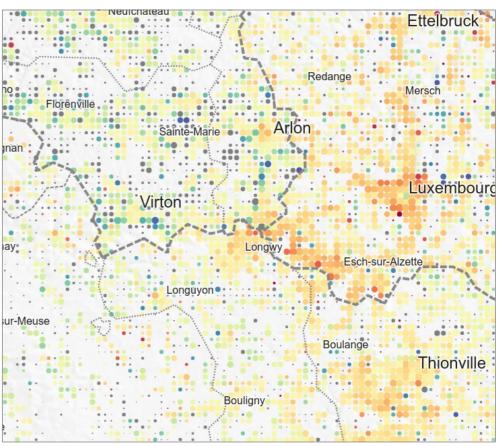
Comparison case	RMS error – Healthcare (min)	RMS error – Education (min)
NMCA vs. TomTom multinet	3.15	1.32
TomTom multinet vs. OpenStreetMap	13.27	4.37
OpenStreetMap vs. NMCA vs.	12.05	4.11



#### Data quality and spatial analysis

- Building demography: Various indicators on building surface, by use (residential, industrial, etc.)
- 1km resolution grid
- NMCA data, OpenStreetMap.

Comparison case	RMS error – Residential (m²)		RMS error – Commercial & services (m²)
NMCA vs. OpenStreetMap	85 142	20 548	15 927



#### Outline

- 1. Quality requirements
- 2. Quality control
- 3. Quality influence on spatial analyse



# Thank you



© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the <u>CC BY 4.0</u> license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

