Relevance of Data Integration Line of Work

Synopsis

UN-GGIM: Europe | Data Integration | Line of Work

Version 1.0

2025-03-17

DOCUMENT HISTORY

Version	Date	Comment	Modified Pages	Responsibility
0.1		Drafting	all	Markus Jobst
0.2		Editing	all	Pier-Giorgio Zaccheddu
0.3		Editing	all	Nathalie Delattre
1.0		Final editing	all	Sabine Afflerbach-Thom

CONTENT

INTRODUCTION

The UN-GGIM: Europe Working Group on Data Integration started its work with the main focus on effects and requirements of the integration of geospatial data with other information in 2014. The members are coming from National Mapping and Cadastral Authorities and National Statistical Institutes from about 20 European UN Member States and also includes Eurostat and the Joint Research Center (JRC) as well as the European Environment Agency (EEA).

The published outcome of the past ten years deals with the activities of the UN-GGIM: Europe Line of Work on Data Integration (until 2022 known as Working Group on Data Integration – including both topics: effects and requirements of Data Integration and SDG analysis), particularly in the context of European statistics and spatial data infrastructure.

Main points of the data integration Line of Work are…

- There is a growing need for closer cooperation between NSI and NMCA to address the challenges of data integration and to optimize the production and use of geospatial data.
- Regular consultations at the highest level (ministry) are less frequent, which can complicate the strategic direction of cooperation.
- Geospatial data play an important role in monitoring the SDGs, especially for indicators with a spatial dimension.
- Challenges include harmonizing data sources, defining thresholds and the availability of geospatial data at different spatial levels.
- License agreements for the use of geospatial data are heterogeneous and range from free use to paid licenses for commercial purposes.
- Challenges include documenting and standardizing technical and organizational methods, interoperability of systems and ensuring data quality.
- Involving users and taking their needs into account are essential for the success of data integration.
- Exchanging best practices and cooperation between Member States are essential for progress in data integration.

In summary, the UN-GGIM: Europe Line of Work on Data Integration provides important impetus for the further development of the European geospatial knowledge infrastructure (EU-GKI) and the use of spatial data for political decision-making processes as well as the monitoring of the UN-SDGs.

TECHNICAL AND ORGANIZATIONAL REQUIREMENTS AND RECOMMENDATIONS FOR SUCCESSFUL GEOSPATIAL DATA INTEGRATION IN EUROPE

To ensure successful geospatial data integration in Europe, various technical and organizational requirements must be met. The sources emphasize the need for close cooperation between the authorities responsible for geospatial data and statistics, the harmonization of standards and data, and the establishment of a European geospatial data infrastructure.

Organizational requirements and recommendations

Close cooperation: One of the key recommendations is formal cooperation between statistical offices and geospatial data authorities [1, 2]. This can be achieved through joint projects, data exchange, and the development of common standards and methods for data integration [13]. A common use case for all countries should be the work on the 2030 Agenda, which typically involves both statistical offices and geospatial data authorities [3].

Clear governance: Establishing strong governance with laws to support data integration that apply to all data producers is essential [4]. This includes signing a national data integration charter that applies to all actors and establishes clear responsibilities and processes for data integration.

Financial and policy support: Sufficient financial resources and policy incentives are needed to enable the transformation of business models and the provision of data to support data integration [4].

Capacity building: Strengthening capacities for data integration at local, national and international levels, including geospatial data and technologies, is crucial [5]. This can be achieved through training, workshops and sharing of best practices.

User engagement: Developing a user engagement strategy and communication plan on data integration issues is important [6]. Users should be involved in the data integration process to ensure that their needs are met.

Technical requirements and recommendations

Harmonisation of standards: Evaluating, implementing and supporting international, European and nationally agreed standards is crucial [7]. This includes using open standards and common data models to ensure geospatial data interoperability. In that perspective we should refer to the INSPIRE Directive and the UN-GGIM core data.

Data quality and availability: Improving the quality and availability of geospatial data is a prerequisite [7]. This includes ensuring the accuracy, completeness, timeliness and consistency of data.

Overarching data infrastructure: Building a European spatial data infrastructure based on national spatial data strategies is crucial [8, 9]. This infrastructure should enable easy access to harmonised spatial data and ensure interoperability of different datasets at national and European level.

Using innovative technologies: Using innovative technologies such as cloud computing, big data analytics, machine learning and integrated information systems can significantly improve data integration [7, 10].

Geocoding: An agreed and documented method for geocoding is essential for successful data integration [11, 12]. This includes using unique and stable identifiers for spatial objects and well stated lifecycle information. Recommendation is to use when appropriate the geocoding of the INSPIRE Directive as a preliminary base for any further adaptation. The INSPIRE Directive has also provided generic guidance enabling unique and stable identifiers for spatial objects.

Linked data: Using linked data as a key technology for data integration is recommended [13-15]. Linked data allows data from different sources to be linked together, thus providing a more comprehensive picture.

Additional challenges and recommendations

The different institutional and legal frameworks in European countries can make data integration difficult [16]. It is important to recognise these challenges and find solutions that enable collaboration and data sharing between different actors.

Data protection and data security are important aspects of data integration [17]. It is important to ensure that data integration is carried out in accordance with data protection rules and that data is protected from unauthorised access.

The establishment of a European spatial data agency responsible for pan-European data could improve the coordination and harmonisation of data integration at European level [18-20].

The successful implementation of the technical and organisational requirements outlined is crucial to realise the full potential of geospatial data integration in Europe.

STRENGHTEN NATIONAL AND INTERNATIONAL COLLABORATIONS TO IMPROVE DATA INTEGRATION IN EUROPE

The publications of the UN-GGIM data integration working group emphasize the importance of collaborations between national and international actors to improve data integration in Europe. In particular, collaboration between National Mapping and Cadastral Authorities (NMCAs) and National Statistical Institutes (NSIs) is highlighted as crucial for the effective use of geospatial data to support policy decisions. [1-3] Users should also be involved to insure that data integration proecss and applications will meet their needs.

Recommendations to strenghten collaboration

Strengthening communication: An open dialogue between all stakeholders, including governments, data providers and data users, is essential to communicate the benefits of data integration and identify challenges. [4, 5] Regular meetings and conferences, such as INSPIRE and the European Forum for Geography and Statistics (EFGS), provide platforms for knowledge exchange. [6]

Institutional and governance structures: Establishing formal collaboration agreements between NSIs and NMCAs, such as Memoranda of Understanding (MoUs), can promote collaboration. [7-9] Defining clear responsibilities, developing common methodologies and standards and ensuring sufficient funding are crucial. [8, 10] Establishing a European coordination body could promote the harmonisation of data and methodologies at European level. [11]

Improving standards, quality and accessibility: Harmonising data, standards and classifications is crucial for data interoperability. [12-14] Adhering to quality standards and ensuring data accessibility through open data principles are also important. [15, 16]

Using new technologies and data sources: Incorporating Earth observation (EO) data and other new data sources offers great potential for producing innovative indicators and improving the spatial resolution of statistics. [17, 18] Developing capacities in NSI to use EO data is crucial. [19]

Development of a Geospatial Knowledge Infrastructure (GKI): The further development of National Spatial Data Infrastructures (NSDI) towards a GKI that links different data domains and sources is an important step to improve data integration. [20, 21] The GKI should be based on linked data and common ontologies to improve data interoperability and usability. [22]

User engagement: Data integration applications should mainly serve user's needs and requirements considering other domains like climate changes, mobility, health and the Green deal data spaces [1]. When meetings and conferences are organised such as INSPIRE and the European Forum for Geography and Statistics (EFGS), this should include the possibility to exchange with key users dealing in these varied domains. In counterpart, promoting spatial data integration applications in their specific conferences and meetings might open the door to new opportunities. At national level, NSIs

[1] https://green-deal-dataspace.eu/

and NMCAs should collaborate with the other national departments in order to investigate and serve their needs.

Challenges and obstacles

Despite the benefits of data integration, challenges and obstacles still exist that need to be addressed:

Lack of funding and political support: Funding for data integration initiatives is often inadequate and there is a lack of political support to implement the necessary changes. [8]

Complex institutional and legal environment: Existing institutional structures and legal frameworks can hinder collaboration between data providers and data sharing. [14, 23]

Lack of capacity and expertise: Some institutions lack the expertise and resources to effectively apply data integration methods and exploit new technologies. [24]. Capacity building might be organised in cooperation with well stated organisations like EuroSDR and with universities.

Data protection concerns: The increasing use of big data and EO data raises data protection issues that need to be carefully considered. [25]

Examples of collaborations

The sources provide several examples of successful collaborations to improve data integration:

Finland: The IGALOD pilot project demonstrates the integration of geospatial data and statistical classifications in RDF and their delivery via SPARQL endpoints. [26]

Finland: EuroStat funded the project “GSGF in Finland”, which addresses methods and practices of the GSGF framework implementation in a national context, has resulted in many aspects of data integration.

Sweden: The Swedish geodata strategy (geodatastrategi) has already implemented many of the measures recommended in the sources and serves as a model for other countries. [27]

Spain: The collaboration between the National Statistics Institute (INE) and the Directorate-General for Transport (DGT) has resulted in improved territorial segmentation of statistical data and the development of new indicators. [7]

In conclusion, improving data integration in Europe requires joint efforts at national and international level. By promoting communication, strengthening institutional structures, harmonising standards and leveraging new technologies, the effective use of data to support policy decisions and address societal challenges can be achieved. The sources show that progress has already been made, but more efforts are needed to overcome the obstacles and exploit the full potential of data integration.

GEODATA INTEGRATION MECHANISMS TO SUPPORT AI APPLICATIONS

Although the sources primarily focus on the integration of geodata and statistics, some important points can be derived on how these mechanisms can support new AI applications.

1. Data availability and accessibility

AI applications require large amounts of high-quality data. Integrating geodata into a networked system, such as a Geospatial Knowledge Infrastructure (GKI) [1-3], can facilitate access to relevant training data for AI models [3].

Through standardized interfaces and open APIs [3-5], such as OGC API processes, AI systems can more easily access and process the data they need. [5]

The increasing availability of geodata, also through initiatives such as INSPIRE, Copernicus and the Open Data movement, is helping to develop AI applications in areas such as environmental monitoring, urban planning and disaster management. [6-9]

2. Semantic interoperability

Semantic interoperability is essential for AI systems to understand and process the meaning of geodata. [10, 11]

Ontologies and vocabularies play an important role in defining the relationships between different data sets and thus enabling a common understanding of the data. [10-12]

This is particularly important for AI applications that need to combine and analyse data from different sources to recognize complex relationships. [10]

3. Georeferencing and spatial analysis

Geodata integration enables precise georeferencing of data from different sources. This is the basis for spatial analyses that can be used by AI systems.

AI algorithms can recognize spatial patterns and relationships that are difficult to identify using conventional analysis methods. [13, 14]

Examples of applications include the optimization of logistics processes, the prediction of traffic flows or the analysis of environmental changes. [15, 16]

4. Linking geodata with other data domains

Geodata can serve as a bridge between different data domains. By linking geodata with socioeconomic, demographic or environmental data, AI systems can gain a more comprehensive picture of reality. [13, 17-20]

This opens up new opportunities for the development of AI applications that address complex issues in areas such as health, smart cities or sustainable development. [21-23]

5. Scalability and automation

Geodata integration and the use of cloud technologies enable AI applications to be scaled to large amounts of data. Automated processes for data processing and analysis can increase the efficiency of AI applications.

Challenges

Despite the potential of geodata integration for AI applications, there are challenges that need to be overcome. Data protection and data security must be ensured when sensitive data is incorporated into AI systems. The development of AI applications often requires specialized expertise and interdisciplinary collaboration.

Summarizing, geospatial data integration plays a key role in supporting new AI applications. By improving data availability, semantic interoperability and georeferencing, the mechanisms described in the sources create the basis for the development of innovative AI solutions in various fields.

OBSERVATIONS OF THE DATA INTEGRATION LINE OF WORK

Cooperation between National Statistical Offices (NSI) and National Mapping and Cadastre Offices (NMCA): Many of the sources emphasize the importance of cooperation between NSI and NMCA for effective data integration. This includes sharing data, developing common standards and aligning workflows. [Example: "2019-02-06_UNGGIM-Europe_ReportQuestionnaire_v2.pdf" shows that 53% of countries have a national spatial data management policy, but only 29% of these policies also cover aspects of data integration with statistics.]

Open data and license agreements: The increasing importance of open data for spatial data is highlighted in several documents. It is discussed how license agreements can be designed to make data accessible both within the government and for citizens and businesses. [Example: "2016-12-21_WGB2.1-Report-Final-version_1.0.pdf" analyses various national and international frameworks for sharing geospatial data and compares their licensing models.]

Using geospatial data for monitoring the Sustainable Development Goals (SDGs): The documents show how geospatial data can be used to calculate and analyse SDG indicators. Examples include indicators for access to public transport, land sealing rate and the share of forest area. [Example: "2019-05__UN_GGIM_08_05_2019-The-territorial-dimension-in-SDG-indicators-Final.pdf" provides detailed analyses of selected SDG indicators and shows how geospatial and statistical data can be integrated to gain meaningful insights.]

Data integration methods and technologies: Various documents address the technical aspects of data integration. These include geocoding, linked data and the use of standards such as ISO/TC11, IGC and SDMX. [Example: "2021-11-22__UNGGIM-Europe_WG_DataIntegration_Subgroup-II_DataIntegrationMethods.pdf" provides an overview of different data integration methods, including geocoding, table linking and linked data.]

Data integration challenges: The documents also identify challenges that can arise during data integration. These include privacy concerns, interoperability of different datasets and ensuring data quality. [Example: "2019-02-06_UNGGIM-Europe_ReportQuestionnaire_v2.pdf" identifies statistical confidentiality, data duplication and lack of interoperability as challenges for data use.]

Key findings:

- Integrating geospatial and statistical data offers great potential for generating meaningful insights and supporting policy decisions.
- Collaboration between different stakeholders, including NSI, NMCA and other authorities, is critical to the success of data integration.
- Open data & clear licensing agreements are essential to promote access to/use of geospatial data.
- Developing and implementing standards and methodologies for data integration is critical to ensure interoperability and data quality.
- Data Integration is a core functionality für use cases with artificial intelligence.

Recommendations:

- Promote national strategies and action plans to support data integration, e.g. with the governance of the UN-IGIF.
- Strengthen collaboration and knowledge sharing between NSI, NMCA and other relevant authorities and user communities.
- Develop and implement open data policies and clear licensing agreements for geospatial data.
- Invest in developing and implementing standards and methodologies for data integration.

- Increase the capacity building in matter of data integration methods with the cooperation of universities and other associations

OUTLOOK: RELEVANCE OF DATA INTEGRATION FOR AI

The integration of geospatial data plays a significant role in artificial intelligence (AI) use cases.

Foundation for data-driven innovation: Accessible and interoperable data, including geospatial data, form the core of data-driven innovation. This data, combined with a digital infrastructure and artificial intelligence solutions, facilitates evidence-based decision-making.

Linked data methods contribute to the existence of fundamental geospatial data themes on the web and form a foundation for building a geospatial knowledge infrastructure, which can be understood as a set of spatial knowledge graphs. The consideration of linked data within national, regional, or international spatial data infrastructures (SDI) will contribute to better and more efficient discoverability, accessibility, exploration, and use of geospatial data over the internet and enable the development of a next generation of SDIs.

Semantic interoperability: A general harmonization of schemas has already been achieved through INSPIRE, but a more semantics-based approach is still pending. Geospatial data is integrated through the use of common and domain-specific semantics. The data spaces are designed for specific questions and use cases, e.g., environmental reporting structures. Geospatial data is made interoperable according to the requirements of the respective data space – including the provision of ontologies and vocabularies for each data space, which at least enable geospatial-thematic data integration.

Better decision-making and problem-solving: GKI applications demonstrate that the concept of spatial knowledge is used in many fields, such as education, land monitoring, ecological transition, and mobile mapping applications, to support the decision-making process with the right information or to solve a specific problem.

In summary, the integration of geospatial data, particularly through linked data approaches and within data spaces and geospatial knowledge infrastructures, forms an essential foundation for AI applications by expanding the data basis for analyses, improving semantic interoperability, and thus contributing to more precise insights and more informed decisions in various fields.

REFERENCES

[1,2,3] UN-GGIM: Europe (2015). *Definition of the priority user needs for combinations of data*. Version 1.1.
[4] UN-GGIM: Europe (2016). *Report on Managing Side-Effects*. Version 1.0.
[5] UN-GGIM: Europe (2016). *Report on task B2.1 "The methods of implementing the prioritised combinations of data: Review of current European Interoperability Frameworks and geospatial and statistical integration projects regarding methods of combinations of data"*. Version 1.0.
[6] UN-GGIM: Europe (2017). *Recommendations for methods implementing the prioritised combinations of data*. Version 1.0.
[7] UN-GGIM: Europe (2019). *Report on questionnaire on data integration*. Version 2.0.
[8] UN-GGIM: Europe (2019). *The territorial dimension in SDG indicators: geospatial data analysis and its integration with statistical data*. Instituto Nacional de Estatística
[9] UN-GGIM: Europe (2019). *The integration of statistical and geospatial information — a call for political action in Europe*. 2019 Edition
[10] UN-GGIM: Europe (2020). *Scoping Paper: Subgroup I: SDG Indicator Calculation and Recommendations*. Version 1.0.
[11] UN-GGIM: Europe (2020). *Scoping Paper: Subgroup II: Data Integration Methods*. Version 1.0.
[12] UN-GGIM: Europe (2020). *Evaluation of the Policy Outreach Paper*. Version 1.0.
[13] UN-GGIM: Europe (2021). *Data Integration Methods*.
[14] UN-GGIM: Europe (2022). *SDG Indicator 11.2.1 Calculation and Recommendations*
[15] UN-GGIM: Europe (2022). *SDG Indicator 15.1.1 Calculation and Recommendations*
[16] UN-GGIM: Europe (2022). *SDG Indicator 15.3.1 Calculation and Recommendations*
[17] UN-GGIM: Europe (2022). *Evaluation Report*. Version 1.0
[18] UN-GGIM: Europe (2023). *Evaluation Report: Webinar 1: Data Integration in EU Context*. Version 1.0
[19] UN-GGIM: Europe (2023). *Evaluation Report: Webinar 2: Data Integration and GKI*. Version 1.0
[20] UN-GGIM: Europe (2023). *Evaluation Report: Webinar 3: Data Integration and Linked Data*. Version 1.0
[21] Keeley, L., Walter, H., Pikkel, R., & Quinn, B. (2013). *Ten Types of Innovation: The Discipline of Building Breakthroughs*. Wiley
[21] Plattner, H. et al. (2010). *Design Thinking: Understand, Improve, Apply*. Springer
[22] Lerche, I., & Glaesser, W. (2006). *Environmental risk assessment: quantitative measures, anthropogenic influences, human impact*. Springer
[22] O'Brien, M. (2002). *Making better environmental decisions: an alternative to risk assessment*. MIT Press
[23] Shrader-Frechette, K., & Westra, L. (Eds.). (1997). *Technology and values*. Rowman & Littlefield.
[24] Panu Muhli, National Land Survey Finland, https://www.maanmittauslaitos.fi/en/research/gsgf-finland-integration-geospatial-and-statistical-information-finland-gsfi