



Data Integration Methods: Scoping Paper

UN-GGIM: Europe | Work Group on Data Integration | subgroup II

Final

2020-06-18



Change controls

Version	Author	Revision Date	Comments, summary of changes
0.1	Nathalie Delattre	13/03/2020	Initial draft for completion with BKG
0.2	Pier-Giorgio Zaccheddu, Sabine Afflerbach-Thom, Esa Tiainen	27/03/2020	Updated draft
0.3	Bianka Fohgrub, Pier-Giorgio Zaccheddu	31.03.2020	Updated draft
0.4	Dorus Kruse	15.04.2020	Updated draft
0.5	Bianka Fohgrub Kathrin Gebers	24.04.2020 28.04.2020	Updated draft
0.6	Bianka Fohgrub	05.05.2020	Adaption of discussed changes and comments
1.0	Subgroup II	18.06.2020	Accepted by the subgroup



Content

WG on Data Integration – Background for current line of work.....	4
Framework	5
Connections / Links	5
Aim	6
Outputs.....	8
Work outline.....	8
Planning.....	8
Participants.....	10



WG on Data Integration – Background for current line of work

The new 2019-2022 work plan for UN-GGIM: Europe, approved at the 6th Plenary Meeting held in Brussels between 5 and 6 June 2019¹, recognises that the Working Group (WG) on Data Integration “has formed a network of interested members from both National Mapping Agencies and National Statistical Institutions” and that, over the years, “has successfully explored the challenges and benefits of how the integration of statistical and geospatial and other information can meet and satisfy user needs and requirements”. In this context, the new work plan defined, as the next logical step, “to increase the scope of data integration beyond just geospatial and statistical, and to also include earth observations, environmental data and other themes” and to assess how these can contribute to address the 2030 Agenda challenges at global, national and regional levels.

Against this background, the following primary and secondary tasks have been put forward to be completed in the next 2-3 years:

Primary Tasks:

1. Analysing further SDG indicators – focusing on Earth Observation, and can include as part of a subtask ‘Requirements and practices from National Statistical Offices for the use of earth observation data for national statistics’;
2. Advisory Group for global and European data integration issues;
3. Analysis of (future) trends in data capture, creation, maintenance and management – using e.g. Linked (Open) Data and other methods to enhance data integration

Secondary Task:

1. Consider options on how the Regional Committee can support Post-Census 2020 activities and use synergies with UNECE/WG Data Integration to explore ways on how to manage data integration post Census 2020.

Two subgroups have been established at the kick-off meeting held at BKG (Frankfurt, Germany) between 30 and 31 October 2019, to address these tasks. This document presents the scope of work to be developed by subgroup II dealing with primary task 3 ‘*Analysis of (future) trends in data capture, creation, maintenance and management – using Linked (Open) Data methods to enhance data integration*’

¹ <https://un-ggim-europe.org/wp-content/uploads/2019/11/UN-GGIM-Europe-WorkPlan-2019-2022-FinalAdopted.pdf>.



Framework

The activities carried out within the UN-GGIM: Europe WG on Data integration have been taking into account the developments of the participating countries in relation to data integration methods, e.g. Table Joining Service (TJS) and Linked Data. In several countries, e.g. The Netherlands, Finland, the developments are quite progressive; in other countries the expertise is not existent and therefore not used at all.

The subgroup II refers to the task "Analysis of (future) trends in data capture, creation, maintenance and management - using e.g. Linked (Open) Data and methods to enhance data integration" as described in the Work Plan 2019-2022 that can be found via the following link:

<https://un-ggim-europe.org/wp-content/uploads/2019/11/UN-GGIM-Europe-WorkPlan-2019-2022-FinalAdopted.pdf>

Connections / Links

The work of this subgroup II will be closely connected to the global UN Expert Group on the Integration of Statistics and Geospatial Information (ISGI), as the Global Statistical and Geospatial Framework (GSGF) is an important document for ongoing work. At the last UN EG ISGI meeting in October 2019 in Manchester, this Expert Group discussed, that further information on data integration methods based on location will be a benefit to the global perspective and can be used for future work in developing countries. The work of both groups will be streamlined and a frequent information exchange established. As a general, overarching framework, the UN-GGIM Integrated Geospatial Information Framework (IGIF) will be reflected as well.

As the GEOSTAT-3 project produced a European version of the global GSGF, the UN-GGIM: Europe Working Group on Data Integration will have a close connection to the new GEOSTAT-4 team who continues to work on this topic.

The subgroup will closely follow and possibly contribute to the development of EU data spaces.



Aim

In a first step the UN-GGIM: Europe WG on Data Integration has agreed to evaluate the written deliverables of work phase 2014-2017 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”² and B2.2/2.3 “The methods of implementing the prioritised combinations of data: Provide best practice guidance to the interactions between National Mapping and Cadastral Agencies (NMCAs) / National Statistical Institutions (NSIs) / Environmental Agencies and other relevant organisations. Review current use of data from multiple sources to identify case studies and best practices relevant for combinations with core data”³.

The subgroup II will focus on collecting “Data Integration Methods” to enhance data integration by location in general. The goal is to questioning the described methods and to evaluate and assess them in consultation with the appropriate countries.

Data integration may cover two different aspects: data integration across border, which implicates the integration of national data components of a common theme in order to set up a pan-European coverage; and data integration across themes, which implicates the combination of datasets on a same location and originated from different domains: geospatial and statistical data, such as socioeconomic, environmental, transport, etc.

The subgroup will target the data integration methods based on national examples across themes:

- spatial data sets delivered from the National Mapping and Cadastral Agencies or through the national geospatial data infrastructure (NSDI);
- statistical data sets owned by the National Statistical Institutions;
- alternative data sources that can be geo-enabled, e.g. big data (mobile phone data).

Data integration methods are already existing, applied and well documented. Moreover the technology is rapidly evolving. The objective of the subgroup II on “Data Integration Methods” is therefore to summarise the situation and provide a European perspective with recommendations for possible use, completed with an impact assessment at social, economic and organisational level. This will be an incentive for discovering, encouraging and initiating the use of data integration methods and facilitate

² https://un-ggim-europe.org/wp-content/uploads/2018/11/20161221_WGB2.1-Report-Final-version_1.0.pdf

³ https://un-ggim-europe.org/wp-content/uploads/2018/11/20160909_ReportonManagingSideEffects_UNGGIM-Europe_WGB3_v1.0.pdf



the decision-making process in countries, which are seeking for a new approach to capture, create, maintain and manage their data.

The task comprises a quick scan regarding different data integration methods used across Europe. “Table Joining Services” (TJS), “Linked Data” (LD) are already known and used methods, but all methods and best practices (if available) will be considered in addition if they can be used in this publication. The benefits and obstacles, adequacy of each method to use cases will be scanned and (national) best practices will be used in the report. Next to the adequacy of each method to use cases, the subgroup II will show realistic examples of these methods.

Moreover than a technical assessment of the different methods, an impact assessment at economic and organisational level will be investigated in order to facilitate the decision-making process.

The subgroup II will have the same target group as in their previous tasks, i.e. the high level management / decision makers of NMCAs and NSIs in Europe.

For this purpose, the subgroup II will collect (national) best practices and examples among the WG members before extending the collection to other countries (if needed). Based on the review of this first collection of information, a template will be created in order to structure the information that may be used for collecting similar information from other countries.

The summaries about national examples / projects will be drafted answering the following questions:

- How can the data integration method be described?
- What has been the motivation to execute the method? (advantages, benefits)
- What are the needs / the lacks identified which have still to be tackled?
- What might be the main message / recommendations for the management level from the example?

Moreover, the (national) best practices will consider existing guidelines and rules in that matter, for example: common standards and vocabularies, and on geocoding processes using addresses.

In that perspective, the subgroup II will take into consideration other similar works and projects on data integration involving as well the environmental domain like the development about Data Ecosystem Accounting Project (Destatis/BKG/Eurostat) / Geospatial Environmental Accounting Team (EEA), or any other previous European projects (like the GEOSTAT projects) or in the European institutions.



Outputs

The expected outputs or deliverables of this subgroup will be

1. a **short report** resulting in a management summary,
2. a **short leaflet**, to be published, easy to read as an outcome to attract the management level derived from (1),
3. **the publication of Confluence Wiki pages** used, amongst others, for the collection of the national best practices

The output and political message should be condensed in a clear and appealing way. It should not hold recommendations only, but more input to show the work, that was done behind the outcomes.

Finally, the outcome and findings from subgroup I “SDG Indicator Analysis” and subgroup II “Methods of Data Integration” will be conciliated and a **conclusive report for future actions** will be prepared.

Work outline

The workload will be the collection of national best practice examples by each partner per country. This means the conduction of 1 or 2 short meetings nationally with the NSI/NMCA counterparts to discuss and compile these national examples.

The subgroup II will preferably conduct webinars for the information exchange and national examples will be collected within the wiki.

After having collected examples, the “best” (criteria to be defined) will be selected and a first compilation / report drafted. A physical meeting for consolidation and decision might be organised for a smaller task team assigned with this task.

Planning

The work is expected to be developed in a time-frame of 2 years, from 2020 to 2021/2022

Task	Topic
1	Review of the deliverable B.2.1 and B2.2/2.3 from the work plan 2014-2017
2	Evaluate links to the GSGF and GSGF-Europe to avoid double work (GEOSTAT-4)
3	Collect national examples/projects
4	Identify best practices (max 3 most important, most used methods) for data input, creation, maintenance and management (Methods in general, TJS, LOD)
5	Consider economic, political and social level of each method
6	Determine lessons learned of each methods



Task	Topic
7	Derive recommendations (if deemed appropriate)
8	Roadmap to implementation
9	Create the output

**PHASE 1:
REVIEW AND EVALUATE**
Until 04/2020

- Review of the deliverable B.2.1 and B2.2/2.3 from the work plan 2014-2017
- Evaluate links to the GSGF and GSGF-Europe to avoid double work (GEOSTAT-4)

**PHASE 2:
COLLECT AND IDENTIFY**
Until 06/2020

- Collect national examples/projects
- Identify best practices (max 3 most important, most used methods) for data input, creation, maintenance and management (Methods in general, TJS, LOD)

**PHASE 3:
SELECT AND STRUCTURE /
ANALYSE**
Until Q4/2020

- Consider economic, political and social level of each method
- Determine lessons learned of each methods
- Derive recommendations (if deemed appropriate)

**PHASE 4:
DOCUMENT**
Until Q2/2021

- Roadmap to implementation
- Create the output



Participants

Function	Name	Country	Affiliation	
Subgroup leader	Nikolaos Roubanis	EU	Eurostat	EC
Chief editor	Nathalie Delattre	BE	IGN Belgium	NMCA
Member	Esa Tiainen	FI	National Land Survey of Finland	NMCA
	Kathrin Gebers	DE	Federal Statistical Office of Germany	NSI
	Romain Douard	CH	Swiss Federal Statistical Office	NSI
	Dorus Kruse,	NL	Dutch Kadaster	NMCA
	Berny Kersten	NL	Dutch Kadaster	NMCA
	Paloma Abad Power	ES	Instituto Geográfico Nacional	NMCA
	Anna Slawinska	PL	Statistics Poland	NSI
	Amalia Velasco Martín-Varés	ES	Spanish Directorate General for Cadastre	NMCA
	Jana Pressová	CZ	Land Survey Office, Czech Republic	NMCA
	Rina Tammisto	FI	Statistics Finland	NSI
	Antti Santaharju	FI	Statistics Finland	NSI
	Mirosław Migacz	PL	Statistics Poland	NSI
	Vlado Cetl	EU	Joint Research Centre (JRC)	EC
	Markus Jobst	AT	BEV	NMCA
	Regine Elling	DE	BKG	NMCA
Pier-Giorgio Zaccheddu	DE	BKG	NMCA	
Bianka Fohgrub	EU	Eurostat	EC	
Contributor	Thomas Ellett von Branch			NOSIN LOD Group