# Evaluation of the Webinar

## #3: Linked Data in the European spatial data infrastructure – an operational view

7 November 2023

---

Synopsis

---

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

**Version 1.0**

**2024-01-26**

# CONTENT

# FIGURES

# INTRODUCTION

The UN-GGIM Europe Working Group on Data Integration has analysed current and future trends in geospatial data integration. Examples of best practices have been collected and data integration methods that are especially relevant in a European context have been identified. With the latest developments and use cases in the geospatial domain, these principles of data integration support common data spaces, a single digital market, artificial intelligence (AI) and a geospatial knowledge infrastructure (GKI). A basis for the establishment of a geospatial knowledge infrastructure, which could be understood as a set of spatial knowledge graphs, are the methods of linked data, which contribute to the existence of fundamental geospatial data themes on the Web.

By observing actual developments of AI and the creation of common data spaces three main questions occur:

- How far are the methods of linked data from an operational implementation?
- Are these concepts productive in the geospatial domain?
- What are the experiences, observed pitfalls and feasible approaches that are missing?

In this webinar, three selected examples highlighted an operational view of linked data as well as the indexing and standardizing of linked data in a common data space. Main experiences will be shared and pitfalls discussed. The discussion and interaction within the webinar should indicate possible gaps, identify focal points of actual implementations and highlight some requirements for maturity building.

# PARTICIPATING PARTIES

## 1. National Land Survey of Finland (NLS) and ISO Subgroup on EU Data Spaces

The National Land Survey of Finland (NLS) safeguards the land ownership and credit system by maintaining information about properties and housing company shares in its registers and takes care of the registration of ownership and mortgages. Other tasks of the agency include spatial data research and application. The National Land Survey's mission and the cornerstone of the operations is the collection and provision of information about the Earth. Information is the foundation for everything, service and research are a central part of the NLS's activities.

The ad-hoc group “Input to EU Data Spaces” of the ISO Technical Committee 211 (Geographic Information/Geomatics) prepares a document describing the importance of standards for the use and integration of high-value datasets as well as in accordance with the European Strategy for data (https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020DC0066).

This presentation revealed the use of open standards with RESTful principles to deliver information for common European data spaces.

**Presentation title:**
Preparing data integration and the common data spaces using OGC API's

## 2. Central Statistics Office (CSO) of Ireland and European Forum for Geography and Statistics (EFGS)

The Central Statistics Office (CSO) is Ireland's national statistical office and its purpose is to impartially collect, analyse and make available statistics about Ireland’s society and economy. At national level, CSO official statistics inform decision making across a range of areas including construction, health, welfare, the environment and the economy. At European level they provide an accurate picture of Ireland’s economic and social performance and enable comparisons between Ireland and other countries.

European Forum for Geography and Statistics (EFGS) started as a voluntary cooperation between National Statistical Institutes (NSIs) in the Nordic countries in 1998 on use of geographic information systems (GIS) and statistics. Today EFGS has national contact persons from more than 40 countries and territories, having annual conferences and meetings. EFGS’s (previously called European Forum for Geostatistics) activities are mainly concentrated on the development of the best practices in the production of geostatistics in Europe.

Using persistent identifiers and supporting data linkages are main requirements for common data spaces. This talk highlighted experiences with establishing linked data and spatial analysis even with protected identifiers and pseudonymised georeferences.

**Presentation title:**
Enabling data linkage on pseudonymised address data

## 3. Kadaster Netherland

Data is the engine of Kadaster Netherland. That is why Kadaster has been continuously working on data-oriented innovation for many years. In 2019, Kadaster started a Data Science Team. With this team, Kadaster wants to stimulate data innovation, aims at gaining more experience with Data Science, and tries also to investigate the possibilities of Data Science for Kadaster and its partners.

A rich collection and increasing availability of fundamental geospatial data themes lead to enhanced linking of the semantics as well as data down to the feature level across data spaces and therefore creates a geospatial knowledge infrastructure. This presentation shares experiences and highlights the most important methods by example of the Kadaster Knowledge Graph.

**Presentation title:**
A pragmatic example for Linked Data: the Kadaster Knowledge Graph

## OBJECTIVES

- Encourage opportunities for future cooperation and joint work.
- Identify the operational implementation of linked data methods.
- Discover productivity and actual use of linked data concepts in the geospatial domain.
- Provide an up-to-date overview of the application and perspectives of linked data.
- Provide a forum for discussion on key challenges of a real-time linked data space as source for a geospatial knowledge infrastructure.

## OUTCOME OF THE DISCUSSION

**Questions discussed:**

***From which sector do you come from?***

*Purpose:*

*The topic of data integration connects to different data spaces and bridges these domains with the data sources used. One main aspect is the relation of the geospatial and statistical domain. But also any other data space (mobility, health, agricutlure, …) makes use of the data integration methodology.*

*Knowing roughly about the interest group around UN-GGIM, this question aimed at identifying the main areas of activity of the audience.*

*Result:*

*About 87% of the participants came from the public sector covering the geospatial, statistical and cadastral domains. Based on the question, 8% assigned themselves to private sector (technology) and 5% others.*

*In the group of the public sector more than the half sourced in the geospatial domain, a quarter in statistics and about a sixth from cadaster.*

***What is your experience in the sector?***

*The topic of linked data and data integration on one hand counts on long lasting experience in collaboration, data science and domain-specific requirements, and on the other hand needs young experts and fresh views to result in innovative pathways and implementations.*

*Result:*

*More than half of the participants answered with 11 or more years experience in this sector. Almost a fifth showed an experience of 1 to 3 years.*

***What do you think how ISO/OGC standardisation is supporting (your) use cases in common data spaces?***

*Purpose:*

*Many use cases, especially those that cover different domains and provoke collaboration or interaction via computer interfaces, or especially by exchanging data structures, depend on a common specification. This common consenus could be called "standard". Different standardisation organisations try to establish a common consenus as technological guidance and therefore support the creation of widely usable software products. Standards on a specific encoding or on data structures may be critical because this "regulation" (concenus) may influence existing use cases, product acceptance, legal requirements or usability.*

*Possible answers:*

- *ISO/OGC standardisation for dataspace-related standards is an agreed consensus that makes the use cases work.*
- *For an initial implementation of the common dataspace use cases NO standardisation is needed. Considering standardisation is needed only in the long run.*
- *We do not consider ISO/OGC standardisation, except when legally obliged.*
- *ISO/OGC standardisation is not needed.*
- *I have no position.*
- *Comment: …..(Free text comment)*

*Result/Discussion:*

*In the discussion more than 60% of participants agreed that ISO/OGC standardisation as consensus for dataspace-related aspects makes use cases work. 6% are in the opinion that no standardisation is needed for an initial implementation, but should be considered on the long run of the implementation. 3% of participants will only consider ISO/OGC standardisation when it is legally obliged.*

*It is important to note that more than a quarter of participants were not able to deliver a position or were uncomfortable with the given answer possibilities.*

*It was noted that the agreement about ISO/OGC standards does not include eGovernment and Smart City standardisation ecosystems.*

**Do you index your data in order to support retrieval of records of your dataspace?**

*Purpose:*

*A smart indexing mechanism for data enables performant retrieval of records, enhances discovery and at least supports data integration. The indexing mechanism can be specific for an ecosystem, for a nationawide infrastructure or for local instances. These different specification levels offer diverse retrieval and data integration characteristics. The main purpose of this question was to observe consciousness to the topic of data indeces.*

*Possible answers:*

- A specific index for data sources in the data space has been created.
- The data index of our local instances (databases) is used.
- I have no idea of the technical implementation.
- Comment: …..(Free text comment)

*Result/Discussion:*

More than 40% of participants confirm that the data index of the local database instances is used, instead of 15% who create a specific index for data sources in the data space.

It is worth to note that 44% have no idea of the technical implementation.

***Which actions do you take to establish linked data?***

*Purpose:*

*The concept of linked data is a modern and meaningful approach to establish automated data integration and to support a common data space. The concept of linked data requires more than technical components and their knowledge base. It includes governance, standardisation, partnerships, governance, innovation, capacity, communication, finances and the data. Therefore, it follows the nine pathways, which are described in UN-IGIF (https://ggim.un.org/un-ggim-intergrated-geospatial-information-framework/). This question explored the status of linked data concerning stakeholder embedding, license management, identifier management, standardisation and embedded vocabularies.*

*Possible answers:*

- *We have our stakeholders prepared.*
- *We have specific datasets selected (at least one) and given a useful license for unrestricted linking the data.*
- *We have established good URI´s for linked data.*
- *We use and connect to standard vocabularies.*
- *We are using our own data vocabulary according to our use-case driven data.*
- *I have no idea of the technical implementation.*
- *Comment: …..(Free text comment)*

*Result/Discussion:*

*This question seems to be very specific. About 50% of participants have no idea of the technical implementation and its requirements.*

*Another 46% make use of vacabularies. 25% of them have their own data vocabulary established according to their use-case driven data. 21% connect to standard vocabularies.*

*14% have their stakeholder prepared for the linked data use case. Another 14% have specific datasets selected and assigned a license for unrestricted linking of the data. 11% have created a specific and good URI schema for linked data.*

***Do you pseudonymise data and therefore replace unique identifiers with protected identifier keys?***

*Purpose:*

*Linked open data are often in criticism of data protection. Therefore, data protection regulation has to be considered in linked data concepts. Possible ways are anonymization of data and their identifiers or to leave critical data out of scope. This question observed the situation of enhanced pseudonymising of data or to rely on persistent unique identifiers for linking data.*

*Possible answers:*

- *Yes, we need to pseudonymise our data in order to enable wider processing of these data.*
- *No, we use persistent unique identifiers in the best case.*
- *No, we use unique identifiers, but not persistent ones.*

*Result/Discussion:*

*More than the half of participants responded that persistent unique identifiers are used in the best case for data integration or linking mechanisms. 24% stated the need for a more detailed pseudonymisation of the data in order to be possible to process the data in a wider use case. Further 24% said that no persistent identifiers are used in their data.*

***If yes, what was the main challenge to keep your data linkable?***

*Result/Discussion:*

*Those participants who confirmed the need for an enhanced pseudonymisation of the data,*

- *the adding of new unique identifiers,*
- *the PID management and its definition when a new object or only a new version is created,*
- *the PID and URI handling in the registry software,*
- *the unique identifer management between different data institutions,*

*were the most challenging issues.*

***Do you provide your own data vocabulary?***

*Purpose:*

*Ecosystems, use cases, domains or data spaces have their specific need of data structures and the naming of the content. This semantic description of the content is expressed in a vocabulary. There are public and standardized vocabularies, which e.g. are needed on the Web to make webpages and their content work, or controlled vocabularies which aim at harmonizing specific content, e.g. data specifications in the legislative framework of EU INSPIRE, or local vocabularies that are directly related to the data structures of a provider. This question tried to find out the status of embedded vocabularies.*

*Possible answers:*

- *Yes, standard vocabularies are embedded.*
- *No. Specific data vocabularies according to our use-cases have been created.*
- *No. We do not use vocabularies.*
- *I do not know.*

*Result/Discussion:*

*About the half of answering participants confirmed the use of specific data vocabularies according to their local use cases. 14% do not use vocabularies at all. 10% embed common standardized vocabularies.*

*About one third of participants do not know the status of data vocabularies in their institution.*

***Do you connect the semantic of your data with other ontologies or thessaurii?***

*Purpose:*

*After creating a vocabulary, the creation of an ontology helps to formally define semantics of data in an service-oriented architecture. A machine needs to understand provided values (geographical features) in order to link and process them. This question tried to explore the status of ontologies and thessaurii in participating institutions.*

*Possible answers:*

- *Yes. We use existing ontologies on the Web and refer to it.*
- *No. We maintain our own ontology.*
- *No. We do not use ontologies to describe the semantic of our data.*
- *I do not know.*

*Result/Discussion:*

*Less than a quarter of participants (22%) stated that existing ontologies on the Web are used for the semantic description and are related to the data. Another quarter of participants do not use ontologies to describe the semantic of the data. 11% maintain their own ontology.*

*About the half of participants (44%) have no idea of the use of ontologies in their institution.*

# CONCLUSION AND NEXT STEPS

The webinar "Linked Data in the European spatial data infrastructure – an operational view" (Recording at https://un-ggim-europe.org/events/webinar-linked-data-in-the-european-spatial-data-infrastructure-an-operational-view/) as seventh meeting in the UN-GGIM: Europe webinar series focused on an operational view of linked data in the geospatial domain. Three selected examples were used to highlight maturity of linked data for production in the geospatial domain. The examples covered OGC API´s for the preparation of data integration and common data spaces, pseudonymised address data for enabling data linkage and the Kadaster Knowledge Graph as very pragmatic example to show requirements in the cadastral domain.

About 110 colleagues from the European continent, but also from Australia and Africa, joined the webinar. One third of the participants were active in the SLIDO interaction, responded and discussed the different perspectives. In the active group on SLIDO more than 85% came from the public sector,

from which more than the half sourced in the geospatial domain, a quarter in statistics and about a sixth from cadastre. More than the half were very well experienced colleagues with more than 11 years of involvement in this domain. Unfortunately, only few participants contributed to the observations and help to identify possible gaps for a future work program. Therefore, the conclusions of this webinar may be biased.

The webinar with its focus on linked data could be categorized as very specific, technical and innovative topic. Throughout all the questions it could be observed that on one hand there is a group of colleagues that is involved in technologies and very well informed, on the other hand there is another group of colleagues that is not informed about technical implementations, -architectures and specifications in detail. As observer of this occurrence we could imagine that participants can be roughly divided into managers and operational experts. Due to the basic importance of the questions in the webinar also for strategic decisions in common data spaces, the situation for a single digital market, artificial intelligence (AI) and the creation of geospatial knowledge infrastructures (GKI), we can observe an urgent knowledge gap that needs to be bridged between managers and innovative implementers.

The operational implementation of linked data seems to be on its way, although for a significant part the main principles for the linked data concept are not fulfilled. E.g. only 11% use persistent unique identifiers for their data, or only 10% make use of standardised common vocabularies.

The main findings are:

- The support and promotion of capacity building at all levels of stakeholders is needed to support the creation of common data spaces, ecosystems and geospatial knowledge infrastructures.
- Development initiatives, competence centres and knowledge hubs need to transmit requirements in an easy and understandable way.
- More specific information about controlled vocabularies, thesauri, ontologies and their requested software products for a sustainable participation of the geospatial domain in data science and for spatial data on the Web is needed.
- The importance of Linked Data approaches, the impact on AI and the influence on the Digital Single Market, has to be emphasized and better explained to decision makers, national offices and geospatial agencies.
- The impact of integrated geospatial knowledge infrastructures on environment and society has to be identified and promoted amongst all stakeholders.
- More information on the different Linked Data methodologies, the governance, its decision-processes and the integrated geospatial supply chain need to be elaborated and its impact communicated.

With the use of this assessment and evaluation report, UN-GGIM: Europe plans to identify common issues, outcomes and findings, gaps and requirements which have not been tackled yet by other stakeholders and to gain valuable feedback for the future work and direction of UN-GGIM: Europe.

# ANNEX I

+++Results of the slido poll++++

## Table of contents

**From which country are you joining us from today?**

0 5 4

## From which sector do you come from?

0 3 8

Public sector - Geospatial

50 %

Public sector - Statistics

24 %

Public sector - Cadastral

13 %

Private sector - Technology

8 %

Other

5 %

## What is your experience in the sector?

0 3 9

1 to 3 years

18 %

4 to 6 years

15 %

7 to 10 years

13 %

11 or more

54 %

## What do you think how ISO/OGC standardisation is supporting (your) use cases in common data spaces?
(1/2)

0 3 4

ISO/OGC standardisation for dataspace-related standards is an agreed consensus that makes the use cases work.

62 %

For an initial implementation of the common dataspace use cases NO standardisation is needed. Considering standardisation is needed only in the long run.

6 %

We do not consider ISO/OGC standardisation, except when legally obliged.

3 %

ISO/OGC standardisation is not needed.

0 %

**What do you think how ISO/OGC standardisation is supporting (your) use cases in common data spaces?**
(2/2)

0 3 4

I have no position.

26 %

Other

3 %

**If Other, please specify:**

0 0 2

- no
- Agreement about standards does not include eGovernment and Smart City standardisation ecosystems

## Do you index your data in order to support retrieval of records of your dataspace?

027

A specific index for data sources in the data space has been created.

15 %

The data index of our local instances (databases) is used.

41 %

I have no idea of the technical implementation.

44 %

Other

0 %

## Which actions do you take to establish linked data?

(1/2)

028

We have our stakeholders prepared.

14 %

We have specific datasets selected (at least one) and given a useful license for unrestricted linking the data.

14 %

We have established good URI´s for linked data.

11 %

We use and connect to standard vocabularies.

21 %

We are using our own data vocabulary according to our use-case driven data.

25 %

**Which actions do you take to establish linked data?**
(2/2)

0 2 8

I have no idea of the technical implementation.

50 %

Other

0 %

**Do you pseudonymise data and therefore replace unique identifiers with protected identifier keys?**

0 2 1

Yes, we need to pseudonymise our data in order to enable wider processing of these data.

24 %

No, we use persistent unique identifiers in the best case.

52 %

No, we use unique identifiers, but not persistent ones.

24 %

## If yes, what was the main challenge to keep your data linkable?

0 0 7

- Periodically adding new unique identifiers
- PDI Management. Clear Object Definition. What ist a New Object, or only a new version
- PID, URI handle, registry (software)
- A unique identifier between different data institutions
- When is an object 'changed'
- -
- PIDs

## Do you provide your own data vocabulary?

0 2 1

Yes, standard vocabularies are embedded.

10 %

No, Specific data vocabularies according to our use-cases have been created.

48 %

No, We do not use vocabularies.

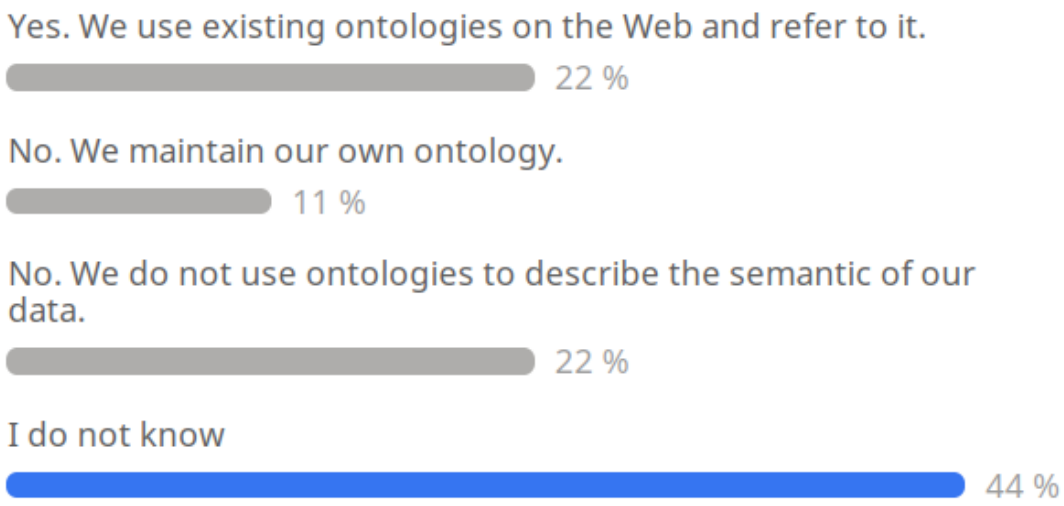
14 %

I do not know

29 %

## Do you connect the semantic of your data with other ontologies or thessaurii?

0 1 8



# ANNEX II

**Further reading and links to consider for this webinar topic**

Best Practices for Publishing Linked Data https://www.w3.org/TR/ld-bp/

NL Kadaster Knowledge Graph https://data.labs.kadaster.nl/dst/kkg
and https://labs.kadaster.nl/thema/Knowledge_graph

Trustworthy Artificial Intelligence https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/excellence-and-trust-artificial-intelligence_en

Real-time linked data spaces https://dataspaces.info/common-european-data-spaces/

SEMIC definition of data space https://joinup.ec.europa.eu/collection/semic-support-centre/data-spaces

International Dataspace Association https://internationaldataspaces.org/

Our common data space EU https://ourcommondataspace.eu/

European Linked Data Taskforce https://pro.europeana.eu/project/linked-data-task-force

List of ontologies https://www.w3.org/wiki/Lists_of_ontologies

European Forum for Geography and Statistics (EFGS)

GSGF Europe

UN-GGIM Europe – Working Group on Data integration