Semantic Interoperability solutions for data spaces

Moderator: Pavlina Fragkou

Objectives of the workshop

- Interoperability in data spaces from vision to reality / from inception to implementation
- How interoperability is achieved in the different data spaces and/or between data spaces?
- What are the main challenge challenges in achieving interoperability?
- Which assets meaning semantic tools oral models are used for this purpose?

Objectives of the workshop

The workshop will be divided into two parts:

The first part focuses on the work that is conducted in the European Commission.

The second part presents the work that has been done so far in different domain data spaces.

Agenda

- Presentations of the first part :
  - Data Act article 33
  - Data.Europa.eu- High Value Datasets
  - Governance, Technical and Semantic Assets from INSPIRE
  - Smart middleware platform (Simpl): Enabling Data Spaces
- Presentations of the second part :
  - HealthDCAT-AP : from vision to reality
  - Language Data Space: Tackling semantic interoperability at the metadata level
  - Public Procurement Data Space (PPDS): A paradigm shift in public procurement data
  - European Data Space for Smart and Sustainable Cities and Communities

Semantic Interoperability solutions for data spaces

Join us on Slido!

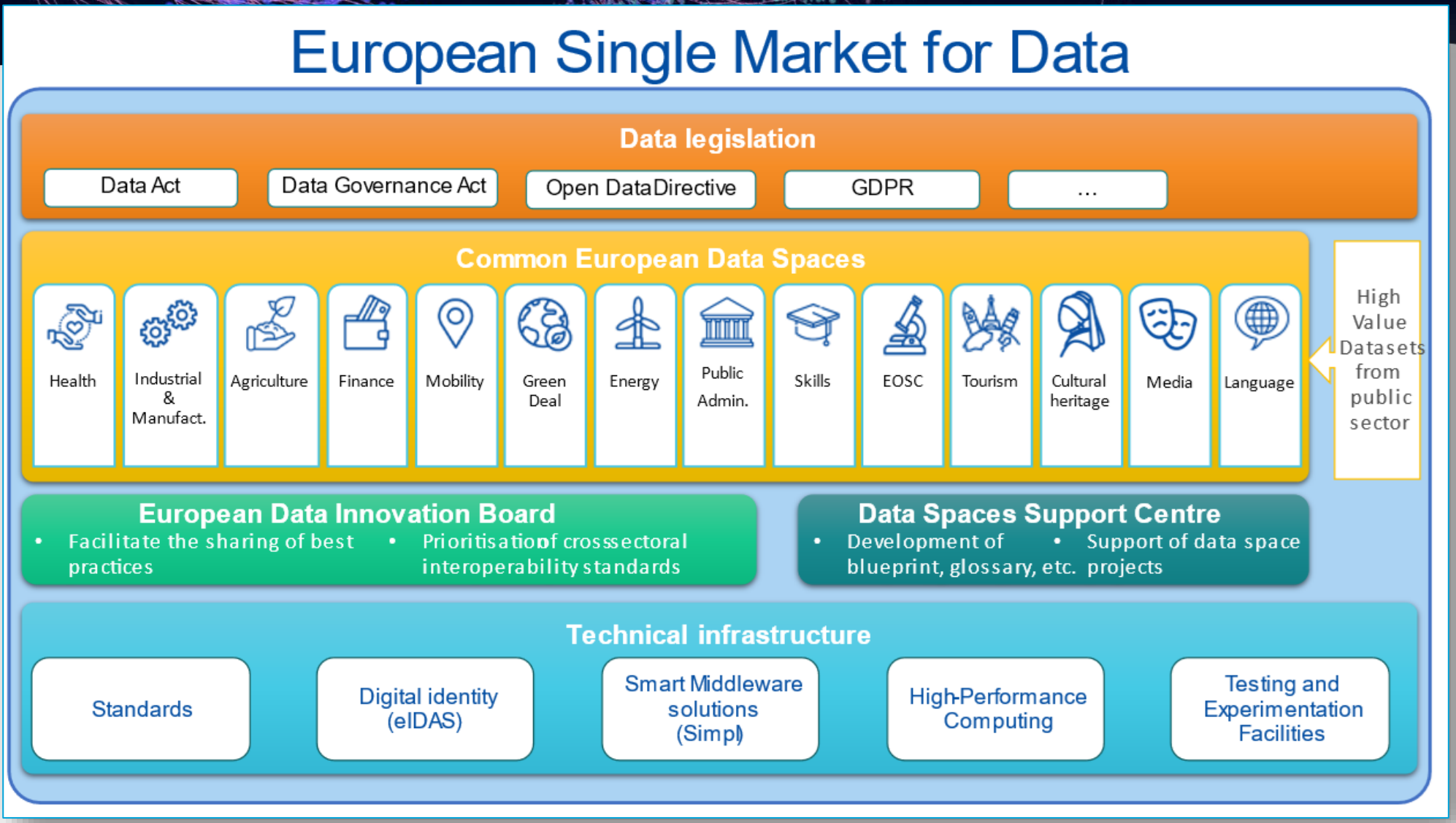
- Use the QR code
- Or go on slido.com
  #SEMIC2024Workshops
- Select the correct workshop

Data Act article 33

Coen Janssen

European Commission, DG Connect, Data Policy & Innovation

European Single Market for Data



Data Act article 33

Essential requirements regarding the interoperability of data, data sharing, and common European data spaces

- FAIR principles + Semantic interoperability
- Point 3 introduces harmonised standards as way to conform to the essential requirements
- Point 4 introduces a standardisation request

1. Participants in data spaces that offer data or data services to other participants shall comply with the following essential requirements (…) :

(a) the dataset content, use restrictions, licences, data collection methodology, data quality and uncertainty shall be sufficiently described, (…) to allow the recipient to find, access and use the data;

(b) the data structures, data formats, vocabularies, classification schemes, taxonomies and code lists, where available, shall be described in a publicly available and consistent manner;

(c) the technical means to access the data, such as application programming interfaces, and their terms of use and quality of service shall be sufficiently described to enable automatic access and transmission of data between parties (…)

(d) where applicable, the means to enable the interoperability of tools for automating the execution of data sharing agreements, such as smart contracts shall be provided.

Approach for semantic interoperability

Selection criteria for trustworthy data models, vocabularies, …

- Addressing aspects such as quality, completeness, development and long-term maintenance
- Leverage SEMIC framework

Documentation requirements on shared data

- Specify minimum requirements
- Semantic annotation
- JSON-LD, but also annotation of other formats, e.g. CSV, XML, JSON (without requiring transformations)

Based on recommendations of Workstream 14 on Data Interoperability of the High-Level Forum on European Standardisation.

Report: https://ec.europa.eu/docsroom/documents/58914

Thank you!

26 JUNE 2024

data.europa.eu – High-Value Datasets

Dr. Simon Steuer,
Head of Sector, Publications Office of the EU

SEMIC2024

interoperable europe

Purpose of High-Value Datasets

Forerunner for future datasets

HVD regulation

Article 3

Arrangements for publication applicable to all categories of high-value datasets

1. Public sector bodies holding high-value datasets listed in the Annex shall ensure that the datasets described or referenced in the Annex are made available in machine-readable formats via APIs corresponding to the reasonable needs of re-users. Where indicated in the Annex, the datasets shall also be made available as a bulk download.

2. Those public sector bodies referred to in paragraph 1 shall set out and publish the terms of use of the API and the quality of service criteria on its performance, capacity and availability. The terms of use shall be available in a human-readable and machine-readable format. Both the terms of use and the quality of service criteria shall be compatible with the arrangements for the re-use of high-value datasets laid down in accordance with Article 4.

3. API terms of use shall be accompanied by API documentation in a Union or internationally recognised open, human-readable and machine-readable format.

4. Public sector bodies referred to in paragraph 1 shall designate a point of contact for questions and issues related to the API with a view to ensure the availability and maintenance of the API and ultimately the smooth and effective publication of the high-value datasets.

5. Public sector bodies holding high-value datasets listed in the Annex shall ensure that the datasets are denoted as high-value datasets in their metadata description.

Article 4

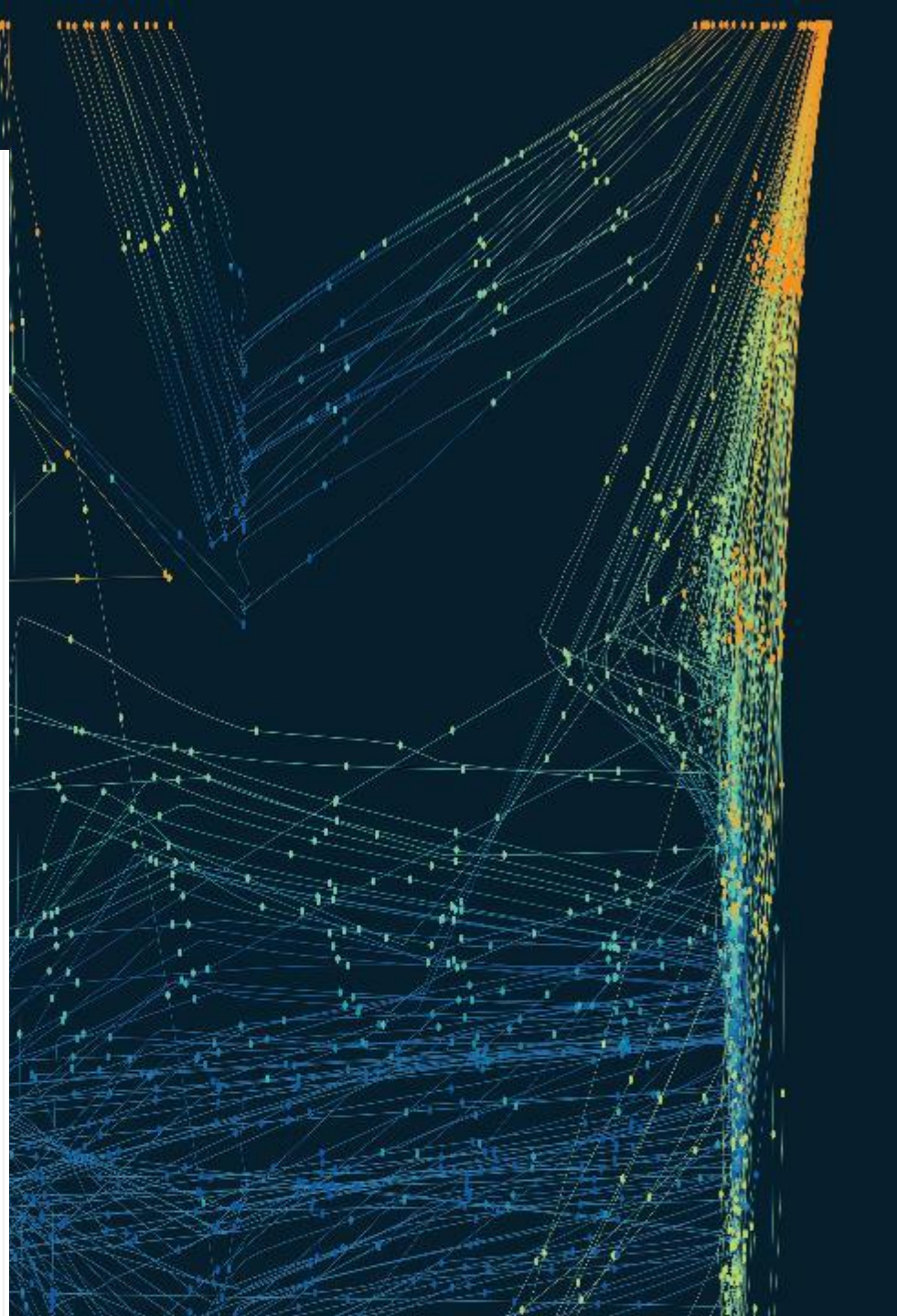
Arrangements for re-use applicable to all categories of high-value datasets

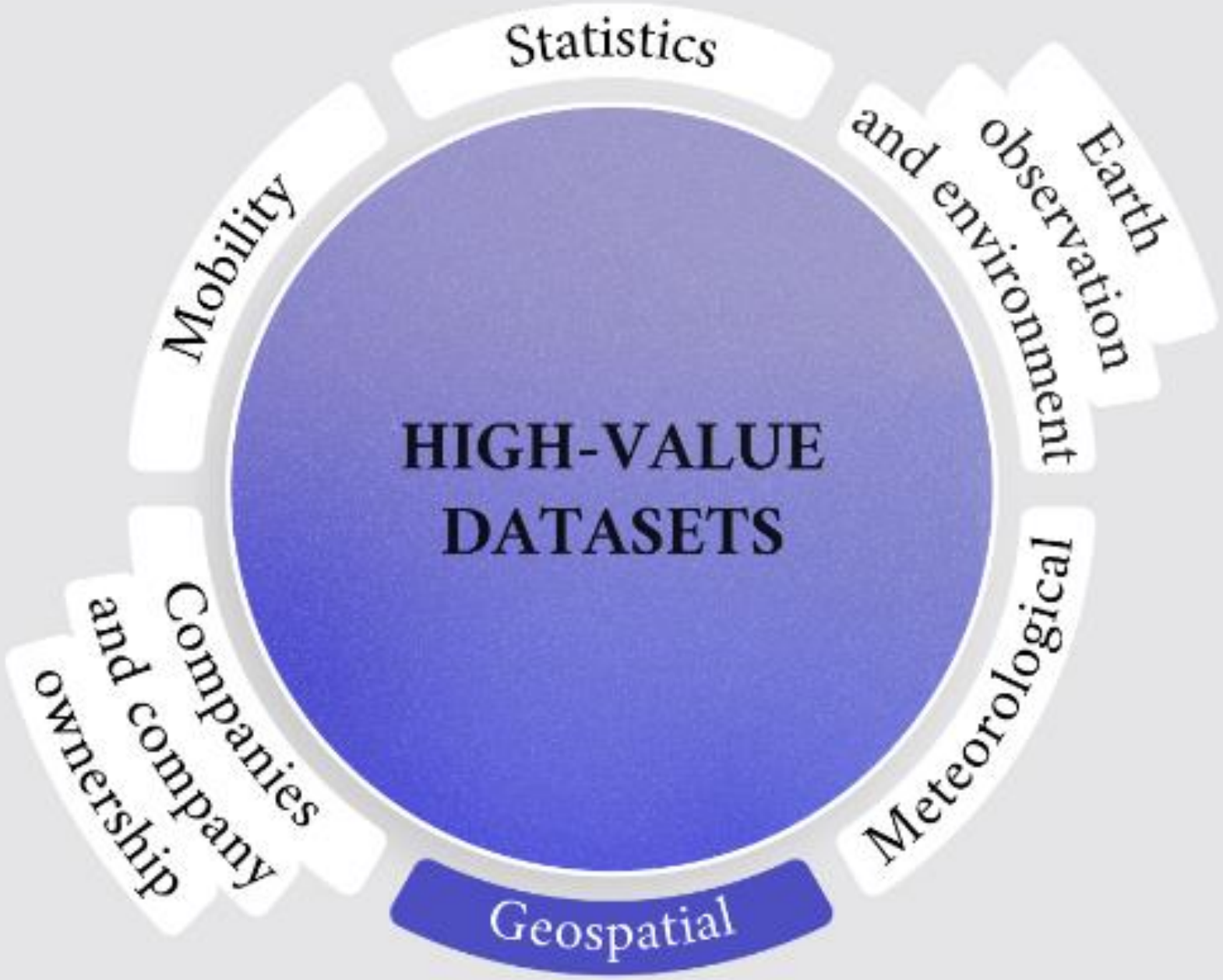
1. An exemption granted by a Member State in accordance with Article 14(5) of Directive (EU) 2019/1024 shall be published online, in the same way as the list of public sector bodies referred to in Article 6(3) of Directive (EU) 2019/1024.

2. To facilitate the availability of datasets for re-use covering longer periods of time, the obligations imposed under this Regulation shall also apply to existing machine-readable high-value datasets created before the date of application of this Regulation.

3. High-value datasets shall be made available for re-use under the conditions of the Creative Commons Public Domain Dedication (CC0) or, alternatively, the Creative Commons BY 4.0 licence, or any equivalent or less restrictive open licence, as set out in the Annex, allowing for unrestricted re-use. A requirement of attribution, giving the credit to the licensor, can additionally be required by the licensor.

4. High-value datasets shall be made available in accordance with the arrangements for the publication and re-use as set out in the Annex.



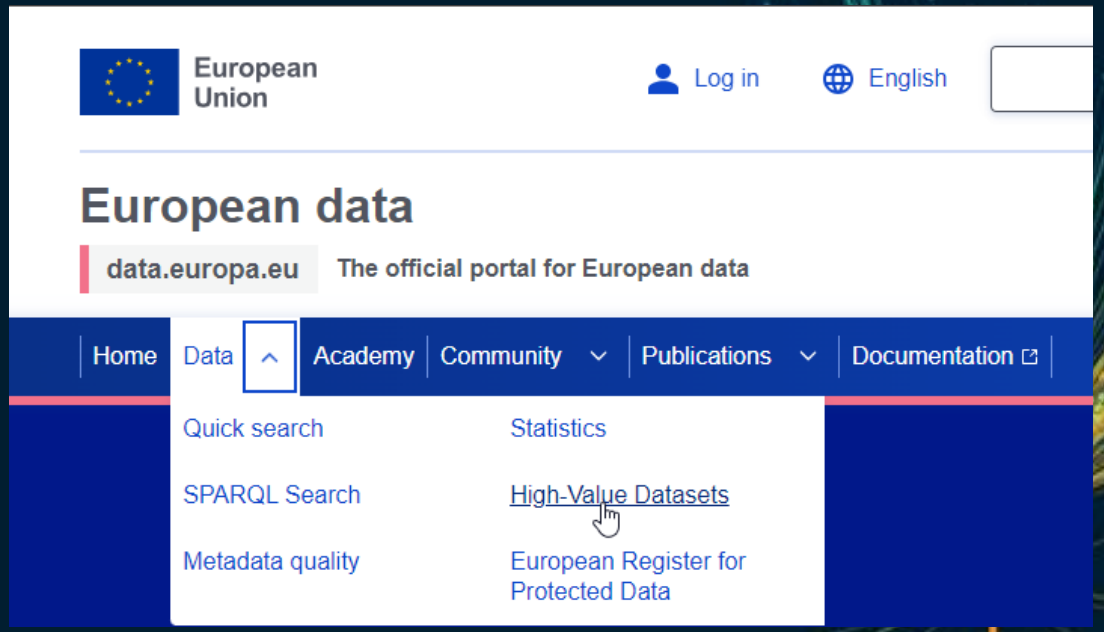

Statistics
Earth observation and environment
Meteorological
Geospatial
Companies and company ownership
Mobility
HIGH-VALUE DATASETS

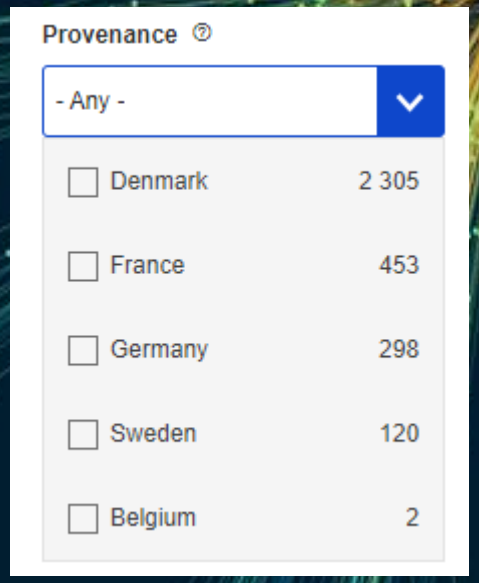
Activities around HVD

- HVD’s are defined in DCAT-AP 2.2
- European Commission organised multiple webinars to present and clarify HVDs
- 5 Member States are publishing HVDs on data.europa.eu already
- All HVD metadata is currently automatically translated
- European Commission is developing SPARQL queries to fulfil reporting requirements for HVDs

§ **6.2.3 HVD category**

Property	*HVD category*
URI	http://data.europa.eu/r5r/hvdCategory
label	HVD category
definition	A data category defined in the High Value Dataset Implementing Regulation.
usage	For the possible values consult the regulation at http://data.europa.eu/eli/reg_impl/2023/138/oj. Or consulted the controlled vocabulary derived from it.
domain	rdfs:Resource
range	skos:Concept





European data

data.europa.eu The official portal for European data

Home | Data | Academy | Community | Publications | Documentation

Home > Datasets

Datasets

High-Value Dataset

HVDs only ◉ Yes ○ No

High-Value Dataset Theme

- Any -

Data scope

- Any -

Provenance

- Any -

Publisher

- Any -

More filters

Search | Datasets

Datasets found (3 178) Sort by: Relevance

High-Value Dataset: true ×

Certain public sector data are particularly interesting for creators of value-added services and applications and have important benefits for society, the environment and the economy. That is why they should be made available to the public under conditions that make their reuse easier. Commission Implementing Regulation laying down a list of specific high-value datasets is set up under the Open Data Directive, which defines six categories of such high-value datasets: geospatial, earth observation and environment, meteorological, statistics, companies and mobility. These datasets have to be available in machine-readable format, via an Application Programming Interface and, where relevant, as bulk download. For further information on the implementation of the EU legislation on open data go to Open data - Shaping Europe´s digital future (europa.eu) and to Open data and high-value datasets: step-by-step access guide.

Bauwerke

Dieser Datensatz umfasst die Bauwerke in der Hanse- und Universitätsstadt Rostock mit Informationen zu Art und Objekthöhe in der Einheit Meter. Die…

UNKNOWN Esri Sha GeoJSON CSV Excel XL GML KML

GovData

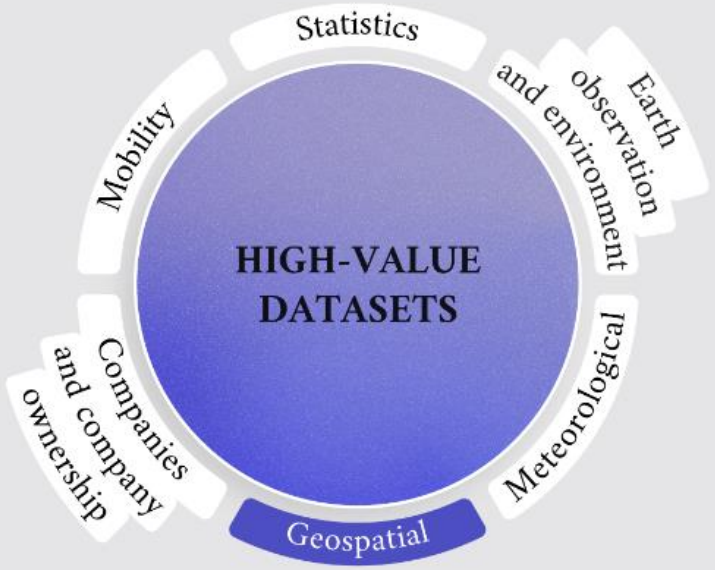
Light signalling systems

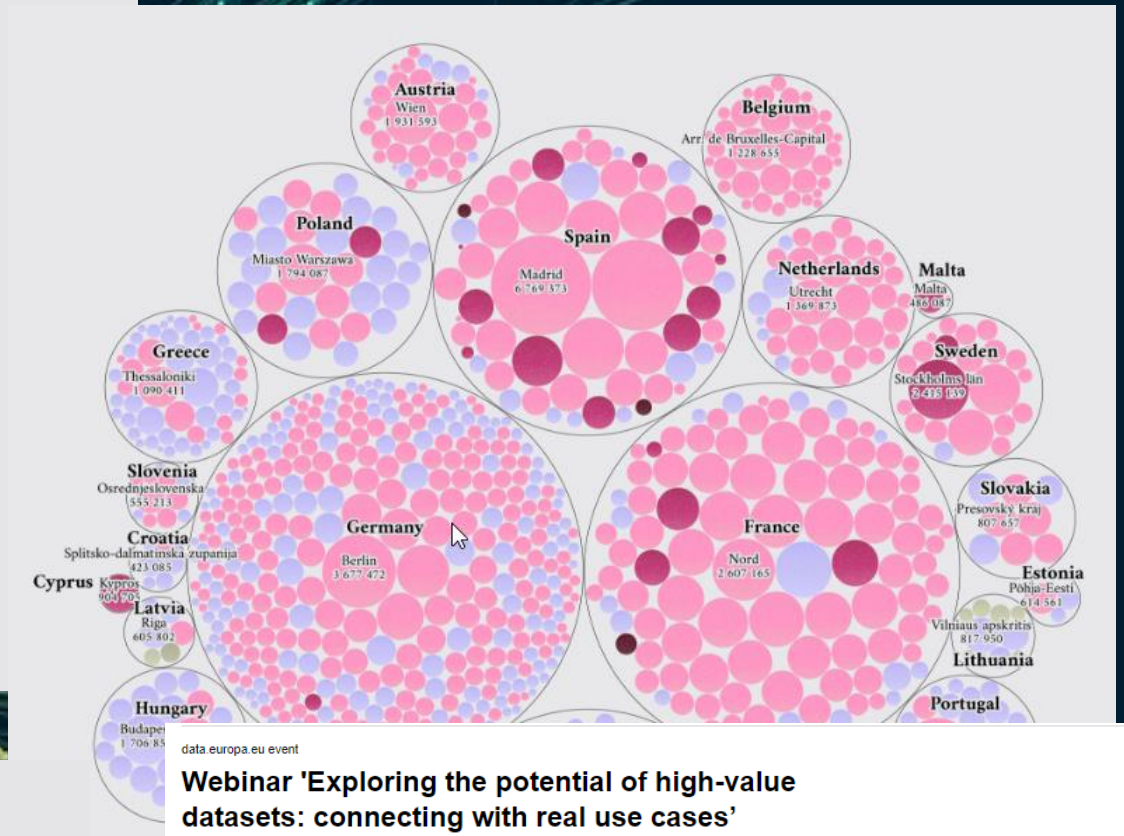
This data set includes the locations of the light signal control units owned by the Hanseatic and University City of Rostock, managed by Stadtwerke Rostock AG, wi…

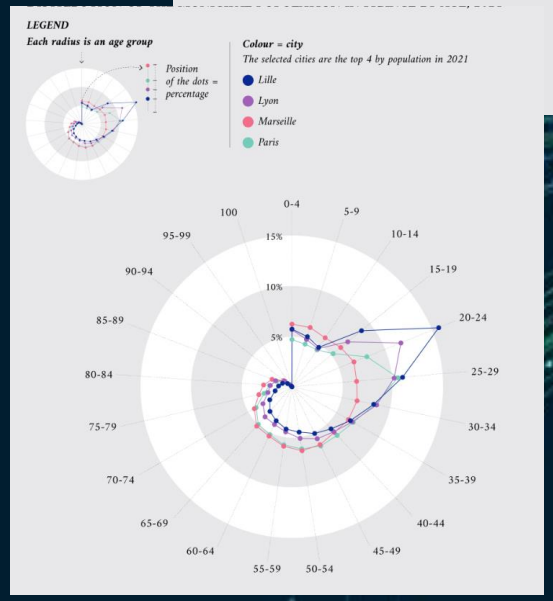
UNKNOWN KML Excel XL CSV GeoJSON GML

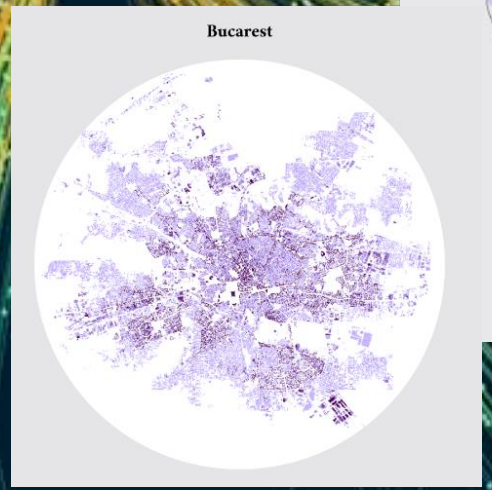
GovData

HIGH-VALUE DATASETS









data.europa.eu event

Webinar 'Exploring the potential of high-value datasets: connecting with real use cases'

14 June 2024

Online

Are you interested in leveraging high-value datasets to enhance your research or business? Join our webinar, 'Exploring the potential of high-value datasets: connecting with real use cases', and learn from experts about the current landscape and practical applications of these datasets across Europe.

Scheduled for 14 June from 10:00 to 11:30 CET, this webinar is ideal for researchers, business professionals, and data enthusiasts. It will review the report 'High-Value Datasets Best Practices in Europe', highlighting methodologies and key insights from leading Member States. You'll discover how high-value datasets drive decision-making and operational efficiency.

Hear real-world examples from European countries showcasing their best practices with open data. These stories will provide actionable insights and inspiration for your projects.

Curious to learn more? Mark your calendars for 14 June from 10:00 to 11:30 CET and join us for an engaging discussion on high-value datasets and their role in the digital age!

Learn more about Webinar 'Exploring the potential of high-value datasets: connecting with real use cases'

ANALYTICAL REPORTS |

High-value Datasets Best Practices in Europe

This report explores Commission Implementing Reg
(EU) 2023/138, which complements the open data d
by establishing specific requirements for high-value
(HVDs). With the regulation taking effect on 9 June
focuses on highlighting best practices and strategies
EU Member States in meeting these requirements.
interviews with representatives from seven Member
the analysis identifies also common challenges, and
provides a roadmap for Member States to achieve
compliance, aiming to improve public sector efficien
economic opportunities, and cross-border collaborat
findings reveal varying degrees of adoption and cha
HVD availability and use.

(0.97 MB - PDF) Download

14 June 2024

Unveiling the high-value datasets: HVD Best Practices Report

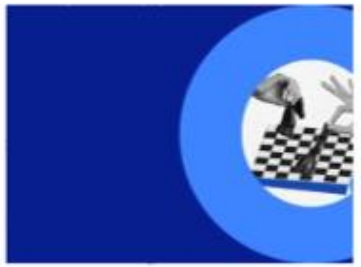
In an era of digital transformation, the European Union is at the forefront of unlocking the potential of public sector data, with high-value datasets (HVDs) being particularly impactful due to their substantial social and economic benefits. As these datasets are essential for enhancing digital services, the HVD Implementing Regulation was introduced to standardise the publication and

spite…

Events

Webinar 'Exploring the potential of high-value datasets: connecting with real use cases'

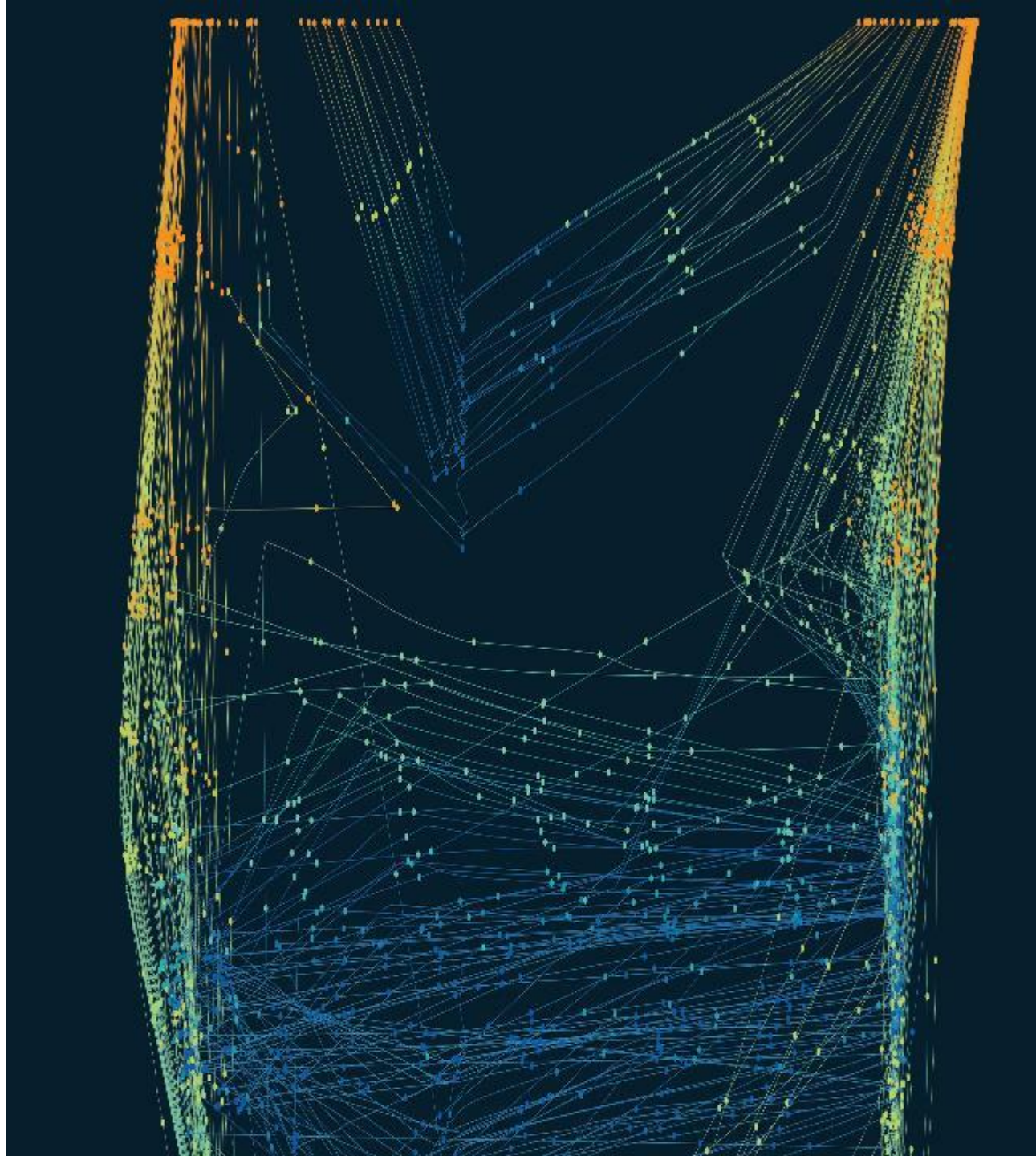
High-value datasets are essential for creating a European data-driven market but what is the state of play? Join our webinar to learn about best practices from European countries and see how businesses and researchers are leveraging this data to create insights and value-added services. Don't miss this chance to connect with experts and explore the potential for growth and impact in the digital age.

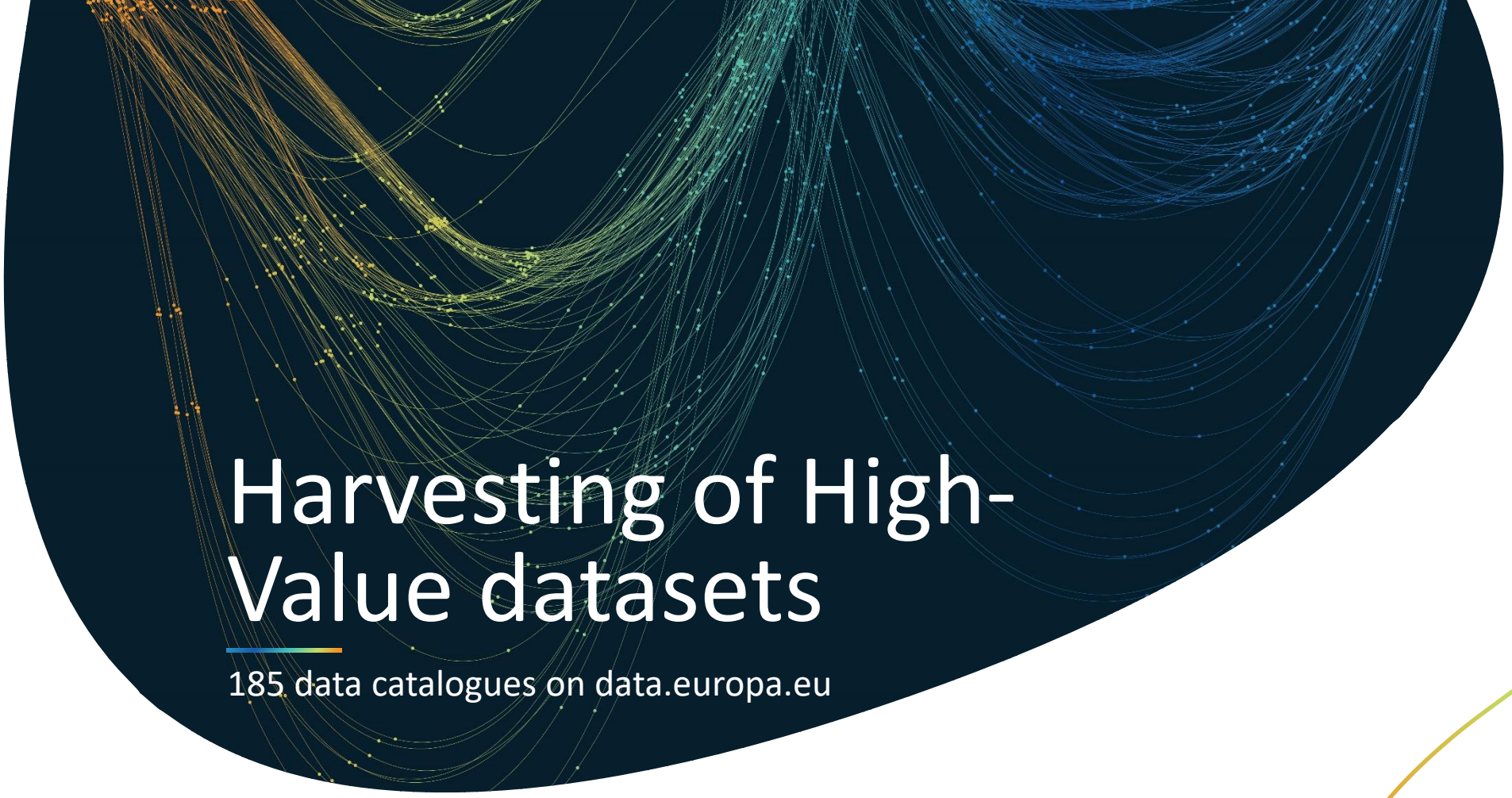


Date	14/06/2024
Registration	Register here

HVD and data spaces

- HVD’s can be a blueprint for other valuable data assets to be published
- Important to apply same metadata standard and availability via API and bulk download
- Mobility HVD's could be part of the Mobility Data Space (Interoperability is key)
- Decentral approach to data publishing and data sharing is crucial

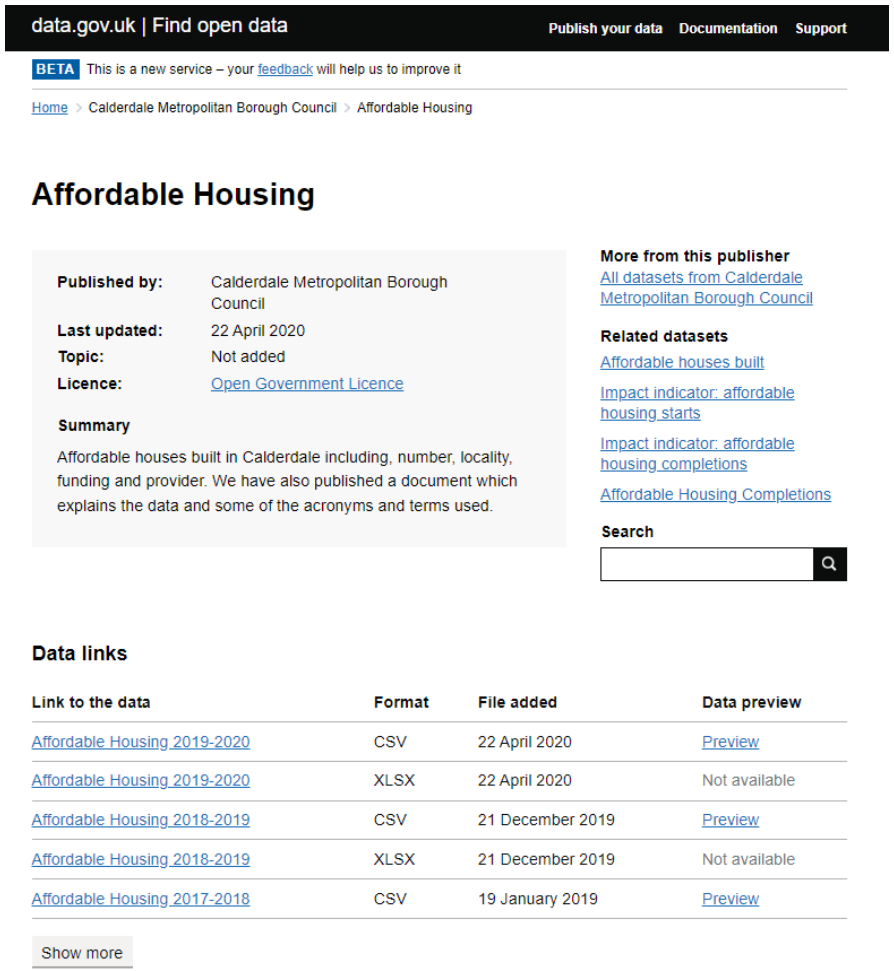


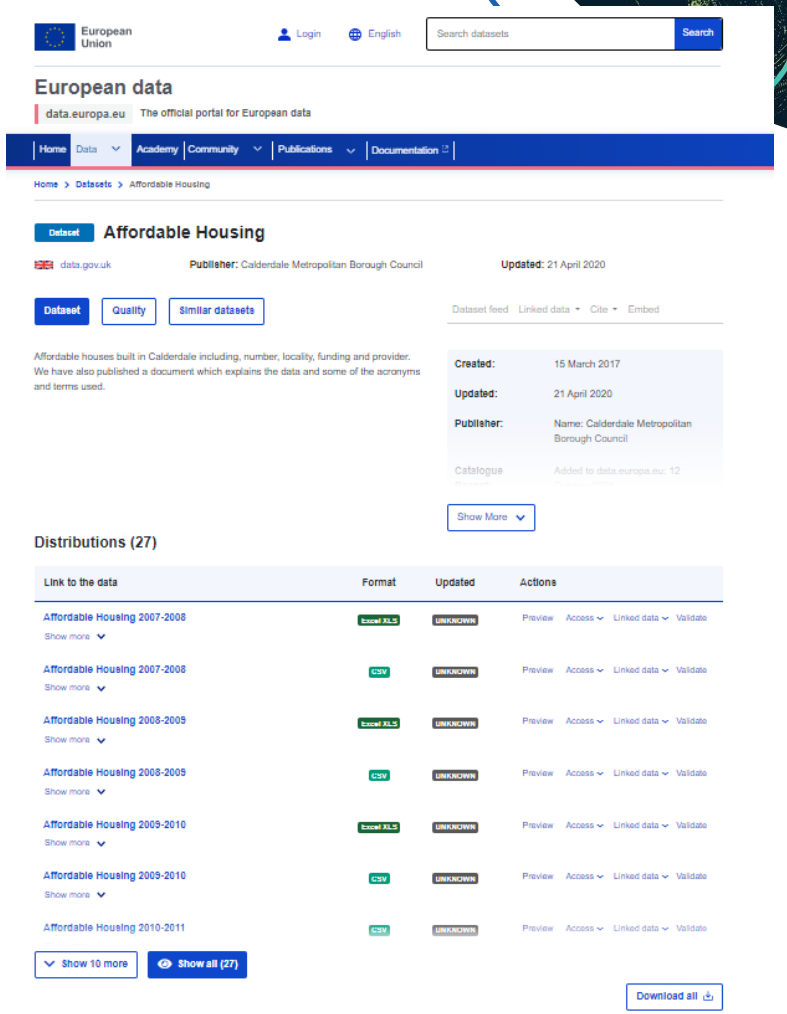


Harvesting of High-Value datasets

185 data catalogues on data.europa.eu

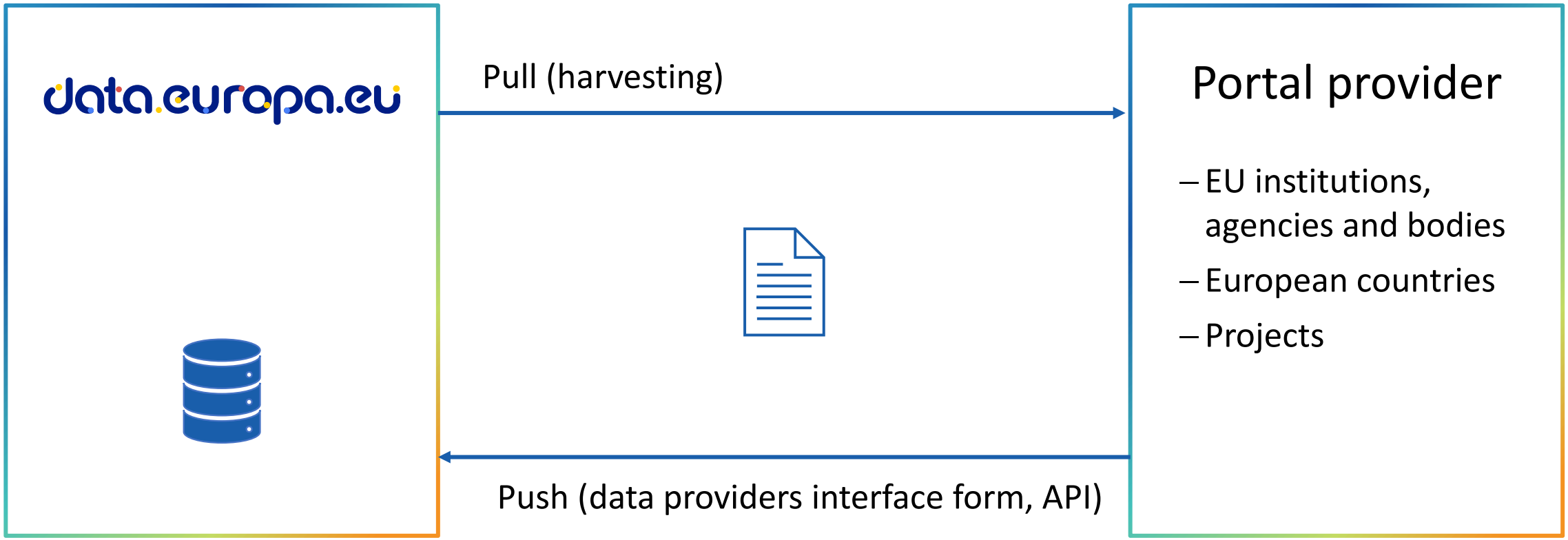
Current Data harvesting Processes



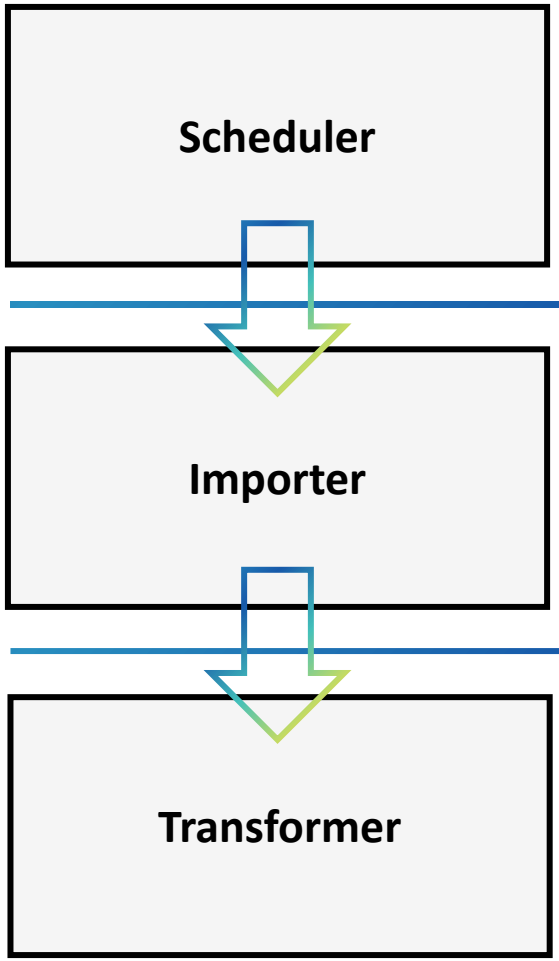


European Commission

Metadata retrieval



Data acquisition

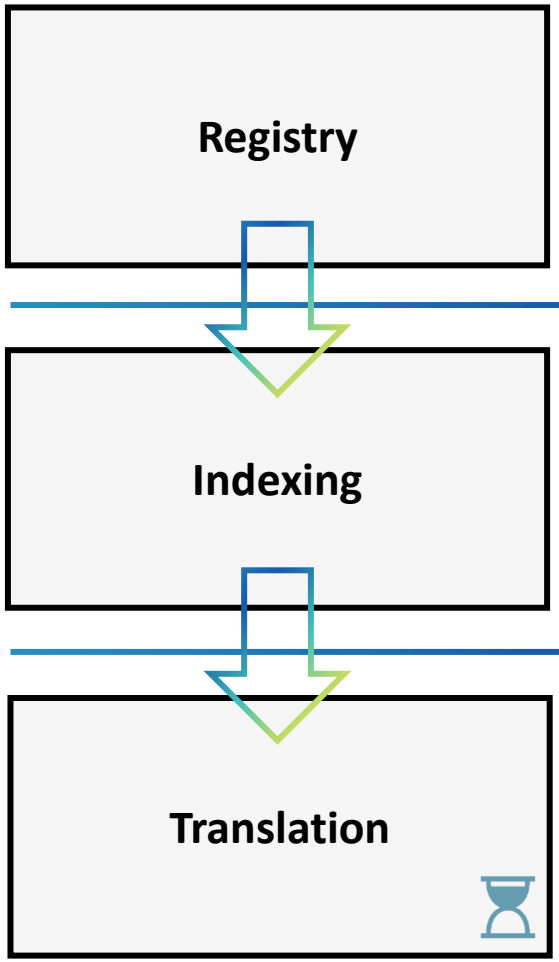


- Main entry point for the service orchestration.
- Periodically triggers the harvesting process, defined as a **pipeline descriptor.**
- Frequency: hourly, daily, weekly… depends on data provider

- Retrieves the metadata from the source portal(s).
- Support for a variety of interfaces and data formats:
- **Responsive API that provides DCAT-AP and supports paging is preferred (e.g. RDF / XML)**

- Applies lightweight scripting-based transformation rules.
- Rules are written in JavaScript or XSLT.
- The final output is “DCAT-AP-compliant” RDF.
- The scripts can be managed externally (e.g. in Git) to ensure maintainability.

Processing and storing

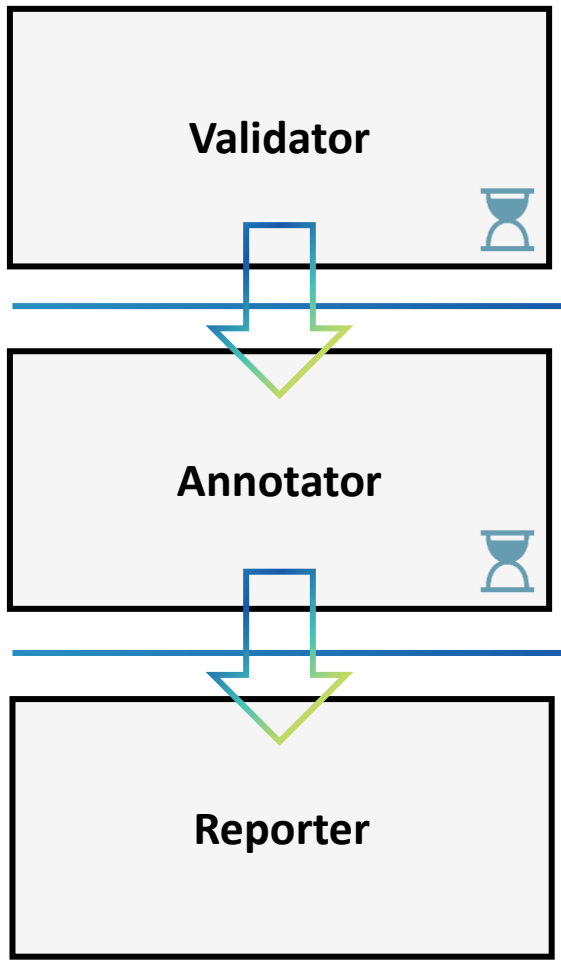


- Middleware and abstraction layer to interact with the triple store (**Virtuoso**).
- RESTful interface for RDF (**Turtle, JSON-LD, N-Triples, RDF/XML, Notation3**).
- Application of URI schemata, generation of unique IDs and inter-linking.

- Responsible for managing the high-performance search index (**Elasticsearch**).
- “Flattening” of the DCAT RDF to simple JSON.
- Extracting literals from the data, e.g. from properties like title and description.
- Supports the use of existing and **vocabularies and ontologies**.

- Middleware to eTranslation
- Bundling literals from multiple datasets to an integrated request.
- Returns the translation by applying the native multi-language features of RDF.
- Translates description and title from datasets and distributions.

Quality evaluation



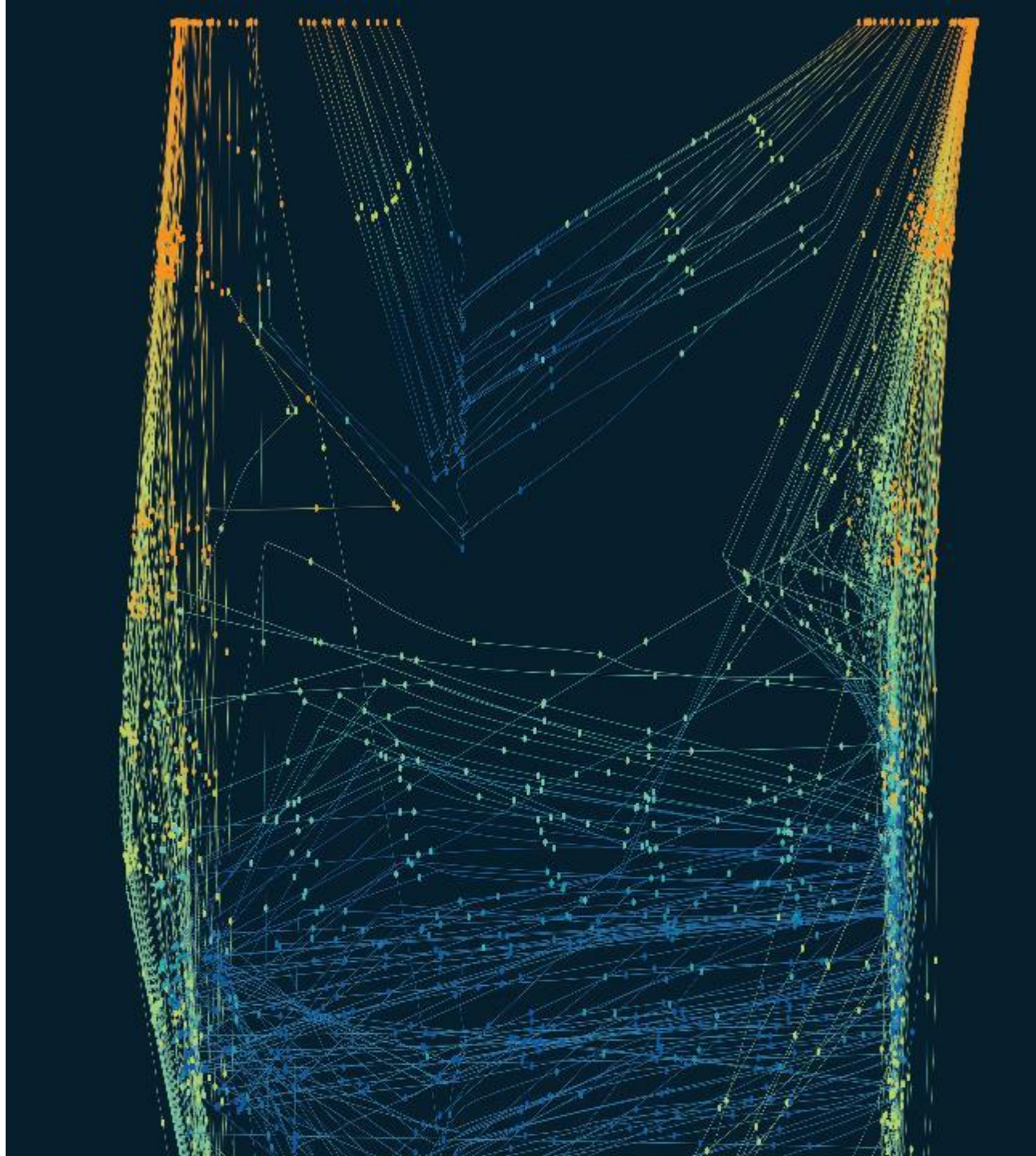
- Application of the **W3C SHACL**.
- Results include detailed information violations.
- Applied rules can also be extended or replaced (Built-in **DCAT-AP**).
- Accessibility tests on each linked distribution (the actual data).

- Quality assessment for each dataset with a custom metrics scheme.
- Inspired by the **FAIR** principles.
- Completeness of the metadata, evaluating the format and type of data, availability of licensing information and linked distributions.

- Applies W3C Data Quality Vocabulary (**DQV**) for creating quality reports.
- Based on the results of the Validator and Annotator.
- Attached as RDF to the concerned dataset in the triplestore.
- Offers a variety of human-readable versions (PDF, XLS, ODS, HTML).

Identifier handling

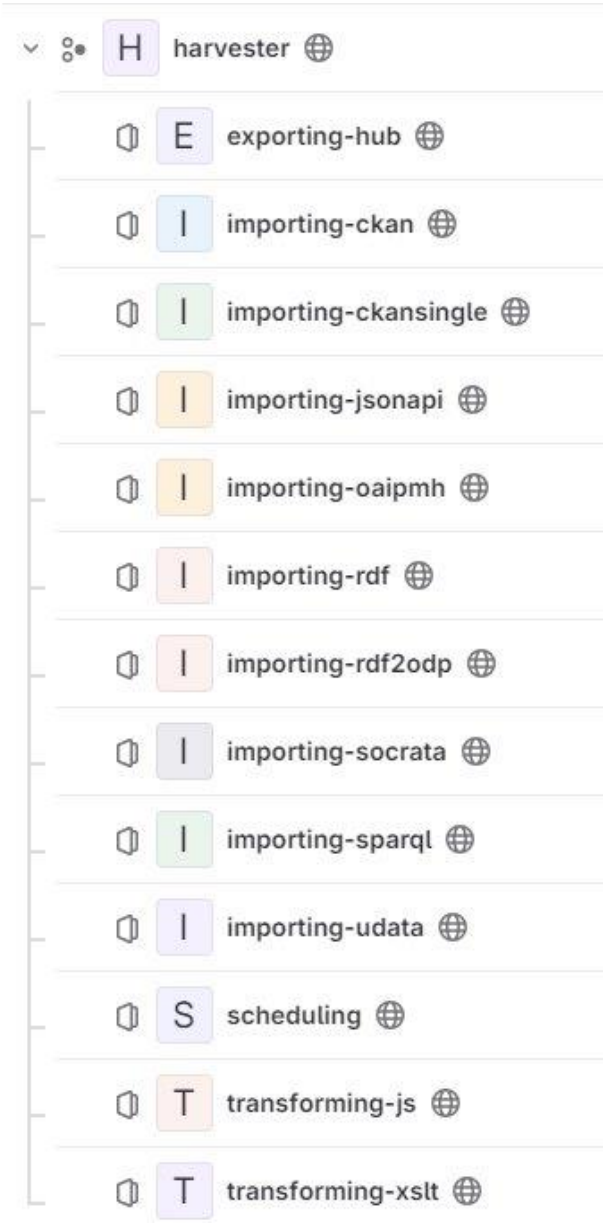
- When harvesting, we always store the original identifier in "dct:identifier"
- For internal handling, we create an additional identifier
- A URIRef based on the original identifier (our baseURI + “normalised” identifier)
- If the new identifier already exists, we add an increment at the end
- Property "applicable:legislation" helps to spot HVDs

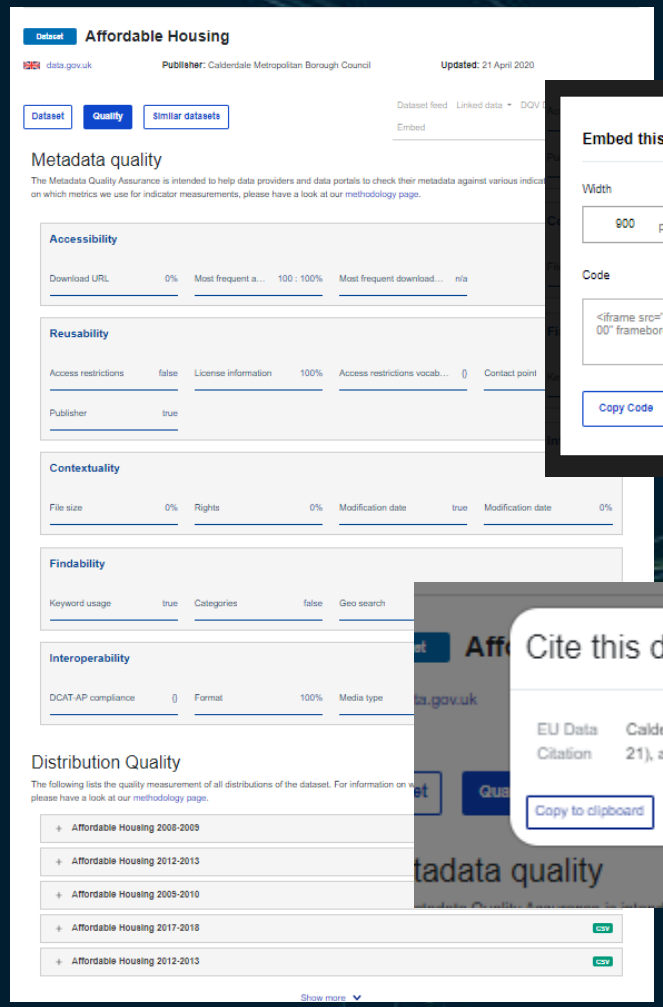


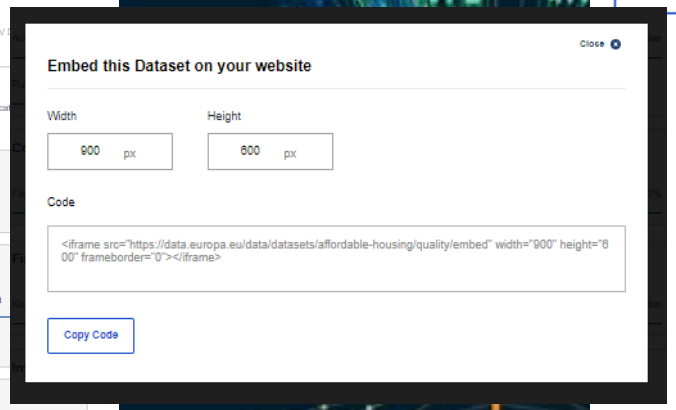
- A pipeline orchestration is described by a **descriptor**: a plain JSON document.
- It includes list of **segments**, where each segment describes a step aka a service.
- The descriptor is a compilation and self-contained description of a data processing chain.
- Each microservice must expose an endpoint to receive the descriptor and must be able to parse and execute its content.
- Data itself can be embedded directly into the descriptor or passed via a pointer to a separate data store.

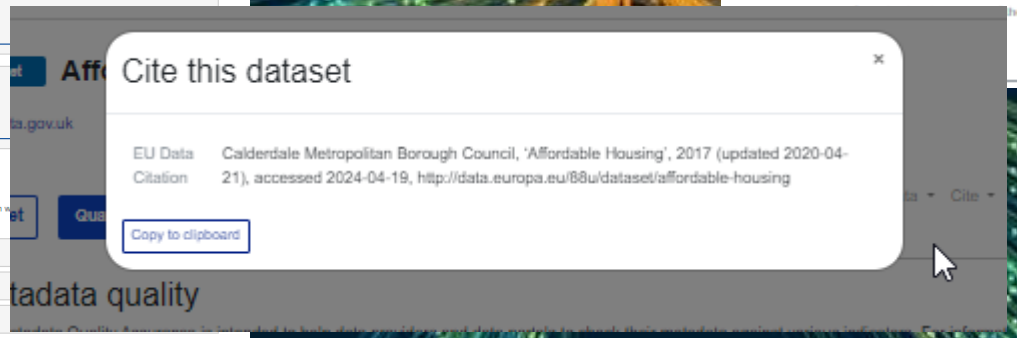
Software stack

- Reactive Java framework **Vert.x** and employment of an asynchronous programming paradigm
- DevOps-based **Microservice** approach
- Deployment via **Docker** and support for container-orchestration like **Kubernetes**
- Virtuoso triple store a primary database and **Elasticsearch** as search server
- Modern Single-Page-Application frontend based on **Vue.js**

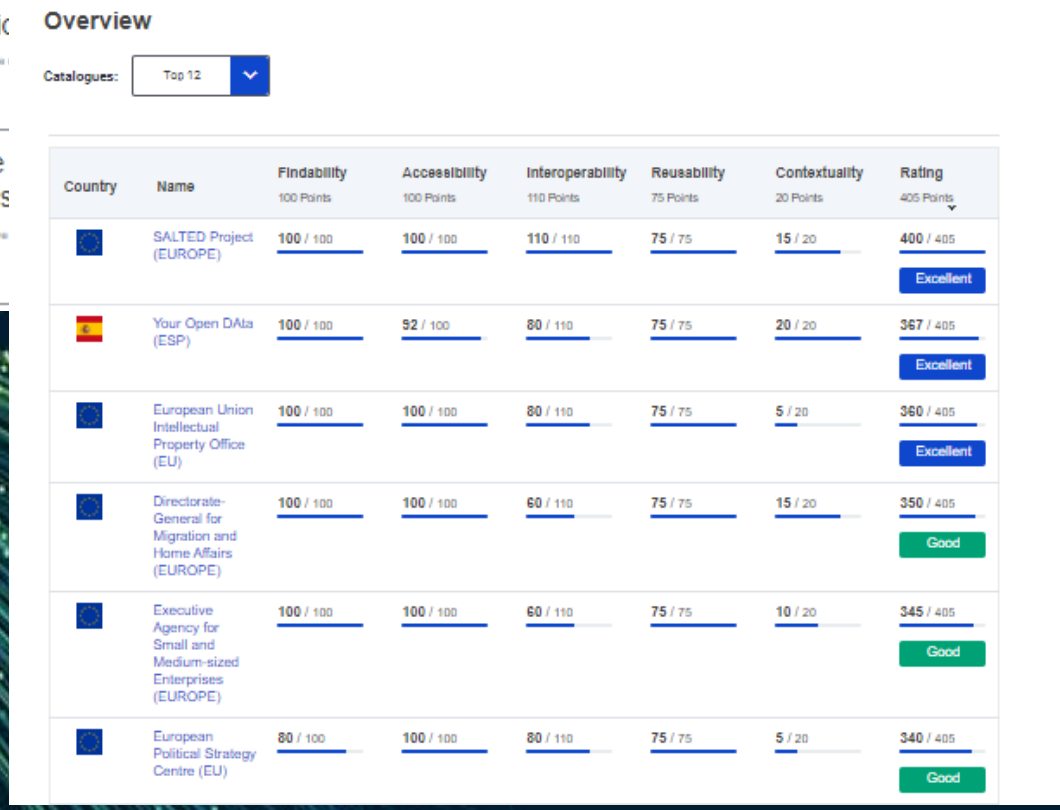


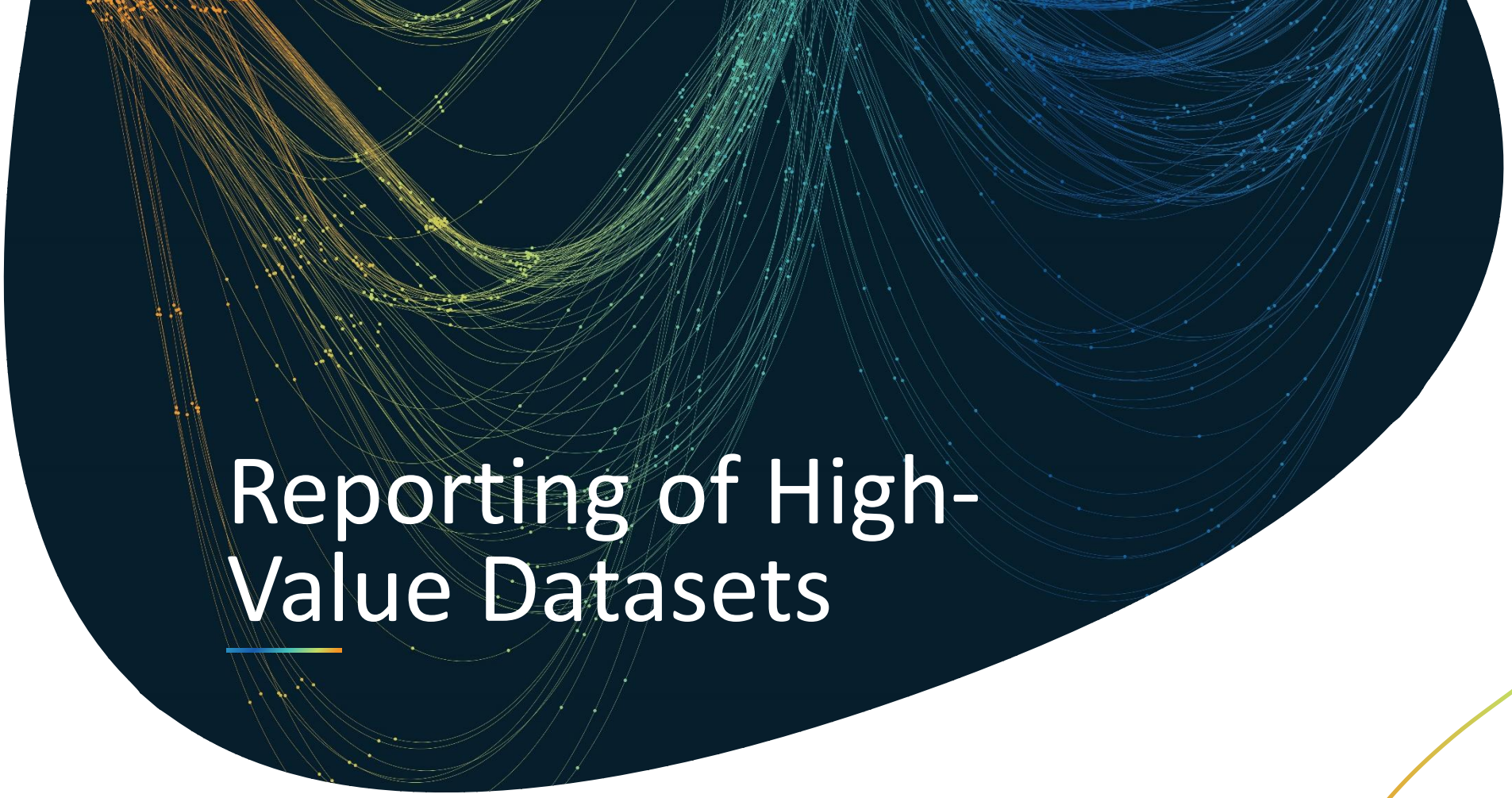

Dataset
Affordable Housing
data.gov.uk
Publisher: Calderdale Metropolitan Borough Council
Updated: 21 April 2020
Dataset
Quality
Similar datasets
Metadata quality
Accessibility
Reusability
Contextuality
Findability
Interoperability
Distribution Quality
Affordable Housing 2008-2009
Affordable Housing 2012-2013
Affordable Housing 2009-2010
Affordable Housing 2017-2018
Affordable Housing 2012-2013
Show more


Embed this Dataset on your website
Close
Width
900 px
Height
600 px
Code
<iframe src="https://data.europa.eu/data/datasets/affordable-housing/quality/embed" width="900" height="600" frameborder="0"></iframe>
Copy Code


Cite this dataset
EU Data Citation
Calderdale Metropolitan Borough Council, 'Affordable Housing', 2017 (updated 2020-04-21), accessed 2024-04-19, http://data.europa.eu/88u/dataset/affordable-housing
Copy to clipboard

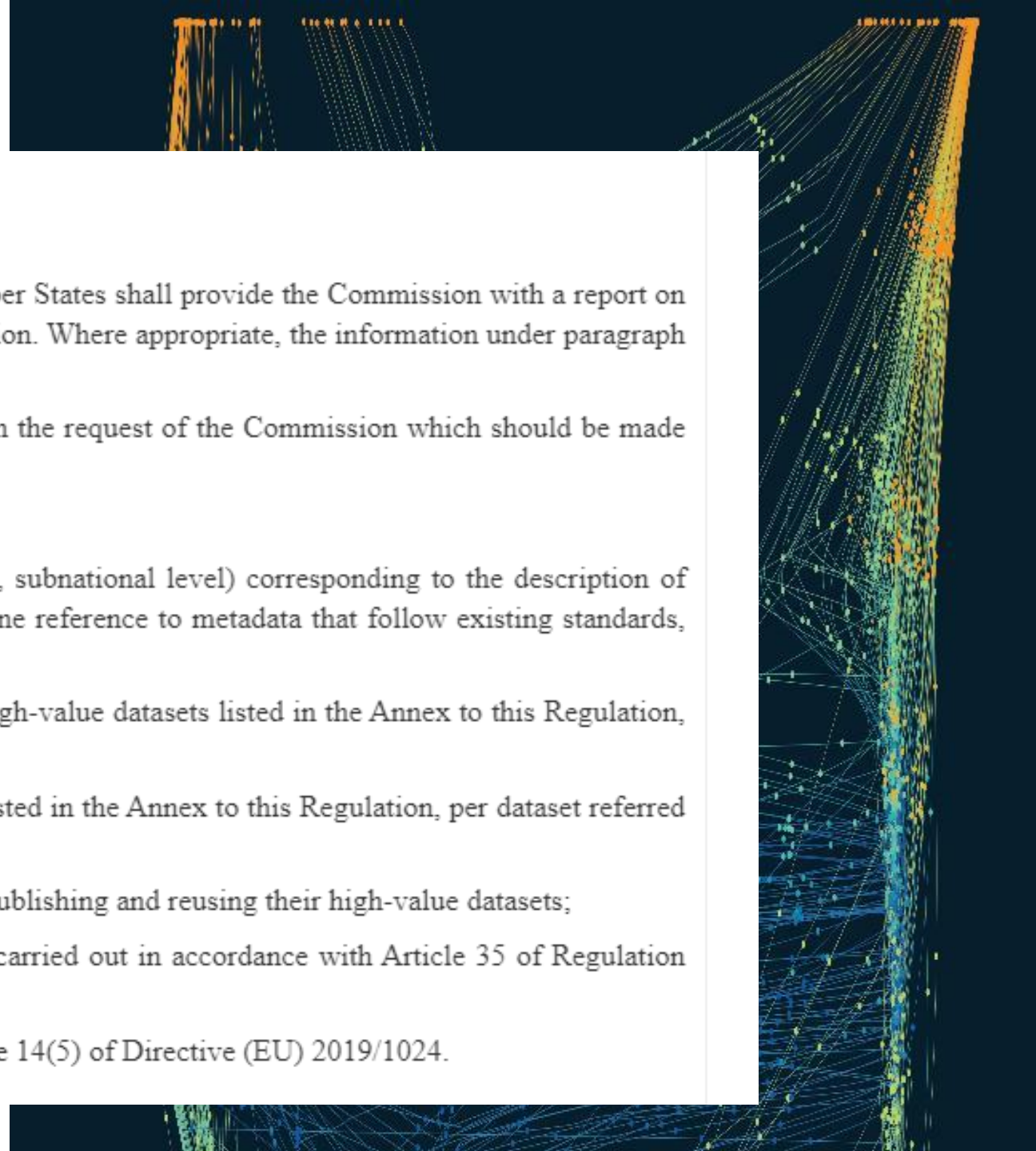

Dataset
Affordable Housing
data.gov.uk
Publisher: Calderdale Metropolitan Borough Council
Updated: 21 April 2020
Dataset
Quality
Similar datasets
Dataset feed
Linked data
Cite
Embed
Similar datasets
Main dwellings according to useful surface by town size (API identifier: /t20/e244/viviendas/p01/l0/01002.px)
Marginal Similar


Overview
Catalogues:
Top 12
Country
Name
Findability
100 Points
Accessibility
100 Points
Interoperability
110 Points
Reusability
75 Points
Contextuality
20 Points
Rating
405 Points
SALTED Project (EUROPE)
100 / 100
100 / 100
110 / 110
75 / 75
15 / 20
400 / 405
Excellent
Your Open DAta (ESP)
100 / 100
92 / 100
80 / 110
75 / 75
20 / 20
367 / 405
Excellent
European Union Intellectual Property Office (EU)
100 / 100
100 / 100
80 / 110
75 / 75
5 / 20
360 / 405
Excellent
Directorate-General for Migration and Home Affairs (EUROPE)
100 / 100
100 / 100
60 / 110
75 / 75
15 / 20
350 / 405
Good
Executive Agency for Small and Medium-sized Enterprises (EUROPE)
100 / 100
100 / 100
60 / 110
75 / 75
10 / 20
345 / 405
Good
European Political Strategy Centre (EU)
80 / 100
100 / 100
80 / 110
75 / 75
5 / 20
340 / 405
Good



Reporting of High-Value Datasets

HVD regulation



Article 5

Reporting

1. By 2 years after entry into force of this Implementing Regulation Member States shall provide the Commission with a report on the measures they have carried out to implement this Implementing Regulation. Where appropriate, the information under paragraph 3 can be provided through references to relevant metadata.

2. Each Member State shall provide an updated version of the report upon the request of the Commission which should be made every 2 years.

3. The report shall contain the following information:

(a) a list of specific datasets at Member State level (and, where relevant, subnational level) corresponding to the description of each high-value dataset in the Annex to this Regulation and with online reference to metadata that follow existing standards, such as a single register or open data catalogue;

(b) persistent link to the licensing conditions applicable to the re-use of high-value datasets listed in the Annex to this Regulation, per dataset referred to in point a);

(c) persistent link to the APIs ensuring access to the high-value datasets listed in the Annex to this Regulation, per dataset referred to in point a);

(d) where available, guidance documents issued by the Member State on publishing and reusing their high-value datasets;

(e) where available, the existence of data protection impact assessments carried out in accordance with Article 35 of Regulation (EU) 2016/679;

(f) the number of public sector bodies exempted in accordance with Article 14(5) of Directive (EU) 2019/1024.

HVD reporting

- Reporting HVDs with key information
- Bulk downloads
- APIs for HVDs
- Legal information on distributions and APIs
- Legal information on licences

```
prefix dct: <http://purl.org/dc/terms/>
prefix r5r: <http://data.europa.eu/r5r/>
prefix dcat:  <http://www.w3.org/ns/dcat#>

select distinct ?d ?api ?title ?desc ?category ?endpointURL ?endpointDesc where {
<?MSCat?> ?cp ?d.
?d r5r:applicableLegislation <http://data.europa.eu/eli/reg_impl/2023/138/oj>.
{
?d dcat:distribution ?dist.
?dist r5r:applicableLegislation <http://data.europa.eu/eli/reg_impl/2023/138/oj>.

?dist dcat:accessService ?api.
?api r5r:applicableLegislation <http://data.europa.eu/eli/reg_impl/2023/138/oj>.
}
union {
?api dcat:servesDataset ?d.
?api r5r:applicableLegislation <http://data.europa.eu/eli/reg_impl/2023/138/oj>.
}

optional { ?api dct:title ?title.
     FILTER ( lang(?title) = "en" )
 }
optional { ?api dct:description ?desc.
     FILTER ( lang(?desc) = "en" )
 }
optional { ?api r5r:hvdCategory ?category. }
optional { ?api dcat:endpointDescription ?endpointDesc. }
optional { ?api dcat:endpointURL ?endpointURL. }
}
```


Brussels 27 June
SEMIC 2024 conference
interoperable europe
from Vision to Reality
be EU belgium24.eu
European Commission

Governance, Technical and Semantic Assets from INSPIRE

Alexander Kotsev, Jordi Escriu, Marco Minghini (EC JRC)

Outline

- INSPIRE value proposition
- Community and asset governance
- Technical and semantic assets (demos)
- Conformance and compliance
- Benefits for data spaces (discussion)

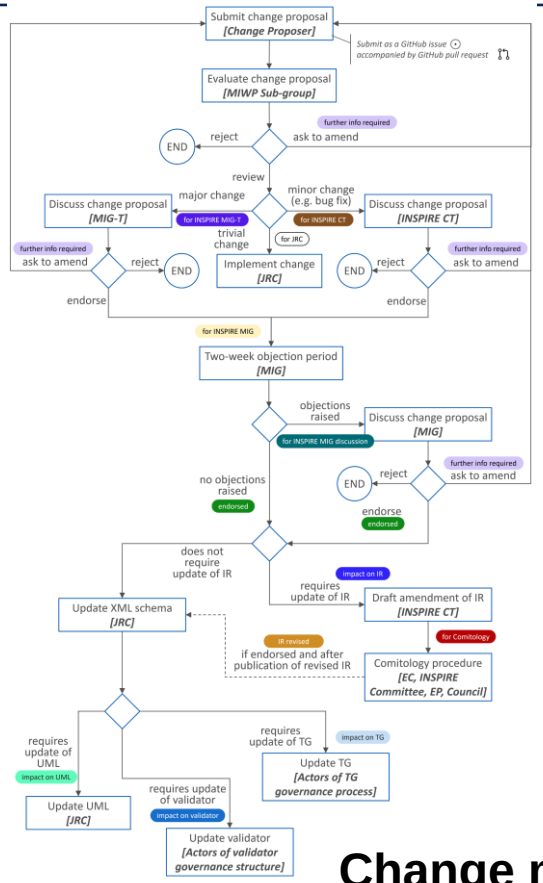
INSPIRE value proposition

- 17 years of implementation experience
- Upcoming revision under the GreenData4all initiative
- Stakeholders
  - 7000+ data providers
  - Governance structure & technical and political arm (MIG/MIG-T)
  - Community of adopters, open source and proprietary vendors, SDO
- 120K datasets documented through metadata for reuse
- Regulated semi-automated monitoring process

Semantic and Technical Assets

- Fully transparent & community driven approach
  - 28 GitHub repositories
  - Increasing number of issues & commits
- Technical guidelines (asciidoc, html, pdf)
  - Horizontal: Metadata, network services, data encoding
  - Theme specific
- Data models (uml, app schemas)
- Open source software components

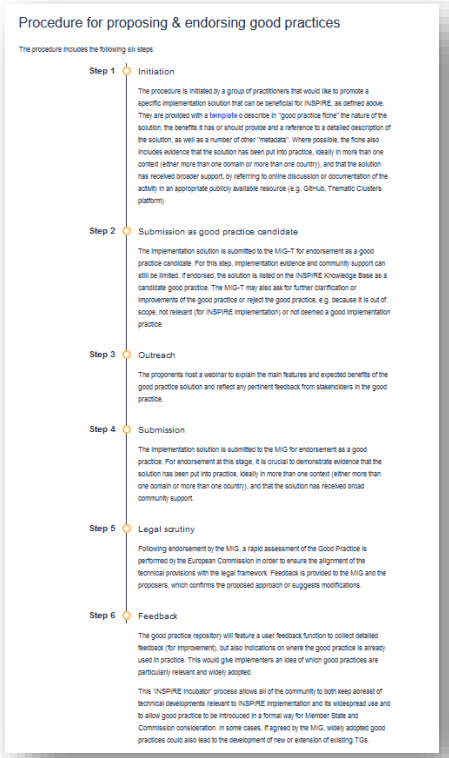
Asset governance



Change management

https://github.com/INSPIRE-MIF/application-schemas/blob/main/governance-release-process

Good practices



Procedure for proposing & endorsing good practices

The procedure includes the following six steps:

Step 1 Initiation

The procedure is initiated by a group of practitioners that would like to promote a specific implementation solution that can be beneficial for INSPIRE, as defined above. They are provided with a template o describe in "good practice fiche" the nature of the solution, the benefits it has or should provide and a reference to a detailed description of the solution, as well as a number of other "metadata". Where possible, the fiche also includes evidence that the solution has been put into practice, ideally in more than one context (either more than one domain or more than one country), and that the solution has received broader support, by referring to online discussion or documentation of the activity in an appropriate publicly available resource (e.g. GitHub, Thematic Clusters platform)

Step 2 Submission as good practice candidate

The implementation solution is submitted to the MIG-T for endorsement as a good practice candidate. For this step, implementation evidence and community support can still be limited. If endorsed, the solution is listed on the INSPIRE Knowledge Base as a candidate good practice. The MIG-T may also ask for further clarification or improvements of the good practice or reject the good practice, e.g. because it is out of scope, not relevant (for INSPIRE implementation) or not deemed a good implementation practice

Step 3 Outreach

The proponents host a webinar to explain the main features and expected benefits of the good practice solution and reflect any pertinent feedback from stakeholders in the good practice.

Step 4 Submission

The implementation solution is submitted to the MIG for endorsement as a good practice. For endorsement at this stage, it is crucial to demonstrate evidence that the solution has been put into practice, ideally in more than one context (either more than one domain or more than one country), and that the solution has received broad community support.

Step 5 Legal scrutiny

Following endorsement by the MIG, a rapid assessment of the Good Practice is performed by the European Commission in order to ensure the alignment of the technical provisions with the legal framework. Feedback is provided to the MIG and the proposers, which confirms the proposed approach or suggests modifications.

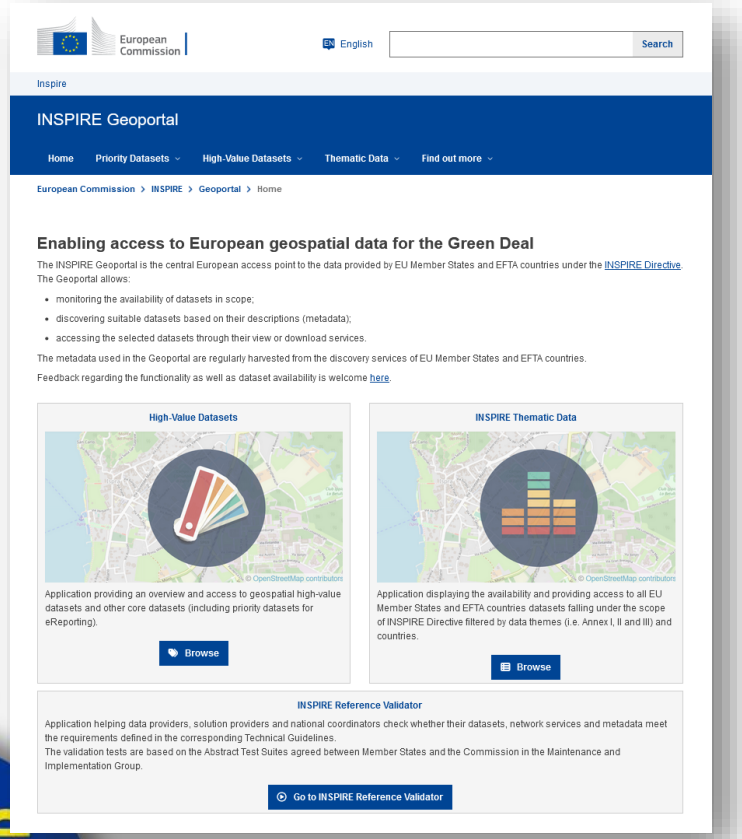
Step 6 Feedback

The good practice repository will feature a user feedback function to collect detailed feedback (for improvement), but also indications on where the good practice is already used in practice. This would give implementers an idea of which good practices are particularly relevant and widely adopted.

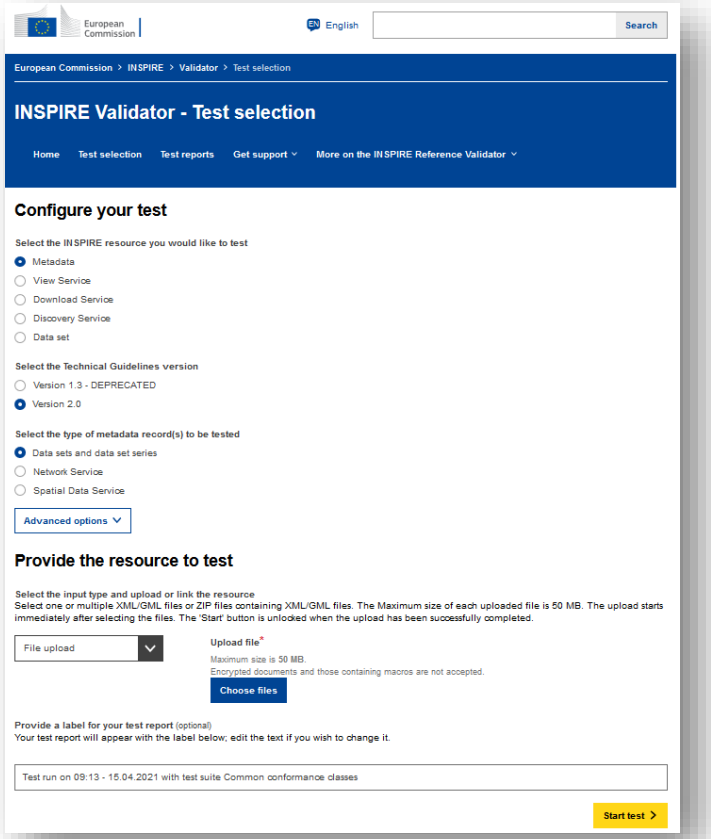
This 'INSPIRE incubator' process allows all of the community to both keep abreast of technical developments relevant to INSPIRE implementation and its widespread use and to allow good practice to be introduced in a formal way for Member State and Commission consideration. In some cases, if agreed by the MIG, widely adopted good practices could also lead to the development of new or extension of existing TGs.

Endorsed
GeoJSON encoding of INSPIRE datasets
GeoDCAT-AP
SDMX for Human Health and Population Distribution
GeoPackage encoding of INSPIRE datasets
Data-Service Linking Simplification
OGC compliant INSPIRE coverage data and service implementation
Guidance for the integration of dispersed WMS sources
Guidelines for making spatial data downloadable via WMS services
Setting up an INSPIRE Download service based on the OGC API-Features standard
INSPIRE download services based on OGC SensorThings API

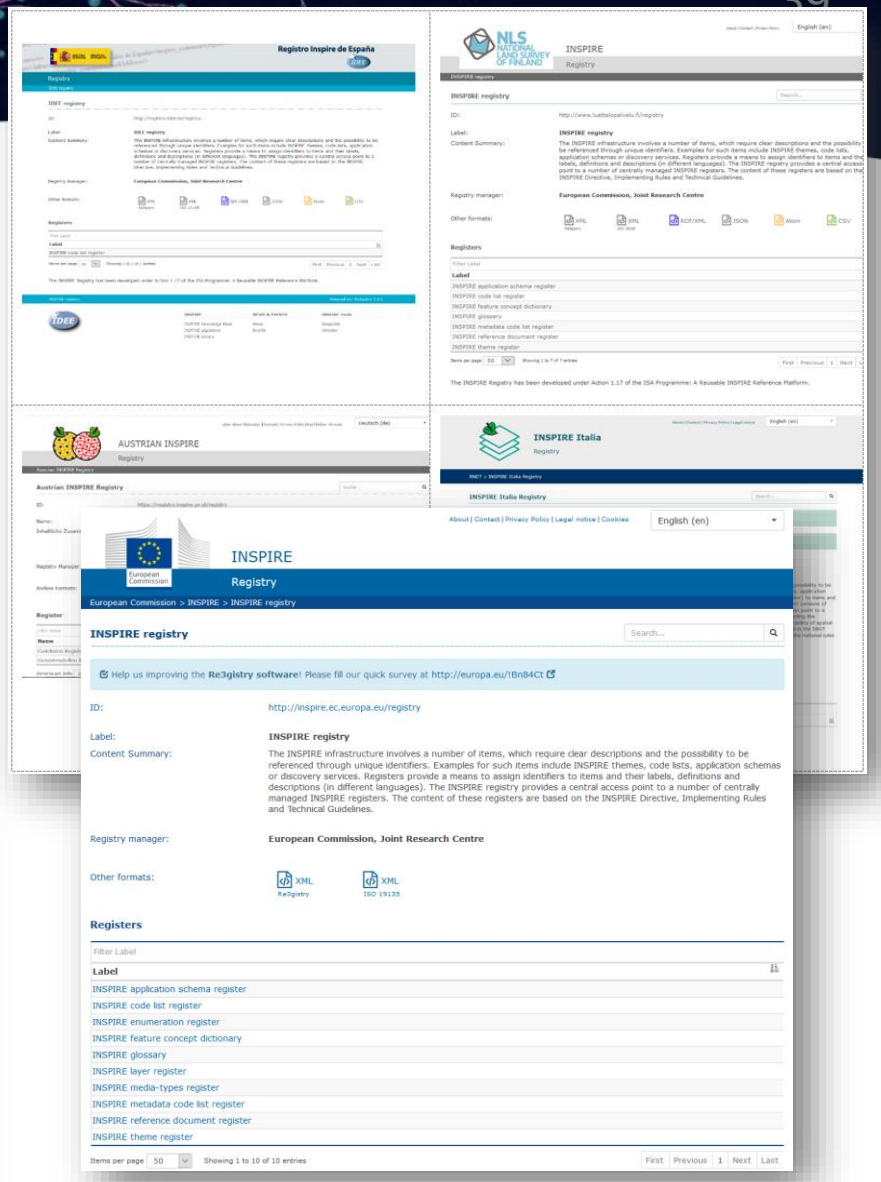
Software components



INSPIRE Geoportal



INSPIRE Reference Validator



INSPIRE Registry

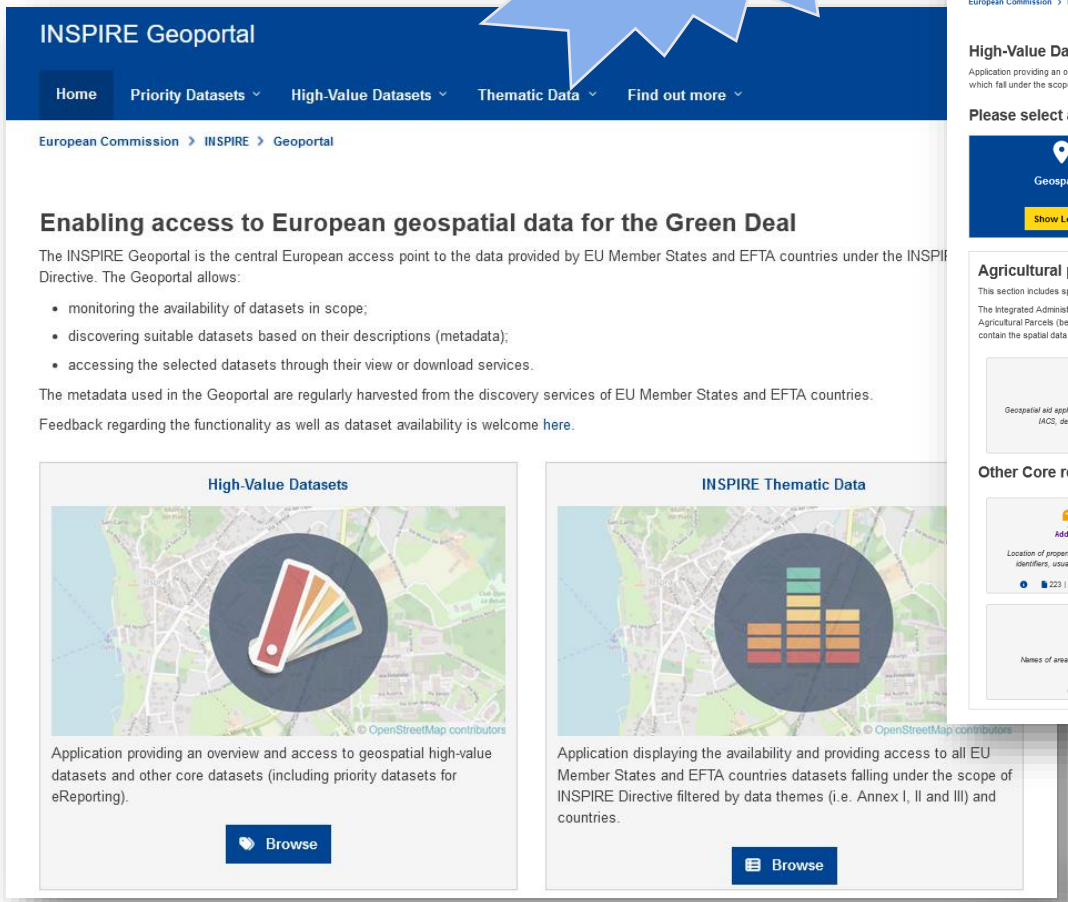
Data discoverability

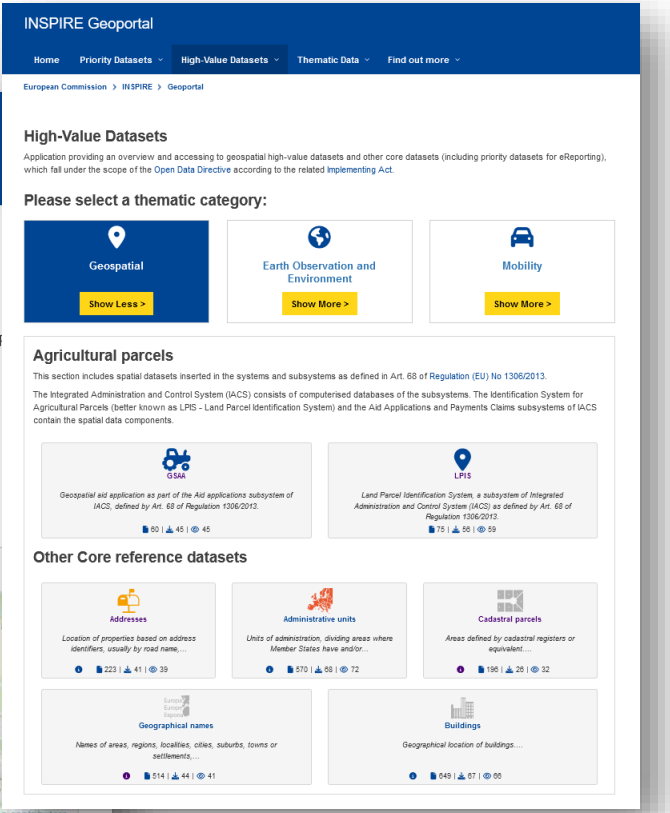
Revamped 2023.11.24

- Central EU access point to data provided by MS and EFTA countries under the INSPIRE.
- **DEMO**: Metadata management & exposure.

- **Software:**

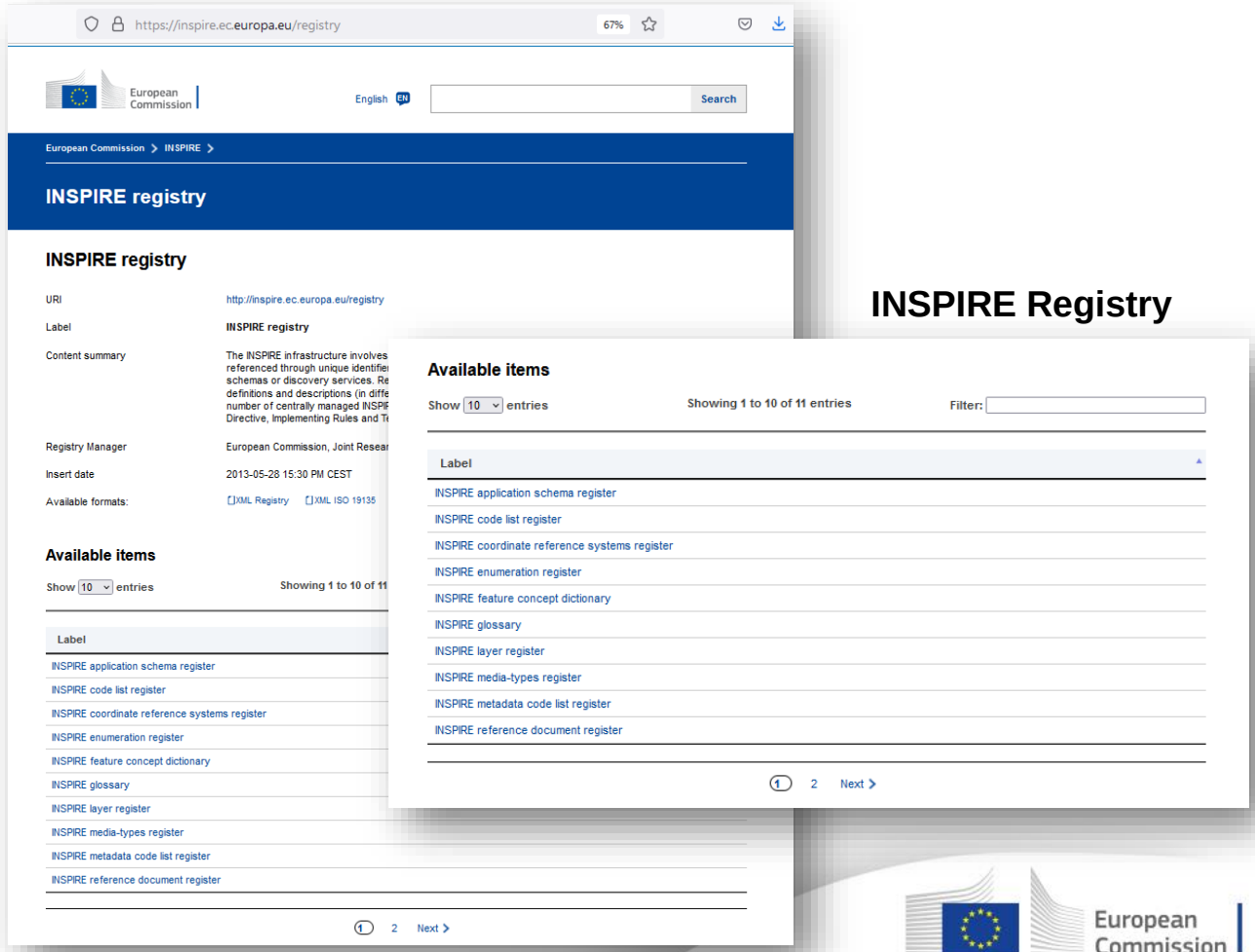
GeoNetwork opensource





Register management and reuse

- **INSPIRE Registry**
  - Central registry to manage and share common INSPIRE reference codes.
  - 11 registers available.
  - **DEMO**: Codelists & Controlled vocabularies.
- **Software: open source Re3gistry software**

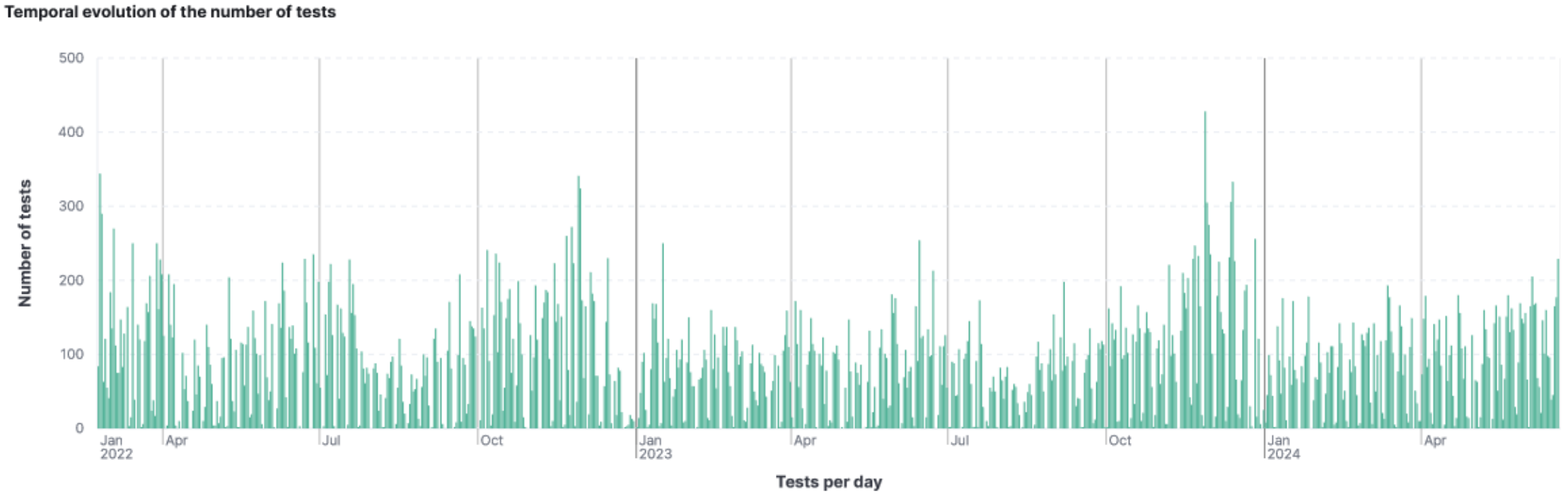


INSPIRE Registry

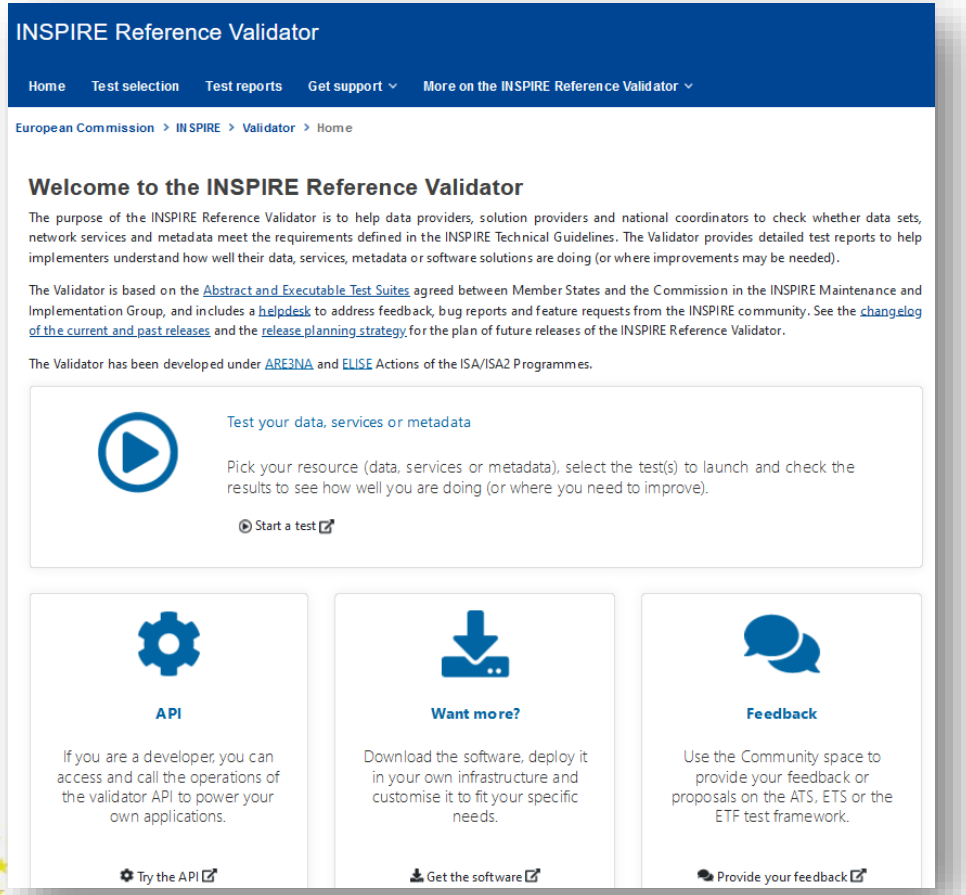
Reference validator

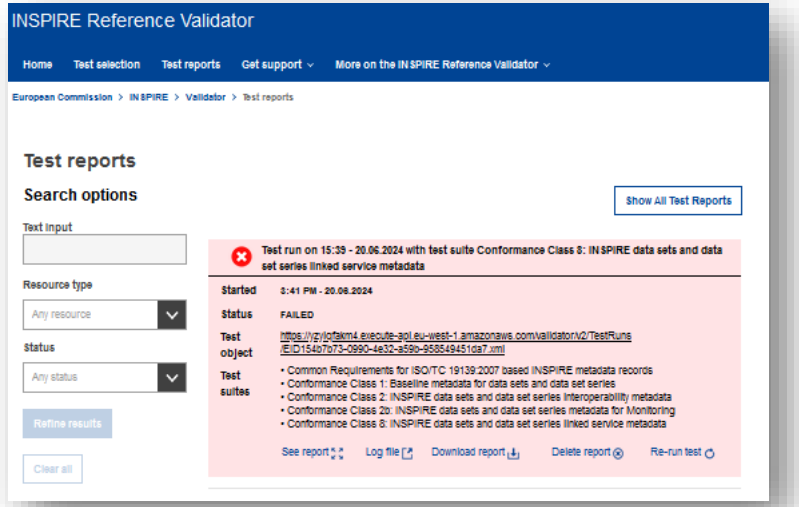
Software: open source ETF software

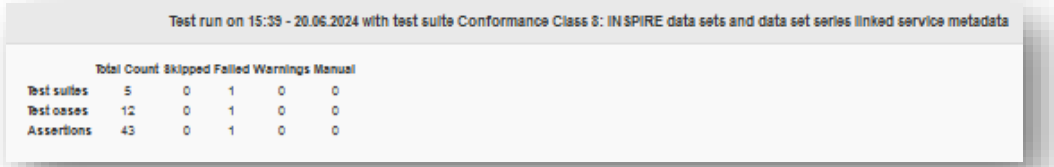
- **Validation** of metadata, services and data
- Provides **certainty** on the status of implementations
- **Central EU instance** and **Docker instances** deployed on the national level



Reference validator DEMO







Benefits for data spaces (discussion)

- INSPIRE tools, governance procedures and assets are reusable in different data spaces, e.g.
  - smart cities (e.g. MIM-7), agriculture (e.g. IACS data sharing), Green Deal, mobility, health
- Geospatial data is horizontal and can feed different data spaces and solve part of interoperability issues.
- Do we need a location strategy for data spaces?

Thank you!


Brussels 27 June
SEMIC 2024 conference
interoperable europe
from Vision to Reality
be EU belgium24.eu
European Commission


Brussels 27 June
SEMIC 2024 conference
interoperable europe
from Vision to Reality
be EU belgium24.eu
European Commission

Smart middleware platform (Simpl): Enabling Data Spaces

Leire Orue-Echevarria, Dominique Roelants
European Commission
Jean Michel Guilmot, Sovereign-X

- **What is Simpl?**
- **What are we doing now in Simpl?**
- **Demo / Proof of Concept**

Overview of European Cloud Policies

By 2030: > 10 000 Edge nodes & 75% of cloud uptake by EU

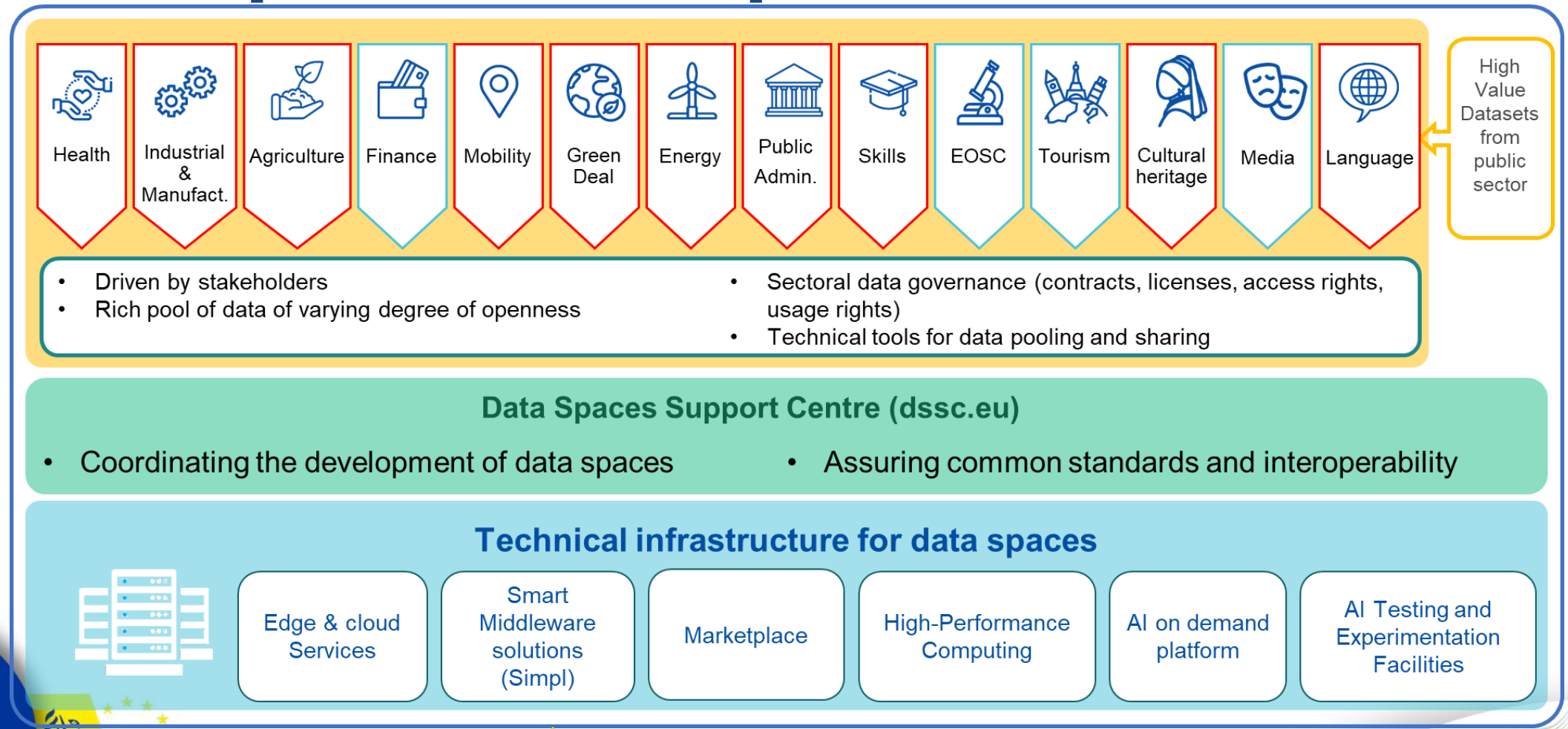
- European Data Spaces
- Data Governance Act
- Data Act

EU Data Strategy

Digital Decade Policy

- IPCEI – Cloud Infrastructure & Services
- European Alliance on Industrial Data, Edge and Cloud
- Cloud Rulebook

European Data Spaces: The vision



SEMIC conference

What is Simpl?

Simpl is the **open-source secure middleware** that will support data access and interoperability among European data spaces and other cloud-to-edge federations

be EU belgium24.eu

Simpl is made of three products

Simpl - Open

The Open - Source middleware itself

Simpl - Labs

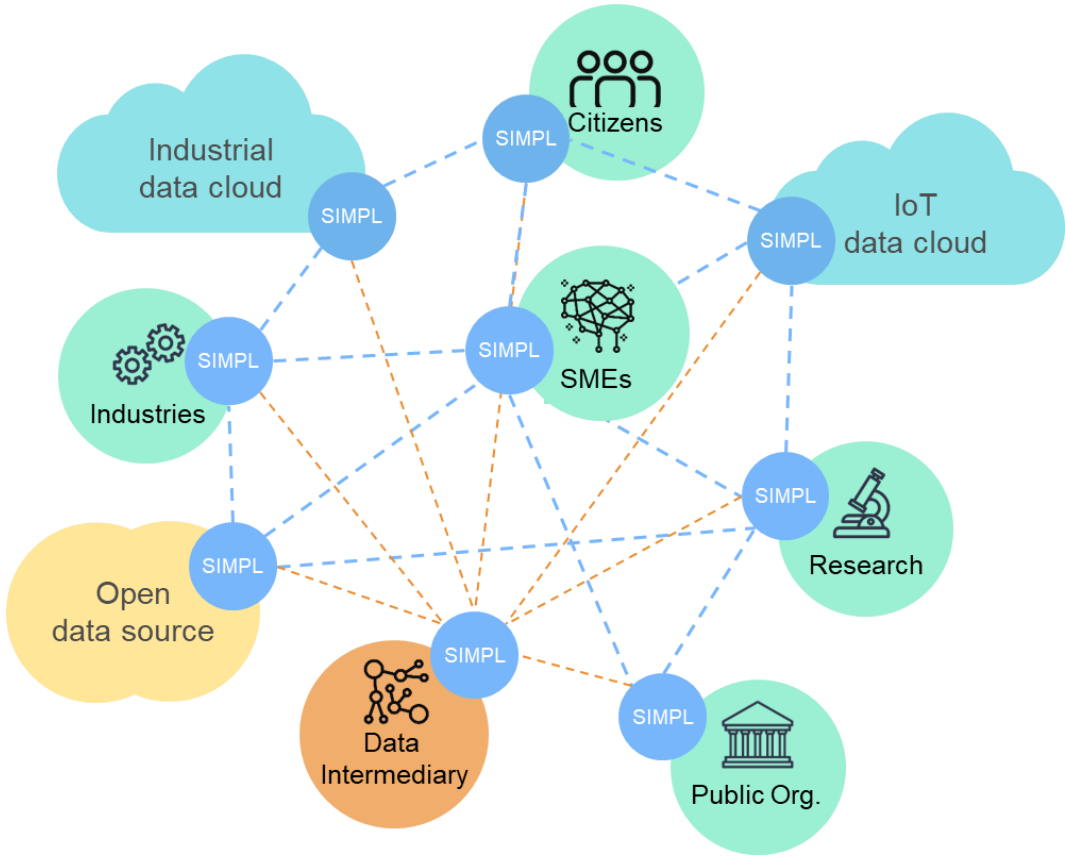
Playground environment for Simpl Open
+
Interoperability test for existing data spaces

Simpl - Live

Instances of Simpl-Open for sectoral data spaces

Simpl is the common software behind the data spaces

1. **Federated Data Ecosystem with shared policies and rules**
2. **Secure, transparent, trusted, easy and unified access to data**
3. **Data holders remain in control of their data**



Why is Simpl important?

- Support the emergence of a European cloud and data federation
- Embedded business intelligence services for multiple uses
- High degree of interoperability and portability
- Allow EU to develop its own cloud-to-edge supply chain. Increase EU open strategic autonomy

How can Simpl help the public sector?

1. Provide the means to deploy your own data space and facilitate the federation with other data spaces in an easier manner
2. Interoperable, compliant with existing vocabularies, extensible, customizable and adaptable
3. Secure, trust Access and control of your data

From an initial baseline to an evolving set of Simpl-Open Requirements

Preparatory Study

1

Data spaces

2

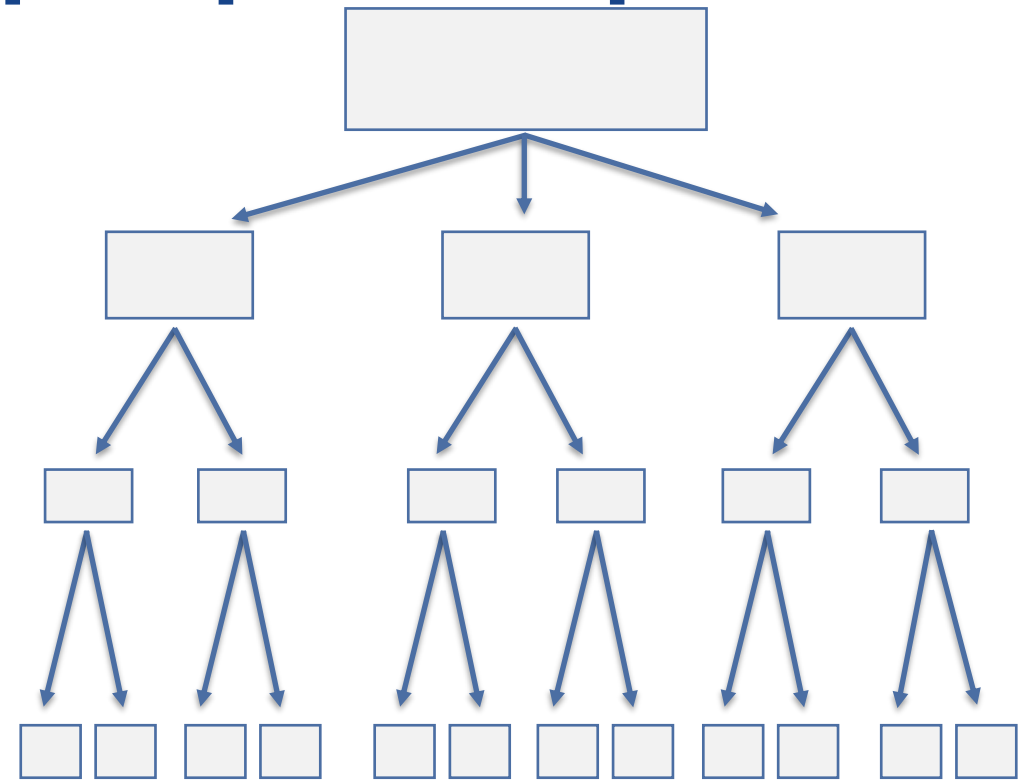
First Refinement

3

Evolving requirements

...

Simpl-Open Requirements Organisation



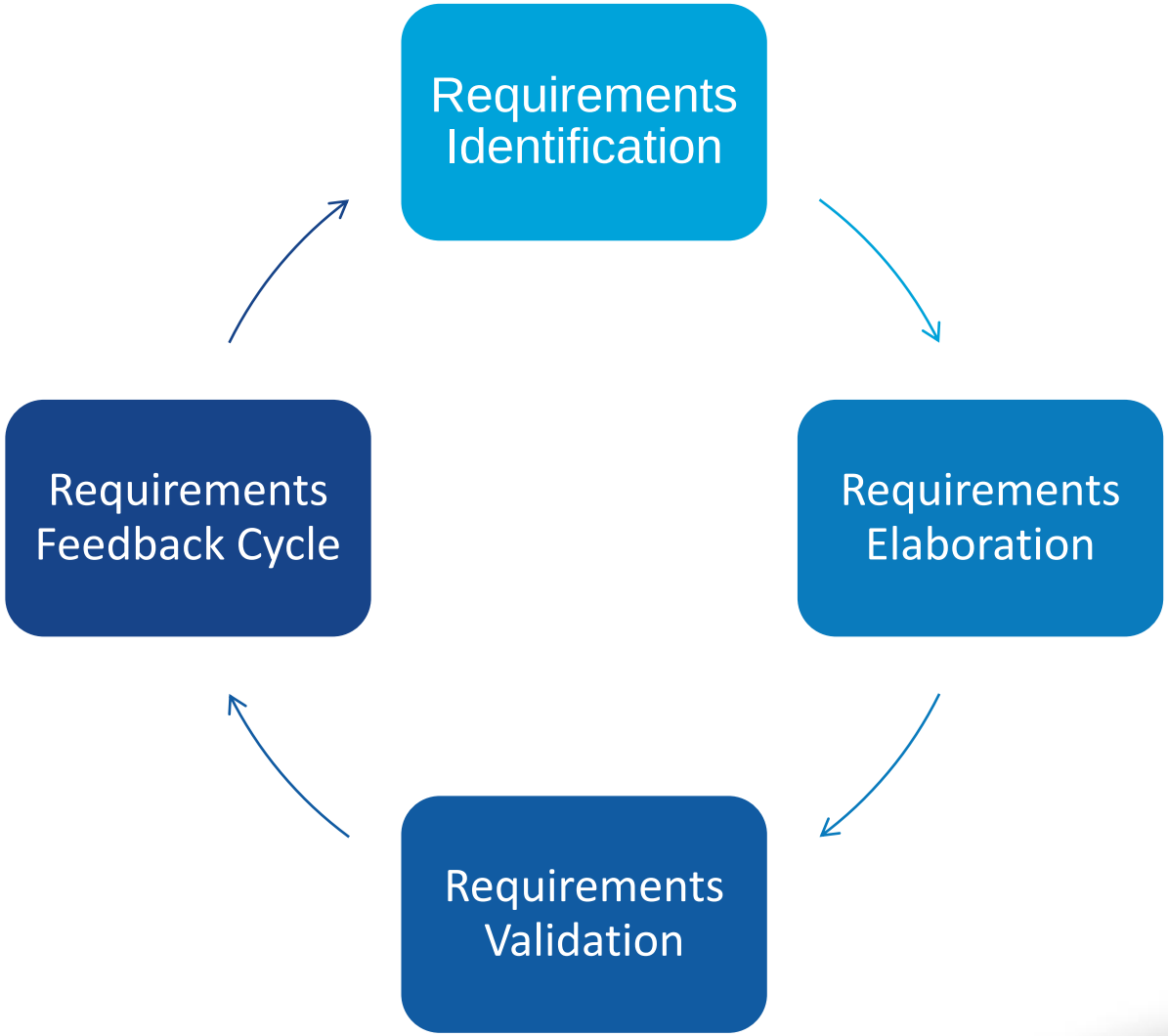
Business Processes (L0)

High-level requirements (L1)

Detailed requirements (L2)

User stories

Agile Requirements Elicitation



Main actors

Dataspace Governance Authority	• Oversees the dataspace governance and operations • A.o. defines the rules and policies for participating to the dataspace
Provider	• Publishes and shares resources and services • Data, Applications, Infrastructure
Consumer	• Searches for, accesses, and uses resources

Supported Business Processes

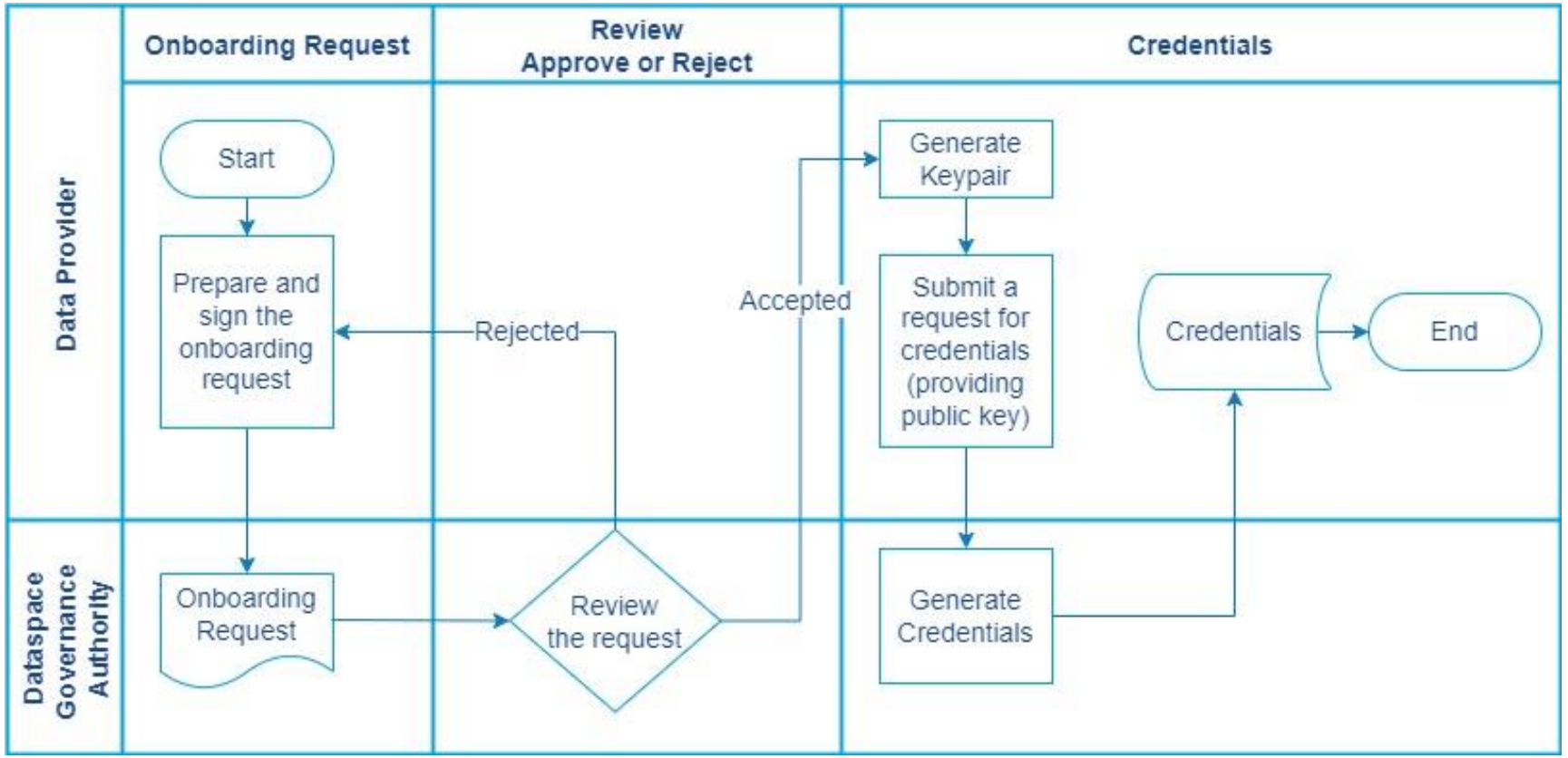
- Setup of Dataspace: Role of Governance Authority
- Setup of Catalogues and Vocabulary
- Onboarding of a Dataspace Participant: Provider of Data, Application, Infrastructure
- Onboarding of a Dataspace Participant: Consumer
- Publish on the Catalogue: Data, Applications, Infrastructure
- Consumers search the Catalogue
- Consumers establish a Contract with a Provider
- Consumers use a Resource from a Provider
- Consumers decommission a Resource
- Management and Operations - Business
- IT Administration

Demo Introduction

Onboarding of a Provider - summary

- Before a provider of data, applications, or infrastructure can participate in a dataspace, they must be formally onboarded
- The Dataspace Governance Authority will have set the rules that the providers must fulfil to be onboarded
  - E.g. which documents to provide, security to set up

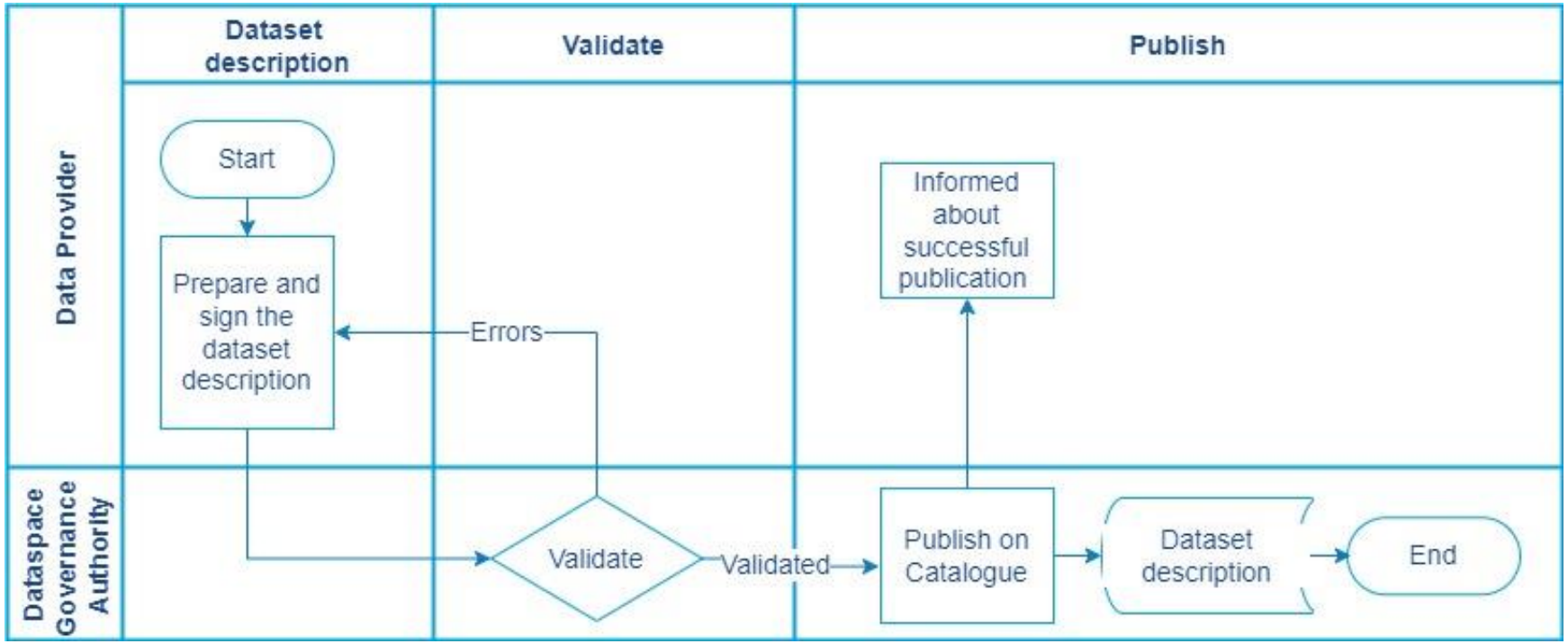
Onboarding of a Provider – main steps



Publishing on the Catalogue - summary

- Providers can publish resources on the dataspace catalogue
- Resources can be:
  - Datasets
  - Applications
  - Infrastructure
- Resources published on the catalogue are discoverable by consumers

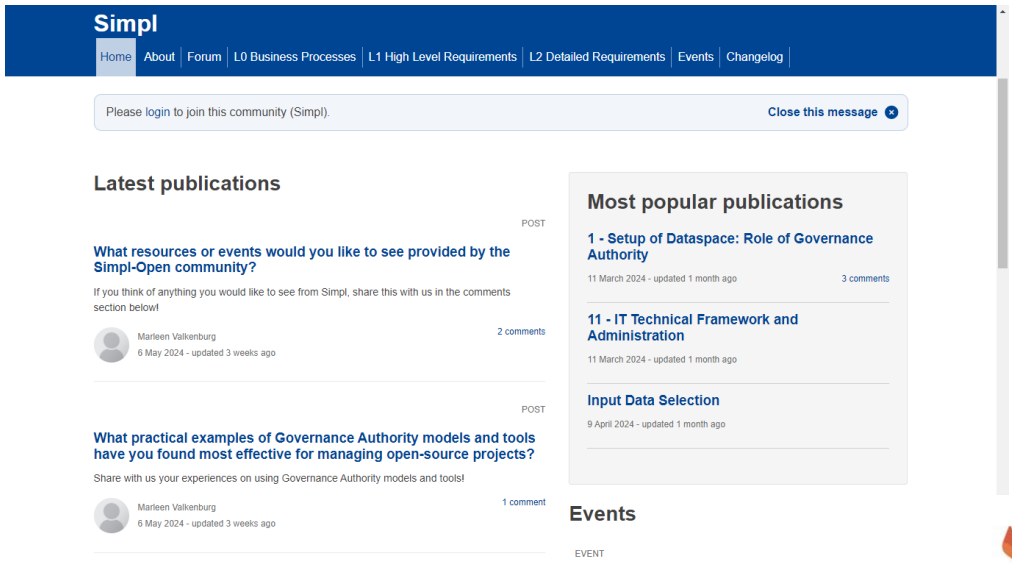
Publishing on the Catalogue – main steps

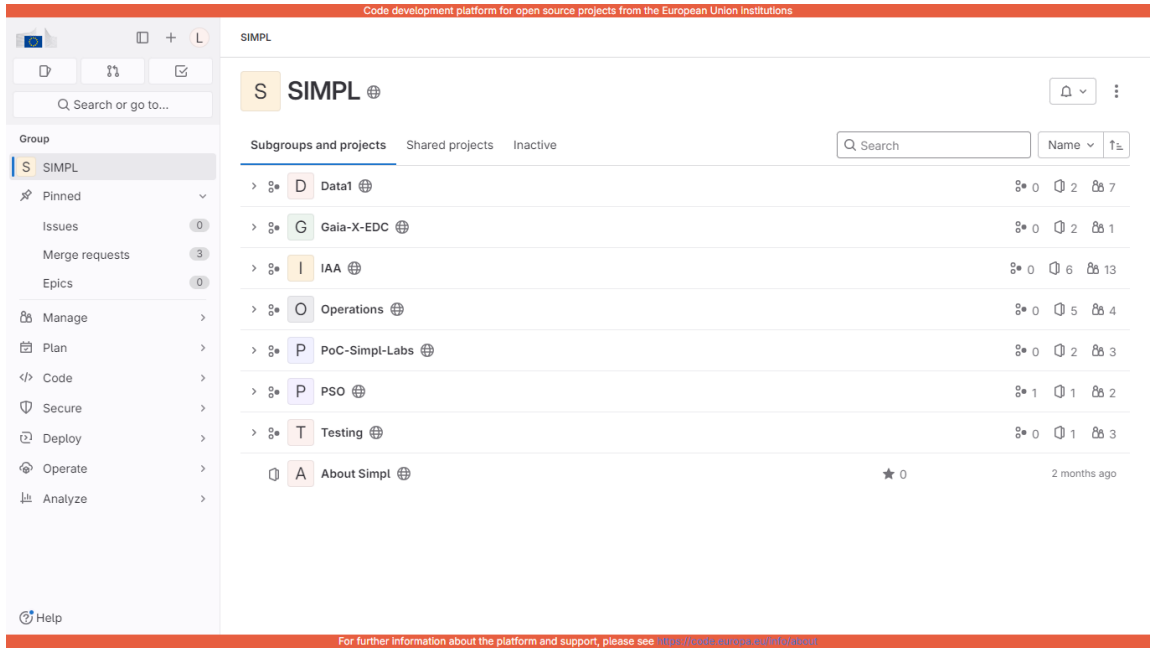


Simpl is Open Source

European Commission

Futurium

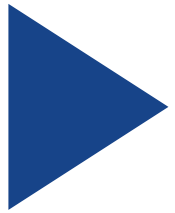




GitLab code.europa.eu

https://futurium.ec.europa.eu/en/simpl

Demo

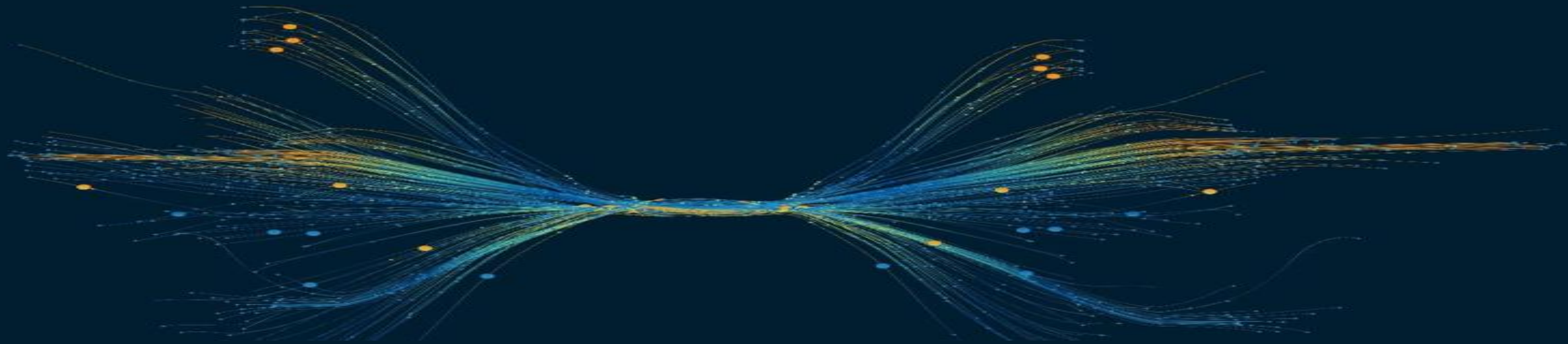


Thank you!


Brussels 27 June
SEMIC 2024 conference
interoperable europe
from Vision to Reality
be EU belgium24.eu
European Commission

HealthDCAT-AP : from vision to reality

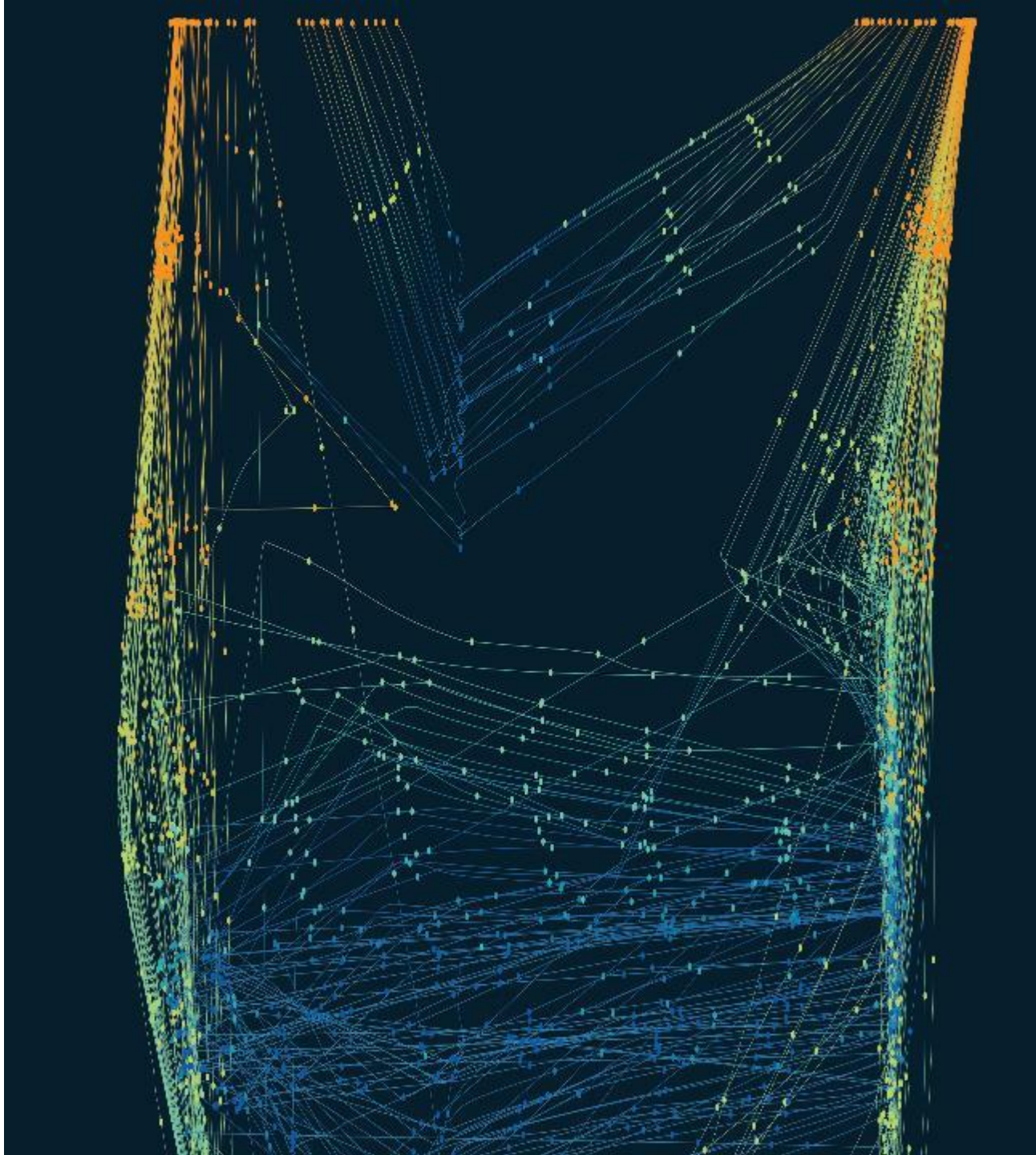
European Health Data Space



Contents

1

Policy context



The European health data space (EHDS)

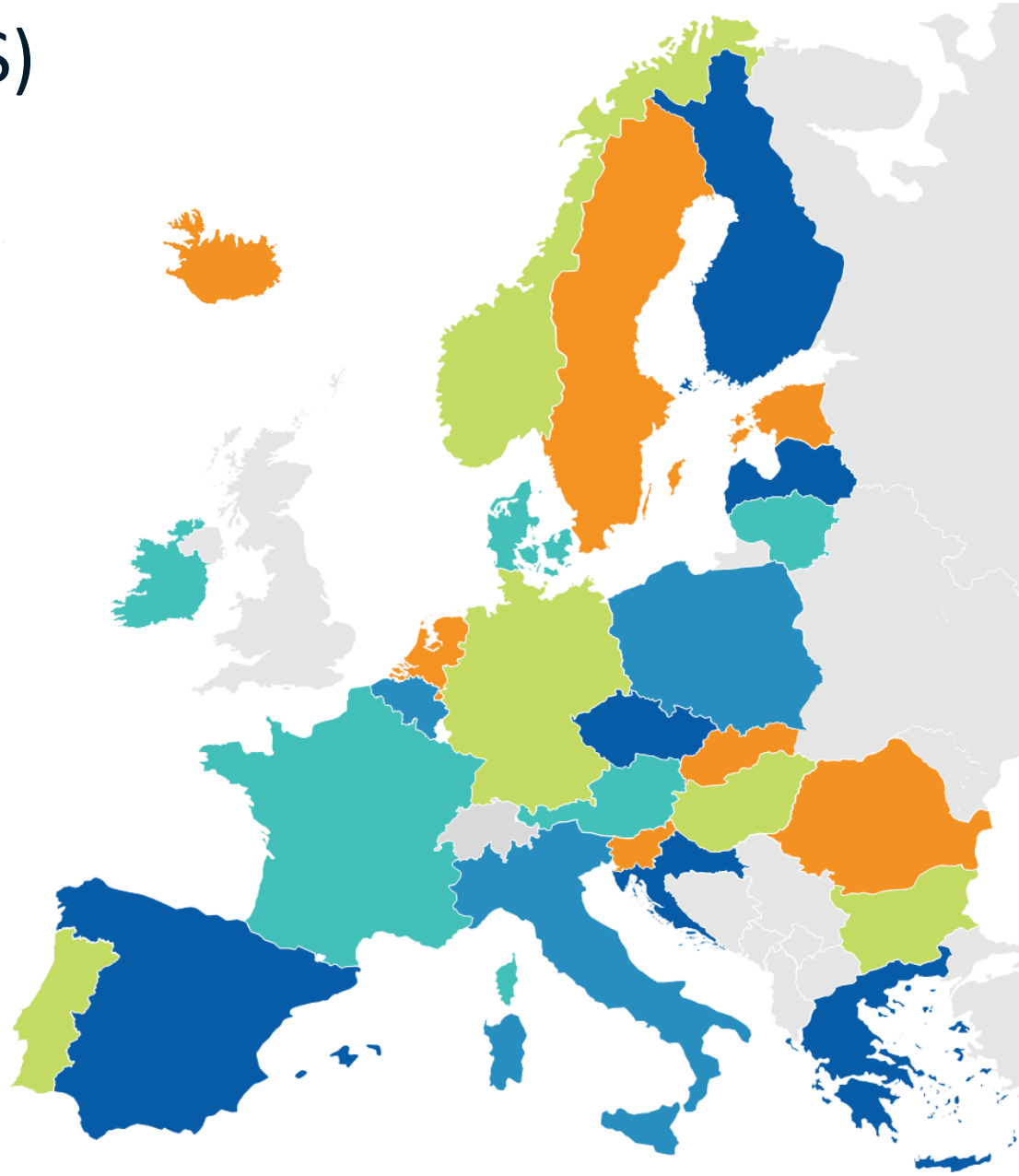
What ? To create a unified framework to facilitate the exchange of electronic health data across Europe for secondary use.

Why ?

- Enable large-scale health research and innovation.
- Support evidence-based policy making and public health interventions.
- Cross-Border collaboration, promote interoperability and data sharing across EU member states.

How ?

- Data governance: clear policies and regulations for data access and use.
- Technical infrastructure: robust platforms for secure data exchange and analysis.
- Stakeholder engagement: involvement of healthcare providers, researchers, policymakers, and patients.

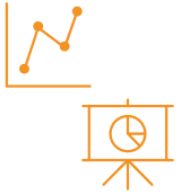


The need for a health-specific metadata standard

Data discoverability

Enables efficient search and retrieval of datasets.

Data usability



Ensures data is well-described and actionable.

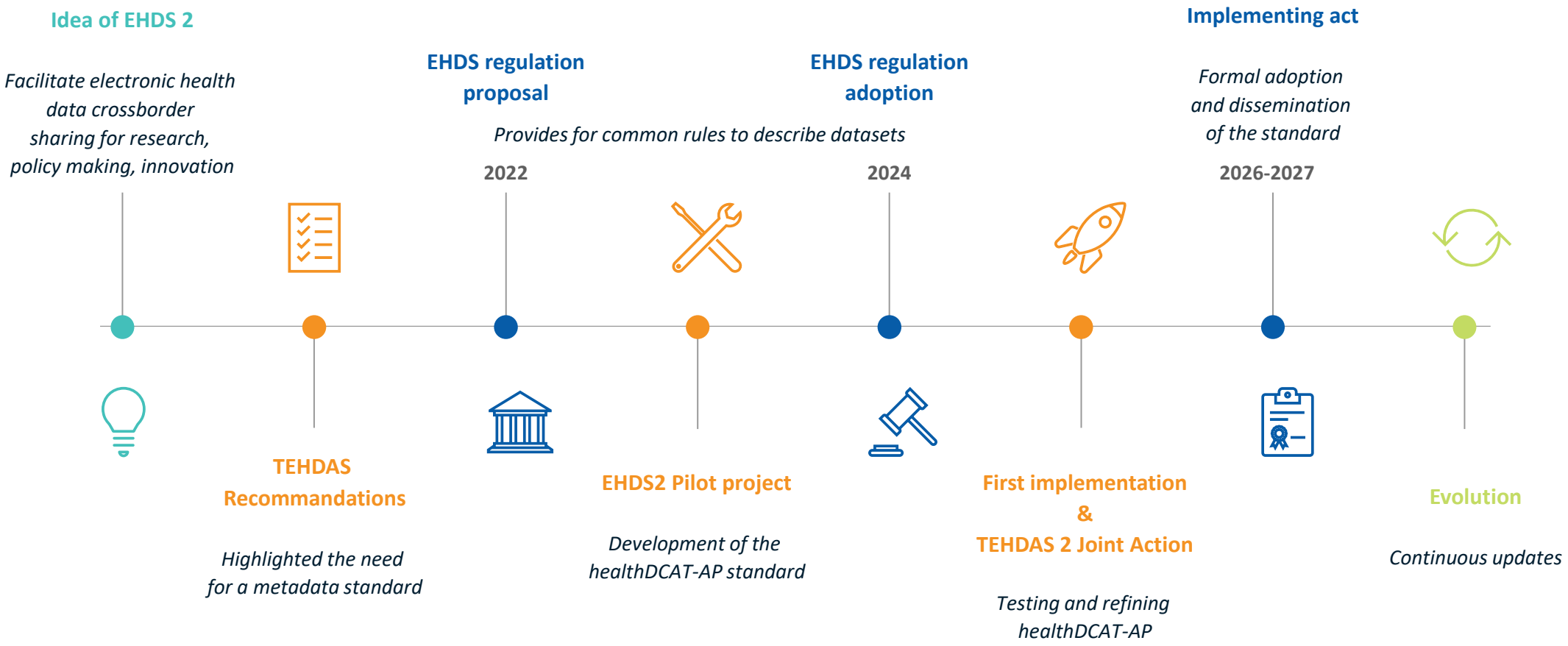
Interoperability

Facilitates data integration across systems.

Support for Advanced Technologies

Enables advanced analytics and AI processing.

The path towards healthDCAT-AP



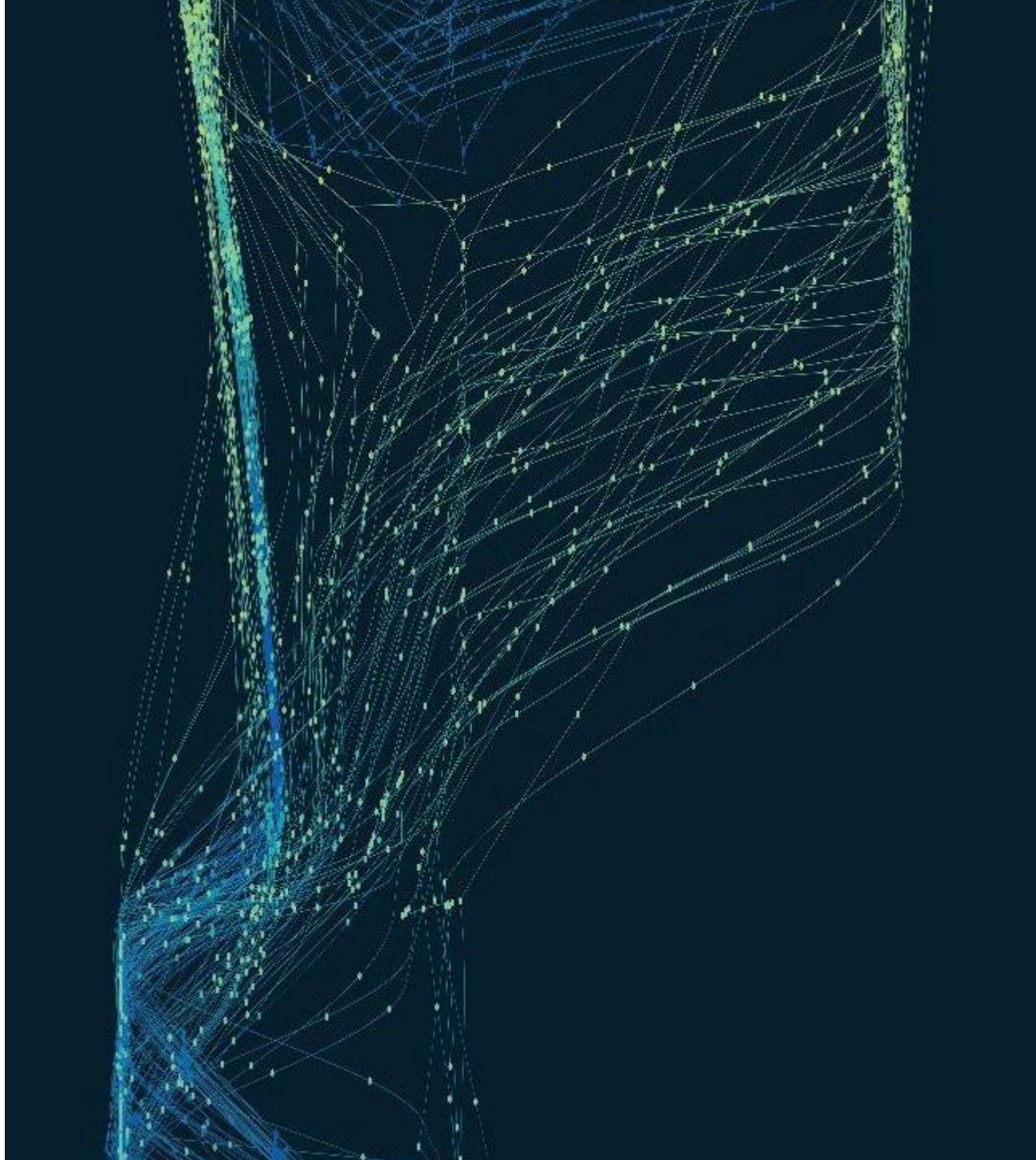
Choosing and extending DCAT-AP

- **Proven framework**
  - Established Standard: widely adopted for data cataloging, ensuring reliability and robustness.
  - Flexible / easily adapted to meet specific needs.

- **Interoperability and compatibility**
  - Ensures compatibility with existing data catalogs and systems.
  - Facilitates data sharing and promotes data exchange within and across sectors.

- **HealthDCAT-AP**
  - Adds health-specific properties and vocabularies to DCAT.
  - Improved Metadata Quality: Provides detailed and relevant metadata for health datasets.

2 Developing Health-DCAT AP



Roadmap

Oct. 2022 to … 2026

healthDCAT-AP roadmap

from inception to implementation

DEFINITION OF THE REQUIREMENTS

- Landscape analysis of existing health metadata catalogues and health metadata models
- Analysis of existing DCAT Application Profiles: geoDCAT, statDCAT, JRC-DCAT-AP
- Release of a DCAT-AP Sandbox catalogue (AS-IS)

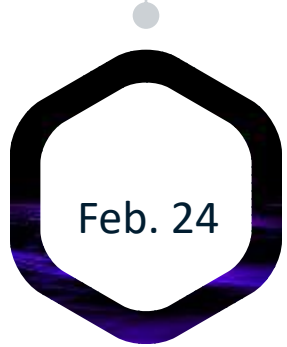
TRAINING AND ONBOARDING OF A TECHNICAL WORKING GROUP OF HEALTH EXPERTS

- Workshop on DCAT-AP rationales provided by DIGIT/SEMIC
- Rules for extending DCAT-AP

DEFINITION OF THE HEALTHDCAT-AP DOMAIN MODEL

Iterative and interactive sessions and publication of surveys to collect feedback and content on a continuous basis from the Technical Working Group

Further validation and implementation of healthDCAT-AP concepts



DRAFT healthDCAT-AP

- Publication of the draft healthDCAT-AP according to the ReSpec template https://healthdcat-ap.github.io

Implementation in HealthData@EU EU Dataset Catalogue

- First implementation of healthDCAT-AP

- healthDCAT-AP Sandbox catalogue TO-BE
- Workshops on creating healthDCAT-AP records with support of tutors and a healthDCAT-AP editor.
- Mapping to healthDCAT-AP exercises and development of heaDCAT-AP compliant endpoints

Definition of the requirements

Jan. to Jun. 2023

Definition of the requirements:

Identification of the metadata elements by functional groups:

- Data Discovery
- Data Access
- Data Provenance
- Data Ownership
- Temporal Coverage
- Spatial Coverage
- Population Coverage
- Data analytics
- Data quality
- Variables
- Data categorisation

Review of use cases

AS « persona »
I WANT …
SO THAT …

Requirements derived from the FAIR data principles

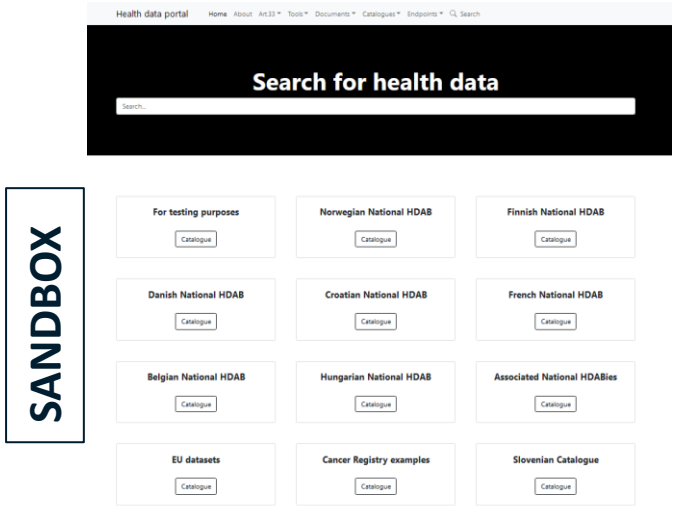
AS a metadata catalogue
I WANT TO …
SO THAT …

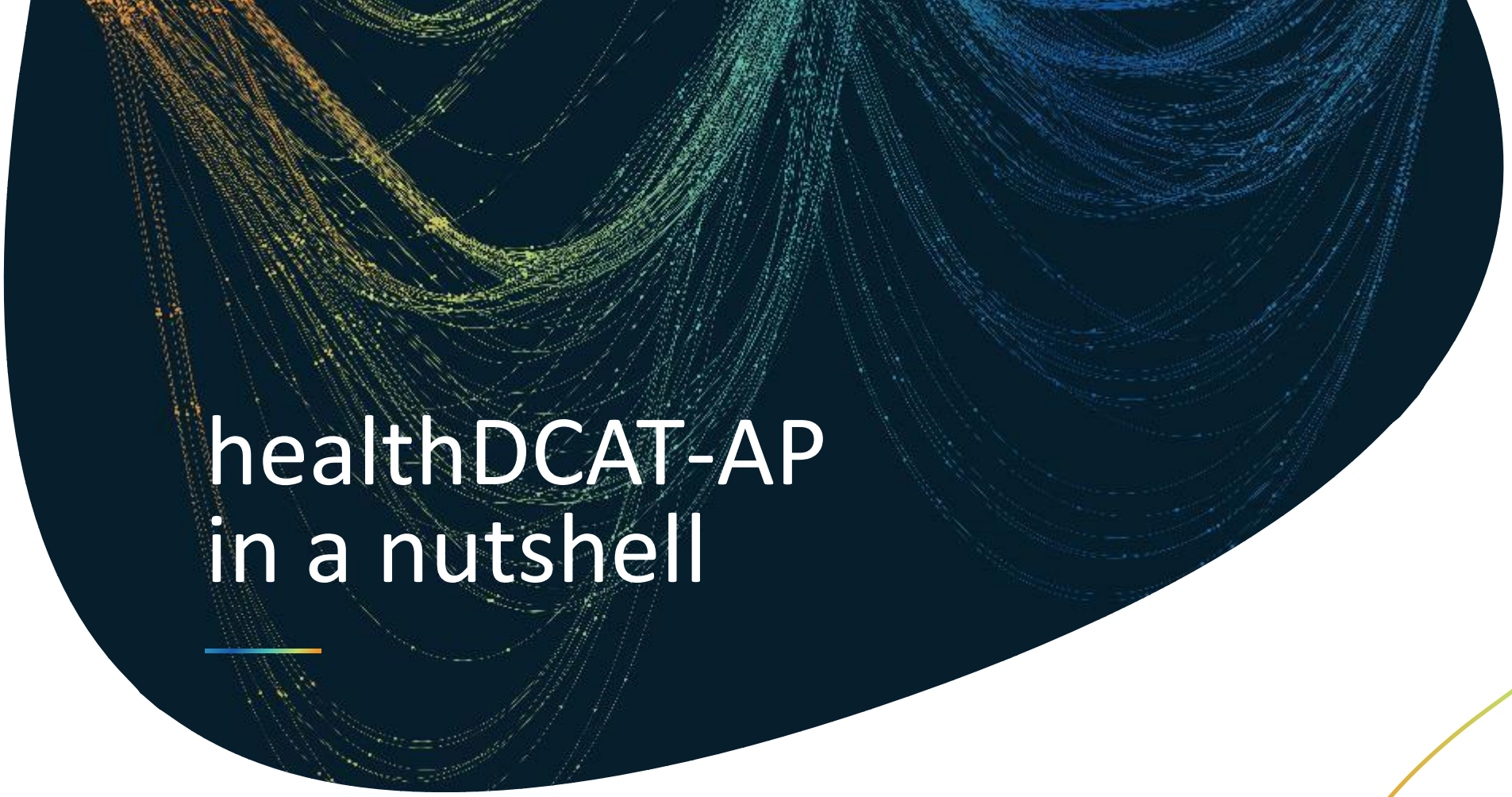
Requirements derived from EU Policies

- European Health Data Space
- Digital Governance Act (NSIP requirements)
- HVD

Requirements derived from technical considerations

- Keyword, Faceted, Full-Text Search
- Semantic Search
- Natural Language Processing (NLP) and GenAI
- Geospatial Search
- Metadata Management





healthDCAT-AP in a nutshell

Feb. 2024
Draft healthDCAT-AP
healthdcat-ap.github.io

Draft healthDCAT-AP in a nutshell

NEW PROPERTIES

alternative
analytics
code values
coding system
health category
health data access body
health theme
legal basis
max typical age
min typical age
number of records
number of unique individuals
personal data
population coverage
publisher note, publisher type
purpose
quality annotation
retention period

NEW CONTROLLED VOCABULORIES

Based on Art.33

Contact Point Registry of the Health Data Access Bodies, the Union data access service and authorised participants of the healthData@EU infrastructure

Wikidata as an ontological health database

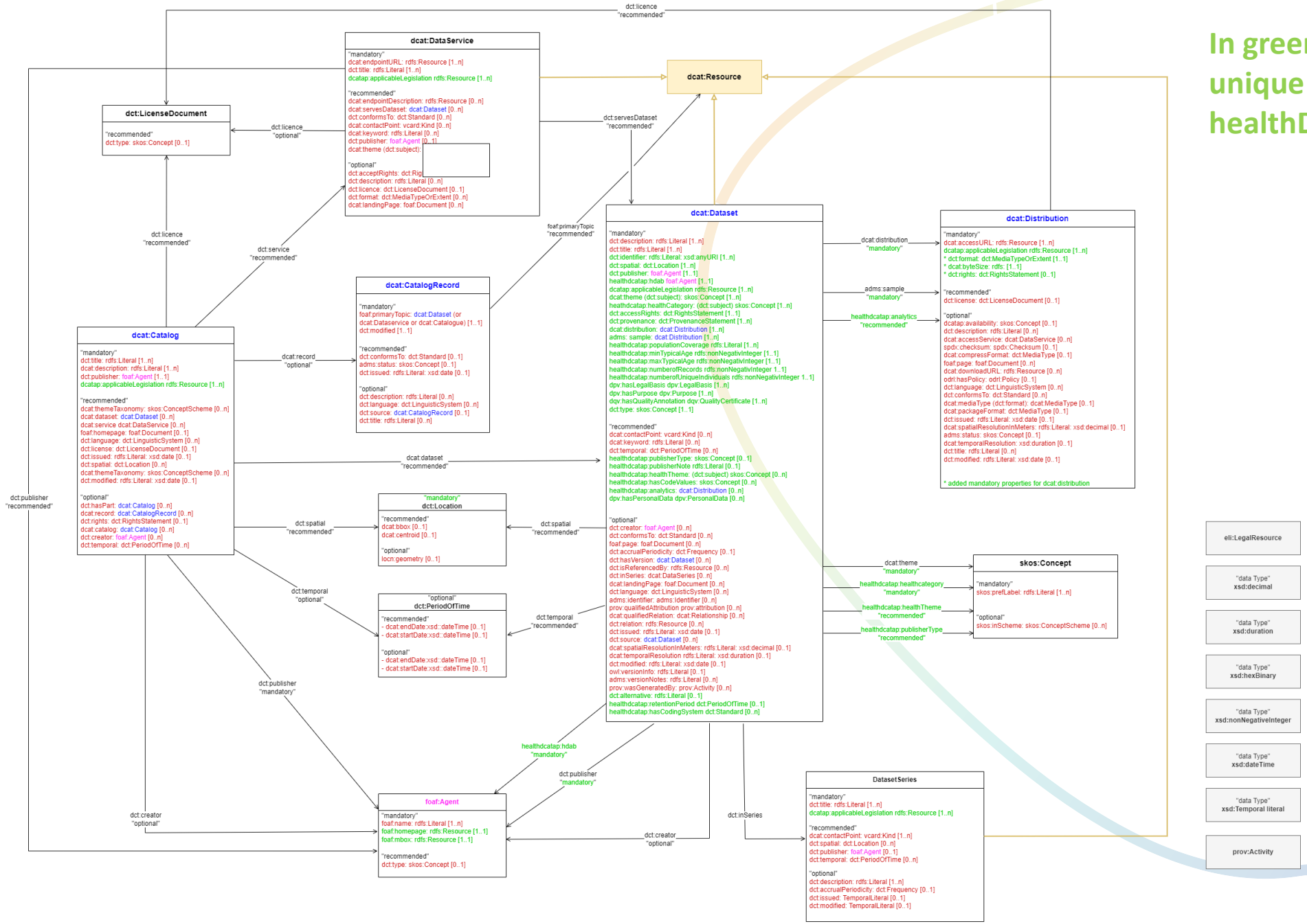
Health Publisher type

CARDINALITIES & USAGE NOTES

General increase of the cardinalities to create rich metadata

Technical properties as mandatory properties to operationalise the healthData@EU infrastructure

Recommended Sample distributions:
- use of synthetic or anonymized subsets
- use of CSVW terms for RDF-izing variable descriptions



In green unique to healthDCAT-AP

Key examples

dct:identifier

Property	Range	Card	Definition
identifier	Literal	1..*	The main identifier for the Dataset, e.g. the URI or other unique identifier in the context of the Catalogue. RDF: dct:identifier

Usage: The use of permanent dereferenceable URIs is mandatory in the healthDCAT Application Profile for the dataset identifier property.

Ex: *dcterms:identifier "http://.../dataset/6d034a3c-8562-4205-b0a9-da5cdfecd427"^^xsd:anyURI;*

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the European Health Data Space
Art. 41 Duties of health data holders
2. The health data holder shall, at a minimum, on an annual basis check that its dataset description in the national catalogue is accurate and up to date.

dct:theme

Property	Range	Card	Definition
theme	Concept	1..*	A category of the Dataset. RDF: dcat:theme

Usage: A Dataset may be associated with multiple themes.

The authority table for Data Themes, maintained by the Publications Office of the European Union is the mandatory controlled vocabulary for dcat:theme. It must have at least the value NAL:data-theme "HEAL" to annotate health datasets.

Ex: *dcat:theme <http://publications.europa.eu/resource/authority/data-theme/HEAL>;*

healthdcatap:healthCategory

Property	Range	Card	Definition
health category	Concept	1..*	The health category to which this dataset belongs as described in the Commission Regulation on the European Health Data Space laying down a list of categories of electronic data for secondary use, Art.33. RDF: healthdcatap:healthCategory

Usage: A mandatory controlled vocabulary denoting health data within the scope of the Commission Regulation is provided.

Article 33
Minimum categories of electronic data for secondary use

1. **Health data holders shall make the following categories of electronic data available for secondary use in accordance with the provisions of this Chapter:**

(a) electronic health data from EHRs;
(b) data on factors impacting on health, including socio-economic, environmental and behavioural determinants of health;
(ba) aggregated data on healthcare needs, resources allocated to healthcare, the provision of and access to healthcare, healthcare expenditure and financing;
(c) pathogen data, impacting on human health;
(d) healthcare-related administrative data, including dispensation, claims and reimbursement data;
…
17 categories are defined

healthdcatap:healthTheme

Property	Range	Card	Definition
health theme	Concept	0..*	A category of the Dataset or tag describing the Dataset. RDF: healthdcatap:healthTheme

Usage: healthDCAT-AP relies on *Wikidata* as large-scale, human-readable, machine-readable, multilingual, multidisciplinary, centralised, editable, structured, and linked knowledge-base.

Ex: *healthdcatap:healthTheme <https://www.wikidata.org/wiki/Q63391344>, <https://www.wikidata.org/wiki/Q18975322>, <https://www.wikidata.org/wiki/Q4227886>, …*

Corresponding tags: ['antimicrobial resistance'] ['Acinetobacter infectious disease'] ['Escherichia coli infectious disease']…

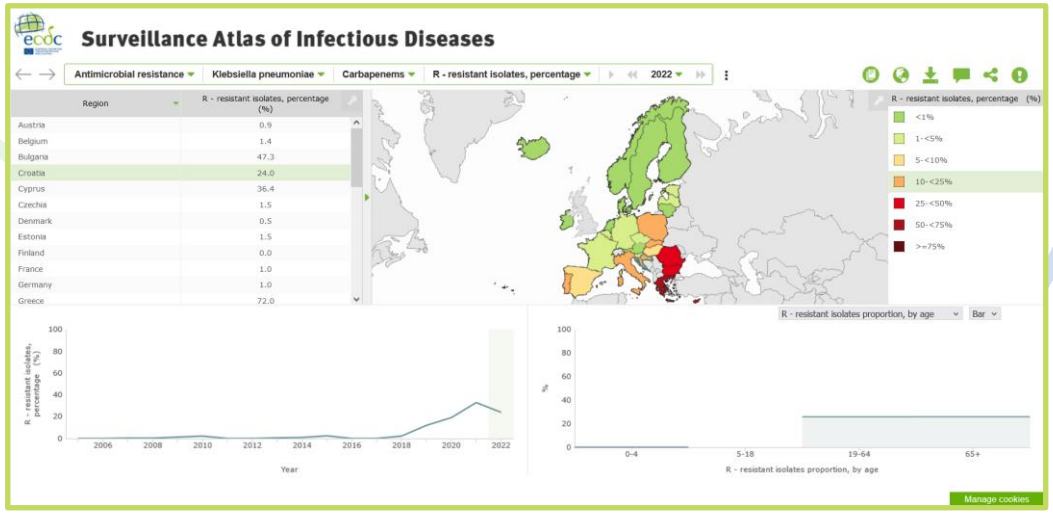
Comment: Wikidata URIs *MUST* be used for the following properties coding system, code values, conform to, health theme.

healthdcatap:analytics

Property	Range	Card	Definition
analytics	Distribution	0..*	An analytics distribution of the dataset. RDF example: healthdcatap:analytics

Usage: Publishers are encouraged to provide URLs pointing to API endpoints or document repositories where users can access or request associated resources such as technical reports of the dataset, quality measurements, usability indicators,... or analytics services.

Example of analytics dashboards:



EDCD Atlas

Statistics Austria

New full text properties are included alongside dct:description

Complementing the "description" property, healthDCAT-AP introduces the following new "full text" properties:

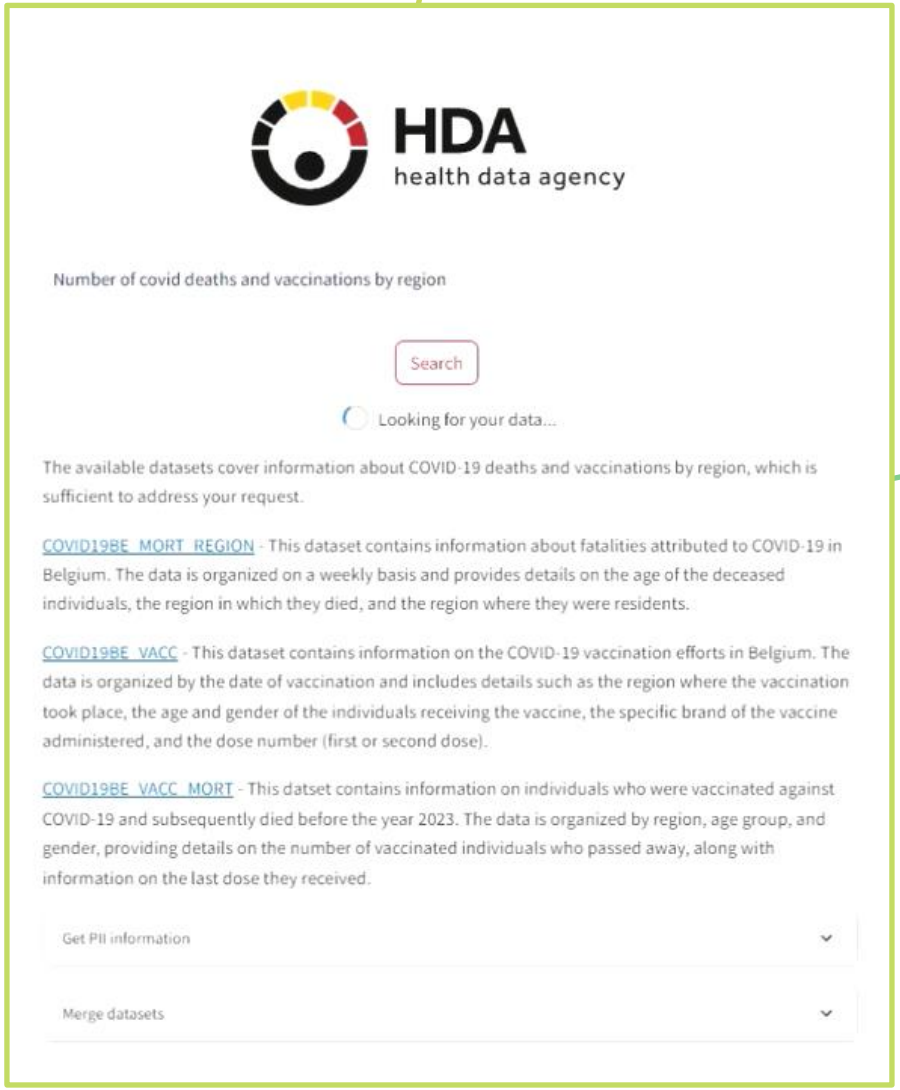
- "Purpose" (dpv:hasPurpose)

- "Population Coverage" (healthDCAT-AP)

- "Publisher Note" (healthDCAT-AP)

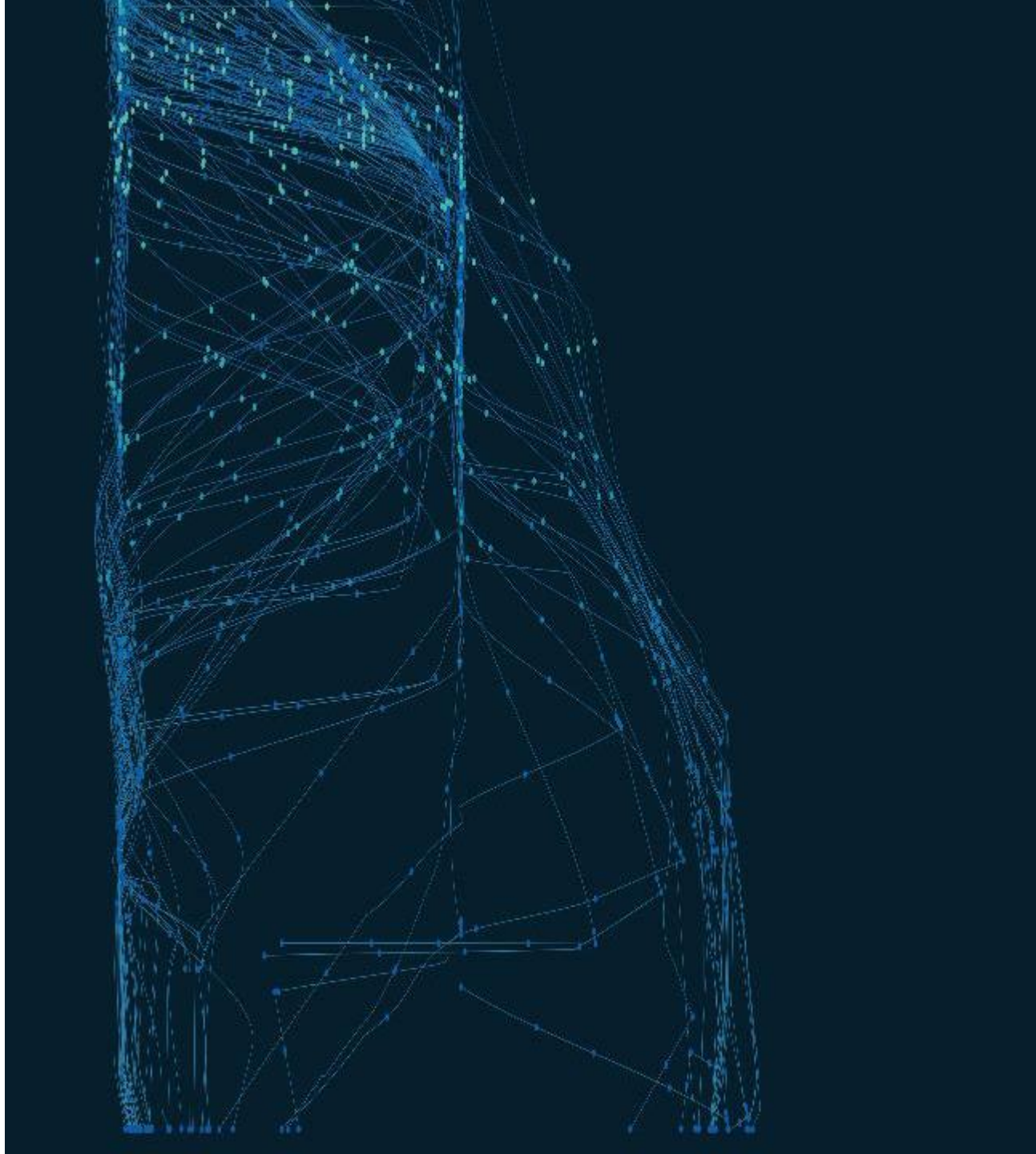
Additionally, it makes "Provenance" (dct:provenance) a mandatory property to better support genAI.

Screenshot of POC genAI demo by Belgian Health data agency

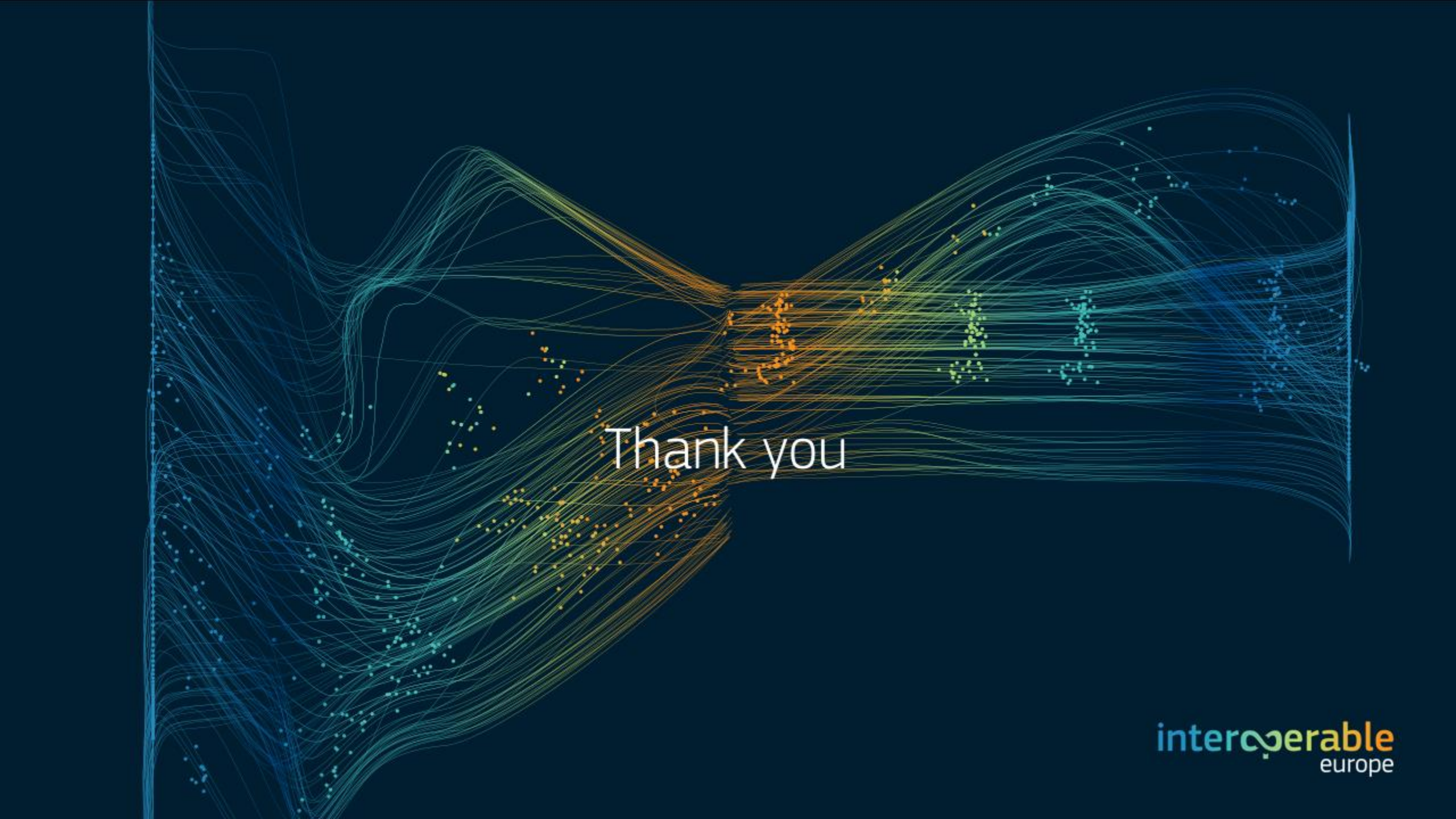


3

Implementing



Demo



interoperable europe

Natalia ZYLINSKA-PUTA

Team Leader
IT Portfolio Manager European Health Data Space
European Commission
Directorate-General for Health and Food Safety
R.4 Information Systems

Pascal DERYCKE, PhD

Researcher, data & semantic engineer
EU health information system unit
Epidemiology and public health

sciensano

www.sciensano.be

Mélodie BERNAUX

Policy Officer

European Commission
Directorate-General for Health and Food Safety
C.1 Digital Health

Stay in touch ec.europa.eu/InteroperableEurope

@ IoP_EU

Interoperable Europe

IoP_Europe_COMM@ec.europa.eu

Language Data Space: Tackling semantic interoperability at the metadata level

Penny Labropoulou, ILSP/Athena RC

Why “Language Data Space”?

- Large Language Models are the most disruptive breakthrough in AI in recent history (BERT, GPT-3, ChatGPT, GPT-4 etc.)
- LLMs are trained on vast amounts of training data (language data)
- LLMs use dozens, some even hundreds of terabytes (trillions of tokens) of language and also image, video, audio etc. training data
- Europe’s languages are vastly under-resourced, except English
- A concerted effort for the **collection of enormous amounts of language data for all European languages** is very much needed

Language Data Space: What is it?

- One of the 14 official EU data space projects
- Objective: Develop and deploy a European platform and marketplace for the collection, creation, sharing and re-use of **multilingual** and **multimodal** language data
- Type of action: procurement (CNECT/LUX/2022/OP/0026)
- Runtime: 36 months (+ 12 months if renewed)
- Salient features: governance framework, technical architecture and infrastructure, openness, promotion
- Stakeholders: **industry**, research, public administration, cultural associations, NGOs and citizens

EUROPEAN LANGUAGE DATA SPACE

Target data

- Very large data volumes
- Generic and domain specific language data
- For pretraining LLMs but also for fine-tuning and continuous training
- Multiple language tasks (machine translation, summarization, Q&A, …)
- Already available (e.g. regular corpora, web crawls) but mainly **new fresh data** from publishing houses, media industry, libraries, call centres, etc.
- Text, audio, video, image - any modality
- In different languages - monolingual / multilingual / parallel
- In different domains, different communication scenarios
- Emphasis is on **raw data** but also
  - Annotated datasets
  - Language models
  - Lexical, terminological resources, ontologies, semantic lexica, knowledge graphs, …
  - Language data processing services

Semantic interoperability before LDS (1)

Semantic interoperability before LDS (2)

- Most popular metadata models for exchange: Dublin Core, OLAC (OAI-PMH harvesting protocol)
- Component Metadata Infrastructure (CMDI) ➔ ISO 24622-2:2019
- Mainly in XSD
- ISOcat ➔ CLARIN Concept Registry
- META-SHARE model ➔ META-SHARE ontology (MS-OWL, as an initiative of the W3C LD4LT community group)

Revisiting semantic interoperability: Requirements

- **Simple and flexible** metadata model for **new providers** (with less technical knowledge around language data and technologies)
- Interoperability **across** data spaces (DSSC recommendations, Data Space Protocol, etc.)
- **Simplify** mappings and **conversion** process (and support tools)
- Support **multilinguality** of the model and of the descriptions
- As before, cover description requirements for language data (and language models)
- Allow for continuous updates and enrichments

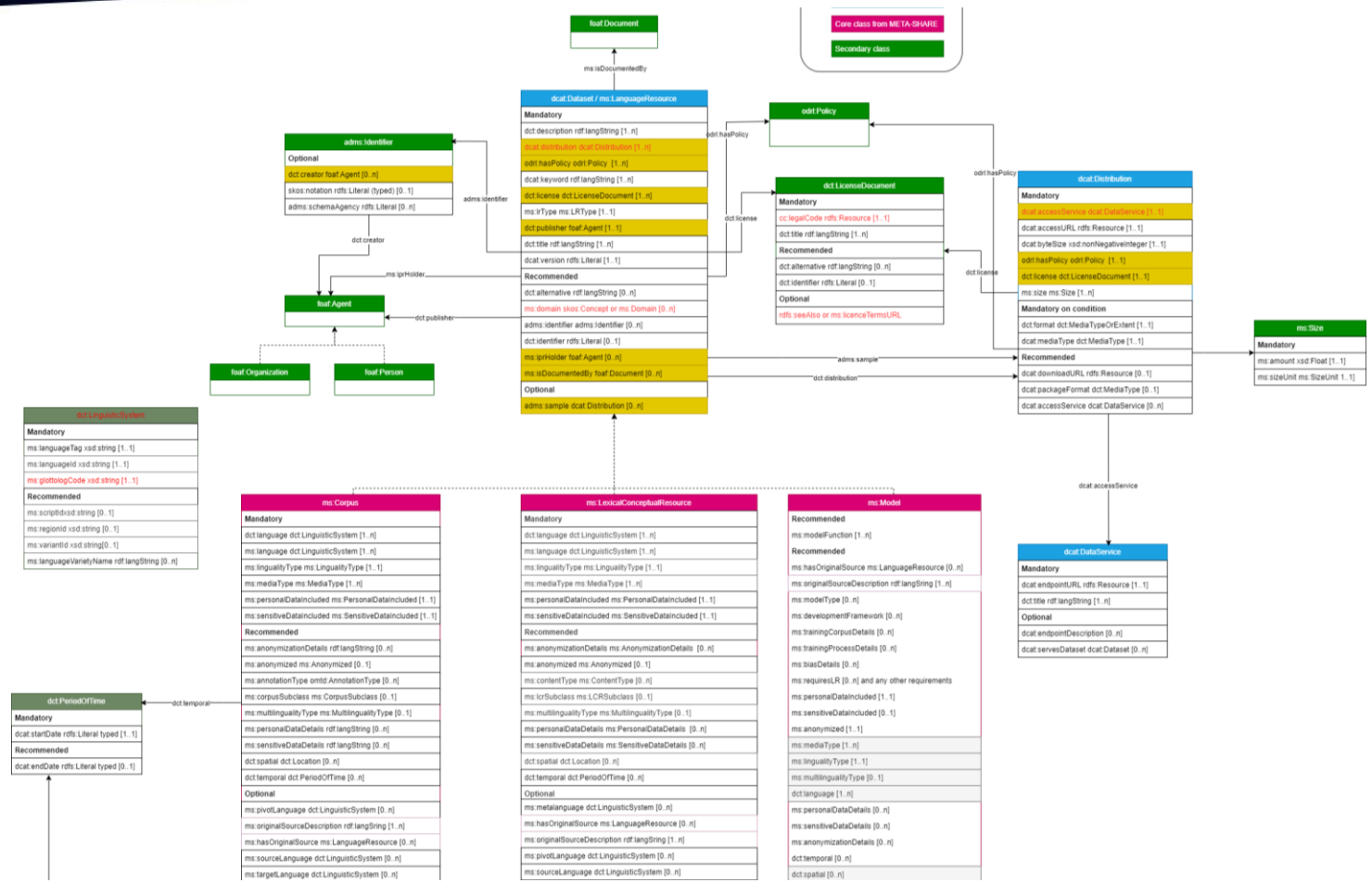
Revisiting semantic interoperability: Adopted approach

- Adoption of a "common" metadata model
- Base the LDS metadata model on a commonly accepted standard: DCAT and DCAT-AP
- Customize the model to meet the community-specific description needs

➔

LanguageDCAT-AP profile based on

- DCAT & DCAT-AP properties for generic concepts
- META-SHARE ontology for language data concepts



LanguageDCAT-AP v0.8

Focus on data resources:

- Corpus
- Language Model
- Lexical/Conceptual resource

("Data Service" in next versions)

DCAT-AP vs. LanguageDCAT-AP (1)

- Re-using DCAT-AP properties, classes and recommendations as much as possible
- Minimize to-the-extent possible mandatory properties (trying to balance between the **consumer's** requirements and the **provider's** wishes)
- Adding community properties: types and subtypes of dataset/language resource, GDPR aspects, content (e.g. annotation types, model functions, etc.)
- "Changing" some data types: e.g., free text properties with range rdf:langstring instead of rdf:Literal

DCAT-AP vs. LanguageDCAT-AP (2)

- Mandatory/Recommended controlled vocabularies:
  - keep where possible, e.g. "dct:spatial" (EU Vocabularies Continents, EU Vocabularies Countries, EU Vocabularies Places, Geonames), "dcat:mediatype" (IANA Media Types)
  - for language data specific properties, use our own vocabularies (e.g. size units, model functions, standards, etc.)
  - some cases with our vocabularies (but we're planning to ask for additions in the EU vocabularies and, if this is done, adopt the EU vocabulary): e.g. "dct:format"

"language": between standards (1)

- Various standards and vocabularies/ontologies
  - ISO 639: ISO 639-1(*en*) & ISO 639-3 (*eng*) and linked data representations (e.g. https://vocabs.dariah.eu/en/)
  - BCP 47 (*en-US*)
  - lexvo, glottolog, wikidata, ...
  - EU Vocabularies Languages Named Authority List for dct:language

BUT we need

- more detailed descriptions: *Brazilian Portuguese*, *Viennese German*, *greeklish* (Greek in Latin script), specific dialects
- translation equivalents of language labels in EU official languages

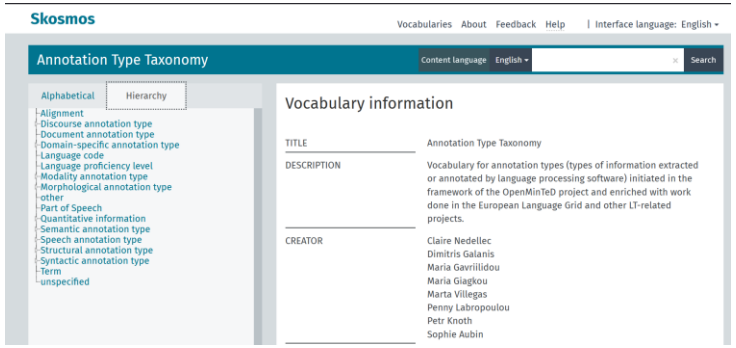
"language": between standards (2)

APPROACH ADOPTED

- 'ms:language': combining properties implementing BCP47 (mandatory *language* & optional *region, script, variant*) and optional property 'ms:languageVarietyName' (free text)
  - ➔ for use in LDS
- 'dct:language': EU languages authority list & lexvo ontology
  - ➔ for use across data spaces
  - ➔ automatically computed from 'ms:language'

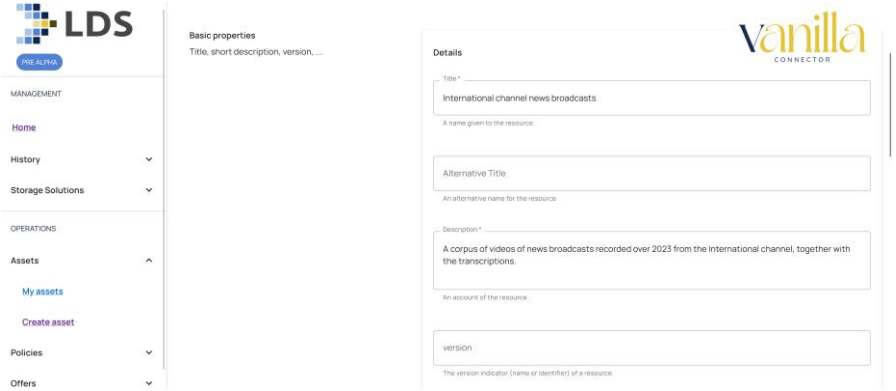
Governance & Support tools (1)

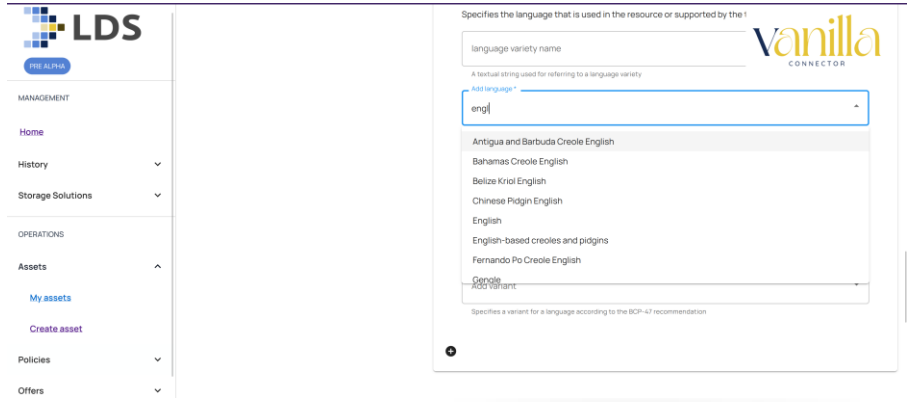
- For the model
  - V0.8 in RDF/OWL and SHACL; vocabularies in SKOS
  - Documentation: work in progress
  - Vocabularies published in Skosmos
  - Updates:
    - of the model: at regular intervals by a restricted set of editors following feedback received
    - of the vocabularies: using Vocbench; discussing collaborative editing
  - Translation: using an MT system; output to be validated by native language experts
  - Mappings and crosswalks
    - to be made available over GitHub and/ or where possible, integrated in the RDF/OWL model
  - Vocabulary Hub (under investigation)
    - for addition of models and mappings and for consultation through the catalogue



Governance & Support tools (2)

- For metadata descriptions
  - GUI *(and planned API)* for creating and validating metadata descriptions integrated in the LDS connector, including
    - API for the import of controlled vocabularies
    - SHACL for validation purposes
  - SHACL validator service to be made available over GitHub
  - Converters with metadata models to be made available over GitHub; already for two infrastructures (ELG & ELRC-SHARE), ongoing work for other infrastructures (CLARIN)





Semantic interoperability for policies (1)

- DCAT-AP: 'dct:license' & 'odrl:hasPolicy': both properties allowed (recommended) but not connected
- DSP – Catalog specification: *A Dataset must have 1..n 'odrl:hasPolicy' attributes*
- ODRL vocabulary: supports semantic interoperability, facilitates policy enhancement (where computationally possible) BUT
  - is not human user-friendly
  - consensus on the representation of licensing terms is not always guaranteed

Semantic interoperability for policies (2)

- In LDS, we want to support both licences and policies, i.e.
  - (re-)use of standard (open) licences ➔ LDS connector prefilled with ODRL representations for them
  - For new licences/policies, ongoing implementation of a policy wizard combining “policy classes” (subset from IDSA policy classes and common practices in our community) with the addition of attributes (e.g., dct:title, dct:description, etc.)
  - Copying the relevant property-value pairs from "policy" to "licence" (to ensure consistency) for standard licences
- Needed: a central point for all data spaces to ensure proper semantic interoperability between licensing terms

Future work

- Focus on data models
  - Collection of vocabularies (cf. lexical/conceptual resources)
  - Collection and documentation of popular data models
  - Collection and documentation of converters between data models
- Vocabulary hub
  - For adding new metadata and data models that are not publicly available
  - For creating mappings between (meta)data model properties / classes / individuals
  - Across data spaces
    - for understanding (meta)data models from other data spaces
    - for sharing controlled vocabularies

Thank you!

Public Procurement Data Space (PPDS)

A paradigm shift in public procurement data

Semantic Interoperability solutions for data spaces

Marc Christopher – DG-GROW
José Pedro Revenga Arias – NTT Data

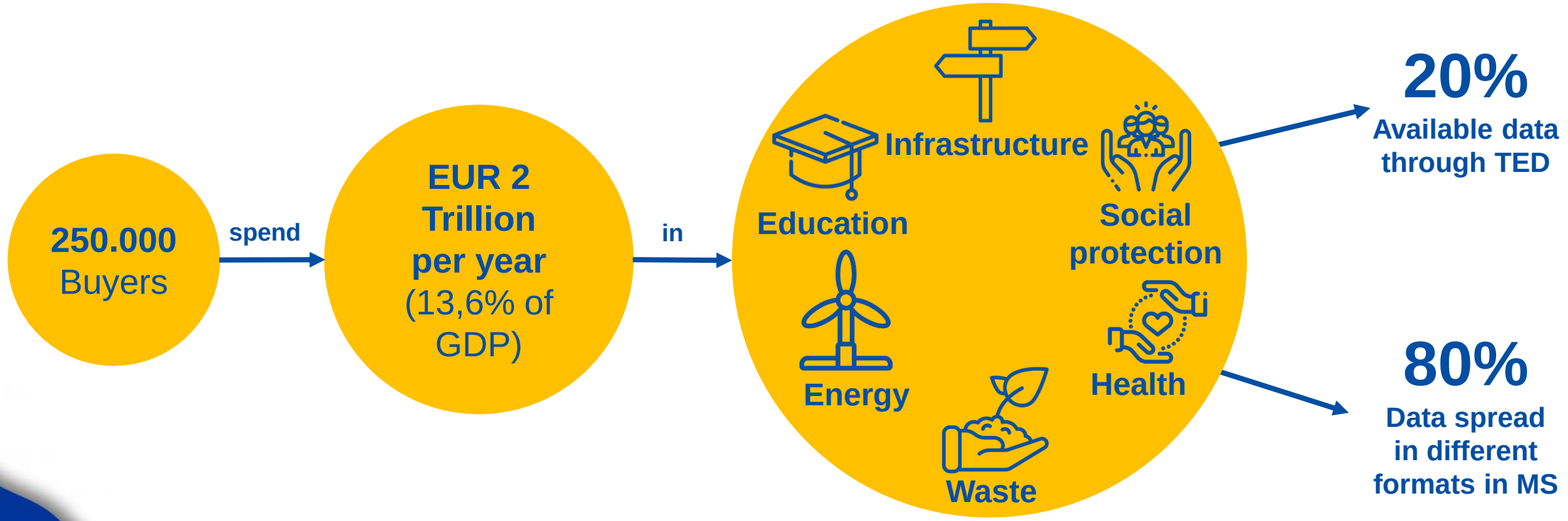
The Business Side

- Why we need the PPDS?
- Use cases of the PPDS
- Way forward

The Business Side

Of the PPDS

Why we need the PPDS?



The PPDS

- Enables **easy access to public procurement data** to all user groups
- Improves **monitoring of public procurement** in the EU
- **Reinforces transparency**
- Supports **improving data quality**
- **Raises awareness** of the level of competition in public procurement
- **Supports cooperation** between Member States and Commission
- **Is open source** allowing Member States to re-use the solution

Use cases of the PPDS in the beginning

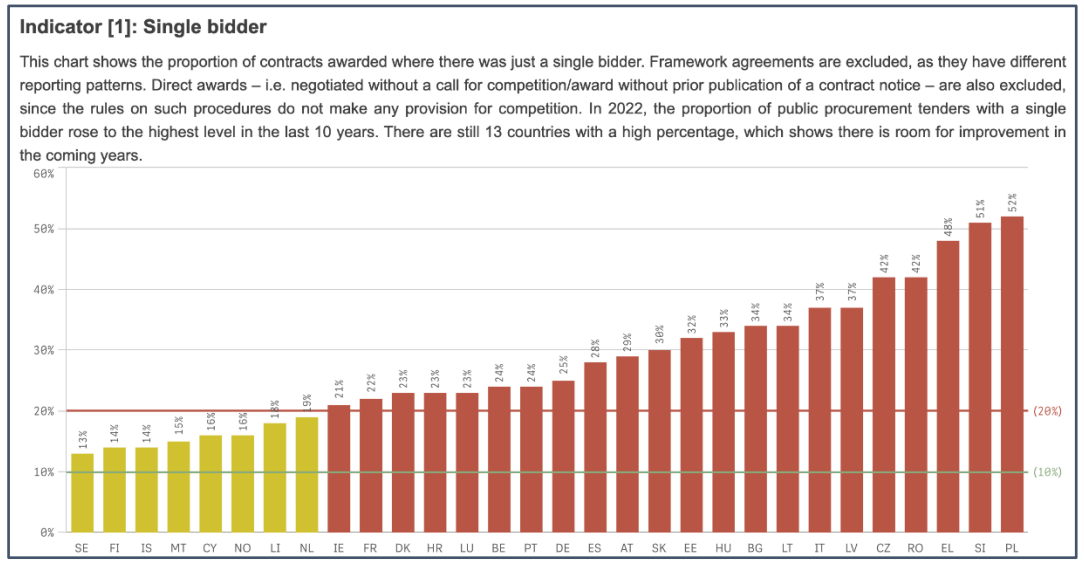
Monitoring of Public Procurement market

Data Quality measurement

Strategic Procurement measurement

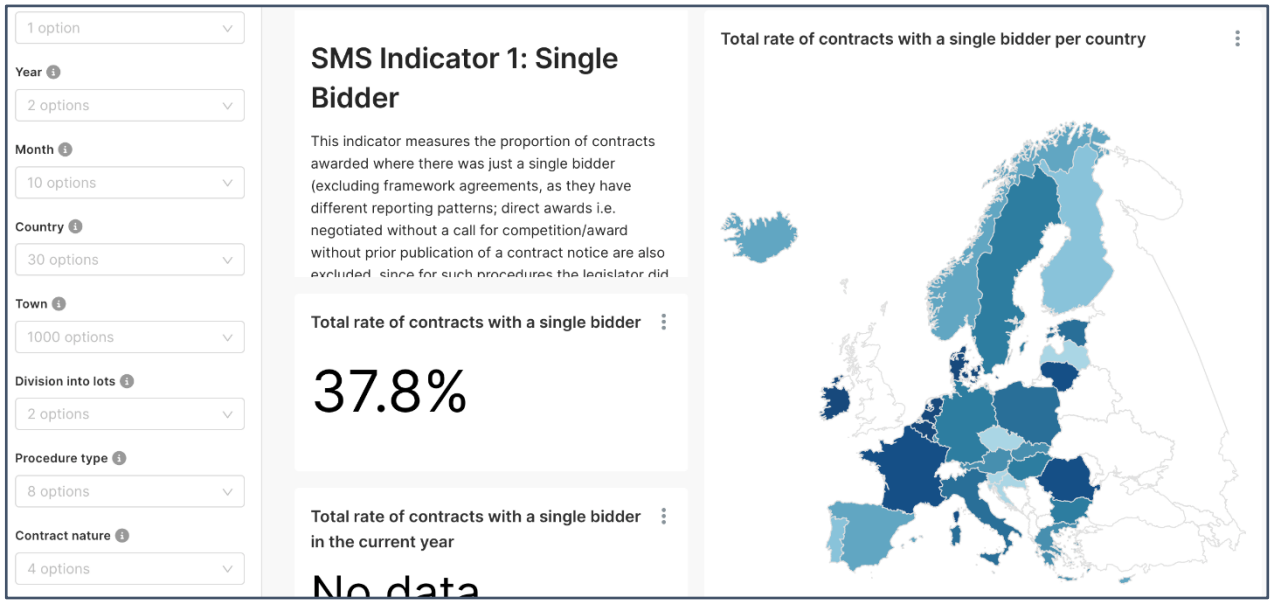
Use Case - Monitoring of Market

Current Single Market Scoreboard



- No maps or filtering possible

PPDS Dashboards



- Maps and filtering makes analysis possible
- Fulfils parts of the recommendation from ECA

Use Case – Data Quality

PPDS indicators on data quality

Completeness

Business rule	Issue rate
NUTS code is provided (Place of performance)	45.0%
Legal type description of the Buyer	44.9%
Main activity of the Buyer is provided	44.9%

Validity

Business rule	Issue rate
'Procurement procedure is divided into lots' indicator follows a Boolean format	45.0%
Award Criteria Weight is ≥ 0	17.8%

Consistency

Business rule	Issue rate
If winner is from an EU Member State, the number of received EU tenders must be ≥ 1	20.0%
If winner is SME, number of received SME tenders must be ≥ 1	4.2%
Sum of award criteria weight must be 100	1.8%
If winner is from a non-EU Member State, the number of received non-EU tender must be ≥ 1	0.4%
Number of received tenders ≥ number of received SME tenders	0.0%

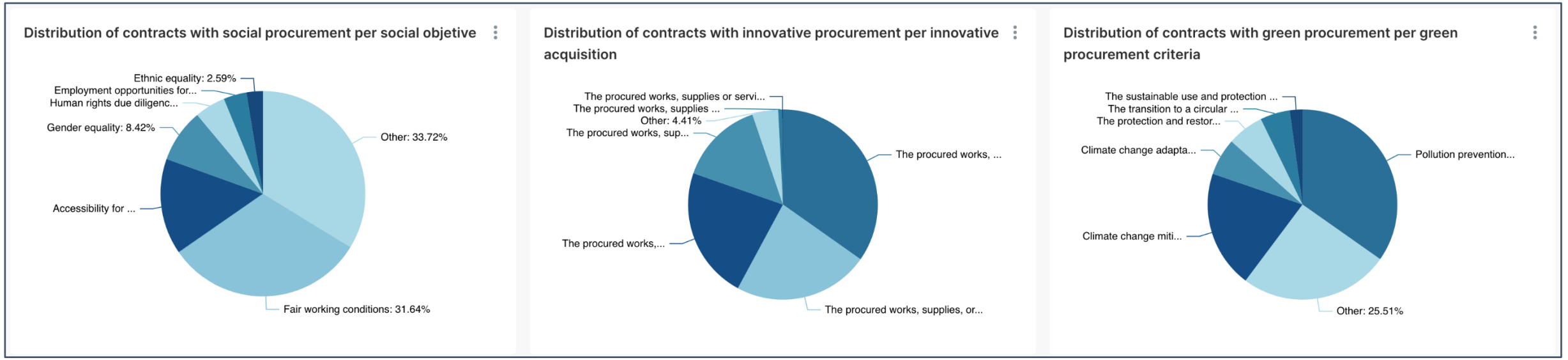
Accuracy

Business rule	Issue rate
For procurement procedures with estimated value above a given threshold, the Estimated Value should not follow an unrealistic number pattern (e.g. 111,111,1111; 222,222,222; 333,333,333; 444,444,444; 555,555,555; 666,666,666; 777,777,777; 888,888,888; 999,999,999 or 123,456,789)	0.0%
Number of received tenders should be within a reasonable range (e.g. below 250 bids)	0.0%

- All notices are checked against business rules on different angles
  - Completeness
  - Validity
  - Consistency
  - Accuracy

Use Case – Strategic Procurement

PPDS indicators on distribution of social, green and innovative objectives



- eForms allows buyers to indicate if the procedure includes social, environmental or innovative objectives
- For each objective buyers can provide more details like “Fair working conditions”
- It is not mandatory to provide this information and currently not all countries are making use of it

Way forward

Q3 2024

- Stabilising the PPDS and improving reliability
- Go live of the PPDS in Summer 2024
- PPDS Day 2024 on 24 September – The Grand Opening

Q4 2024 and Q1 2025

- Continue onboarding other countries and further data mappings
- Continue to work together with the community

Q2 2025 and beyond

- Enhance analytical capabilities, including AI
- Include services to support buyers to improve data quality (buyers AI companion)

The Semantic Side

- The Challenge
- The Solution
- The Standards
- Why BRegDCAT-AP?
- The Data Model
- Next steps

The Semantic Side

Of the PPDS

The challenge

Harmonise the data under a common data model (eProcurement Ontology - ePO)

- 5 data sources
- 31 different data models

Ensure the data meet the minimum quality requirements

- 2,850 billion triples
- 65 business rules

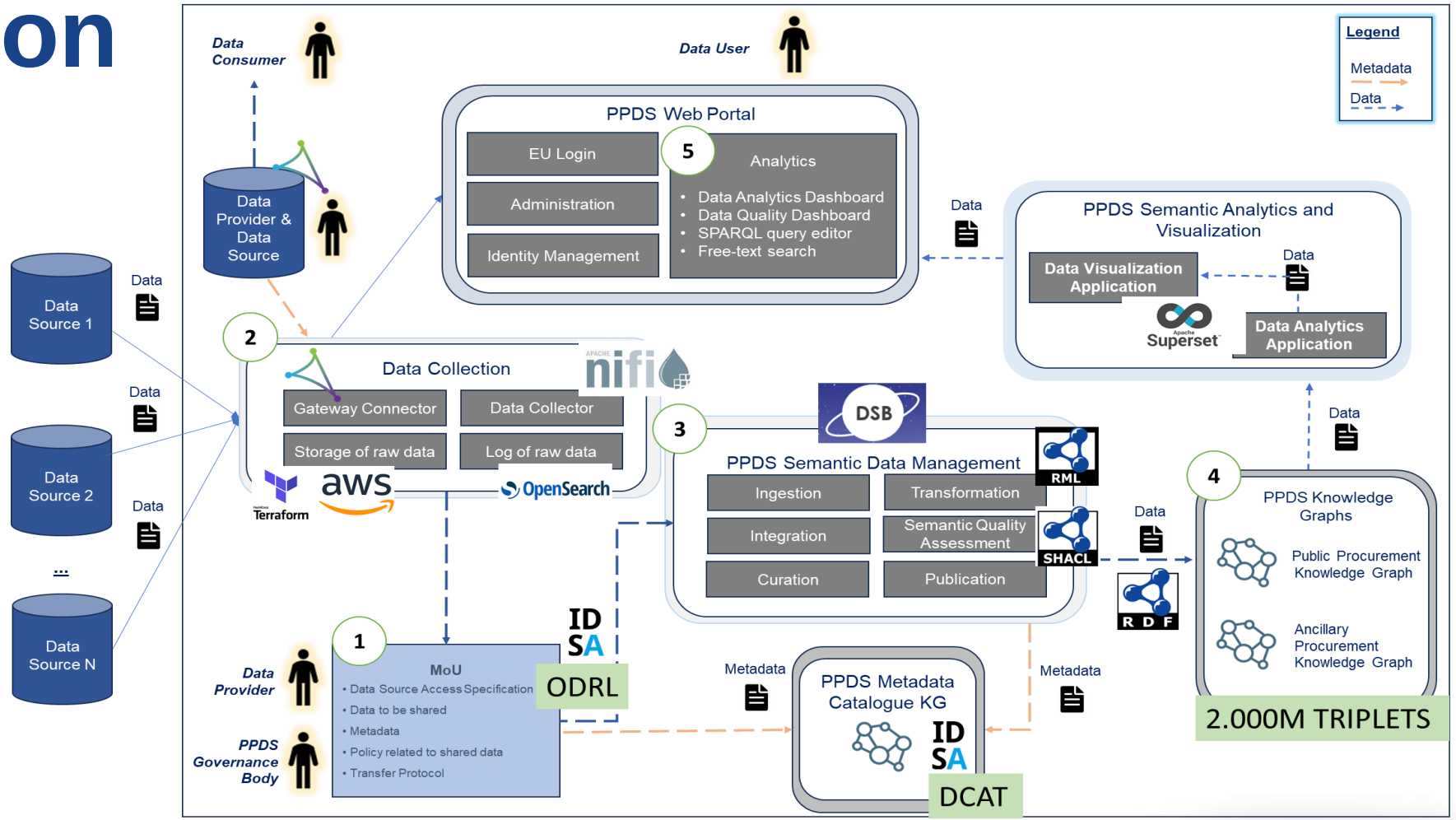
Guarantee the correct integration of the data

- 1,000 billion entities

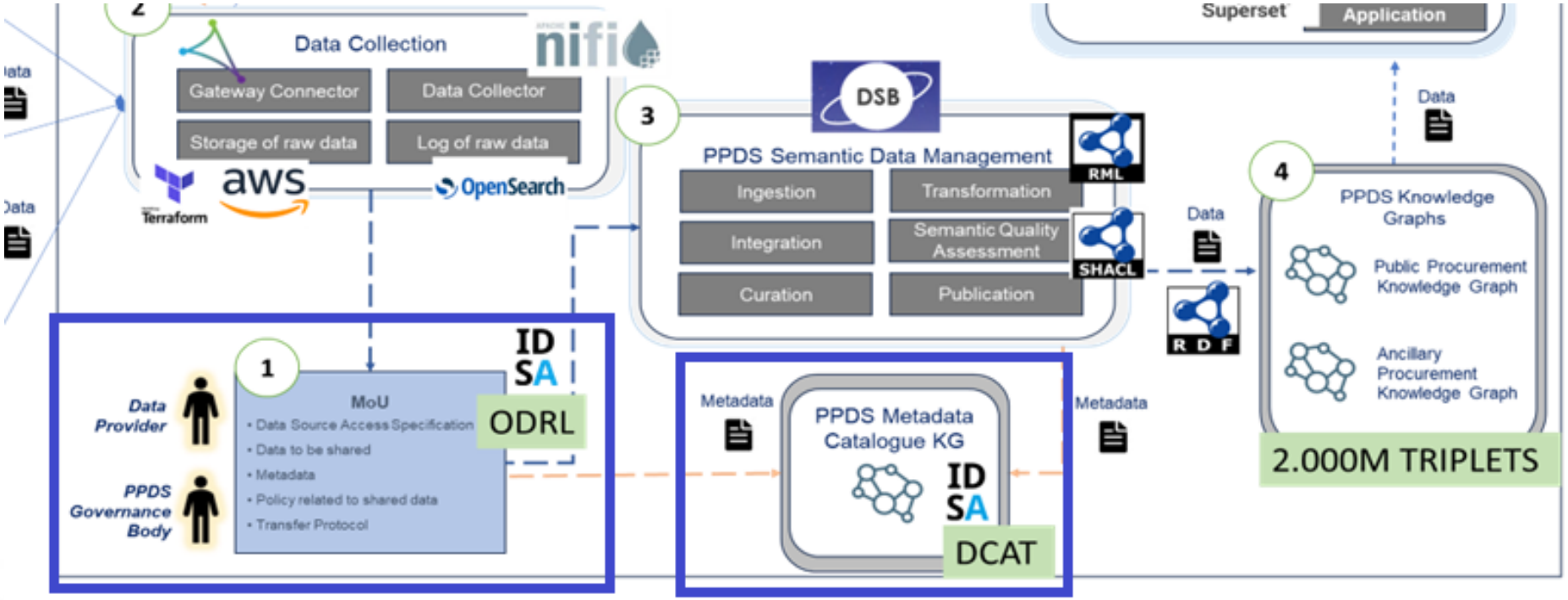
Inventory all resources and enable data traceability

- 5 data catalogues

The Solution



The Standards

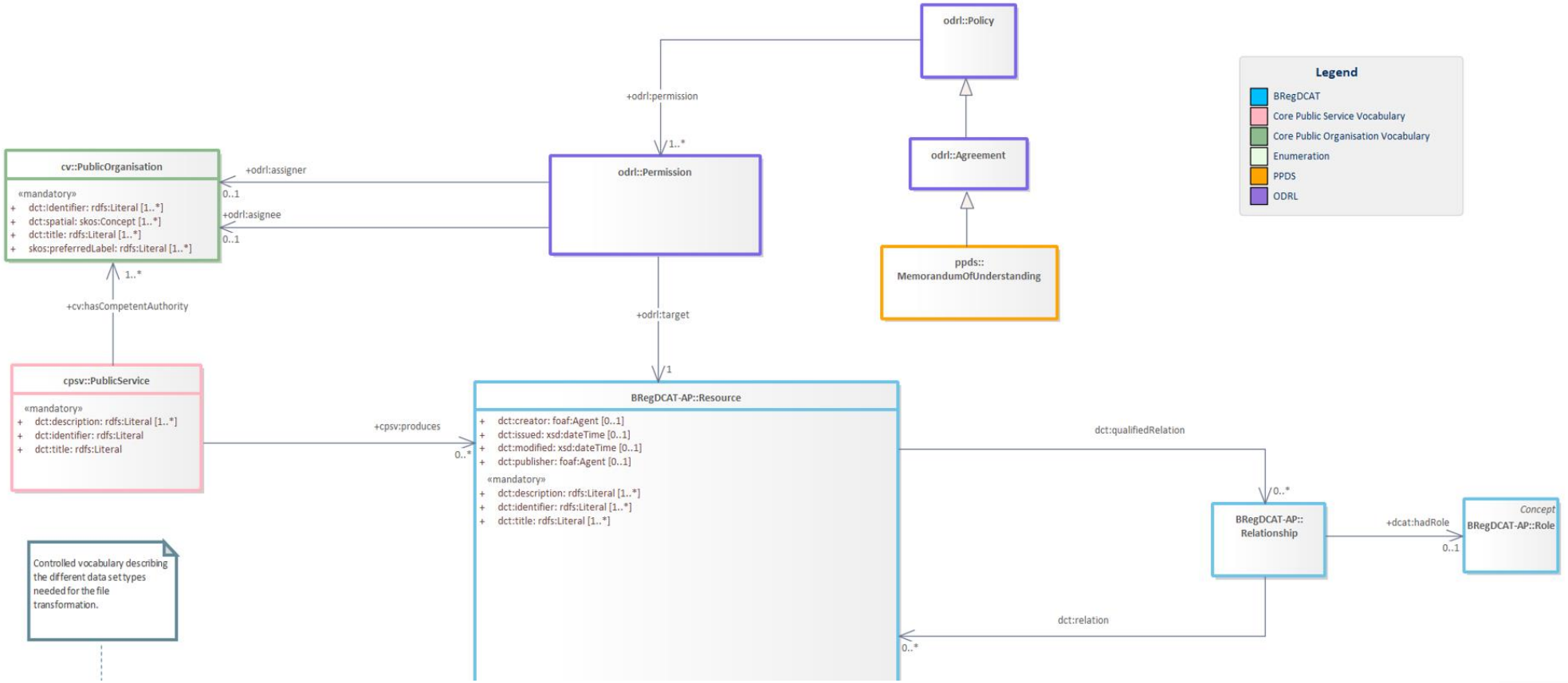


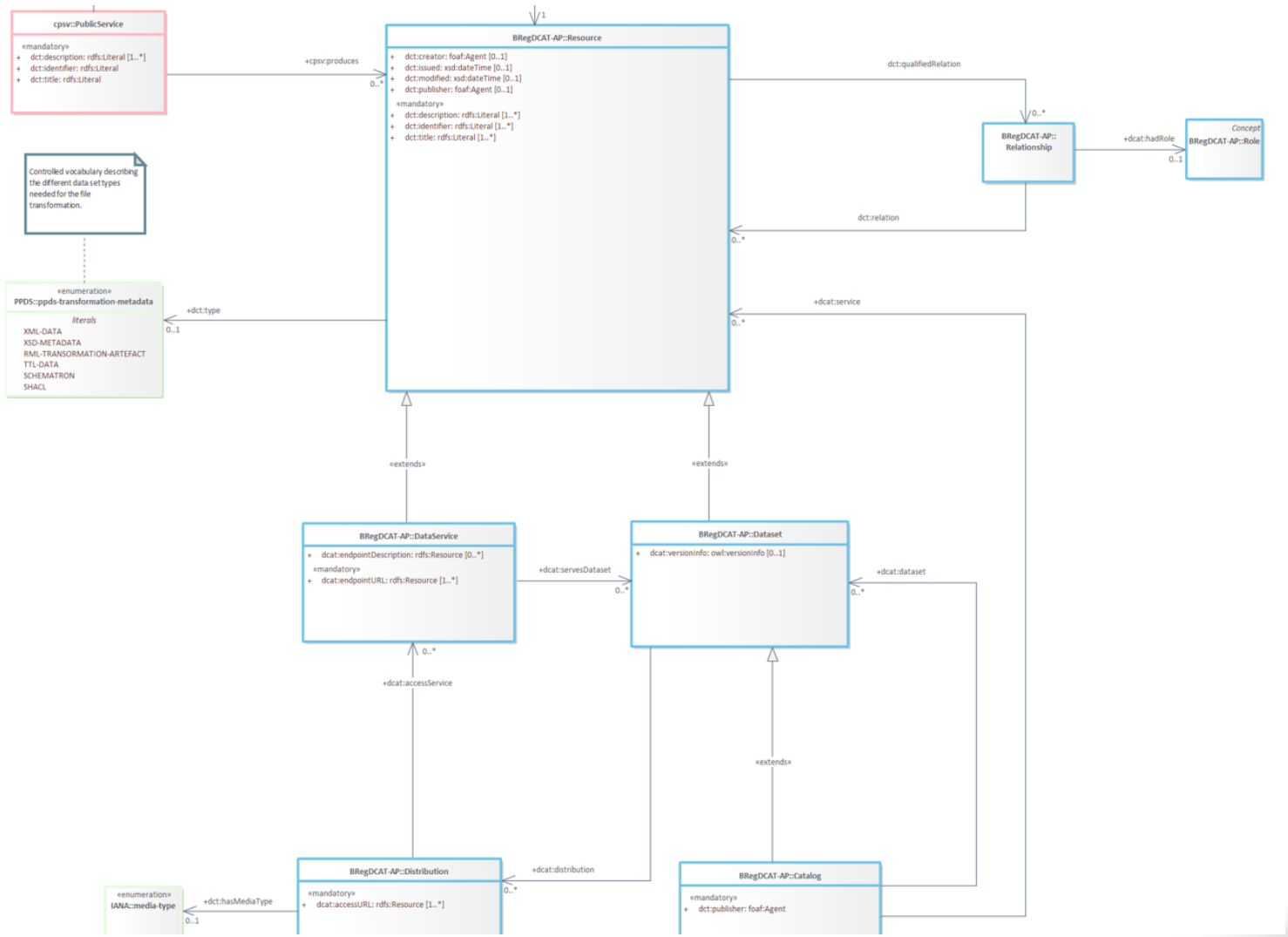
- DCAT-AP for Base Registries (BRegDCAT-AP)
- Open Digital Rights Language (ODRL)

Why BRegDCAT-AP?

- Based on DCAT:
  - Reliable and widely used
  - Compatible with IDSA, Gaia-X & other data spaces pioneers vision
- Tailored for European base registries:
  - Extended classes and properties unique to describing metadata of base registries
- Enabling and profiting from the Access to Base Registries Action:
  - Facilitates the interconnection with national base registries across Europe
  - Facilitates the creation of a one-stop platform for accessing and managing base registries across the EU

The Data Model




cpsv::PublicService
«mandatory»
+ dct:description: rdfs:Literal [1..*]
+ dct:identifier: rdfs:Literal
+ dct:title: rdfs:Literal
+cpsv:produces
0..*
1
BRegDCAT-AP::Resource
+ dct:creator: foaf:Agent [0..1]
+ dct:issued: xsd:dateTime [0..1]
+ dct:modified: xsd:dateTime [0..1]
+ dct:publisher: foaf:Agent [0..1]
«mandatory»
+ dct:description: rdfs:Literal [1..*]
+ dct:identifier: rdfs:Literal [1..*]
+ dct:title: rdfs:Literal [1..*]
dct:qualifiedRelation
0..*
BRegDCAT-AP::
Relationship
+dcat:hadRole
0..1
Concept
BRegDCAT-AP::Role
Controlled vocabulary describing the different data set types needed for the file transformation.
dct:relation
0..*
«enumeration»
PPDS::ppds-transformation-metadata
literals
XML-DATA
XSD-METADATA
RML-TRANSORMATION-ARTEFACT
TTL-DATA
SCHEMATRON
SHACL
+dct:type
0..1
+dcat:service
0..*
«extends»
«extends»
BRegDCAT-AP::DataService
+ dcat:endpointDescription: rdfs:Resource [0..*]
«mandatory»
+ dcat:endpointURL: rdfs:Resource [1..*]
+dcat:servesDataset
0..*
BRegDCAT-AP::Dataset
+ dcat:versionInfo: owl:versionInfo [0..1]
+dcat:dataset
0..*
0..*
+dcat:accessService
«extends»
«enumeration»
IANA::media-type
+dct:hasMediaType
0..1
BRegDCAT-AP::Distribution
«mandatory»
+ dcat:accessURL: rdfs:Resource [1..*]
+dcat:distribution
0..*
BRegDCAT-AP::Catalog
«mandatory»
+ dct:publisher: foaf:Agent

Next steps

- BRegDCAT-AP v3.0.0
- Core Vocabularies evolution
- Integration of new kind of participants

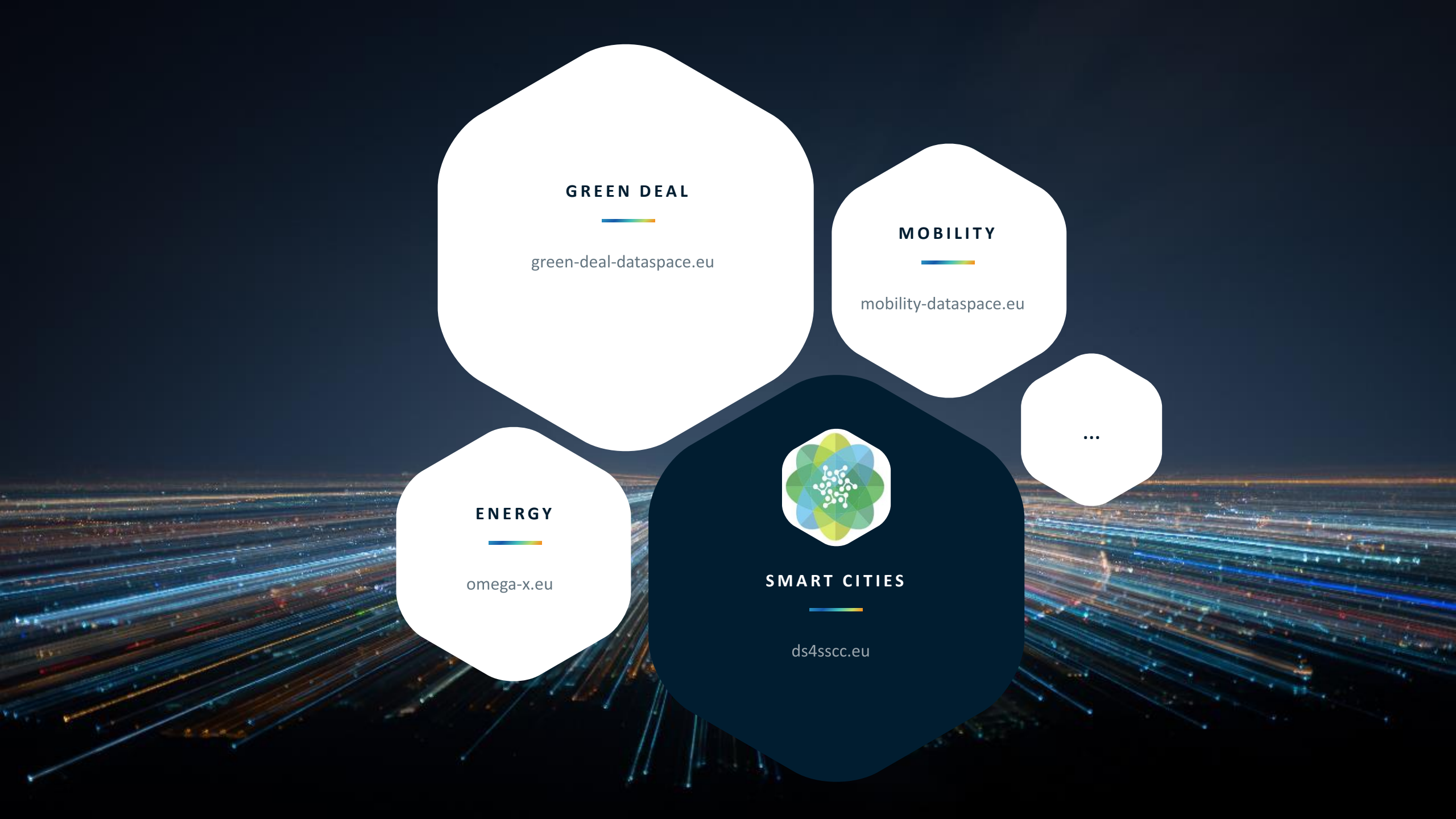
Thank you!


Brussels 27 June
SEMIC 2024 conference
interoperable europe
from Vision to Reality
be EU belgium24.eu
European Commission

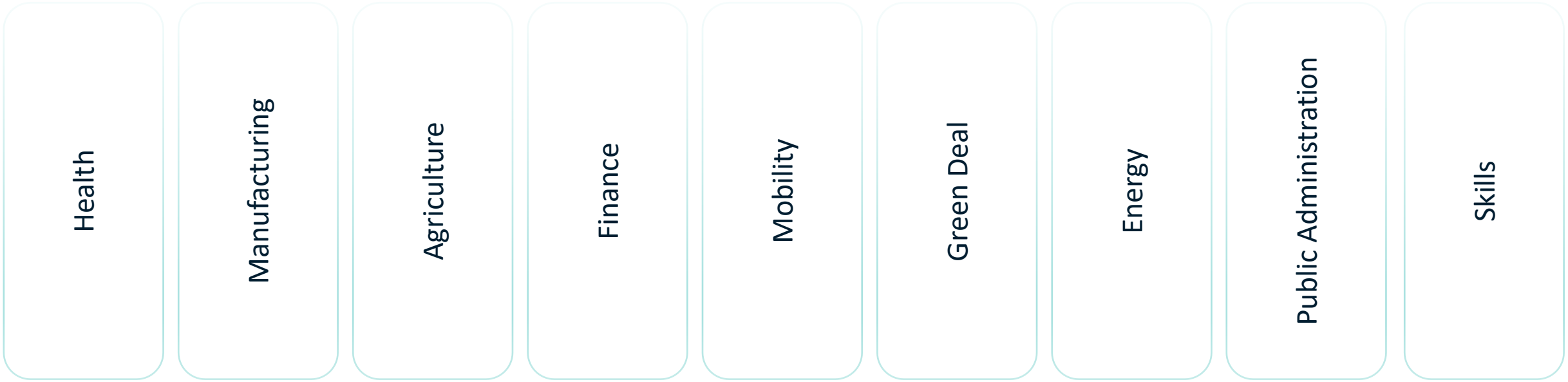
interoperable europe

innovation ∞ govtech ∞ community

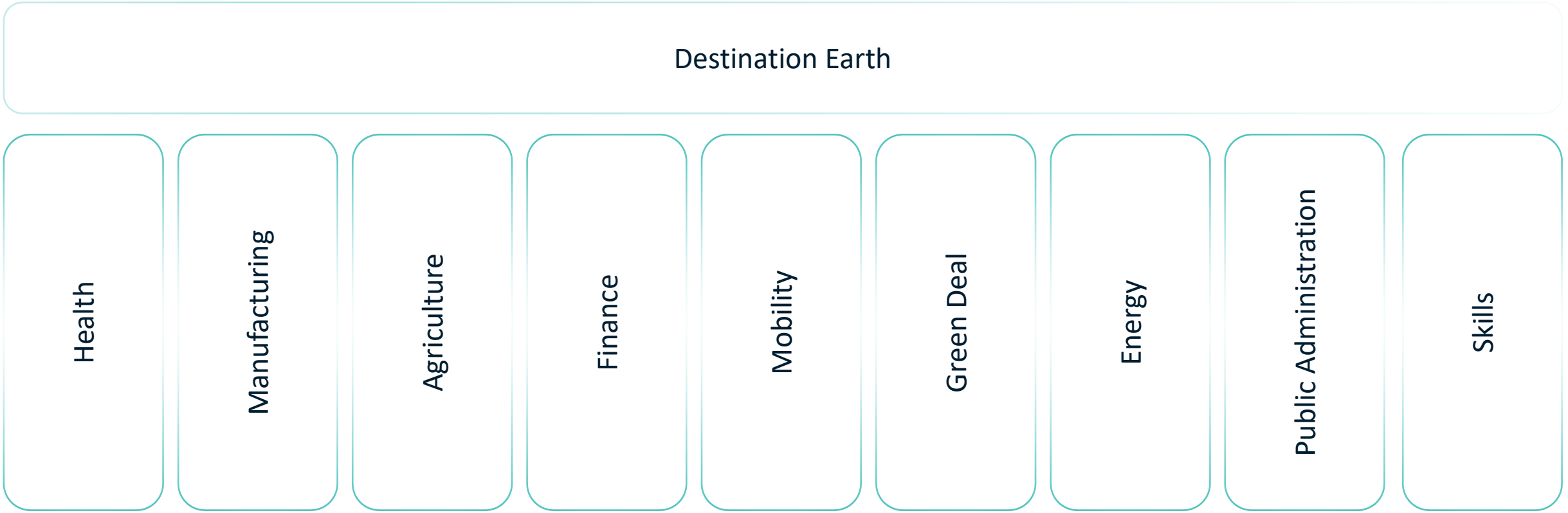
EUROPEAN DATA SPACE FOR SMART AND SUSTAINABLE CITIES AND COMMUNITIES


GREEN DEAL
green-deal-dataspace.eu
MOBILITY
mobility-dataspace.eu
...
ENERGY
omega-x.eu
SMART CITIES
ds4sscc.eu

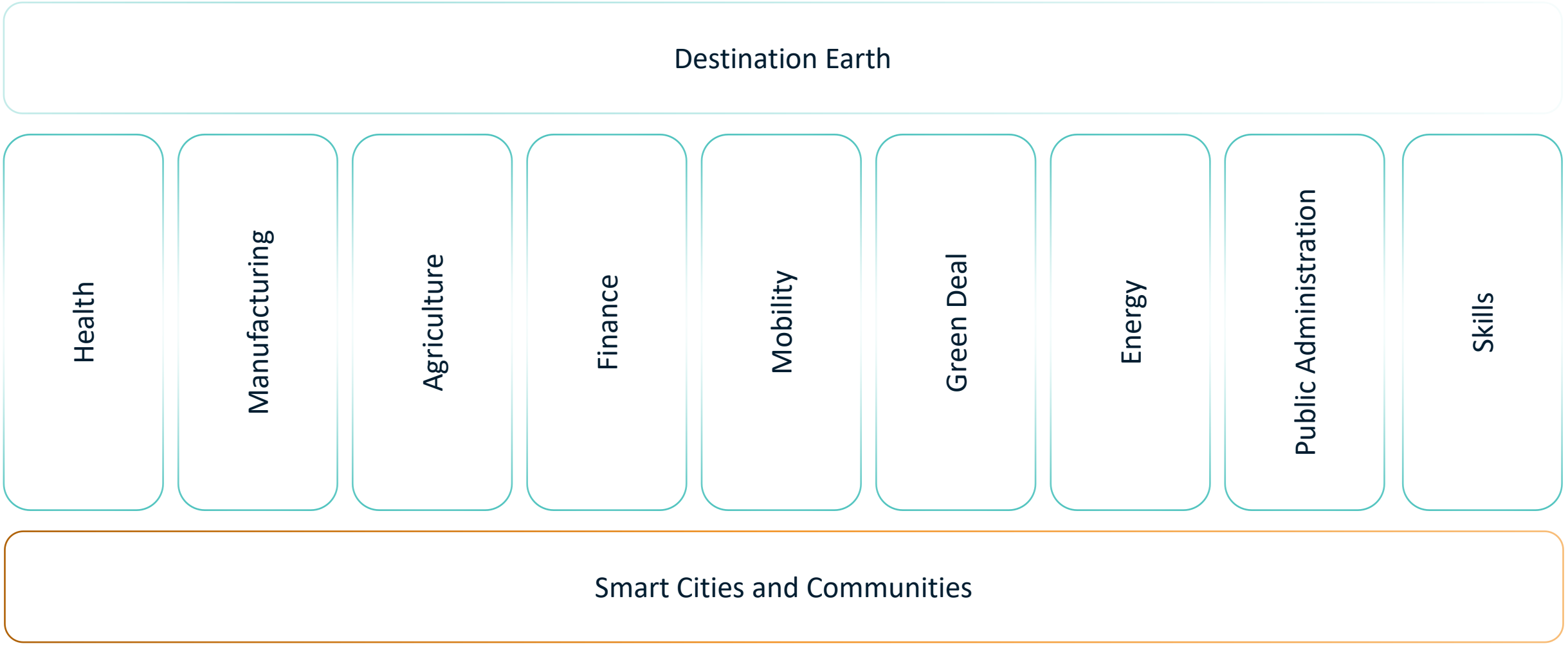
Data Space Ecosystem



Data Space Ecosystem



Data Space Ecosystem




Destination Earth
Health
Manufacturing
Agriculture
Finance
Mobility
Green Deal
Energy
Public Administration
Skills
Smart Cities and Communities

Managing Semantics within local gov’t

A 4-step approach

USE AN INTERNAL DATA CATALOG

Ideally use DCAT-AP

TAKE OWNERSHIP OF DATA MODELS

Create a dictionary, list vocabularies and enforce data models in procurement

REUSE EXISTING VOCABULARIES

Core Vocabularies, Schema.org, OSLO, Smart Data Models, …

INTRODUCE SEMANTIC DATABASES

Have a graph db or triplestore to underpin services

A Case Study of Ghent

Federated semantic public services using CPSV

After the COVID pandemic the City Council wanted to stimulated tourism by extending terrace permits for local bars and restaurants

A Case Study of Ghent

Federated semantic public services using CPSV

However, these new rules did not apply in other cities and for business owners it was hard to find the information online

Data Space approach

INFORMATION ON LOCAL SERVICES IS DESCRIBED SEMANTICALLY

This is done using the CPSV Ontology

THIS INFORMATION IS HARVESTED AND FEDERATED ON THE REGIONAL LEVEL

The Flemish government runs a project called “IPDC” or “Intergovernmental Products and Services”

FLANDERS REPUBLISHES THE ACCURATE SERVICE DESCRIPTIONS

This way, data is collected straight from the local source, and people can easily find new or different services and regulations on the local government level

Managing Semantics Applied To Ghent

Description of Public Services

01

USE AN INTERNAL DATA CATALOG

DCAT-AP is used internally and for the open data portal

02

TAKE OWNERSHIP OF DATA MODELS

CPSV is used internally and cross-government to describe local public services

03

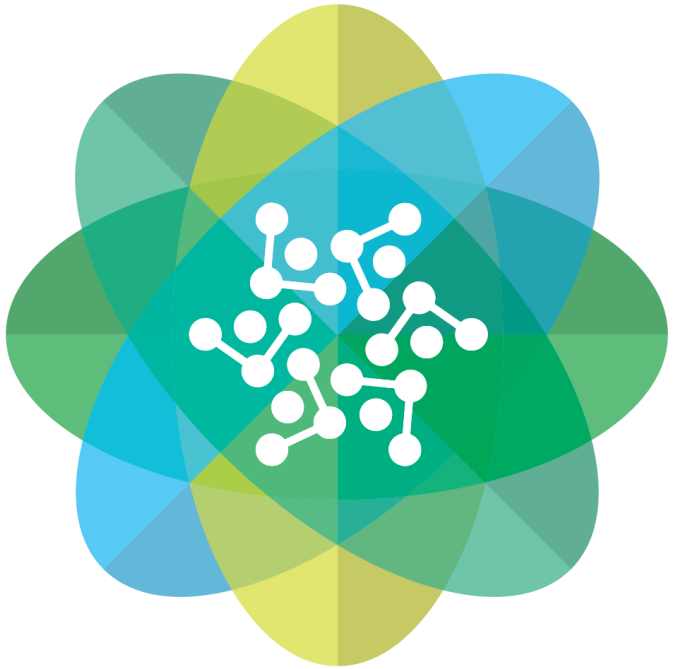
REUSE EXISTING VOCABULARIES

CPSV is one of the many ontologies Ghent uses. Many of them are Flemish, co-created within the OSLO programme

04

INTRODUCE SEMANTIC DATABASES

Ghent runs a Virtuoso triple store which also feeds information to the public facing website and the open data portal



DS4SSCC

The European Data Space for Smart and Sustainable Cities and Communities

The European data space for smart communities is an EU-wide action creating a federated cross-sectorial data space for governments on all levels and their providers to deliver the best possible services to their citizens by enabling interoperability to reach critical goals, including the Green Deal.

10 – 12 Pilots

Shaping the European Data Space for cities and communities

Open for local and regional public administrations in the EU and partners (e.g., companies, academia, civil society organisations) working with them.

Up to 1.5 million euro co-funded for 50%

CROSS - SECTORAL

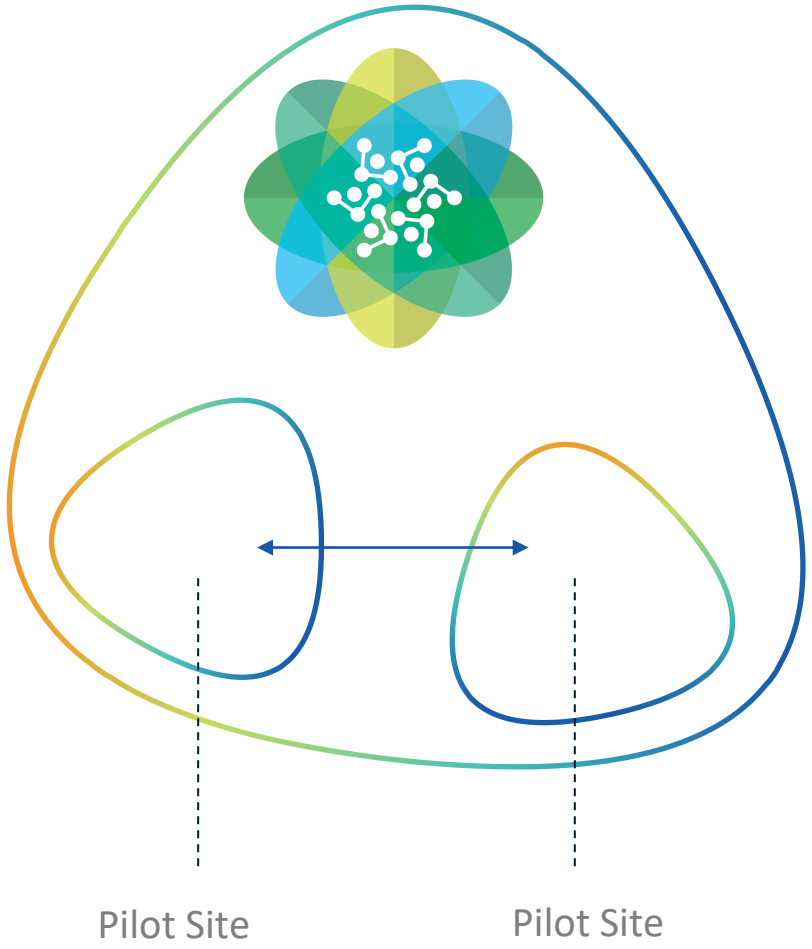
The pilot consortia must create new data services from the data sharing across at least two heterogeneous areas specified below, but can also include other European Green Deal domains

CROSS BORDER

At least two local or regional public administrations from different EU member states and/or Digital Europe Programme Associated countries.

EVOLVING THE BLUEPRINT

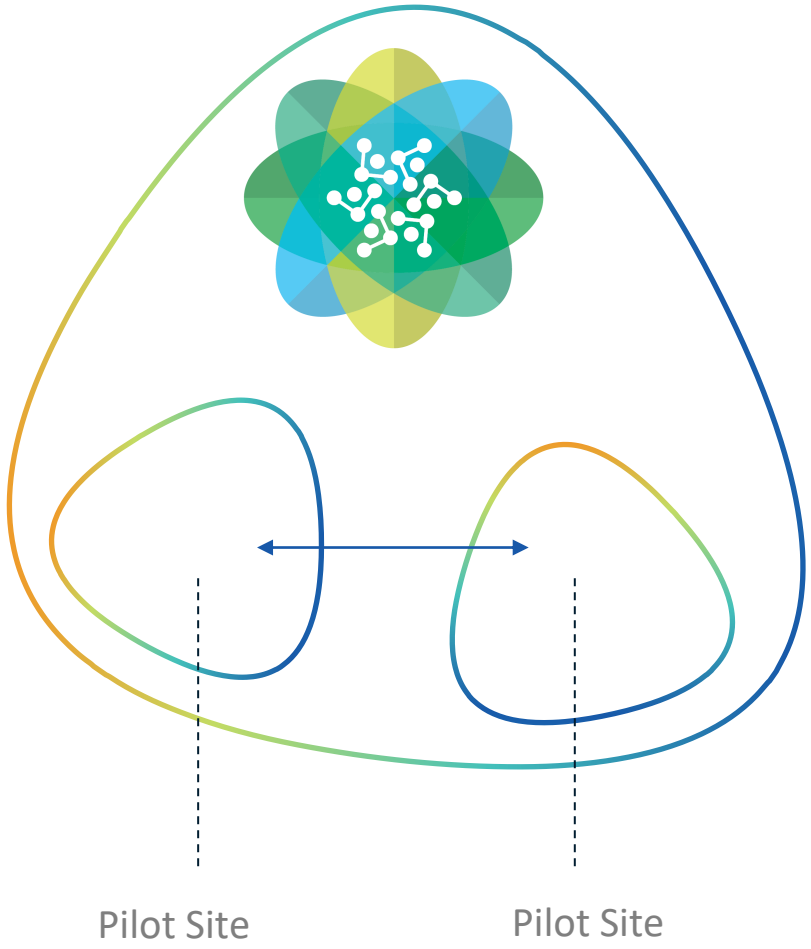
Experiences gained through these pilots will allow the European Data Space to evolve and expand



Data Exchange within a pilot

Each pilot site will require at least:

- Exposure of a data catalog
- Authentication and Authorization policies

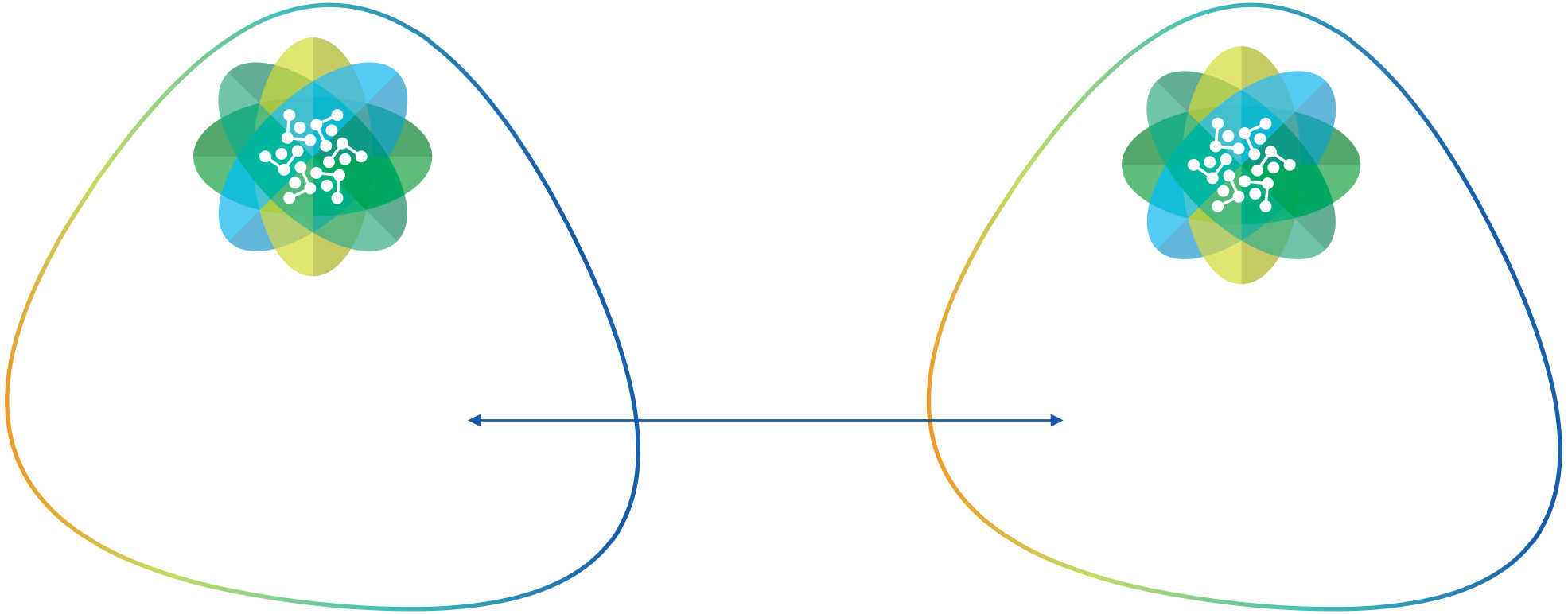


Use the Blueprint

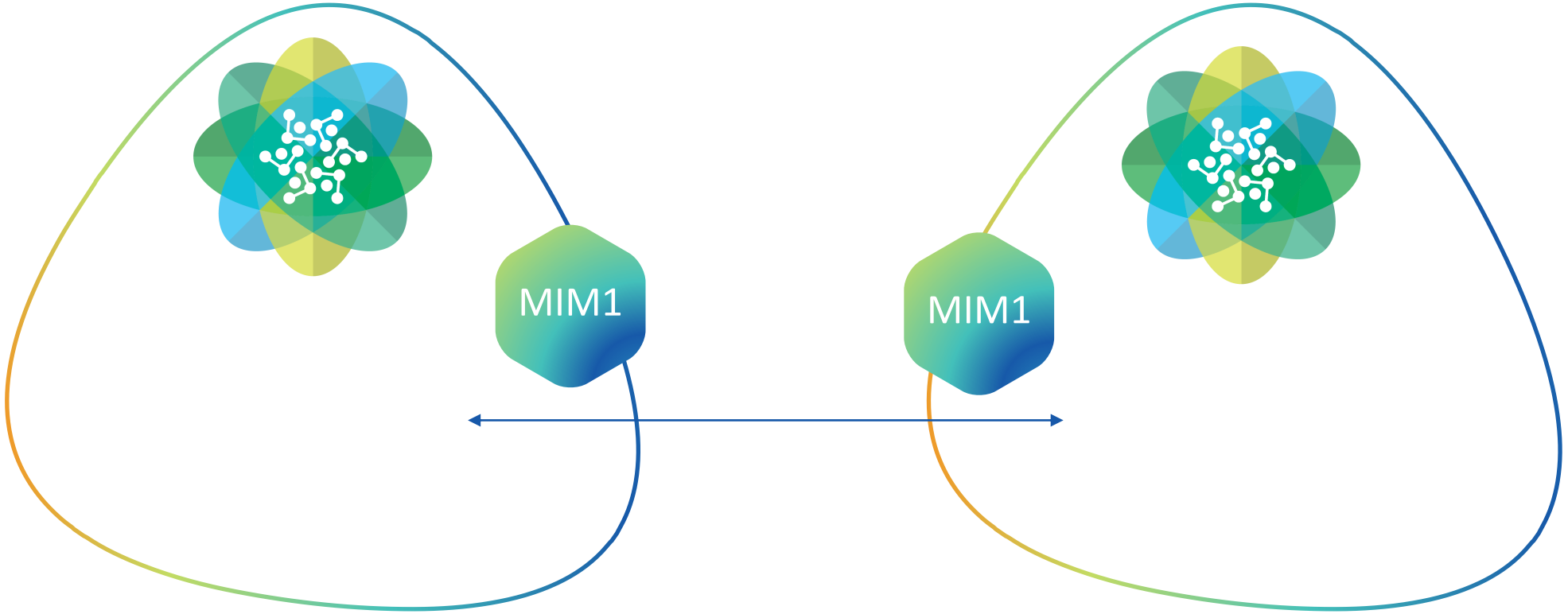
Catalog of building blocks and services available at

https://www.ds4sscc.eu/inventory

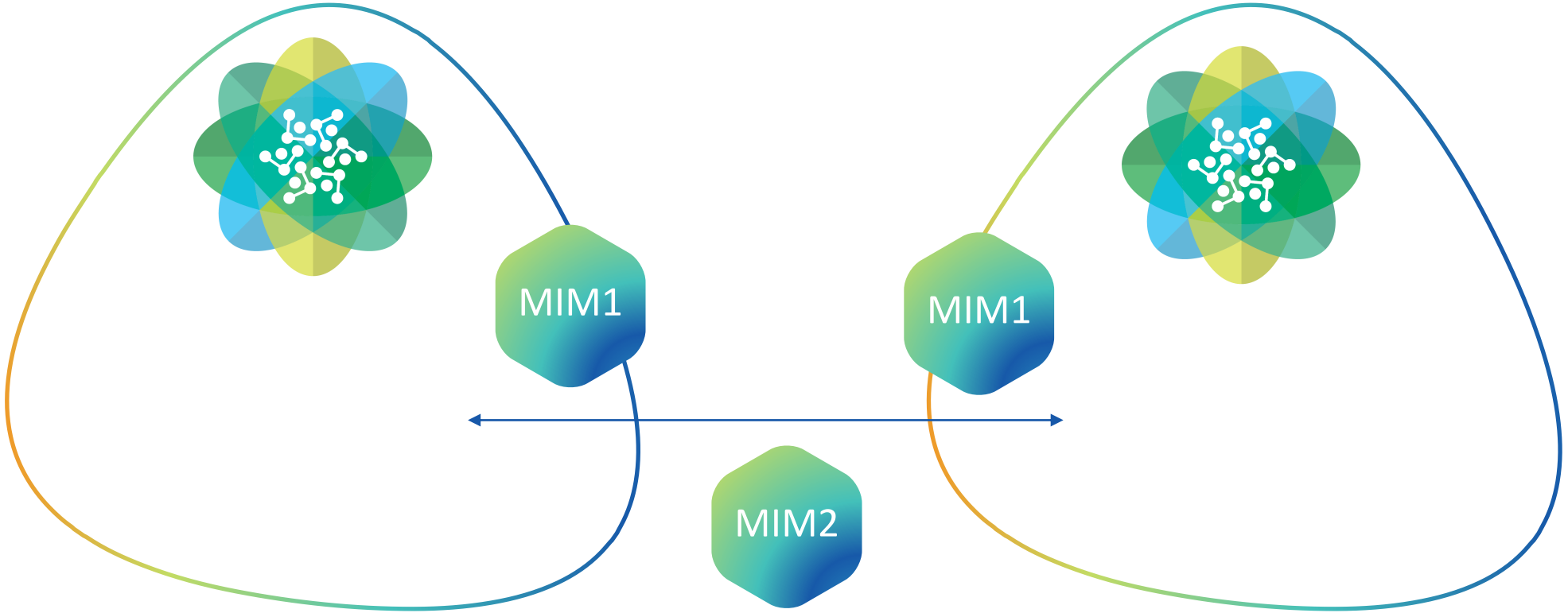
Federating Pilots across Europe



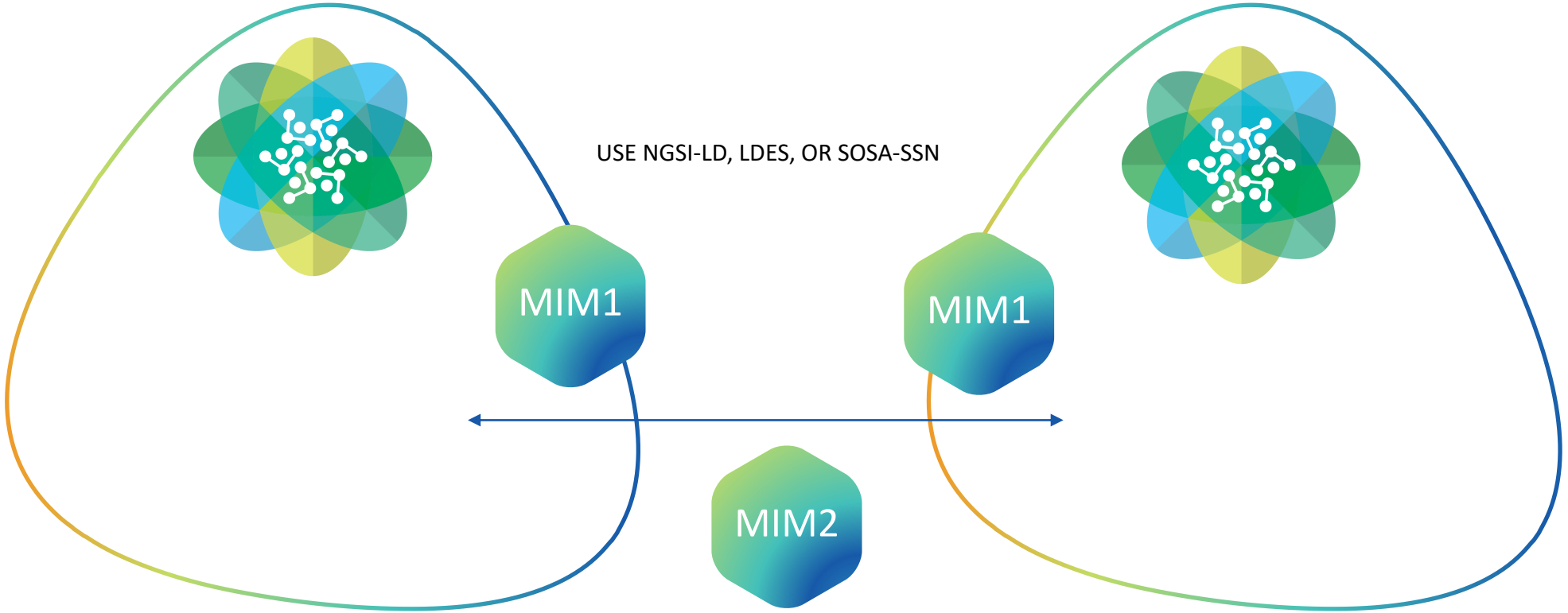
Federating Pilots across Europe



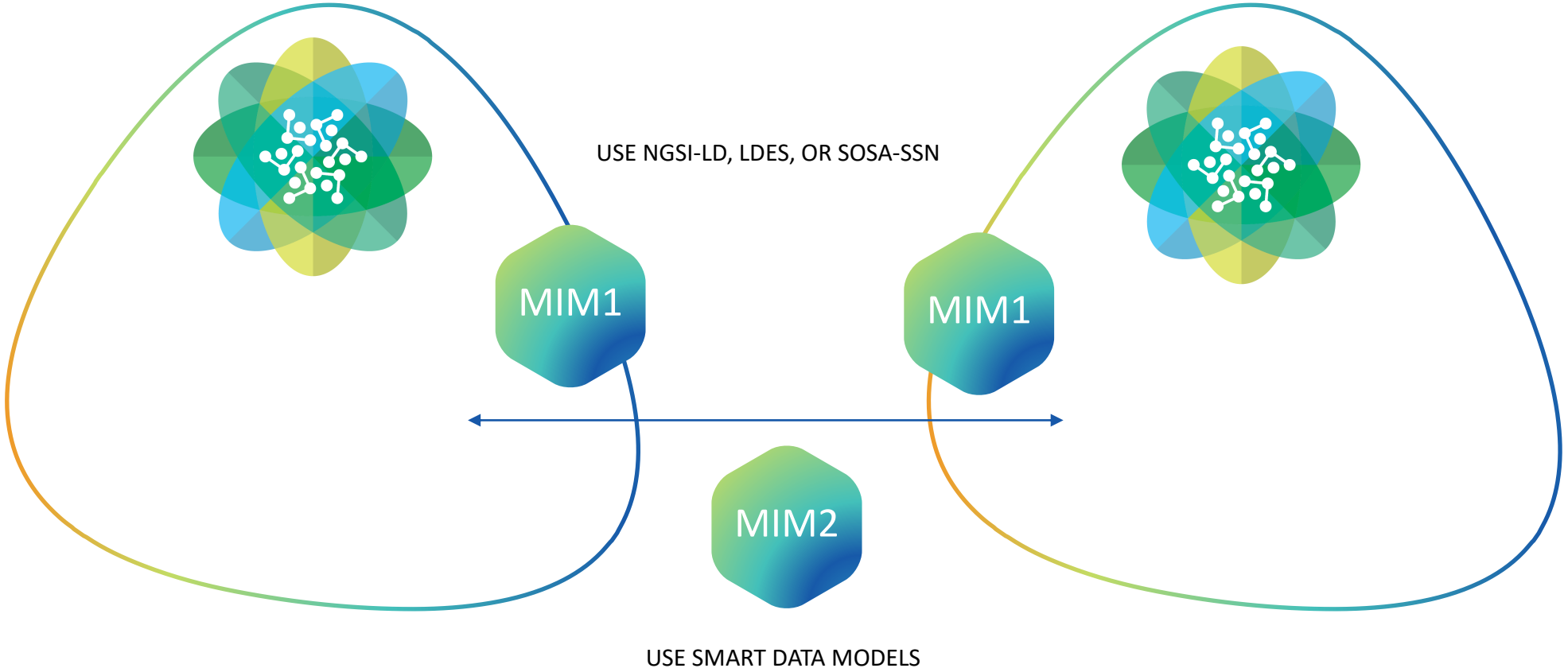
Federating Pilots across Europe

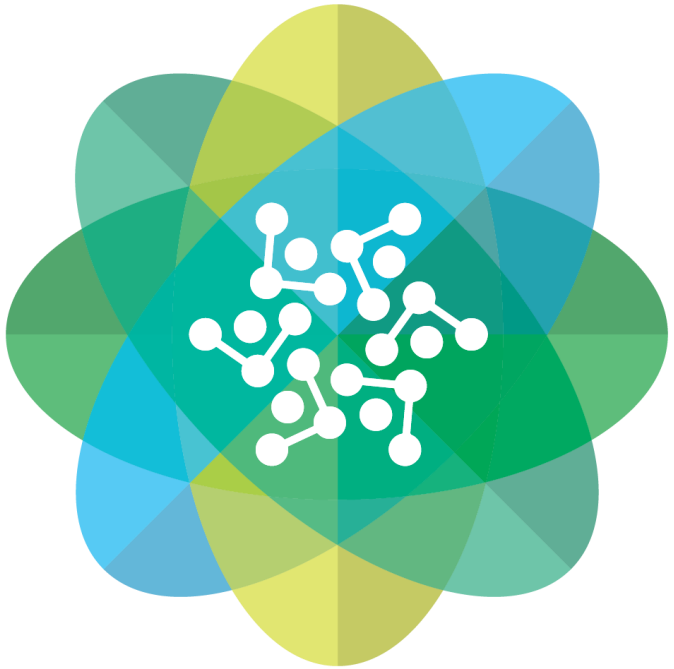


Federating Pilots across Europe



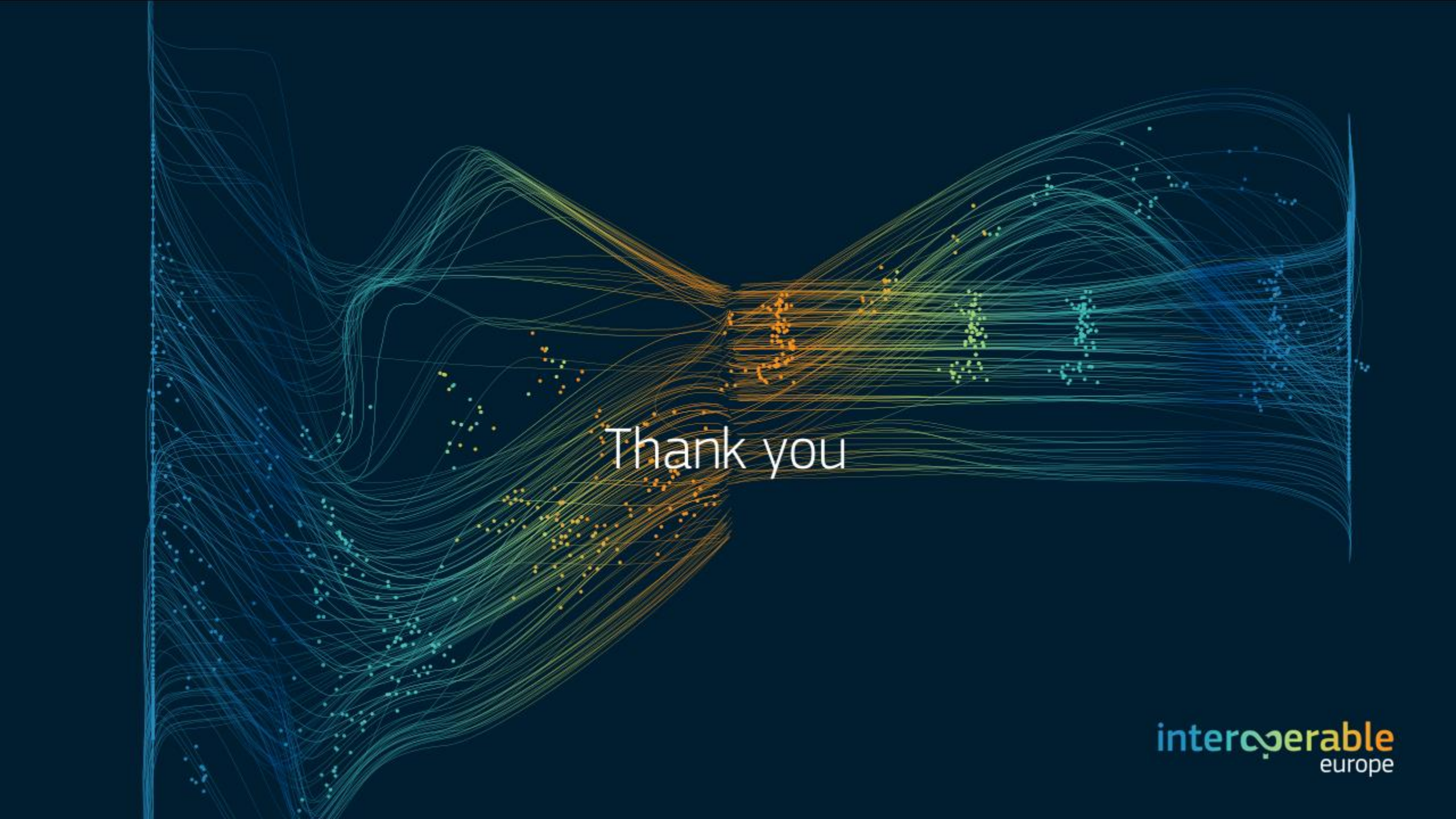
Federating Pilots across Europe





Open Call 1 now open!

https://www.ds4sscc.eu/cfp-one



interoperable europe