European Commission

1 October 2024

Solutions for establishing Semantic Interoperability across data spaces: SEMIC approach and offering

@3rd Workshop on Semantic Interoperability in Data Spaces

interoperable europe

DIGIT.B2 - Interoperability.

Pavlina Fragkou, George Lobo, Marcello Grita

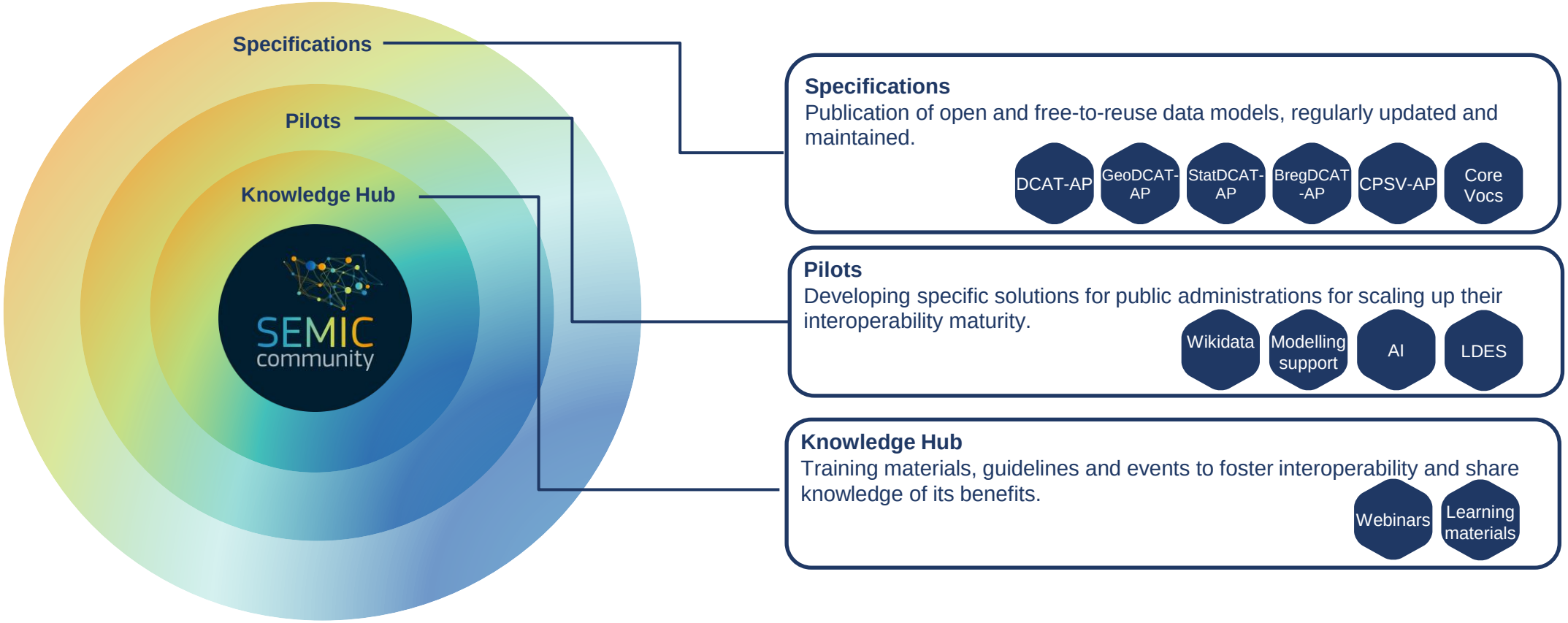
Introduction to SEMIC

The objectives of the SEMIC action are to promote Semantic Interoperability amongst the EU Member States by:

- Promoting the share and reuse of semantic assets, experience and tools and facilitating agreements in key areas.
- Identifying opportunities for alignment on semantic definitions, metadata and reference data sources with special focus on identification and definitions of Core Concepts / Vocabularies.
- Raising awareness on the importance of data and metadata management.

SEMIC

SEMIC's goal is to deliver pragmatic support to help build an interoperable Europe.



Specifications

SEMIC specifications enable interoperability:

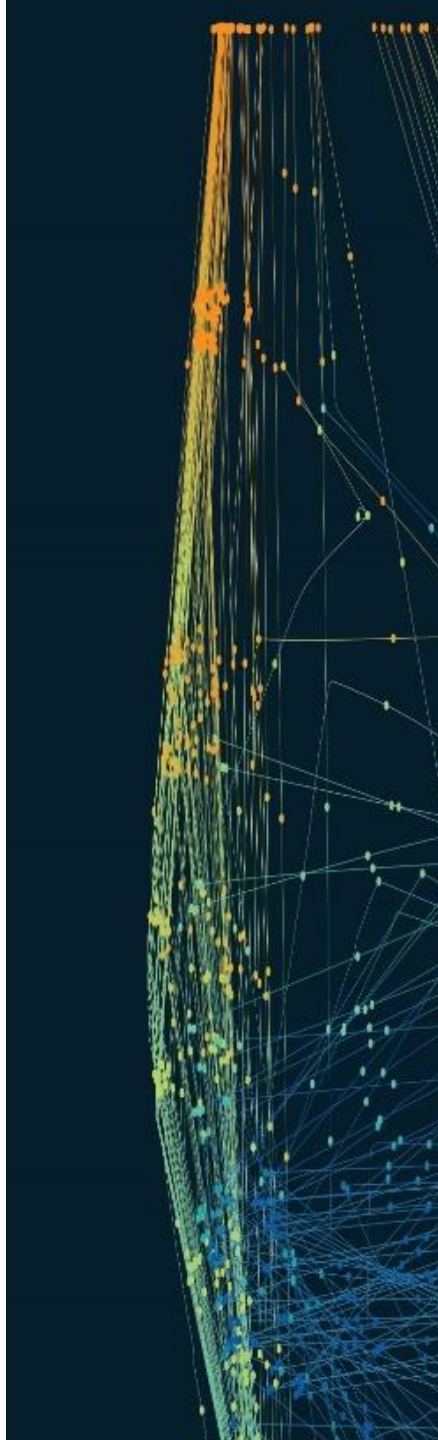
- They make data **transparent** and **available**
- They support the **coherent** implementation of laws and policies
- They help implement **cost efficiencies**
- They help **digitalisation** and **harmonising** processes

Core Vocabularies

Core Vocabularies are a cornerstone element of semantic interoperability. They provide a standardised approach for describing key concepts such as locations, businesses, organisations and natural persons.

Application Profiles

Application Profiles make use of vocabularies for a detailed set of use cases to define mandatory relations, constraints and relationships.



European Commission

SEMIC specifications

Vocabularies

CORE PERSON VOCABULARY

A person's name(s), date and place of birth/death, identifier, addresses, citizenship, etc.

CORE BUSINESS VOCABULARY

The legal name, address, identifier, company type, and activities of a legal entity.

CORE LOCATION VOCABULARY

The different ways of describing a location, e.g. via an address, a geographic name, or a geometry, in alignment with INSPIRE.

CORE CRITERION & EVIDENCE VOCABULARY

The requirements and evidence of a procedure or formal process.

CORE PUBLIC ORGANISATION VOCABULARY

The administrative information, hierarchy, identifiers, events and classification of a public organisation.

CORE PUBLIC EVENT VOCABULARY

A public event, its time, audience, location, etc.

Application Profiles

CORE PUBLIC SERVICE VOCABULARY Application Profile

DCAT-AP FOR DATA PORTALS IN EUROPE

BRegDCAT-AP FOR BASE REGISTRIES

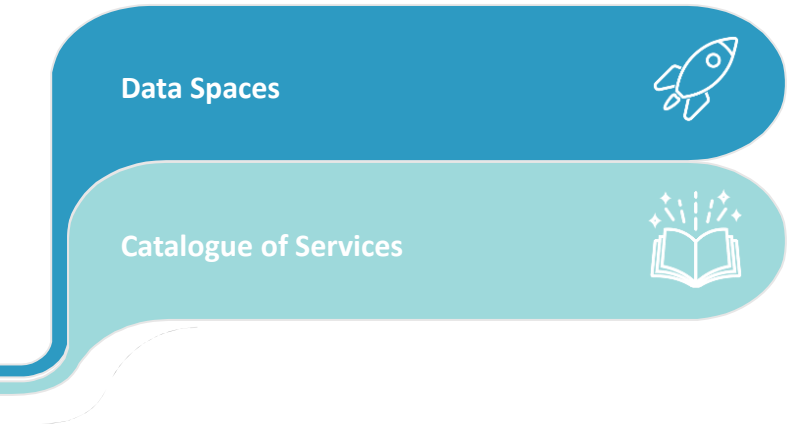
GeoDCAT-AP FOR GEOSPATIAL DATASETS

StatDCAT-AP FOR STATISTICAL DATASETS

ADMS ASSET DESCRIPTION METADATA SCHEMA

Data spaces

Data spaces aim to create a single market for data that will ensure Europe's global competitiveness and data sovereignty.



DCAT-AP

DCAT-AP is the standard solution to ensure metadata is exchanged smoothly across all data spaces.

Core Vocabularies

The e-Government Core Vocabularies are simplified, reusable and extensible data models that capture the fundamental characteristics of a data entity in a context-neutral fashion.
For example, Core Business complies with the BRIS directive and the HVD regulation.

Modelling support

SEMIC provides support to organisations developing their own semantic data models or extending existing ones. The support may consist of sharing of best practices, reviewing models, co-develop the model, mappings and advice on alignments.

LDES

A publishing strategy by which a data provider allows multiple third parties to stay in sync with the latest or historical versions of the data source in a cost-effective manner.

Semantic registry

SEMIC paves the way for increased reusability and discoverability of semantic assets throughout Europe.

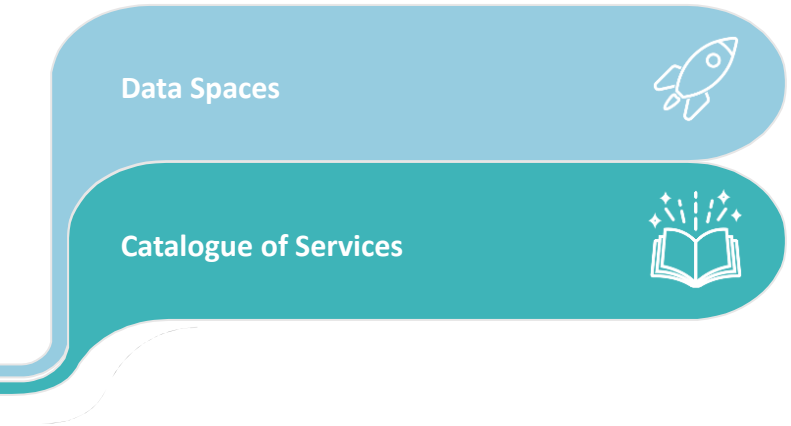
Learning materials

Some of the training developed are particularly relevant to data spaces:

- DCAT-AP: Introduction, Basic user, Advanced user, DCAT-AP for Data Spaces
- Publishing data with Linked Data Event Streams: why and how

Catalogue of Services

The Catalogue of Services Action supports public administrations in building their digital catalogues of public services.



A comprehensive list of services offered by a public administration

- ✓ Obtaining a passport or driver's license
- ✓ Applying for a loan or tax relief
- ✓ Getting a vaccination

Enabling citizens and businesses to find and access services

- ✓ Information on the name of the service, description, required documents.

CPSV-AP

CPSV-AP is a simplified, reusable and extensible data model that captures the fundamental characteristics of a public service.

Tools

SEMIC develops and proposes tools and guidelines to help you cataloguing your public services

Learning materials

Some of the training developed are particularly relevant to Catalogue of Services.

- Introduction to CPSV-AP

Linked Data Event Stream

- **Linked Data Event Streams (LDES)** apply the linked data concept to data event streams.
- A data stream is typically a constant flow of distinct data points, each containing information about an event or change of state that originates from a system that continuously creates data.
- Some examples of data streams include sensor and other IoT data, financial data, and so forth.
- A Linked Data Event Stream is a collection of immutable objects (such as version objects, sensor observations or archived representations). Each object/member is described in RDF.
- The Linked Data Event Stream (LDES) specification (ldes:EventStream) allows data publishers to publish their dataset as an append-only collection of immutable members in its most basic form. Consumers can host one or more in-sync views on top of the default (append-only) view.

Combining data across data spaces

As a researcher I want to find the impact of road related emissions on work absence…

Findability of Catalogues

If there are three different portals…
How does the researcher find them?

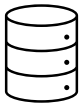
Understanding dataset descriptions

discoverability of datasets to maximise their reuse

If portals could share the metadata description…
Is the metadata compatible?

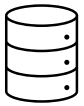
Example of Data Space Solutions

Agency for maintenance of roads



Publishes the road network in the area A.

Agency of environment



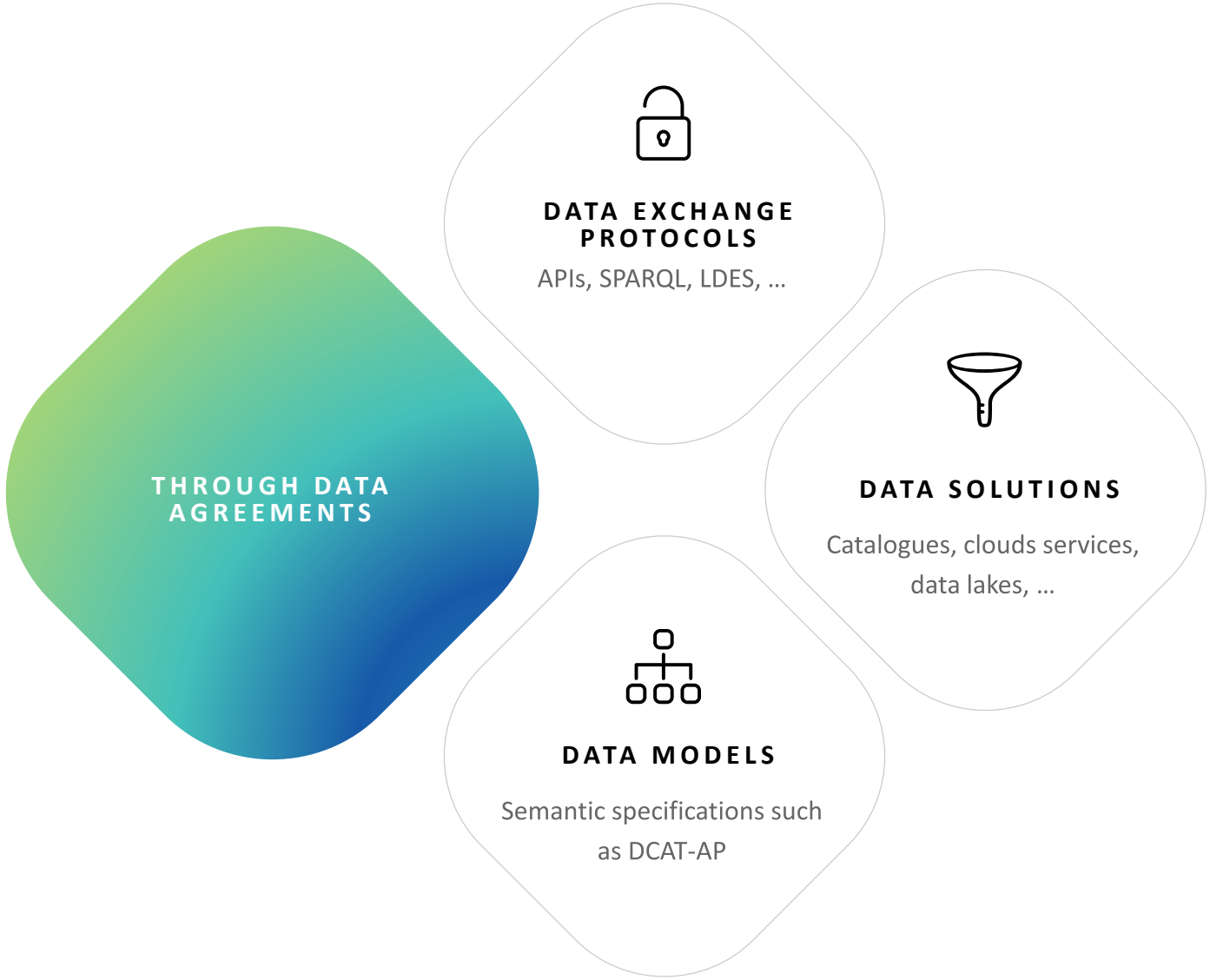
Publishes the emission of small dust particles in the area A.

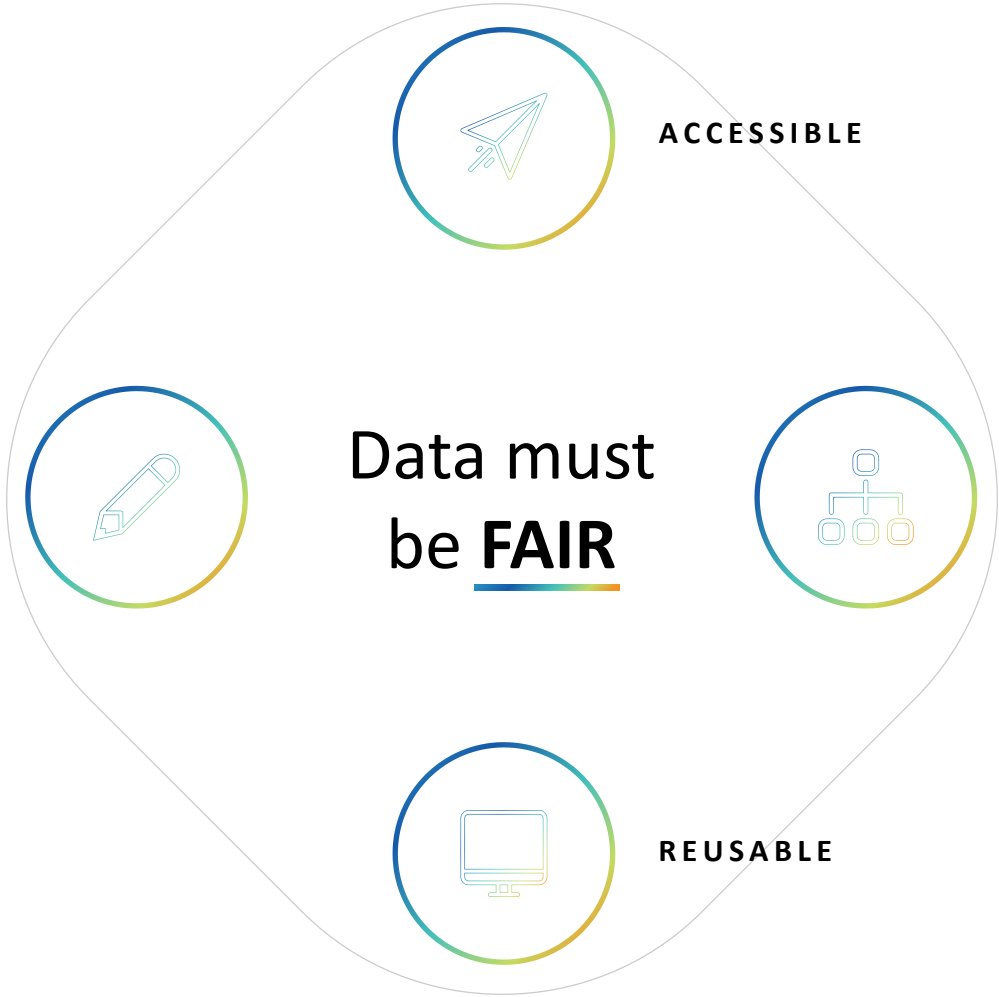
Agency of health

Publishes the absence due to illness in the area A.

HOW?

How to make data **more "*accessible*"**?





European Commission

DCAT-AP as a solution

An EU-wide specification that enables interoperability across data spaces through

A common set of constraints

Description of metadata of datasets in a unified manner

Possibility to go beyond the common specification and create domain-specific data space extensions in a stackable way

Resulting in:

Stable and long-term management of metadata and data

Reduced amount of metadata management

SEMIC Specification: DCAT-AP

- Provides a common specification for describing public sector datasets in Europe to enable the exchange of descriptions of datasets among data portals.
- Support the discovery of/access to (open) data in a cross-border and cross-domain environment

by **describing metadata to be harvested** across a distributed network of portals and capturing the requirements denoting for example **high-value datasets** (HVDs).

In the form of an application profile of W3C DCAT, by

- expressing constraints and usages on DCAT properties and classes, and
- including additional properties and usages of controlled vocabularies

Domains of application

Open data portals
with an extension for statistics and geospatial data.

Base registries
metadata descriptions

BregDCAT-AP,
GeoDCAT-AP,

StatDCAT-AP

Data spaces

- NAPCORE-Mobility
- HealthDCAT-AP
- Language DCAT-AP

High Value Datasets,

DCAT-AP for HVDs

Benefits of DCAT-AP

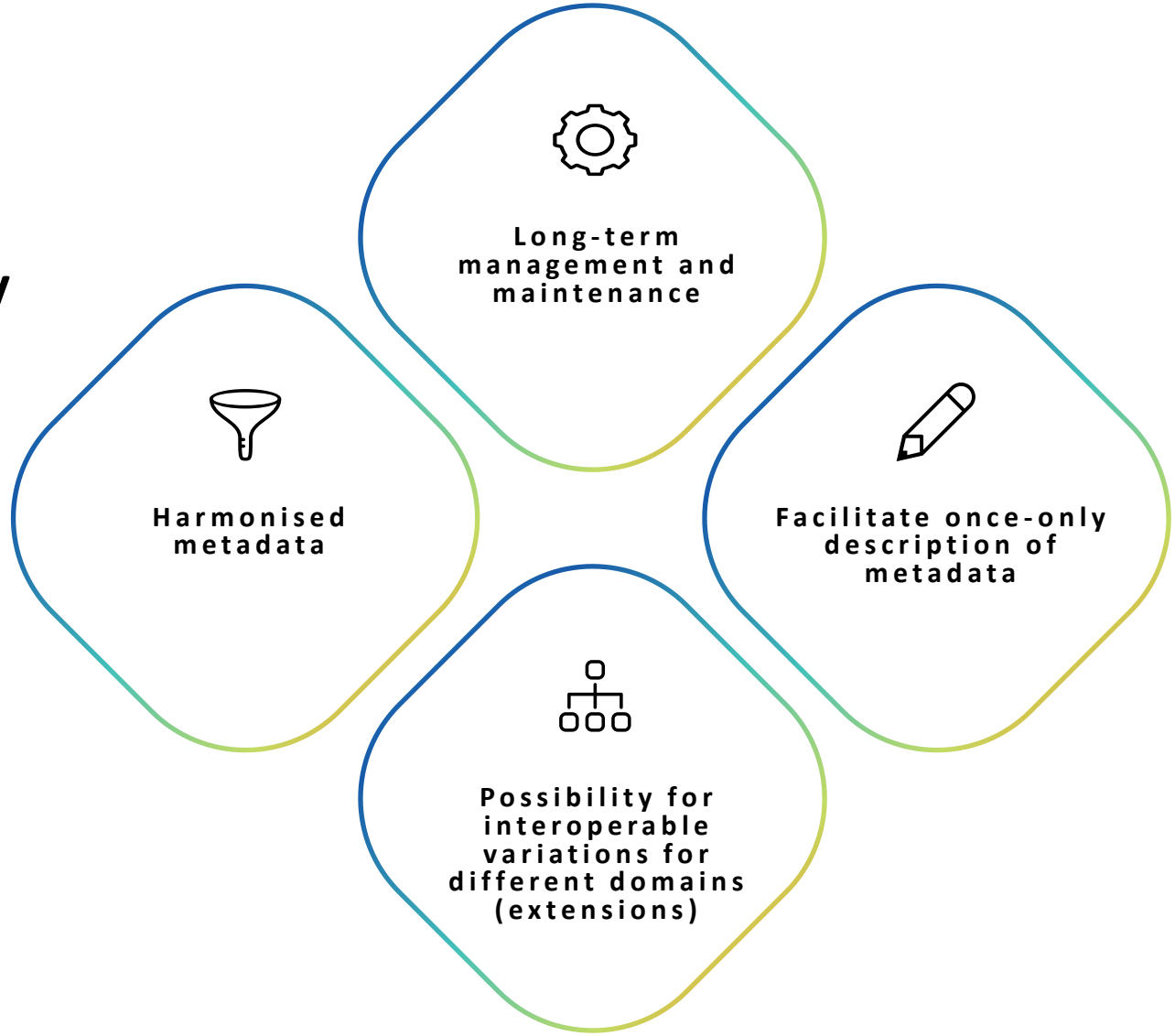
Strategic benefits

- Aligned with **best practices**
- **Higher efficiency** in the development of your own semantic assets
- Possibility to make national or domain-specific **extensions**
- Supports **intra-domain and cross-domain** interoperability
- Create **common understanding** of core elements

Technical benefits

- Fully aligned with key **semantic technologies** (semantic web, linked data, …)
- Expressions in **RDF** and **JSON-LD**
- Validation tools based on **SHACL**
- **Automated data processing**

Interoperability for Dataset sharing

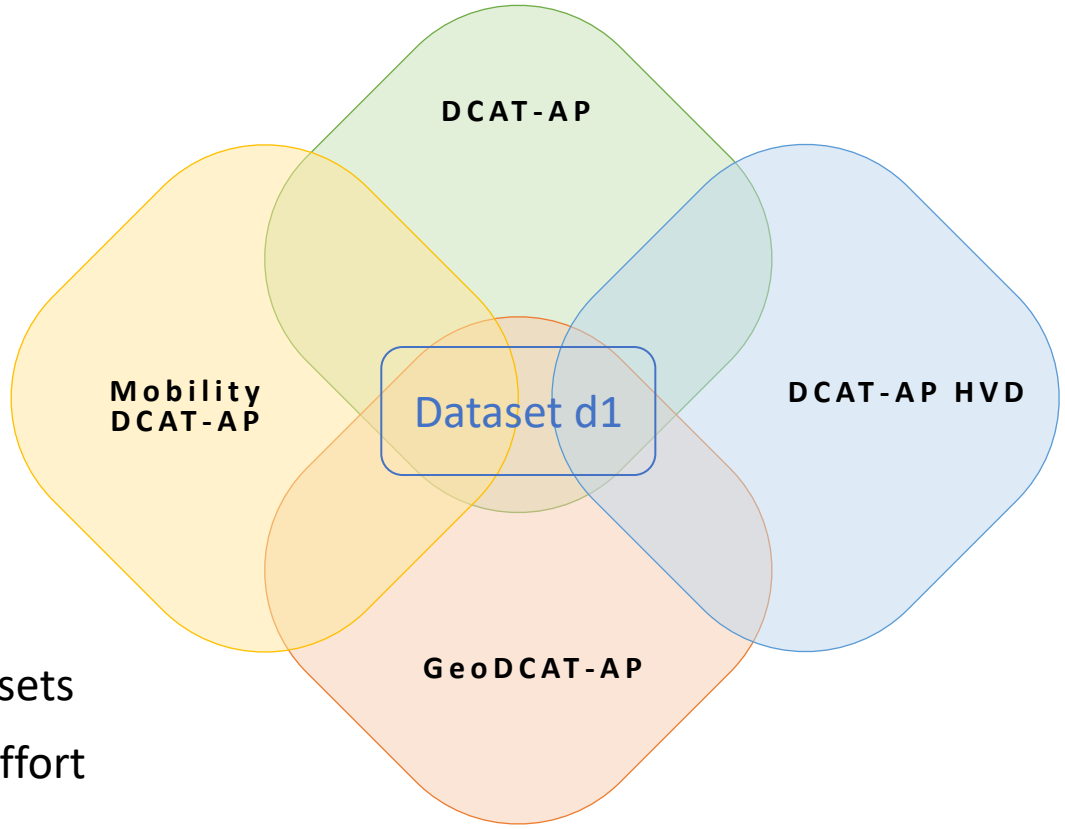


Multi-domain collaboration

Interoperable Profiles

- Concise
  - easiness to read, editorial effort
- Once-only effort for publishers of datasets

resulting in acceptable implementation effort



DCAT-AP a growing ecosystem

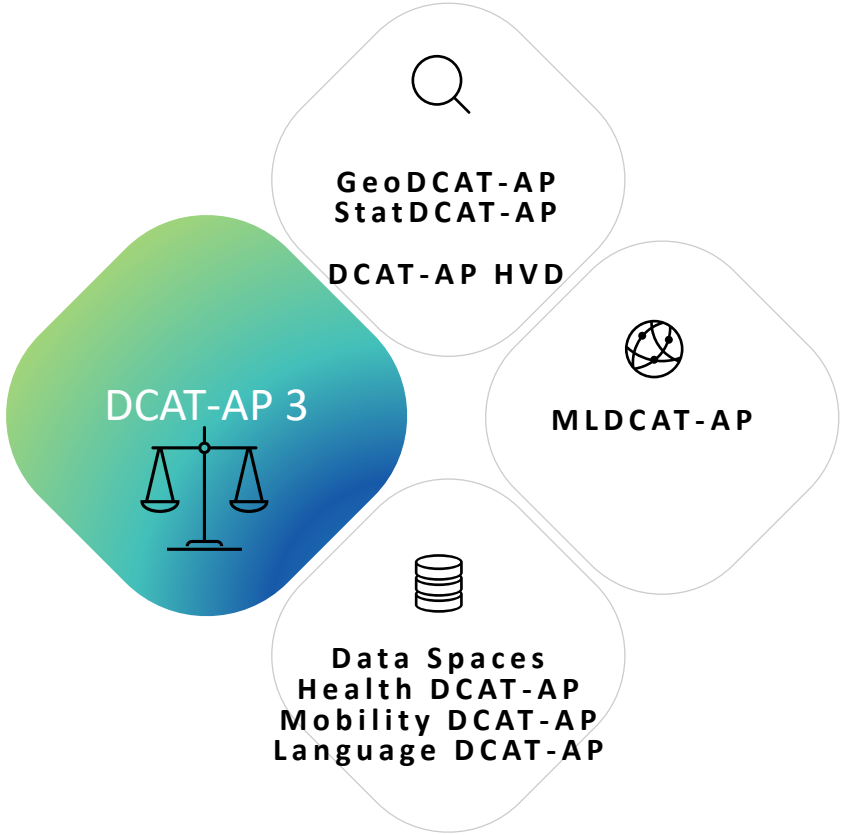
Supporting the discovery of/access to (open) data in a cross-border and cross-domain environment, by describing metadata to be harvested across a distributed network of portals.

Open community: webinars

Interconnecting communities

Support for profiling

Support for publishers and portals



COMMUNITY

- Vibrant DCAT-AP community
- GitHub DCAT-AP
- Joinup

MATURE IMPLEMENTATIONS

Such as

data.europa.eu

https://joinup.ec.europa.eu/collection/semic-support-centre

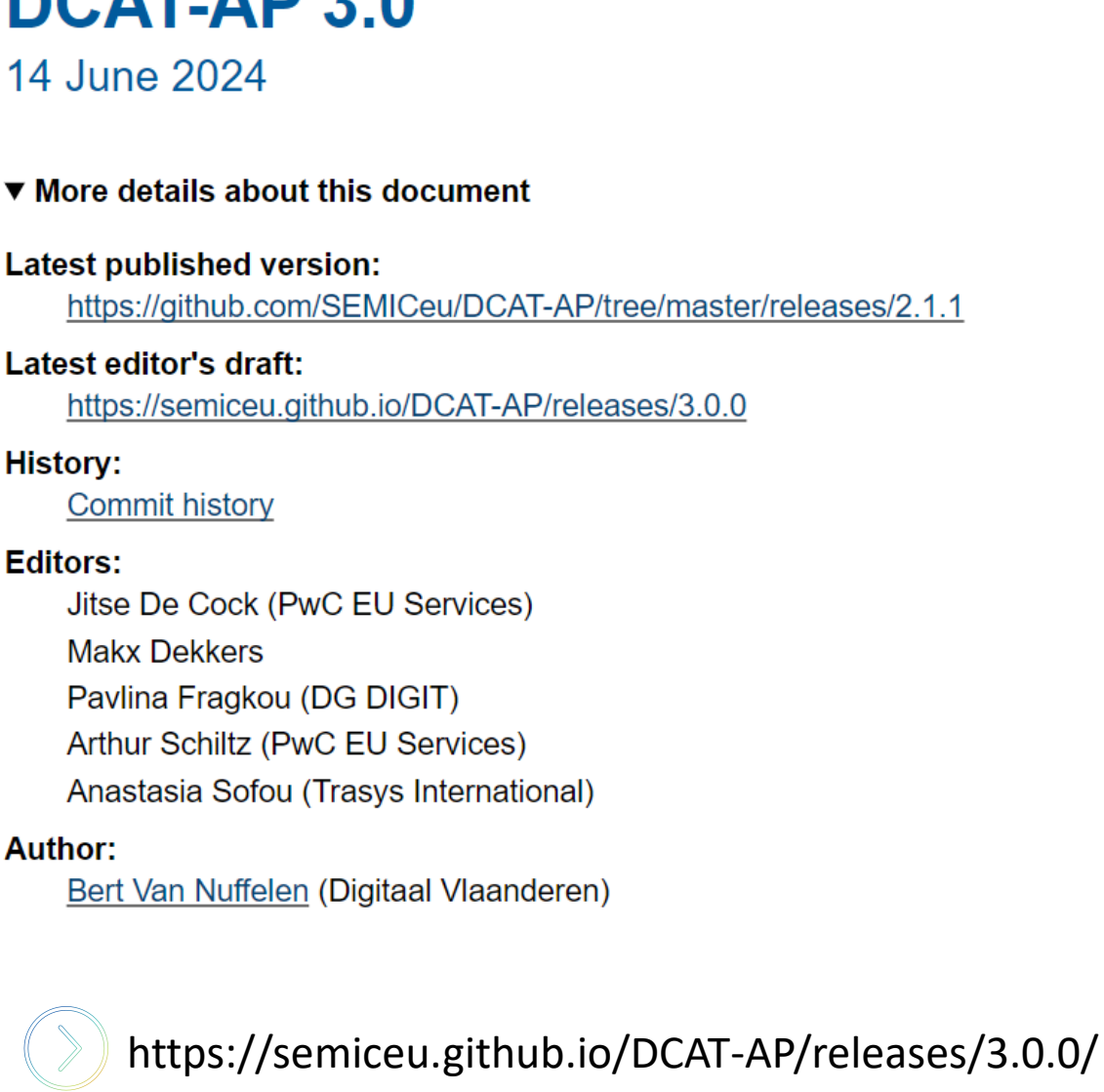
DCAT-AP 3 status

DCAT-AP 3 was released in June 2024

- Aligned with W3C DCAT 3
- Introduces Dataset Series
- Major editorial update
- Alignment with SEMIC style guide

GeoDCAT-AP 3 in active revision

- https://joinup.ec.europa.eu/collection/semic-support-centre/news/upcoming-events-may-2024
- https://github.com/SEMICeu/GeoDCAT-AP



DCAT-AP 3.0

14 June 2024

▼ **More details about this document**

Latest published version:
https://github.com/SEMICeu/DCAT-AP/tree/master/releases/2.1.1

Latest editor's draft:
https://semiceu.github.io/DCAT-AP/releases/3.0.0

History:
Commit history

Editors:
Jitse De Cock (PwC EU Services)
Makx Dekkers
Pavlina Fragkou (DG DIGIT)
Arthur Schiltz (PwC EU Services)
Anastasia Sofou (Trasys International)

Author:
Bert Van Nuffelen (Digitaal Vlaanderen)

https://semiceu.github.io/DCAT-AP/releases/3.0.0/

SEMIC Style Guide

Question?

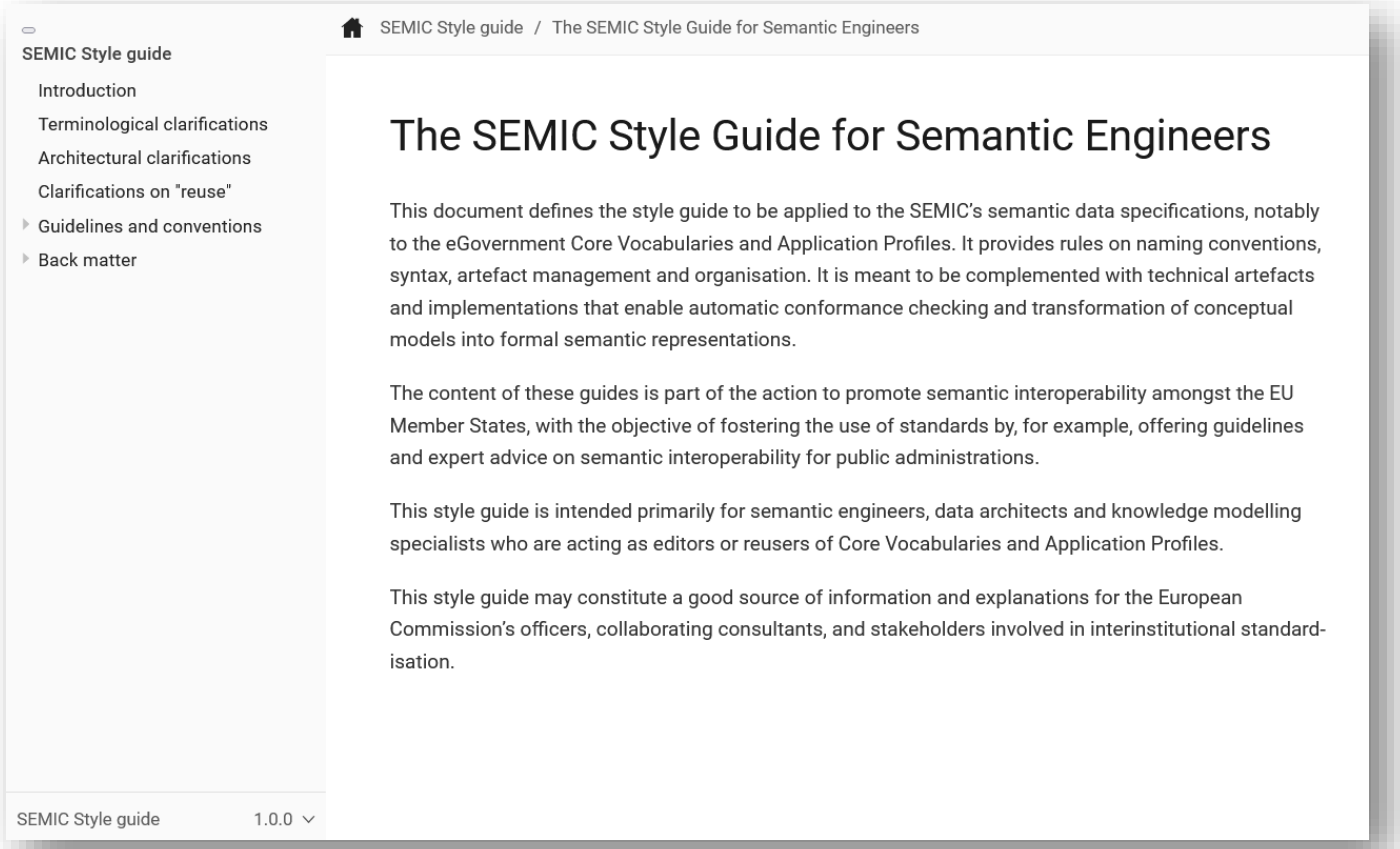
What are the rules to create good, interoperable data specifications.

Reuse types

- as-is
- with terminological changes
- with semantical adaptions

Coherency & editorial effort

- internal coherent
  - all artifacts express the same knowledge
- external reusable
  - artifacts ready for reuse



https://semiceu.github.io/style-guide/1.0.0/index.html

Interoperability test bed

The Interoperability Test Bed is a service to facilitate conformance testing of IT systems. The Test Bed is a software system that can be both downloaded and installed locally, but also reused through a shared online installation operated by DIGIT.

How does it work?

Test Bed is a complete platform consisting of both software and hardware components to facilitate testing.

When to use it?

The particular focus is conformance and interoperability testing, ensuring that tested systems conform to a specification's requirements and can interoperate consistently with conformant peer systems.

Validators

The Test Bed also has standalone validation services, operated independently from the Test Bed, which focus specifically on receiving content via various channels and validating it against specifications to produce reports.

Case studies

Data.Europa.eu

Data.Europa.eu offers uniform, central access to open data from public authorities. The data published on data.Europa.eu is harvested from various data providers who are expected to make their datasets available in DCAT-AP. As a means of verifying the quality of provided data and reporting issues, data.Europa.eu is using the DCAT-AP validator, a validator instance hosted by the Test Bed.

CEF eInvoicing Building Block

The validation of invoices against the European standard's core invoice is the central objective of the eInvoicing conformance testing service. Given that the supported syntaxes are XML-based, this was realised by means of the Test Bed's core XML validation service that was used to create a new validator supporting the validation of invoices and credit notes for each syntax.

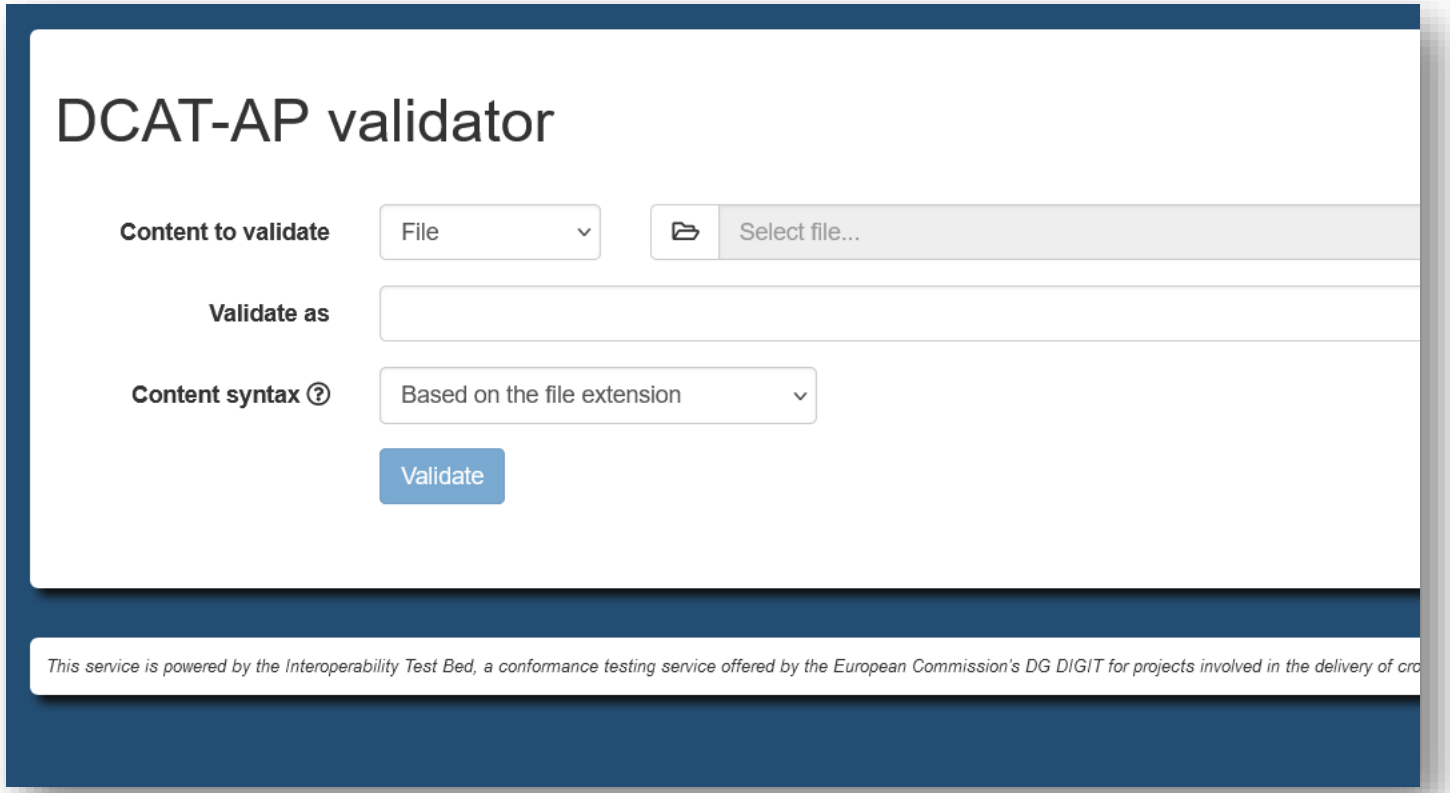
DCAT-AP validator

Question?

How can publishers validate their efforts?

DCAT-AP provides SHACL shapes which are activated in the DCAT-AP validator.

Based Interoperability Test Bed (https://www.itb.ec.europa.eu/)



https://www.itb.ec.europa.eu/shacl/dcat-ap/upload

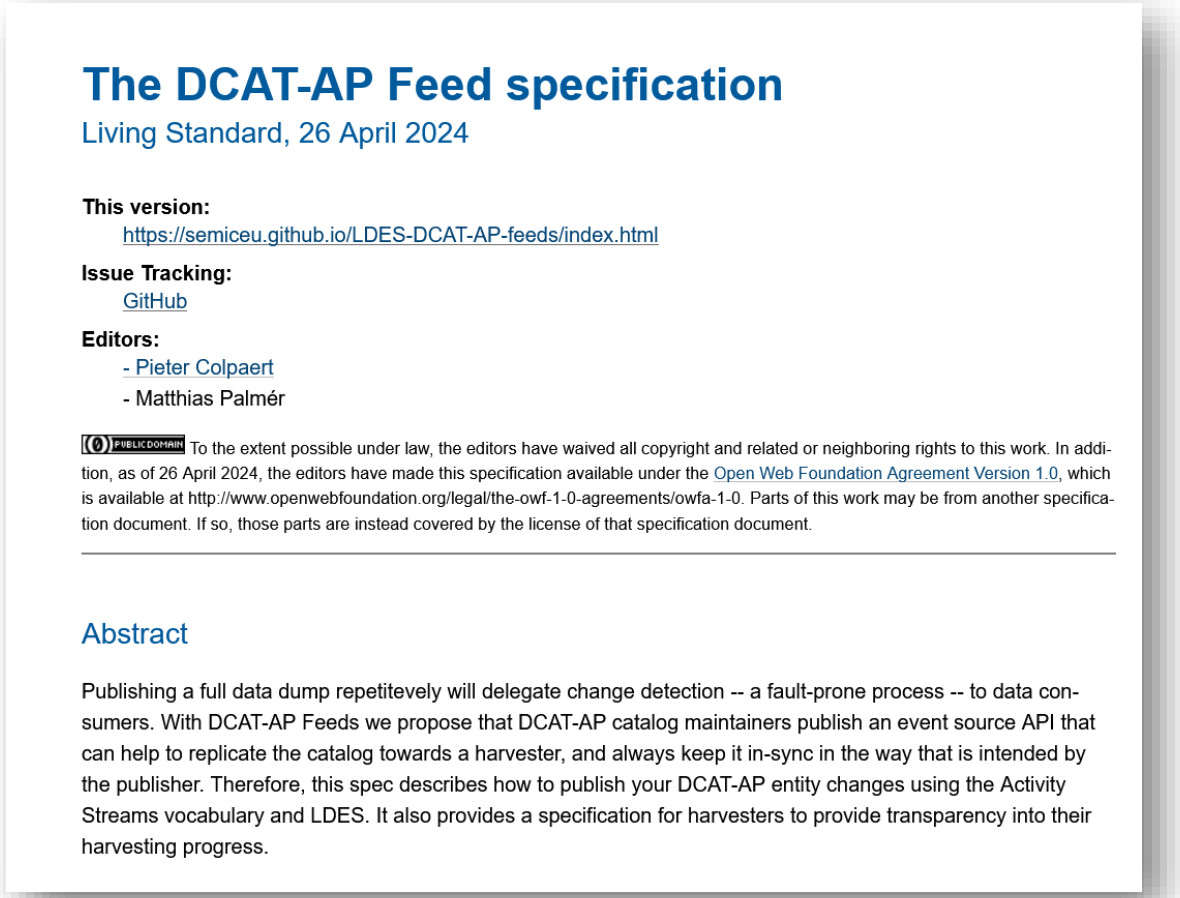
SEMIC DCAT-AP feeds specification

Question?

How to share DCAT-AP in a harvesting network?

Linked Data Event Streams (LDES)

- https://semiceu.github.io/LinkedDataEventStreams/
- An protocol to share a collection of immutable objects in Linked Data native format
- LDES apply the linked data concept to data event streams. A data stream is typically a constant flow of distinct data points, each containing information about an event or change of state that originates from a system that continuously creates data.



The DCAT-AP Feed specification

Living Standard, 26 April 2024

This version:
https://semiceu.github.io/LDES-DCAT-AP-feeds/index.html

Issue Tracking:
GitHub

Editors:
- Pieter Colpaert
- Matthias Palmér

Abstract

Publishing a full data dump repetitevely will delegate change detection -- a fault-prone process -- to data consumers. With DCAT-AP Feeds we propose that DCAT-AP catalog maintainers publish an event source API that can help to replicate the catalog towards a harvester, and always keep it in-sync in the way that is intended by the publisher. Therefore, this spec describes how to publish your DCAT-AP entity changes using the Activity Streams vocabulary and LDES. It also provides a specification for harvesters to provide transparency into their harvesting progress.

https://semiceu.github.io/LDES-DCAT-AP-feeds/index.html

Webinar 30 Apr 2024 : https://joinup.ec.europa.eu/collection/semic-support-centre/event/ldes-webinar

SEMIC Levels of Service Offering

- Provide documentation
- SEMIC Specification

- Provide documentation
- SEMIC specification
- Review your specification

- Provide documentation
- SEMIC specification
- Review your specification
- Participate in development of specification

- Provide documentation
- SEMIC specification
- Review your specification
- Co-organise development of the specification

DCAT-AP Profiles

A growing ecosystem

2012

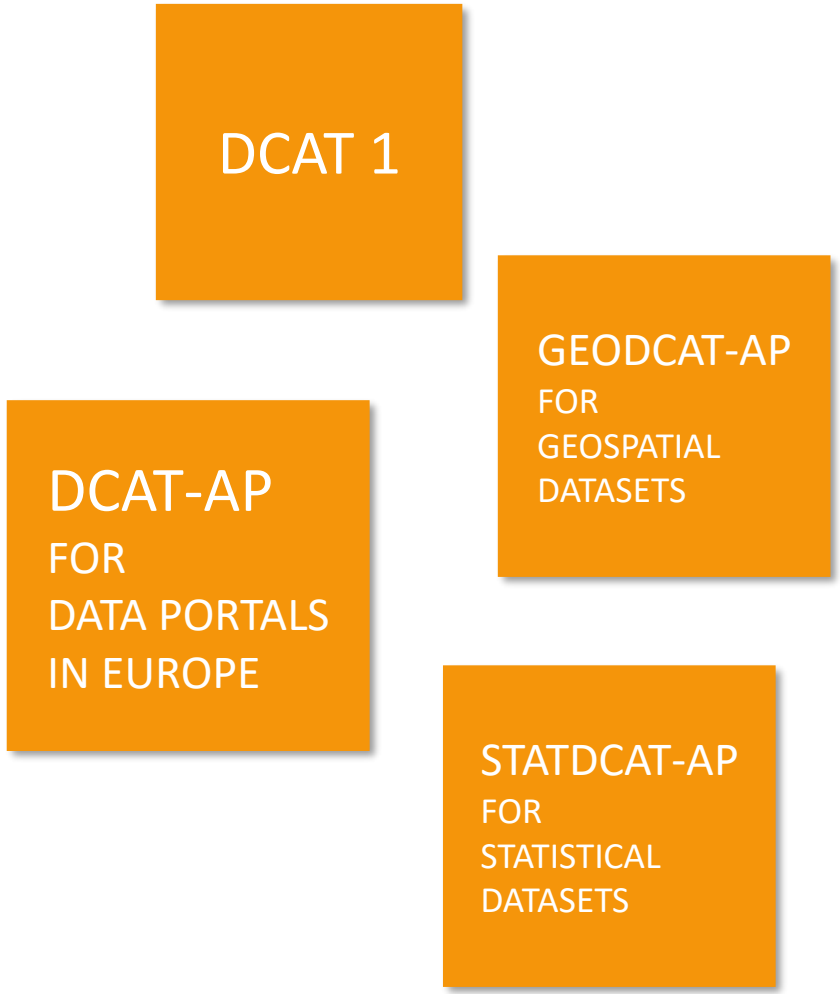
DCAT 1

DCAT-AP
FOR
DATA PORTALS
IN EUROPE

DCAT-AP Profiles

A growing ecosystem

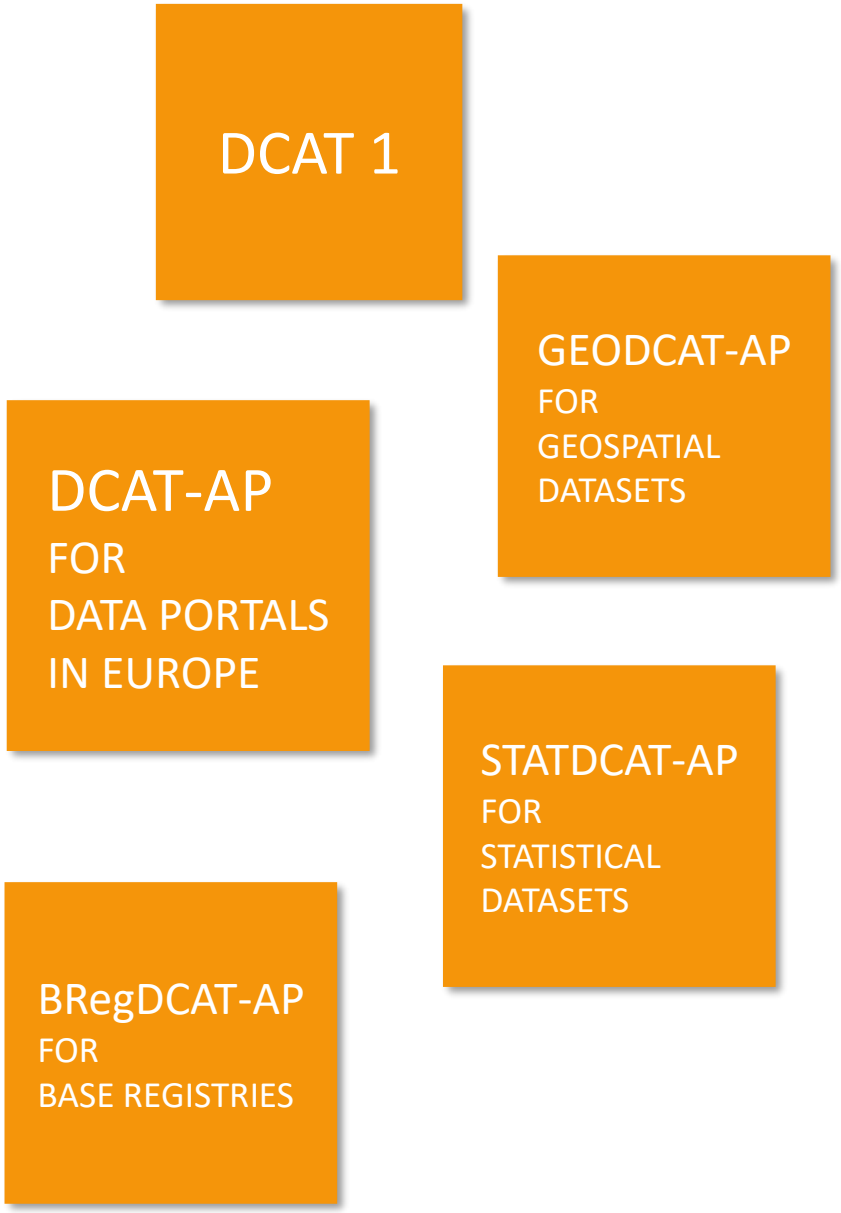
2015



DCAT-AP Profiles

A growing ecosystem

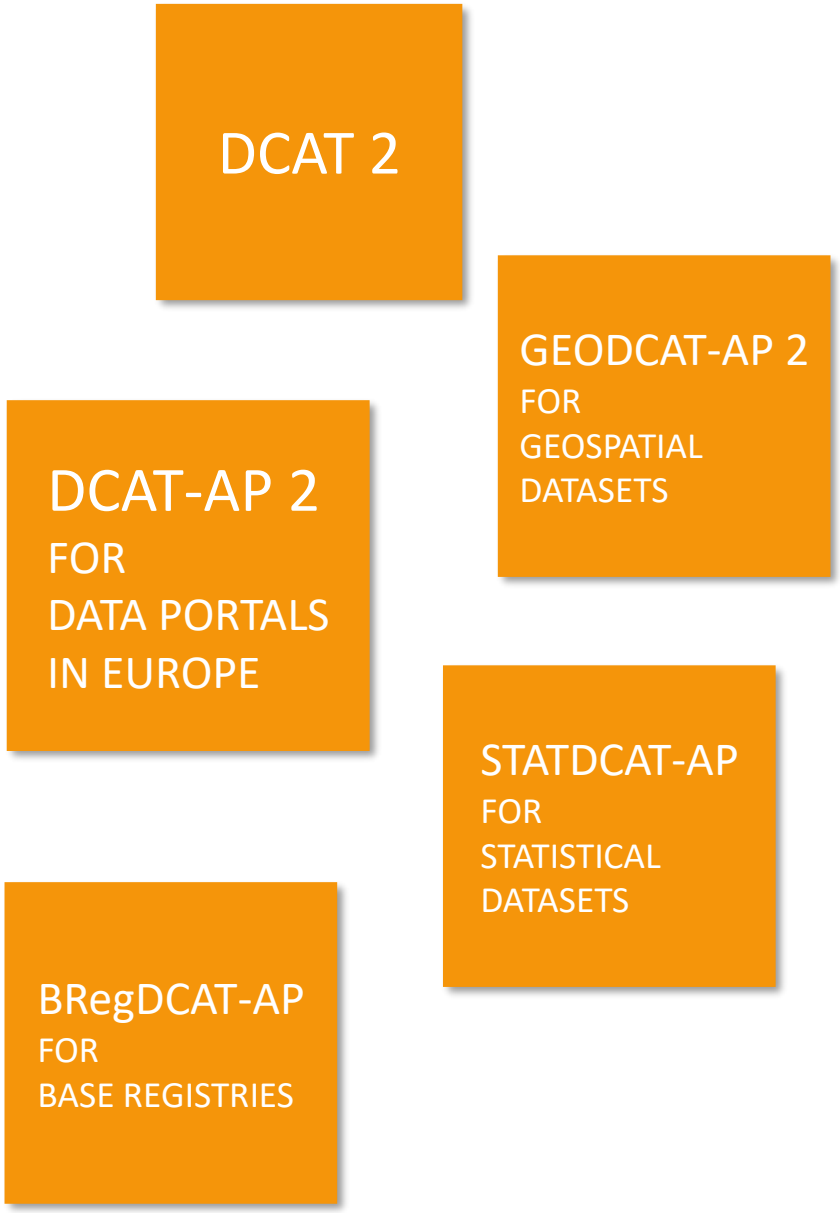
2018



DCAT-AP Profiles

A growing ecosystem

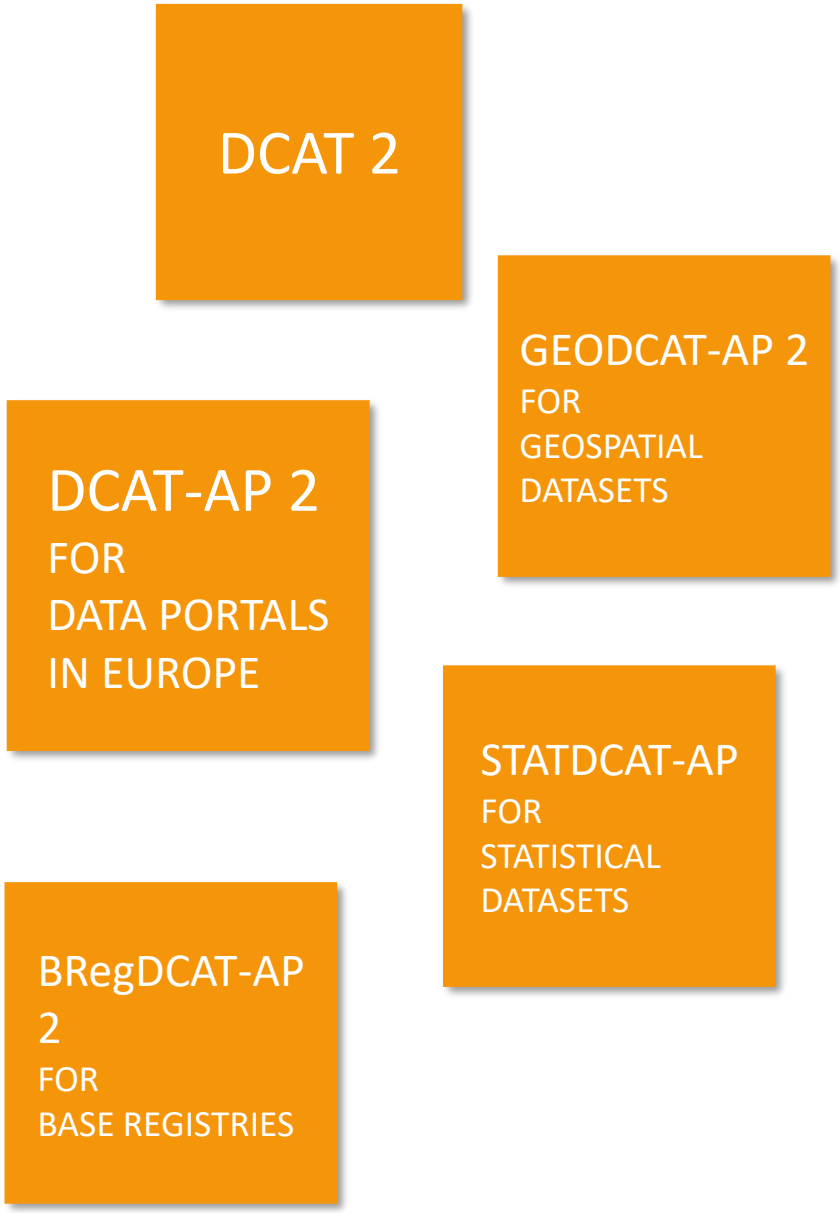
2020



DCAT-AP Profiles

A growing ecosystem

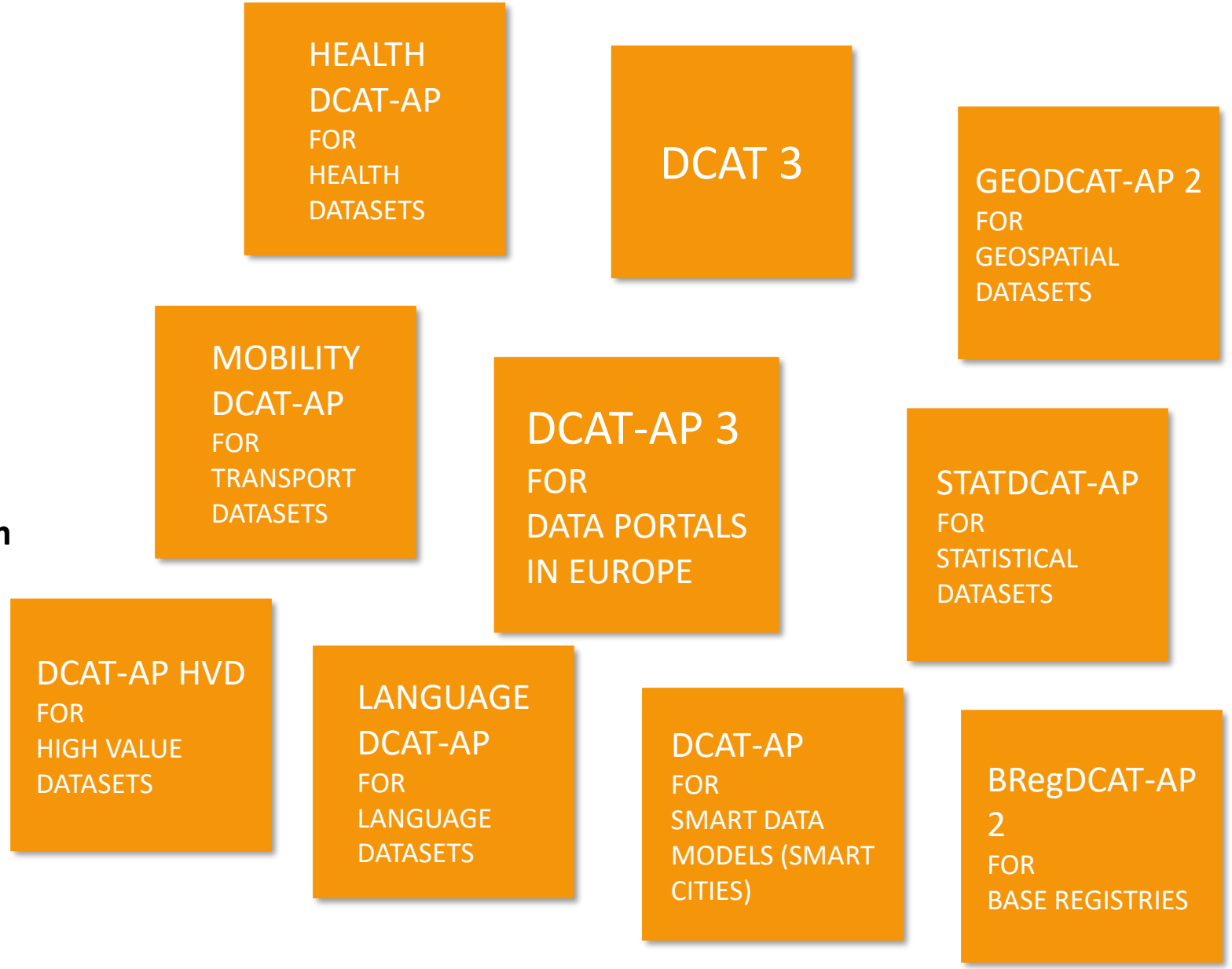
2021



DCAT-AP Profiles

A growing ecosystem

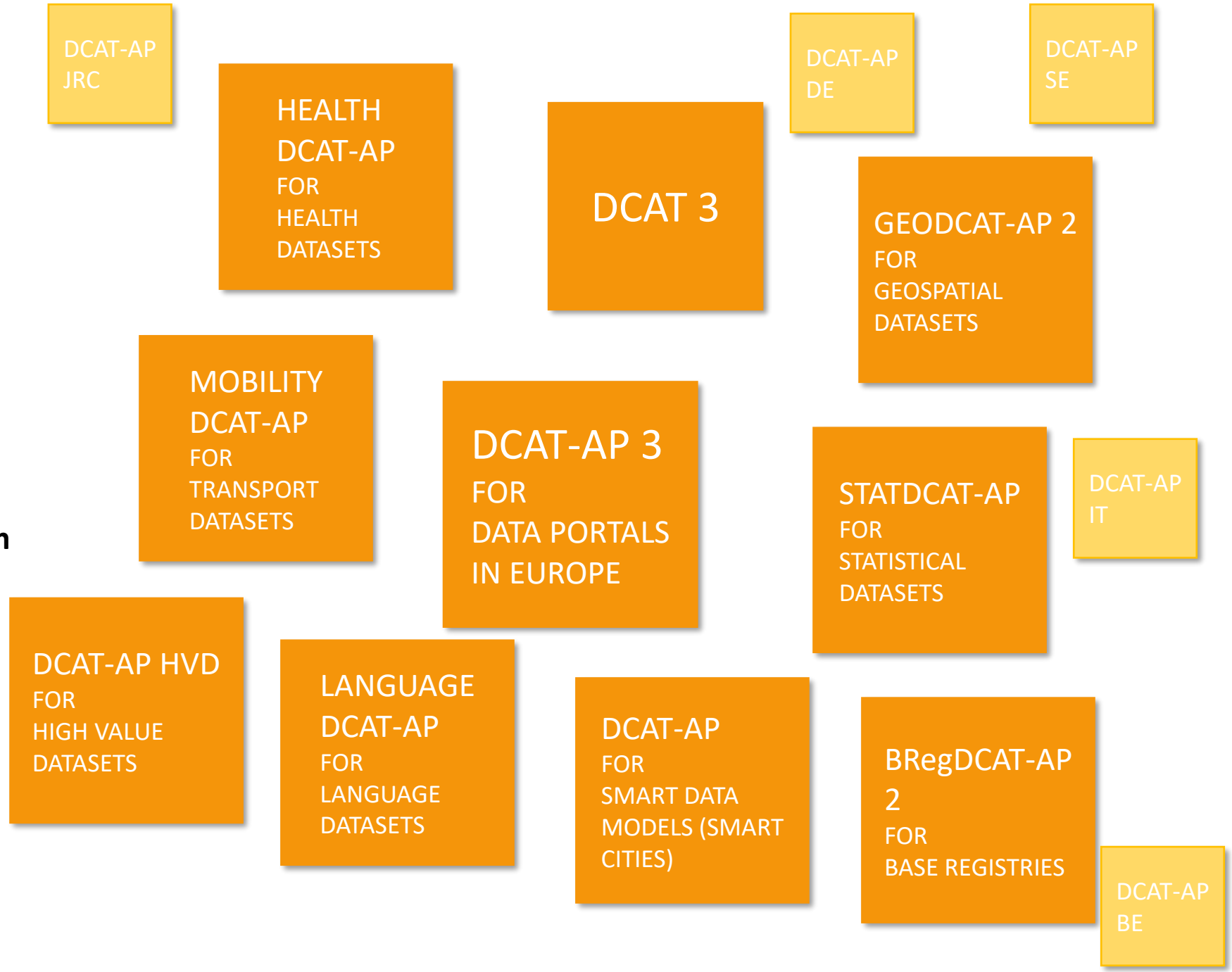
2023



DCAT-AP Profiles

A growing ecosystem

2023



GeoDCAT-AP

GEODCAT-AP
FOR
GEOSPATIAL
DATASETS

Basic use case: Make spatial datasets (services, series) **searchable on general data portals**.
→ **improve findability** across borders and sectors.

Spatial data natively described by
- core profile of ISO 19115:2003 and ISO 19119
- INSPIRE metadata

General data portals use
- W3C DCAT
- DCAT-AP in Europe

GeoDCAT-AP provides an alternative representation of INSPIRE metadata in 2 levels
- GeoDCAT-AP **Core**: Mapping to existing DCAT-AP terms
- GeoDCAT-AP **Extended**: Extensions for INSPIRE metadata with no direct DCAT-AP counterpart

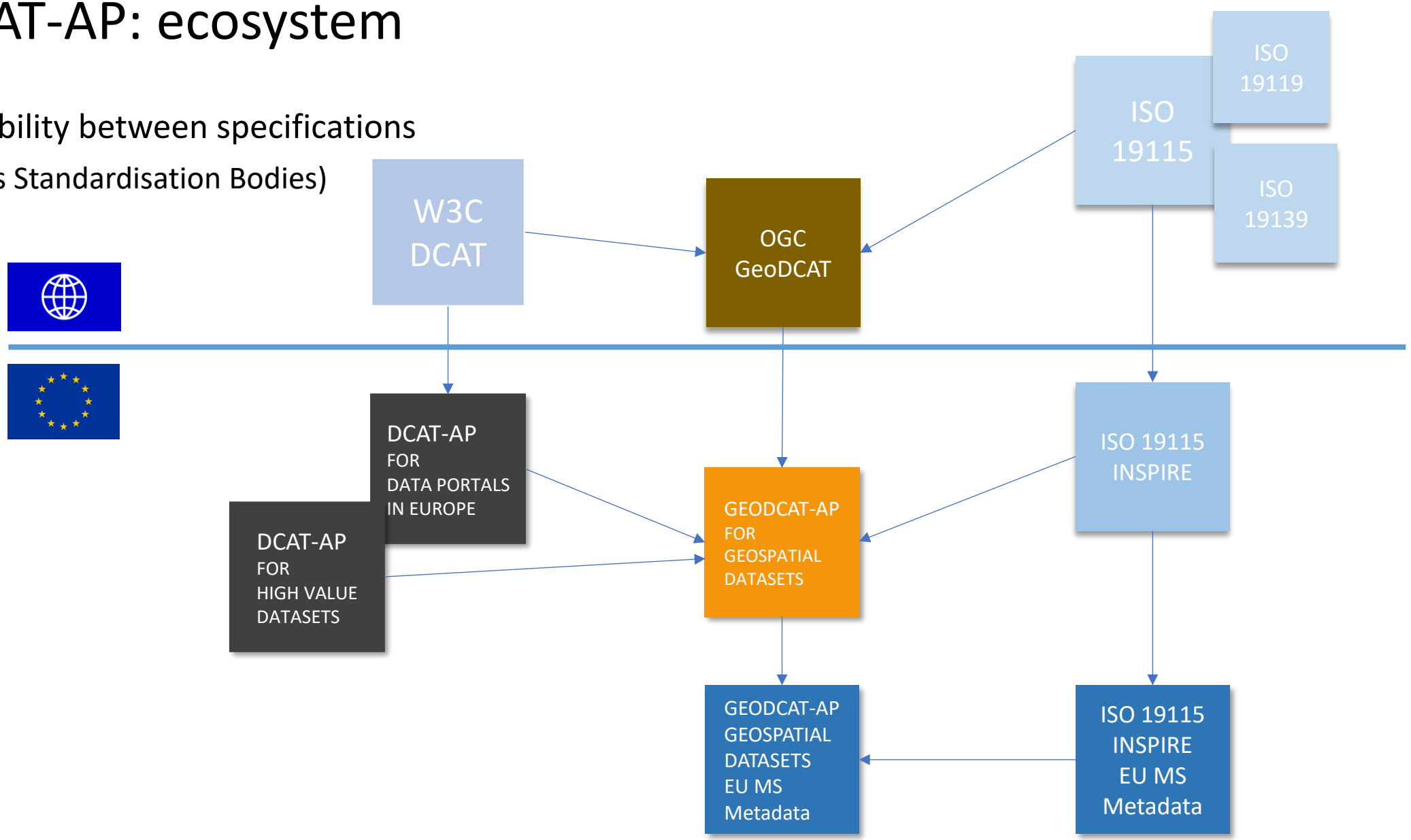
Current status: GeoDCAT-AP 2.0.0 (**2020**)
aligned with W3C DCAT 2, DCAT-AP 2.0.1, INSPIRE Metadata Technical Guidelines 2.0.1
Upcoming: GeoDCAT-AP 3.0.0 (**2024**)
alignment with W3C DCAT 3, DCAT-AP 3.0, DCAT-AP HVD, INSPIRE Metadata Technical Guidelines 2.2.0

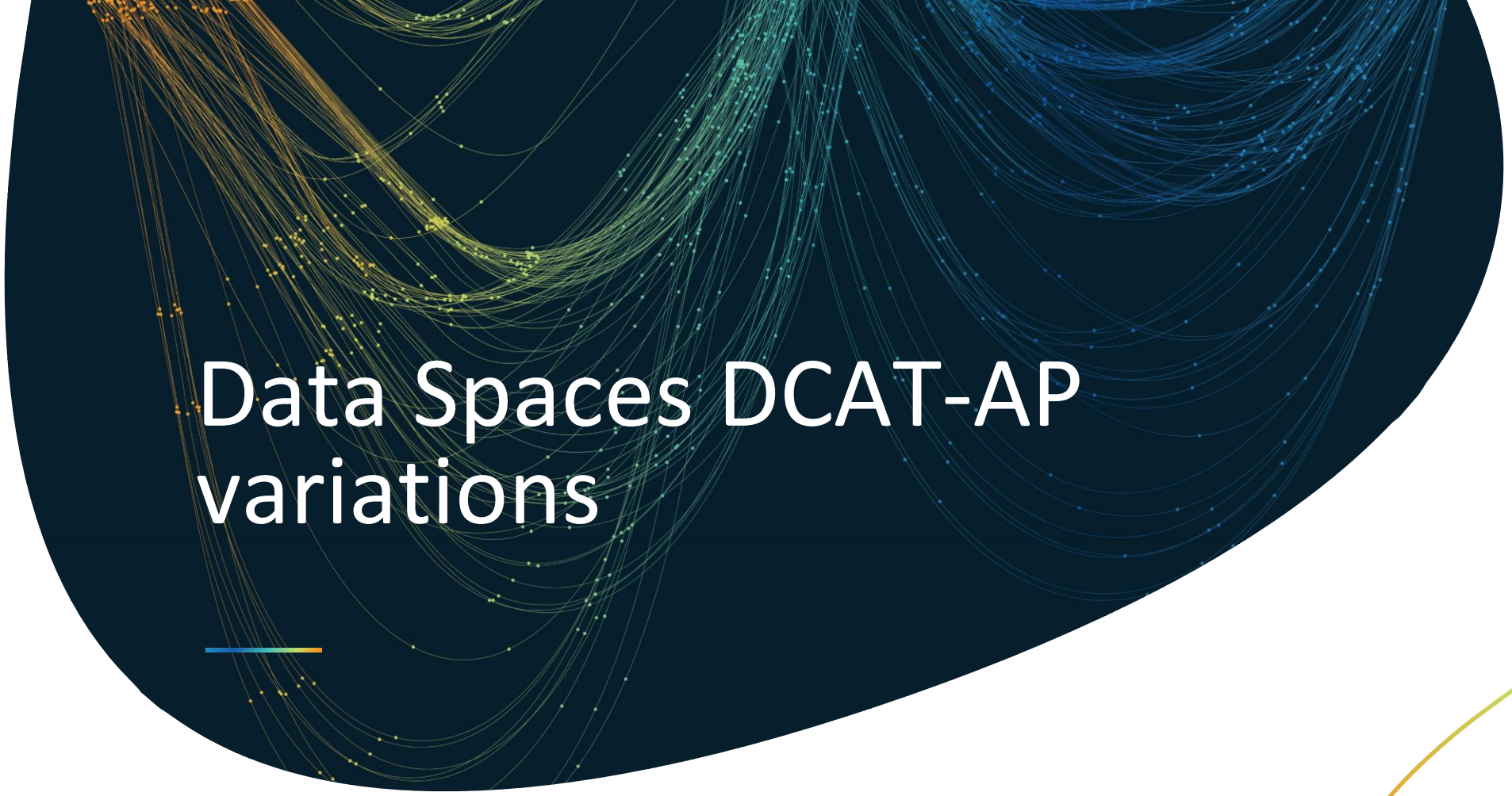
https://semiceu.github.io/GeoDCAT-AP/releases/2.0.0/

GeoDCAT-AP: ecosystem

Interoperability between specifications

(even across Standardisation Bodies)





Data Spaces DCAT-AP variations

BLE OF CONTENTS

mobilityDCAT-AP - Version 1.0.1

A mobility extension for the DCAT application profile for data portals in Europe

ReSpec

mobilitydcat-ap

NAPCORE Editor's Draft 13 March 2024

This version:
https://w3id.org/mobilitydcat-ap/drafts/latest/

Latest published version:
https://w3id.org/mobilitydcat-ap/releases/

Latest editor's draft:
https://w3id.org/mobilitydcat-ap/drafts/latest/

Latest Recommendation:
https://w3id.org/mobilitydcat-ap/releases/1.0.0/

Editors:

Lina Molinas Comet (Fraunhofer Institute for Applied Information Technology FIT)
Peter Lubrich (Federal Highway Research Institute (BASt))
Mario Scrocca (Cefriel)

Extension of DCAT-AP: HealthDCAT-AP

Expected in early 2024

More information coming soon

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HaDEA. Neither the European Union nor the granting authority can be held responsible for them.

Domains

Smart Cities	Smart Agrifood	Smart Water	Smart Energy	Smart Logistics
Smart Robotics	Smart Sensoring	Cross sector	Smart Health	Smart Destination
Smart Environment	Smart Aeronautics	Smart Manufacturing	Incubated	Harmonization

python package pysmartdatamodels

pip install pysmartdatamodels

python

DCAT-AP catalogue service in beta versi

20/10/2023

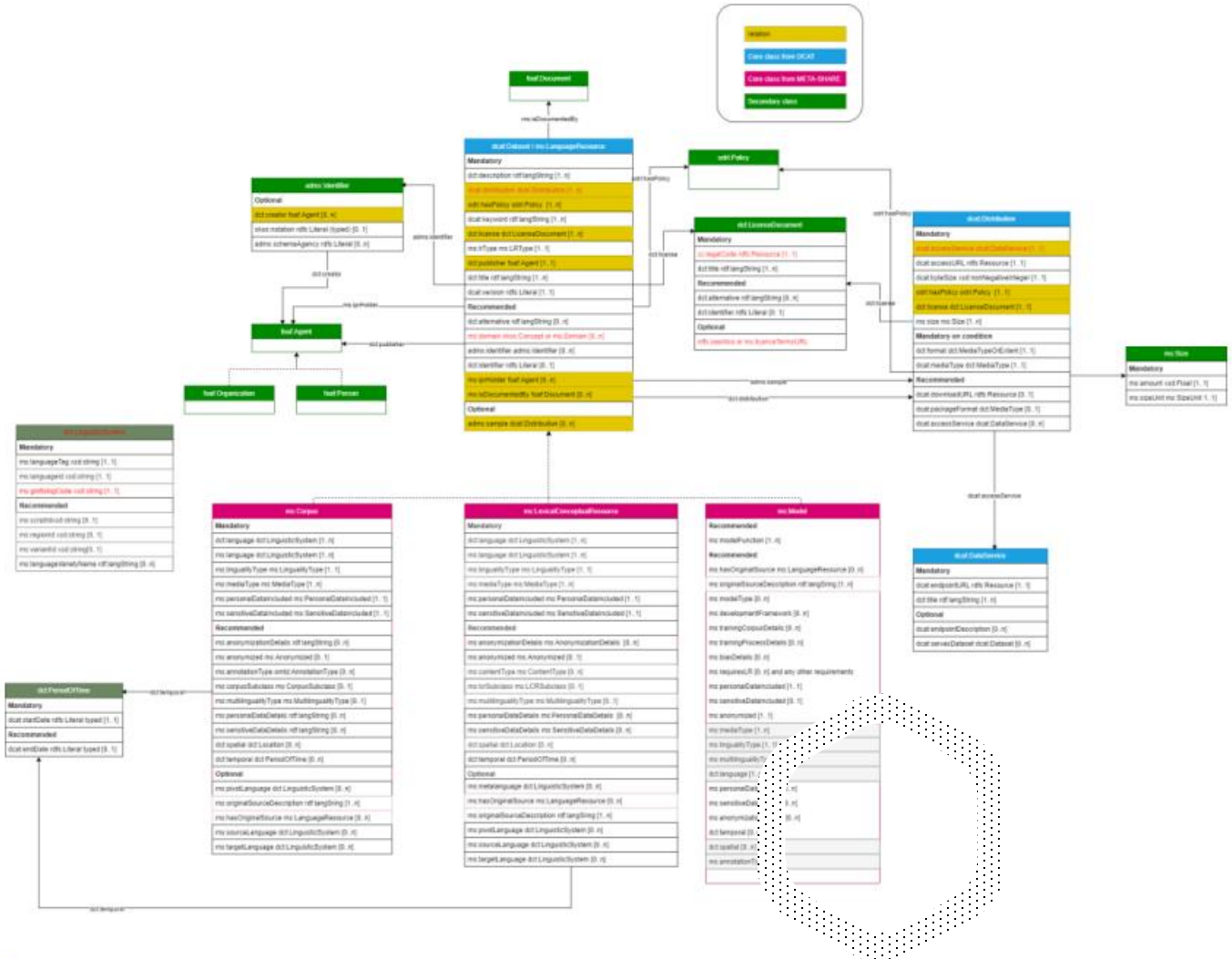
Smart Cities domain, Smart Energy domain, Smart Environment [illegible]ain, Smart Manu[illegible]turing domain, Smart Robotics domain, Smart Water Domain, Smart-Sensoring doma[illegible]mart [illegible]onautics, SmartAgrifood domain, SmartDestinations, SmartHealth, SmartLogistics

DCAT-AP is, possibly, the most relevant standard of a catalogue of datasets (even resources as well).

MasterSheet

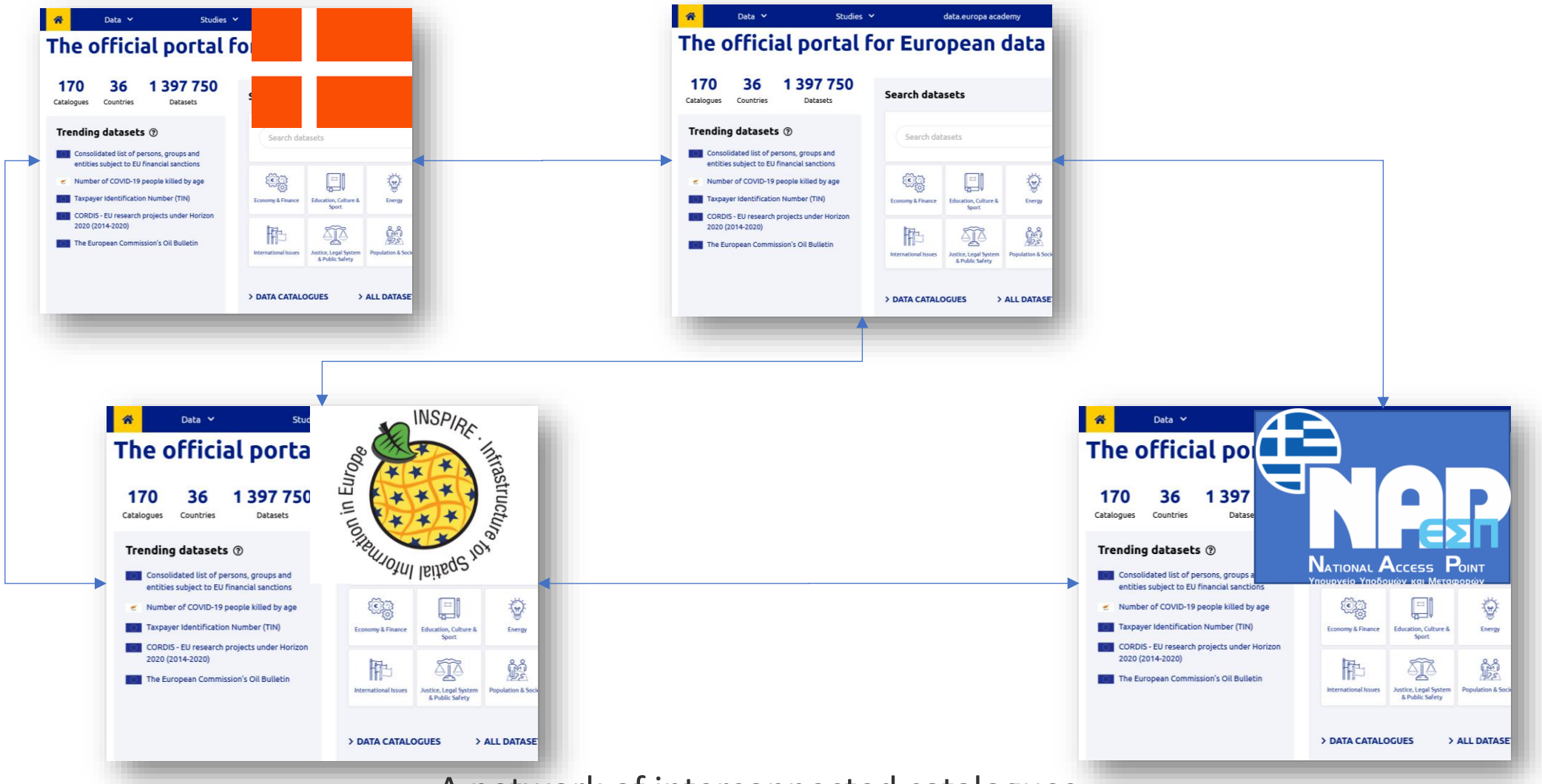
Create your simple data model with this sheet(it creates a copy)

LanguageDCAT-AP v0.7



- Currently focusing on data resources: *Corpus*, *Lexical/Conceptual resource*, *Model (language description)*
- Based on the minimal version of the ELG-SHARE model
- General properties from DCAT and DCAT-AP (with some differences on optionality and recommended controlled vocabularies)
- Language data concepts from META-SHARE ontology
- LanguageDCAT-AP v0.7 proposal: to be evaluated & approved by CELT/CELT+

Use Case Harvesting



A network of interconnected catalogues

Training Material “DCAT-AP for Data Spaces”

- **Goal**: to provide learners with a comprehensive understanding of DCAT-AP and its practical application in improving the findability and reusability of data within Data Spaces.
- No prerequisites of prior knowledge of semantics or data cataloguing solutions.
- Emphases on the importance of metadata specifications in breaking down data silos and improving data findability and reusability.
- Focuses on the practical application of DCAT-AP in the realm of Open Data by providing standardized metadata descriptions.
- Highlights the enhanced findability and accessibility of data as a significant benefit of DCAT-AP.
- Provides a comprehensive guide for setting up a DCAT-AP extension and explain the support that SEMIC can offer.

Your Data Spaces DCAT-AP specification

https://academy.europa.eu/courses/dcat-ap-for-data-spaces

Governance of DCAT-AP variations

- **Current status**: Creation of different variations of DCAT-AP each for data space

- **Challenges**:
  - how to ensure compliance of each of those with DCAT-AP i.e., interoperability between data spaces
  - How to have an overview of the different needs and escalate them to DCAT-AP?
  - How to provide guidance and compliance validation ?

- **Current attempt/proposal**:
  - Creation of a coordination group at a policy level with representatives from various DG and stakeholders
  - Establish an indirect connection with WGs of various data spaces through their policy member in the coordination group.

Forthcoming Event

Webinar: SEMIC Specifications for Data Space interoperability

Topic: how SEMIC specifications can support interoperability in Data Spaces

Date: 09/10/2024 from 10.00 until 12.00 CET.

Registration: https://ec.europa.eu/eusurvey/runner/semic-data-space-webinar-registration

European Commission

interoperable europe

innovation ∞ govtech ∞ community

Stay in touch

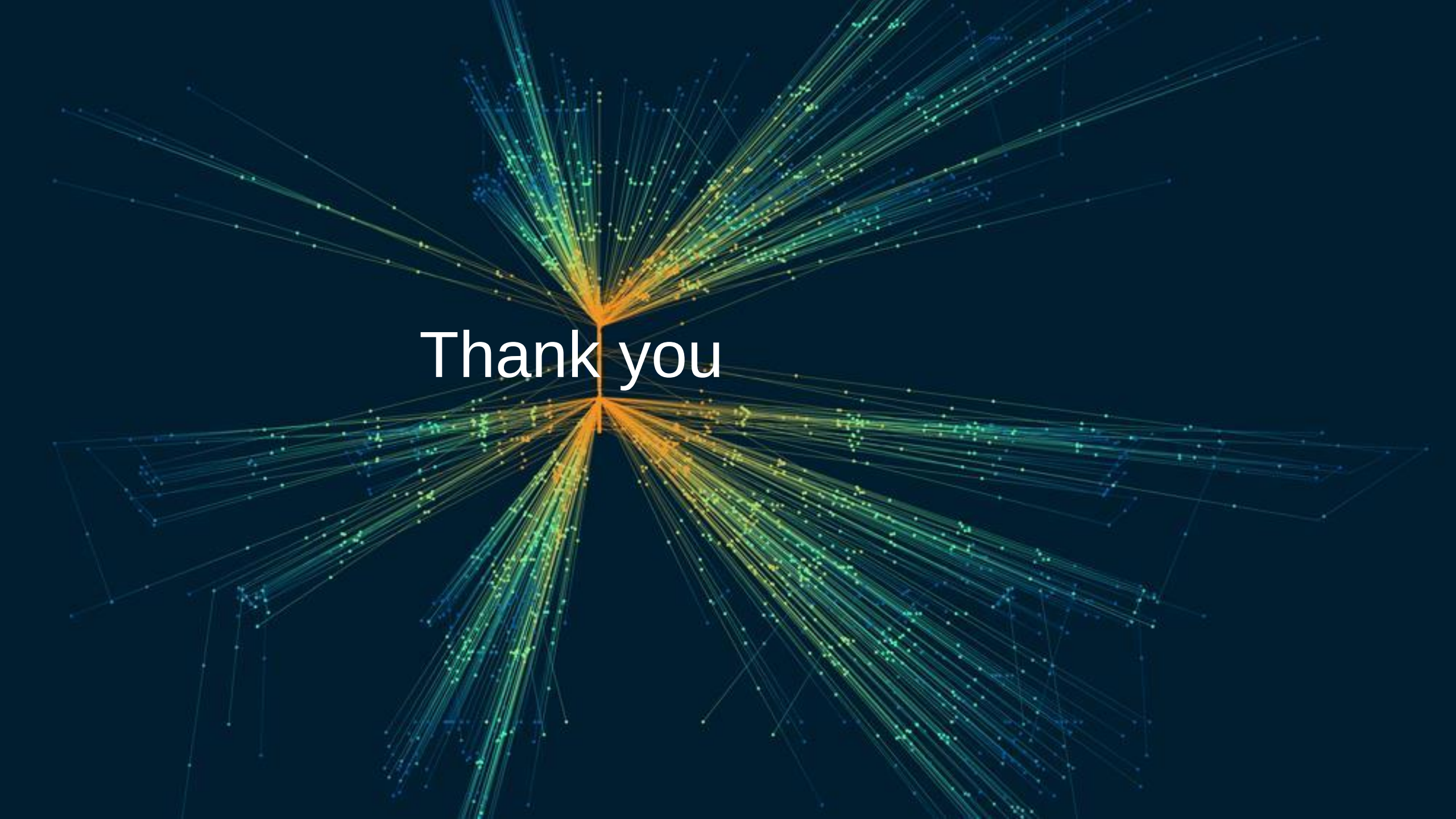
(@InteroperableEU) / Twitter

Interoperable Europe - YouTube

Interoperable Europe | LinkedIn

DIGIT-INTEROPERABILITY@ec.europa.eu

https://joinup.ec.europa.eu/collection/interoperable-europe/interoperable-europe


Thank you