WEBINAR

‘Unlocking data interoperability: exploring ways of standardisation with real use cases’

data europa academy

25 April 2025

10:00 – 11:30 CEST

Rules of the game

The webinar will be recorded and published on the data.europa academy

For questions, please use the ClickMeeting chat

Please reserve 3 min after the webinar to help us improve by filling in our feedback form

data europa academy

Agenda

10.00 – 10.05	Opening and introduction – *Inmaculada Farfan Velasco*
10.05 – 10.25	DCAT-AP: creating interoperable profiles – *Bert van Nuffelen*
10.25 – 10.45	Czechia’s approach to metadata quality – *Lenka Kováčová*
10:45 – 11:05	France’s approach to data quality – *Pierlou Ramade*
11:05 – 11:25	Panel Q&A discussion
11:25 – 11:30	Closing remarks

data europa academy

Today’s speakers

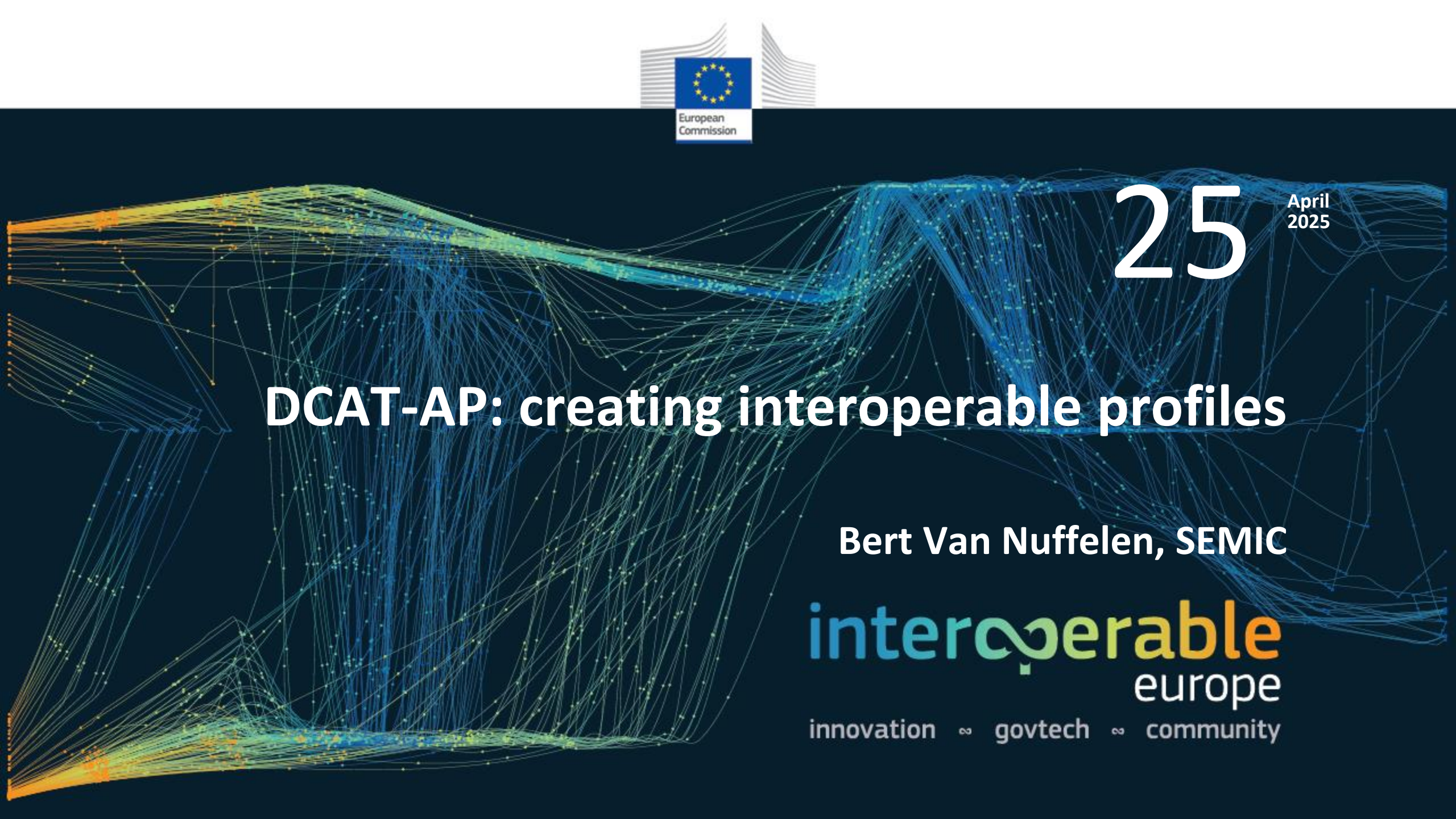
Inmaculada Farfan Velasco
data.europa.eu,
Publications Office of the European Union

Bert van Nuffelen
Semantic Data Interoperability Expert,
Digital Flanders

Lenka Kováčová
National Open Data Coordinator,
Czech Digital and Information Agency

Pierlou Ramade
Data Engineer,
French Interministerial Digital Directorate (DINUM)

data europa academy



25 April 2025

DCAT-AP: creating interoperable profiles

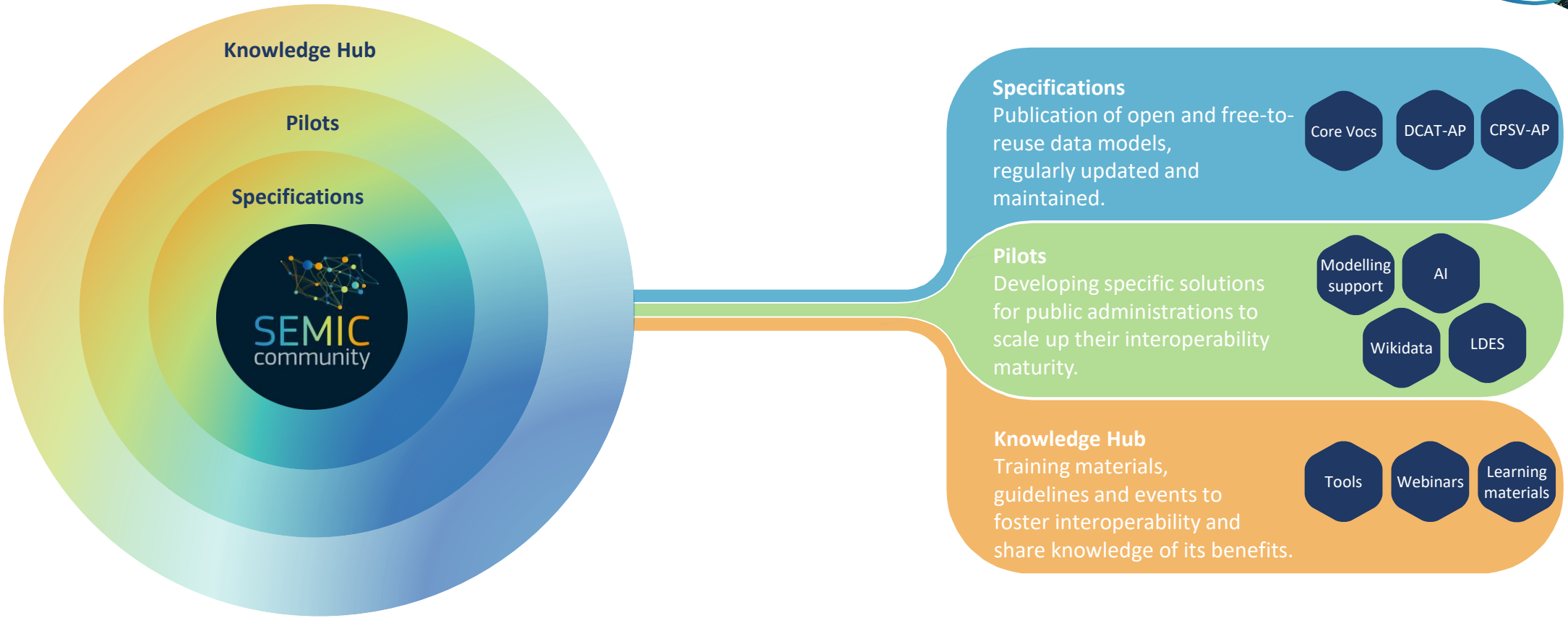
Bert Van Nuffelen, SEMIC

interoperable europe

innovation ∞ govtech ∞ community

SEMIC

SEMIC's mission is to promote Semantic Interoperability amongst the EU Member States and deliver pragmatic support to help build an Interoperable Europe.



Specifications

SEMIC specifications enable interoperability:

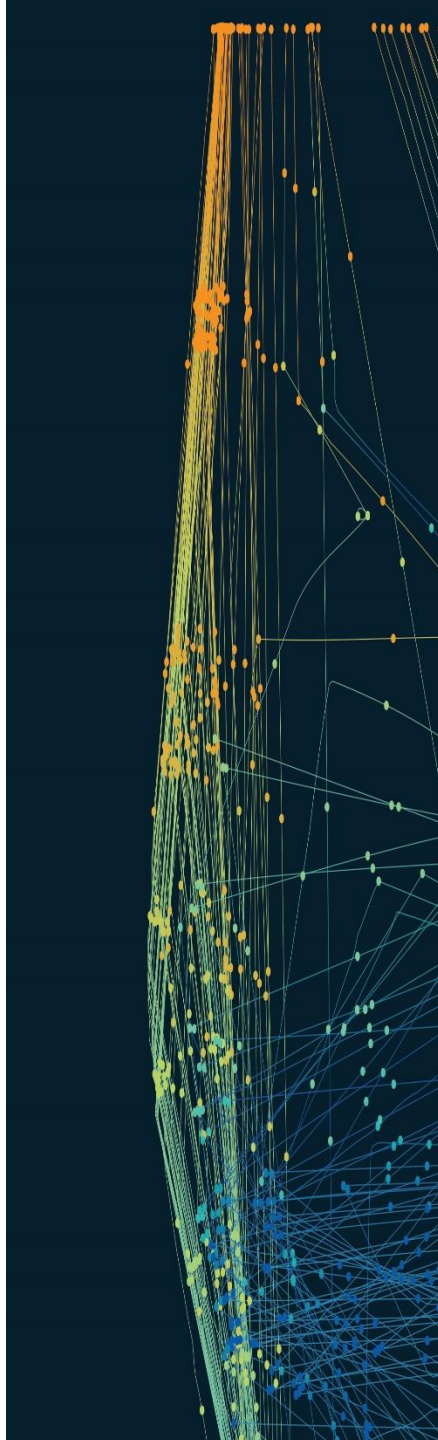
- They make data **transparent** and **available**
- They support the **coherent** implementation of laws and policies
- They help implement **cost efficiencies**
- They help **digitalisation** and **harmonising** processes

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

An official website of the European Union How do you know?

European Commission

interoperable europe

Policy Solutions Support Centre Governance Academy Sign in Get started

SEMIC Support Centre

Join this collection

Welcome Services Catalogue of Services Base Registries AI for Interoperability Data Spaces SEMIC Conference See more

Welcome

Bookmark Translate

Enabling Interoperable Europe

The SEMIC Support Centre is the first step on a journey to become an interoperable European administration. In line with the Interoperable Europe Act, the SEMIC community stands for co-creation and sharing best practices. The Support Center is also home to a community of adopters that share tooling, standards and good practices to promote semantic interoperability.

Who is it for?

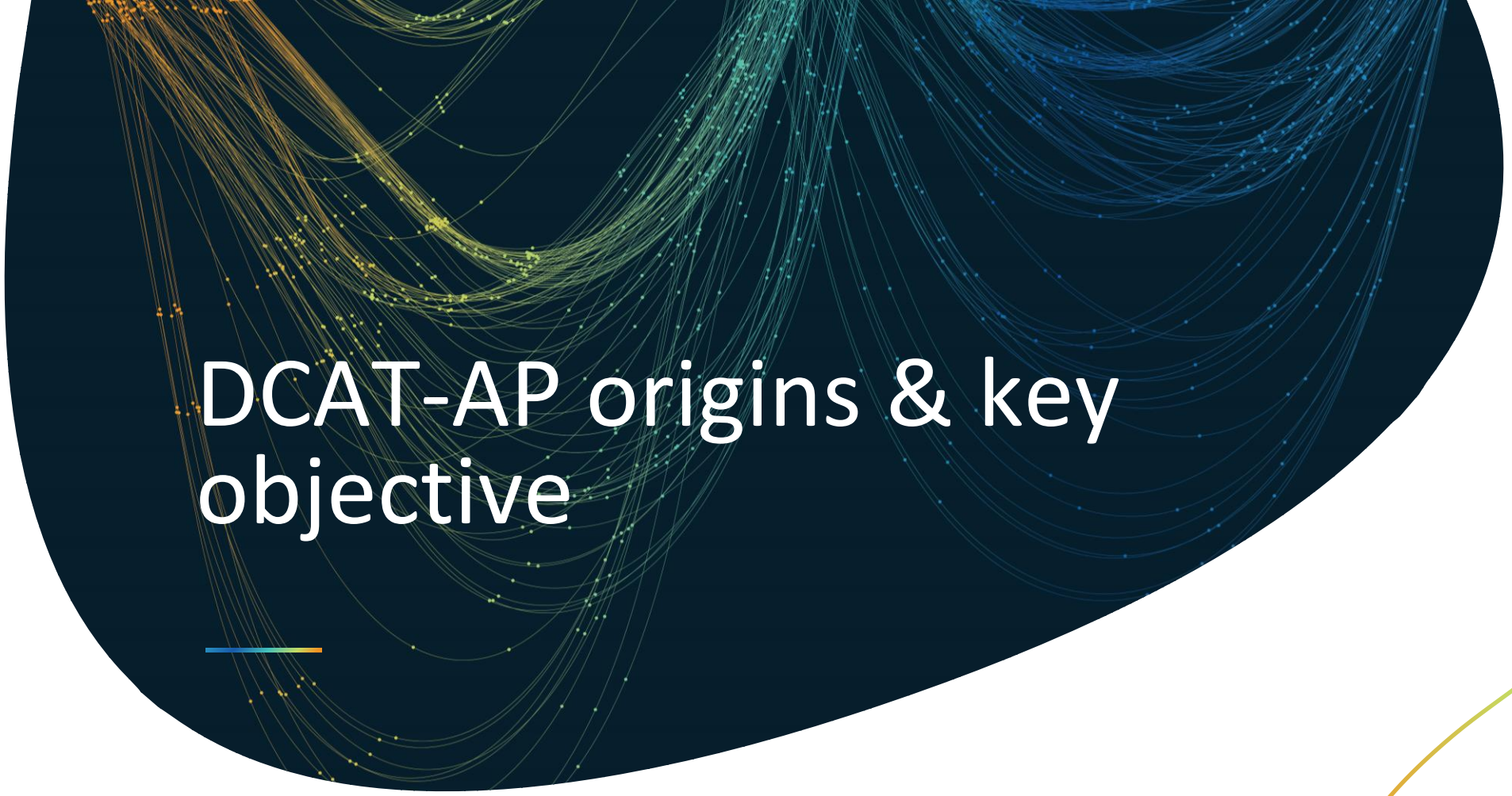
Are you a policy officer or a technical expert working on data sharing and reuse? SEMIC and its support centre offers several services that aim to help public administrations share their experience with each other and help each other out in making the single

In the spotlight: DCAT-AP for Data Spaces!

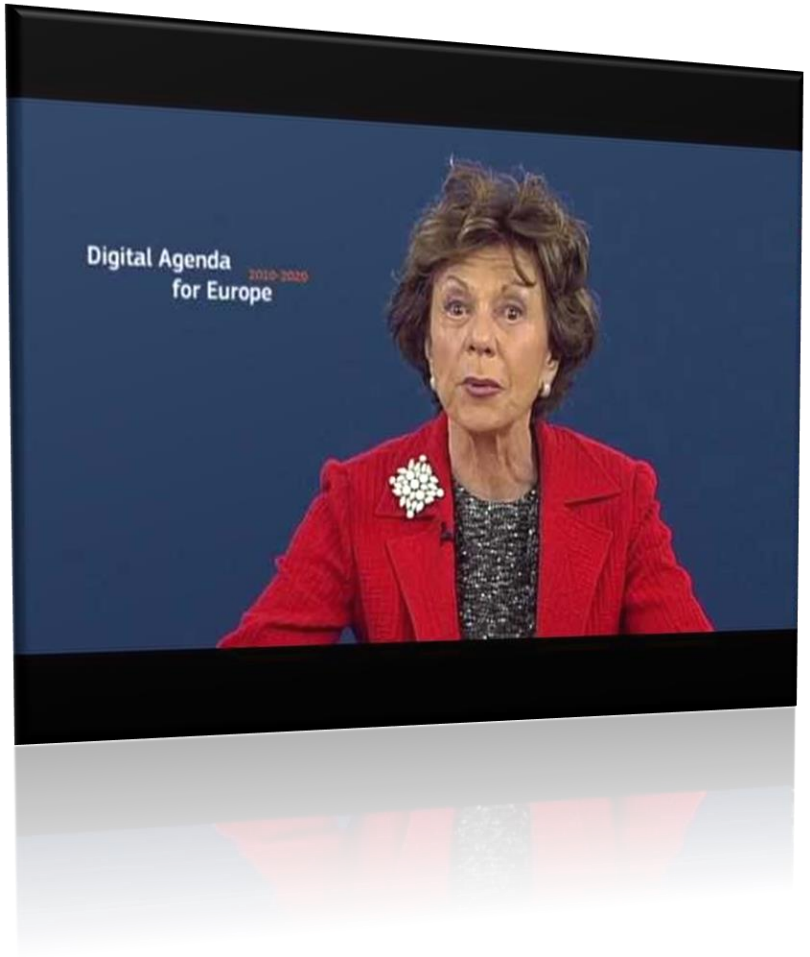
Discover how DCAT-AP is revolutionising data interoperability across European Data Spaces. Learn how this harmonised semantic specification facilitates data findability, accessibility, and sharing, driving collaborative innovation and breaking down data silos.

Explore real-world applications in health, mobility, and geospatial data, showcasing the versatility and impact of DCAT-AP in uniting data infrastructures for a competitive European market.

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



DCAT-AP origins & key objective



"Data in the new Oil" – Neelie Kroes (**2012**)

Foster Open Data and ensure access to (public sector) data.

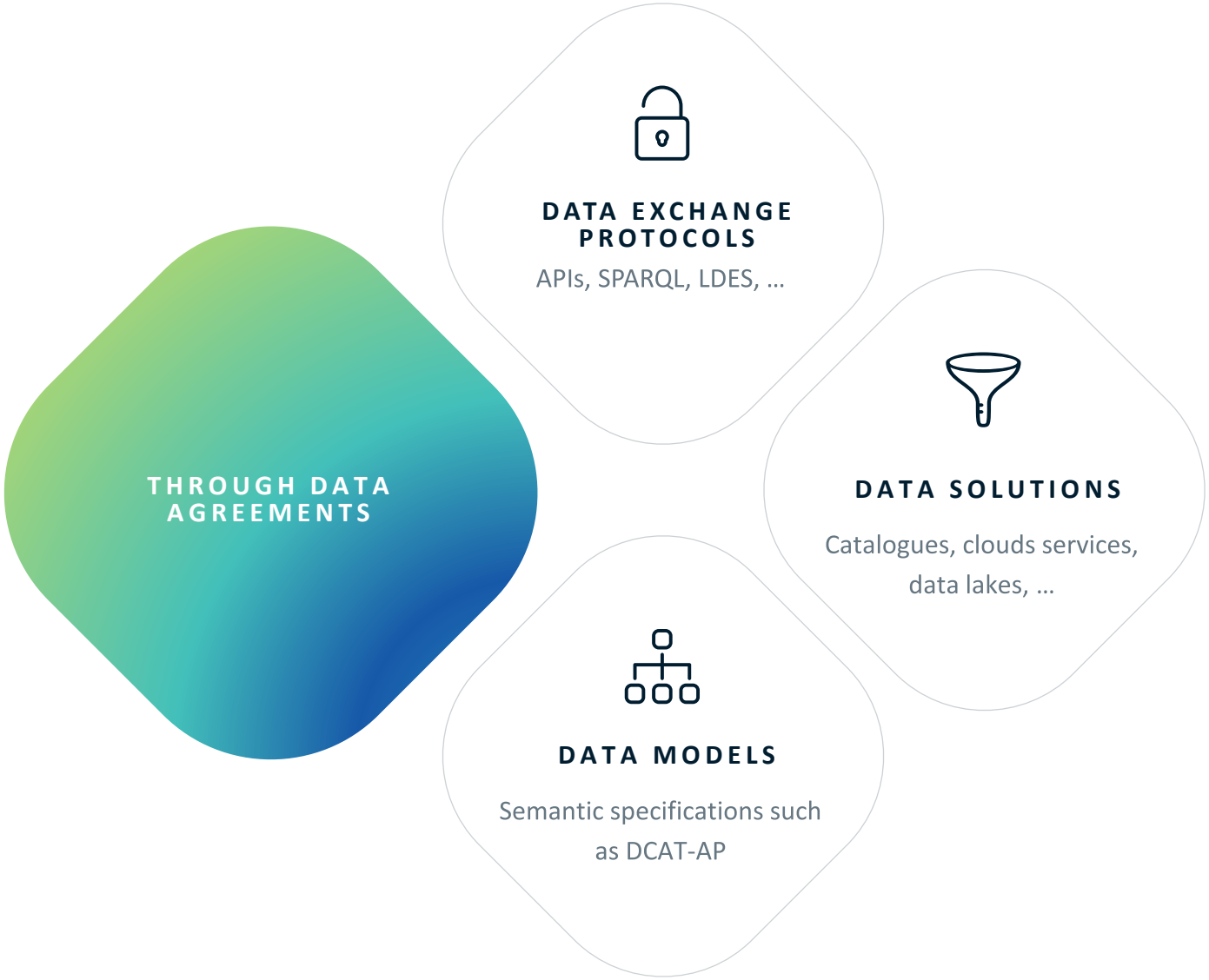
(Open) Data Catalogues emerged everywhere (in the public sector).

European Commission

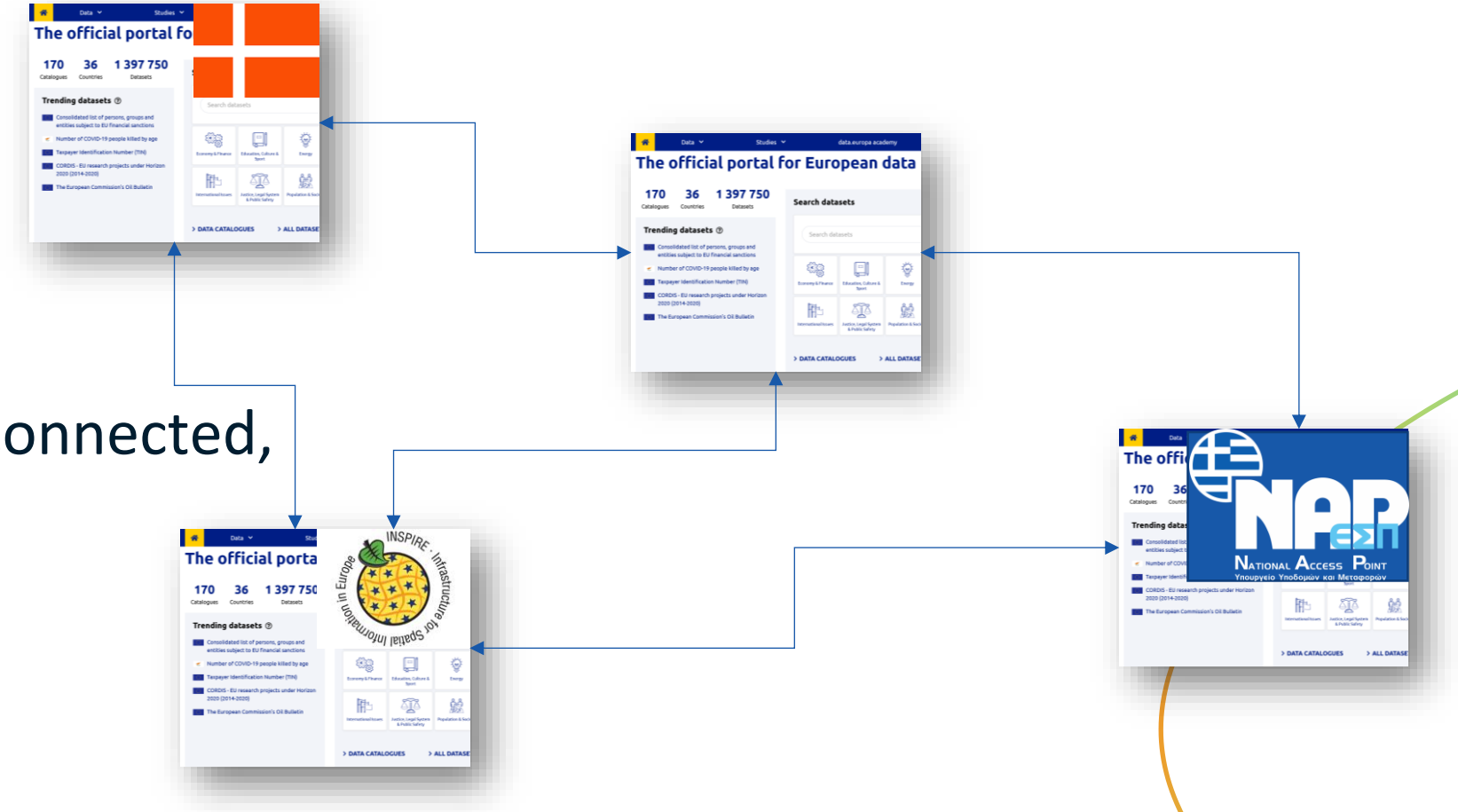
HOW?

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

Applies to making data **more "*findable*"**?



Data Catalogues make data findable, ...



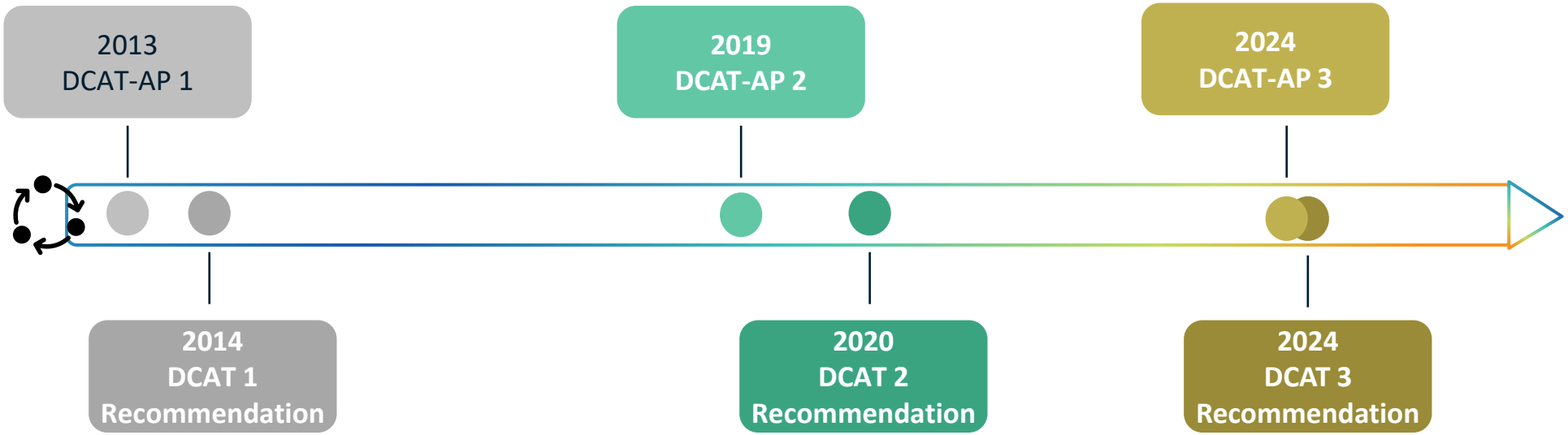
.. are interconnected,

.. and have their own needs

European Commission

DCAT (W3C) and DCAT-AP (EU)

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

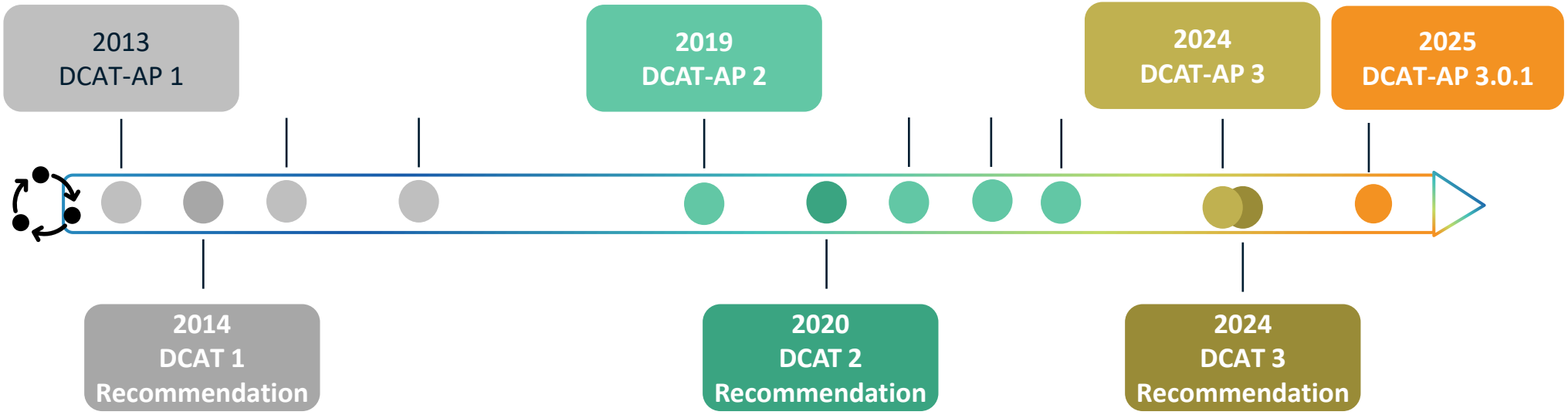


https://www.w3.org/TR/vocab-dcat-3/

European Commission

DCAT (W3C) and DCAT-AP (EU)

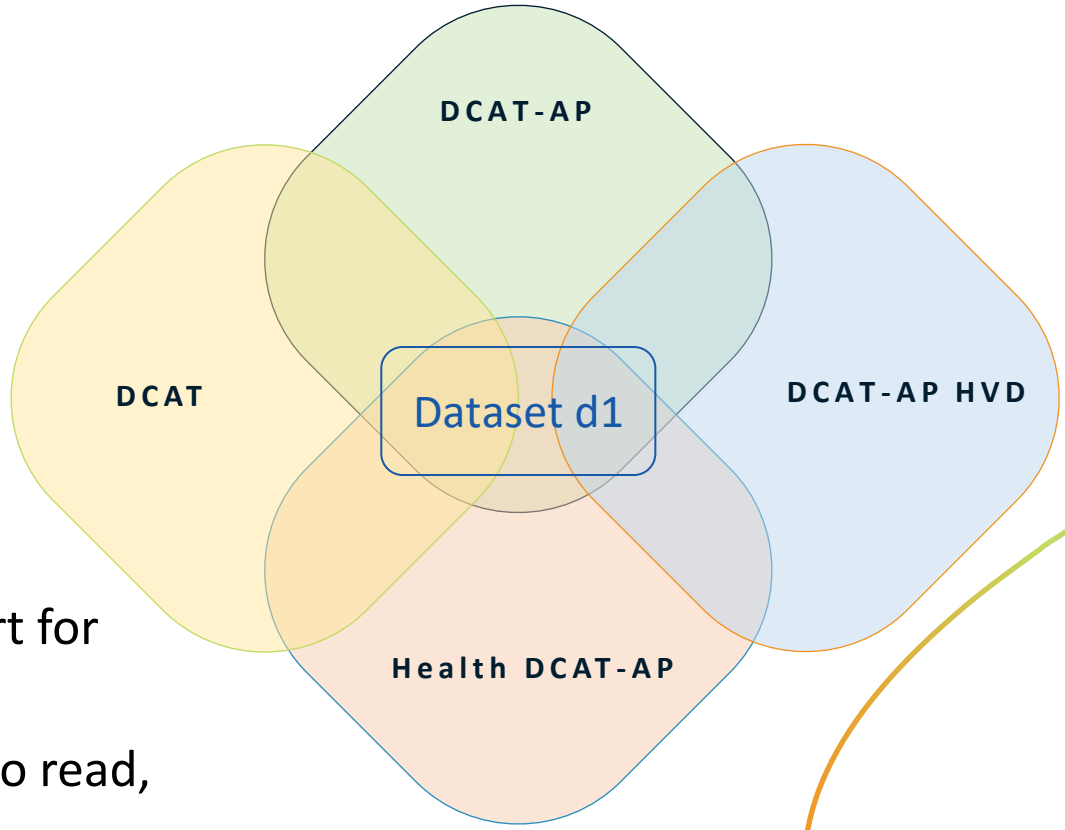
https://semiceu.github.io/DCAT-AP/releases/3.0.1-draft/



https://www.w3.org/TR/vocab-dcat-3/

European Commission

Increasing importance for multi-domain collaboration



Key aspects

- (target = publishers) Once-only effort for publishers of datasets
- (target = editors) Concise, easiness to read, minimal editorial effort

European Commission

Domains of application

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

- DCAT-AP HVD
- GeoDCAT-AP
- StatDCAT-AP

National, regional and local (Open) data portals

(too many to list)

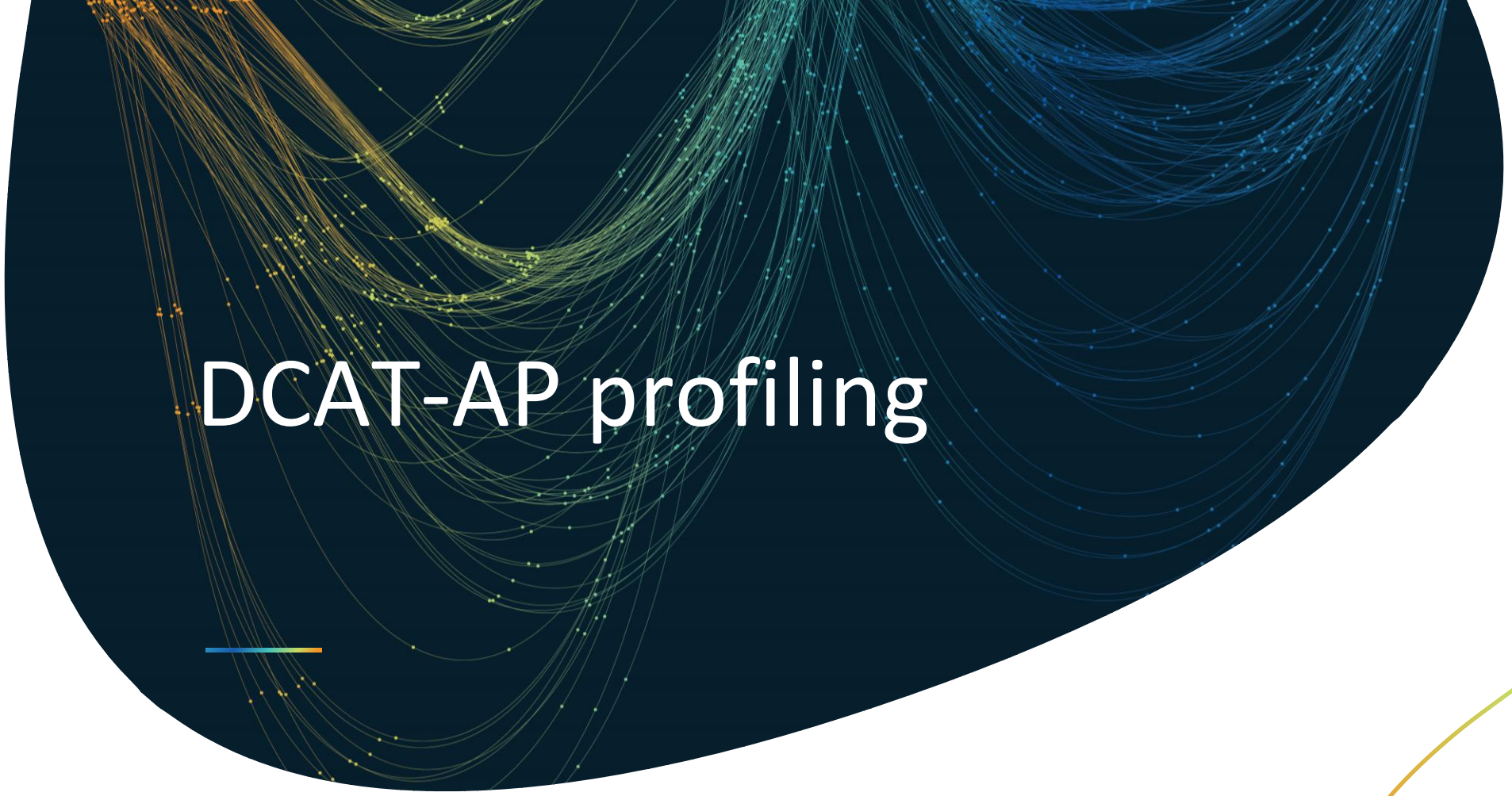
Data spaces

- Mobility DCAT-AP (NapCore)
- Health DCAT-AP
- Language Data Space
- ...

Base registries metadata descriptions

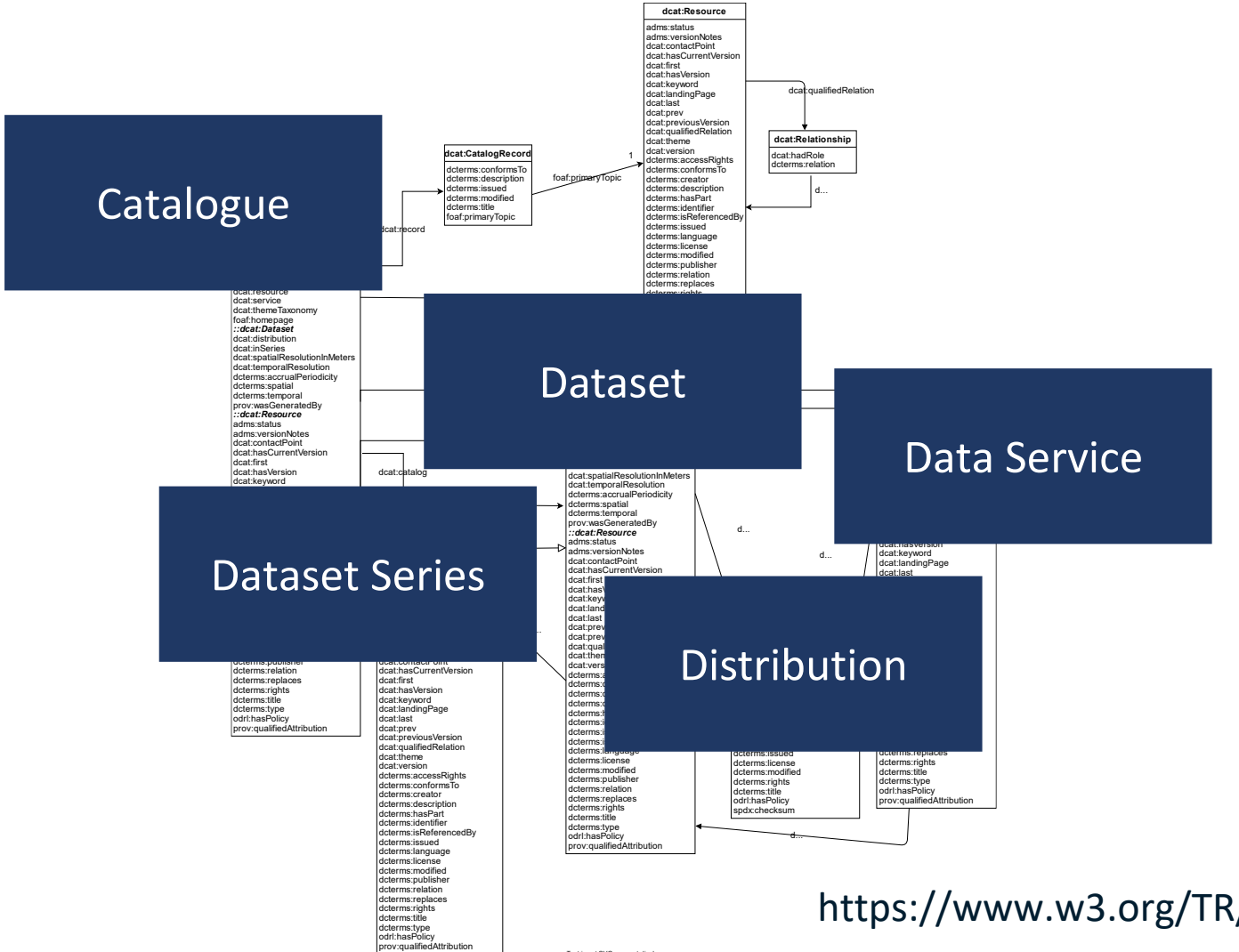
Machine Learning with MLDCAT-AP

European Commission



DCAT-AP profiling

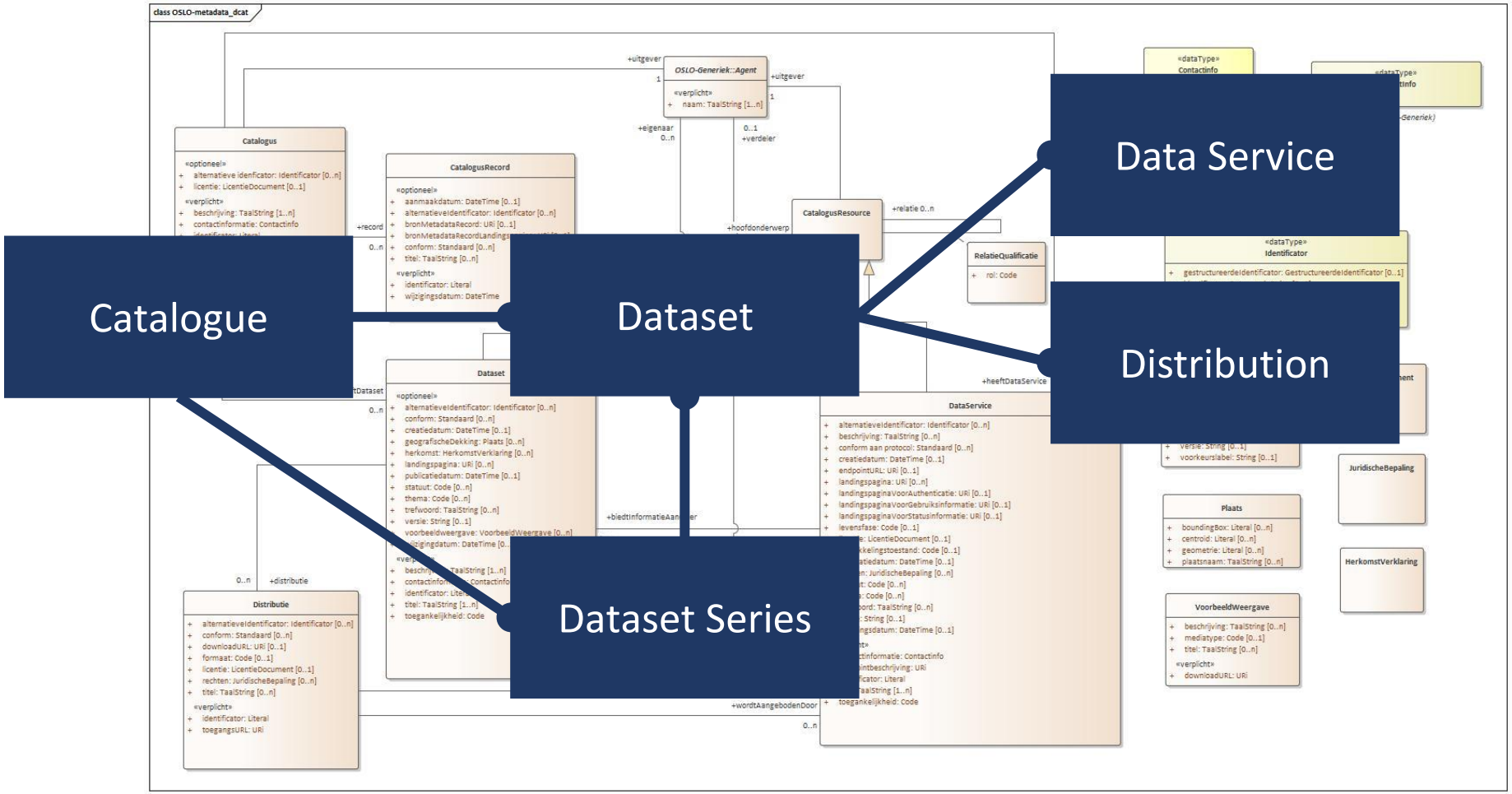
From DCAT vocabulary to …



https://www.w3.org/TR/vocab-dcat-3/

European Commission

DCAT-AP application profile



https://github.com/SEMICeu/DCAT-AP/

European Commission

Profiling DCAT-AP

= *The process of making the usage of DCAT-AP more precise for the* ***intended scope*** *(i.e. Health data).*

Based on **use cases** (scope, motivation, support) to decide property (re)use, cardinalities, codelists, etc.

Key consideration: why creating a DCAT-AP profile?

Never forget: the profile becomes part of this multi-domain ecosystem.

European Commission

Modeling decisions Application Profile

§ 7.2 Catalogue

Definition

A catalogue or repository that hosts the Datasets or Data Services being described.

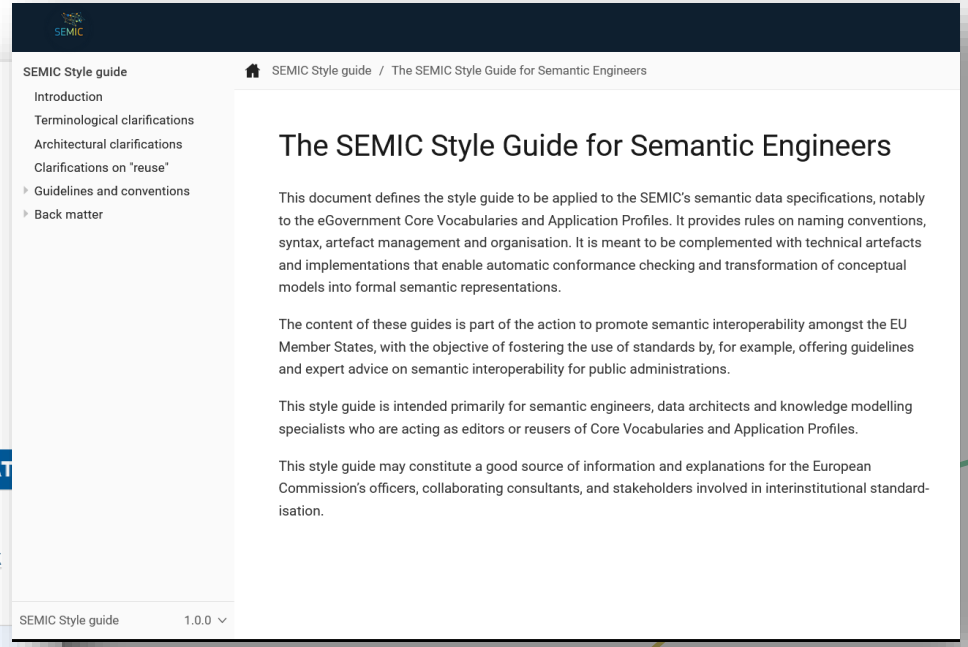
Reference in DCAT

Link

Properties

For this entity the following properties are defined: catalogue , creator , dataset , description , geographical coverage , has part , homepage , is part of , language , licence , modification date , publisher , record , release date , rights , service , temporal coverage , themes , title .

Property	Range	Card	Definition	Usage	DCAT
catalogue	Catalogue	0..n	A catalogue whose contents are of interest in the context of this catalogue		Link
creator	Agent	0..1	An entity responsible for the creation of the catalogue		Link
dataset	Dataset	0..n	A Dataset that is part of the Catalogue.	As empty Catalogues are usually indications of problems, this property should be combined with the next property service to implement an empty Catalogue check.	Link
description	Literal	1..n	A free-text account of the Catalogue.	This property can be repeated for parallel language versions of the description.	Link



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

DCAT-AP Profile Alignment principles

Recurring Questions

- Q1: How to indicate that the properties (constraints) of one class apply to another class?
- Q2: How to reuse a property with conflicting constraints?
- Q3: Is profile A coherent/consistent with profile B and what are the changes?

Proposed principles

- P1: DCAT-AP is master profile
- P2: use subproperties to avoid conflicts
- P3: diagram subclass arrows means adopt all constraints of superclass
- P4: indicate reuse with extra column
- P5: expressing local scopes as additional classes is allowed
- P6: use the codelist usage qualifiers as defined in DCAT-AP

Additional support

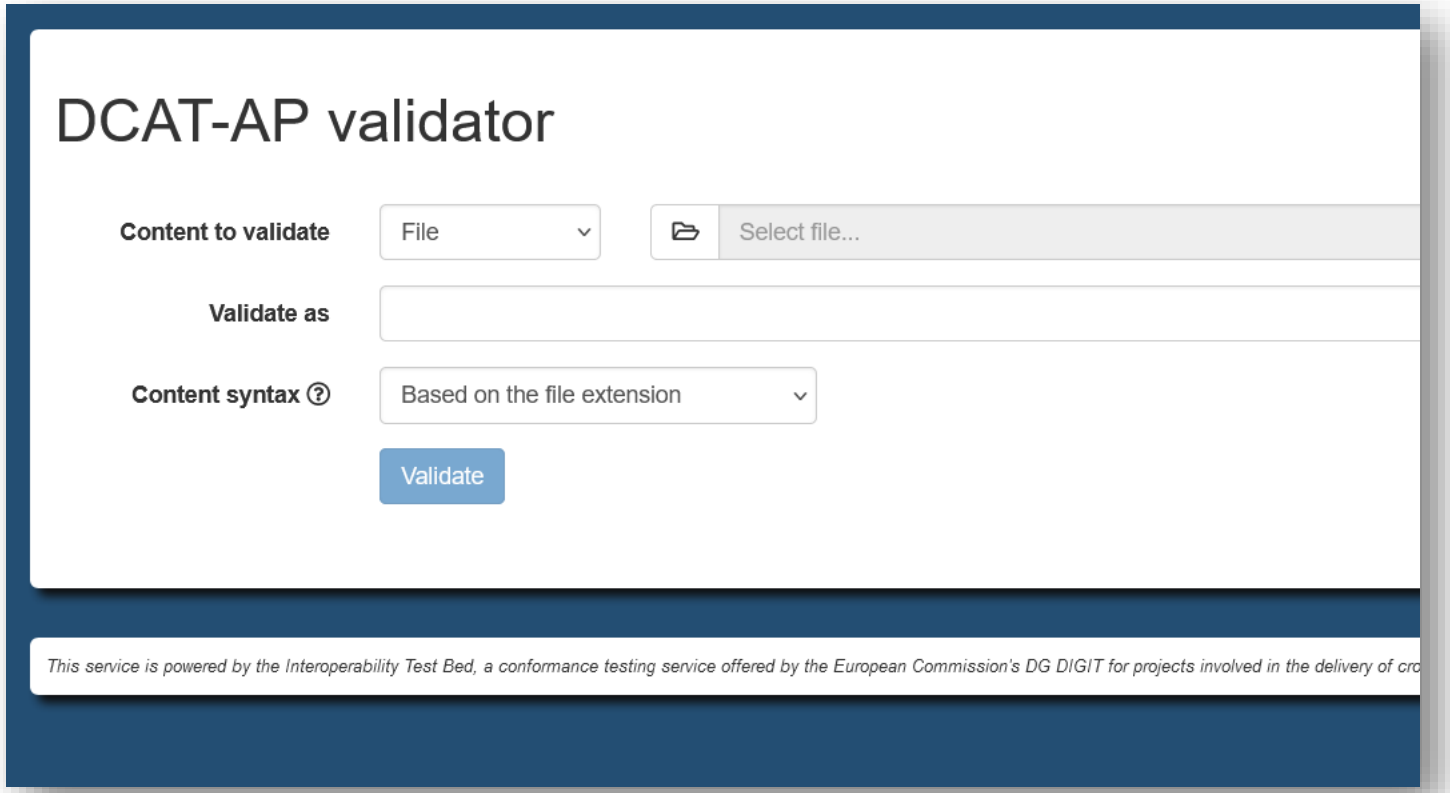
DCAT-AP validator

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

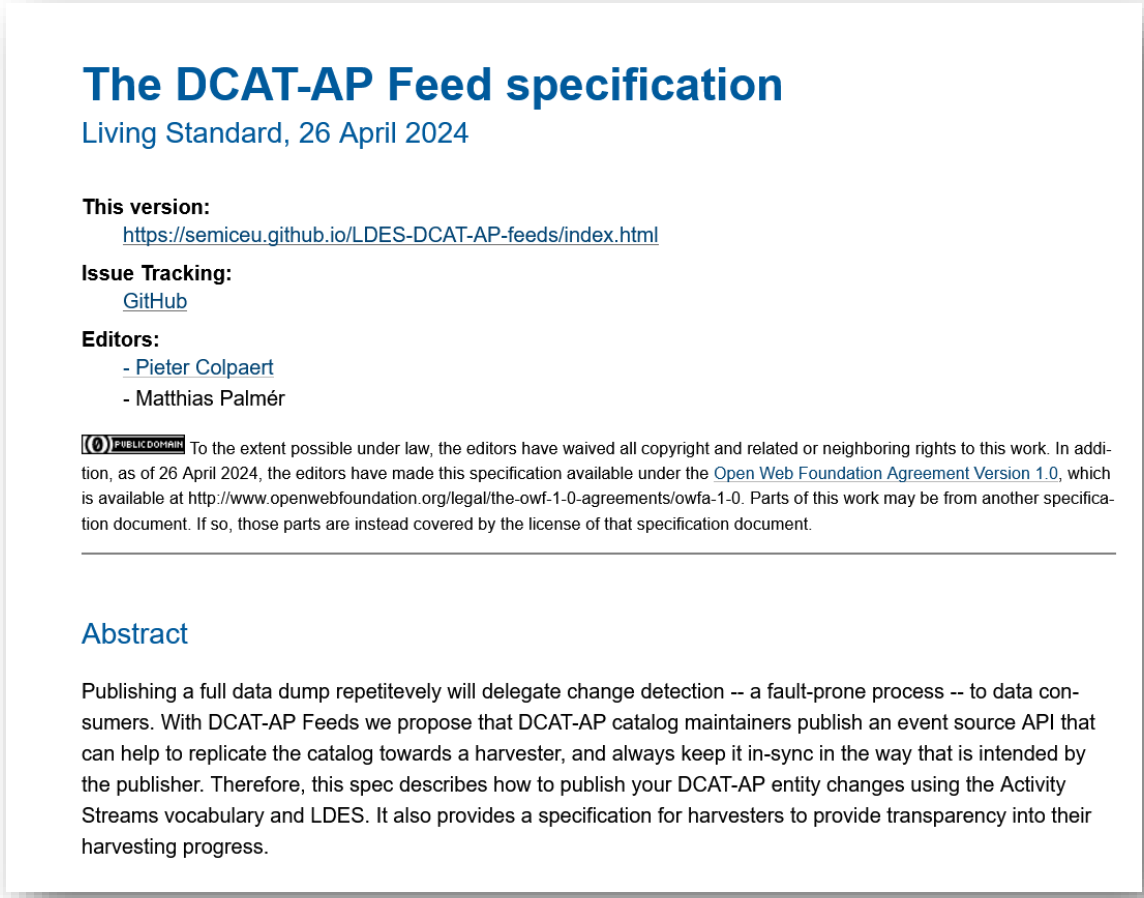
Question

How to share DCAT-AP in a harvesting network?

Linked Data Event Streams

- https://semiceu.github.io/LinkedDataEventStreams/
- A protocol to share a collection of immutable objects in Linked Data native format

Pilot with OpenData Portal Sweden



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

PUBLIC DOMAIN To the extent possible under law, the editors have waived all copyright and related or neighboring rights to this work. In addition, as of 26 April 2024, the editors have made this specification available under the Open Web Foundation Agreement Version 1.0, which is available at http://www.openwebfoundation.org/legal/the-owf-1-0-agreements/owfa-1-0. Parts of this work may be from another specification document. If so, those parts are instead covered by the license of that specification document.

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

Thank you

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

METADATA_AND_DATA_INTEROPERABILITY_ CZECHIA_

WHY_WHAT_HOW_

LENKA_KOVACOVA_

Czech Digital and Information Agency

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

CONTENT_

- Who are we
- Metadata interoperability
- DCAT-AP-CZ
- Metadata quality metrics
- Data interoperability

WHAT_OUR_TEAM_DOES?

- Digital and Information Agency – **Data Cataloging and Sharing Unit**
- Coordination of the data cataloging and sharing in public administration
  - vision and **strategy** for data
  - methodological, technical and legal **support** to public authorities
  - management of open data portal/**national catalog**
  - leading of public sector data **working groups**

data.gov.cz

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

WHAT_DO_WE_WANT_TO_ACHIEVE?

1. To have high-quality data about data - **metadata**
2. To ensure easy data **findability** and **interoperability**

- If data is not properly cataloged, then:
  - there's a risk that users **won't find** the datasets
  - users will have **doubts** about their validity and origin
  - due to poorly accessible documentation, there's a risk that the data will be **misinterpreted**
  - users won't have an overview of the **terms of use** for the respective dataset

WHAT_DO_WE_WANT_TO_ACHIEVE?

1. To have high-quality data about data - **metadata**
2. To ensure easy data **findability** and

DATA interoperability
METADATA interoperability

- If data is not properly cataloged, then:
  - there's a risk that users **won't find** the datasets
  - users will have **doubts** about their validity and origin
  - due to poorly accessible documentation, there's a risk that the data will be **misinterpreted**
  - users won't have an overview of the **terms of use** for the respective dataset

METADATA_INTEROPERABILITY_

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA

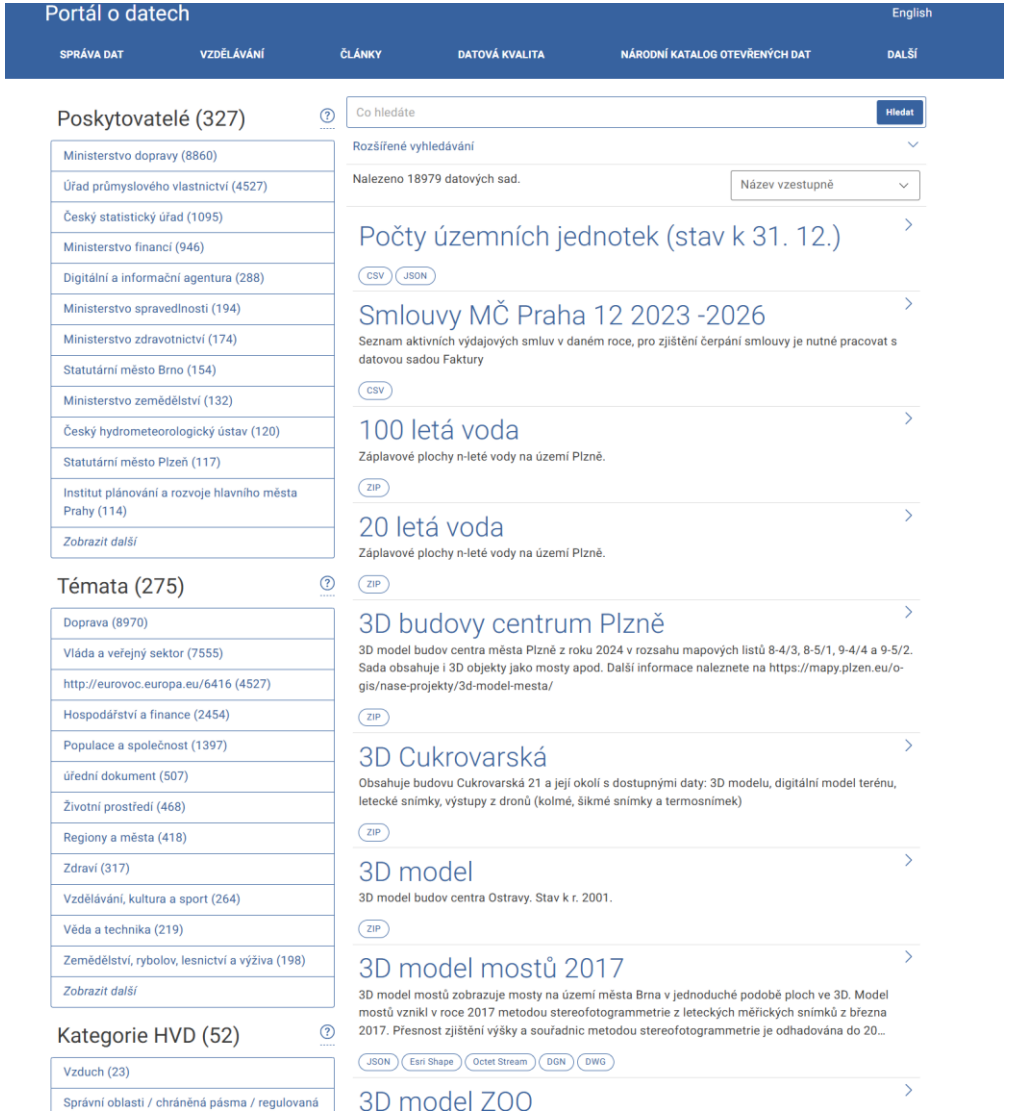
WHEN_IT_COMES_TO_METADATA_...
ONE_TO_RULE_THEM_ALL_...

- **ONE catalog**
  - legally binding to register open data in the National Open Data Catalog

- **ONE standard – DCAT-AP-CZ**
  - open formal standard (ODD)
  - every open formal standard is according to law legally binding for public administration bodies – metadata standards included

NATIONAL_OPEN_DATA_CATALOG_

- It aggregates metadata about open datasets across the **entire** Czechia
- It allows to **search** and **filter** datasets, e.g., by providers, keywords, formats, etc.



DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

Absolventi vysokých škol podle oborů

Královéhradecký kraj

§

Vysoké školy | Absolventi | Školství | Královéhradecký kraj

Datová sada s prostorovou lokalizací absolventů vysokých škol podle oborů za rok 2022. Zdrojem dat je Agentura pro podporu podnikání a investic CzechInvest.

Téma
Vzdělávání, kultura a sport
Věda a technika

Související geografické území
Královéhradecký kraj

Periodicita aktualizace
roční

Distribuce datové sady

SQLite Geodatabase
http://publications.europa.eu/resource/authority/file-type/UNKNOWN

Podmínky užití distribuce	Soubor ke stažení
Neobsahuje ✓ Autorské dílo	Stáhnout
Neobsahuje ✓ Originální databáze	Typ média
Není chráněna ✓ Zvláštní právo pořizovatele databáze	
Neobsahuje Osobní údaje	

CSV
CSV

Podmínky užití	Soubor ke stažení
Autorské dílo	Stáhnout
Neobsahuje ✓ Originální databáze	text/csv
Není chráněna ✓ Zvláštní právo pořizovatele databáze	Typ média
Neobsahuje Osobní údaje	

GeoPackage
http://publications.europa.eu/resource/authority/file-type/UNKNOWN

Podmínky užití distribuce	Soubor ke stažení
Neobsahuje ✓ Autorské dílo	Stáhnout
Neobsahuje ✓ Originální databáze	Typ média
Není chráněna ✓ Zvláštní právo pořizovatele databáze	
Neobsahuje Osobní údaje	

Excel
http://publications.europa.eu/resource/

ArcGIS Hub Dataset
HTML

KML
KML

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

HOW_TO_GET_YOUR_DATA_ TO_THE_NATIONAL_CATALOG?

- **Simple form**
  - Recommended for providers with few datasets
  - The provider registers the chosen dataset; thanks to the form, the **metadata record complies with the standard**
- **Local OD catalogue**
  - A systematic approach to cataloging open data
  - Simple app for managing metadata of data providers
  - The provider registers the local catalog that must follow **the standard**

DCAT-AP-CZ

1 Datová sada — 2 Distribuce — 3 Shrnutí

! Nedělejte stejné chyby jako řada poskytovatelů před vámi. Před registrací datové sady se ujistěte, že neděláte nějaké z nejčastějších chyb v otevřených datech!

Název datové sady česky

Název datové sady anglicky (volitelné)

Popis datové sady česky

Popis datové sady anglicky (volitelné)

Klíčová slova česky

Klíčová slova anglicky

Téma datové sady

Téma datové sady (volitelné)

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

1 Katalog — 2 Shrnutí

! Nedělejte stejné chyby jako řada poskytovatelů před vámi. Před registrací lokálního katalogu se ujistěte, že neděláte nějaké z nejčastějších chyb v otevřených datech! Také se ujistěte, že rozhraní vašeho katalogu splňuje Otevřenou formální normu pro rozhraní katalogů otevřených dat.

Název lokálního katalogu otevřených dat česky

Název lokálního katalogu otevřených dat anglicky (volitelné)

Jméno správce katalogu

E-mail správce katalogu

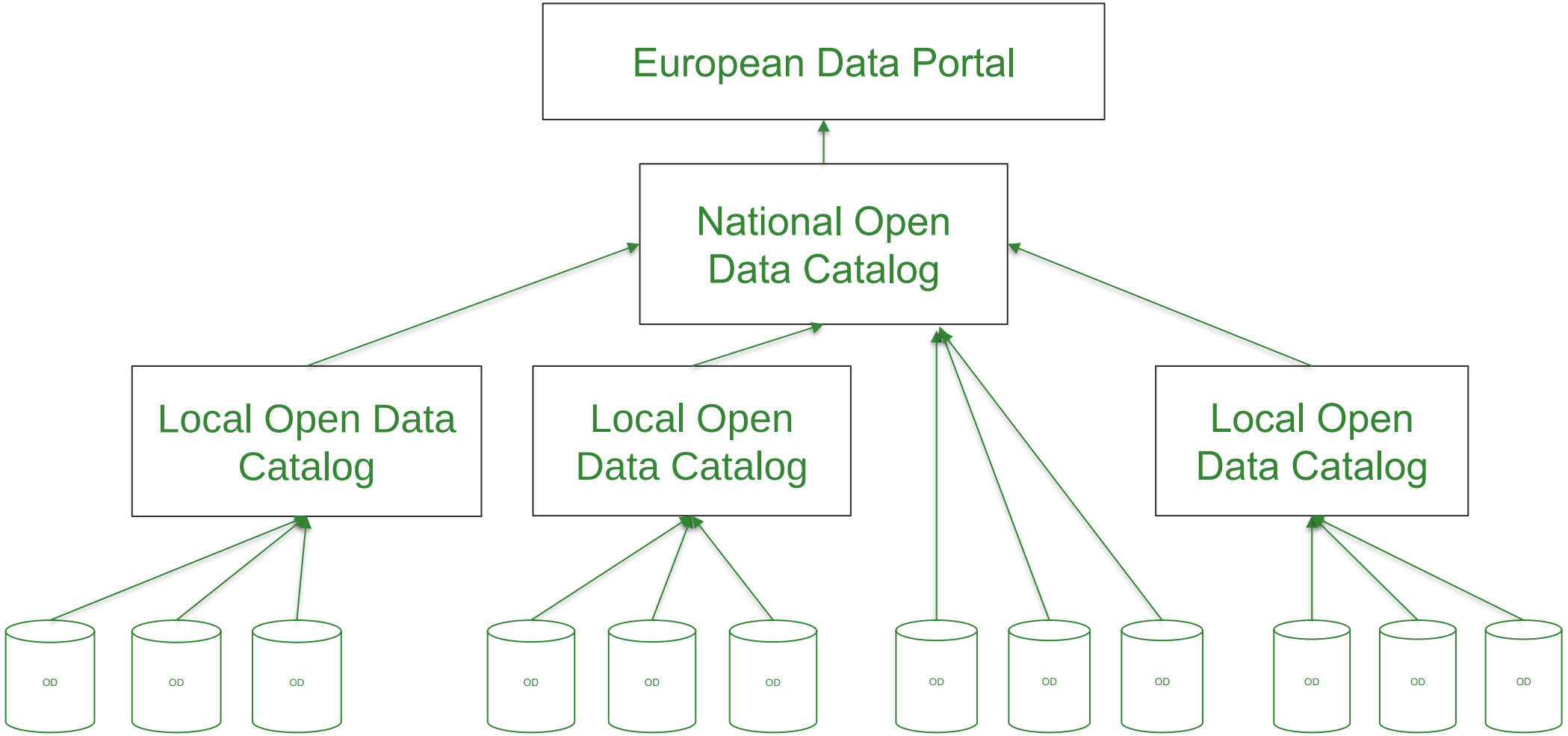
Typ API lokálního katalogu

URL LKOD API

Domácí stránka katalogu (volitelné)

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

DATA_CATALOG_SYSTEM_



DCAT-AP-CZ_

- Basic model for data catalogs of the **Czech application profile** of the European DCAT-AP standard for data cataloging
- **Czech specificities**
  - more detailed way of representing the **terms of use** for dataset distribution
  - a national way of representing the **territorial coverage** of dataset using Czech territorial base register (LOD)
  - an identification of **public administration bodies** according to the Czech base registries (LOD)
  - the possibility of annotating the **topic** of a dateset using the EuroVoc thesaurus
  - some of the metadata entities (e.g., keywords, periodicity) are obligatory
- Metadata according to standard uploaded into an RDF database and made accessible via a so-called SPARQL Endpoint

OBSAH

1. **Úvod**

2. **Co je nového**

3. **Položky metadatového záznamu**

3.1 Použité prefixy

3.2 Třída: Katalog

3.2.1 Název

3.2.2 Popis

3.2.3 Poskytovatel

3.2.4 Kontaktní bod - jméno a email

3.2.5 Domovská stránka

3.2.6 Vazba: Datová sada

3.3 Třída: Datová sada

3.3.1 Název

3.3.2 Popis

3.3.3 Poskytovatel

3.3.4 Téma

3.3.5 Periodicita aktualizace

3.3.6 Klíčová slova

3.3.7 Související geografické území - prvek z RÚIAN

3.3.8 Související geografické území

3.3.9 Časové pokrytí

3.3.10 Kontaktní bod - jméno a email

3.3.11 Odkaz na dokumentaci

3.3.12 Odkaz na specifikaci

3.3.13 Klasifikace dle EuroVoc

3.3.14 Prostorové rozlišení v metrech

3.3.15 Časové rozlišení

3.3.16 Příslušný právní předpis

DIA_

NÁRODNÍ PLÁN OBNOVY

DCAT-AP-CZ: Rozhraní katalogů otevřených dat

Doporučení 28. května 2024

Tato verze:
https://ofn.gov.cz/dcat-ap-cz-rozhraní-katalogů-otevřených-dat/2024-05-28/

Nejnovější verze:
https://ofn.gov.cz/dcat-ap-cz-rozhraní-katalogů-otevřených-dat/

Minulé doporučení:
https://ofn.gov.cz/rozhraní-katalogů-otevřených-dat/2021-01-11/

Editor:
Jakub Klímek (Digitální a informační agentura)

Zúčastněte se:
GitHub datagov-cz/otevrene-formalni-normy
Nahlaste bug
Historie změn
Pull requesty

Odkazy:
Validátor lokálních katalogů
Portál otevřených dat
Kontaktní e-mail

Abstrakt

Cílem tohoto dokumentu je specifikovat rozhraní lokálních katalogů otevřených dat (LKOD) pro jejich harvestování Národním katalogem otevřených dat (NKOD). Dokument je založen na základním datovém modelu

MANDATORY_DCAT-AP-CZ_ADVANTAGES_

- **Responsibility** for metadata lies with the provider, not with central integrator
- A harvesting catalog could be done by **anyone** (anti vendor lock-in)
- To **pre-fill registration** for the provider
- To **validate** compliance with the standard
- **Easier adaptation** to changes from the European level – Data Spaces (HealthDCAT-AP, mobilityDCAT-AP), National Single Access Point (DGA)

Validátor lokálního katalogu

Tento validátor na základě URL rozhraní Lokálního katalogu otevřených dat (LKOD) zvaliduje jeho správnost vůči otevřené formální normě. Lze tedy použít během vývoje LKOD, před jeho registrací do NKOD, nebo po registraci, pokud je lokální katalog vidět v NKOD, ale nenačtou se z něj žádné datové sady. Rozhraní LKOD musí implementovat techniku CORS.

URL vstupního bodu k validaci

https://data.mf.gov.cz/lod/sparql

SPUSTIT VALIDACI

Formát vstupního bodu
DCAT-AP SPARQL Endpoint
URL katalogu
https://data.mf.gov.cz/lod/katalog/
https://monitor.statnipokladna.gov.cz/api/opendata/monitor
Název katalogu
Katalog otevřených dat MF
MONITOR Státní pokladny katalog

⚠ Nalezeno více, než jeden jeden katalog.

Odkazané záznamy datových sad s chybou
0
Celkový počet nalezených datových sad
965
Počet nalezených datových sad s varováním, ale bez chyby
23
Počet nalezených datových sad s chybou
281
Počet nalezených HVD
0

SKRÝT DETAILY

Minimální závažnost zobraných výsledků
Informace

Datová sada
https://data.mf.gov.cz/lod/katalog/8cbd6d7d-ba7f-4499-b80a-abd7d26f7d15

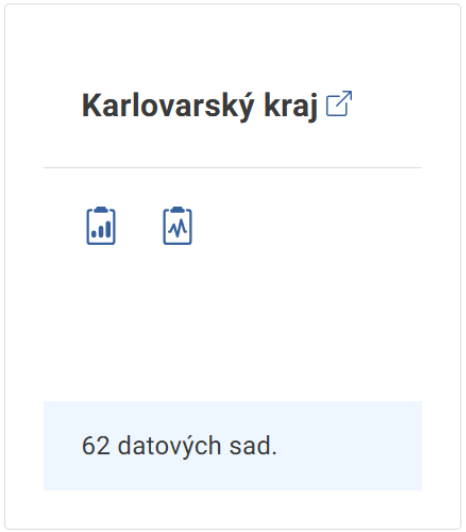
- Chybí dcat:Dataset.
- Chybí český název.
- Chybí český popisek.
- Chybí poskytovatel.

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

METADATA_QUALITY_

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA

METADATA_QUALITY_DASHBOARDS_



Distribuce datové sady

CSV
CSV

Podmínky užití distribuce	Soubor ke stažení
Neobsahuje ✓ Autorské dílo	Stáhnout
Neobsahuje ✓ Originální databáze	text/csv Typ média
Není chráněna ✓ Zvláštní právo pořizovatele databáze	
Neobsahuje Osobní údaje	

Soubor byl 22. 4. 2025 23:31:08 nedostupný. Read timed out

Shapefile
ZIP

Podmínky užití distribuce	Soubor ke stažení
Autorské dílo	zip Typ média
Neobsahuje ✓ Originální databáze	
Není chráněna ✓ Zvláštní právo pořizovatele databáze	
Neobsahuje Osobní údaje	

Excel
http://publications.europa.eu/resource/authority/file-type/UNKNOWN

Podmínky užití distribuce	Soubor ke stažení
Neobsahuje ✓ Autorské dílo	Stáhnout
Neobsahuje ✓ Originální databáze	Typ média

ArcGIS Hub Dataset
HTML

Podmínky užití distribuce	Soubor ke stažení
Neobsahuje ✓ Autorské dílo	Stáhnout
Neobsahuje ✓ Originální databáze	text/html Typ média
Není chráněna ✓	

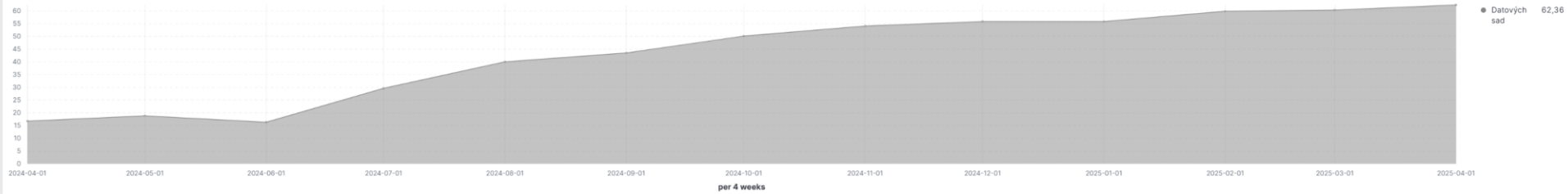
Poskytovatel
Karlovarský kraj

Metriky v čase jsou aritmetické průměry hodnot za daný měsíc.

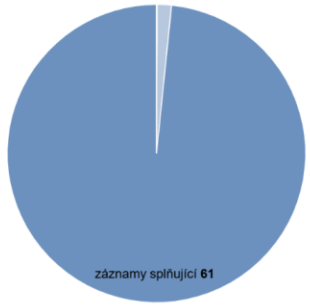
Yesterday

Datových sad
62

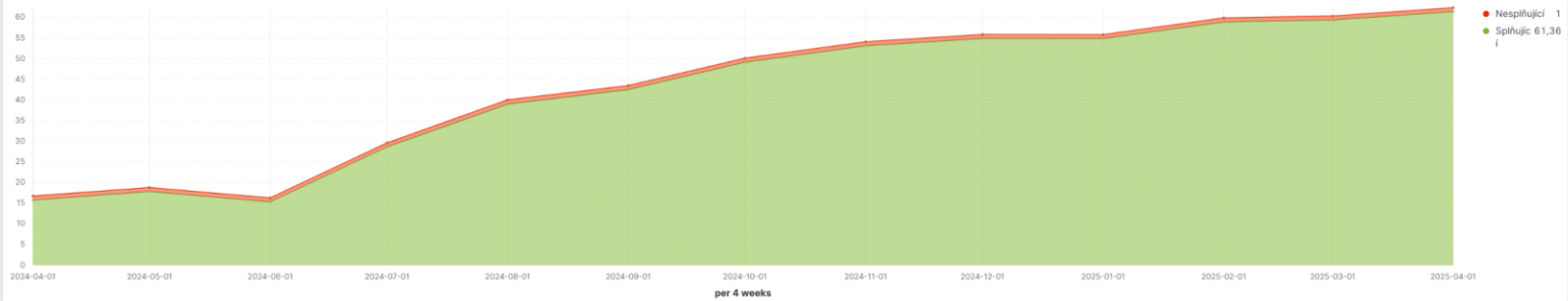
Datové sady v čase - měsíční průměr



Povinné atributy záznamů datových sad Yesterday



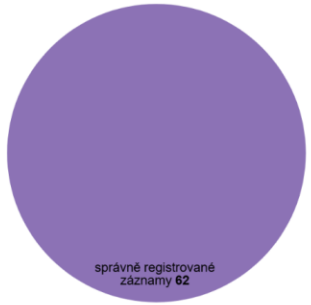
Povinné atributy záznamů datových sad v čase - měsíční průměr



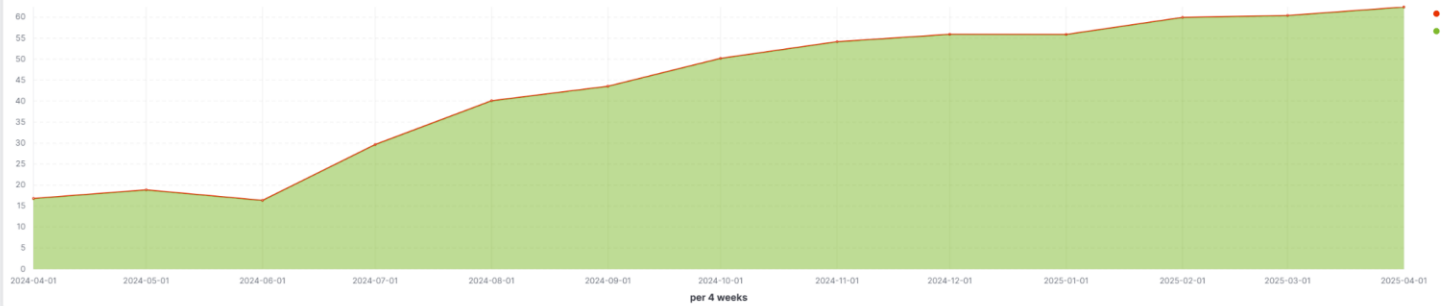
Povinné atributy

Záznamy datových sad v NKOD se skládají z množiny položek, kde některé jsou povinné. Pokud zde vidíte záznamy nesplňující povinné položky, najděte je v seznamu a zkontrolujte vůči Otevřené formální normě.

Správnost registrace záznamů datových sad v NKOD Yesterday



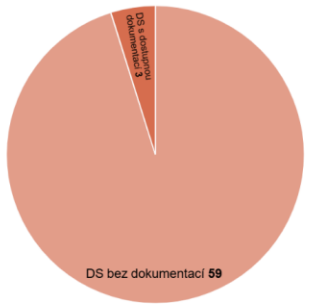
Záznamy datových sad dle souladu s OFN v čase - měsíční průměr



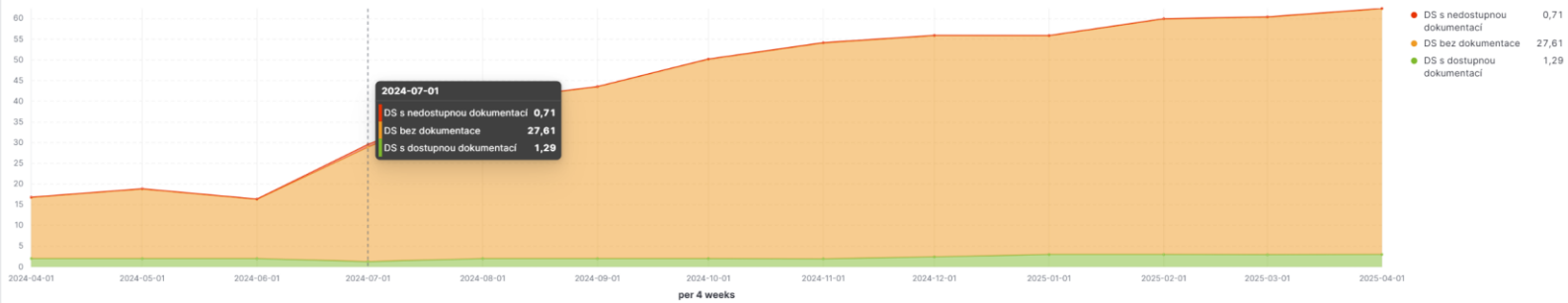
Soulad záznamů s OFN

Rozhraní lokálních katalogů otevřených dat se řídí Otevřenou formální normou. Pokud zde vidíte některé záznamy v nesouladu s OFN, zřejmě používáte již nepodporované rozhraní LKOD CKAN API. Uveďte svůj LKOD do souladu s OFN použitím jednoho z DCAT-AP rozhraní specifikovaného v OFN.

Dostupnost dokumentací datových sad Yesterday



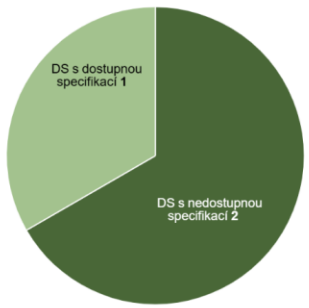
Dostupnost dokumentací datových sad v čase - měsíční průměr



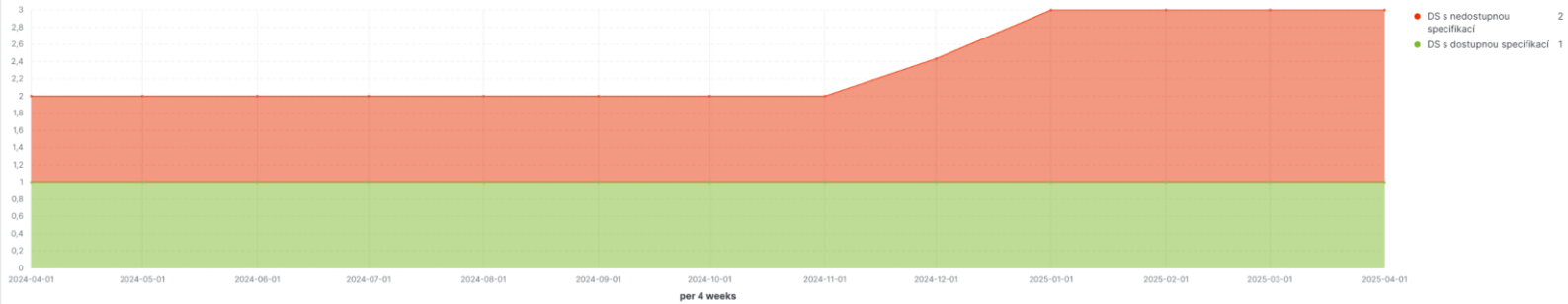
Dokumentace datové sady

Pokud máte některé dokumentace datových sad nedostupné, najděte je v seznamu. Zkontrolujte, že se nejedná o nějakou z technických překážek přístupu k otevřeným datům. Zkuste k URL přistoupit nástrojem curl pomocí metody HTTP HEAD.

Dostupnost specifikací datových sad Yesterday



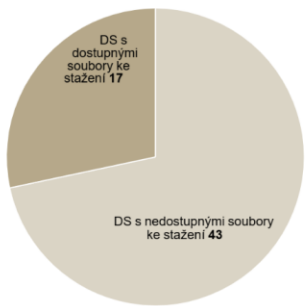
Dostupnost specifikací datových sad v čase - měsíční průměr



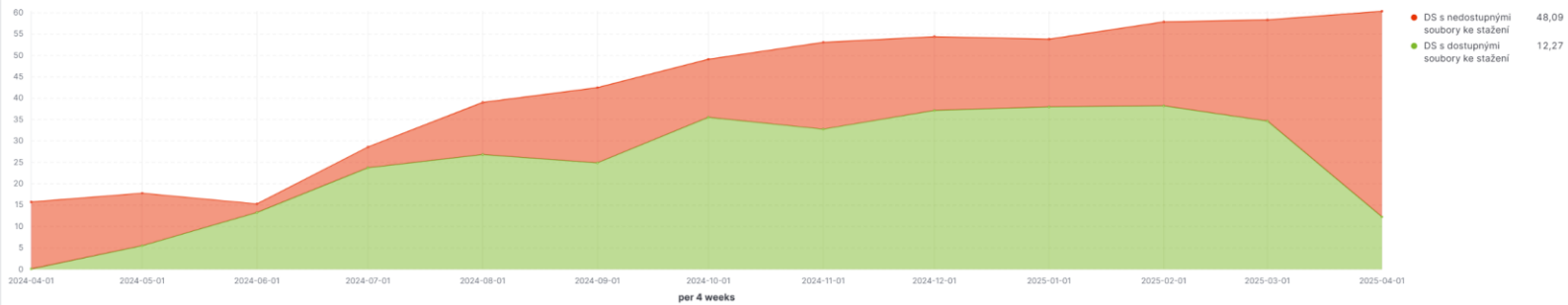
Specifikace datové sady

Pokud máte některé specifikace datových sad nedostupné, najděte je v seznamu. Zkontrolujte, že se nejedná o nějakou z technických překážek přístupu k otevřeným datům. Zkuste k URL přistoupit nástrojem curl pomocí metody HTTP HEAD.

Dostupnost souborů ke stažení Yesterday



Dostupnost souborů ke stažení v čase - měsíční průměr



Soubor ke stažení

Pokud máte některé soubory ke stažení nedostupné, najděte je v seznamu. Zkontrolujte, že se nejedná o nějakou z technických překážek přístupu k otevřeným datům. Zkuste k URL přistoupit nástrojem curl pomocí metody HTTP HEAD. V případě nedostupnosti techniky CORS opět zkontrolujte dle seznamu a případu špatné praxe. Zkuste k URL přistoupit nástrojem curl pomocí metody HTTP OPTIONS.

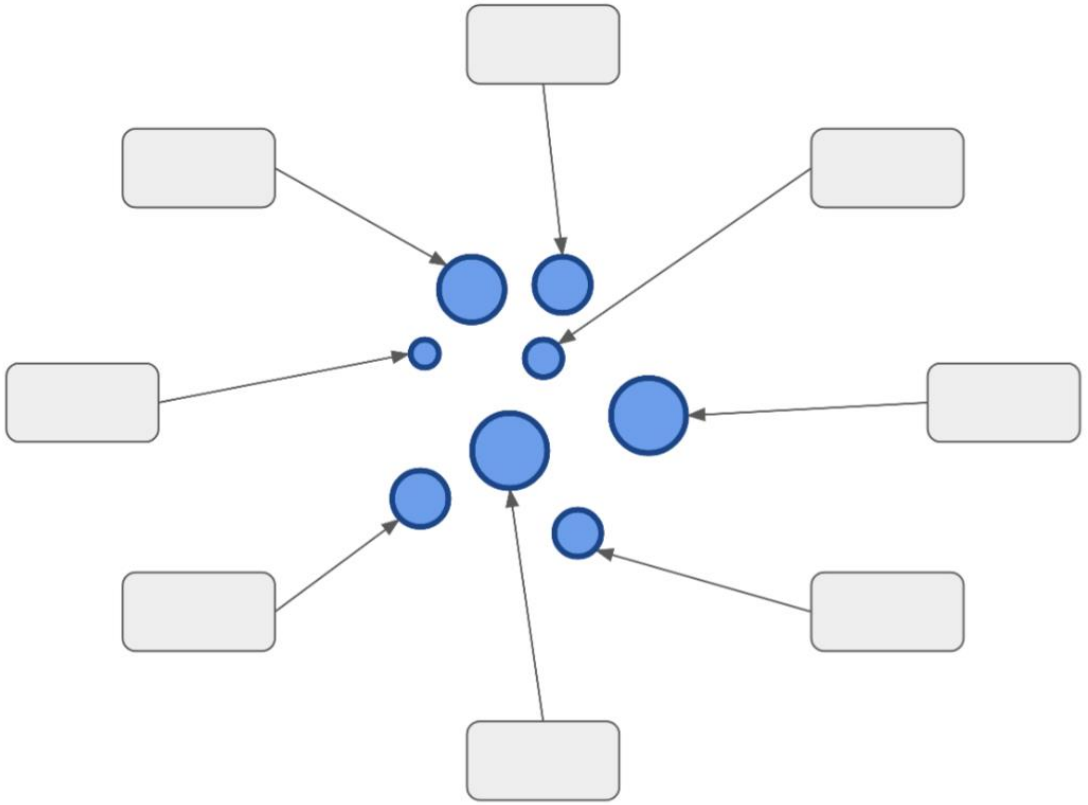
DATA_INTEROPERABILITY_

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA

DATA_INTEROPERABILITY_

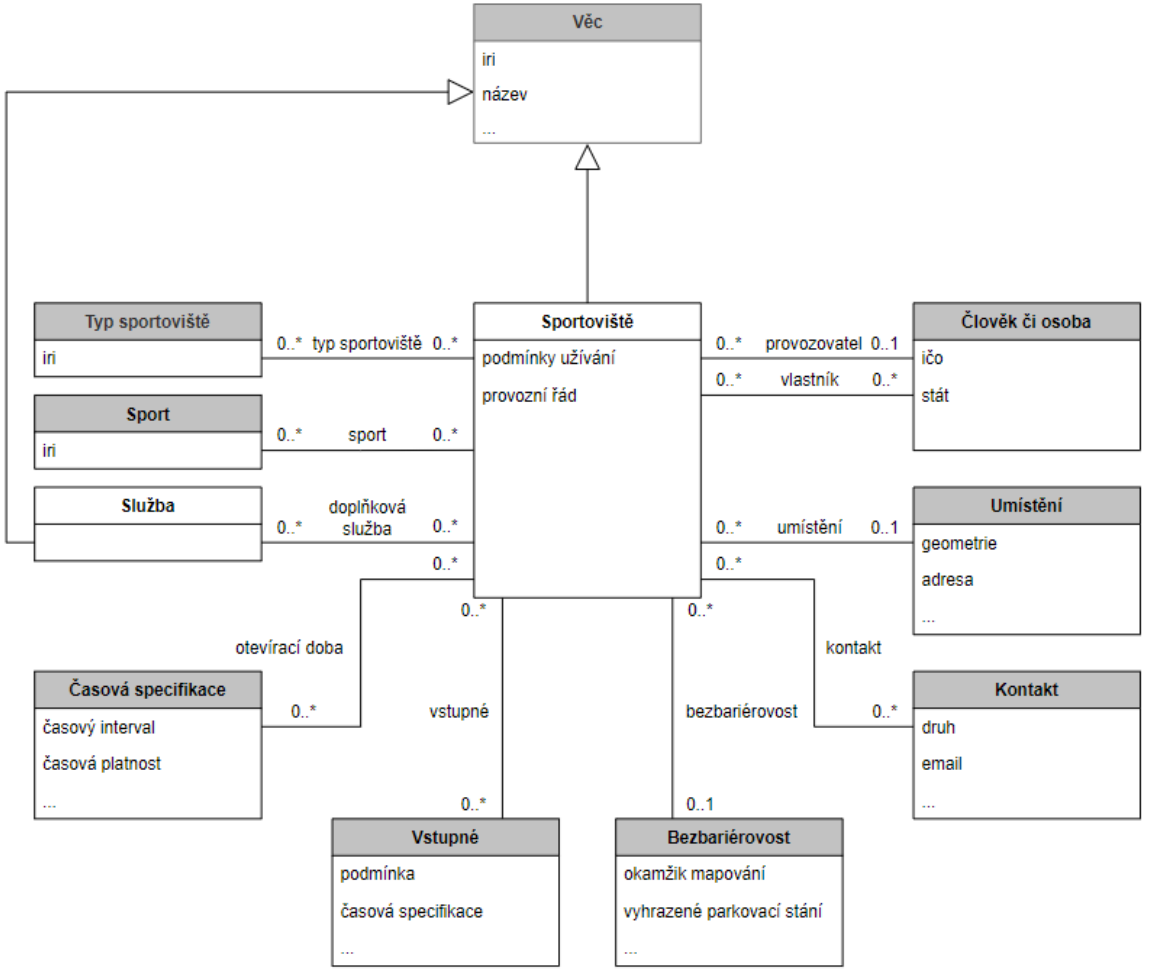
- A key step towards an effective solution is **standardization**

- Providers publish the same types of data in the same way, which significantly simplifies the processing of their data sources



OPEN_FORMAL_STANDARDS_

- The aim of the standards:
  - Unify the factual **content** (conceptualization)
  - Unify the **structure** and **format** of data
- A conceptual data model in the form of a diagram that describes concepts and the relationships between them is included



DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

OPEN_FORMAL_STANDARDS_

- Official notice board
- Code lists
- News
- Events
- Tourist attractions
- Sports facilities
- Waste collection centers
- Recycling bins

- Mandatory data structure
- Interoperability ensured
- Users can use the data more easily
- Mostly focused in municipalities

Standards - ofn.gov.cz

THANK YOU!

otevrenadata@dia.gov.cz

lenka.kovacova@dia.gov.cz

DIGITÁLNÍ
A INFORMAČNÍ
AGENTURA_

data.gouv.fr

Data quality enhancement at data.gouv.fr

A quick tour of our tools and methods

Pierlou Ramade

Data engineer at data.gouv.fr

March 3rd 2025

What is data.gouv.fr?

data.gouv.fr

- Open and communautary platform which aims at **centralizing and structuring French (open) data**
- Raw data that can be **freely downloaded and reused** by everyone
- 60k datasets, more than 300k files, across many topics and from more than 5k producers, including administrations (ministries, local entities…), private actors and citizens
- An active community of 130k users, interacting through thousands of discussions and publishing their reuses (websites, apps, dashboards…)

Our 3 main goals:

Simplify the discoverability of French public data

Improve open data quality

Foster the reuse of open data

Improving data quality before publication: creating standards

data.gouv.fr

- schema.data.gouv.fr: a catalog of French (public) data standards
- standards ideally follow international guidelines (TableSchema, JSONSchema…)
- standards are created:
  - by entities who have use cases and technical abilities to formalize them
  - by our team following a request for standardization of an entity that doesn't have the technical assets to formalize it

  always with as many involved actors as possible
- recently, some French laws have been requiring actors to publish their data (repairability index and other kinds of reporting); these **laws contain the URLs to the dedicated publication standards**

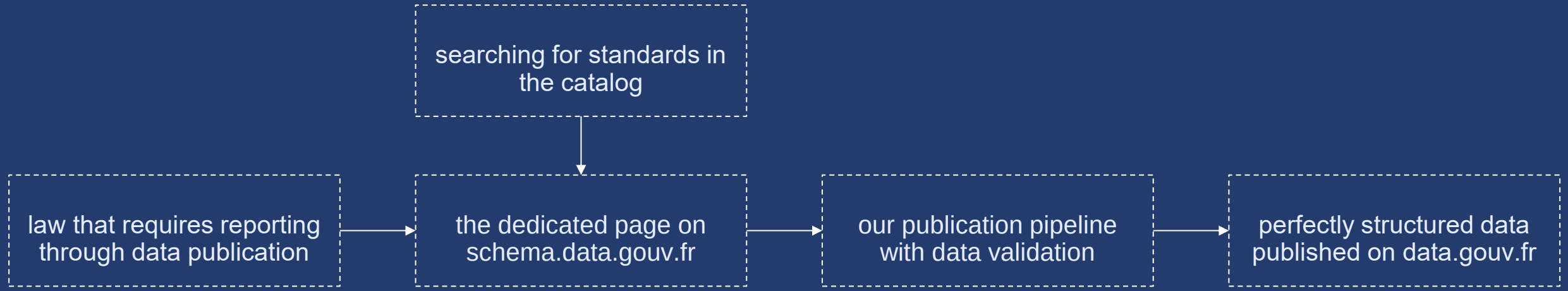
Push producers towards good practices

data.gouv.fr

1. Capitalize on standards: data publication tool

- create a clear publication funnel: wherever the producers land, they will eventually end up on the right page to validate and publish their data
- have validation tools in the back, with user-friendly layouts that encourage producers to fix the potential errors in their data

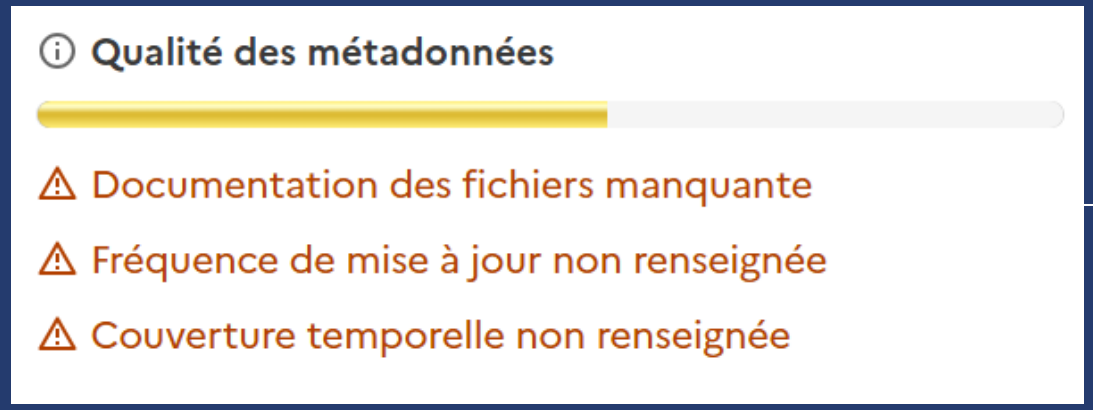
Our funnel looks like this:

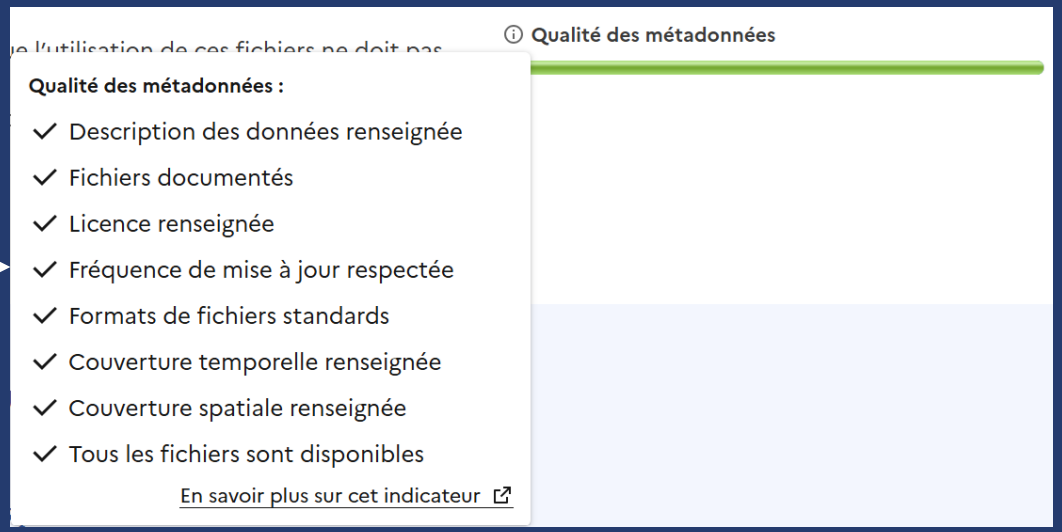


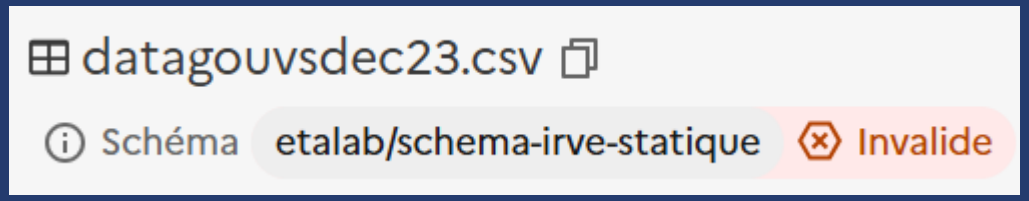
Push producers towards good practices

data.gouv.fr

2. People don’t like red: a metadata quality score







Reprocess published data

data.gouv.fr

1. When published data do not meet minimal quality standards and are considered of high interest, we proactively process and expose them alongside the original source, to facilitate their reusability

Examples:

A dataset is updated each month with a file that contains the previous month's data

↓

We publish an aggregated file

Files are not properly organized (missing headers,uneven structure across occurrences…)

↓

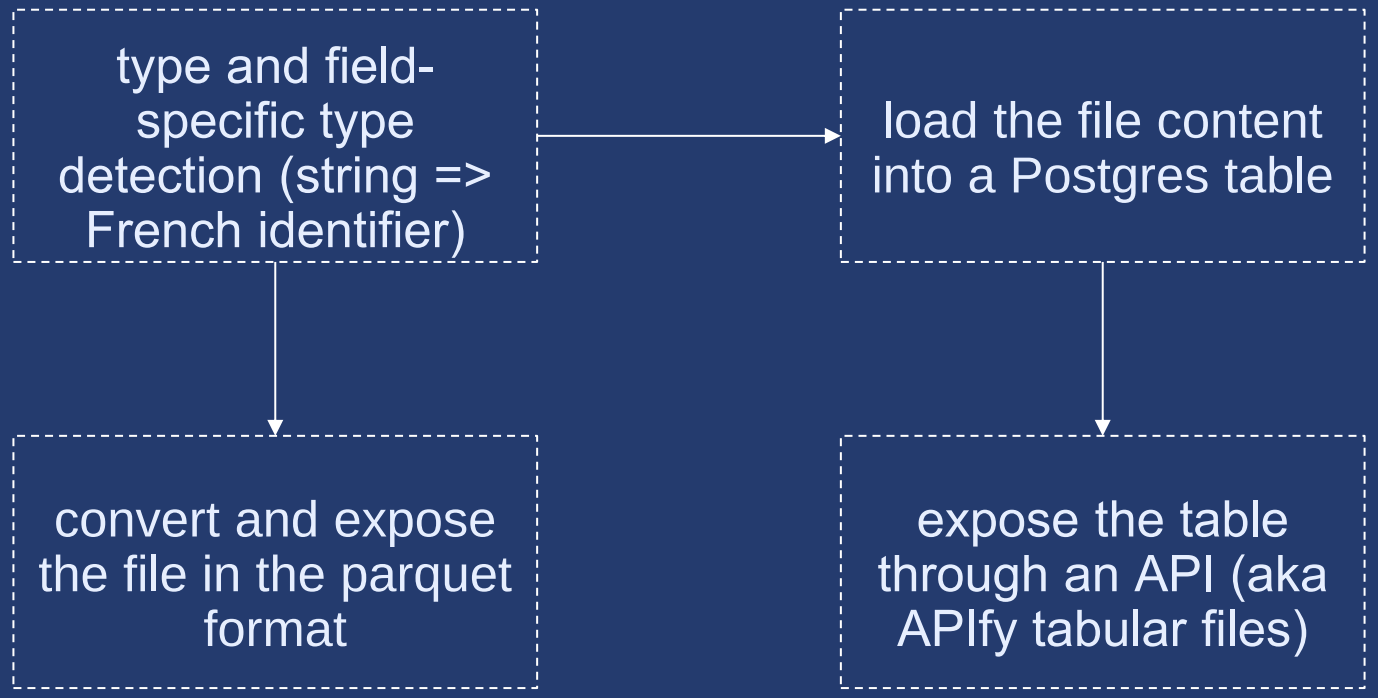
We publish a cleaner version

Our main ETL tool

Reprocess published data

data.gouv.fr

2. For well-structured tabular files (csv, xls-like), we have a automatic processing pipeline that does:



Working towards metadata interoperability

data.gouv.fr

There are many existing open/geographical data platforms in the French ecosystem: by themes, territories, publishers, etc.

- For years, there has been a massification of DCAT as the main interoperable standard used in the open data context

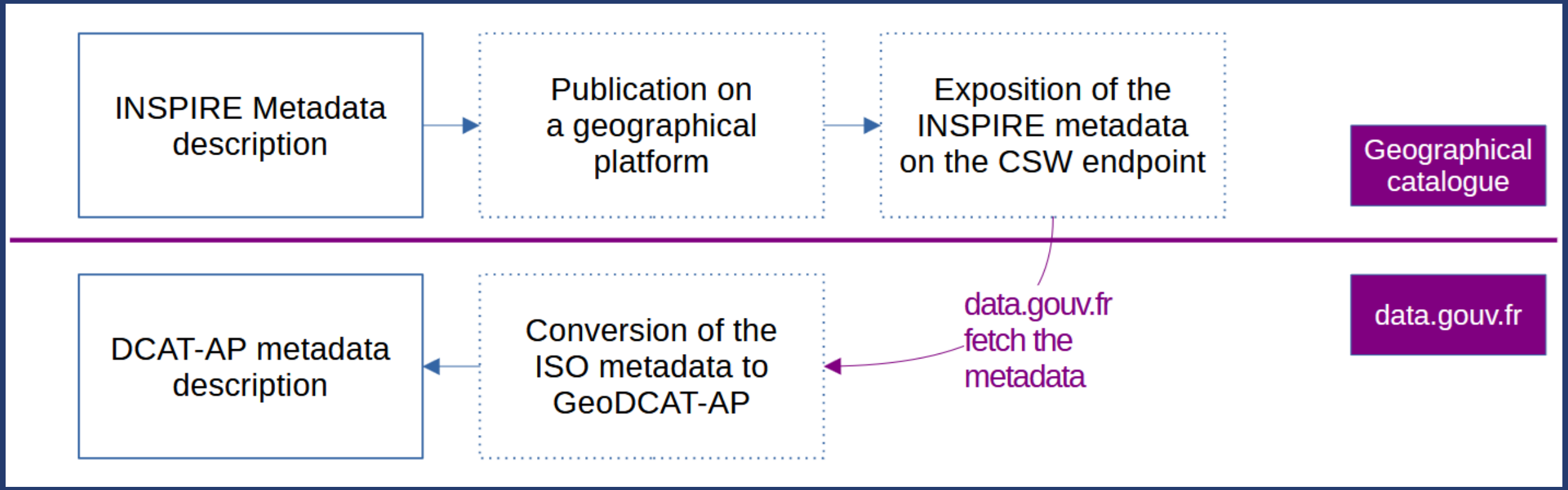
- data.gouv.fr is meant to be the national platform for open AND geographical data: all French open data have to be referenced on it

- Geographical metadata interoperability with open data standards used in France (DCAT) was unsatisfactory at the French level until recently

Working towards metadata interoperability

data.gouv.fr

Usage of the SEMIC ISO INSPIRE → GeoDCAT-AP XSLT:



Working towards metadata interoperability

data.gouv.fr

- Most platforms are now harvested by the french national platform (data.gouv.fr) in (Geo)DCAT-AP 🚀
  - Almost ~400 harvested sources
  - The French national platform re-exposes all these metadata description as DCAT-AP

- We are now focusing on:
  - the way licence, access and usage rights are mapped
  - Proper distributions identifications

- Our approach is to base our transformation on the XSLT and to fix metadata upstream
  - by giving clear recommendations
  - adding tooling to use small and specialized XSLT

    > Isomorphe: https://github.com/ecolabdata/ecospheres-isomorphe

    >XSLTs:https://github.com/ecolabdata/ecospheres-xslt

- A common place to share these specialized XSLT between member states may be useful

Q&A

Inmaculada Farfan Velasco
data.europa.eu,
Publications Office of the European Union

Bert van Nuffelen
Semantic Data Interoperability Expert,
Digital Flanders

Lenka Kováčová
National Open Data Coordinator,
Digital and Information Agency Czech Republic

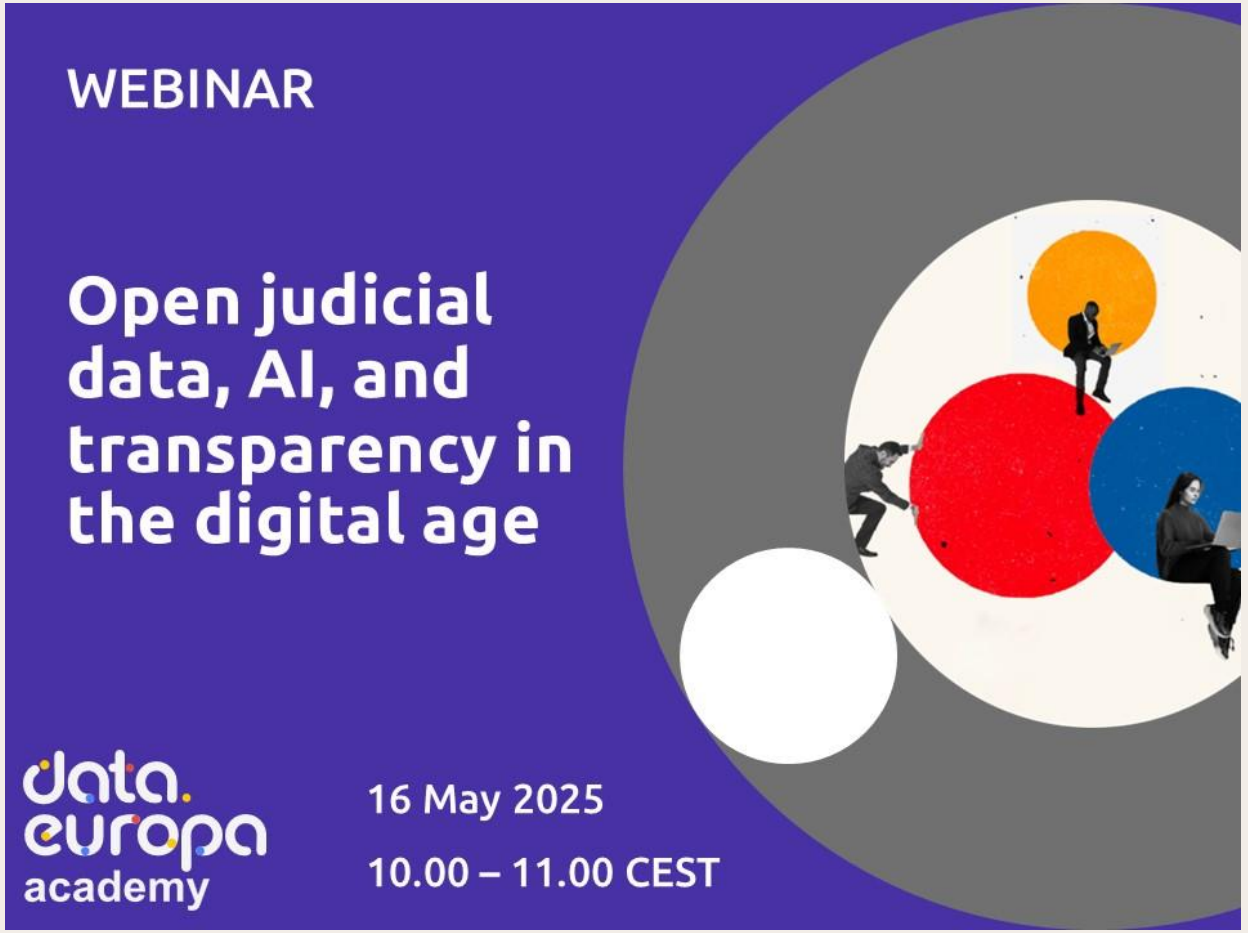
Pierlou Ramade
Data Engineer,
data.gouv.fr

data europa academy

Stay up-to-date on our
2025 activities!

data.europa academy

Register now for our next webinar!



data europa academy

Your opinion is important to us!

data.
europa
academy

Thank you!

data europa academy