EU OPEN DATA DAYS

Georges Lobo

Programme manager – SEMIC Team Leader

Pavlina Fragkou

Project manager, European Commission

SEMIC: An introduction to DCAT-AP and LDES

19 – 20 MARCH 2025

What is SEMIC?

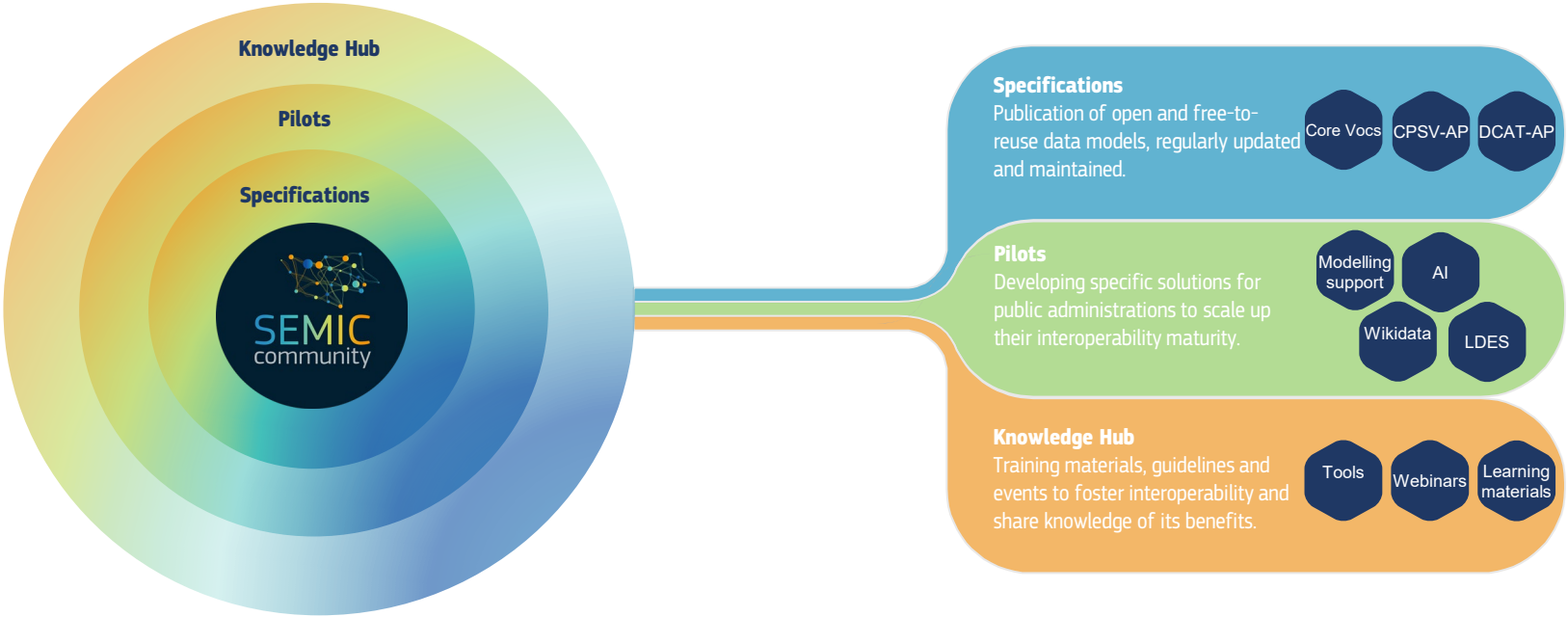
Introduction to SEMIC

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

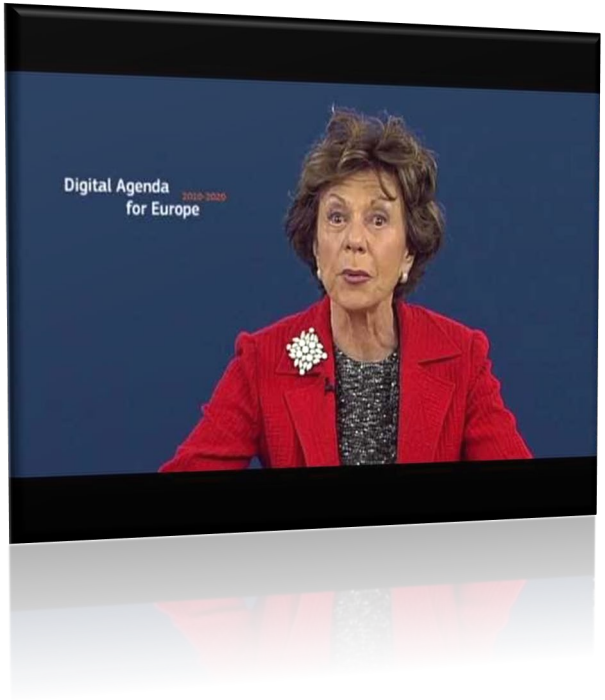
- Promoting the share and reuse of semantic assets, experience and tools and facilitating agreements in key areas.
- Identifying opportunities for alignment on semantic definitions, metadata and reference data sources with special focus on identification and definitions of Core Concepts / Vocabularies.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



DCAT-AP origins & key objective



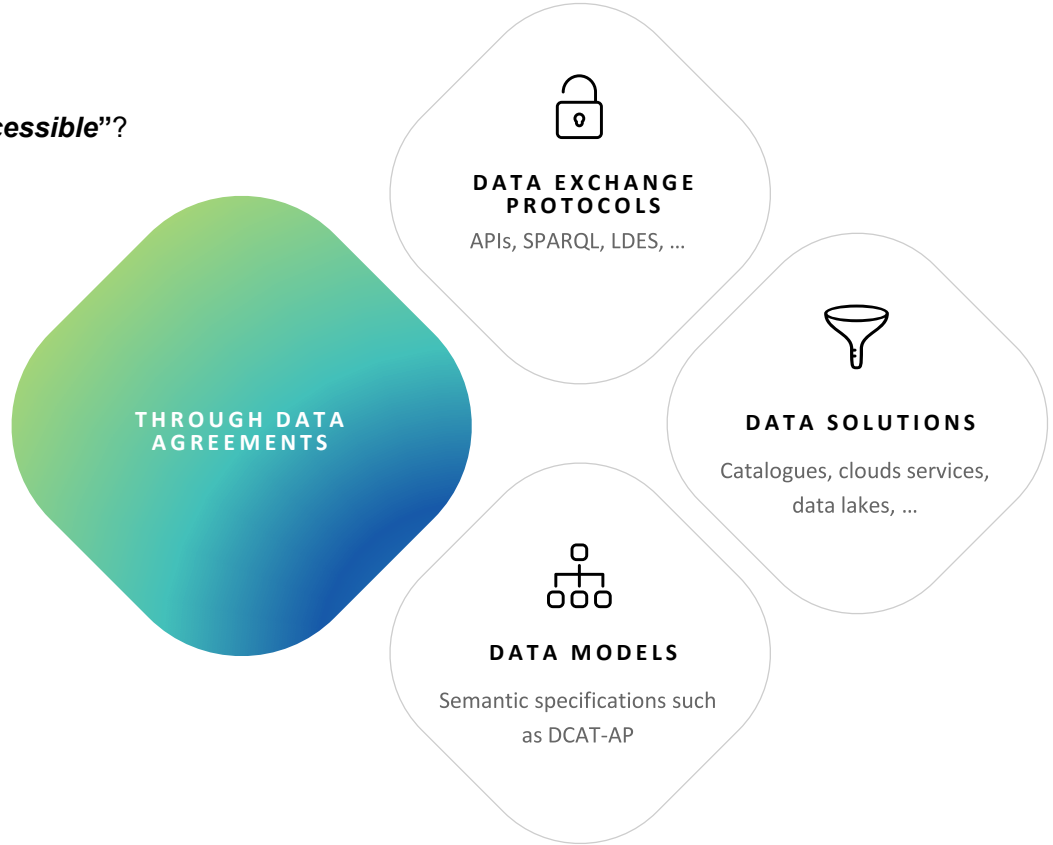
"Data in the new Oil" – Neelie Kroes (**2012**)

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

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

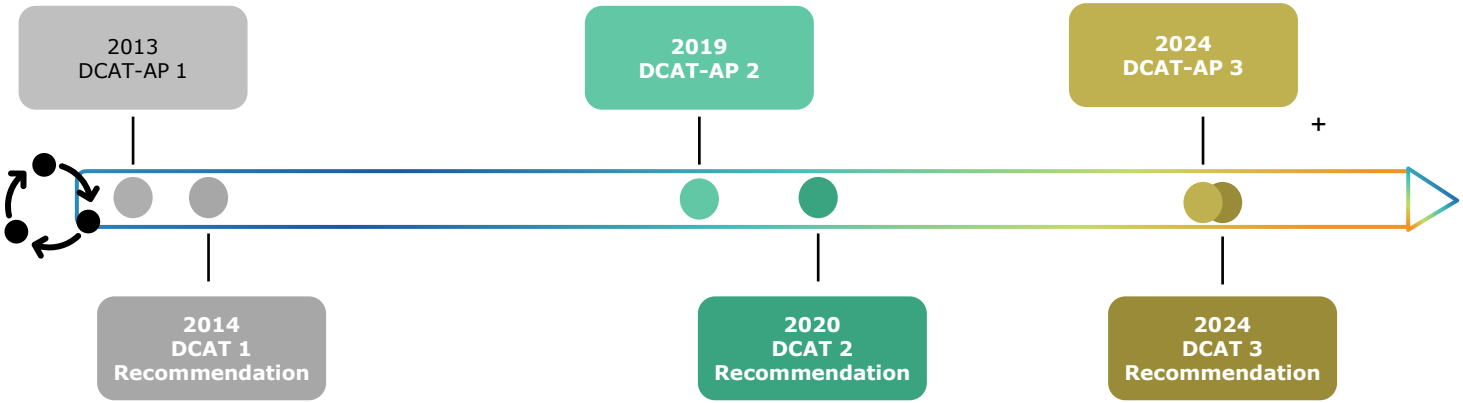
How?

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



DCAT (W3C) and DCAT-AP (EU)

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



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

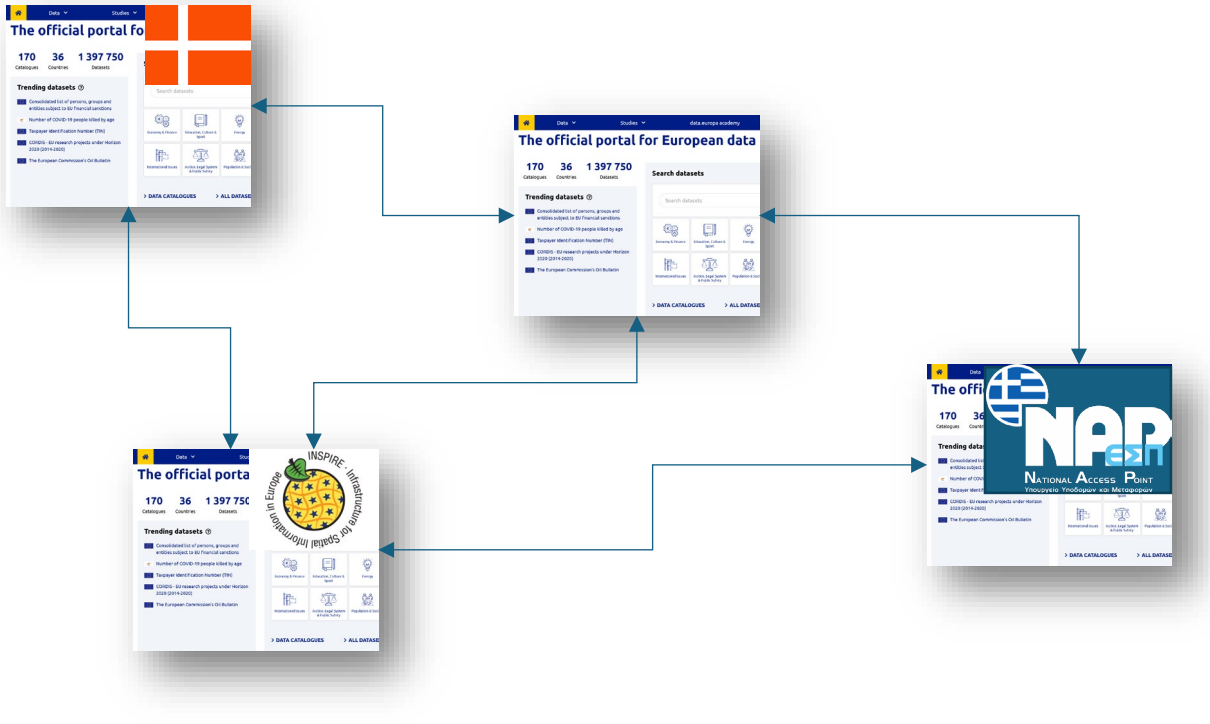
DCAT

- A W3C application profile/vocabulary for facilitating interoperability between **data catalogs** published on the Web.
- Is a metadata standard: namely describing information about the shared entities.
- https://w3c.github.io/dxwg/dcat/

DCAT-AP

- DCAT-AP provides a common specification for describing public sector datasets
- DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogs published on the Web.
- DCAT enables a publisher to describe datasets and data services in a catalog using a standard model and vocabulary that facilitates the consumption and aggregation of metadata from multiple catalogs.
- This can increase the discoverability of datasets and data services.

Data Catalogues are interconnected



Objectives of DCAT-AP

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.

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 applications

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

Base registries metadata descriptions

Data spaces

- MobilityDCAT-AP
- HealthDCAT-AP
- ...

Machine Learning with MLDCAT-AP

Benefits of DCAT-AP

Strategic benefits

- Able to base yourself on best practice
- Less time/effort spent on the development of your own semantic assets
- Possibility to make your own national or domain specific extensions
- Supports intra-domain and cross-domain interoperability
- Create common understanding of core elements

Technical benefits

- Fully aligned with semantic technologies (semantic web, linked data, …)
- Expressions in RDF & JSON-LD
- Validation service based on SHACL
- Automated processing of data

Proven track record

W3C STANDARD

Using DCAT (W3C) as a basis

MATURE IMPLEMENTATIONS

Such as data.europa.eu

COMMUNITY

Vibrant DCAT-AP community

- GitHub DCAT-AP
- Webinars
- Interoperable Europe Portal

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

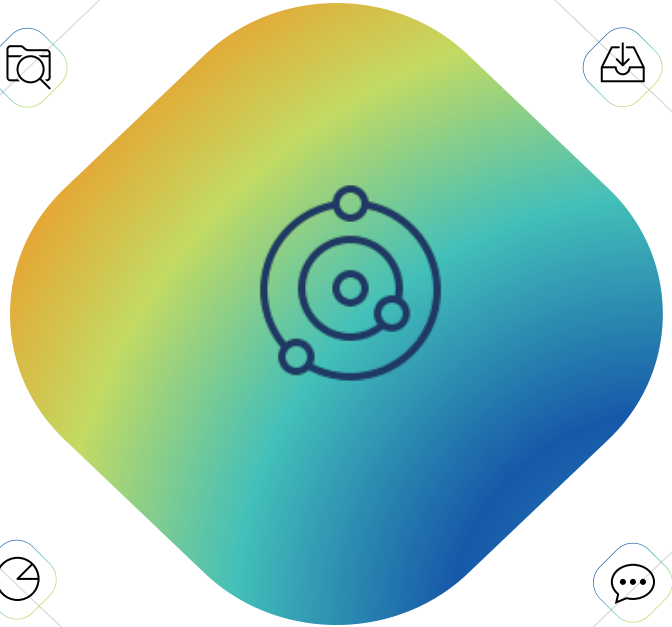
Possibility to go beyond the common specification and create domain-specific data space extensions in a stackable way

Resulting in:

Stable and long-term management of metadata and data

Reduced amount of metadata management

Benefits of the DCAT-AP ecosystem



Enhances the findability and accessibility of data

Enables data spaces to make **data catalogues findable**
→ A harvesting network is made possible

Comes with a decade of **experience** of documenting, maintaining metadata records; sharing through harvesting, etc.

Enables data spaces to express their **metadata in a common language**

Provides tooling to **validate the implementation** data.

Collaborative environment that allows data spaces to express their needs and additional requirements (specialisations)

DCAT-AP a growing ecosystem

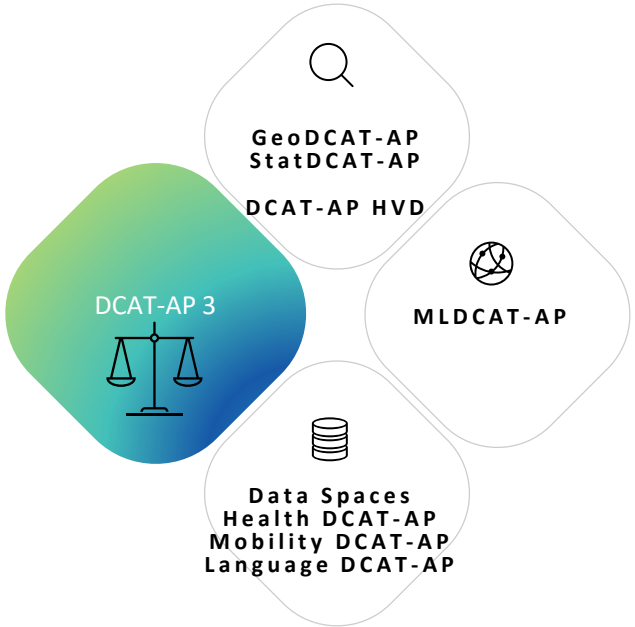
Supporting the discovery of/access to (open) data in a cross-border and cross-domain environment, by describing metadata to be harvested across a distributed network of portals.

Open community: webinars

Interconnecting communities

Support for profiling

Support for publishers and portals



COMMUNITY

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

MATURE IMPLEMENTATIONS

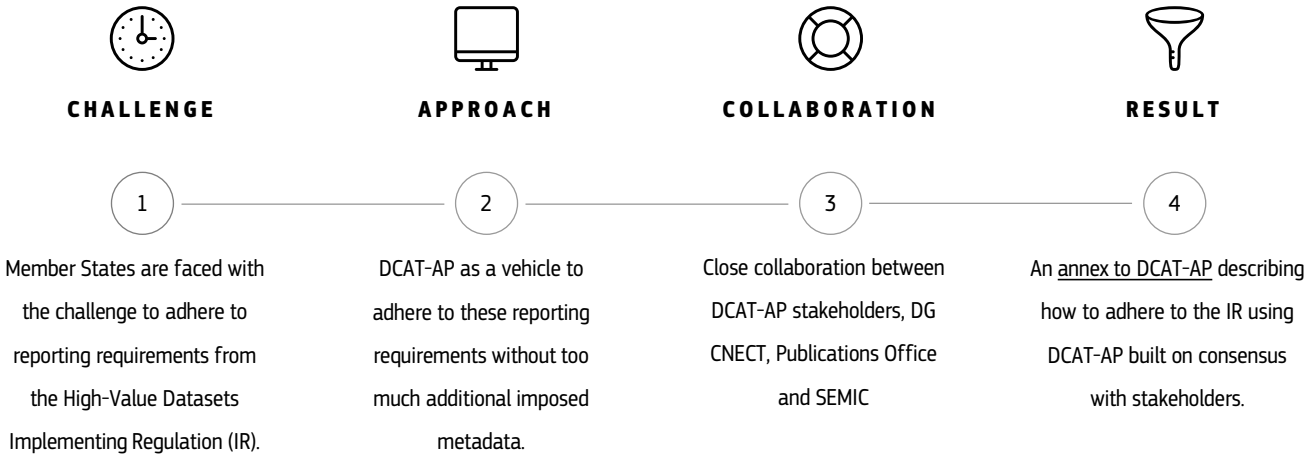
Such as data.europa.eu

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

Community & collaboration

- Strong open community
- Both DCAT and DCAT-AP are publicly accessible for anyone on the web
- Anyone can contribute through GitHub issues
- DCAT-AP offers discussion webinars in which topics are addressed

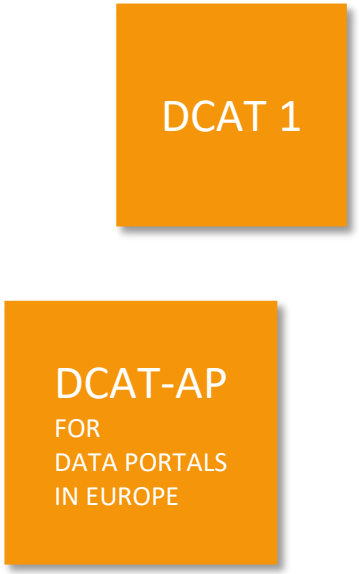
What do we do? DCAT-AP HVD as an example



DCAT-AP profiles

A growing ecosystem

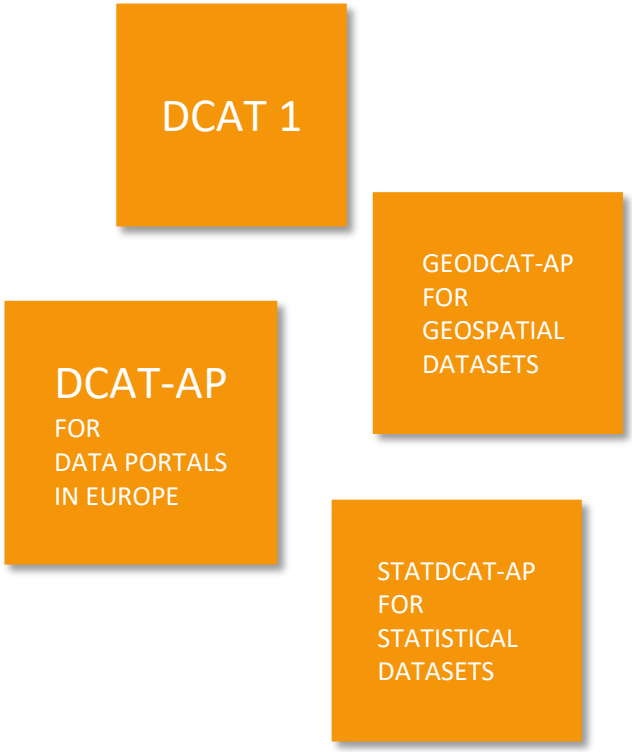
2012



DCAT-AP profiles

A growing ecosystem

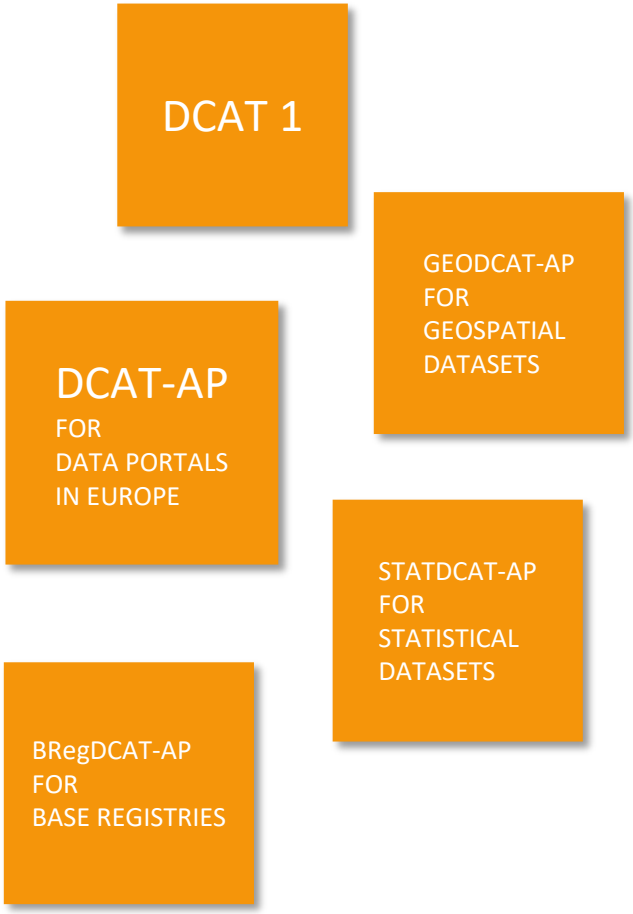
2015



DCAT-AP profiles

A growing ecosystem

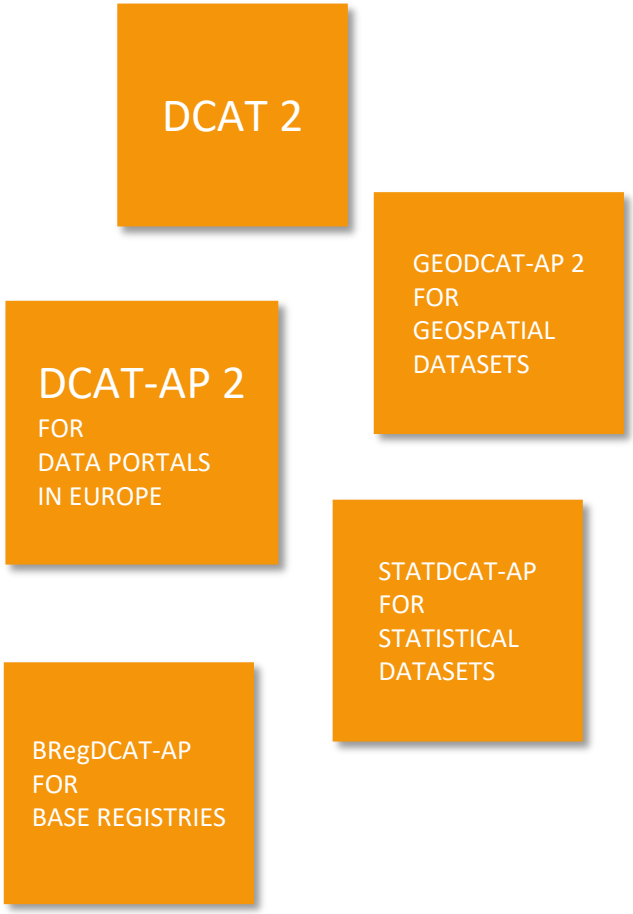
2018



DCAT-AP profiles

A growing ecosystem

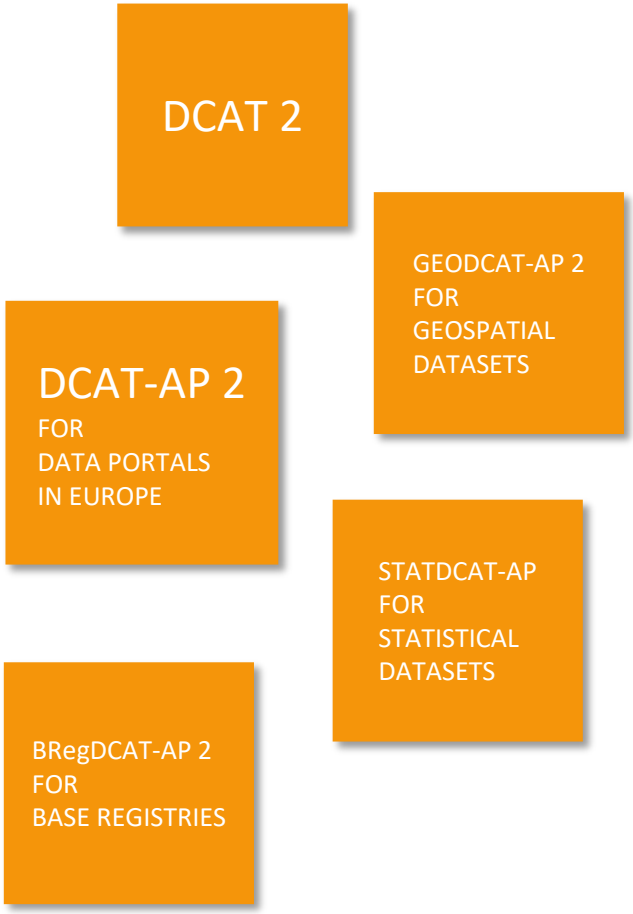
2020



DCAT-AP profiles

A growing ecosystem

2021



DCAT-AP profiles

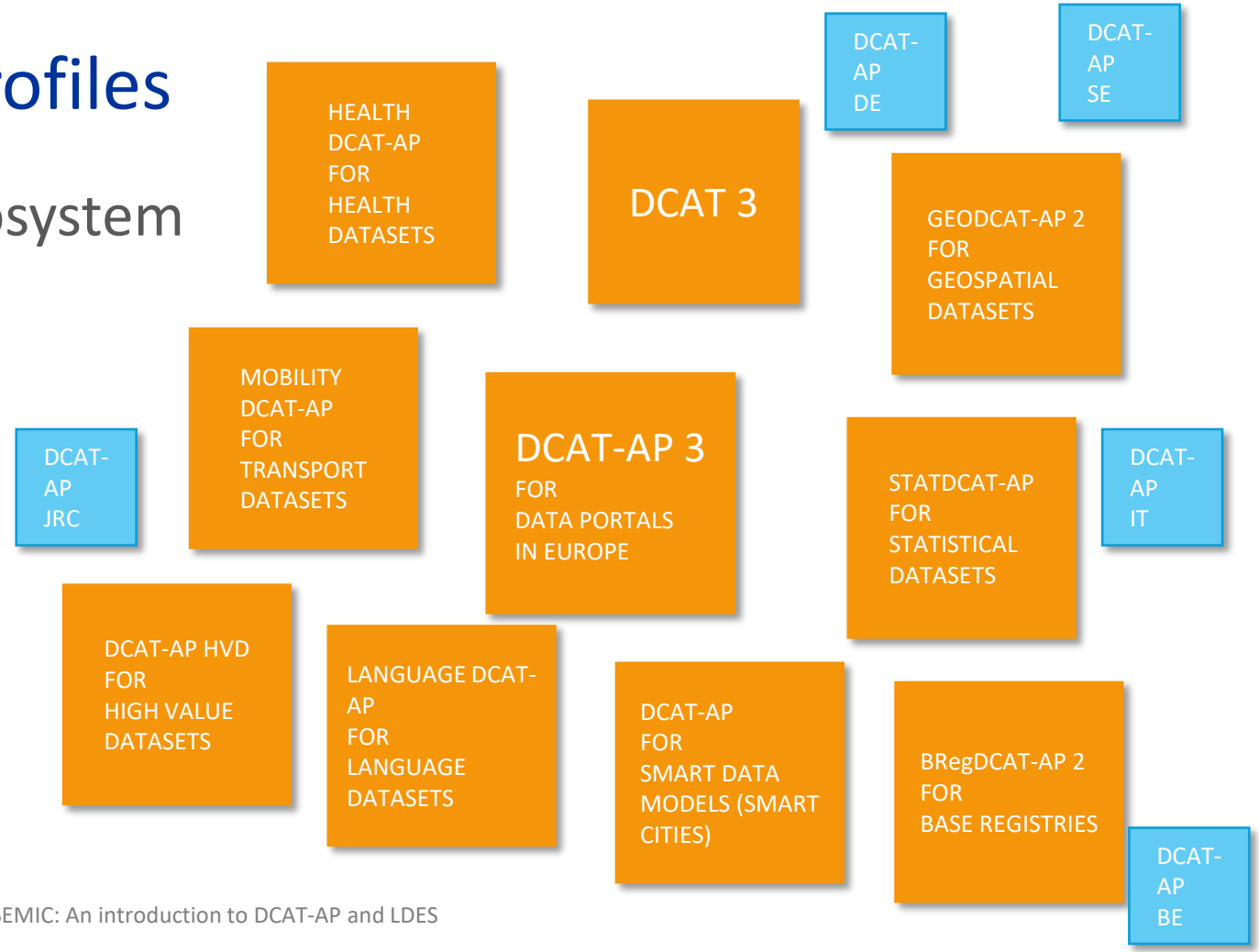
A growing ecosystem

2023

DCAT-AP profiles

A growing ecosystem

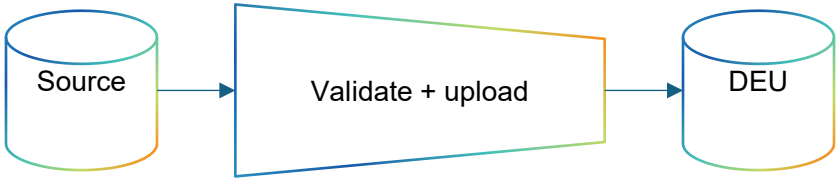
2023



Data.Europa.eu

The official portal for European data

The largest aggregated collection of European datasets.

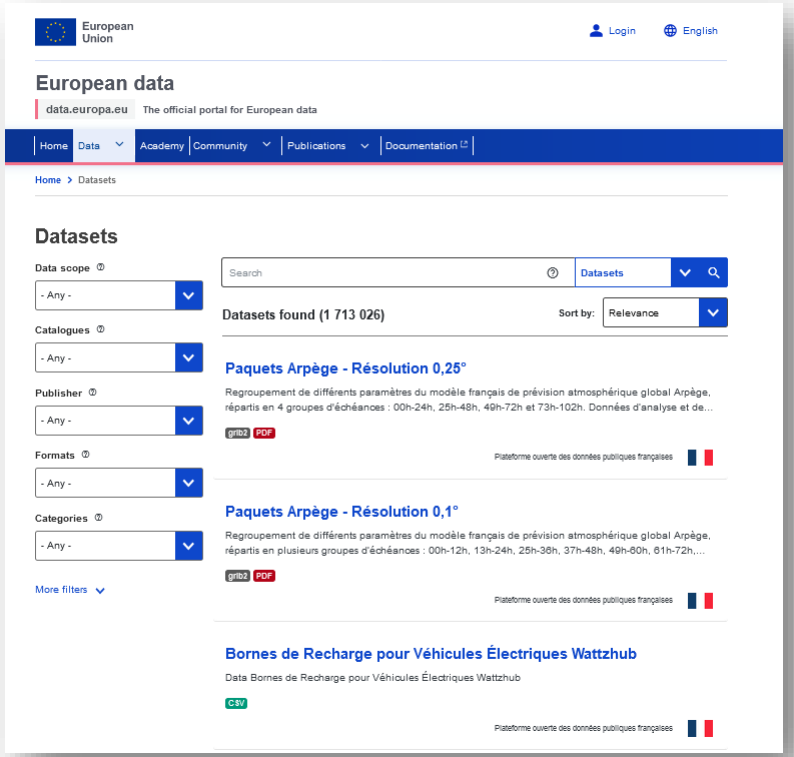


DCAT-AP
(RDF)

Open Source Software

Steps:

1) Provide metadata information about the catalogue to be harvested to DEU technical team either manually (via a web form) or through an API
2) The DEU technical team will initiate the harvester for your catalogue (when and API is provided)
3) The catalogue owner can follow-up the harvesting with the DEU team.



https://data.europa.eu

Open challenges

- Reporting obligations from MS raising from High Value Datasets
  - Use the cross-domain metadata catalogue data.europa.eu (DEU) as source of the HVD reporting.
- Future of Geo-DCAT-AP and Inspire Directive (ISO Standard)
- Update of STAT-DCAT-AP
- Proposed governance document of DCAT-AP variations
- CEN-CENELEC involvement on DCAT-AP governance under the context of data spaces raising from the European Data Act.

How can we share DCAT-AP data? LDES

A Linked Data Event Stream (LDES)

A publication technology to share or aggregate information with or from multiple parties

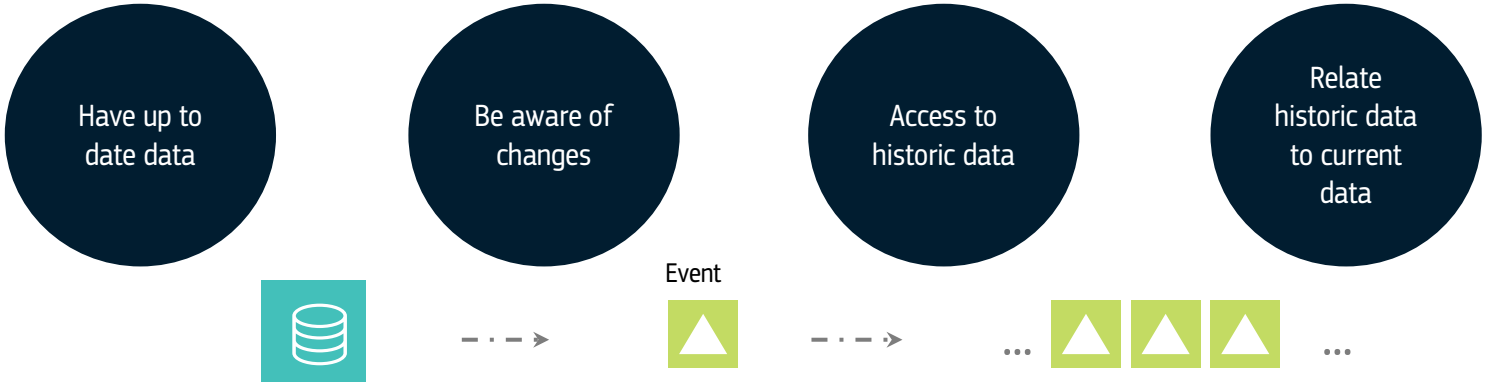
Allowing everyone to replicate and stay up-to-date regarding the unique source of truth

What is a LDES?

A Linked Data Event Stream (LDES) is a collection of immutable objects whereby you do not change the data itself but simply add new data records to the stream. It represents a publication strategy to publish and make data discoverable in a cost-effective and flexible manner.

LDES helps you to structure your data as stream data, enabling you and your users to keep track of what changed at the data level, independently from the data format.

It allows data users to:



What is a LDES? - example

Classical model

Data supplier

2019

Name	Living place
Willem	Paris

copy →

2020

Name	Living place
Willem	Paris
Ann	Lyon

copy →

2021

Name	Living place
Willem	Strasbourg
Ann	Lyon

copy →

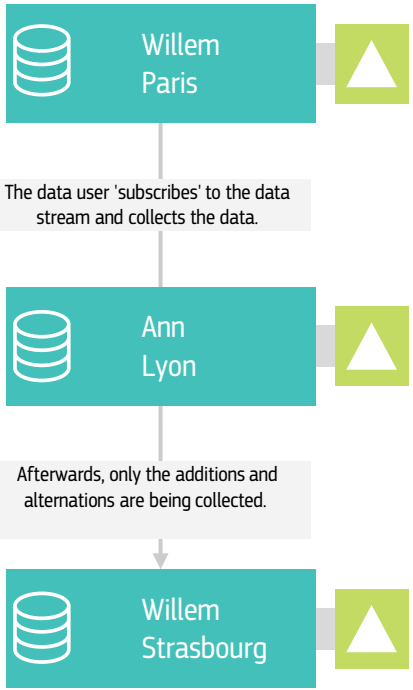
Data user

Name	Living place
Willem	Paris

Name	Living place
Willem	Paris
Ann	Lyon

Name	Living place
Willem	Strasbourg
Ann	Lyon

Linked Data Event Stream



EU-wide multimodal travel planning

EU regulation* on multimodal travel information services:

Provision of accurate and reliable travel information for travelers across the Union.

Ensure the optimal use and full interoperability of EU standards (NeTEx and SIRI) between Member States.

Each Member State shall set up a national access point.

COMMISSION DELEGATED REGULATION (EU) 2017/1926

of 31 May 2017

supplementing Directive 2010/40/EU of the European Parliament and of the Council with regard to the provision of EU-wide multimodal travel information services

(Text with EEA relevance)

* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R1926

National Access Points (NAPs)

- A report on all known NAPs (until 11/22) is available online*.
- Current NAPs suffer from the "data portal syndrome": Data discovery and integration remains eminently manual work.
- Additional efforts are being taken to align and coordinate NAPs: https://napcore.eu/

Map source: https://nap.cnadnr.ro/index.php

* https://transport.ec.europa.eu/transport-themes/intelligent-transport-systems/road/action-plan-and-directive/national-access-points_en

Cross-border travel planning in the EU

Automatic data exchange

Harmonizing data models

Scalable interfaces for data access

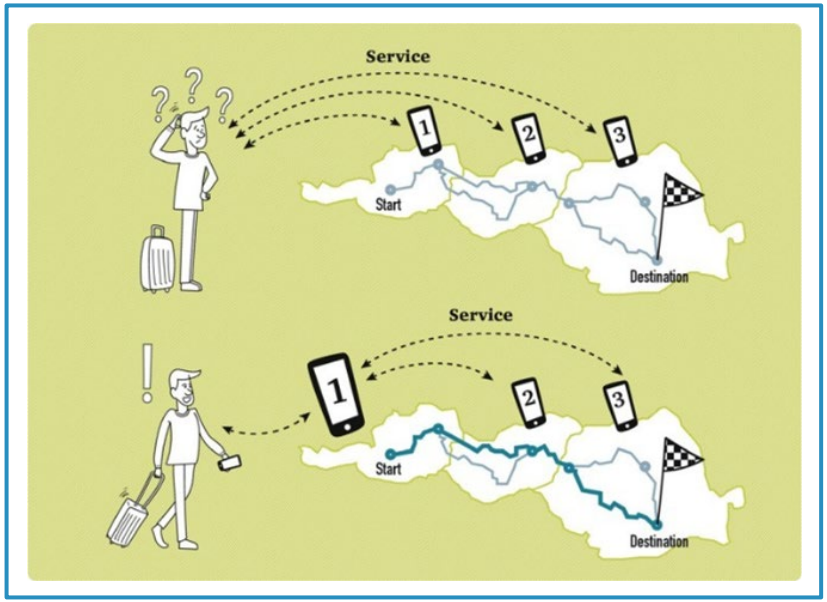
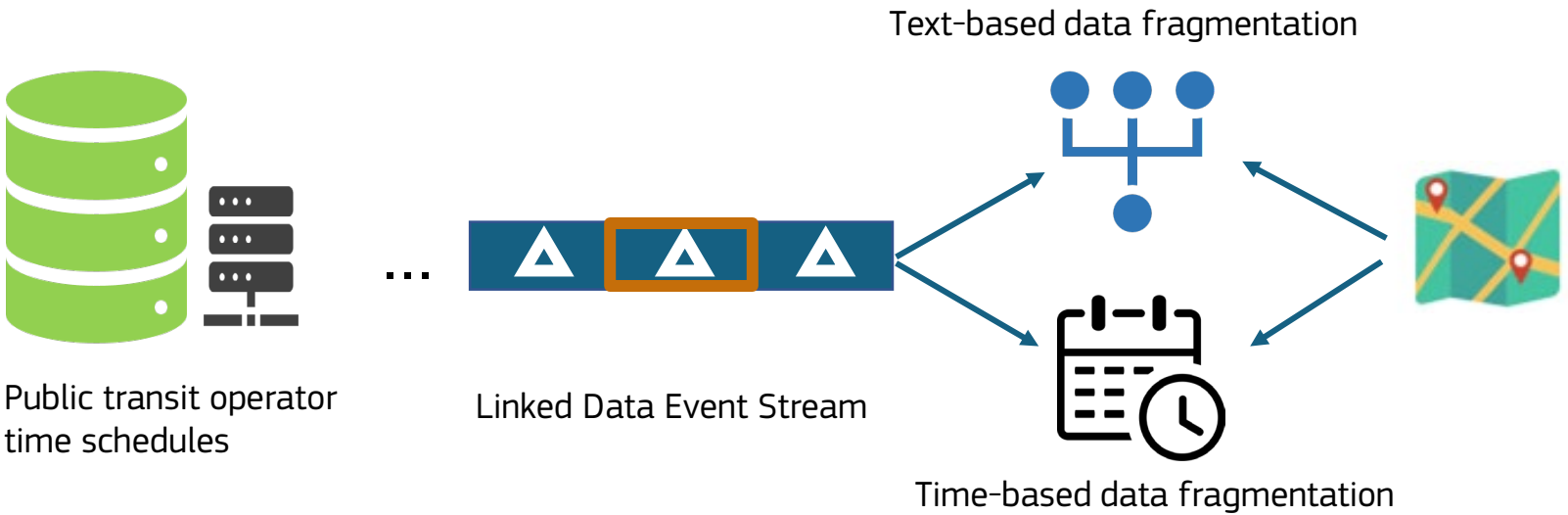


Figure source: https://www.linkedin.com/pulse/how-enable-eu-wide-multimodal-travel-information-s%C3%B8ren-s%C3%B8rensen/

Data architecture based on LDES



A semantic data model

JSON-LD context

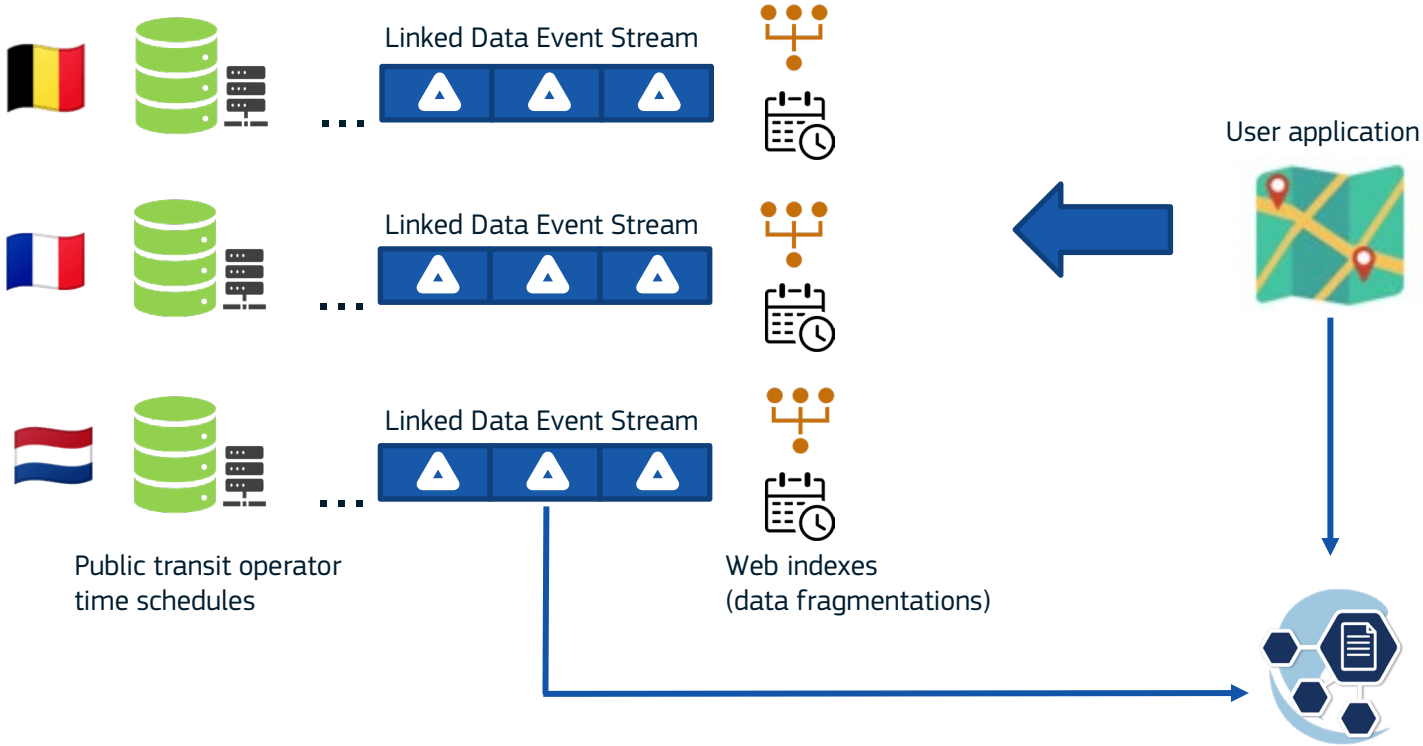

```
{
  "@context": {
    "lc":
"http://semweb.mmlab.be/ns/linkedconnections#",
    "gtfs":"http://vocab.gtfs.org/terms#",
    "xsd":"http://www.w3.org/2001/XMLSchema#",
    "Connection" : "lc:Connection",
    "departureStop": { "@id": "lc:departureStop",
"@type":"@id"},
    "arrivalStop": { "@id": "lc:arrivalStop", "@type":"@id"},
    "departureTime": {"@id":"lc:departureTime",
"@type":"xsd:datetime"},
    "arrivalTime": {"@id":"lc:arrivalTime",
"@type":"xsd:datetime"},
    "departureDelay": {"@id":"lc:departureDelay",
"@type": "xsd:integer"},
    "arrivalDelay": {"@id":"lc:arrivalDelay", "@type":
"xsd:integer"},
    "gtfs:headsign": {"@type":"xsd:string"},
    "gtfs:trip": {"@type": "@id"},
    "gtfs:route": {"@type": "@id"},
    "gtfs:pickupType":{"@type": "@id"},
    "gtfs:dropOffType": {"@type": "@id"}
  }
}
```


Linked Connection JSON-LD example

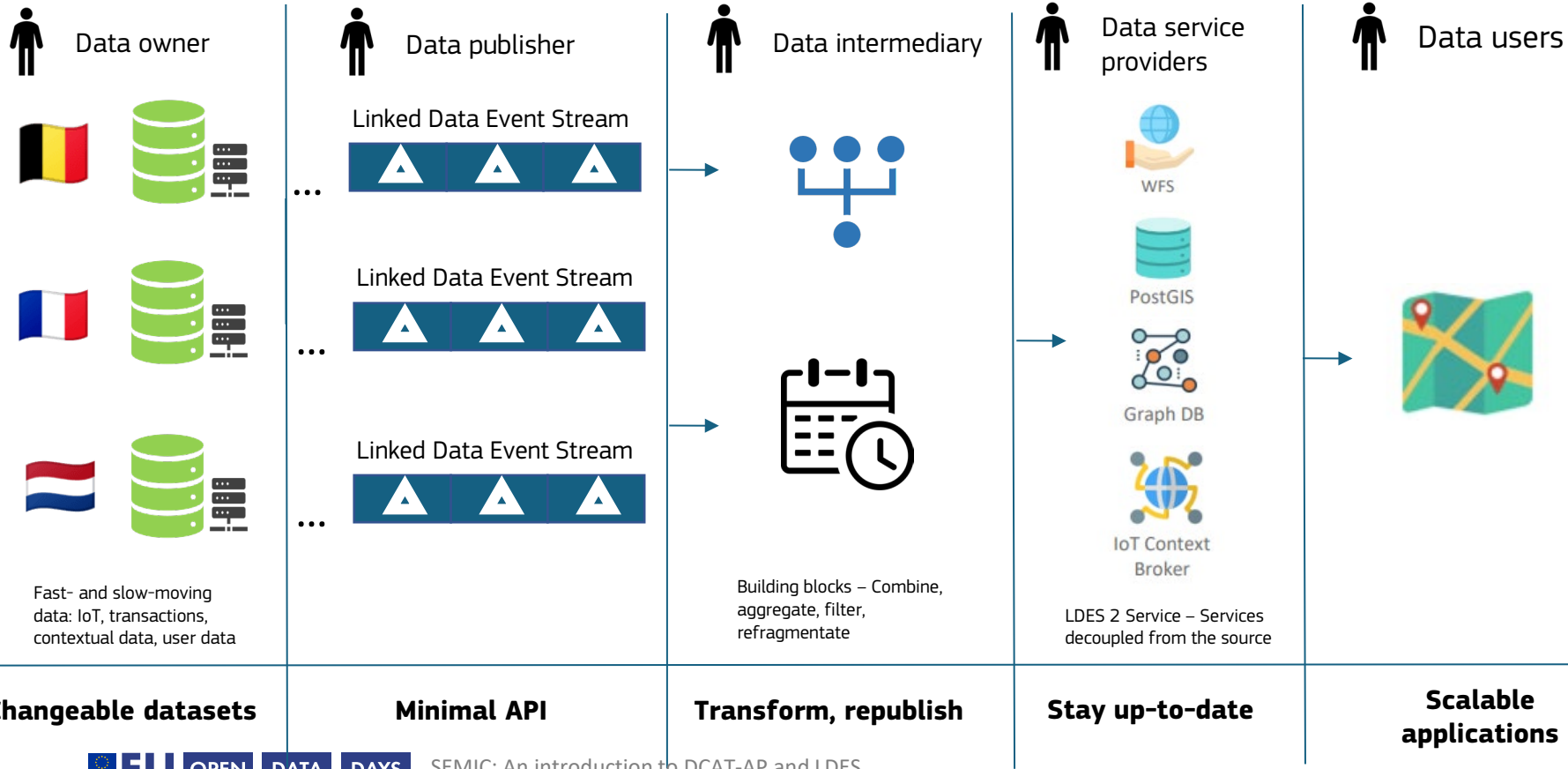

```
{
  "@id":
"http://irail.be/connections/8863008/20171219/P8671",
  "@type": "Connection",
  "departureStop":
"http://irail.be/stations/NMBS/008863008",
  "arrivalStop":
"http://irail.be/stations/NMBS/008863354",
  "departureTime": "2023-04-19T15:50:00.000Z",
  "departureDelay": 60,
  "arrivalTime": "2023-04-19T16:20:00.000Z",
  "arrivalDelay": 60,
  "gtfs:trip": "http://irail.be/vehicle/P8671/20171219",
  "gtfs:pickupType":"gtfs:Regular",
  "gtfs:dropOffType":"gtfs:Regular"
}
```


Linked Connections

Data architecture based on LDES



Data architecture based on LDES



Problems when publishing data

Synchronise data with users

Different clients request different release schedules.

Collect or distribute

Do you need different approaches to collect data and then distribute it again?

Users

Do you have too many users for your current setup?

Is it difficult to discover your data?

How can user explore the data without opening it?

Different usage Profiles?

Are you under pressure to provide use different publishing technologies for different stakeholders?

Existing set up?

Are you dependent on your current system and afraid to change?

Users request older versions?

Some Data Users prefer to work with an older data set or want to roll back to a previous state

Is collecting and updating data from multiple sources a nightmare?

How to achieve interoperability with other data sources and platforms?

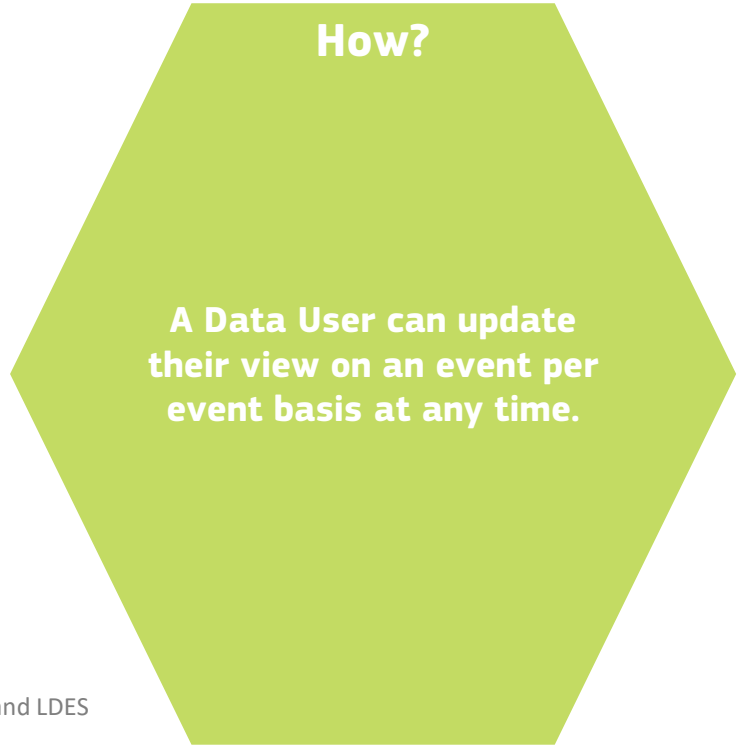
Update speed

Data Users and Data Providers can independently access/change the data.

No effort required to synchronise with users

Update speed

Data Users and Data Providers can independently access/change the data.



Decentralised

Due to the decentralised nature and uniform approach, it can be used to collect data from multiple sources and share data to multiple users.

One approach for multiple usage types

Ideal approach for Collecting and sharing Base Registry Data

Cost effective scalability

With an LDES only one publishing system is required. Costs are independent from quantity of usage.

Reduced maintenance cost

Cost effective scalability

With an LDES only one publishing system is required. Costs are independent from quantity of usage.

Reduced maintenance cost

How?

Only one end point is required serving multiple users.

No need for multiple types of APIs.

A query on the data by the user no longer results in an additional API call.

Customisable by the data user

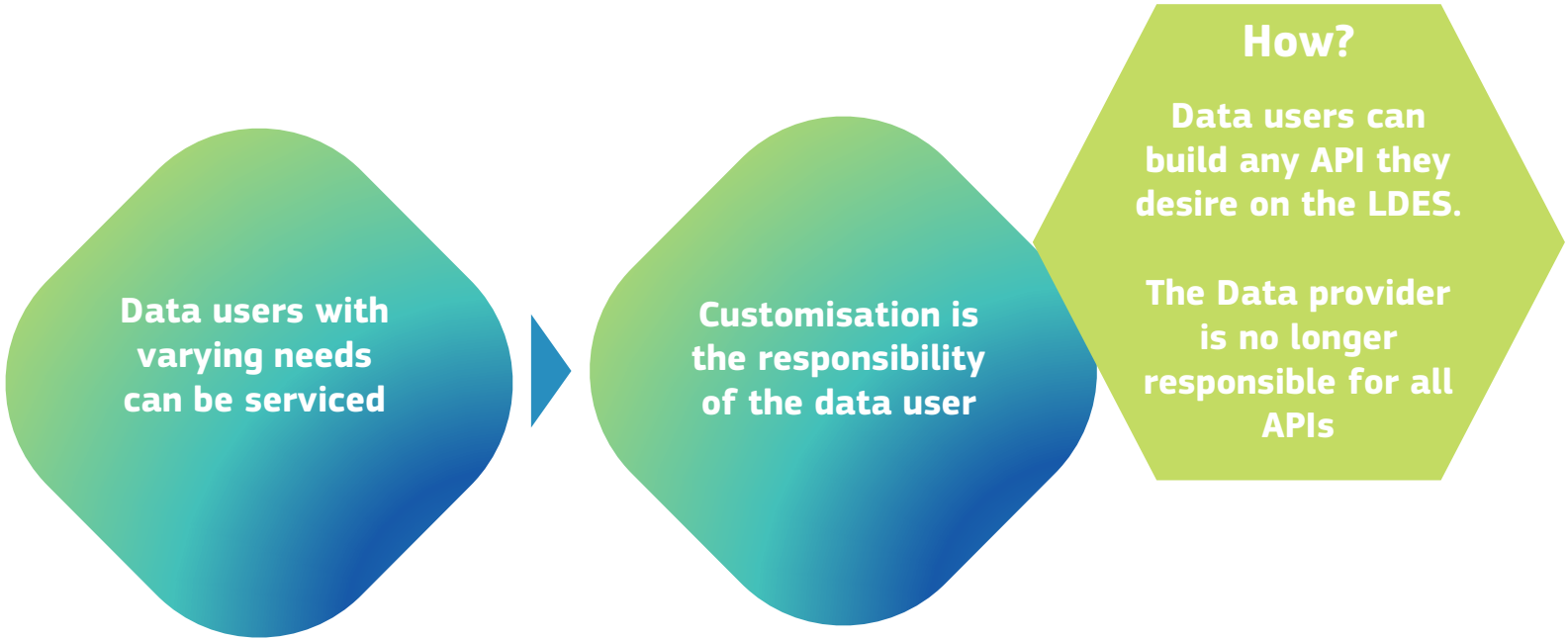
LDES allows to recreate the data. Allowing data users to attach intermediary publishing systems.

Data users with varying needs can be serviced

Customisation is the responsibility of the data user

Customisable by the data user

LDES allows to recreate the data. Allowing data users to attach intermediary publishing systems.



Extendable and Flexible

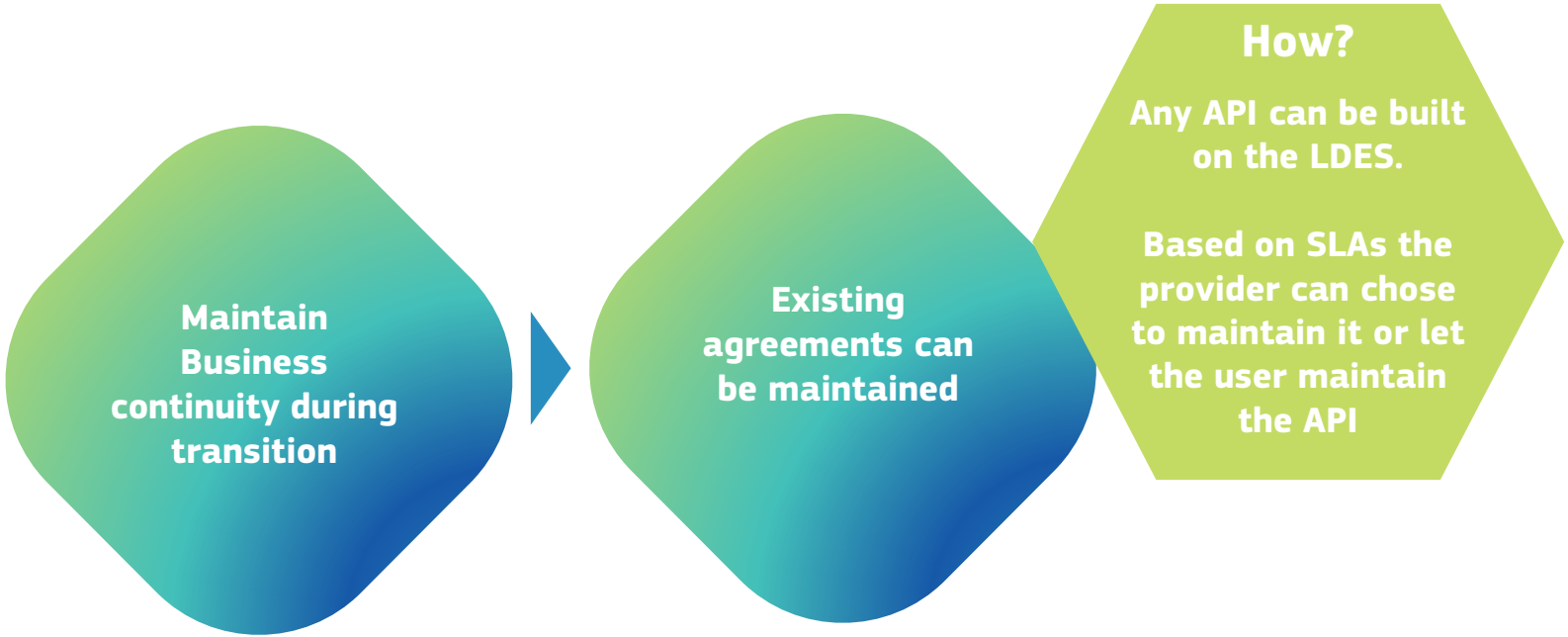
LDES can be build next to existing publishing systems. Existing publishing systems can rebuild/reconfigured on the LDES.

Maintain Business continuity during transition

Existing agreements can be maintained

Extendable and Flexible

LDES can be build next to existing publishing systems. Existing publishing systems can rebuild/reconfigured on the LDES.



Standards facilitate LDES rollout

LDES requires a common data structure, if this in place the effort required to implement an LDES is significantly reduced.

Reduced implementation cost

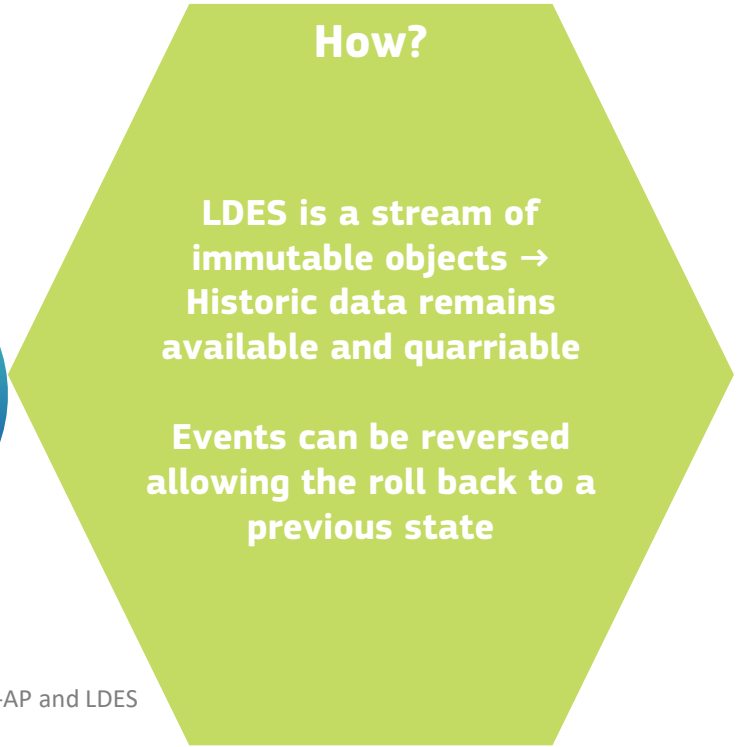
Historic Search and Roll Back

Due to the streaming of versions anybody can use the LDES to access a specific historic version of the data.

No separate access point needs to be maintained for the historic versions of the data

Historic Search and Roll Back

Due to the streaming of versions anybody can use the LDES to access a specific historic version of the data.



Discoverability

LDES is rich in meta data allowing the datasets to be more easily discovered using Data portals.

The right user will more easily find the correct Dataset for them

Discoverability

LDES is rich in meta data allowing the datasets to be more easily discovered using Data portals.

The right user will more easily find the correct Dataset for them

How?

Only the root node needs to be exposed it contains all meta data to navigate the LDES both manually and automatically.

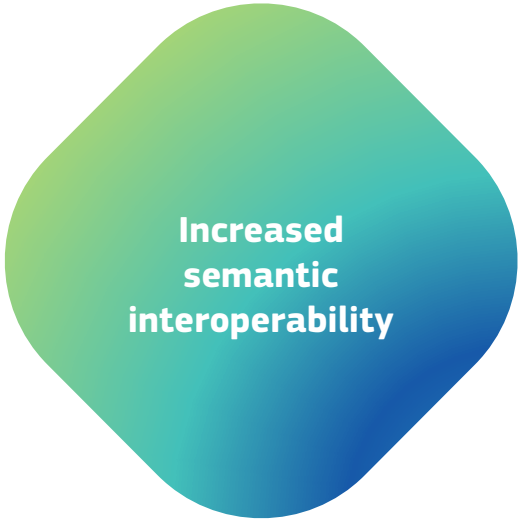
Semantic interoperability

A LDES is build on common Linked Data standards. Using these standards increases interoperability with other data platforms.

Increased semantic interoperability

Semantic interoperability

A LDES is build on common Linked Data standards. Using these standards increases interoperability with other data platforms.



How?

Reusing existing linked data models and concept to define the LDES data model establishes interoperability with systems reusing the same concepts.

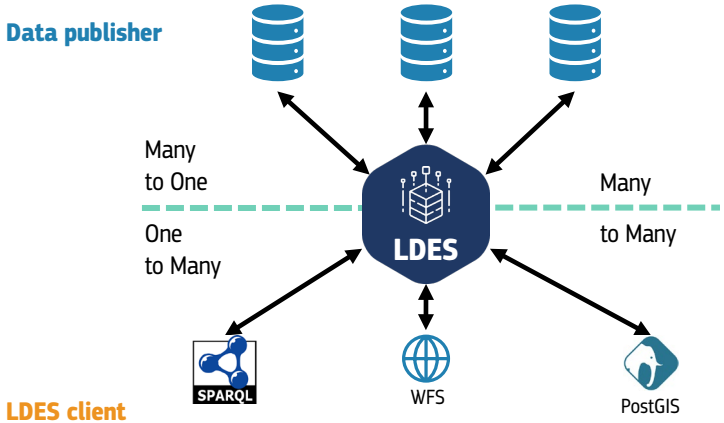
SEMIC advises the reuse of the SEMIC Core Vocabularies and Application Profiles

LDES: Pilot use cases

A decentralised application for planning cross-border public transit journeys in the EU based on official timetables.

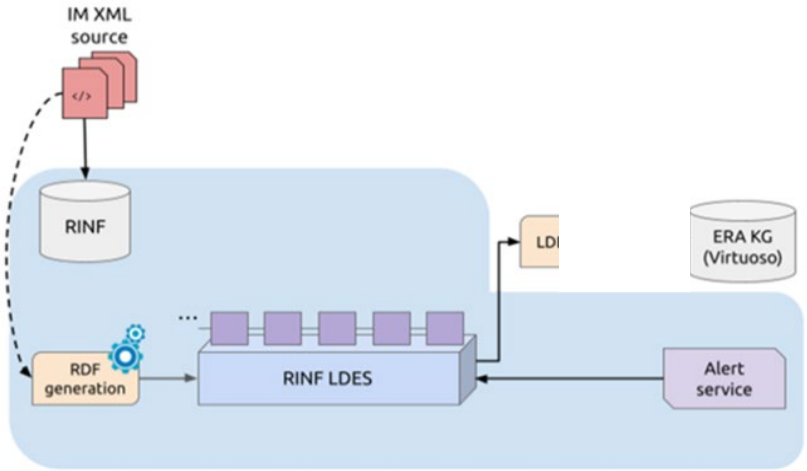
DCAT-AP feeds

- A new specification built on LDES which simplifies and standardises LDES' usage to DCAT-AP.
- Focussed on the aggregation and redistribution of DCAT-AP data
- Successful PoCs have been made for **Swedish** and **Belgian** registries
- Initial steps have been tested for LDES as a data harvester for **Data.Europa.EU**



LDES @ European Union Agency for Railways

- SEMIC supported ERA to create a Register of Infrastructure (RINF) LDES
- The RINF LDES allows for the notification of changes to infrastructure enabling informed and safe train operation across regions
- Due to the nature of LDES operators can only ingest the part of the LDES relevant to their routes
- ERA further rolled out LDES to support their ERA Knowledge Graph connecting further technical data to the system via LDES



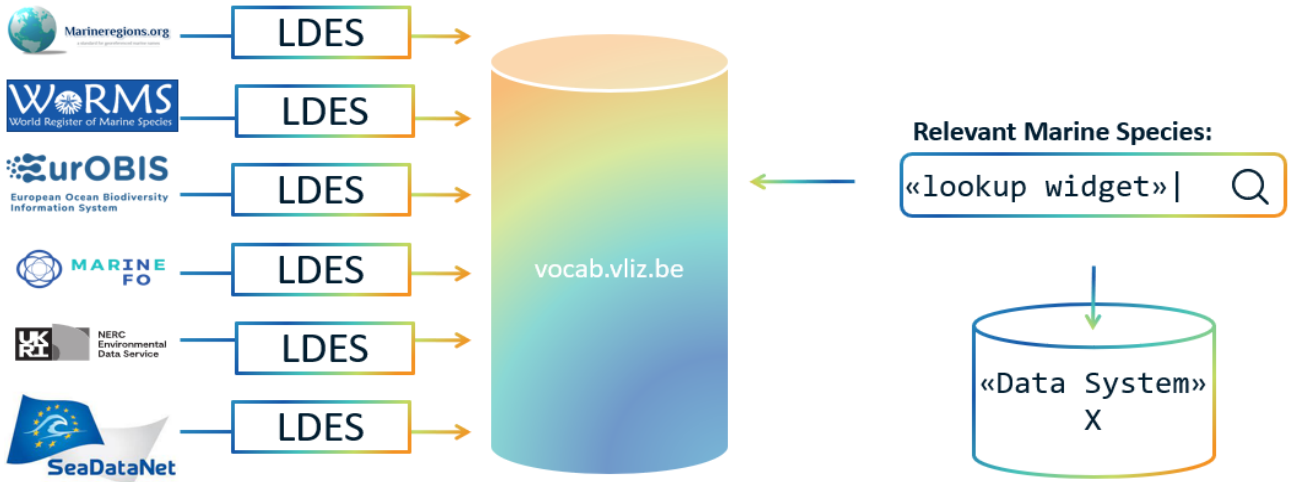
LDES @ MareGraph

- SEMIC supported the MareGraph project. MareGraph aims to connect existing data in the marine domain and make it available as Linked Open Data
- Using LDES to aggregate the data allows MareGraph to be synchronised and keep up to data with its sources

LDES for Maritime Vocabulary alignment

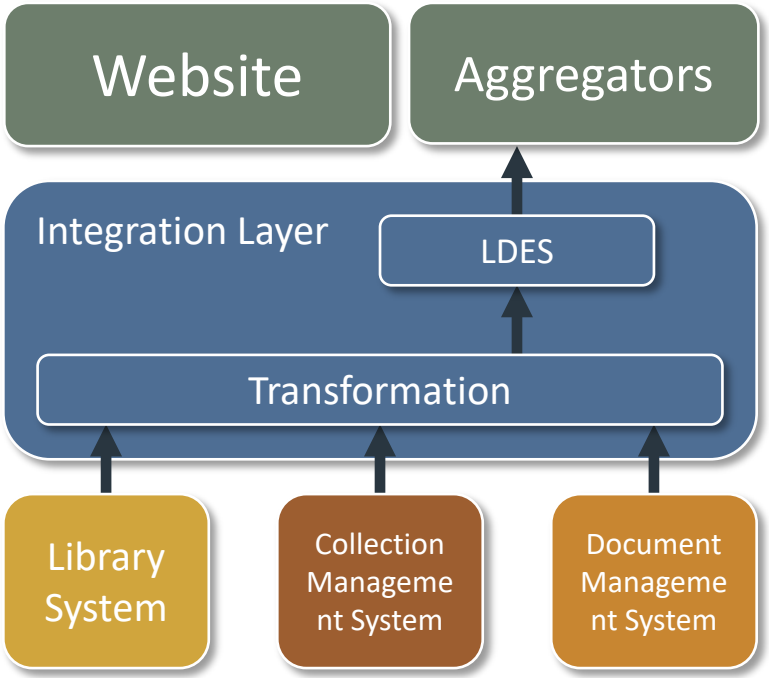
The Flemish Institute for the Sea (VLIZ) in collaboration with EMODnet, SeaDataNet and MareGraph created a Vocab Terms Service within the MArine domains

Using LDES VLIZ aggregates all vocabularies used by 6 authoritative sources in the domain to align the vocabularies used by them and enable the reuse of these existing Vocabularies



LDES for colonial cultural heritage

- SEMIC supported the Rijksmuseum and Against Opacity to support the sharing of cultural data with a colonial past.
- Against Opacity has the task of disclosing cultural artefacts with a colonial past in the Netherlands to allow former Colonies to reclaim their lost artefacts.
- The MetaData included in LDES allows Against Opacity to only harvest data on the relevant objects from the Rijksmuseum
- The equal distribution of efforts between publisher and aggregator allows LDES to be suitable for both big and small museums in NL, while remaining scalable for Against Opacity.



Will you be the next Pilot?

Contact the SEMIC team: anastasia.sofou@ext.ec.europa.eu

EU OPEN DATA DAYS

19 – 20 MARCH 2025

Thanks!


@ IoP_EU


Interoperable Europe


Interoperable Europe


ec.europa.eu/InteroperableEurope


IoP_Europe_COMM@ec.europa.eu