EC DATA CATALOGUE

a data discovery journey with Alma and Primo VE

Quinten Van Gorp, Data Librarian

IGeLU 2024, 10 September, Copenhagen

EC DATA CATALOGUE

- European Union
  - EU institutions
    - European Commission
      - Publications Office of the European Union
        - EC Library

EC DATA CATALOGUE

- What?
- Why?
- How?

Who we are: European Union

Who we are: European Union

27 Member States "united in diversity"

- 4 million km^2 (2023)

- 448.4 million inhabitants   (2023)

- 24 official EU languages

Who we are: EU institutions

Who we are: EU institutions

- 7 institutions

- 8 bodies

- > 30 agencies

- 4 interinstitutional services

Who we are: EU institutions

European Union

European Commission

European Parliament

European Council

Council of the European Union

Court of Justice of the European Union

European Central Bank

European Court of Auditors

Who we are: Publications Office of the European Union

Who we are: Publications Office of the European Union

Official provider of publishing services to all EU institutions, bodies and agencies.

Central point of access to:
- EU law
- Publications
- Data
- Research
- …

Who we are: Publications Office of the European Union

Its mission is to support EU policies as a **centre of excellence for information, data and knowledge management**, and to ensure that this broad range of information is available to the public as accessible and reusable data to facilitate transparency, economic activity and the diffusion of knowledge.
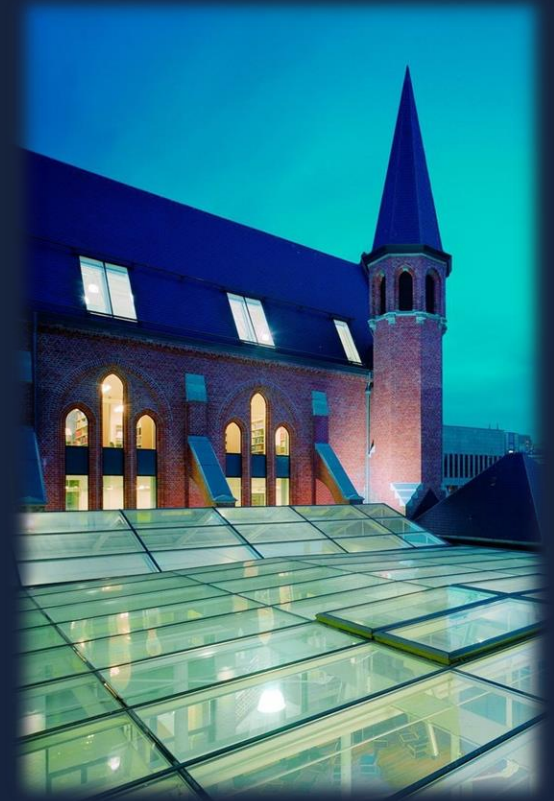
Who we are: EC Library

Who we are: EC Library

- Two sites (Brussels & Luxembourg)

- Serving +33.000 staff

- Interlibary loan services

- Open to external readers

Who we are: EC Library

Through its collections, and related services, specialising in European policy areas, the EC Library **connects Commission staff with data, information and knowledge** to support them in developing and delivering informed, **evidence-based policies** for the benefit of EU citizens whilst providing to all those interested a unique insight into the history of EU integration.

Data and information discovery tools

Find-eR

EC Data Catalogue

LibGuides

Data and information discovery tools

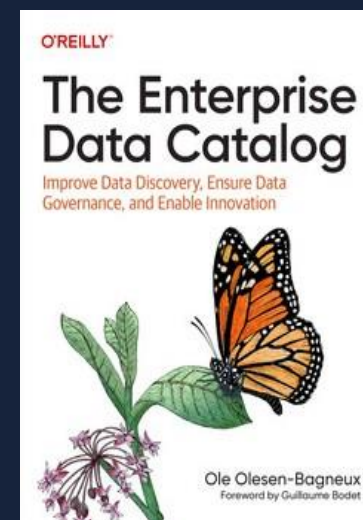Find-eR   EC Data Catalogue   LibGuides
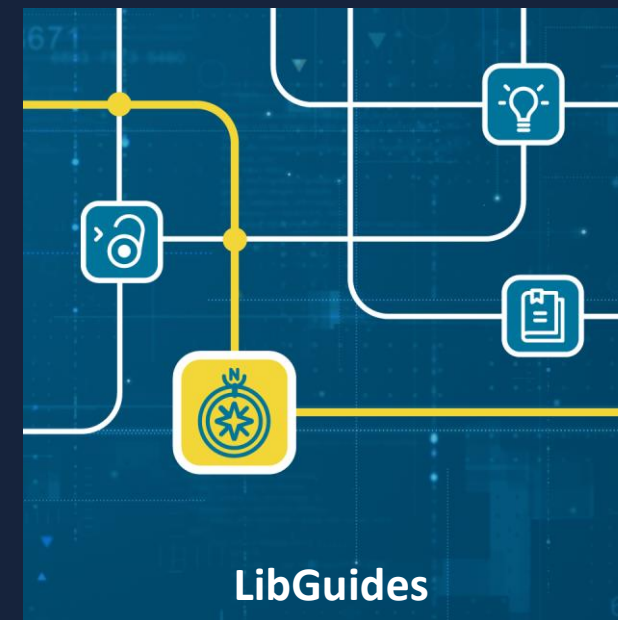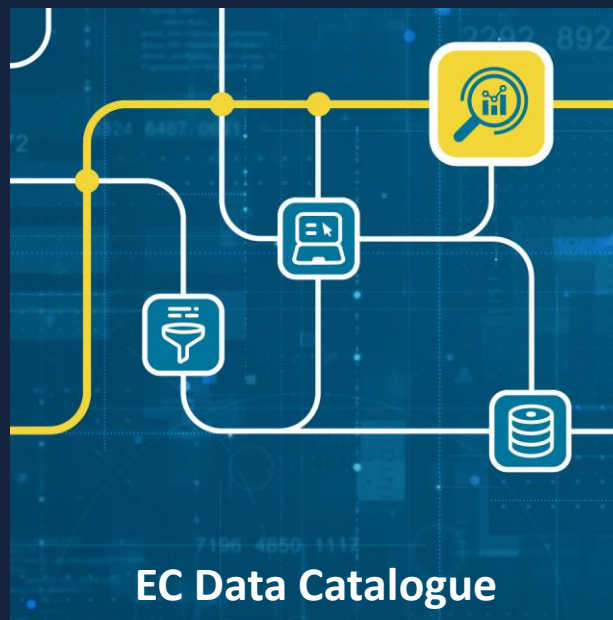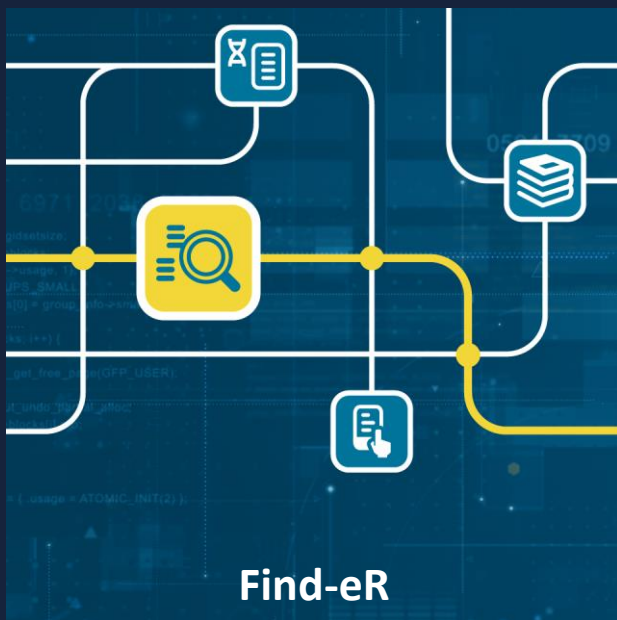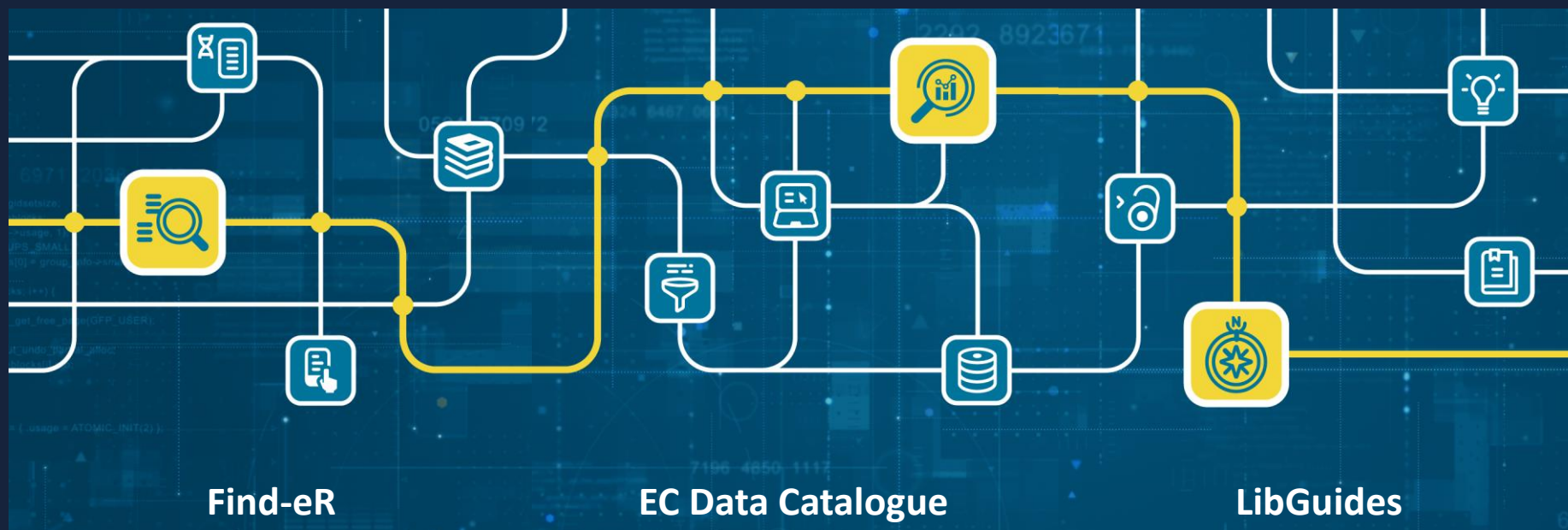
EC Data Catalogue

What is a data catalogue?

"At its core, a data catalog is an organized inventory of the data in your company.
That's it."

Ole Olesen-Bagneux

What is the EC Data Catalogue?

Theoretically

- Metadata catalogue
  - Collection of bibliographic records with links to the actual data

Content wise

- Key data assets
  - Data created by the European Commission
  - Data acquired from a third-party

Technically

- Separate Primo VE Discovery View
- Internal access only

Why a data catalogue?

Data-driven organisation

- Generates and acquires huge amounts of data

Evidence-based policy-making

- Need for reliable data
- Need for traceability

Data discovery and accessibility

- Efficiently retrieving data

Why a data catalogue?

Collaboration and interoperability

- Data-sharing and reuse

Data ecosystem

- EC Data Catalogue is part of complete data ecosystem

Why Alma and Primo VE?

Market exploration

- Temporary open-source solution

Need for a long-term solution

- Proposal by the Publications Office of the European Union to leverage Alma and Primo VE
  - Proven discovery tool
  - Extra out of the box functionalities
  - Customisation

Why the EC Library?

Experienced in:

- Data acquisition and licenses

- (Meta)data management and curation

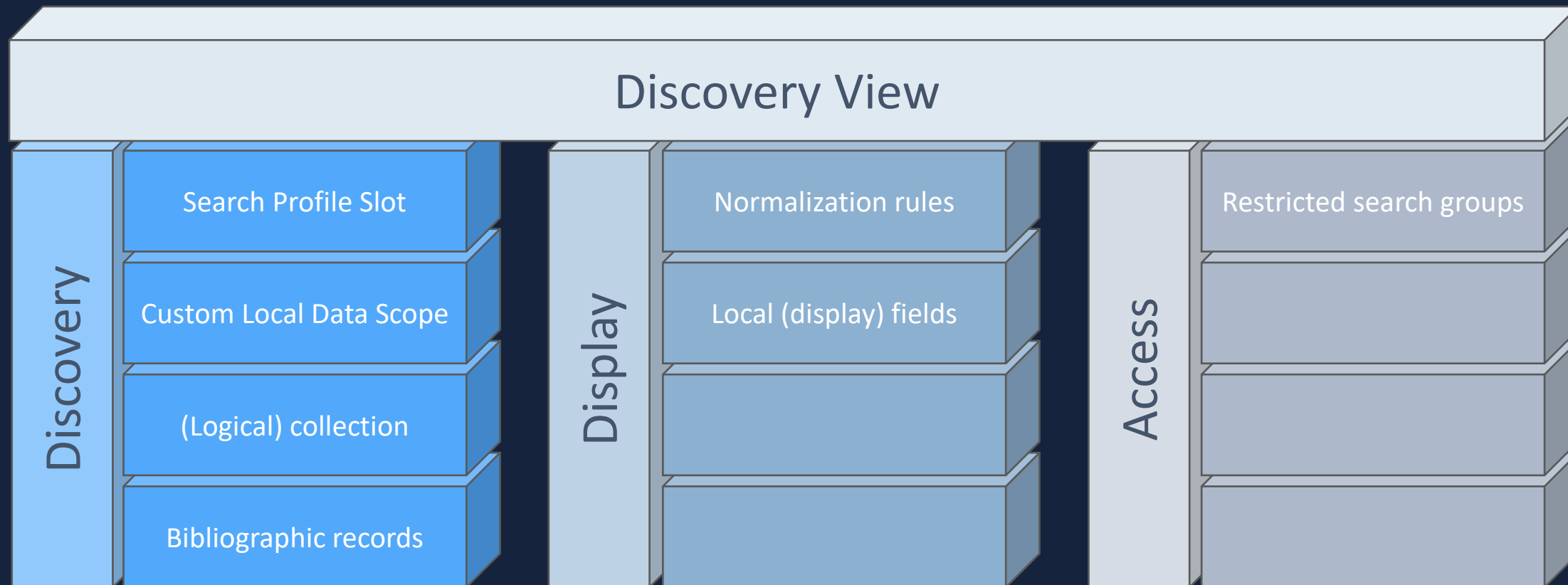- User support, training, promotion and communication

How?

Configuration of the Discovery View

- Discovery
  - Dedicated set of records

- Display
  - Present specific metadata

- Access
  - Restrictions

Data asset publication process

- Content

- Roles

- Tools

Configuration of the Discovery View

Discovery View

Discovery
- Search Profile Slot
- Custom Local Data Scope
- (Logical) collection
- Bibliographic records

Display
- Normalization rules
- Local (display) fields

Access
- Restricted search groups

Data asset publication process

Key data asset → Data owner → Data steward → Alma → Primo VE → Librarian → User Interface
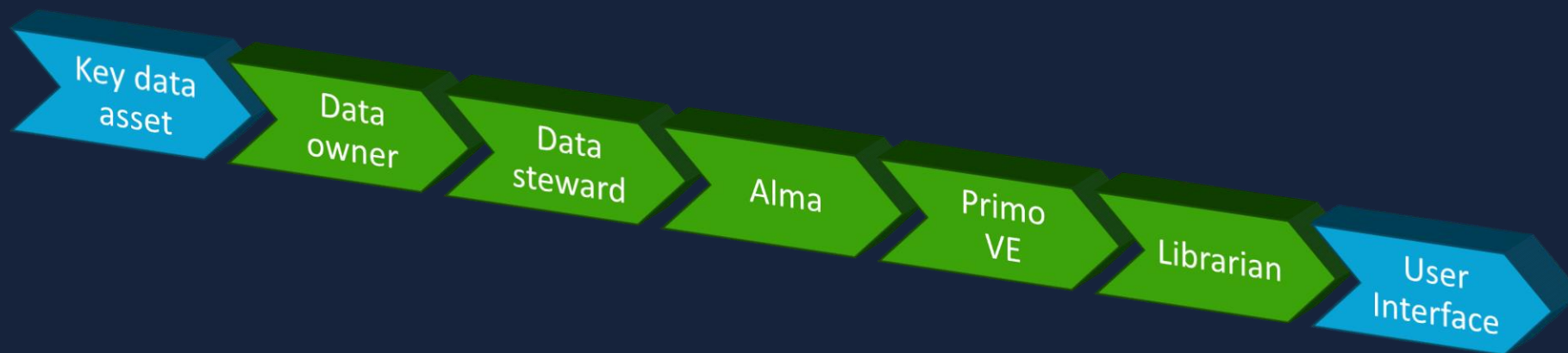
Key data asset definition

Key data asset

- Owned by the European Commission or to which it has access

Key
- Decision-making process
- Functioning of the organisation

Data asset
- Not just datasets, but a broad variety in terms of type, format, location, ownership,…
- *datasets, databases, web portals, studies, statistics,…*

Roles

Data steward

- Not library staff

Librarian

- Provide guidance and support
- Curation

Data asset submission

Alma → Primo VE

Library management system
- Holds the bibliographic records
- Requires access/credentials
- Knowledge of MARC21 required

Submission form

- NodeJS application
  - Standard input fields to MARC21

Create data asset

☐ = required/invalid

Data theme

| Population and society ⌄ | + |

Title

Population by age group

Description

Member States send population data to Eurostat data as on of 31 December for the reference year under Regulation 1260/2013 on European demographic statistics. The data are conventionally published by Eurostat as population on 1 January of the following year (reference year + 1).

Publisher

ESTAT Eurostat ⌄

Submission form

- NodeJS application
  - Standard input fields to MARC21 in XML format

⬇

- Ex-Libris API
  - Create bibliographic record

⬇

- Alma and Primo VE

⬇

- User Interface

Title

Population by age group

```
<record>
    <leader>#####nmi#a22#####7a#4500</leader>
    <controlfield tag="008">######q20uu####xx#k|#w|o||||#0|||||eng||</controlfield>
    <datafield tag="046" ind1=" " ind2=" ">
        <subfield code="k">20xx-xx-xx</subfield>
    </datafield>
    <datafield tag="899" ind1=" " ind2=" ">
        <subfield code="a">ECDC</subfield>
        <subfield code="c">DATASET</subfield>
    </datafield>
    <datafield tag="084" ind1=" " ind2=" ">
        <subfield code="1">http://publications.europa.eu/resource/authority/data-theme/SOCI</subfield>
        <subfield code="a">Population and society</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="0">
        <subfield code="a">Population by age group</subfield>
    </datafield>
    <datafield tag="520" ind1="8" ind2=" ">
        <subfield code="a">Member States send population data to Eurostat data as on of 31 December for the reference year
    </datafield>
    <datafield tag="710" ind1="2" ind2=" ">
        <subfield code="4">pbl</subfield>
        <subfield code="a">European Commission</subfield>
        <subfield code="b">Eurostat [ESTAT]</subfield>
        <subfield code="1">http://publications.europa.eu/resource/authority/corporate-body/ESTAT</subfield>
    </datafield>
</record>
```

Standardisation

- Controlled vocabularies
- DCAT-AP

```
<record>
    <leader>#####nmi#a22#####7a#4500</leader>
    <controlfield tag="008">######q20uu####xx#k|#w|o||||#0|||||eng||</controlfield>
    <datafield tag="046" ind1=" " ind2=" ">
        <subfield code="k">20xx-xx-xx</subfield>
    </datafield>
    <datafield tag="899" ind1=" " ind2=" ">
        <subfield code="a">ECDC</subfield>
        <subfield code="c">DATASET</subfield>
    </datafield>
    <datafield tag="084" ind1=" " ind2=" ">
        <subfield code="1">http://publications.europa.eu/resource/authority/data-theme/SOCI</subfield>
        <subfield code="a">Population and society</subfield>
    </datafield>
    <datafield tag="245" ind1="1" ind2="0">
        <subfield code="a">Population by age group</subfield>
    </datafield>
    <datafield tag="520" ind1="8" ind2=" ">
        <subfield code="a">Member States send population data to Eurostat data as on of 31 December for the reference year
    </datafield>
    <datafield tag="710" ind1="2" ind2=" ">
        <subfield code="4">pbl</subfield>
        <subfield code="a">European Commission</subfield>
        <subfield code="b">Eurostat [ESTAT]</subfield>
        <subfield code="1">http://publications.europa.eu/resource/authority/corporate-body/ESTAT</subfield>
    </datafield>
</record>
```

Consistency

Interoperability

Mapping

- DCAT
  - Standard for data sets

- DCAT-AP
  - DCAT profile for data sets and data services in Europe
    (maintained by SEMIC action, Interoperable Europe)

- Ensures data integrity, interoperability and quality
  - Controlled vocabularies
  - Cardinality constraints

A quick guided tour

A quick guided tour

data.europa.eu

European Commission, Publications Office of the European Union [OP]

public

Edit   Clone   Delete

Details

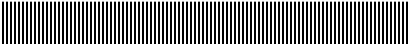|	
Record created	2021-05-07
Record modified	2023-11-30
Title	data.europa.eu
Description	Catalogue of open data descriptions, published by national, European, regional, local and geo portals.
Publisher	European Commission, Publications Office of the European Union [OP]
Contact point	**unit:** OP.C.1. - European Data and Related Services **contact person:** ▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓▓ **e-mail:** op-data-europa-eu@publications.europa.eu **role:** Data steward
Associated DG/Agency	European Commission, OP Publications Office of the European Union
Data ownership category	Public Domain *Open data*
Access rights	public
Homepage	https://data.europa.eu
Documentation	https://dataeuropa.gitlab.io/data-provider-manual/ *Documentation to access and to publish datasets*

Subjects	dissemination of EU information >
	data collection >
	open data >
	European Union >

Data management platform/system	Database
	Virtuoso

Send to

SHARE	QR	EMAIL	PRINT	CITATION	PERMALINK	EXPORT TO EXCEL

EU Data Citation

European Commission, Publications Office of the European Union [OP], 'data.europa.eu' 2021 (updated 2023-11-30), accessed 2024-09-04, https://ec-data-catalogue.primo.exlibrisgroup.com/permalink/32EUC_INST/1u0ms6n/alma997229722229

COPY CITATION TO CLIPBOARD

Remember to check citations for accuracy before including them in your work

Send to

SHARE QR EMAIL PRINT CITATION PERMALINK EXPORT TO EXCEL

Linked data

RDF/XML   Turtle   Notation-3   N-triples   JSON-LD

SANDBOX

What's in it for you

- Experiences can be shared
  - Configuration of Discovery View (settings, normalization rules,…)

- Mapping will be made publicly available
  - EU Vocabularies

- Libraries and librarians have their role to play in data management

Interesting links

- Publications Office of the European Union

- EC Library

- SEMIC

- data.europa.eu

- DCAT-AP 3.0

- Data Citation: A guide to best practice

THANK YOU

quinten.van-gorp@pubications.europa.eu