Documentation of data.europa.eu (DEU)

Version 1.0

data.europa.eu

Table of contents

1. Who we are

> ✎ **Information**
>
> You can download and share the latest version of this documentation in the PDF format: Download

The portal is a central point of access to European open data from international, European Union, national, regional, local and geodata portals. It consolidates the former EU Open Data Portal and the European Data Portal.

The portal is intended to:

1. give access and foster the reuse of European open data among citizens, business and organisations.
2. promote and support the release of more and better-quality metadata and data by the EU's institutions, agencies and other bodies, and European countries, enhancing the transparency of European administrations.
3. educate citizens and organisations about the opportunities that arise from the availability of open data.

The two former portals EU Open Data Portal and European Data Portal, launched respectively in 2012 and 2015, were originally established on the basis of Directive 2003/98/EC to promote accessibility to and the reuse of public sector information. The successor directives Directive 2013/37/EU and Directive (EU) 2019/1024 confirmed and extended its action. The latter invites all the EU Member States to make available

their public data resources, contributing to the smooth functioning of the internal market and the proper development of the information society in the EU.

Currently in its third iteration, the portal merges the activities of the European Data Portal (which focused exclusively on EU Member States and other European countries) and of the EU Open Data Portal (which served data from the EU institutions, agencies, and bodies) into one.

It is funded by the EU and managed operationally by the Publications Office of the European Union in cooperation with the Directorate-General for Communications Networks, Content and Technology of the European Commission, responsible for EU open data policy.
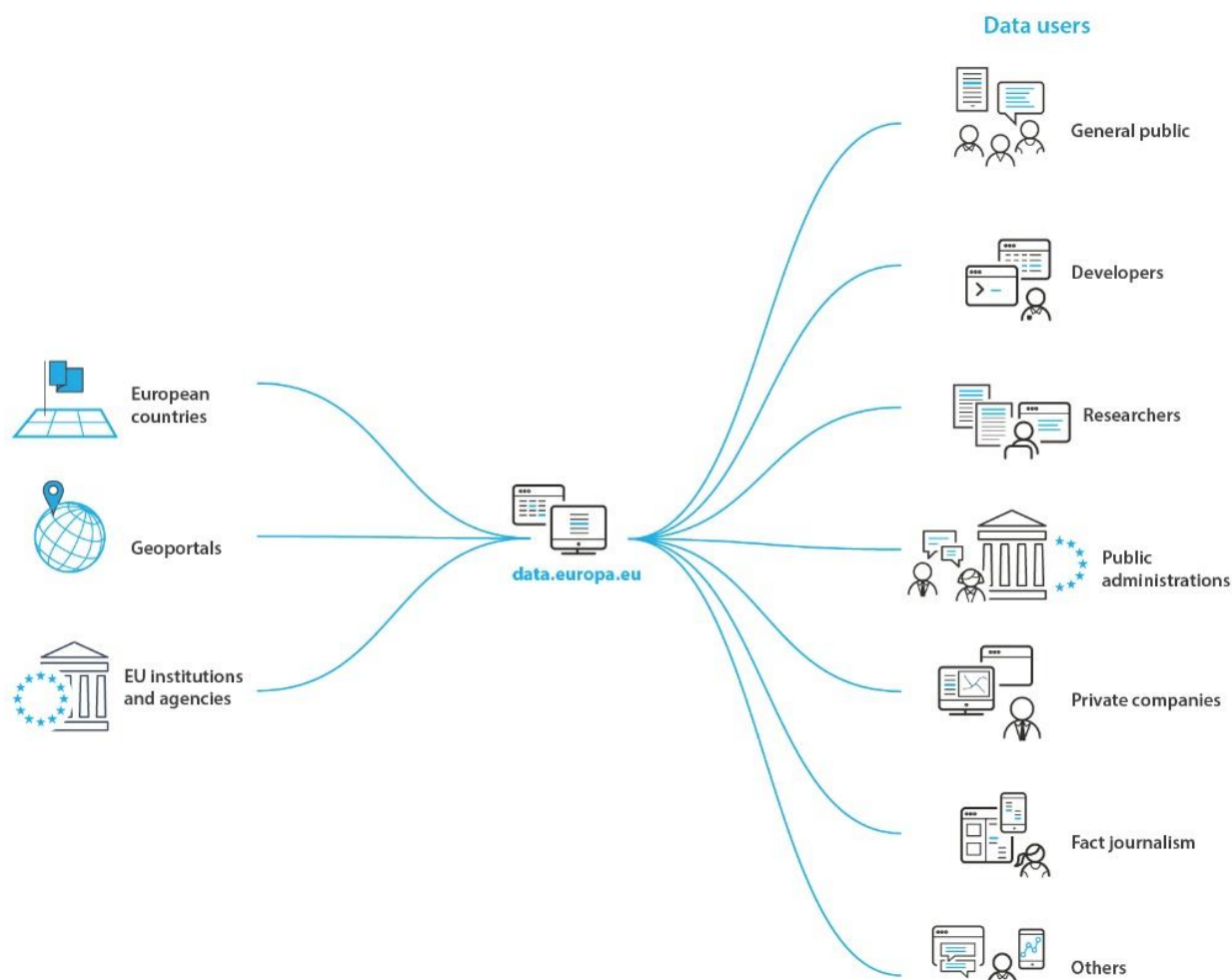
1.1 Our mission, vision and values

The mission, vision and values of data.europa.eu are part of those from the Publications Office of the European Union (OP) with respect to open data.

OP mission. The Publications Office of the European Union is the official provider of publishing services to all EU institutions, bodies, and agencies. As such, it is a central point of access to EU law, publications, open data, research results, procurement notices and other official information. Its mission is to support EU policies and

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.

OP vision. A well-informed EU, empowered by timely and effective access to trustworthy information and knowledge and benefiting from all the opportunities this brings to society and the economy.

OP values.

1. Transparency -- we facilitate transparency throughout the policy cycle of the EU institutions to enhance evidence-based decision-making, accountability, civic participation, and democracy.

2. Trustworthiness -- we strive to ensure that the content we provide is accurate and reliable so that citizens trust the EU as a provider of information.

3. Accessibility -- we believe access to information is a human right that all citizens should enjoy regardless of language, culture, disability, social status, location, technology, or the way they understand information.

4. Service orientation -- we are committed to continuously improving our services to both our institutional stakeholders and EU citizens because we want to contribute to the European project in the best possible way.

a	b	c	d	e	f
European legislative frameworks for open data	Awareness raising of open data topics	Collecting/ receiving metadata	Metadata normalisation/ transformation/translation	Metadata as Linked Open Data based on the DCAT-AP standard	Promotion
Implementation of the European legislative frameworks for open data (EC reuse decision for Commission open data, and public sector information directive for Member States' open data).	Raising awareness of open data topics and capacity building among data providers, at the EU and national levels.	Collecting/ receiving metadata (information about data) from data providers (EU institutions, national portals of the EU Member States and European countries, international organisations).	Normalisation, transformation and automatic translation of the metadata (into all official 24 EU languages) and quality checks.	Storing metadata as LOD (based on the DCAT-AP standard) and making it available for search and usage to the users (experts as well as the public). The use of persistent uniform resource identifiers (data.europa.eu/88u) which is the cornerstone of linked data.	Promotion of available open datasets and open data standards via outreach activities (i.e. webinars, EU Open Data Days, EU Datathon competition, EU DataViz conferences) and direct contacts with the stakeholders.

1.2 Translated metadata catalogue

The portal is a metadata catalogue. To foster the comparability of data published across borders, it presents metadata references using the application profile for data portals in Europe (Data Catalogue Vocabulary (DCAT-AP)), using Resource Description Framework (RDF) technology. It provides translations of metadata descriptions in all 24 official EU languages using machine-translation technologies (eTranslation).

In some situations, metadata machine-based translation might not be as efficient as human translation.

Unique features of data.europa.eu

The EU component of metadata makes specific use of the **EU reference data (data themes)** to describe topic of datasets.

EuroVoc is a powerful tool for linking data as it is aligned to many resources such as major thesauri and wikidata.

Multilingual and multidisciplinary thesaurus

data.europa.eu

Multilingualism
Metadata is **translated in 24 languages.**

DCAT-AP metadata schema

Benefit for data providers and data user groups (or reusers): citizens, businesses, policy makers, etc.

Adds value to harvested metadata and makes them **discoverable** by search engines.

The **DCAT application profile** for data portals in Europe (DCAT-AP) is a specification based on the World Wide Web Consortium's (W3C) data catalogue vocabulary (DCAT) for describing metadata of public sector datasets in Europe.

1.3 From raw data to services

Our services are focused around three main spheres:

- Data providers -- any entity that gives access and distributes data to the public. To ensure the reliability of the published resources in the portal, the data providers of the portal are official representatives from supranational, national and local public administration.

- Data users -- any person or entity who accesses and consumes data for any purpose. These include a wide range of data users who visit the portal for different purposes: non-governmental organisations, international organisations, private sector, academia, students, etc.

- Data literacy -- the ability to read, understand, create, and communicate data as information. Much like literacy as a general concept, data literacy focuses on the competencies involved in working with data.

The 'data providers' (EU institutions, agencies and EU bodies, Member States, and other European countries) are autonomous in publishing their metadata (which gives you access to their data) in data.europa.eu. The portal also publishes datasets of organisations beyond the EU. The portal is being updated whenever new datasets and content are available.

In addition to the vast collection of public datasets, most of them published in open formats, we offer additional contents and related services:

- training materials on open data in the data.europa academy

- studies such as Open Data Maturity or The Economic Impact of Open Data;

- news and data stories

- a metadata quality dashboard assessing metadata quality against various findable, accessible, interoperable, and reusable (FAIR) indicators

- a licencing assistant

- the possibility of storing your data under certain conditions (please contact us)

- the social media channels (X, LinkedIn, YouTube, Facebook)

- a list of events with the most relevant events taking place

- a monthly Newsletter

New features for registered users will be added soon to data.europa.eu.

The whole source code of the portal is available at GitLab for free reuse.

How to contribute to the portal: - Give us feedback, for example, suggest new functionalities you would like to see on the portal - Suggest a portal to be harvested by us - Share your data use story in which we can promote the use of data

1.4 Open data in the European Union

Several policy and legal developments on EU open data have taken place since 2003.

- **2003** -- Directive 2003/98/EC on the re-use of the public sector information (PSI) directive.
- **2011** -- Commission Decision (2011/833/EU) on the reuse of Commission documents, established the basis for the former EU Open Data Portal, operated by the Publications Office of the EU (2012--2021).
- **2012** -- launch of the EU Open Data Portal, central point of access to open data from EU institutions, agencies, and bodies.
- **2013** -- Directive 2013/37/EU amending Directive 2003/98/EC on the re-use of public sector information.
- **2015** -- launch of the European Data Portal publishing data from national, regional, and local open data portals.
- **2019** -- Directive 2019/1024 on open data and the re-use of public sector information recasts Directives 2003/98/EC (PSI directive) and 2013/37/EU setting up a legal framework of minimum requirements for Member States regarding the accessibility of public data resources.
- **2020** -- Communication on a European strategy for data (COM/2020/66) aiming at creating a single market for data for the benefit of businesses, researchers, and public administrations to ensure Europe's global competitiveness.
- **2021** -- data.europa.eu consolidates the former European Data Portal (data from the EU Member States and other European countries) and the former EU Open Data Portal (data from the EU institutions, agencies, and bodies) into the official portal for European data.
- **2022** -- the Data Governance Act sets a framework for fostering voluntary data sharing beyond open data
- **2023** -- Entry into force of a Commission implementing regulation on high-value datasets, those that have a high potential for economic and societal impact
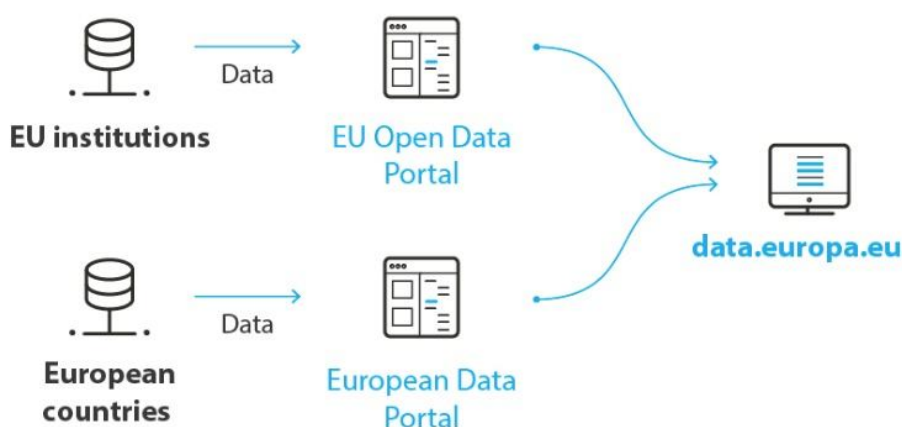
1.5 European open data space

The European open data space is an essential element of the single market for data -- an EU-wide interoperable data space that will enable the development of new products and services based on public data

and industrial and scientific applications. It focuses on the implementation of EU open data and reuse policies under the legal acts adopted by the EU institutions. The Publications Office of the EU works on all four European open data space building blocks and their objectives: - providing a comprehensive catalogue of open data and citizen-centric reuse services; - improving the interlinking and interoperability of open data with other sources of public-sector information, such as legislation, publications and digital content; - fostering the use of data from EU content through the organisation of EU Datathon competitions and data visualisation events; - contributing to the implementation of data governance and policies across the EU institutions.

Federated European Open Data

The open data landscape of Europe could be seen as a system of federated open data portals, data.europa.eu being the highest layer, where all metadata arrive in the end.

data.europa.eu
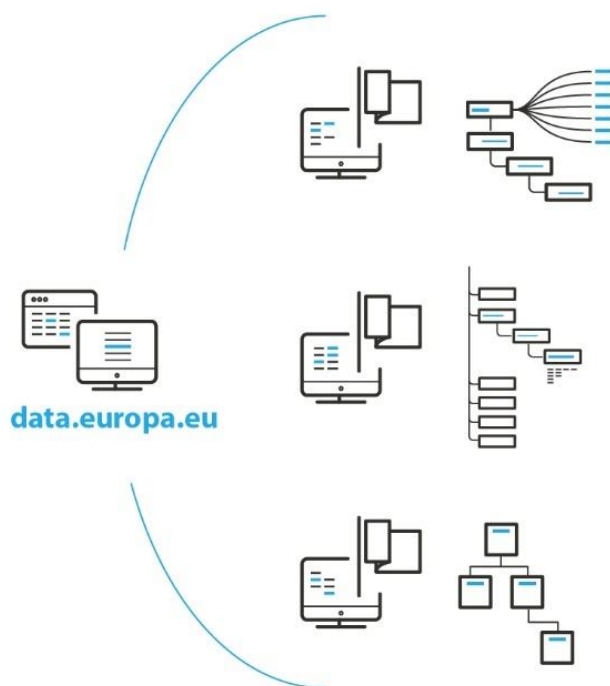
National open data portals

The main differences are in the scope, in the number of languages provided and in the number of metadata they are storing – and that only a small number of them already use DCAT-AP natively.

2. Our metadata model

It is important to distinguish between metadata and data in data.europa.eu. Most information you discover on the portal constitutes metadata, i.e. information about data (title, description, publisher, etc). The metadata then links to the actual data, in most cases a downloadable file. The entirety of metadata and data is often called a dataset. The metadata is stored in the databases of data.europa.eu, while the data usually remains with the original data publisher. However, data.europa.eu is capable of storing both, metadata and data. With the data provider interface (DPI), both metadata and actual data can be inserted: metadata by filling in the relevant fields the actual data by uploading the data files (up to 10 Gigabytes each).

The (meta)data model used in data.europa.eu is DCAT-AP, the application profile for data portals in Europe. The specification of the DCAT-AP was a joint initiative of the Directorate-General for Communications Networks, Content and Technology, the Publications Office of the European Union and the interoperable Europe programme. The specification was elaborated by a multi-disciplinary working group with representatives from 16 EU Member States, some European institutions, and the United States.

DCAT defines a dataset as 'a collection of data, published or curated by a single agent, and available for access or download in one or more representations' (by data it is implicitly understood metadata and the actual data).

The portal collects the metadata from the data providers even when it contains deviations from the common data model. With the scope of improving the quality of the metadata and data, the portal constantly assesses the quality of the metadata errors and communicates them back to the owners of the datasets. Metadata quality evaluation is visible for each dataset and for entire collections as a dashboard. More information is provided in the Metadata Quality section.

The portal collects all datasets from the portals it harvests. The actual data is collected and made available in the file format provided by the data provider and shown under the distribution(s) of the dataset.

data.europa.eu is a **catalogue of metadata** providing a single point of access to European countries and EU institutions' open data for re-use.

Anyone can easily search, explore, link, download and reuse the data for commercial or non-commercial purposes.

users **data.europa.eu** **data**

Metadata catalogue *Data provider's websites*

users **data.europa.eu** **data**

reuse of data
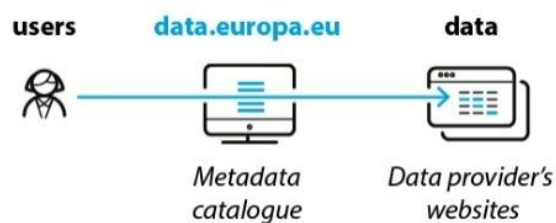
Users can suggest datasets, give feedback on the quality of data and share the use made with the data from the portal.

2.1 The DCAT-AP version used

The current version of DCAT-AP in the portal is version 2.1.1. These are the improvements of this version:

• improved Unified Modelling Language (UML) diagram in accordance with the agreed profile reading;

• improved coherency between the UML diagram and the specification text;

• a usage guide on the relationships between dataset, distribution and data service; and the consequences of this clarification on the model;

• various editorial fixes;

• consolidation of the SHACL shapes;

• minor specification updates:

• introduction of the named authority list (NAL) planned-availability, NAL access-right and NAL dataset-type,

• lift of the max-cardinality for dataset dct:type,

• lift of the max-cardinality for property dct:creator,

• allow other than SHA1 checksum algorithms,

• the range for temporal properties is enlarged to contain any temporal XSD (XML schema definitions) datatype,

• alignment of usage notes for used property adms:status with W3C DCAT,

• addition of max-cardinality 1 for dcat:temporalResolution and dcat:spatialResolutionInMeters to align with the usage note.

A complete list of the issues and their resolutions can be found on the DCAT-AP GitHub. The issue tracker for DCAT-AP can be consulted in the GitHub repository.

2.2 How to export the metadata of the dataset

The metadata of a dataset can be exported via the details page of the requested dataset. The Linked Data tab allows the user to download the metadata in various representations. In particular, RDF/XML, Turtle, Notation3, N-Triples and JSON-LD are exportable. Alternatively, the user can extend the dataset ID in the URL with the intended file extension. In particular and in same order, .rdf, .ttl, .n3, .nt and jsonld.

The metadata of the distribution of a dataset is located under the same hood. The user can export the metadata of a dataset to extract the metadata of a corresponding distribution. Analogously, the user can export the metadata of a catalogue under the catalogue details page. Go to the datasets page and select catalogues. From there select the requested catalogue and export metadata via Catalogues Metadata as Linked Data dropdown.

The available formats can be consulted with this query. This is a non-exhaustive list of the formats offered (as this is updated on a daily basis):

- Tabular/Text data:
- CSV
- Excel
- HTML (Hyper Text Markup Language)
- PDF
- TSV
- Application/Script data:
- ATOM
- JSON
- JSON-LD
- OCTET STREAM
- RDF-N3
- RDF-Turtle
- RDF-XML
- RSS
- XML (eXtensible Markup Language)
- Geospatial data:
- GML
- KML (Keyhole Markup Language)
- SHP
- WFS
- WMS
- Images/Graphics:
- JPEG
- GIF (Graphic Interchange Format)
- PNG
- SVG (scalable vector graphics)

3. How to search for datasets

The metadata catalogue of datasets can be explored through a search engine (data tab), through a SPARQL endpoint and API endpoint.

An up-to-date list of all catalogues can always be retrieved by this SPARQL query.

3.1 Search manually

3.1.1 Basic Search

You can search for datasets containing one or more keywords by typing the keywords into the search field and then clicking on the search button. The search process will reduce your keywords to their root form, to ensure variants of your keyword also match. For example, "Europa" can also match "Europe", or "walking" can also match "walk" and "walked".

3.1.2 Advanced Search

When entering more than one search term it is also possible to use the logical operators (connectors) **AND**, **OR** as well as **parentheses ()**. Furthermore, it is also possible to do a **wildcard** search and search for an **exact phrase**.

Logical Operator OR

The **OR** operator can be used in order to

- connect two or more similar concepts (synonyms)
- broaden your results, telling the database that ANY of your search terms can be present in the resulting dataset

Example: population OR education OR science

All three circles represent the result set for this search. It is a big set because any of those words are valid using the OR operator.
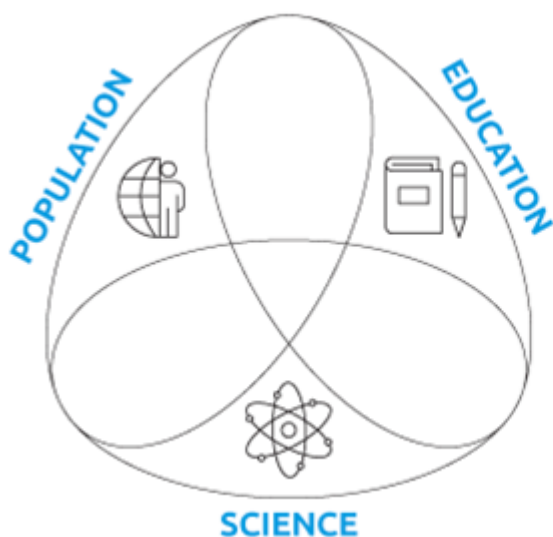
POPULATION OR
EDUCATION OR
SCIENCE

Logical Operator AND

The **AND** operator can be used to:

- find sources containing two or more ideas
- narrow the search

The database will only retrieve items containing both keywords. The **AND** operator can be used multiple times in one query.

Example: population AND education AND science
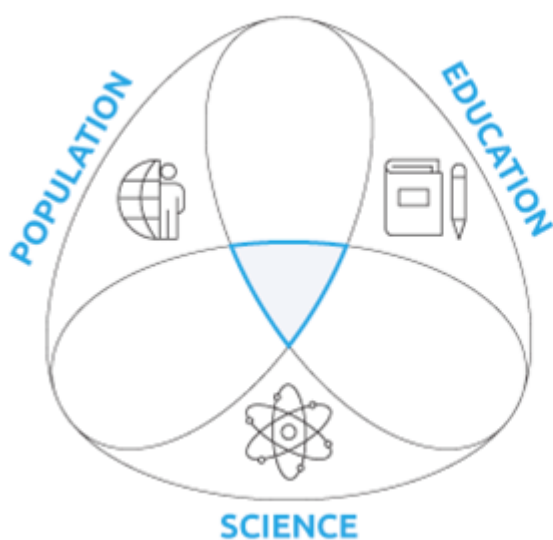
POPULATION AND
EDUCATION AND
SCIENCE

Parantheses ()

Parentheses are used to:

- perform a more complex search using both AND and OR by placing parentheses around synonyms
- save time by searching multiple synonyms at once

Example: (environment OR nature) AND (refurbish OR reuse)

This avoids the need to perform multiple searches for combinations of keywords.

| (environment OR nature) AND (refurbish OR reuse) | Datasets ∨ | 🔍 |

Datasets found (1 981)   Sort by: | Relevance ∨ |

Basic rules for using AND, OR operators and parentheses ()

- The OR is implicit: search function automatically puts an OR in between your search terms. It means that the search for population OR education OR science gets the same matches as the search for population education science.

| cat OR dog | Datasets ∨ | 🔍 |

Datasets found (72 670)   Sort by: | Relevance ∨ |

| cat dog | Datasets ∨ | 🔍 |

Datasets found (72 670)   Sort by: | Relevance ∨ |

- Always enter **AND** and **OR** operators in uppercase letters.
- Never translate **AND** and **OR** operators into any other language. It does not matter in what language your keywords are the operators always must be **AND** and **OR**.

| katze OR hund | Datasets ∨ | 🔍 |

Datasets found (2 290)   Sort by: | Relevance ∨ |

WRONG! Never

| katze ODER hund   translate | Datasets ∨ | 🔍 |

Datasets found (17 833)   Sort by: | Relevance ∨ |

- Please keep in mind to always close **parentheses ()**. The combination of logical operators with an odd number of **parentheses** in the query leads to incorrect results in the search.

Wildcard

Wildcard search can be used to optimize the search result when you do not know the entire keyword by using ? (question mark) to replace a single character and * (asterisk) to replace zero or more characters.

Example: ois?au dat*

The example above will return datasets that can contain oisiau data, oisoau dataset, or oiseau datasets. Wildcard search does not reduce a keyword to its root form, like in basic search, since it would be wrong to do so on a term that some of its letters are unknown.

Please note that the more keywords need to be checked, just in case they match (e.g. a*, or b*, or c*), the heavier and the poorer the wildcard search performance can be. Having a wildcard at the beginning of a keyword (e.g. *ing, or ?iseau) is ignored and the search term will be treated as it is to avoid such a heavy and expensive process. Invalid wildcard search terms (e.g. ois?au **OR** (data* **AND** organization (unbalanced parentheses)) will be also treated as it is.

Exact Phrase

When you place your keywords in double quotes, they will be considered as a phrase, not case sensitive and characters are taken literally as it is. The search will return datasets containing the keywords in exactly the same order.

Example: "Manual public space"

The example above returns datasets containing exactly "manual public space". It will not return datasets containing "manual space public", or "space public manual", or any other results with the search terms appearing in different sequences than "manual-public-space". If you search for "public spa*", the results will only list datasets containing "public spa*".

3.2 SPARQL search (graph search)

SPARQL search enables even more advanced users to find datasets using Resource Description Framework (RDF) query language. SPARQL can help to find specific information from a large amount of RDF dataset, even if it organized in a complex way. For more information see the data.europa.eu SPARQL, and the SPARQL search.

3.3 Terms of reuse

Most of the data accessible via data.europa.eu is released by the respective data providers using an open licence. Data can be used for free for commercial and non-commercial purposes, provided the source is acknowledged. Specific conditions for reuse, relating mostly to the protection of data privacy and intellectual property, apply to a small amount of data. A link to these conditions can be found for each dataset.

The terms of use can be found in the data.europa.eu copyright notice. Most data is covered by open licences. As of September 2021, the most common open licences were the Creative Commons 'CC-BY-4.0' licence, the 'Data licence Germany – attribution' licence or Etalab's Open Licence (used by the French government).

data.europa.eu has been set up by the European Commission to **implement EU open data** and reuse policies under the legal acts adopted by the EU institutions and EU Member States.

Legal basis

European countries

EU institutions and agencies

Open data

data.europa.eu

Legal basis

Directive on open data (Member States)

Commission reuse decision

Council decision

Creative commons
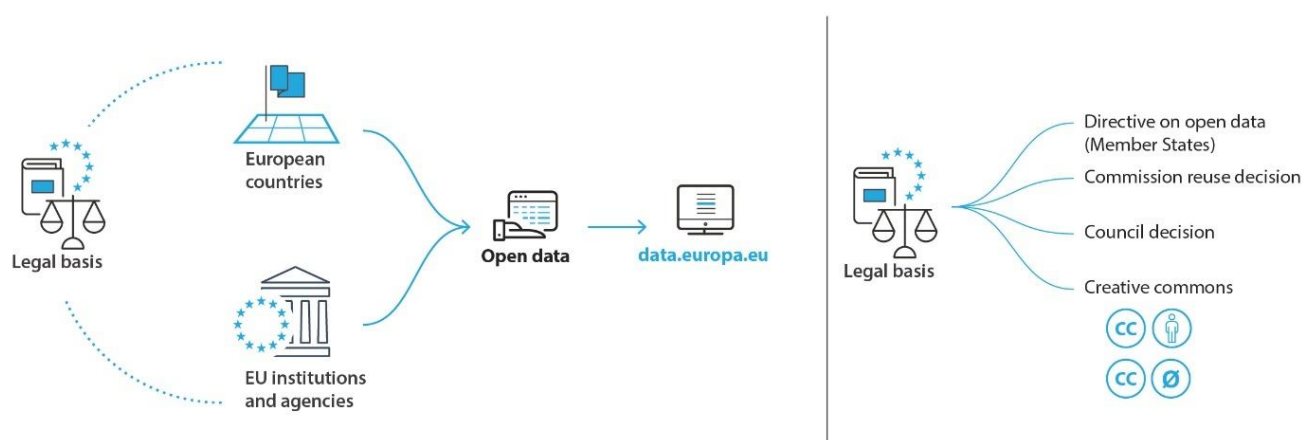
4. How to publish on data.europa.eu

4.1 How to publish on data.europa.eu

The core (meta)data model of data.europa.eu is DCAT-AP.

Essentially, DCAT-AP consists of three principal data classes: catalogues, datasets and distributions. Each data provider is represented by a catalogue. Each catalogue consists of datasets that constitute the general metadata of the data, and each dataset can have multiple distributions, where each distribution describes the actual data of the dataset in detail. All this data is serialised in the RDF format. Therefore, the DPI converts all user input into RDF. As a data provider you are concerned with the creation of datasets and distributions. The catalogues are managed by the administrators of data.europa.eu.

For more information, see the Our metadata model section.

4.2 Manual publication

4.2.1 Access control model

The provision of datasets follows a straightforward and simple access control model. Let us assume a data provider organisation consists of multiple users (i.e. data providers), who want to manage datasets on behalf of the organisation. Each user is granted write access to one or more catalogues that belong to the data provider organisation. This access allows the user to create, update, delete, and execute any available function on any dataset in that catalogue. All users of one data provider organisation have the same view on the datasets and their state. It is up to the internal processes of a data provider organisation to manage the detailed publication process and individual responsibilities. The write access to catalogues is set by administrators of data.europa.eu.

4.2.2 State of a dataset

Datasets can have two states: draft or public. A draft dataset is not publicly available via the frontend, API or SPARQL interface of the data section. It is only visible to permitted data providers. A public dataset is available like any other dataset on the data section. Datasets can be directly created as draft or public. It is possible to toggle the state of a dataset at any time.
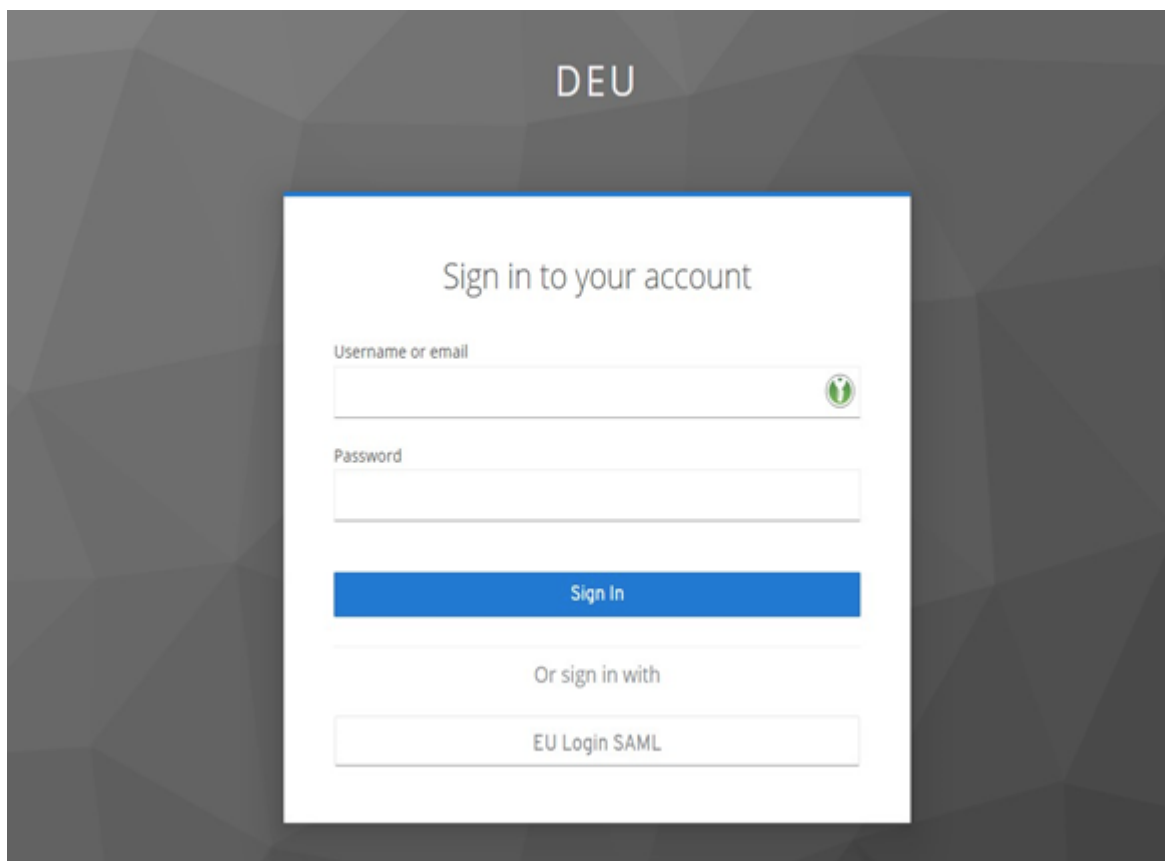
4.2.3 Registration and login

There is no self-registration to use the Data Provider Interface (DPI). Please contact OP-DATA-EUROPA-EU@publications.europa.eu for requesting access to the DPI. You will be provided with access credentials (username and password).

You will find the Data Provider Interface Login in the footer under Other Service. The data provider interface is only available in English.

data.europa.eu

This site is managed by the Publications Office of the European Union.

Contact us

Contact the data.europa.eu team
subscribe to our newsletter.

Ressources

Copyright Notice
Persistent URIs

Follow us on

Facebook   Twitter   LinkedIn
Youtube

Legal

Accessibility

About us

The Publications Office of the European Union is the official provider of publishing services to all EU institutions, bodies, and agencies. As such, it is a central point of access to EU law, publications, open data, research results, procurement notices and other official information.

Other Services

EU law
European data
EU tenders
EU research results
EU Whoiswho
EU publications
Data Provider Interface Login

After clicking, the page redirects to the login form.

You have to enter your username and password and click on **Sign In**. If it is your first login process, you will be redirected to another form to change your initial password. Upon success, you will be automatically redirected to the data section.

You can logout by simply clicking on **Logout** in the DPI menu. You will be redirected to the data section.

4.3 Data Provider Interface (DPI) structure and function

This section provides an overview of the structure and individual functionalities of the DPI as accessible via the DPI menu.

The menu gives you access to high-level pages for data providers and dataset-specific functions. Some functions are context sensitive, so they are only available on specific pages, such as a dataset details page. Functions that aren't available are grey in the menu.

High-level menu

Function	Description
My Draft Datasets	Gives you access to the list of draft datasets of the current user. Further functions regarding the draft datasets are available on that page.
My Catalogues	Gives you access to the list of catalogues that are assigned to the current user. An assignment implies the right to create, edit and delete datasets in the respective catalogues.
User Profile	Gives you access to the profile of the current user.
Logout	Logs the current user out.

Dataset sub-menu

Function	Description
Create Dataset	Navigates to the form for creating a new dataset.
Delete Dataset	Deletes the current dataset. *
Edit Dataset	Navigates to the form for editing the current dataset. *
Set to draft	Sets the current dataset as draft, so it is not publicly visible anymore. *
Register DOI	Registers a DOI (Digital Object Identifier) for the current dataset. *

* Only available on a dataset details page.

4.3.1 Create a dataset

You can create a dataset with a wizard-like form that guides you through the provision of the metadata and data. Just click on **Dataset** then **Create Dataset** in the DPI menu.

w do you know? ⌄

European Commission

◯ Anmelden   **DE** Deutsch   Nach Datensätzen suchen   **Search**

data.europa.eu - The official portal for European data

Home   **Daten** ⌄   **Documentation** ⬀   **Publications** ⌄   **data.europa academy**   **Meldungen** ⌄   **Kontakt**

Create a new Dataset | Clear   Next Step

| Essential Properties | Advised Properties | Additional Properties | Distribution Overview | Dataset Overview |

Title of the Dataset* ⓘ

A short descriptive title for the data set.

🌐 English (en)   Add a title

Please provide a title in your chosen language

Add Title in another language

Dataset ID* ⓘ

Create a unique ID

This property contains the main identifier for the Dataset.

Description of the Dataset* ⓘ

The description of the Dataset

🌐 English (en)

Dataset Description placeholder text

The form is divided into five main steps:

- define essential properties of the dataset
- define advised properties of the dataset
- define additional properties of the dataset
- create distributions (distribution overview)
- dataset overview

The creation of distributions is divided into several sub steps. You can always switch between the steps by clicking on the step titles or using the 'Previous Step' or 'Next Step' buttons. There might be cases when a direct access to a step is not possible, for example, a mandatory field is missing.

In order to prevent accidental data loss, your input is constantly stored in the local storage of the browser.

You can clear the entire form by clicking on 'Clear'.

You will find additional information in the i-tooltips, simply hover over them and the information will be displayed.

4.3.2 Special input fields

The form consists of specialised input fields, supporting the various properties of DCAT-AP.

Multi-lingual fields

Some properties can be provided in multiple languages. This is supported with the following kind of input field:

You can just add more languages by clicking on the blue button and remove them by clicking on the small minus sign.

Vocabulary fields

Many properties depend on controlled vocabularies. You can select the fitting value(s) from these vocabularies with a search-based dropdown field. Just type in some characters to find a suitable match. Below is an example for the language property.

For properties where multiple values can be selected from a vocabulary, you can easily repeat the process for each value and your selection is displayed under the input form. Please be aware that for usability reasons, the dropdown will only display a selection of ten values. If you are looking for a different value that is not displayed, simply type in the name and click it inside of the dropdown.

4.3.3 Filling the form

By stepping through the wizard you are able to provide all DCAT-AP properties to describe your dataset. However, only a few properties are mandatory, such as the title and description. You will get a clear warning if a mandatory property is missing.
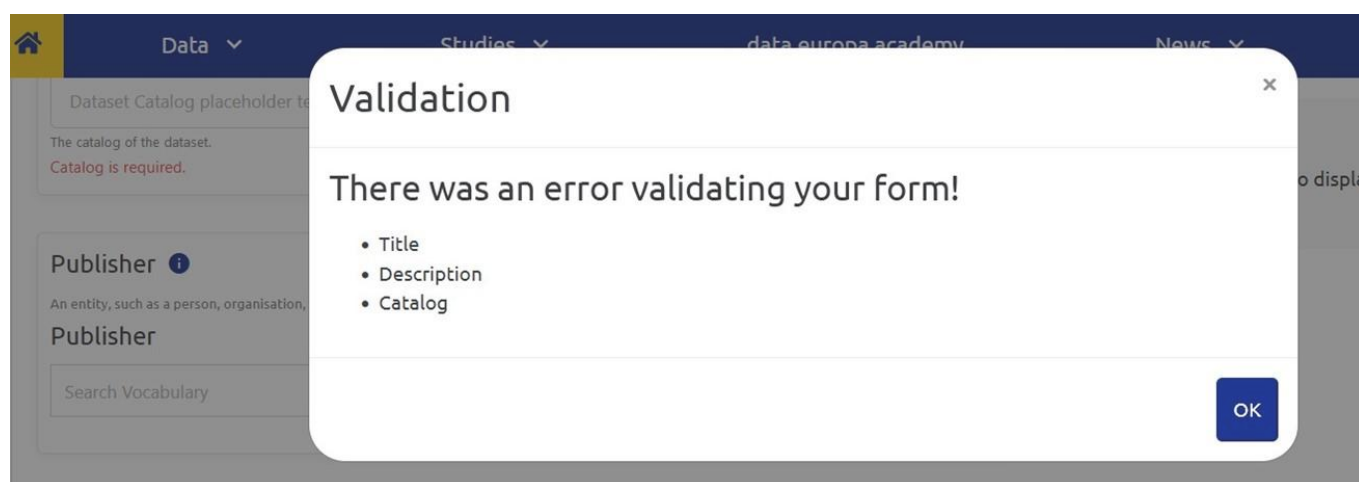
4.3.4 New dataset

In the following, some important details about the form are presented. However, not every property is discussed. Please consult the DCAT-AP documentation for detailed information about every property.

Basic metadata and metadata ID

In the first step, you provide the very basic metadata about the dataset. An important property is the dataset ID, which will be used in the URL to resolve the dataset after publication (http://data.europa.eu/88u/dataset/ [dataset-id]). You can enter it yourself or it will be automatically generated based on the provided title. It can only contain lowercase letters, numbers and dashes. Its uniqueness is checked on-the-fly to avoid any clashes with existing datasets.

You must select a catalogue, which the dataset will be part of. You can only select catalogues that you have access to.

By clicking on 'Next Step' you will be directed to the second step.

Define dataset properties

In the second and third steps, the remaining properties can be provided.

Create distributions

You can provide all possible distribution data in four steps. You can repeat these steps for each distribution you want to add. To navigate through the steps, use the buttons 'Previous Step' and 'Next Step' again, or click directly on the step names.

A central property is the access URL, which gives you access to the actual data of the dataset. Each distribution can have several access URLs. If your data is already hosted and publicly available, you can just provide the URL by selecting the **Type** 'Provide a URL'.

You can also upload your data directly here, by selecting the **Type** 'Upload a file'.

If you do not provide a separate download URL, the download URL is automatically set to the access URL, after saving the dataset.

On the last page of the distribution wizard you will find an overview of your created distributions. You can delete, edit them or add another distribution by clicking on **Add**

Distribution

When you click on **Next Step** you will be redirected to the final overview of your dataset.

Dataset overview and storing

The final step provides you with an overview of your dataset. Note that the layout here is different than the final dataset detail page. You can still go back to previous steps and make changes to your data. If you want to finish the process, you have two options. By clicking on 'Publish Dataset', your dataset will be published immediately and publicly, visible to all users of the portal. You will be redirected to the public dataset details page. By clicking on 'Save as Draft', the dataset will be stored separately and will not be publicly available.

You will be redirected to the draft overview page. You can later edit or publish the draft dataset.

Since the access control is catalogue-based, all users that have access to the catalogue of your dataset can view and edit your draft datasets.

4.3.5 Managing datasets

You can edit, publish and delete datasets. Depending on whether the dataset is public or a draft, access to the available functions differs.

Managing public datasets

You can manage all datasets that are part of any catalogue you have access to. You can check and access these catalogues and their datasets by clicking on 'My Catalogues'.

If you are on a dataset details page you use the sub-menu to access the options.
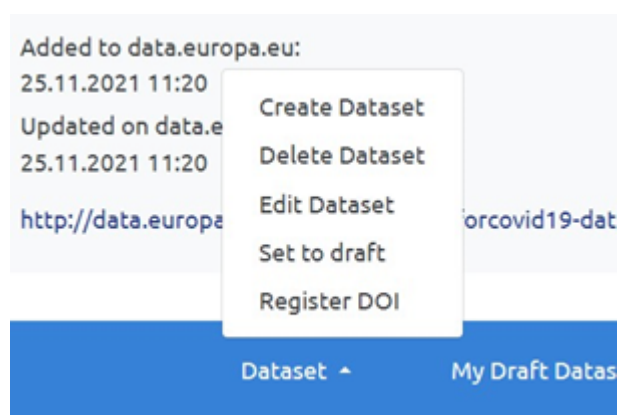
'Delete Dataset' allows you to delete the current dataset. A final confirmation is required. 'Edit Dataset' will redirect you to the dataset wizard with already prefilled form fields. You can apply any changes, as if it were a new dataset. 'Set to draft' will un-publish the dataset and add it to the draft dataset pool.

Managing draft datasets

You can manage all draft datasets of your catalogues by clicking on 'My Draft Datasets'. Note that you will also see datasets here that were not created by you in person. Other users may have access to the same catalogues.

'Delete' allows you to delete the draft dataset. A final confirmation is required and this action cannot be undone. 'Edit' will redirect you to the dataset wizard with already prefilled form fields. 'Publish' will make the draft dataset publicly available. After that, it will not appear in the draft list anymore. The 'Linked data' button gives you access to the raw RDF representation in different formats.

4.3.6 DOI Registration

You can easily register a DOI for your dataset. We use the registration agency of the OP to issue DOIs with the prefix 10.2906. Therefore, your dataset will be available under 'https:// doi.org/10.2906/[id]', where [id] is a randomly assigned number.

The registration of a DOI is permanent and should only be considered for finalised datasets. You can only register one DOI for a single dataset.

Requirements

Since the dataset needs to be public, a DOI can only be registered for published datasets and not for drafts. In addition to the mandatory DCAT-AP properties, you must set the following fields in your dataset: publisher, creator and the issue date and time. Without this information, the registration process will fail.

Register a DOI

You can register a DOI for all datasets you have access to. Just browse to a dataset details page, open the dataset sub-menu and click on 'Register DOI'.
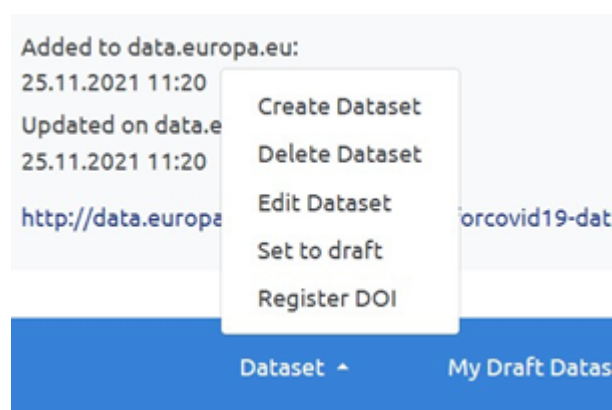
You will need to acknowledge the registration again.

After successful registration you can reload the dataset details page and find your DOI under additional information section.

It is possible to repeat the process, when you have updated the metadata of the dataset, such as the title. In that case, no new id is generated, but the existing one is updated accordingly.

4.4 Publication and editing via API

Alternatively, you can create, update, and delete datasets programmatically via the API of data.europa.eu. For further information on this, please refer to the API documentation section.

4.5 How to be harvested

EU institutions, agencies and other bodies, and the Member States (the 'data providers') are autonomous in publishing their open data. Harvesting is the recommended method for publishers who manage their data in a data catalogue.

4.6 How to request harvesting

Data.europa.eu harvests metadata from available national open data portals. To initiate the onboarding of your portal onto data.europa.eu, you will need to undertake two sequential steps: check that your portal is suitable for harvesting and issue a harvesting request via the data.europa.eu contact form.

4.6.1 Required information

The very first step is to go through the following checklist to gather all required information. The purpose of this checklist is to guide you in gathering and summarising the main requirements to enable the successful harvesting. Before contacting data.europa.eu please make sure that you can answer all listed questions. Of course, if anything is unclear, you can always reach out to us via the contact form.

Please remember that the preferred harvesting interface is OAI-PMH.

	Requirement	Value
1	Which country does your portal cover?	Free text
2	Is your portal already being harvested by another portal?	Free text
3	What is the Uniform Resource Locator (URL) to your portal's interface / endpoint?	URL
4	What is the default language of the metadata from your portal?	Free text
5	Which metadata standard is supported by your portal?	E.g. DCAT-AP, CKAN JSON, or INSPIRE
6	Which representation of the metadata can be used?	XML, JSON, or any RDF representation
7	Which API/Protocol is used to retrieve the data?	OAI-PMH (Recommended) RDF dump file CKAN API SPARQL endpoint
8	Is authentication required for you to access your API?	yes/no
9	Does your data use standard date/time formats as specified by the ISO8601?	Using the ISO8601 standard is mandatory.SO8601
10	How often can/should the site be harvested?	E.g. daily/weekly/monthly/etc.
11	Are there any times when the site should not be harvested (e.g. scheduled maintenance)?	Free text
12	Who is the publisher of the portal (name and email address)?	Free text. e.g. Federal Open Data Agency; info@open-data.gv.eu
13	What is the URL to the homepage of the portal?	URL

Harvesting request via contact form

Once you have gathered all answers to the checklist, the second step is to contact the data.europa.eu to initiate the harvesting onboarding of your portal. Please submit a request via the form, selecting 'Get harvested by data.europa.eu' as the issue type. Once we receive your request, we will assess it and keep you informed on progress.

In the contact request, please provide information on all questions listed in the checklist.

4.6.2 Technical requirements/constraints

The harvester accesses the endpoints of all catalogues on a daily basis by default. Depending on the total size of data provided by each portal as well as depending on other factors such as available resources, other

harvesting intervals can be negotiated on an individual basis. The harvester is configured individually for each harvested portal. Metadata data is processed usually overnight. Every incoming non-DCAT-AP-format will be transformed to the most recent version of DCAT-AP.

4.6.3 Access to harvested sites

Authentication

Some source sites require authentication, meaning we need a login name and password before we can access the data (here data.europa.eu). If this applies to your portal, please state this in your message when using our contact form. API access to harvested site. For harvesting to take place, the source site needs to have in place one of the interfaces as. described in detail in the Interface supported for harvesting section.

FTP access to harvested site

Data.europa.eu does not support FTP for downloading datasets from a source site.

4.6.4 Operational requirements

Harvesting frequency

Due to the high volume of metadata that will be harvested from a growing list of data suppliers and the required runtime for the harvesting processes, data supplier sites are harvested daily by default. Furthermore, the harvesting processes must be clustered and scheduled on a fixed time schedule (e.g. during the night) in order to avoid any load impacts on the harvested sites during regular business hours usage. Other factors and circumstances permitting, harvesting intervals that are more or less frequent can be agreed individually.

Data source site API / endpoint

The data source endpoint should accept queries with, for example, offset / limit parameters for resumption, partitioning, and pagination of the datasets to be harvested.

Quality of the harvested datasets

Metadata quality is an important concern for data.europa.eu. This section explains the basic procedures implemented by data.europa.eu to monitor and improve metadata quality.

Ensuring uniquely identifiable datasets

Only when the same dataset always has the same unique id it can be ensured that it will be recognized as the same dataset on data.europa.eu and that it will not be duplicated.

Error reporting on harvested metadata

The MQA module provides a graphical report on the quality of the harvested datasets' metadata by providing access to a dashboard that summarises the main quality indicators, for example, availability and accessibility of distributions, compliance of datasets to metadata formats, and source of violations.

The MQA dashboard can be opened directly from the portal homepage.

User feedback on datasets

Users will be able to provide feedback on a dataset directly from the dataset detail page.

The system makes it possible to gather and extract all feedback received for all datasets and group those by data supplier, so that the feedback can be sent to the data supplier.

4.6.5 Categorisation

The data.europa.eu categories are based on the EU controlled data theme vocabulary. The following are the categories used on data.europa.eu. When providing data, publishers should always use these terms to thematically categorise the datasets. If a different vocabulary is used, it should be aligned (i.e. mapped) to these categories.

Acronym	Definition
AGRI	Agriculture, fisheries, forestry and food
ECON	Economy and finance
EDUC	Education, culture and sport
ENER	Energy
ENVI	Environment
GOVE	Government and public sector
HEAL	Health
INTR	International issues
JUST	Justice, legal system and public safety
REGI	Regions and cities
SOCI	Population and society
TECH	Science and technology
TRAN	Transport

4.6.6 Auto translation

Harvested datasets are automatically translated in all official languages of the European Union. Only title and description of datasets and distributions are translated. The machine translation service eTranslation of the European Commission is used. The translation is asynchronously applied. Therefore, an appropriate disclaimer is displayed on the front end. The translated text is added to the DCAT-AP metadata to the corresponding properties with a language tag as specified in the DCAT-AP specification, e.g. "fr-t-en-t0-mtec", which means "translated from English to French using machine translation service by EC".

4.6.7 Supported formats and protocols

DCAT-AP via OAI-PMH is the preferred way of data harvesting. We can also accept data provided via CKAN APIs. However, we recommend that this is solution is only used for legacy systems.The following sections describe the list of interfaces that data suppliers (e.g. national portals, public data portals in the Member States, portals from international organisations etc.) must have in place in order to be harvested by data.europa.eu.

The main supported interfaces are the following:

- DCAT-AP;

- Comprehensive Knowledge Archive Network (CKAN) compliant sites;

- CSW/Inspire catalogue services (for geospatial datasets).

4.6.8 DCAT-AP

Providing data via a DCAT-AP interface is the official recommended method and will always be preferred for harvesting.

DCAT-AP is a metadata specification for describing public sector datasets in Europe. It's based on the data catalogue vocabulary rdf/xml, n-triples or turtle is allowed.

Metadata model

For general information on the metadata model, please refer to the official documentation. The respective qualifiers (mandatory, recommended and optional) need to be adhered to.

The following is an example dataset with all the mandatory properties in rdf/xml.

```xml
<?xml version="1.0"?>
<rdf:RDF xmlns:edp="https://europeandataportal.eu/voc#"
        xmlns:dct="http://purl.org/dc/terms/"
        xmlns:spdx="http://spdx.org/rdf/terms#"
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:j.0="http://data.europa.eu/88u/ontology/dcatapop#"
        xmlns:adms="http://www.w3.org/ns/adms#"
        xmlns:dqv="http://www.w3.org/ns/dqv#"
        xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
        xmlns:skos="http://www.w3.org/2004/02/skos/core#"
        xmlns:schema="http://schema.org/"
        xmlns:dcat="http://www.w3.org/ns/dcat#"
        xmlns:foaf="http://xmlns.com/foaf/0.1/">
    <dcat:Dataset rdf:about="http://data.europa.eu/88u/dataset/measures-of-solidarity-with-ukraine-eur">
        <dct:alternative xml:lang="en">EU solidarity with Ukraine</dct:alternative>
        <dct:subject rdf:resource="http://eurovoc.europa.eu/1914"/>
        <dcat:distribution>
            <dcat:Distribution rdf:about="http://data.europa.eu/88u/distribution/4608d86c-fe4e-4ae4-9fe7-a550b8d44b28">
                <dcat:accessURL rdf:resource="https://europa.eu/!jb8pvX"/>
                <dct:title xml:lang="en">List of measures</dct:title>
                <dcat:downloadURL rdf:resource="https://europa.eu/!jb8pvX"/>
                <dct:type rdf:resource="http://publications.europa.eu/resource/authority/distribution-type/DOWNLOADABLE_FILE"/>
                <dct:rights>
                    <dct:RightsStatement rdf:about="http://publications.europa.eu/resource/authority/access-right/PUBLIC"/>
                </dct:rights>
                <dct:license>
                    <dct:LicenseDocument rdf:about="http://publications.europa.eu/resource/authority/licence/CC_BY_4_0"/>
                </dct:license>
                <dct:description xml:lang="en">List of measures in csv format</dct:description>
                <dcat:mediaType>
                    <dct:MediaType rdf:about="http://www.iana.org/assignments/media-types/text/csv"/>
                </dcat:mediaType>
                <dct:format>
                    <dct:MediaTypeOrExtent rdf:about="http://publications.europa.eu/resource/authority/file-type/CSV"/>
                </dct:format>
                <dct:identifier>http://data.europa.eu/88u/distribution/813268f8-95e0-46b5-97fc-bc59b093e3d0</dct:identifier>
            </dcat:Distribution>
        </dcat:distribution>
        <dcat:distribution>
            <dcat:Distribution rdf:about="http://data.europa.eu/88u/distribution/6e39a598-5732-4204-b7c3-4e3abb3ef006">
                <dct:format>
                    <dct:MediaTypeOrExtent rdf:about="http://publications.europa.eu/resource/authority/file-type/HTML"/>
                </dct:format>
                <dct:title xml:lang="en">List of measures</dct:title>
                <dct:rights rdf:resource="http://publications.europa.eu/resource/authority/access-right/PUBLIC"/>
                <dcat:accessURL rdf:resource="https://europa.eu/!bCfRRj"/>
                <dcat:downloadURL rdf:resource="https://europa.eu/!bCfRRj"/>
                <dct:identifier>http://data.europa.eu/88u/distribution/6ae2d223-11d6-469e-8f88-7436f77d34fa</dct:identifier>
```

```
                <dct:type rdf:resource="http://publications.europa.eu/resource/authority/distribution-type/DOWNLOADABLE_FILE"/>
                <dct:license rdf:resource="http://publications.europa.eu/resource/authority/licence/CC_BY_4_0"/>
                <adms:status rdf:resource="http://purl.org/adms/status/Completed"/>
                <dcat:mediaType>
                    <dct:MediaType rdf:about="http://www.iana.org/assignments/media-types/text/html"/>
                </dcat:mediaType>
                <dct:description xml:lang="en">List of measures in html format with actionable links</dct:description>
            </dcat:Distribution>
        </dcat:distribution>
        <dct:accessRights rdf:resource="http://publications.europa.eu/resource/authority/access-right/PUBLIC"/>
        <dcat:landingPage>
            <foaf:Document rdf:about="https://eur-lex.europa.eu/content/news/eu-measures-solidarity-ukraine.html">
                <foaf:topic rdf:resource="http://data.europa.eu/88u/dataset/measures-of-solidarity-with-ukraine-eur"/>
                <schema:url>https://eur-lex.europa.eu/content/news/eu-measures-solidarity-ukraine.html</schema:url>
            </foaf:Document>
        </dcat:landingPage>
        <dct:publisher>
            <foaf:Organization>
                <foaf:mbox rdf:resource="mailto:example@europa.eu"/>
                <foaf:name>
                    Landesamt für Digitalisierung, Breitband und Vermessung
                </foaf:name>
                <foaf:homepage rdf:resource="http://www.europa.eu"/>
            </foaf:Organization>
        </dct:publisher>
        <dct:issued rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2022-03-17</dct:issued>
        <dct:description xml:lang="en">This dataset contains a list of documents published on EUR-Lex that bring together the
measures the EU has taken in solidarity with Ukraine. This list includes measures of assistance to Ukraine as well as related
restrictive measures. Acts amended in the above mentioned context are also included. The list is updated regularly.</dct:description>
        <dct:accrualPeriodicity>
            <dct:Frequency rdf:about="http://publications.europa.eu/resource/authority/frequency/IRREG"/>
        </dct:accrualPeriodicity>
        <dct:title xml:lang="en">Measures in solidarity with Ukraine</dct:title>
        <dct:language>
            <dct:LinguisticSystem rdf:about="http://publications.europa.eu/resource/authority/language/ENG"/>
        </dct:language>
        <dcat:contactPoint>
            <vcard:Kind rdf:about="http://www.w3.org/2006/vcard/ns#Kind/be870795-ffe2-4117-885d-191faca58bb6">
                <vcard:hasEmail rdf:resource="mailto:EURLEX-HELPDESK@publications.europa.eu"/>
                <vcard:hasURL>
                    <foaf:Document rdf:about="https://eur-lex.europa.eu/contact.html">
                        <schema:url>https://eur-lex.europa.eu/contact.html</schema:url>
                        <foaf:topic rdf:resource="file:///usr/verticles/"/>
                    </foaf:Document>
                </vcard:hasURL>
            </vcard:Kind>
        </dcat:contactPoint>
    </dcat:Dataset>
</rdf:RDF>
```

Requests

The harvester currently supports harvesting from an open archives initiative protocol for metadata harvesting (OAI-PMH) compliant source or from reading a dump file containing the RDF/XML representation of the datasets or directly reading DCAT-AP from a SPARQL endpoint. If datasets are provided as a dump file, it is recommended to split the file into pages, for example, by using the hydra core vocabulary.

For OAI-PMH-compliant sources, only the verb "ListRecords" is used.

Responses

As indicated above, the response must be DCAT-AP-compliant to be understood by the harvesting component.

Error handling

The OAI-PMH protocol provides methods for error handling that the harvester can understand. When using this protocol, these error methods should be used.

Service information for integration

As stated above, a categorisation mapping should be provided. Apart from that, the URL for the OAI-PMH endpoint or the dump file is needed.

4.6.9 CKAN API

The open-source data portal platform CKAN is still used by various data portals. Its Remote-Procedure-Call-style API (action API) is supported as an interface for data suppliers of data.europa.eu. The following options for using that interface are available:

- The data supplier uses CKAN for providing its open data metadata. It is important that the used CKAN version supports the action API. The legacy APIs of CKAN are not supported.
- The data supplier offers a CKAN compliant API, where the necessary endpoints reproduce the exact API behaviour.

Requests and responses

Only the 'package_search' API endpoint is needed in order to harvest the metadata. Its specifications are described in detail in the official documentation. This endpoint is used to get the metadata in a paginated way. Therefore, it accepts query parameters in a request and returns a dictionary with datasets as a result. The high-level use of this endpoint must be offered as specified in the CKAN documentation.

Example call: `GET http://www.example.com/api/3/action/package_search?rows=50`

Metadata model

Although the CKAN API can be used as is, the basic CKAN data schema was extended and modified to meet the requirements of the underlying data structure (DCAT-AP) of the data.europa.eu. The response of the 'package_search' action exposes a 'results' field, which holds a list of dictised datasets. The data structure of such a dataset differs from the one of a plain CKAN installation.

Please note:

- Fields marked with an asterisk (*) are CKAN standard. Further information in the official documentation.
- Type specifications according to official JSON standard (http://json.org/).
- Besides the mandatory fields, the field names and types are not strict, but data suppliers must make sure an obvious mapping is possible.
- For a detailed explanation of each field, refer to the DCAT-AP specifications.

Dataset schema

The following fields are mandatory.

Field	Type	DCAT-AP dataset equivalent
title *	string	dct:title
notes *	string	dct:description

The following fields are optional but **highly recommended**.

Field	Type	DCAT-AP dataset equivalent
contact_point	array of objects (allowed members: type, name, email, resource)	dcat:contactPoint
tags *	array of objects	dcat:keyword
publisher	object	dct:publisher
groups *	array of objects – the name of each group needs to fit the official categorisation (see Error! Reference source not found..)	dcat:theme
resources *	array	of objects (see distribution schema)

The following fields are optional.

Field	Type	DCAT-AP dataset equivalent
conforms_to	array of objects (allowed members: label, resource)	dct:conformsTo
creator	object	dct:creator
accrucal_periodicity	object	dct:accrucalPeriodicity
identifier	object	dct:identifier
url	string	dcat:landing_page
language	array of objects (allowed members: label, resource)	dct:language
other_identifier	object	adms:identifier
issued	string	dct:issued
dcat_spatial	array of objects (allowed members: label, resource)	dct:spatial
temporal	array of objects (allowed members: start_date, end_date)	dct:temporal
modified	string	dct:modified
version_info	string	owl:versionInfo
version_notes	string	adms:versionNotes
provenance	array of objects (allowed members: label, resource)	dct:provenance
source	array of strings	dct:source
access_rights	object	dct:accessRights
has_version	array of strings	dct:hasVersion
is_version_of	array of strings	dct:isVersionOf
relation	array of strings	dct:relation
page	array of strings	foaf:page
sample	array of strings	adms:sample
dct_type	string	dct:type

Distribution schema

The following fields are mandatory.

Field	Type	DCAT-AP distribution equivalent
url	string	dcat:accessURL

The following fields are optional but **highly recommended**.

Field	Type	DCAT-AP distribution equivalent
description	string	dct:description
format	string	dct:format
license	object	dct:license

Note that the list of licences recognised by data.europa.eu's DCAT-AP parser is available online (https://data.europa.eu/en/training/licensing-assistant). This is also used by our metadata quality assessment (MQA) tool () for assessing the data providers' performance in using known licences.

The following fields are optional.

Field	Type	DCAT-AP distribution equivalent
checksum	object	spdx:checksum
mimetype	string	dcat:mediaType
download_url	array of strings	dcat:downloadURL
issued	string	dct:issued
status	object	adms:status
name	string	dct:title
modified	string	dct:modified
rights	object	dct:rights
page	array of strings	foaf:page
size	number	dcat:byteSize
language	array of objects	dct:language
conforms_to	array of objects	dct:conformsTo

Example

A result of the 'package_search' action looks like this.

```
{
  "help": "http://example.eu/data/api/3/action/help_show?name=package_search",
  "success": true,
  "result": {
    "count": 113948,
    "sort": "score desc, metadata_modified desc",
    "facets": {},
    "results": [
```

```
{
    "issued": "2011-10-20T00:00:00Z",
    "id": "525abe30-ef60-4bf9-824e-916368c1fad8",
    "metadata_created": "2015-09-15T12:08:54.860742",
    "metadata_modified": "2015-09-15T13:17:51.405474",
    "temporal": [
        {
            "start_date": "2011-10-19T22:00:00Z",
            "end_date": "2011-10-19T22:00:00Z"
        }
    ],
    "state": "active",
    "type": "dataset",
    "resources": [
        {
            "package_id": "525abe30-ef60-4bf9-824e-916368c1fad8",
            "id": "7166a1fa-d994-4d88-8e76-3378930b1e16",
            "state": "active",
            "format": "XHTML",
            "mimetype": "application/xhtml+xml",
            "name": "Example",
            "created": "2015-09-15T14:39:43.865240",
            "url": "http://example.com"
        }
    ],
    "tags": [
        {
            "vocabulary_id": null,
            "state": "active",
            "display_name": "Example Tag",
            "id": "06993102-a2ee-4e40-b9e4-ed3e4b86e943",
            "name": "example-tag"
        }
    ],
    "groups": [
        {
            "display_name": "Economy and finance",
            "description": "",
            "title": "Economy and finance",
            "id": "128d0956-4526-440e-a951-f153c190d890",
            "name": "economy-and-finance"
        }
    ],
    "creator_user_id": "0ab3c2ec-c2a2-4eef-b70f-ed093e028063",
    "publisher": {
        "resource": "http://example.com "
    },
    "organization": {
        "description": "Example Organization",
        "created": "2015-09-15T13:56:32.985936",
        "title": "Example Organization",
        "name": "example-orag",
        "is_organization": true,
        "state": "active",
        "image_url": "",
        "revision_id": "ea70fb1f-29a8-4e7b-8527-809e4792a75b",
        "type": "organization",
        "id": "0897b420-3c3d-4a19-9c2c-a9815e2db2be",
        "approval_status": "approved"
    },
    "name": "example-dataset",
    "notes": "Example",
    "owner_org": "0897b420-3c3d-4a19-9c2c-a9815e2db2be",
    "modified": "2011-10-20T00:00:00Z",
    "url": "",
    "title": "Example Dataset",
    "identifier": [
        "http://example-ident.com"
    ]
    }
    ],
    "search_facets": {}
    }
}
```

4.6.10 CSW/INSPIRE catalogue services (for geospatial metadata)

General remarks

This interface represents an INSPIRE compliant catalogue (discovery) service. It is defined as a slightly extended version of the OGC CSW AP ISO.

The GetCapabilities operation (mandatory for all OGC Services) is not needed for running the harvesting. But this operation could be helpful upon registration of the catalogue service within the EU Data Portal as the service's response provides additional information which must otherwise be found out during the registration (e.g. the supported protocol bindings or the support of the 'modified' queryable for selective harvesting).

For the harvesting process only the GetRecords operation will be called. The GetRecordById is not needed.

Operation	Operation description	data.europa.eu usage
GetCapabilities	Retrieve catalogue service metadata	Only for gathering service information upon registration
GetRecords	Retrieval of a bunch of metadata items	Yes
GetRecordById	Retrieval information of single metadata items	No

Metadata model

The metadata model considered is as defined in the INSPIRE Technical Guidance on Discovery Services and on Metadata.

Within a GetRecords query (constraint) just the following metadata model elements (queryables) are used.

Request parameter	Definition a	Used values in data.europa.eu	XPath b
Type	Provides the desired information resources.	Always the following fixed values used: 'dataset', 'datasetcollection','series' and 'service'	/gmd:MD_Metadata/ gmd:hierarchyLevel/ gmd:MD_ScopeCode/ @codeListValue
Modified	The metadata date stamp in case of selective harvesting (if supported), see below.	Date	/gmd:MD_Metadata/ gmd:dateStamp/gco:Date

a: 'Definition' represents the semantic meaning of element in data.europa.eu, it is slightly different from the genetic meaning in OGC CSW.

b: Element's XML path in GetRecords request.

Example query (constraint).

```
<Constraint version="1.1.0">
    <ogc:Filter>
        <ogc:Or>
            <ogc:PropertyIsEqualTo>
                <ogc:PropertyName>Type</ogc:PropertyName>
                <ogc:Literal>dataset</ogc:Literal>
```

```
            </ogc:PropertyIsEqualTo>
            <ogc:PropertyIsEqualTo>
                <ogc:PropertyName>Type</ogc:PropertyName>
                <ogc:Literal>datasetcollection</ogc:Literal>
            </ogc:PropertyIsEqualTo>
            <ogc:PropertyIsEqualTo>
                <ogc:PropertyName>Type</ogc:PropertyName>
                <ogc:Literal>series</ogc:Literal>
            </ogc:PropertyIsEqualTo>
            <ogc:PropertyIsEqualTo>
                <ogc:PropertyName>Type</ogc:PropertyName>
                <ogc:Literal>service</ogc:Literal>
            </ogc:PropertyIsEqualTo>
        </ogc:Or>
    </ogc:Filter>
</Constraint>
```

As defined in the INSPIRE Technical Guidance on Discovery Services the operation must be able to return ISO19139 metadata aligned with the INSPIRE Technical Guidance on Metadata.

Requests

The mandatory GetRecords operation works as the primary means of metadata item discovery with HTTP protocol binding. It executes an inventory search and returns the metadata items. Only OGC Filter XML encoding is supported. For the GetRecords requests a few additional requirements exist. These will be explained in the following.

Bindings

One or more of HTTP POST/XML, POST/XML/SOAP1.1 and POST/XML/SOAP1.2 have to be supported as bindings.

Operation parameters

The following parameters (not the queryables) and parameter values are used in data.europa.eu for the GetRecords requests.

Request parameter	Definition a	Used values in data.europa.eu	XPath b
service	Tells this is a CSW service.	Always fixed value: CSW	/GetRecords@service
version	Tells which version of CSW service is requested.	Always fixed value; 2.0.2	/GetRecords@version
resultType	Specifies the type of result	Always fixed value: 'results'	/GetRecords@resultType
outputFormat	Specifies the output format of GetRecords returned document	Always fixed value: 'application/xml'	/GetRecords@outputFormat
outputSchema	Specifies the schema of GetRecords returned document	Always fixed value (namespace): 'http://www.isotc211.org/2005/gmd'	/GetRecords@outputSchema
startPosition	Specifies the sequence number of first returned record	Used: integer between 1 and returned number Default value is 1	/GetRecords@startPosition
maxRecords	Specifies number of returned records	Supported: positive integer between 1 and N. Default value is: 50	/GetRecords@maxRecords
typeNames	Specifies the query- and elementSetName type	Always fixed value: 'gmd:MD_Metadata' 'gmd' is valid namespace prefix for 'http://www.isotc211.org/2005/gmd'	/GetRecords/Query@typeName and /GetRecords/Query/ElementSetName@typeName
ElementSetName	Specifies the type of GetRecords returned document	As only full metadata sets will be requested by the harvester this parameter will always be set to 'full'.	/GetRecords/Query/ElementSetName

a: 'Definition' represents the semantic meaning of element in data.europa.eu. it is slightly different from the genetic meaning in OGC CSW.

b: Element's XML path in GetRecords request.

Partitioning

For partitioning (pagination) the following parameters are used (see table on GetRecords):

- startPosition;
- maxRecords.

Selective harvesting

Selective harvesting allows harvesters to limit harvest requests to just those portions of the metadata available from a repository which have been changed within a specified time frame.

Selective harvesting often makes sense as this would require that only a few metadata records be harvested daily as only a few metadata records are changed within a day.

For selective harvesting the predefined queryable (usually 'modified' GetRecords) is used.

Responses

As defined in the INSPIRE Technical Guidance on Discovery Services the operation must be able to return ISO19139 metadata aligned with the INSPIRE Technical Guidance on Metadata.

Partitioning

For partitioning (pagination) as part of the search response, it is **mandatory** to have the total count of matching metadata items returned, even if the metadata for these items is not contained in the search response. This parameter, coupled with the ability to specify the startPosition and the number of desired records (maxRecs) from the catalogue , will allow data.europa.eu to implement results paging and reducing the load on both the data.europa.eu system and on the data partners.

Error handling

Useful status and error messages help data.europa.eu manage client sessions effectively. Any limitations on submitted search requests to the inventory systems should be noted in the response (e.g. 'too many records requested', 'search timed out') so that predictable error handling can be managed by data.europa.eu.

Service information for integration

To be able to integrate an INSPIRE Discovery Service /CSW, the following information need to be provided by the data supplier.

Service information	Definition a	Obligation (M=Mandatory, O=Optional, C=Conditional)	Datatype
GetRecords URL	URL of the CSW GetRecords operation	M	URL
GetRecords Binding	URL of the CSW GetRecords operation	M	Codelist (one of): 'POST/XML', 'POST/XML/SOAP1.1' 'POST/XML/SOAP1.2'
Modified (a)	Name of the queryable (if supported) for filtering on the metadata date stamp (for selective harvesting)	Possibly for future use	String. [Namespace":"]QueryableName
MaxRecordsMax	Specifies the maximal number of maximal returned records	Possibly for future use (currently always set to '50')	Integer

a = Value in CSW filter will be formatted as 'MM-DD-YYYY'. Operators: '>=', '<=' will be used.

4.7 Harvesting guidelines for the European Register for Protected Data (ERPD) held by the Public Sector

This section is based on v1.3 of the Harvesting Guidelines for the European Register for Protected Data held by the Public Sector.

4.7.1 Introduction

The Data Governance Act (DGA) is a regulation of the European Parliament and the Council to facilitate data sharing within the EU's Single Market. It came into force on 23rd June 2022 and covers private companies, citizens, and public sector bodies. The Act will be applicable from 24th September 2023. The Data Governance Act includes mechanisms to foster the reuse of public sector data that, for certain reasons, cannot be made available as open data. This could be health or mobility data that are, according to Article 3 of the Data Governance Act, protected on grounds of statistical or commercial confidentiality, intellectual property rights of third parties, or privacy.

The most basic lever to promote data reuse is simply to increase its discoverability, enabling potential users to understand which data public sector institutions hold. For this, the publication of metadata containing information about existing datasets is required. Accordingly, Article 8 of the Data Governance Act instructs Member States to establish so-called National Single Information Points (NSIPs). NSIPs are intended to serve as national one-stop-shop allowing any user, such as citizens, entrepreneurs, or researchers, to search for and find information about the affected public sector data in their Member States. Because the relevant data, such as health or mobility data, is protected, users will not be able to readily access that same data. Instead, they will have to issue an access request. Therefore, NSIPs must not just enable users to search for protected data, but they must also offer means for users to understand how they can gain access and, if applicable, offer a direct procedure to issue access requests to the relevant public sector bodies.

Additionally, the DGA also specifies that the European Commission shall establish a European Register for Protected Data held by the Public Sector (ERPD) to collect, partially mirror, and render discoverable the NSIPs' data in a European register. According to recital 26 of the DGA, NSIPs and the ERPD can be implemented either as entirely new infrastructures or as part of existing (open) data infrastructures.

This document provides an overview of how the ERPD will be implemented by the European Commission. For this, it specifies how metadata must be provided and structured by the NSIPs to interface with the ERPD. Member States may also draw from this document their own conclusions on how to align, by design, their own NSIP with the ERPD, thus facilitating harmonisation across NSIPs. However, the primary purpose of these harvesting guidelines is to formulate specific requirements for the unidirectional harvesting of metadata from NSIPs into the ERPD – and to advise Member States on how to establish their NSIP accordingly.

4.7.2 The European Register for Protected Data held by the Public Sector on data.europa.eu

Based on a common decision by the Publications Office and DG CNECT, the European Register for Protected Data held by the Public Sector will be integrated into and implemented as part of data.europa.eu. This means that the ERPD will leverage the existing infrastructure and general approach to data harvesting and management. However, the advantages of integrating the ERPD into data.europa.eu go well beyond mere technicalities: This approach allows users to easily search and find metadata relating to open data (from open data portals, already available on data.europa.eu) and non-open data (from NSIPs, newly added) in one place,

fostering data uptake and reuse along a single value chain. Data.europa.eu offers rich, multilingual search features, it is well known in the community thanks to its manyfold activities around data, and it is constantly with new, useful features added on an ongoing basis. This means that the ERPD and ERPD users will benefit from a head start and from future developments of data.europa.eu.

The integration of the ERPD in data.europa.eu can go also go ahead with relative ease, offering advantages to NSIPs and data.europa.eu. The implementation on data.europa.eu will ensure that all NSIP data included in the ERPD are arranged in a dedicated catalogue structure and is searchable via filters. This ensures that, on the portal, the non-open NSIP data can be separated from other open data. From the perspective of data publishers, the implementation along the lines of data.europa.eu's existing operating model offers various advantages, too. In particular, the existing harvesting approach and overall data architecture can be adapted relatively easily. This means that relevant national data publishers can build on their experience with data.europa.eu's proven, production-level system rather than having to adapt to an entirely new development.

The Data Governance Act does not provide concrete technical instructions for the implementation of the NSIPs and the ERPD, neither on the technology or data level. The harvesting guidelines presented here are therefore based on an interpretation of the relevant articles of the Data Governance Act. This concerns in particular Articles 5-8 of the Data Governance Act.

The harvesting guidelines presented in this document build largely on the DCAT-AP data schema. Over the past decade, DCAT-AP has become the preferred choice for EU Member States to describe their open data. It is a well-documented, comprehensive, and flexible data schema that is the most convenient choice for data publishers and users, thanks to its widespread adoption and high interoperability. All required metadata presented in section 4 of these guidelines can be represented via existing DCAT-AP properties. Differences are only present in the hierarchy of required metadata. On the level of required metadata, precisely four additional metadata properties are added. These include information on publishers and conditions for the re-use of data (both relating to the dataset metadata level; dct:publisher and dct:rights) as well as information on individual the format and size of individual distributions (dct:format and dcat:byteSize). For further information, please see the section on required metadata.

In keeping with data.europa.eu's general data governance, the harvesting guidelines for the European Register for Protected Data held by the Public Sector demand that metadata is provided in a of DCAT-AP-compliant manner. Metadata that can be identified as being mandatory from the Data Governance Act are therefore mapped into DCAT-AP properties and must be structured correspondingly by NSIPs to enable harvesting. This relates to the requested information on the titles of datasets, their descriptions, their publisher, conditions for reuse and access procedure, format, and size. NSIPs are also required to separate provided metadata into datasets and distributions, in line with the DCAT-AP logic. A dataset can have multiple distributions, e.g. offering a file with the same data in either XML, JSON, or CSV.

The ERPD data architecture follows data.europa.eu's known hierarchical structure with minor modifications. Once harvested, the NSIP metadata on datasets and distributions are automatically grouped into country specific NSIP catalogues in the ERPD. These catalogues are referred to as NSIP country catalogues and are labelled according to the following schema "Country National Single Information Point". This means that for each country, only one NSIP catalogue exists, e.g. "Italy National Single Information Point". The publisher of this catalogue will automatically be the relevant NSIP. Unlike in the case of data.europa.eu's existing open data registry, there is no option to create catalogues for individual institutions below the NSIP level (e.g. a specific catalogue for the statistical office of a given country). Furthermore, to ensure a consistent classification of NSIP data aside of data.europa.eu's open-data-related metadata, all NSIP data are sorted into a meta-catalogue labelled "European Register for Protected Data held by the Public Sector". This data structure allows users to target their search and retrieve metadata either across the entire range of the ERPD

(i.e. discovering data across all NSIPs) or limiting their search to only one NSIP. Furthermore, users will also be able to search across the entire data range of data.europa.eu, locating both open data and ERPD data alongside each other.

The following sections provide detailed technical instructions on the required metadata, formatting, as well as supported interfaces.

4.7.3 Requirements

The following requirements specify which and how data must be provided the NSIPs to enable harvesting by the ERPD.

4.7.4 Segregation of NSIP data

As stated before, the Data Governance Act states in recital 26 that MS could either implement NSIPs as entirely new, standalone infrastructures or that they might repurpose the existing open data (portal) infrastructure to also register the metadata on non-open public sector data.

The requirements stated in this document do not favour either solution. It is evident that, if an NSIP is implemented as a standalone solution, the metadata from that NSIP is delivered by a standalone endpoint as well.

However, if your NSIP is implemented as part of an existing (national) open data infrastructure, it is essential that you offer a means by which NSIP data can be separated from other, open data on the same portal. For this, your NSIP must offer:

- A dedicated endpoint exclusively for the harvesting of NSIP data, **OR**
- A filter mechanism for the endpoint that allows the retrieval of NSIP data only. For the avoidance of doubt, this implies that any filtering must take place at the endpoint. Furthermore, the endpoint must allow filtering for NSIP data only as well as filtering for open data only.

These requirements must be met to ensure that data from your NSIP can be correctly ingested into the ERPD on data.europa.eu. The correct ingestion of data by data.europa.eu relies on a mode of data provisioning that makes, in some way, NSIP data separable from other, open data that might be provided via the same portal.

4.7.5 Technical requirements

The harvester accesses the NSIP endpoints on a weekly basis. Depending on the total size of data provided by each NSIP as well as depending on other factors such as available resources, other harvesting intervals can be negotiated on an individual basis. Metadata data is processed overnight. Every incoming non-DCAT-AP-format will be transformed to DCAT-AP 2.1.1. The harvester is configured individually for each harvested NSIP.

4.7.6 Access to harvested sites

Authentication

Some source sites require authentication, this means we need a login name and password before we can access the data. If this applies to your portal, please state this in your message when using our contact form.

API access to harvested site

To be able to be harvested, the source site needs to expose an endpoint from which the data can be gathered.

This endpoint should, as described in 3.1, offer the ability to only provide the NSIP data. Either by being a dedicated endpoint or by offering a filter which includes only NSIP data.

The harvested NSIP data should have one of the formats described in section 6 and the endpoint should use one of the protocols described in this documentation. The preferred combination is DCAT-AP via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), but others are also possible.

FTP access to harvested site

The ERPD on data.europa.eu does not support FTP for downloading datasets from a source site.

4.7.7 Operational requirements

Harvesting frequency

Due to the high volume of metadata that will be harvested from a growing list of data suppliers and the required runtime for the harvesting processes, data supplier sites are harvested weekly by default. Furthermore, the harvesting processes must be clustered and scheduled on a fixed time schedule (e.g. during the night) in order to avoid any load impacts on the harvested sites during regular business hours usage. Other factors and circumstances permitting, harvesting intervals that are more or less frequent can be agreed individually.

Data source site API / endpoint

The data source endpoint should accept queries with, for example, offset / limit parameters for resumption, partitioning, and pagination of the datasets to be harvested.

Ensuring uniquely identifiable datasets

Only when the same dataset always has the same unique id it can be ensured that it will be recognized as the same dataset on data.europa.eu and that it will not be duplicated.

4.7.8 Required metadata

The following metadata is required for NSIPs and the ERPD according to Articles 5-8 of the Data Governance Act: title, description, publisher, conditions for re-use / access procedure, format, and size. To enable a comprehensive harvesting of required metadata, equivalent metadata is mandatory for ERPD harvesting according to these guidelines.

The following tables show how the metadata required by the Data Governance Act are modelled to corresponding DCAT-AP properties based on DCAT-AP (version 2.1.1). As explained in section 2, the required metadata must be modelled into datasets and distributions. Because the Data Governance Act requires metadata that pertains to the datasets and distribution classes, metadata on both classes must be provided as specified below to fully meet the Act's legal requirements. The data service class can be used in specific cases to specify metadata on data endpoints. Metadata properties that are generally mandatory according to the DCAT-AP specification are marked by an asterisk (*) in the tables below.

Both dataset and distribution level metadata must be made available for harvesting. Sections 4.1 and 4.2 list mandatory metadata required for dataset and distribution classes. For the avoidance of doubt, information on the underlying distribution(s) is also required via the dcat:Distribution property.

Datasets According to DCAT-AP, a dataset is a collection of data, published or curated by a single agent. Data comes in many forms including numbers, words, pixels, imagery, sound and other multi-media, and potentially other types, any of which might be collected into a dataset.

The following metadata is mandatory for NSIP datasets:

Property	URI	Range	Usage note	Cardinality
Title (M)	dct:title *	rdfs:Literal	This property contains a name given to the Dataset. This property can be repeated for parallel language versions of the name.	1..n
Description (M)	dct:description *	rdfs:Literal	This property contains a free-text account of the Dataset. This property can be repeated for parallel language versions of the description.	1..n
Publisher (M)	dct:publisher	foaf:Agent	This property refers to an entity (organisation) responsible for making the Dataset available.	1..1
Access rights (M)	dct:accessRights	dct:RightsStatement	This property refers to information that indicates whether the dataset is open data, has access restrictions, or is not public. From the controlled vocabulary of the Publications Office of the EU 6, the following codes should be used for NSIP data: "non-public" or "restricted". "open" is prohibited for NSIP data.	1..1
	dct:distribution	dcat:Distribution		1..n

Property	URI	Range	Usage note	Cardinality
Distribution (M)			Distribution(s) available for a dataset.	

Further properties that are either recommended or optional according to the DCAT-AP specification can be provided at the NSIP's or data publisher's discretion.

As specified in the section on the segregation of NSIP data, data.europa.eu will differentiate open data and NSIP-related data in one of two ways: Either NSIP-related data is delivered via a dedicated endpoint or, if an endpoint delivers both open data and NSIP-data, data publishers must specify a mechanism that allows data.europa.eu to filter (and therefore separate) NSIP-data from open data. The exact filter logic can be agreed individually as part of the harvesting on-boarding process. Both mechanisms enable data.europa.eu to differentiate and then label the relevant data as non-open data that has originated from an NSIP.

Ongoing discussions in SEMIC might lead to future extensions or modifications of DCAT-AP that specify additional properties or vocabularies, e.g. to indicate the legislation that led to the release of a dataset. If such additions were to be adopted by SEMIC and implemented by Member States, data publishers will of course be able to provide such data to the ERPD. data.europa.eu's triple store backend will accept and save such data.

However, if NSIP data and open data are provided via the same endpoint, it should be stressed that such tagging will not invalidate the need for a filter mechanism that enables reliable segregation of data.

Distribution

A distribution according to DCAT-AP represents an accessible form of a dataset such as a downloadable file.

The following metadata is mandatory for distributions:

Property	URI	Range	Usage note	Cardinality
Format (M)	dct:format	dct:MediaTypeOrExtent	This property refers to the file format of the Distribution. You can only specify one format per Distribution. If an NSIP offers the same data in different formats, each format must be specified as a separate distribution.	1..1
Size (M)	dcat:byteSize	rdfs:Literal	The size in bytes can be approximated (as a decimal) if the precise size is not known. If data is offered via an API or other endpoint, size should refer to the overall size of the underlying dataset.	1..1
Access procedure (M)	dcat:accessURL *	rdfs:Resource	A URL of a Website that enables either access to the described data or that contains information on how to request the data.	1..n
Conditions for re-use (Rights) (M)	dct:rights	dct:RightsStatement	This property refers to a statement that specifies rights associated with the Distribution.	1..1

Further properties that are either recommended or optional according to the DCAT-AP specification can be provided at the NSIP's or data publisher's discretion.

If your NSIP includes metadata on endpoints (e.g. and APIs) that are directly accessible, we recommend that you use the DCAT-AP DataService class in addition to the Distribution. If you use the DataService class, you must include information on the properties endpointURL10 (i.e. the root location or primary endpoint of the service (an IRI), according to DCAT-AP 2.1.1) and title ((i.e. a name given to the Data Service; this property can be repeated for parallel language versions of the name according to DCAT-AP 2.1.1). Further recommended or optional DataService properties can be added at your discretion.

This means that this information should only be provided if an endpoint exists that can be accessed by external users without e.g. registration. In these cases, we recommend that the endpoint URL is provided in addition to the information on how to request access (as expressed by dcat:accessURL as part of Distributions). For cases where the relevant endpoints cannot be accessed without prior registration, information on the access URL of the relevant distribution is sufficient.

4.7.9 Supported formats and protocols

DCAT-AP via OAI-PMH is the preferred way of data harvesting. We can also accept data provided via CKAN APIs. However, we recommend that this is solution is only used for legacy systems, i.e. only in cases where NSIPs are implemented as part of existing, CKAN-based (open) data infrastructures.

4.7.10 DCAT-AP

Providing metadata as DCAT-AP is the officially recommended method and will always be preferred for harvesting.

General remarks

DCAT-AP is a metadata specification for describing public sector datasets in Europe. It is based on the data catalogue vocabulary. The datasets are provided as linked data and can be represented in multiple ways. For the harvesting process, any common representation like rdf/xml, n-triples, or turtle is allowed.

Metadata model For general information on the metadata model, please refer to the official documentation. The respective qualifiers (mandatory, recommended, and optional) need to be adhered to and, going beyond the core requirements of DCAT-AP, mandatory metadata as specified in the section on required metadata must be included.

Example

The following is an example dataset with all mandatory properties in rdf/xml.

```xml
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:dcat="http://www.w3.org/ns/dcat#"
        xmlns:dc="http://purl.org/dc/terms/"
        xmlns:foaf="http://xmlns.com/foaf/0.1/"
        xmlns:vcard="http://www.w3.org/2006/vcard/ns#"
        xmlns:locn="http://www.w3.org/ns/locn#"
        xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">

  <dcat:Dataset rdf:about="http://data.europa.eu/88u/dataset/ded24b58-a5ab-4d34-8603-23ded830bab2">
    <dc:publisher>
      <foaf:Agent rdf:about="http://publications.europa.eu/resource/authority/corporate-body/CCC">
        <foaf:name>Customs Cooperation Council</foaf:name>
      </foaf:Agent>
    </dc:publisher>
```

```
    <dcat:contactPoint>
      <vcard:Kind>
        <rdf:type rdf:resource="http://www.w3.org/2006/vcard/ns#Individual"/>
        <vcard:hasEmail rdf:resource="mailto:john@doe.de"/>
        <vcard:fn>John Doe</vcard:fn>
      </vcard:Kind>
    </dcat:contactPoint>

    <dcat:keyword>example</dcat:keyword>
    <dcat:theme rdf:resource="http://publications.europa.eu/resource/authority/data-theme/ENVI"/>
    <dc:title>NSIP example metadata dataset</dc:title>
    <dc:temporal>
      <dc:PeriodOfTime>
        <dcat:endDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-06-09T00:00:00</dcat:endDate>
        <dcat:startDate rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2015-06-09T00:00:00</dcat:startDate>
      </dc:PeriodOfTime>
    </dc:temporal>

    <dc:spatial>
      <dc:Location rdf:about="https://piveau.eu/def/example-location">
        <locn:geometry rdf:datatype="http://www.opengis.net/ont/geosparql#gmlLiteral">&lt;gml:Envelope srsName=&quot;http://
www.opengis.net/def/EPSG/0/4326&quot;&gt;&lt;gml:lowerCorner&gt;53.1485 12.915&lt;/gml:lowerCorner&gt;&lt;gml:upperCorner&gt;53.1985
12.9983&lt;/gml:upperCorner&gt;&lt;/gml:Envelope&gt;</locn:geometry>
      </dc:Location>
</dc:spatial>

    <dc:identifier>ded24b58-a5ab-4d34-8603-23ded830bab2</dc:identifier>
    <dc:description>This is an minimal example dataset to showcase the metadata to be offered by an NSIP</dc:description>
    <dc:accessRights>
      <dc:RightsStatement rdf:about="http://publications.europa.eu/resource/authority/access-right/RESTRICTED"/>
    </dc:accessRights>

    <dcat:distribution>
      <dcat:Distribution rdf:about="http://data.europa.eu/88u/distribution/a5be938b-a5ab-4d34-8603-cabf323af6ee">
        <dc:format>
          <dc:MediaTypeOrExtent rdf:about="http://publications.europa.eu/resource/authority/file-type/PDF"/>
        </dc:format>

        <dc:title>NSIP example metadata distribution</dc:title>
        <dc:description>An Example Distribution for the NSIP Example Dataset</dc:description>
        <dc:identifier>https://nsip.data.example.com/dataset/ded24b58-a5ab-4d34-8603-23ded830bab2/resource/a5be938b-a5ab-4d34-8603-
cabf323af6ee</dc:identifier>
        <dcat:accessURL rdf:resource="https://nsip.data.example.com/dataset/ded24b58-a5ab-4d34-8603-23ded830bab2/access"/>
        <dcat:byteSize rdf:datatype="http://www.w3.org/2001/XMLSchema#decimal">18006.0</dcat:byteSize>
        <dc:rights>
          <dc:RightsStatement rdf:about="http://example-rights.com"/>
        </dc:rights>

      </dcat:Distribution>
    </dcat:distribution>
  </dcat:Dataset>

</rdf:RDF>
```

Requests

The harvester currently supports harvesting from an OAI-PMH compliant source or from reading a dump file containing the RDF/XML representation of the datasets or directly reading DCAT-AP from a SPARQL endpoint. If datasets are provided as a dump file, it is recommended to split the file into pages, for example, by using the hydra core vocabulary 3.

For OAI-PMH-compliant sources, only the verb 'ListRecords' is used.

Responses

As indicated above, the response must be DCAT-AP-compliant to be understood by the harvesting component.

Error handling

The OAI-PMH protocol provides methods for error handling that the harvester can understand. When using this protocol, these error methods should be used.

Service information for integration

As stated above, a categorisation mapping should be provided. Apart from that, the URL for the OAI-PMH endpoint or the dump file is needed.

4.7.11 CKAN API

The open-source data portal platform CKAN is still used by various open data portals. Its RPC-style API (action API) is supported as an interface for data suppliers of data.europa.eu. This support will also apply to the ERPD on data.europa.eu. The following options for using that interface are available.

- The data supplier uses CKAN for providing its NSIP metadata. It is important that the used CKAN version supports the action API. The legacy APIs of CKAN are not supported.
- The data supplier offers a CKAN compliant API, where the necessary endpoints reproduce the exact API behaviour.

Requests and responses

Only the 'package_search' API endpoint is needed to harvest the metadata. Its specifications are described in detail in the official documentation. This endpoint is used to get the metadata in a paginated way. Therefore, it accepts query parameters in a request and returns a dictionary with datasets as a result. The high-level use of this endpoint must be offered as specified in the CKAN documentation. Example call:

```
GET http://singleinformationpoint.tld/api/3/action/package_search?rows=50
```

Metadata model

Although the CKAN API can be used as is, the basic CKAN data schema was extended and modified to meet the requirements of the underlying data structure (DCAT-AP) of the data.europa.eu. The response of the 'package_search' action exposes a 'results' field, which holds a list of dictised datasets. The data structure of such a dataset differs from the one of a plain CKAN installation.

Please note:

- Fields marked with an asterisk (*) are CKAN standard. Further information can be found in the official documentation.
- Type specifications according to official JSON standard (http://json.org/).
- Besides the mandatory fields, the field names and types are not strict, but data suppliers must make sure an obvious mapping is possible.
- For a detailed explanation of each field, refer to the DCAT-AP specifications.

Just like metadata provided in DCAT-AP, CKAN metadata must be structured into datasets and distributions. For usage notes of the relevant DCAT-AP properties, please see the section on required metadata.

Dataset schema

The following fields are mandatory for datasets.

Field	Type	DCAT-AP dataset equivalent
Title *	string	dct:title
Notes *	string	dct:description
Publisher	object	dct:publisher
accessRights	object	dct:accessRights
Resources	object	dct:distribution

Distribution schema

The following fields are mandatory.

Field	Type	DCAT-AP distribution equivalent
url	string	dcat:accessURL
size	number	dcat:byteSize
Format	string	dct:format
rights	object	dct:rights

Example

A result of the 'package_search' action looks like this.

```
{
    "help":"http://example.eu/data/api/3/action/help_show?name=package_search",
    "success":true,
    "result":{
        "count":113948,
        "sort":"score desc, metadata_modified desc",
        "facets":{
        },
        "results":[
            {
                "issued":"2011-10-20T00:00:00Z",
                "id":"525abe30-ef60-4bf9-824e-916368c1fad8",
                "metadata_created":"2015-09-15T12:08:54.860742",
                "metadata_modified":"2015-09-15T13:17:51.405474",
                "temporal":[
                    {
                        "start_date":"2011-10-19T22:00:00Z",
                        "end_date":"2011-10-19T22:00:00Z"
                    }
                ],
                "state":"active",
                "type":"dataset",
                "resources":[
                    {
                        "package_id":"525abe30-ef60-4bf9-824e-916368c1fad8",
                        "id":"7166a1fa-d994-4d88-8e76-3378930b1e16",
                        "state":"active",
                        "format":"XHTML",
                        "mimetype":"application/xhtml+xml",
                        "name":"Example",
                        "created":"2015-09-15T14:39:43.865240",
                        "url":"http://example.com"
                    }
                ],
                "tags":[
                    {
```

```
                    "vocabulary_id":null,
                    "state":"active",
                    "display_name":"Example Tag",
                    "id":"06993102-a2ee-4e40-b9e4-ed3e4b86e943",
                    "name":"example-tag"
                }
            ],
            "groups":[
                {
                    "display_name":"Economy and finance",
                    "description":"",
                    "title":"Economy and finance",
                    "id":"128d0956-4526-440e-a951-f153c190d890",
                    "name":"economy-and-finance"
                }
            ],
            "creator_user_id":"0ab3c2ec-c2a2-4eef-b70f-ed093e028063",
            "publisher":{
                "resource":"http://example.com "
            },
            "organization":{
                "description":"Example Organization",
                "created":"2015-09-15T13:56:32.985936",
                "title":"Example Organization",
                "name":"example-orag",
                "is_organization":true,
                "state":"active",
                "image_url":"",
                "revision_id":"ea70fb1f-29a8-4e7b-8527-809e4792a75b",
                "type":"organization",
                "id":"0897b420-3c3d-4a19-9c2c-a9815e2db2be",
                "approval_status":"approved"
            },
            "name":"example-dataset",
            "notes":"Example",
            "owner_org":"0897b420-3c3d-4a19-9c2c-a9815e2db2be",
            "modified":"2011-10-20T00:00:00Z",
            "url":"",
            "title":"Example Dataset",
            "identifier":[
                "http://example-ident.com"
            ]
        }
    ],
    "search_facets":{
    }
  }
}
```

4.7.12 First steps for getting started

The ERPD on data.europa.eu harvests metadata about non-open-data published by public sector bodies of European Union Member States' NSIPs. To initiate the onboarding of your NSIP onto the ERPD on data.europa.eu, you will need to undertake two sequential steps: check that your NSIP is suitable for harvesting and issue a harvesting request via the data.europa.eu contact form.

4.7.13 Required information to prepare for harvesting

The very first step is to go through the following checklist to gather all the required information. The purpose of this checklist is to guide you in gathering and summarising the main requirements to enable the successful harvesting of a data supplier site and to assure a certain quality level of harvested datasets. Before contacting the ERPD on data.europa.eu please make sure that you can answer all listed questions. Of course, if anything is unclear, you can always reach out to us via the contact form.

Please remember that the preferred harvesting interface is OAI-PMH.

	Requirement	Value
1	Which country does your NSIP cover?	Free text
2	Is your NSIP already being harvested by another portal?	Free text
3	Does your National Single Information Point provide the metadata listed in section 4 of this documentation?	See section 4. Only metadata can be harvested, not the data itself.
4	What is the Uniform resource locator (URL) to your NSIPs interface / endpoint?	URL
5	If your NSIP is integrated in your existing open data infrastructure **AND** if your NSIP cannot be harvested via a dedicated endpoint: How can NSIP data be filtered from other (i.e. open) data that is part your infrastructure?	Free text. This is essential to correctly distinguish data provided under the DGA from other data.
6	What is the default language of the datasets from your National Single Information point?	Free text
7	Which metadata standard is supported by your NSIP?	DCAT-AP or CKAN (mapped to DCAT-AP)
8	Which representation of the metadata is used?	XML, JSON, or any RDF representation
9	Which type of API is used to retrieve the data?	OAI-PMH (Recommended) RDF dump file CKAN SPARQL endpoint
10	Is authentication required for you to access your API?	yes/no
11	Does your data include complete vocabulary for categorisation, or other fields that use a defined vocabulary (for example update frequency)?	Free text. Please provide vocabularies with translations, if available.
12	Does your data use standard date/time formats as specified by the ISO8601?	Yes/no. (Please note: Using the ISO8601 standard is mandatory.)
13	How often can/should the site be harvested?	E.g. daily, weekly, monthly. The default harvesting interval is weekly.
14	Are there any times when the site should not be harvested (e.g. scheduled maintenance)?	Free text
15	Who is the publisher of the NSIP (name and email address)?	Free text. e.g. Federal Open Data Agency; info@open-data.gv.example
16	What is the URL to the homepage of the NSIP?	URL

4.7.14 Harvesting request via contact form

Once you have gathered all answers to the checklist, the second step is to contact the ERPD to initiate the harvesting onboarding of your NSIP. Please submit a request via the contact form and please select 'Request to be harvested by data.europa.eu' as the issue type. In the contact request, please provide information on all questions listed in the checklist.

Once we receive your request, we will assess it and keep you informed on progress.

5. API documentation

5.1 General introduction to APIs

In digital form, a machine can exchange data. To provide a seamless interaction for data exchange, APIs are defined by developers of such a machine. Those APIs can guide humans and machines on how to interact and understand the exchanged data. For each API call, the requesting service is defined as the client and the service providing data exchange is defined as the server. In case of a human, the browser or a dedicated application for API calls would represent the humans as the client. Beyond those basic principles, it is important to understand the concepts of HTTP, URL, data representations formats and predefined communication methods.

HTTP

The 'Hypertext Transfer Protocol' (HTTP) allows to exchange hypertext between machines. Unlike standard text, hypertext can be non-linear and can contain links to other text. Nevertheless, hypertext can easily be read and understood by humans. HTTP is not only indispensable for browsing the Web but has also established as a standard for modern APIs. The payload via HTTP can be information of any kind, e.g. executable code or data. Beyond payload, HTTP contains meta information in their headers, e.g. information about the communication session. To provide a secured communication HTTP is combined 'Transport Layer Security' (TLS). Together they form HTTPS, a secure version of HTTP. HTTPS is crucial to provide a private conversion against external entities.

URL

A 'Uniform Resource Locator' (URL) is form of an address on the Web. In the context of APIs, a URL can address entry points, functions or resources provided by a server. In general, URLs are used to varying degrees depending on the API type. Some APIs use URLs only to address an entry point, others use path and query parameters to provide information to the server.

Data representation format

In addition to the communication protocol of the API and where the API is located, an API can define the representation of the data payload. For this, data may be represented as plain text, structured, and encoded data. In practice, data representation formats are used for structured data that are machine-readable to a certain degree, e.g. XML, JSON or YAML.

Predefined methods

When it comes to the management of resource, APIs typically provide create, read, update and delete operations (CRUD). For this, the HTTP protocol has predefined methods that can be used to define an API.

- POST, to create a resource
- GET, to read a resource
- PUT, to create or update a resource
- PATCH, to update a resource
- DELETE, to delete a resource

5.2 What is OpenAPI and why is it relevant?

All APIs on data.europa.eu are documented via OpenAPI. OpenAPI provides a human and machine-readable language agnostic interface description that enables everyone to use the APIs without having to investigate the source code. By using the specified syntax of OpenAPI, tools can be used to automatically create accessible and human readable documentation of our APIs. The screenshot below showcases a human readable documentation of the MQA API.

It explains how to retrieve global quality measurements from the MQA. It gives the user a short description of the context of the provided features and concrete explanations of all required and optional parameters that can be used when using the API. In the case at hand the parameter 'filter' can be used, and it is explained that the value of the parameter must be one of the following: 'findability', 'accessibility', 'interoperability', 'reusability', 'contextuality', 'score'

The right-hand side gives a concrete URL to request for testing the API. Please note that the example URL is not using any parameters. These must be added manually. Below the example URL an example response is given.

5.3 data.europa.eu APIs

data.europa.eu provides the following APIs to read our metadata:

- Search: https://data.europa.eu/api/hub/search/
- SPARQL: https://data.europa.eu/sparql
- Registry: https://data.europa.eu/api/hub/repo/
- Use cases: https://data.europa.eu/en/export-use-cases
- MQA: https://data.europa.eu/api/mqa/cache/
- SHACL metadata validation: https://data.europa.eu/api/mqa/shacl/

5.4 Manage datasets via API

You can manage the metadata in data.europa.eu via the Registry API. To illustrate the basic concepts and the access control methodology, this section guides you through the process of creating, updating and deleting a DCAT-AP dataset. You can apply the same flow to the other endpoints of the API. Please refer to the OpenAPI documentation. The API is mainly designed for programmatic use from your applications, and we recommend you use suitable third-party software libraries for the interaction:

- OpenID Connect is used for access control.
- The API follows the RESTful paradigm.

5.4.1 General

An overview on how to manage (create, update and delete) datasets in data.europa.eu via its API.

You will require at least write access to one catalogue that you are responsible for. Please contact the data.europa.eu team for further information.

The entire API for dataset management is documented with OpenAPI.

The following DCAT-AP dataset will be used as an example for this guide. It is serialised in Turtle. You can provide the dataset in any other common RDF format, such as RDF/XML, JSON-LD, N-Triples, Trig or N3.

```
@prefix dcat:   <http://www.w3.org/ns/dcat#> .
@prefix dct:    <http://purl.org/dc/terms/> .

<https://example.eu/set/data/test-dataset>
    a                         dcat:Dataset ;
    dct:title                 "DCAT-AP 2.1.0 Example Dataset"@en ;
    dct:description           "This is an example Dataset"@en ;
    dcat:theme                <http://publications.europa.eu/resource/authority/data-theme/TECH> ;
    dcat:distribution         <https://example.eu/set/distribution/1> .

<https://example.eu/set/distribution/1>
    a                         dcat:Distribution ;
    dct:format                <http://publications.europa.eu/resource/authority/file-type/CSV> ;
    dcat:accessURL            <https://github.com/ec-jrc/COVID-19/blob/master/data-by-country/jrc-covid-19-countries-
latest.csv> .
```

5.4.2 Access Control and Tokens

data.europa.eu uses Keycloak to implement OpenID Connect as backbone for access control. Hence, every interaction with a restricted API endpoint requires an interaction with Keycloak to obtain an access token (Party Token). It requires two API calls to get the token.

Prerequisites

- You require valid credentials (username and password) for the Keycloak.

- There is no self-registration. Please contact the Publications Office of the European Union data.europa.eu team for further information how to require credentials.

- You need a tool to interact with a HTTP API, such as Postman or curl lfor the command line.

Step 1 - Request User Token

First you will need to require a User Token by performing a x-www-form-urlencoded POST request to the following endpoint:

https://data.europa.eu/auth/realms/DEU/protocol/openid-connect/token

The following form values need to be set:

Key	Value
username	[yourusername]
password	[yourpassword]
grant_type	password
client_id	piveau-hub-ui

Example with curl:

```
$ curl --location --request POST "https://data.europa.eu/auth/realms/DEU/protocol/openid-connect/token" \
--header "Content-Type: application/x-www-form-urlencoded" \
--data-urlencode "grant_type=password" \
--data-urlencode "client_id=piveau-hub-ui" \
--data-urlencode "username=[yourusername]" \
--data-urlencode "password=[yourpassword]"
```

If successful, you will receive a JSON response like this:

```
{
    "access_token": "[yourusertoken]",
    "expires_in": 300,
    "refresh_expires_in": 1800,
    "refresh_token": "[yourrefreshtoken]",
    "token_type": "Bearer",
    "not-before-policy": 0,
    "session_state": "694350c7-38b9-4051-bc2e-e15e34320133",
    "scope": "email profile"
}
```

Step 2 Request Party Token

Now you will need to require a Party Token by again performing a x-www-form-urlencoded POST request to the following endpoint:

https://data.europa.eu/auth/realms/DEU/protocol/openid-connect/token

The following form values need to be set:

Key	Value
grant_type	urn:ietf:params:oauth:grant-type:uma-ticket
audience	piveau-hub-repo

In addition, place the User Token from Step 1 into the header field Authorization with the leading string Bearer.

Example with curl:

```
$ curl --location --request POST "https://data.europa.eu/auth/realms/DEU/protocol/openid-connect/token" \
--header "Content-Type: application/x-www-form-urlencoded" \
--header "Authorization: Bearer [yourusertoken]" \
--data-urlencode "grant_type=urn:ietf:params:oauth:grant-type:uma-ticket " \
--data-urlencode "audience=piveau-hub-repo "
```

If successful, you will get a JSON response like this:

```
{
    "upgraded": false,
    "access_token": "[yourpartytoken]",
    "expires_in": 300,
    "refresh_expires_in": 1800,
    "refresh_token": "[yourrefreshtoken]",
    "token_type": "Bearer",
    "not-before-policy": 0
}
```

[yourpartytoken] can now be used to manage datasets in data.europa.eu.

Token expiry

For security reasons the tokens expire quickly. You can refresh them by just performing Step 1 and 2 again.

5.5 Create or Update a Dataset

Both creating and updating a dataset is performed using the same endpoint:

https://data.europa.eu/api/hub/repo/catalogues/[catalogueId]/datasets/origin?originalId=[dataset_id]

The [dataset_id] can be freely chosen by you and the [catalog_id] determines the catalogue to which the dataset is added. The dataset ID is scoped within the catalogue. If the combination of dataset ID and catalogue ID already exists, the dataset is updated. Otherwise, a new dataset is created.

Pre-creation check

If you intend to create a new dataset, to avoid accidentally updating an existing dataset it is highly recommend to check that the dataset ID is not yet present within the catalogue. To do so perform a GET request for the dataset id.

https://data.europa.eu/api/hub/repo/catalogues/[catalogueId]/datasets/origin?originalId=[dataset_id]

Create/Update

To submit a create or update you perform a PUT request to the endpoint by providing the Party Token and the RDF format in the header. The actual dataset is provided in the body of the request.

Example with curl:

```
$ curl --location --request PUT " https://data.europa.eu/api/hub/repo/catalogues/test-catalog/datasets/origin?originalId=example-
dataset" \
--header "Content-Type: text/turtle" \
--header "Authorization: Bearer [yourpartytoken] \
--data-raw "@prefix dcat:    <http://www.w3.org/ns/dcat#> .
@prefix dct:    <http://purl.org/dc/terms/> .

<https://example.eu/set/data/test-dataset>
    a                              dcat:Dataset ;
    dct:title                      \"DCAT-AP 2.1.0 Example Dataset\"@en ;
    dct:description                \"This is an example Dataset\"@en ;
    dcat:theme                     <http://publications.europa.eu/resource/authority/data-theme/TECH> ;
    dcat:distribution              <https://example.eu/set/distribution/1> .

<https://example.eu/set/distribution/1>
    a                              dcat:Distribution ;
    dct:format                     <http://publications.europa.eu/resource/authority/file-type/CSV> ;
    dcat:accessURL                 <https://github.com/ec-jrc/COVID-19/blob/master/data-by-country/jrc-covid-19-countries-
latest.csv> ."
```

The following Content-Type values are valid:

Format	Value
RDF/XML	application/rdf+xml
Turtle	text/turtle
JSON-LD	application/ld+json
N3	text/n3
Trig	application/trig
N-Triples	application/n-triples

If the request was successful, you will receive a 201 response for a newly created dataset or a 204 for an updated dataset.

5.6 Delete a Dataset

You can delete a dataset by performing a DELETE to the same endpoint.

```
$ curl --location --request DELETE " https://data.europa.eu/api/hub/repo/catalogues/test-catalog/datasets/origin?originalId=example-
dataset " \
--header "Authorization: Bearer [yourpartytoken]
```

If the request was successful, you will receive a 204 response.

5.7 Querying metadata via data.europa.eu APIs

There are several APIs for querying the metadata of data.europa.eu programmatically. Depending on your needs and requirements you can choose a suitable API:

The SPARQL API gives you a all features of the underlying RDF data structure of data.europa.eu and allows you to create complex and specific queries. It may be limited in regard to full text search capabilities.

The Registry API offers a convenient and easy way to directly retrieve the RDF representation of the metadata.

The Search API offers a high-performance full-text search with filtering capabilities over the metadata. Further examples of the Search API usage in Java, Javascript, Python, and Ruby are in the Gitlab repository.

5.7.1 Example of the SPARQL API usage

A simple example to retrieve the first 100 datasets and their title with the SPARQL API would look like this:

```
PREFIX dcat: <http://www.w3.org/ns/dcat#>
PREFIX dct: <http://purl.org/dc/terms/>

SELECT  ?s ?title  WHERE {
    ?s a dcat:Dataset .
    ?s dct:title ?title
}
LIMIT 100
```

This query gives you a simple table with two columns, where the first column represents the dataset URL and the second one the title.

Data.europa has an introduction and relevant links for the SPARQL usage.

5.7.2 Example for the Registry API usage

The Registry API gives you direct access to RDF representations, hence DCAT-AP, of the metadata.

You can get a list of all datasets by executing the following query. It will give you a list with 50 entries, containing the URIs of the datasets/

```
GET https://data.europa.eu/api/hub/repo/datasets?limit=50&valueType=identifiers
```

Based on this result you can query the entire metadata of a specific dataset, by passing one of the IDs to the following call:

```
GET https://data.europa.eu/api/hub/repo/datasets/91f2aec3-1aaf-42d3-8730-c567a46c0116
```

By default, this will return the JSON-LD representation of the dataset, You can request a different serialisation format, by append a file-type suffix. The following example would give you the same dataset as Turtle:

```
GET https://data.europa.eu/api/hub/repo/datasets/91f2aec3-1aaf-42d3-8730-c567a46c0116.ttl
```

Please refer to the OpenAPI documentation for further details how to use the API.

5.7.3 Example of the SearchAPI usage

The simplest possible Search API usage is to just request all metadata without any further parameters that would reduce the search results. Looking at the documentation of the search API this can be achieved by just adding the path 'search'. Our first example API call looks like this:

```
GET https://data.europa.eu/api/hub/search/
```

It provides a limited list of all metadata that is stored in data.europa.eu in JSON This list is limited, because data.europa.eu stores too much metadata and this cannot be transmitted in one single request. If you would like to get all metadata available, multiple requests would be necessary. Here, a method called paging comes into play. With paging you can split the results in 'pages' and simply go through them. The examples below will showcase how to retrieve three pages of metadata via the search API:

```
GET https://data.europa.eu/api/hub/search/search?page=0&limit=10
```

```
GET https://data.europa.eu/api/hub/search/search?page=1&limit=10
```

```
GET https://data.europa.eu/api/hub/search/search?page=2&limit=10
```

Note, that these examples introduce the two parameters page and limit. Basically, the API call is asking for page 0, 1, 2 of the search results with maximum 10 results per page. The page parameter selects the requested page and the limit parameter sets the maximum number of results to be retrieved per page.

For a last example we will add a search term to the request. The following request shall return page 0 with a maximum of 100 results that match with the search term 'water':

```
GET https://data.europa.eu/api/hub/search/search?page=0&limit=100&q=water
```

Please consult the OpenAPI documentation for more information on how to use the APIs.

6. Publications and education

Several publications, ranging from the yearly Open Data Maturity report to specific data stories and open data news are available through data.europa.eu.

The portal offers social media coverage on open data topics, with the purpose of educating and informing citizens about the opportunities that using and reusing public data resources can bring. It also hosts the data.europa academy and publishes a monthly newsletter.

The Use cases section offers a catalogue that includes hundreds of use cases from both publishers and people who reuse open data.

The portal is linked to other activities that demonstrate the potential of open data, such as the EU Datathon (an annual open data competition), the EU Open Data Days (a worldwide conference about open data and data visualisation) and different online seminar series (EU DataViz webinars, EU Open Data Explained webinars and data.europa academy webinars).

The portal also offers documentation on general recommendations for topics like data quality and data citation.

Finally, the portal researches and documents the impact of open data in reports such as the 2020 report on The Economic Impact of Open Data – Opportunities for value creation in Europe. The study forecasts the open data market size and employment growth for 2025. The impact is exemplified by efficiency gains and cost savings due to open data.

6.1 Open data maturity

The study on open data maturity serves as benchmark to gain insights into the development of European countries in the field of open data. This annual assessment, started in 2015, presents the overview to better understand the countries level of maturity on open data. The study captures their progress over time and areas of improvement and benchmarks them against other countries. The report assesses the maturity against four dimensions:

1. Policy
2. Portal
3. Impact
4. Quality

It also provides an overview of best practices implemented across Europe and a set of recommendations tailored to the level of maturity and characteristics of each group. Every year, the data is collected through a questionnaire sent to national open data representatives working in collaboration with the European Commission and the Public Sector Information Expert Group.

6.2 Data.europa academy

The data.europa academy is your knowledge hub on data. It helps you to become more knowledgeable about open data and become more data literate, with different courses hosting e-learnings, reports, webinars and

related material. You will find an extensive learning offer depending on your profile (academia, civil servant, data provider, developer, journalist, NGO, private sector) or your level of expertise on data. They are also grouped per theme (business, communication, impact, legal, policy, portal, and quality).

Whether you are an expert or just getting started, you will find the information you need to upskill your data learnings!

6.3 News and data stories

Under news, you can find short articles about the latest activities, events, projects, initiatives, studies, trends and use cases related directly or indirectly to open data in Europe and beyond. If you would like your event, project, or initiative to be featured in our news section, let us know via the contact form.

6.4 Data stories

Data stories are a set of articles on a variety of topics related to the European Commission's priorities, showcasing relevant open data through data visualisations. The section also includes interviews, highlighting examples of good practices and innovative techniques related to open data across the world.

6.5 Studies

The portal commits to open data research and learning as key milestones to keep updating in this constantly changing data ecosystem. This section provides a full collection of the research developed on several topics.

6.6 Calendar

The events calendar features virtual and in-person events related directly and indirectly to open data in Europe and beyond. If you would like your event to be featured in this section of the portal, let us know via the contact form.

6.7 EU Open Data Days

The EU Open Data Days in 2021, was the first-ever event of its kind and took place from 23 to 25 November 2021. It comprised EU DataViz, an international conference on open data and data visualisation, followed by the finals of EU Datathon, the annual open data competition.

The EU Open Data Days were designed to be relevant to all open data stakeholders and data users, with a special emphasis on the needs of the EU's public sector. The event attracted well over 2 000 registrations from data enthusiasts, data visualisation experts and solution-seekers. On the stage, different sectors of the society were represented, with more than 70 speakers coming from all around the world to bring their expertise for public administrations.

The EU Open Data Days were organised by the Publications Office of the EU, with the active support of over 60 partners, representing the data providers from EU institutions and agencies, European national open data portals and national statistical offices, the key digital players in Luxembourg.

You can access the recordings and presentations online.

6.8 EU Datathon

EU Datathon is an annual open data competition organised by the Publications Office of the EU since 2017. The competitions are organised to create new value for citizens through innovation and promote the use of open data available on the portal.

EU Datathons gather teams of people from all over Europe to stimulate innovation, creating a bridge between citizens and the EU's administration. As a result, new applications are developed to improve existing services or create completely new ones. The competitions have also demonstrated that potential users do need these initiatives for visibility purposes before potential business partners.

Open data demonstrated in every edition how powerful it could be, in every sector of society. It showed us how it could help in the creation of start-ups, innovative products and services; it promotes transparency in the public sector, addresses current societal challenges, boosts economic growth and accelerates scientific progress.

Every EU Datathon competition showed innovation when teams tackled different challenges, such as helping to achieve the digital single market (e.g. Tenderlake and C4P.io), making legislation more interoperable (e.g. The Smartfiles Network and Lexparency), offering innovative technology relating to medicines (e.g. Medicatio and Open Food Facts) or tackling climate change (ODCCI) through a platform which optimises irrigation to reduce water waste (e.g. Chloe irrigation system).

These examples, among others, uncover how accessible and reusable data can help to build a better future.

EU Datathons are organised in collaboration with partners in EU institutions and Member States to address domains of primary importance for citizens, the public sector and businesses: legislation, medications, food safety and public procurement, among others. This interinstitutional collaboration highlights the common goals to be closer to citizens and support innovative start-ups. Additionally, by facilitating cooperation between public entities and the private sector, EU Datathons generate innovative ideas for open government, offering a platform for all sides to exchange information and learn.

Moreover, it is a new, more direct way of interaction between citizens, innovative youth and the EU administration and it can generate a positive atmosphere around the Commission and the EU institutions in general.

6.9 Newsletter

The portal newsletter is sent to subscribers monthly. It provides an overview of the recent and upcoming events and activities of the portal, along with the most important news, events, and developments in the open data domain in Europe.

Subscribe to the newsletter. To unsubscribe, click on the 'Unsubscribe from this newsletter' link at the bottom of the newsletter email.

7. Metadata quality

7.1 Metadata good practices

This section explains concrete examples and good practices to create expressive and high-quality metadata. Metadata guides users to relevant datasets while also providing important context. Without high-quality metadata, even the most comprehensive dataset will be difficult to find and to be identified as relevant. Ensuring that the metadata follows quality standards is essential.

Data and metadata quality training on the data.europa academy

7.1.1 Follow DCAT-AP

In general, we strongly recommend that you make yourself familiar with the DCAT-AP specification and follow it as close as possible. As a minimal requirement, you need to set all mandatory properties. Additionally, we strongly recommend providing all recommended properties as well. The correct application ensures that datasets are described in a standardized way, promoting consistency and interoperability across Open Data repositories as well as for the data in data.europa.eu.

The latest version and documentation of DCAT-AP is always accessible here: https://joinup.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe

A minimal DCAT-AP example in the Turtle format with all mandatory properties:

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .

<https://opendatarepo.org/mydataset>
a dcat:Dataset ;
dct:description "This is an example Dataset"@en ;
dct:title "Example Dataset"@en ;
dcat:distribution <https://opendatarepo.org/mydistribution> .
<https://opendatarepo.org/mydistribution>
a dcat:Distribution ;
dcat:accessURL <http://myaccessurl.org> .
```

7.1.2 Only use controlled vocabularies

DCAT-AP specifies a set of controlled vocabularies that must be used for certain properties, e.g. for media types, data formats, languages, and themes. This ensures interoperability, the correct rendering in the user interface, and guarantees that datasets are uniformly comprehensible. In addition, it supports the creation of precise queries for datasets. Please refer to the DCAT-AP documentation for details. The Publications Office maintains an authority table of EU controlled vocabularies.

The following statement would link the dataset to the data theme 'Agriculture, fisheries, forestry and food'.

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .

<https://opendatarepo.org/mydataset>
a dcat:Dataset ;
dcat:theme <http://publications.europa.eu/resource/authority/data-theme/AGRI> .
```

7.1.3 Apply the correct Data Ranges

For its properties, DCAT-AP indicates concrete data ranges that must be used. This ensures that any provided values are consistent with the expected format. Using the correct data ranges or RDF data types is essential for data integrity, validation, and UI representation. Furthermore, it supports more accurate querying, processing, and interpretation of datasets. Examples for such ranges are rdfs:Literal, xsd:dateTime, and foaf:Agent. Most of these ranges are basic RDF types, with many public resources explaining how these should be applied correctly. Some ranges are further elaborated in the DCAT-AP standard, such as foaf:Agent, where guidelines for sub-properties are provided.

As an example, we add a publisher and an issue date to a dataset:

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

<https://opendatarepo.org/mydataset>
a dcat:Dataset ;
dct:issued "2015-08-28T00:00:00"^^xsd:dateTime ;
dct:publisher [
a foaf:Person ;
foaf:name "John Doe" ;
] .
```

7.1.4 Make your Dataset Identifiable

It is paramount to make your dataset uniquely identifiable within your data catalogue and if possible, even with a globally unique identifier. Assigning distinct identifiers ensures datasets remain distinguishable and facilitates efficient harvesting, updating, referencing, and preventing potential overlaps or duplications. For the identification within a data catalogue, the property dct:identifier must be used and assigned a literal that is unique within your catalogue. For global persistent identifiers, the property adms:identifier must be used allowing the specific type of the identifier (e.g. a DOI). More information about the concepts of persistent identifiers is available.

Here is an example for a dct:identifier and adms:identifier:

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .
@prefix dct: <http://purl.org/dc/terms/> .
@prefix adms: <http://www.w3.org/ns/adms#> .

<https://opendatarepo.org/mydataset>
a dcat:Dataset ;
dct:identifier "b17be550-40b2-11ee-be56-0242ac120002" ;
adms:identifier [
  a adms:Identifier ;
  skos:notation "http://www.doi.org/123456789"^^<http://purl.org/spar/datacite/doi> ;
] .
```

7.1.5 General Remarks

Besides the concrete example, it is good practice to follow some general guidelines for high-quality metadata:

- Within your organisation, push for a uniform application of individual metadata standards across all your datasets. Such consistency is the basis for a systematic structure and presentation of metadata, that promoting familiarity and trust. Organizations that maintain such uniformity emphasize their commitment to quality and reliability.

- Comprehensive descriptions are paramount for open data. Therefore, make use of the property dct:description, since it contains the context of the dataset, methods employed, and content. By offering a thorough overview, data providers ensure users approach the data with an informed perspective, enhancing the accuracy of its reuse.

- Metadata, like all information repositories, evolve over time. Therefore, it is important to ensure that metadata accurately reflects any changes to the datasets it represents. Periodic reviews and consequent updates ensure that metadata remains relevant and accurate.

- The world of data standards is not static. Stay informed about any updates or changes to the DCAT-AP standard to ensure that your metadata remains compliant and to enable you to leverage new features.

7.2 Metadata quality dashboard

The Metadata Quality Assessment dashboard (MQA) is a tool developed by data.europa.eu to check the quality of metadata of published datasets on the portal and provide recommendations to data providers and data portals for improvements. The results are presented via the MQA and are also available to download. The functionality of the MQA and the methodology it uses are detailed in the data.europa.eu MQA service Metadata quality methodology.

7.3 Shacl validation

data.europa.eu offers a stand-alone user-friendly DCAT-AP SHACL validator and a RESTful API that can also be used. The API is described via OpenAPI under the mentioned URL. If you don't know how to call a RESTful API, you can use a tool like Postman and follow the following step-by-step instructions.
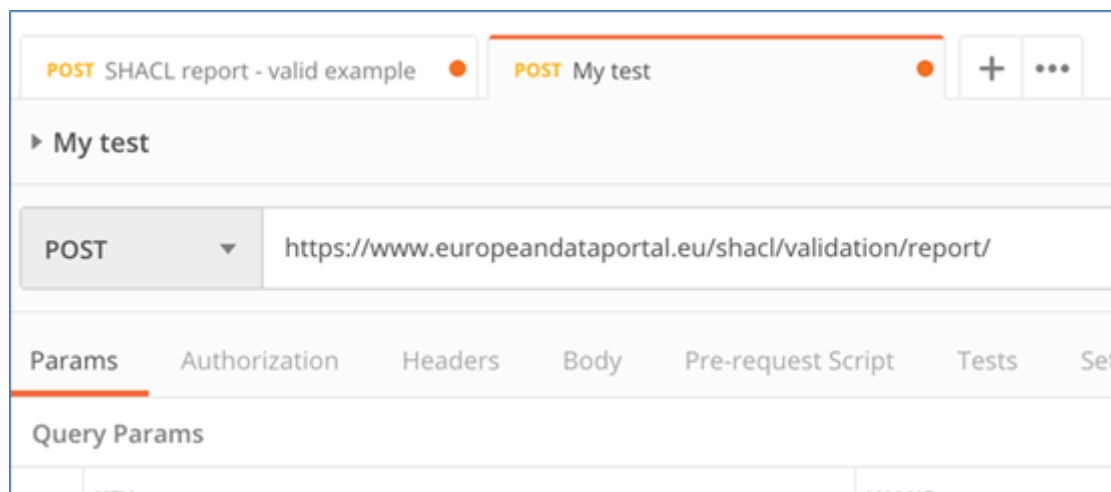
1. Create a new 'request'.
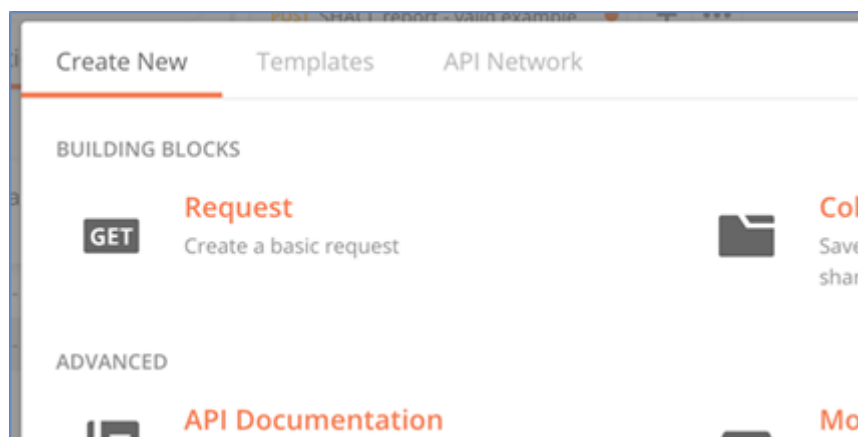
2. Give the request a name and put it in any collection that is convenient for you.
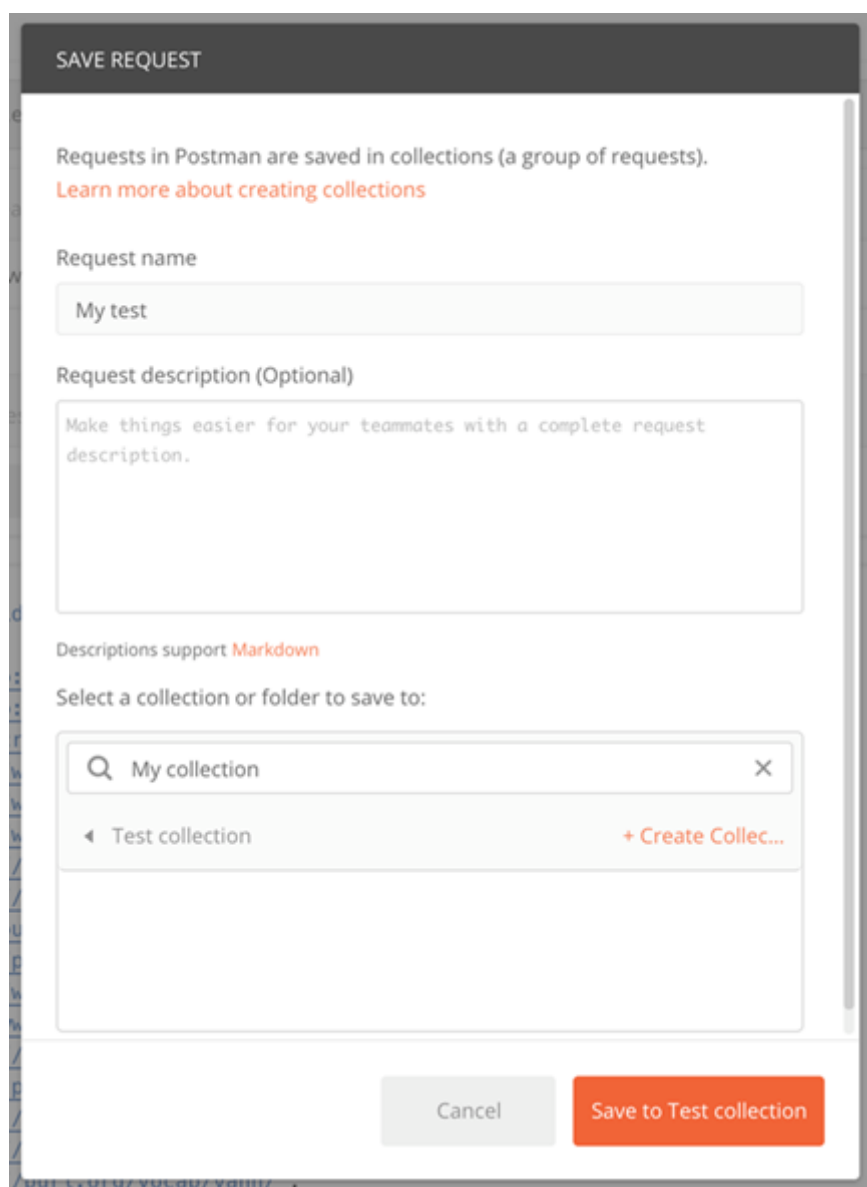
3. Choose to use HTTP POST and write https://data.europa.eu/api/mqa/shacl/validation/report/ as the request URL.

4. Specify the 'Content-Type' variables in the HTTP headers according to the type of DCAT-AP representation you will use, typically 'application/rdf+xml'.
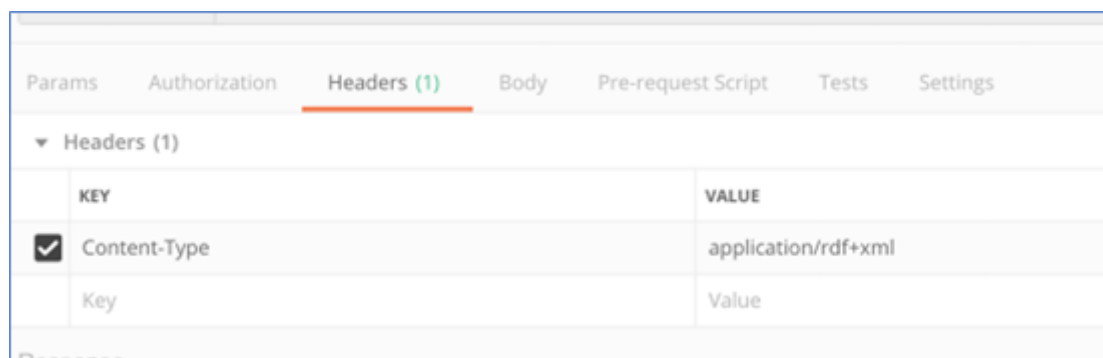
5. Copy and paste your DCAT-AP in the 'body' you are going to send, ensuring to specify that you are using 'raw' input.

6. Click on the 'Send' button.

7. Examine the results in the bottom pane. If your DCAT-AP is valid, you will get an empty report in JSON-LD that looks like this.
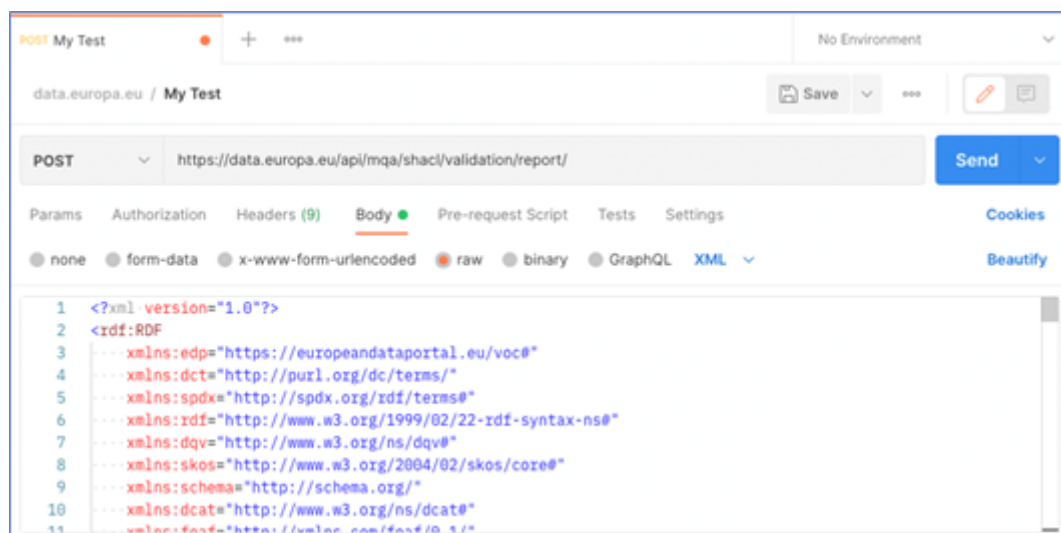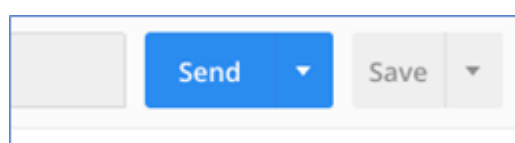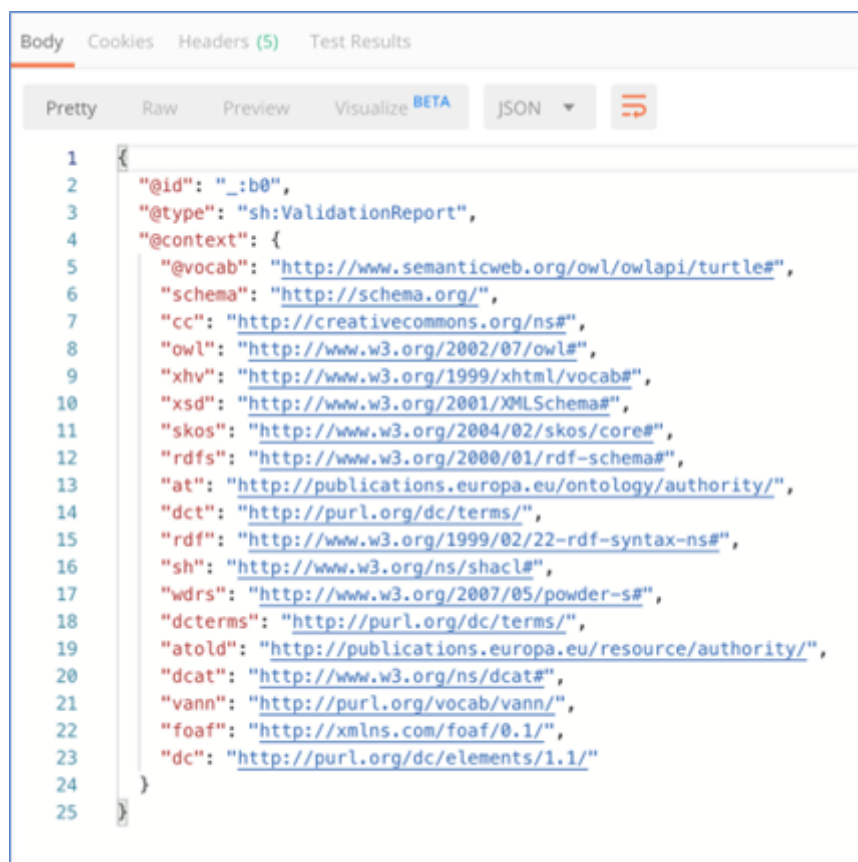
```
Body  Cookies  Headers (5)  Test Results

Pretty   Raw    Preview    Visualize BETA   JSON  ▼   ⇥

  1   {
  2      "@id": "_:b0",
  3      "@type": "sh:ValidationReport",
  4      "@context": {
  5        "@vocab": "http://www.semanticweb.org/owl/owlapi/turtle#",
  6        "schema": "http://schema.org/",
  7        "cc": "http://creativecommons.org/ns#",
  8        "owl": "http://www.w3.org/2002/07/owl#",
  9        "xhv": "http://www.w3.org/1999/xhtml/vocab#",
 10        "xsd": "http://www.w3.org/2001/XMLSchema#",
 11        "skos": "http://www.w3.org/2004/02/skos/core#",
 12        "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
 13        "at": "http://publications.europa.eu/ontology/authority/",
 14        "dct": "http://purl.org/dc/terms/",
 15        "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
 16        "sh": "http://www.w3.org/ns/shacl#",
 17        "wdrs": "http://www.w3.org/2007/05/powder-s#",
 18        "dcterms": "http://purl.org/dc/terms/",
 19        "atold": "http://publications.europa.eu/resource/authority/",
 20        "dcat": "http://www.w3.org/ns/dcat#",
 21        "vann": "http://purl.org/vocab/vann/",
 22        "foaf": "http://xmlns.com/foaf/0.1/",
 23        "dc": "http://purl.org/dc/elements/1.1/"
 24      }
 25   }
```

Otherwise, if there are mistakes, you will see a list of them. Every error has an entry in the JSON-LD file that looks like this.

```json
{
  "@graph": [
    {
      "@id": "_:b0",
      "@type": "sh:ValidationResult",
      "focusNode": "http://data.europa.eu/88u/dataset/ded24b58-a5ab-4d34-8603-2e5b2131a6a2",
      "resultMessage": "ClassConstraint[<http://purl.org/dc/terms/RightsStatement>]: Expected class
        :<http://purl.org/dc/terms/RightsStatement> for <http://publications.europa.eu/resource/
        authority/access-right/PUBLIC>",
      "resultPath": "dct:accessRights",
      "resultSeverity": "sh:Violation",
      "sourceConstraintComponent": "sh:ClassConstraintComponent",
      "sourceSha  Follow link (cmd + click)  h:ClassConstraintComponent",
      "value": "http://publications.europa.eu/resource/authority/access-right/PUBLIC"
    },
    {
      "@id": "_:b10",
      "@type": "sh:ValidationResult",
      "focusNode": "http://data.europa.eu/88u/dataset/ded24b58-a5ab-4d34-8603-2e5b2131a6a2",
      "resultMessage": "Node[<http://data.europa.eu/r5r#DataThemeRestriction>] at focusNode <http://
        publications.europa.eu/resource/authority/data-theme/HEAL>",
      "resultPath": "dcat:theme",
      "resultSeverity": "sh:Warning",
      "sourceConstraintComponent": "sh:NodeConstraintComponent",
      "sourceShape": "_:b11",
      "value": "http://publications.europa.eu/resource/authority/data-theme/HEAL"
    },
    {
      "@id": "_:b12",
      "@type": "sh:ValidationResult",
      "focusNode": "_:b14",
      "resultMessage": "HasValueConstraint[http://spdx.org/rdf/terms#checksumAlgorithm_sha1] : No
        value <http://spdx.org/rdf/terms#checksumAlgorithm_sha1> in [http://spdx.org/rdf/
        terms#checksumAlgorithm_md5]",
      "resultPath": "spdx:algorithm",
      "resultSeverity": "sh:Violation",
      "sourceConstraintComponent": "sh:HasValueConstraintComponent",
      "sourceShape": "_:b13"
    },
```

Finally, amend the DCAT-AP to address the errors and go back to step 5 until all are solved.

8. Data quality

Our team is aware that the demand for high-quality data is still growing, with a focus on data that is publicly available and can be easily reused for different purposes. Poor quality of data is a major barrier to data reuse. Some data cannot be interpreted due to ill-defined, inaccurate elements such as missing values, mismatches, missing data types, lack of documentation about the structure or format availability (HTML, GIF or PDF). Users find poor-quality data harder to understand and may use it less often. The data provider may even appear less reliable as a result.

For those reasons, our team is involved in different initiatives regarding data quality. One of them was the publication of Data.europa.eu Data Quality Guidelines. This publication contains a set of recommendations for delivering high-quality data. They are addressed to data providers to support them in preparing their data, developing their data strategies and ensuring data quality.
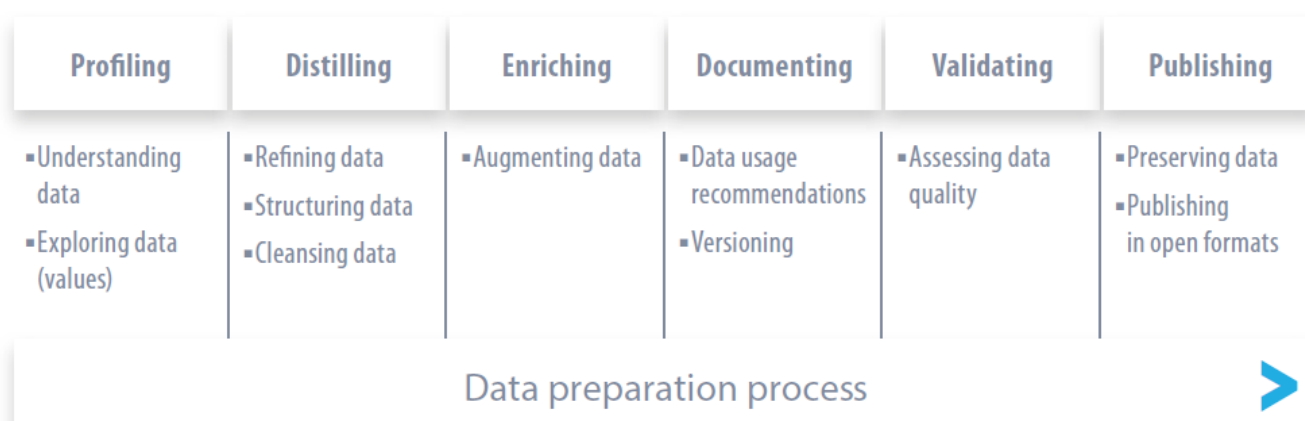
The document is composed of the following four parts.

1. **Recommendations for providing high-quality data.** The recommendations cover general aspects of quality issues regarding the findability, accessibility, interoperability and reusability of data (including specific recommendations for common file formats like CSV, JSON, RDF and XML).
2. **Recommendations for data standardisation (with EU controlled vocabularies) and data enrichment.**
3. **Recommendations for documenting data.**
4. **Recommendations for improving the 'openness level'.**

In the following subsections you will find tips and quick-reference material for providing high-quality data, standardisation and data enrichment, documenting data and improving the 'openness level'.

8.1 Recommendations for providing high-quality data

Data needs to be carefully prepared before publication. Preparation is an interactive and agile process used to explore, combine, clean and transform raw data into curated, high-quality datasets. This process consists of six different phases, illustrated in the image below.

Profiling	Distilling	Enriching	Documenting	Validating	Publishing
▪Understanding data ▪Exploring data (values)	▪Refining data ▪Structuring data ▪Cleansing data	▪Augmenting data	▪Data usage recommendations ▪Versioning	▪Assessing data quality	▪Preserving data ▪Publishing in open formats

Data preparation process

Source: Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

General tip: make use of tooling and create a data management plan.

8.2 Best practices for providing high-quality data (findability, accessibility, interoperability, reusability)

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

Best practices for providing high-quality data – general recommendations

Practice	General recommendations
Findability (red)	• Describe your data with metadata to improve data discovery • Mark null values explicitly as such
Accessibility (green)	• Publish data without restrictions • Provide an accessible download URL
Interoperability (blue)	• Formatting of date and time • Formatting of decimal numbers and numbers in the thousands • Make use of standardised character encoding • Use uniform resource identifiers (URIs) to identify entities
Reusability (yellow)	• Provide an appropriate amount of data • Consider community standards • Remove duplicates from your data • Increase the accuracy of your data • Provide information on byte size

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.3 Format specific recommendations

8.3.1 CSV

• Use a semicolon as a delimiter
• Use one file per table
• Avoid white space and additional information in the file
• Insert column headers
• Ensure that all rows have the same number of columns
• Indicate units in an easily processable way

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.3.2 XML

• Provide an XML declaration
• Escape special characters
• Use meaningful names for identifiers
• Use attributes and elements correctly
• Remove program-specific data

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.3.3 RDF

- Use HTTP URIs to denote resources
- Use namespaces when possible
- Use existing vocabularies when possible

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.3.4 JSON

- Use suitable data types
- Use hierarchies for grouping data
- Only use arrays when required

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.3.5 APIs

- Use correct status codes
- Set correct headers
- Use paging for large amounts of data
- Document the API

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.4 Recommendations for standardisation and data enrichment

- Reuse unambiguous concepts from controlled vocabularies (Findability)
- Harmonise labels (using unique identifiers) (Accessibility)
- Dereference the translation of a label (Reusability)
- Link and augment your data (Interoperability)

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.5 Recommendations for documenting data

- Publish your documentation (Findability)
- Use schemas to specify data structure (Accessibility)

- Document the semantics of data (Reusability)

- Document data changes (Interoperability)

- Deprecate old versions (Findability)

- Link versions of a data set (Accessibility)

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.6 Recommendations for improving the 'openness level'

- Use structured data (Findability)

- Use a non-proprietary format (Accessibility)

- Use URIs to denote things (Reusability)

- Use linked data (Interoperability)

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

File formats and their achievable openness level:

Format	Non-proprietary	Machine readable	Achievable stars
RDF	Yes	Yes	✪✪✪✪
XML	Yes	Yes	✪✪✪
JSON	Yes	Yes	✪✪✪
CSV	Yes	Yes	✪✪✪
ODS	Yes	Predominantly	✪✪✪
XLSX	Yes	Predominantly	✪✪✪
XLS	No	Predominantly	✪✪
TXT	Yes	Predominantly	✪*
HTML	Yes	Predominantly	✪*
PDF	Yes	No	✪
DOCX	Yes	No	✪
ODT	Yes	No	✪
PNG	Yes	No	✪
GIF	No	No	✪
JPG/JPEG	No	No	✪
TIFF	No	No	✪
DOC	No	No	✪

* Strictly according to the five-star model, this format would have to be rated with three stars, since the data may well be designed to be machine readable. However, we only give one star because this format was not originally intended to represent machine-readable but human-readable content. Representing machine-readable content in this format does not meet best practice and is therefore not recommended by the authors.

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

8.7 Checklist for publishing high-quality data

Make your data FAIR(¹)

Define data management plan

Before collecting your data

- Data characteristics: structure, modelling
- Data documentation
- Data quality assurance
- Data sharing
- Data storage and dissemination

Assess data quality

Reusability, accessiblity

- Validate and clean your data
  - > Remove duplicates
  - > Mark null values
- Standardise your data
  - > Format date, time, numbers
  - > Check character encoding

Document your data

Reusability

- Document schemas and data models
- Document data changes and versions
- Document semantics of data
- Document data validation report
- Document validation rules

Enrich your data

Interoperability

- Reuse concepts from controlled vocabularies
- Harmonise labels
- Dereference translation of labels
- Link and augment your data

Publish your data

Reusability, accessiblity

- Offer direct access (accessible download URL)
- Use machine-readable open formats
- Publish under an open licence
- Publish documentation about your data
- Use URIs and linked data

Facilitate data discovery

Findability

- Describe your data with rich metadata

(¹) https://www.go-fair.org/fair-principles/

Source: based on Publications Office, Data.europa.eu Data Quality Guidelines, 2022, https://data.europa.eu/doi/10.2830/333095

9. Data citation

Every day millions of people use data. They search, publish, reuse and analyse. But very often, a major difficulty for the users is finding the right data. Citing data can help to mitigate this.

In this section you will find information about our initiatives on data citation.

9.1 Data citation: A guide to best practice

Data citation: A guide to best practice is a study and collection of best practices and recommendations in the field of data citation. It was published by the Publications Office of the European Union (OP) in 2022. Below you will find the annexes that contain useful quick reference materials including checklists, diagrams and examples. If you are interested in this topic, you can find a dedicated data citation training in the data.europa academy where the author, Paul Jessop, presents the study in an online seminar.

9.1.1 Checklist (long) for footnotes or reference-list entries

The following steps should be a checklist when creating a citation that will appear in a footnote or reference-list entry.

The key initial step is to refer to the metadata of the dataset that has been used. This may include a 'cite this dataset as' function. This will provide most, or all of the information needed, though it may appear in a slightly different format and some editing work may be needed.

Otherwise, the metadata should include all the information needed, though sometimes a little research will be needed to locate some of it.

| Metadata	Description
author(s)	Who does the cited dataset note as its creator or as a contributor to it? Is there a corporate author as well as the individual authors? How many are there and in what order do they appear in the metadata of the original dataset? Decide how many to include if there are more than three – this guide recommends a maximum of three (followed by 'et al.') but allows the inclusion of fewer or more in some circumstances. Check whether the authors have been presented with their family name first (especially important with names from cultures that are not familiar to you).
'title' or title (short code)	What is the accurate title (including capitalisation and punctuation) of the original dataset? Is it necessary to capitalise the title to meet ISG requirements (when the title is presented in italics)? Is there a short code by which the dataset is known, which can be placed in parentheses after the title?
version	What is the version of the dataset that was used in the activity for which it is being cited? Do not just pick up the most recent version unless this is what was used and is appropriate in the citation.
publisher	Who made the dataset available? They might already be included as an author, in which case leave them out here. Do not just assume that a repository is the publisher – there may be an actual publisher in the metadata.
date	When does the metadata say the data was published? For datasets that change, the date of initial publication may not be useful. However, sometimes it helps to distinguish similar but slightly different datasets. Is that the case here? Write the date (and time if needed) in datatypes indicated by DCAT-AP (e.g., 2021-09-13 or 2021-09-13T23:16) or in the Interinstitutional Style Guide format as appropriate.
date of citation 'accessed'	If the data may have changed since it was accessed, or may change in the future, at least note the date it was accessed.
PID	There should be a persistent identifier in the dataset metadata. This will be managed so that people can still use it to find and access the data in the future. Try to locate it and, if possible, write it as a web link so that it can be clicked on to access a landing page. Check that the web link works and gives the expected result before including it in a citation – a link that is already broken is of little use to readers.

Source: Publications Office, Jessop, P., Data citation: A guide to best practice, 2022, https://data.europa.eu/doi/10.2830/59387

9.1.2 Checklist (short) for footnotes or reference-list entries

☐ **author(s)**

☐ **title** (short code)

☐ **version**

☐ **publisher**

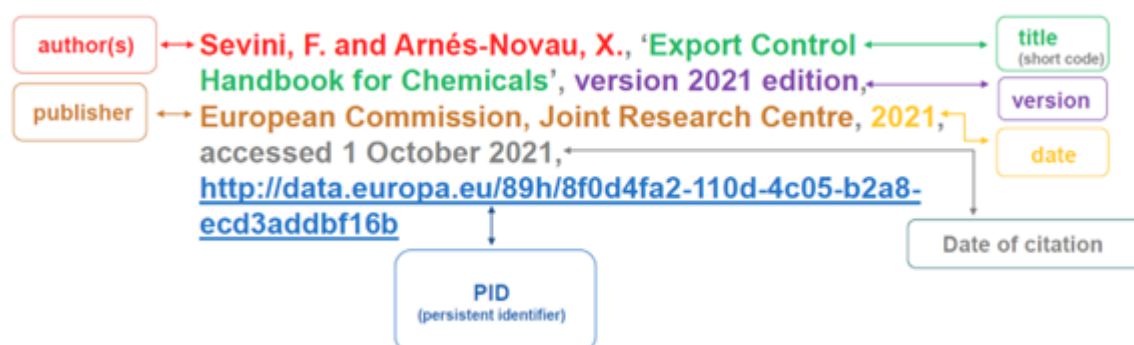☐ **date**

☐ date of citation

☐ **PID (persistent identifier)**

Source: Publications Office, Jessop, P., Data citation: A guide to best practice, 2022, https://data.europa.eu/doi/10.2830/59387

9.1.3 Diagram for footnote and reference citation style with date format compliant with DCAT-AP

Source: Publications Office, Jessop, P., Data citation: A guide to best practice, 2022, https://data.europa.eu/doi/10.2830/59387
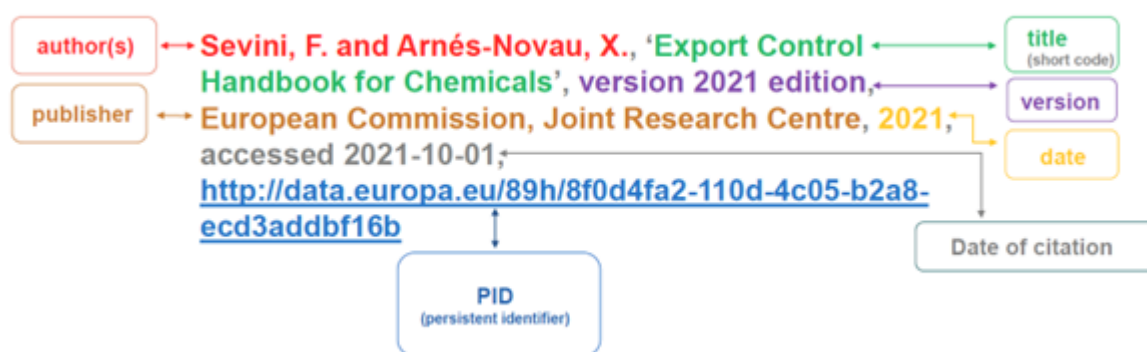
9.1.4 Diagram for footnote and reference citation style with date format compliant with the Interinstitutional Style Guide

Source: Publications Office, Jessop, P., Data citation: A guide to best practice, 2022, https://data.europa.eu/doi/10.2830/59387

9.1.5 Examples of citations with date format compliant with DCAT-AP

European Commission, Directorate-General for Financial Stability, Financial Services and Capital Markets Union, 'Consolidated list of persons, groups and entities subject to EU financial sanctions', 2016 (updated 2020-12-23), accessed 2023-12-22, http://data.europa.eu/88u/dataset/consolidated-list-of-persons-groups-and-entities-subject-to-eu-financial-sanctions

European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, 'Advanced Technologies for Industry - ATI', 2020 (updated 2021-01-31), http://data.europa.eu/88u/dataset/advanced-technologies-for-industry-ati

Sevini, F. and Arnés-Novau, X., 'Export Control Handbook for Chemicals,' version 2021 edition, European Commission, Joint Research Centre (JRC), 2021, accessed 2021-10-01, http://data.europa.eu/89h/8f0d4fa2-110d-4c05-b2a8-ecd3addbf16b

Nijs, W. and Ruiz, P., '01_JRC-EU-TIMES Full model', European Commission, Joint Research Centre, 2019, http://data.europa.eu/89h/8141a398-41a8-42fa-81a4-5b825a51761b European Commission, Eurostat, 'Airport traffic data by reporting airport and airlines' (avia_tf_apal), most recent data 2021-09-01, https://ec.europa.eu/eurostat/databrowser/view/avia_tf_apal/default/table?lang=en

European Commission, Eurostat, 'Real GDP growth rate – volume' (tec00115), updated 2021-09-28, https://ec.europa.eu/eurostat/databrowser/view/tec00115/default/table?lang=en Publications Office of the European Union, 'Country Named Authority List,' 2009 (updated 2021-09-29), http://data.europa.eu/88u/dataset/country

Source: Publications Office, Jessop, P., Data citation: A guide to best practice, 2022, https://data.europa.eu/doi/10.2830/59387

9.1.6 Examples of citations with date format compliant with the Interinstitutional Style Guide

European Commission, Directorate-General for Financial Stability, Financial Services and Capital Markets Union, 'Consolidated list of persons, groups, and entities subject to EU financial sanctions,' 2016 (updated 23

October 2020), accessed 1 October 2021, http://data.europa.eu/88u/dataset/consolidated-list-of-persons-groups-and-entities-subject-to-eu-financial-sanctions-fisma

European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs, 'Tenders Electronic Daily (TED) (csv subset) – public procurement notices,' version 3.3, 2015 (updated 22 November 2020), http://data.europa.eu/88u/dataset/ted-csv

Sevini, F. and Arnés-Novau, X., 'Export control handbook for chemicals,' version 2021 edition, European Commission, Joint Research Centre (JRC), 2021, accessed 1 October 2021, http://data.europa.eu/89h/8f0d4fa2-110d-4c05-b2a8-ecd3addbf16b

Nijs, W. and Ruiz, P., '01_JRC-EU-TIMES full model', European Commission, Joint Research Centre, 2019, http://data.europa.eu/89h/8141a398-41a8-42fa-81a4-5b825a51761b

European Commission, Eurostat, 'Airport traffic data by reporting airport and airlines' (avia_tf_apal), most recent data 1 September 2021, https://ec.europa.eu/eurostat/databrowser/view/avia_tf_apal/default/table?lang=en

European Commission, Eurostat, 'Real GDP growth rate – volume' (tec00115), updated 18 September 2021, https://ec.europa.eu/eurostat/databrowser/view/tec00115/default/table?lang=en

Publications Office of the European Union, 'Country named authority list,' 2009 (updated 29 September 2021), http://data.europa.eu/88u/dataset/country

Source: Publications Office, Jessop, P., Data citation: A guide to best practice, 2022, https://data.europa.eu/doi/10.2830/59387

9.2 Citation button

The citation button will help you retrieve citations for datasets without much effort and will guarantee their correctness.
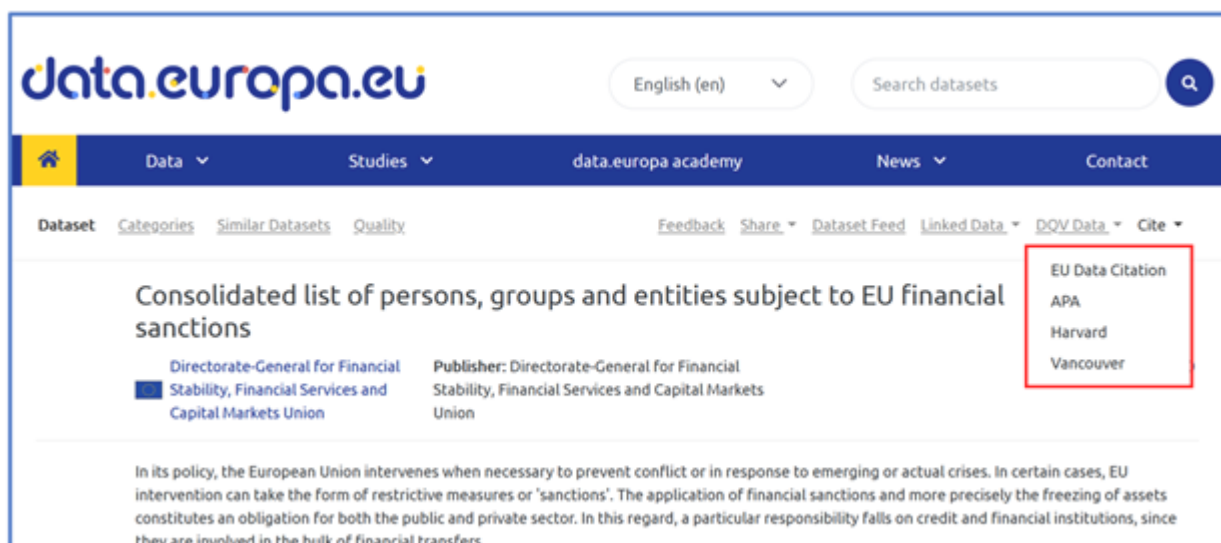
Where can you find it?

View the web page of the dataset you want to cite. You will find the button on the dataset menu on the right above the dataset title.
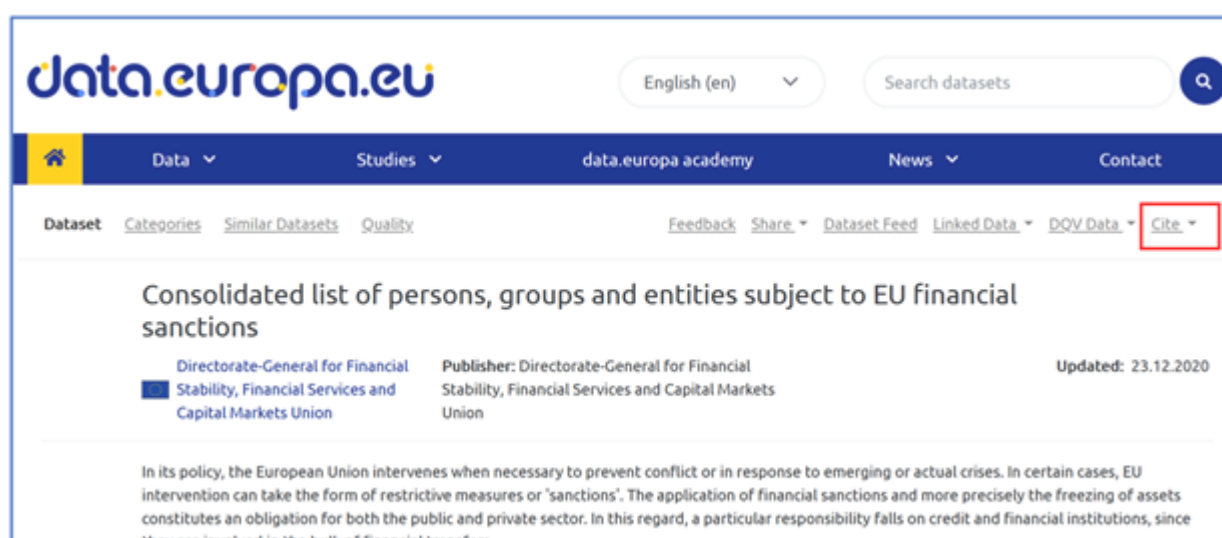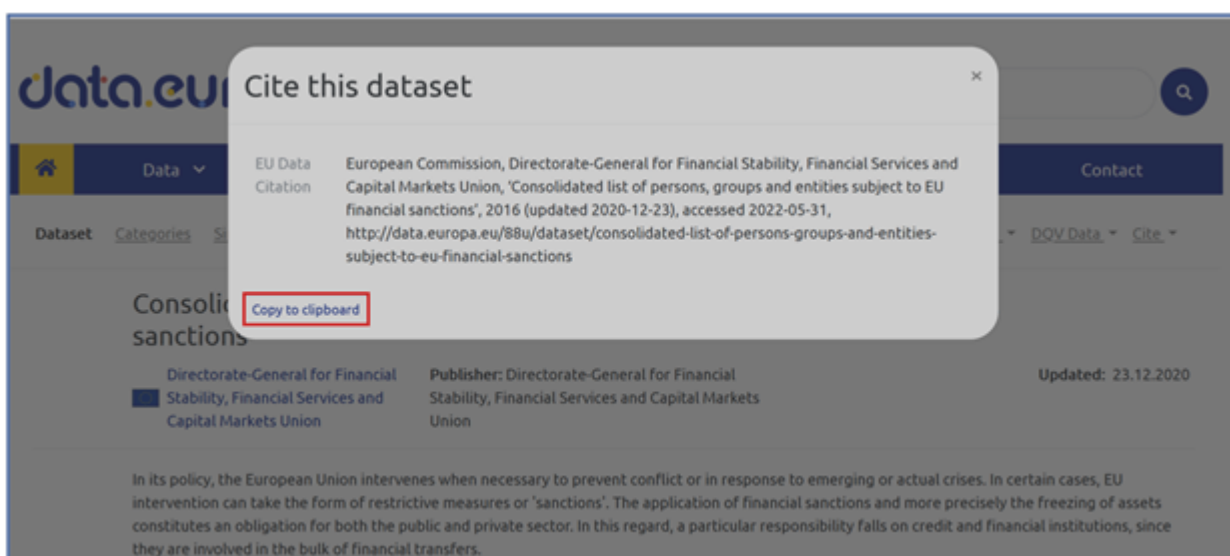
How can you use it?

Click on 'Cite' and choose your preferred citation style. You can choose among the following: EU Data Citation, APA, Harvard, Vancouver.

Once you click on your preferred style, the citation will display on the screen.

Click the 'Copy to clipboard' and you can then paste the citation wherever you need.

10. Data visualisation

Half of the human brain is dedicated to processing things visually. Therefore, visualising information helps people perceive and make sense of concepts and relationships in a complex modern world today. As a result, effective visualisations can contribute to a significantly better understanding of data and policies. They can achieve a considerably higher impact on policymakers and citizens and contribute to considerably higher political support.

The OP offers various services related to this competence, in the portal data.europa.eu, organisation of conferences, webinars, training material, etc.

10.1 Visualisation tool

The **visualisation tool** uses as data source the files as provided by the source. By doing so, it allows users to grasp the data quickly and to understand if the dataset provides the information that the user is looking for. The 'Preview' button is available for all datasets in CSV format (more formats will be added over time). It can show the data in tabular and graphical format without the need for downloading.

Potential errors: it is possible that the tool may not accept the provided file format or that the files may be corrupted at the source. The portal has no influence on the source files.

10.2 Geo-Visualisation

In addition to these functionalities for handling common types of geospatial data, the Geo- Visualisation component is also capable of previewing selected types of real-time data sources. Currently, certain themes provided via FIWARE Context Broker instances are supported. As shown below, the preview of real-time data can be opened in a similar manner as for the other types of geospatial data.

The screenshot below shows an example of live tracking data of public buses in the city of Malaga. After opening the preview of such a data source, the view will be refreshed in regular time intervals (60 seconds). If the user clicks on one of the features displayed on the map, more detailed information is shown (in this case different properties of the tracked bus).
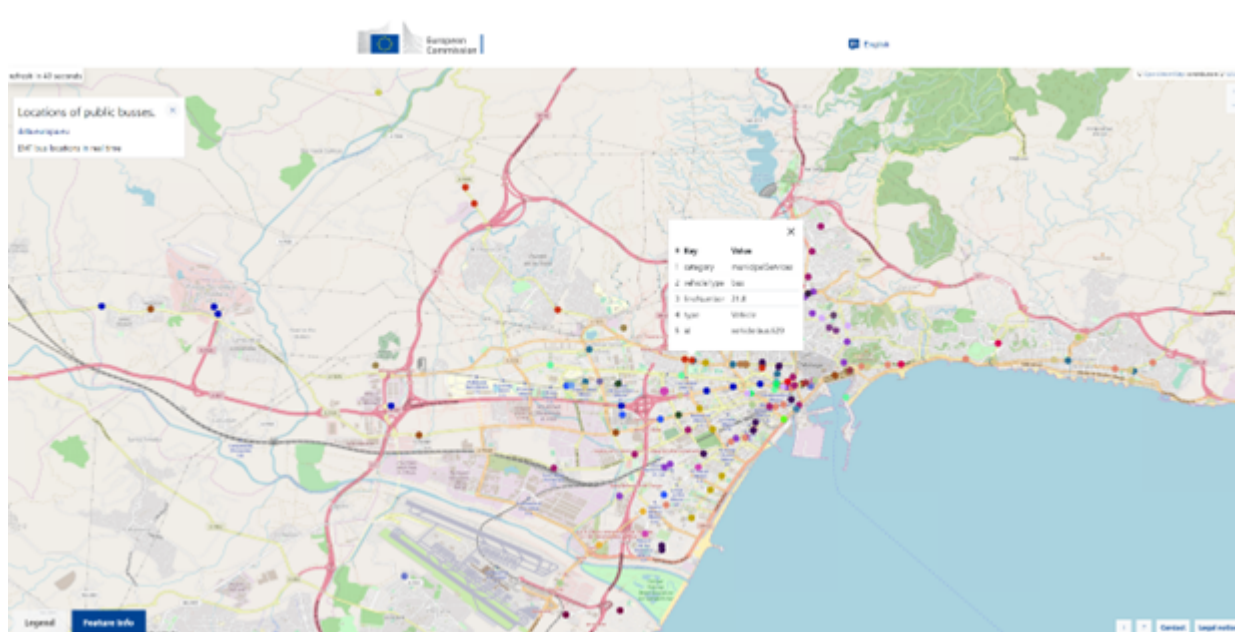
Potential errors: Sometimes a URL to a data source (WMS, GeoJSON file, or real-time data source) is broken or cannot be accessed temporarily. In these cases, the problem needs to be solved by the data provider.

For a better understanding you can watch the introduction to geospatial data webinar, other geospatial topics at the data.europa academy, and a webinar on real time data.

10.3 Further information

EU DataViz is an international conference on open data and data visualisation addressing the needs of the public sector across the EU. It is organised by the OP and the first edition was in 2019, followed by the second edition in 2021 as part of the EU Open Data Days.

In a nutshell, the conference addressed how to translate complex issues and knowledge into digestible and relevant language that resonates with EU citizens and policymakers. Additionally, it helps to explore and analyse data to develop better policies.

Featuring expert speakers, EU DataViz 2021 provided an overview of innovative techniques and best practices used in both the private and the public sectors, offering the participants valuable insights into open-data and data-visualisation techniques and practices. The recordings, programme and presentations are available on the website.

Watch the EU Open Data Days video and discover the EU DataViz event.

The data.europa academy contains a section specifically on data visualisation offering deeper understanding on this topic, its potential benefits and the necessary techniques to visualise open data.

11. Accessibility Statement

This statement applies to content published on the domain: 'https://data.europa.eu/'.

It does not apply to other content or websites published on any of its subdomains. These websites and their content will have their own specific accessibility statement.

This website is managed by the Publications Office of the European Union. It is designed to be used by as many people as possible, including people with disabilities.

You should be able to:

• zoom up to 200% without problems

• navigate most of the website using just a keyboard

• navigate most of the website using a modern screen reader and speech recognition software (on your computer or phone)

This website is designed to comply with the technical standard for websites and mobile apps, EN 301 549, v. 3.2.1. This closely follows level 'AA' of the Web Content Accessibility Guidelines (WCAG) version 2.1.

11.1 Compliance status

This website is partially **compliant** with technical standard EN 301 549 v.3.2.1 and the Web Content Accessibility Guidelines (WCAG) 2.1 Level AA. See 'Non-accessible content' for more details.

The website was last tested on 14 September 2023.

11.2 Preparation of this statement

This statement was reviewed on 30/10/2023.

The statement is based on a review of a representative sample of web pages by an IAAP-qualified accessibility expert, using a combination of manual and automated testing.

11.3 Feedback

We welcome your feedback on the accessibility of the Data.europa.eu website. Please let us know if you encounter accessibility barriers: https://data.europa.eu/en/contact-us

11.4 Compatibility with browsers and assistive technology

The Data.europa.eu website is designed to be compatible with the following most used assistive technologies:

• the latest version of Google Chrome and Apple Safari browsers;

• in combination with the latest versions of JAWS, NVDA, VoiceOver and TalkBack.

11.5 Technical specifications

The accessibility of the Data.europa.eu website relies on the following technologies to work with the particular combination of web browser and any assistive technologies or plugins installed on your computer:

- HTML
- WAI-ARIA
- CSS
- JavaScript

11.6 Non-accessible content

Despite our best efforts to ensure accessibility of the Data.europa.eu website, we are aware of some limitations, which we are working to fix. Below is a description of known limitations and potential solutions. Please contact us if you encounter an issue not listed below.

Known limitations for the Data.europa.eu website:

1. Some buttons and elements can not be accessed (reached) using keyboard only.

2. Some ARIA attributes are used inappropriately, which may result in not correct screen reader presentation

3. Some screen readers will not be able to read buttons or inputs correctly as they do not contain text or labels which screen readers can recognize and present.

4. The focus order of content on some pages is not always logical when navigating with a keyboard. This means that users using keyboard only for navigating the page will not be able to use it and activate functions in logical sequence.

5. Some images may not always have an appropriate alternative attribute describing its function.

6. Some links placed in the sentences may have insufficient colour contrast with the surrounding non-link text.

7. Some buttons may not have sufficient colour contrast which may make them less visible.

8. Some id attributes of the elements may be duplicated which can make it difficult to identify the component.

12. Glossary of terms

12.1 ADMS (asset description metadata schema)

A vocabulary to describe interoperability assets making it possible for ICT developers to explore and search for interoperability assets. ADMS allows public administrations, businesses, standardisation bodies and academia to:

- describe semantic assets in a common way so that they can be seamlessly cross-queried and discovered by ICT developers from a single access point.
- search, identify, retrieve and compare semantic assets, avoiding duplication and expensive design work through a single point of access.
- keep their own system for documenting and storing semantic assets.
- improve indexing and visibility of their own assets.
- link semantic assets to one another in cross-border and cross-sector settings.

Source: https://joinup.ec.europa.eu/asset/adms/description

12.2 API

Application programming interface (API) is a way computer programmes talk to one another. It can be understood in terms of how a programmer sends instructions between programmes.

Source: http://schoolofdata.org/handbook/appendix/glossary

12.3 CKAN (comprehensive knowledge archive network)

A data management system that makes data accessible by providing tools to streamline publishing, sharing, finding and using data. CKAN is aimed at data providers (national and regional governments, companies and organisations) working to make their data open and available.

Source: http://ckan.org/

12.4 Controlled vocabularies

Organised arrangements of words and phrases used to index and/or to retrieve content. A collection of controlled vocabularies is, for example, part of the multilingual metadata registry (http://publications.europa.eu/mdr/index.html). Known also as authority tables, they group concepts like languages, currencies, interinstitutional procedures and many others.

12.5 CORDIS (Community Research and Development Information Service)

The European Commission's primary public repository and portal to disseminate information on all EU-funded research projects and their results.

Source: http://cordis.europa.eu/

12.6 (Data) crawling

A crawler is a programme that visits websites and reads their pages and other information to create entries for a search engine index, among others. All major search engines on the web have such a programme, which is also known as a 'spider' or a 'bot'.

Source: http://searchsoa.techtarget.com/definition/crawler

When extracting data from the web, the term 'crawling' is often also referred to as 'data scraping' or 'harvesting'. There is a difference between these terms: crawling refers to dealing with datasets where someone can develop their own crawlers (or bots), which crawl to the deepest parts of the web pages; data scraping on the other hand refers to retrieving information from any source (not necessarily from the web).

Source: https://www.promptcloud.com/blog/data-scraping-vs-data-crawling

12.7 CSV (comma-separated values)

'Comma-separated values' file format, often used to exchange data between differently similar applications. The CSV file format is useable by KSpread, OpenOffice Calc and Microsoft Excel spreadsheet applications. Many other applications support CSV to import or export data.

Source: http://edoceo.com/utilitas/csv-file-format

12.8 Data citation

Refers to the practice of providing a reference to data in the same way as researchers routinely provide a bibliographic reference to outputs such as journal articles, reports, and conference papers.

Source: https://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation

12.9 Dataset

A collection of related sets of data that is composed of separate elements but that can be processed as a whole and accessed or downloaded in one or more formats.

12.10 DCAT (Data Catalogue Vocabulary)

An RDF vocabulary for interoperability of data catalogues.

See also: http://www.w3.org/TR/vocab-dcat

12.11 DCAT-AP (Data Catalogue Vocabulary Application Profile)

A common vocabulary for describing datasets hosted in data portals in Europe, based on the DCAT. It is the W3C standard.

See also: https://joinup.ec.europa.eu/asset/dcat_application_profile/description

12.12 (Data) dump

A large amount of data transferred from one system or location to another.

Source: http://www.oxforddictionaries.com

12.13 DCMI (Dublin core metadata initiative)

An open organisation supporting innovation in metadata design and best practices across the metadata ecosystem.

Source: http://dublincore.org/

12.14 ELI (European legislation identifier)

It allows to uniquely identify and access national and European legislation online and to guarantee easier access, exchange, and use of legislation for public authorities, professional users, academics and citizens. ELI paves the way for a semantic web of legal gazettes and official journals.

Source: https://en.wikipedia.org/wiki/European_Legislation_Identifier

12.15 FOAF

'Friend of a friend' is a machine-readable descriptive vocabulary of persons, their activities and their relations to other people and objects. FOAF allows groups of people to describe social networks without the need for a centralised database.

Source: https://en.wikipedia.org/wiki/FOAF_%28ontology%29

12.16 GeoDCAT

GeoDCAT-AP is an extension of DCAT-AP for describing geospatial datasets, dataset series and services. It provides an RDF syntax binding for the union of metadata elements defined in the core profile of ISO 19115:2003 and those defined in the framework of the Inspire directive. Its basic use case is to make spatial datasets, data series and services searchable on general data portals, thereby making geospatial information better searchable across borders and sectors. This can be achieved by the exchange of descriptions of datasets among data portals.

The GeoDCAT-AP specification does not replace the Inspire metadata regulation nor the Inspire metadata technical guidelines based on ISO 19115 and [ISO 19119](https://www.iso.org/standard/59221.html. Its purpose is given owners of geospatial metadata the possibility to achieve more by providing an additional RDF syntax binding.

More information: https://joinup.ec.europa.eu/release/geodcat-ap-v10

12.17 IMMC (Interinstitutional Metadata Maintenance Committee)

Interinstitutional Metadata Maintenance Committee. The minimum set of metadata elements, the so-called IMMC core metadata, that is to be used in the data exchange.

Source: https://op.europa.eu/en/web/eu-vocabularies/immc

12.18 Interoperability

The ability of systems to easily exchange information and use the exchanged information.

12.19 JSON (JavaScript object notation)

JavaScript object notation is an open-standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs. It is the most common data format used for asynchronous browser/server communication (AJAX, or asynchronous JavaScript and XML).

Source: https://en.wikipedia.org/wiki/JSON

12.20 Legal notice

An important metadata element is the legal notice for your data. Use of content catalogued in the EU Open Data Portal is permitted free of charge for commercial or non-commercial purposes. According to its copyright notice, 'Reuse is authorised provided the source is acknowledged,' unless otherwise stated. This follows the principles of the reuse policy implemented through Directive 2013/37/EU and Decision 2011/833/EU. In the same way, reuse should fully respect privacy legislation and does not apply to data subject to the intellectual property rights of third parties. In limited cases it can be subject to conditions (Article 2(2) of Decision 2011/833/EU).

12.21 Linked data

Linked data describes a method of publishing structured data so that they can be interlinked. It builds upon standard web technologies such as HTTP and URI, but rather than using them to serve web pages for human readers it extends them to share information in a way that can be automatically read by computers.

Source: https://en.wikipedia.org/wiki/Linked_data

Linked data is one of the core pillars of the 'semantic web', also known as the 'web of data'. The semantic web is about making links between datasets that are understandable not only to humans, but also to machines, and linked data provides the best practices for making these links possible. In other words, linked data is a set of design principles for sharing machine-readable interlinked data on the web.

Source: https://ontotext.com/knowledgehub/fundamentals/linked-data-linked-open-data/

12.22 Linked data principles

Linked data principles provide a common API for data on the web that is more convenient than many separately and differently designed APIs published by individual data suppliers. Tim Berners-Lee, the inventor of the web and the initiator of the linked data project, proposed the following principles upon which linked data is based:

- use URIs to name things.
- use HTTP URIs so that things can be referred to and looked up (dereferenced) by people and user agents.
- when someone looks up a URI, provide useful information using open web standards such as RDF or SPARQL.
- include links to other related things using their URIs when publishing on the web.

Source: http://www.w3.org/TR/ld-glossary/#linked-data-principles

12.23 Machine readable

Machine-readable data are data in a format that can be interpreted by a computer program. There are two types of machine-readable data:

- human-readable data that are marked up so that they can also be understood by computers, for example, microformats and RDFa.
- data formats intended principally for computers, for example, RDF, XML and JSON.

12.24 Mashup

The combination of multiple datasets from multiple sources to create a new service, visualisation, or information.

12.25 Metadata

Metadata is structured information that describes, explains, locates or otherwise makes it easier to retrieve, use or manage an information resource. Metadata is often referred to as data about data. Metadata is important for many reasons, most specifically to:

- facilitate discovery.
- enable high ranking of search results.
- enable refinement of a search.
- help organise electronic datasets.
- provide digital identification.
- support archiving and preservation.
- facilitate interoperability, that is, the ability of systems to exchange information and use the exchanged information.

Why is the quality of metadata important?

Metadata is the first indicator for a qualitative assessment of a dataset, as it provides information about the content and the quality of the data. In short, metadata enables users to discover the data and understand the structure of the data, the terms under which it can be reused and its origin.

Source: http://www.niso.org/publications/press/UnderstandingMetadata.pdf

12.26 Metadata registry

The metadata registry is an important interoperability and standardisation tool. It registers and maintains definition data (metadata elements, named authority lists, schemas, etc.) used by the different European institutions.

Source: http://publications.europa.eu/mdr/index.html

12.27 (Data) mining

The practice of examining large pre-existing databases to generate added information.

Source: http://www.oxforddictionaries.com

12.28 Multilingual thesauri

A standard vocabulary (i.e. EuroVoc) which can be easily translated to other languages. For international interoperability it is useful to use multilingual thesauri.

12.29 Ontology

A formal model that allows knowledge to be represented for a specific domain. An ontology describes the types of things that exist (classes), the relationships between them (properties) and the logical ways those classes and properties can be used together (axioms).

Source: http://www.w3.org/TR/ld-glossary/#ontology

12.30 Open government data

Data collected, produced or paid for by the public bodies and made freely available for use for any purpose.

12.31 Open standards

Generally understood as technical standards that are free from licencing restrictions. They can also be interpreted to mean standards that are developed in a vendor-neutral manner.

Source: http://schoolofdata.org/handbook/appendix/glossary

12.32 (Data) parsing

Breaking a data block into smaller chunks by following a set of rules so that it can be more easily interpreted, managed or transmitted by a computer.

Source: http://www.businessdictionary.com/definition/parsing.html

12.33 PDF (Portable Document Format)

A file format used to present and exchange documents independently of software, hardware, or operating systems. It is an open standard maintained by the International Organisation for Standardisation.

Source: https://acrobat.adobe.com/be/en/products/about-adobe-pdf.html

12.34 PSI (public sector information)

It is the wide range of information that public sector bodies collect, produce, reproduce and disseminate in many areas of activity while accomplishing their institutional tasks. It can be made available under a variety of (not always open) licences.

12.35 Raw data

An expression that refers to data in their original state, not having been processed, aggregated or manipulated in any other way. It is also defined as 'primary data'.

12.36 RDF (Resource Description Framework)

A family of international standards for data interchange on the web. RDF is based on the idea of identifying things using web identifiers or HTTP URIs and describing resources in terms of simple properties and property values.

Source: http://www.w3.org/TR/ld-glossary/#rdf

12.37 RDFa

Resource description framework in attributes is a W3C recommendation that adds a set of attribute-level extensions to HTML, XHTML and various XML-based document types for embedding rich metadata within web documents.

Source: https://en.wikipedia.org/wiki/RDFa

12.38 Resource

The physical representation of a dataset. Each resource can be a file of any kind, a link to a file elsewhere on the web or a link to an API. For example, if the data is being supplied in multiple formats or split into different areas or time periods, each file is a different 'resource' that should be described individually.

12.39 Semantic web

An evolution or part of the web that consists of machine-readable data in RDF and an ability to query that information in standard ways (e.g. via SPARQL).

Source: http://www.w3.org/TR/ld-glossary/#semantic-web

12.40 (Data) scraping

The process of extracting data in machine-readable formats of non-pure data sources, for example webpages or PDF documents. The term is often prefixed with the source (e.g. web scraping, PDF scraping).

Sources: http://en.wikipedia.org/wiki/Data_scraping

12.41 SDMX (statistical data and metadata exchange)

An international initiative that aims at standardising and modernising the mechanisms and processes for the exchange of statistical data and metadata among international organisations and their member countries.

Source: https://en.wikipedia.org/wiki/SDMX

12.42 SHACL

Shapes Constraint Language (SHACL) is a W3C specification for validating graph-based data against a set of conditions. Among others, SHACL includes features to express conditions that constrain the number of values that a property may have, the type of such values, numeric ranges, string matching patterns and logical combinations of such constraints. SHACL also includes an extension mechanism to express more complex conditions in languages such as SPARQL.

Source: https://en.wikipedia.org/wiki/SHACL

12.43 SPARQL

SPARQL protocol and RDF query language (SPARQL) defines a query language for RDF data, analogous to the Structured Query Language (SQL) for relational databases.

Source: http://www.w3.org/TR/ld-glossary/#sparql

12.44 SPARQL endpoint

A service that accepts SPARQL queries and returns answers as SPARQL result sets. It is a best practice for dataset providers to give the URL of their SPARQL endpoint to allow access to their data programmatically or through a web interface.

Source: http://www.w3.org/TR/ld-glossary/#sparql-endpoint

12.45 StatDCAT

The StatDCAT application aims at providing a commonly agreed dissemination vocabulary for statistical open data. StatDCAT-AP defines a certain number of additions to the DCAT-AP model that can be used to describe datasets in any format, for example, those published in SDMX, a standard for the exchange of statistical data.

The principal objective of the development of the StatDCAT-AP, which is funded under the ISA2 action of the European Commission on 'Promoting semantic interoperability amongst the European Union Member States (SEMIC)', is to facilitate a better integration of the existing statistical data portals within open data portals, thus improving the discoverability of statistical datasets across domains, sectors and borders. This will be beneficial for the general data portals, enabling enhanced services for the discovery of statistical data.

More information: https://joinup.ec.europa.eu/solution/statdcat-application-profile-data-portals-europe

12.46 Structured data

Data that reside in fixed fields within a record or file. Relational databases and spreadsheets are examples of structured data. Although data in XML files are not fixed in location like traditional database records, they are still structured because the data are tagged and can be accurately identified.

Source: http://www.pcmag.com/encyclopedia/term/52162/structured-data

12.47 Triplestore

A triplestore is a purpose-built database for the storage and retrieval of triples through semantic queries. A triple is a data entity composed of subject-predicate-object, like 'Bob is 35' or 'Bob knows Fred'. Much like a relational database, information is stored in a triplestore and retrieved via a query language. Unlike a relational database, a triplestore is optimised for the storage and retrieval of triples. In addition to queries, triples can usually be imported/exported using RDF and other formats.

Source: http://en.wikipedia.org/wiki/Triplestore

12.48 URI (uniform resource identifier)

A string that uniquely identifies virtually anything, including a physical building or more abstract concepts such as colours. It may or may not be resolvable on the web.

Source: http://www.w3.org/TR/ld-glossary/#uniform-resource-identifier

12.49 URL (uniform resource locator)

A global identifier commonly called a 'web address'. A URL is resolvable on the web. All HTTP URLs are URIs; however, not all URIs are URLs.

Source: http://www.w3.org/TR/ld-glossary/#uniform-resource-locator

12.50 Vocabulary

A collection of terms for a particular purpose. Vocabularies can range from simple, such as the widely used RDF schema, FOAF and DCMI element set to complex vocabularies with thousands of terms, such as those used in healthcare to describe symptoms, diseases and treatments. Vocabularies play an especially significant role in linked data, specifically to help with data integration. The use of this term overlaps quite often with that of 'ontology'.

Source: http://www.w3.org/TR/ld-glossary/#vocabulary

12.51 XML (Extensible Markup Language)

It is a markup language that defines a set of rules for encoding documents in a format which is both human readable and machine readable. See the standard here: https://www.w3.org/XML/

Source: https://en.wikipedia.org/wiki/XML