

WEBINAR

From data to metadata: enhancing quality across borders

data.
europa
academy

7 March 2025

10:00 – 11:30 CET



Rules of the game



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



For questions, please use the ClickMeeting chat



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



Agenda

10.00 – 10.10	Opening and introduction – <i>Inmaculada Farfan Velasco</i>
10.10 – 10.20	The approach of the Open Knowledge Foundation – <i>Renata Ávila</i>
10.20 – 10.30	Sneak peek of the Open Data Editor – <i>Romina Colman</i>
10:30 – 10:45	The approach of the Open Data Charter – <i>Natalia Carfi</i>
10:45 – 11:05	Metadata quality approach from Danish Agency for Digital Government – <i>Christian Nicolai Larsen</i>
11:05 – 11:25	Q&A session
11:25 – 11:30	Closing remarks

Speakers



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



Renata Ávila
CEO,
Open Knowledge
Foundation



Romina Colman
Product Owner,
Open Data Editor



Natalia Carfi
CEO,
Open Data Charter



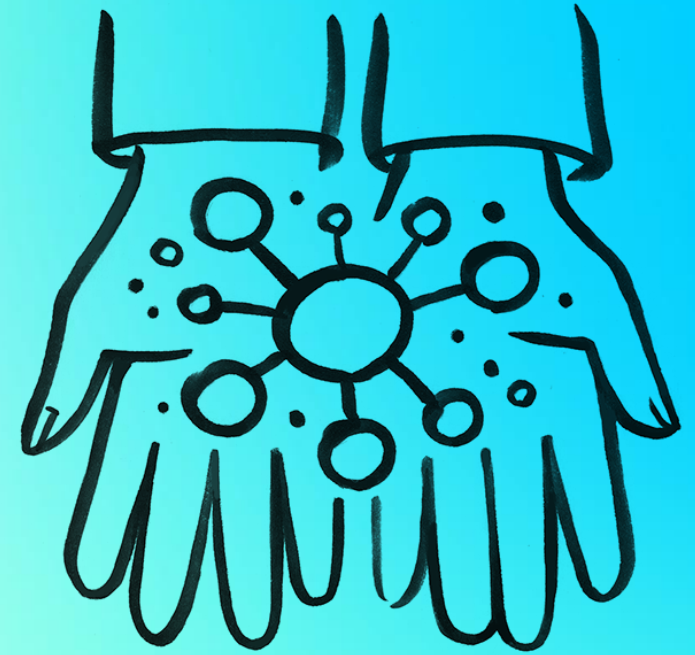
Christian Nicolai Larsen
Policy Officer,
Danish Agency for Digital
Government

The approach of the Open Knowledge Foundation

Renata Ávila
CEO, Open Knowledge Foundation



**We are building a
world open by design
where all knowledge is
accessible to
everyone.**



Open Knowledge
FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

okfn.org
@okfn

About OKFN

The **Open Knowledge Foundation** (OKFN) is the world's ultimate reference in open digital infrastructures and the hub of the open movement. As a global not-for-profit, we have been **establishing and advocating for open standards for the last 20 years**. We provide services, tools and training for institutions to adopt **openness as a design principle**.

Learn more – okfn.org/who-we-are



About OKFN

Our current activities are focused on 3 interrelated pillars:



Community

A Network present in
40+ countries



Open Tech

Tools and Services
for opening data



Advocacy

Digital infrastructure
to be open by design





Network

The Open Knowledge Network connects those in the open knowledge movement **across 40+ countries**.

OKFN provides tools and infrastructure, and manages the community. Chapters and local groups work together to support each other, to share expertise, and to amplify open knowledge activities around the globe.

Learn more – okfn.org/network

Countries where we are present →



A few Network projects



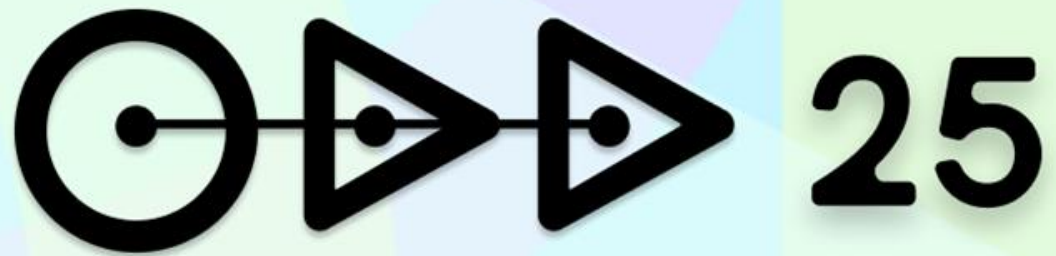
Open Data Day [↗](#)



Project Repository [↗](#)



Global Directory [↗](#)



Open Data Day

opendataday.org



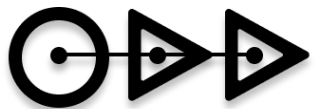
Open Knowledge
FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

okfn.org
[@okfn](https://twitter.com/okfn)

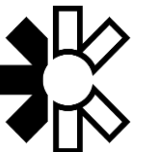
About

Open Data Day (ODD) is an annual celebration of open data all over the world, where groups and communities gather to reach out to new people and build new solutions using open data.

It is an opportunity to show the benefits of open data and encourage the adoption of **open data policies by government, business, and civil society.**



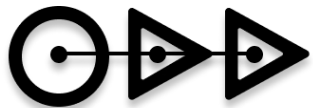
Open Data Day



About

Open Data Day is a bottom-up, decentralised initiative created in 2010 and led by the **Open Knowledge Foundation** (OKFN) for the last 9 years.

Participating and supporting these actions is one of our ways to promote the **sustainability of the open movement** and help circulate knowledge globally in a fair, sustainable and open way.



Open Data Day



Data Quality and its future

Challenges for data interoperability and data reuse.



Open Knowledge

FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

Challenges in open data quality

1. Siloed Data

- Fragmentation across departments
- Inconsistent formats (e.g., CSV vs. proprietary schemas)

2. Data Freshness

- Keeping real-time data current (e.g., transit schedules)
- Manual updates strain resources

3. Resource Constraints

- Limited staff, funding, and tools
- Quality often deprioritized

4. Governance Gaps

- Lack of ownership and accountability
- No centralized standards

Going beyond the basics



Consistency



Traceability



Granularity



Metadata
quality



Best Practices for High-Quality Open Data



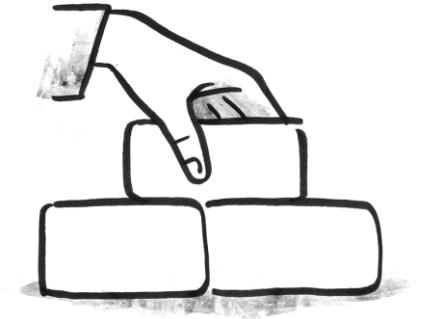
A clear data governance framework



Automatised quality assurance and validation

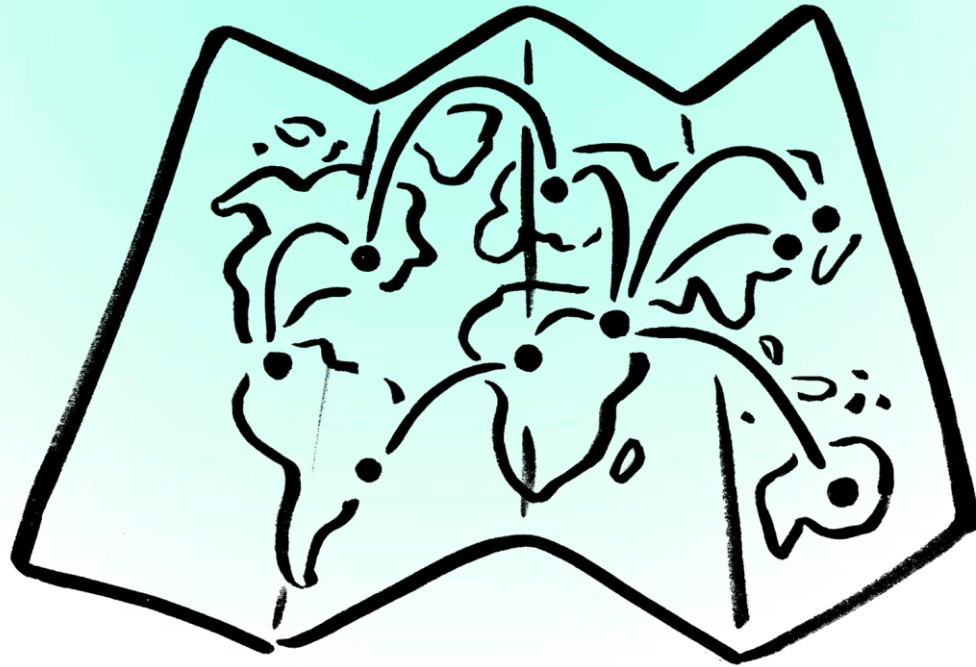


Metadata management



Standardisation

Examples

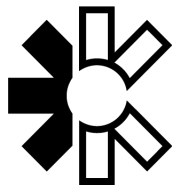


- **Kenya Open Data Initiative**
 - Centralized portal with standards
 - Community feedback improves quality
- **Brazil's INEP Education Data**
 - Automated checks for school census
 - Real-time API monitoring

Result: Better school resources
- **India's Open Government Data Platform**
 - Metadata standards and validation
 - Harmonized departmental data

Example: Agricultural market insights
- **South Africa's Stats SA**
 - Consistent, granular census data
 - Detailed documentation

Result: Improved policy-making

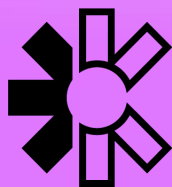


Open Knowledge

FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

The future of Data Quality and Artificial Intelligence

- ◆ Automated Quality Checks: Real-time validation (e.g., TensorFlow Data Validation)
- ◆ Predictive Maintenance: Forecasts update needs (e.g., public health data)
- ◆ Enhanced Interoperability: Maps diverse datasets (e.g., environmental data)



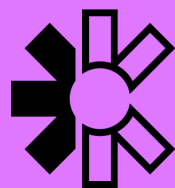
Open Knowledge

FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

Conclusions

- Quality is a strategic asset
 - Address silos, governance, and resources smartly
 - Metadata and AI are key enablers
-
- ★ Assess your data quality maturity
 - ★ Pilot a best practice (e.g., metadata management)
 - ★ Share with the open data community
 - ★ Test Open Data Editor!

Thanks!



Open Knowledge

FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

okfn.org
[@okfn](https://twitter.com/okfn)

Sneak peek of the Open Data Editor

Romina Colman
Product Owner, Open Data Editor

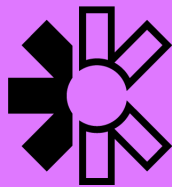




Open Data Editor

The Open Data Editor (ODE) is a desktop app to help **non-technical** data practitioners to **explore and detect errors** in their tables.

opendataeditor.okfn.org



Open Knowledge
FOR A FAIR, SUSTAINABLE AND OPEN FUTURE

March 2025

Tables

Local files: no clouds; no vendor lock-in

Use AI e.g. "create csv with 10 most populated cities in the world"

Project: current

table-cars.csv

SQL VIEW CHART VALID

Editor Metadata Report Source Errors

	brand	model	price	kmpl	bhp	type
2	Volkswagen	Vento	785	16.1	104	Sedan
3	Hyundai	Verna	774	17.4	106	Sedan
4	Skoda	Rapid	756	15	104	Sedan
5	Suzuki	Ciaz	725	20.7	91	Sedan
6	Renault	Scala	724	16.9	98	Sedan
7	Suzuki	SX4	715	16.5	103	Sedan

Page 1 of 1 Results per page 50 Showing 1 - 42 of 42

```
1 brand,model,price,kmpl,bhp,type
2 Volkswagen,Vento,785,16.1,104,Sedan
3 Hyundai,Verna,774,17.4,106,Sedan
4 Skoda,Rapid,756,15,104,Sedan
5 Suzuki,Ciaz,725,20.7,91,Sedan
6 Renault,Scala,724,16.9,98,Sedan
7 Suzuki,SX4,715,16.5,103,Sedan
8 Fiat,Linea,700,15.7,112,Sedan
9 Nissan,Sunny,699,16.9,98,Sedan
10 Fiat,Linea Classic,612,14.9,89,Sedan
11 Toyota,Etios,603,16.8,89,Sedan
12 San,Storm,595,16,59,Sedan
13 Chevrolet,Sail,551,18.2,82,Sedan
14 Volkswagen,Polo,535,16.5,74,Hatchback
15 Hyundai,i20,523,18.6,82,Hatchback
16 Honda,Amaze,519,18,87,Sedan
17 Suzuki,Swift DZire,508,19.1,86,Sedan
18 Ford,Classic,506,14.1,100,Sedan
19 Skoda,Fabia,503,16.4,75,Hatchback
20 Toyota,Etios Liva,500,17.7,79,Hatchback
```

Fields

ADD FIELD GRID VIEW Search

id	INTEGER	REMOVE
name	STRING	REMOVE

table-invalid.csv (invalid)

Type Error

1	LatD	LatM	LatS	NS	LongD	LongM
124	45	38	23	N	54	25
125						

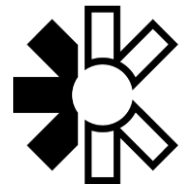
CREATE MANAGE DELETE SAVE AS PUBLISH REVERT SAVE

Sort, pagination, etc

Maps

Use AI e.g. “create a map with all the capitals in Latin America”

The screenshot shows the Frictionless Application interface. At the top, there's a blue header with 'Frictionless Application' and a 'beta' badge. To the right of the header are buttons for 'MAP', 'map.geojson @map', 'CREATE', and 'VALID'. Below the header is a navigation bar with 'Editor', 'Metadata', 'Report', 'Source', and 'Chat AI'. On the left, a sidebar shows a list of files: 'article.md @article', 'chart.json @chart', 'datapackage.json @datapackage', 'image.jpg @image', 'json.yaml @json', 'map.geojson @map' (highlighted), 'oil-prices.json @oil_prices', 'schema.json @schema', 'script.py @script', 'table-cars.csv @table_cars', 'table-invalid.csv @table_invalid', 'table.csv @table', 'text.txt @text', and 'view.json @view'. The main area is a map of Europe with pins on London and Paris. At the bottom, there's a toolbar with 'ADD', 'MANAGE', 'DELETE', 'SAVE AS', 'PUBLISH', 'REVERT', and 'SAVE' buttons.



Articles

Use AI e.g. “write an article about the history of Portugal”

The screenshot shows the Frictionless Application interface. The top navigation bar includes 'Frictionless Application beta', 'ARTICLE', 'article.md @article', 'CREATE', and 'VALID'. Below the navigation bar, there are tabs for 'Editor', 'Metadata', 'Report', 'Source', 'Chat AI', 'Undo', 'Redo', and 'Clear'. The main editor area displays a markdown document with the following content:

```
1 # History of Portugal
2
3 @image
4
5 Portugal, located in southwestern Europe on the Iberian Peninsula, has a rich and diverse history that spans over thousands of years. From ancient civilizations and Roman rule to a powerful maritime empire and modern democracy, Portugal's history is filled with significant events and cultural developments.
6
7 ## Ancient Civilization and Roman Rule
8
9 @map
10
11 The history of Portugal dates back to prehistoric times, with evidence of human presence as early as 30,000 BC. Throughout the centuries, various civilizations inhabited the region, including the Celts and the Phoenicians. In 218 BC, the Romans conquered the Iberian Peninsula and established the province of Lusitania, comprising the territory of modern-day Portugal.
12
13 ## Moorish Influence and the Reconquista
14
15 @chart
16
17 In the 8th century, the Moors, an Islamic civilization from North Africa, invaded the Iberian Peninsula, including Portugal. The Muslim rule lasted for over four centuries and left a lasting impact on the language, culture, and architecture of the region. However, in the 11th century, the Christian kingdoms of Portugal initiated the
```

The preview area on the right shows the rendered article with the following sections:

Ancient Civilization and Roman Rule

The history of Portugal dates back to prehistoric times, with evidence of human presence as early as 30,000 BC. Throughout the centuries, various civilizations inhabited the region, including the Celts and the Phoenicians. In 218 BC, the Romans conquered the Iberian Peninsula and established the province of Lusitania, comprising the territory of modern-day Portugal.

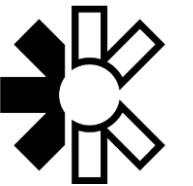
Moorish Influence and the Reconquista

Group	Value
x	0.7
y	0.9
z	1.1

The interface also features a sidebar on the left with a file list, a bottom bar with 'ADD', 'MANAGE', and 'DELETE' buttons, and a bottom right area with 'SAVE AS', 'PUBLISH', 'REVERT', and 'SAVE' buttons. A large text 'Insert files' is overlaid on the editor area with arrows pointing to the code.

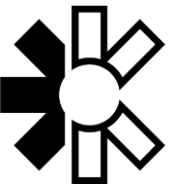


***What did we discover during
the user research phase?***



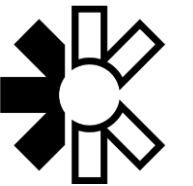


People working with data experience the same old problems.



Let's take a look at how the tool works!

Challenges...





pdelboca commented on Oct 30 • edited ▾

Member ...

@romicolman I think this is a question related to `frictionless-py` since we are using it to read the data and the schema. I would venture to answer that ODE is loading the column as a string because the date is not in ISO8601 format (`YYYY-MM-DD`).

A common data friction is that dates are usually written in different formats like `YYYY/MM/DD` or `DD/MM/YY` or `MM/DD/YY` or even `YYYY-MM-DD` (notice that I changed `/` for `-`). According to the [Table Schema of the Data Package](#) dates needs to be in `YYYY-MM-DD` format. Since this data is `DD/MM/YY` then it is considered as a string.









This is not an error nor a problem with the tool but rather a **data friction** that we need to solve/explain.

The specification of Data Package has a definition on what a `date` should look like, so in order to be able to define it as a `date` in Open Data Editor (that it is based in Data Package) the user should change the values to respect the format of the specification (`YYYY-MM-DD`).



[Upload your data](#)

 Create folder

-  Base de datos - Magis ...
-  Distribucio.de.la.rend ...
-  Ethical Tech Job Resc ...
-  Ethical Tech Job Resc ...
-  Radios precensales 2 ...
-  Wrong data type.csv ...
-  magistrados-justicia-f ...
-  matriculas_por_escol ...

Resource

Integrity

Licenses

Contributors

Sources

> Dialect

> Schema

Resource

Name

Path

Type

Scheme

Format

Title

Media Type

Description

Encoding

Help

Resource

resource

A simple format to describe and package a single data resource such as a individual table or file.

[LEARN MORE](#)

 User guide

 Report an issue

Upload your data

Create folder

- Base de datos - Magis ...
- Distribucio.de.la.rend ...
- Ethical Tech Job Resc ...
- Ethical Tech Job Resc ...
- Radios precensales 2 ...
- Wrong data type.csv ...
- magistrados-justicia-f ...
- matriculas_por_escol ...

🔴 Users also struggled to understand if ODE found errors in their files. In the app the Errors report gets activated only if errors are found, but this was not clear for the users.

🔴 The ODE takes errors explanations from Frictionless. Since Frictionless is mostly used by technical people, the potential users of the ODE found errors description difficult to comprehend.

💡 We decided to include an error counter to clearly show problematic files. Since this change was implemented while user testing sessions were in progress we had the opportunity to test the change and the people we reached out to provide a positive reaction.

🔴 Errors Report 3

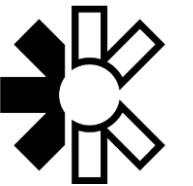


User guide

✅ Almost all users agreed that the ODE's user guide was clear and help them understand how the app and its features work.

❌ The main challenge that we face here is that, although participants read the documentation before the user testing session, we know this is not a common practice when people use apps, so we should not rely on documentation as an element that will address all questions users may have. There are some clarifications that will imply technical changes and user interface work.




ODE main changes



Upload your data

Create folder

- Base de datos - Magis ...
- Distribucio.de.la.rend ...
- Ethical Tech Job Resc ...
- Ethical Tech Job Resc ...
- Radios precensales 2 ...
- Wrong data type.csv ...
- magistrados-justicia-f ...
- matriculas_por_escol ...

-  **Rename**
-  **Open file Location**
The ODE folder where this file exists
-  **Delete file**
Only removes this element from the ODE folder

 User guide

 Report an issue



The ODE supports Excel & csv files

Upload your data

You can also add links to online tables





Upload your data

 Create folder


- Base de datos - Magis ...
- Distribucio.de.la.rend ...
- Ethical Tech Job Resc ...
- Ethical Tech Job Resc ...
- Radios precensales 2 ...
- Wrong data type.csv ...
- magistrados-justicia-f ...
- matriculas_por_escol ...

 User guide

 Report an issue





[From Your Computer](#) Add External Data



Add one or more
Excel or csv files

Select



Add one or more
folders


Select



Upload your data

Create folder

- Base de datos - Magis ...
- Distribucio.de.la.rend ...
- Ethical Tech Job Resc ...
- Ethical Tech Job Resc ...
- Radios precensales 2 ...
- Wrong data type.csv ...
- magistrados-justicia-f ...
- matriculas_por_escol ...

 User guide

 Report an issue


From Your Computer **Add External Data**


Link to the external table:

Add



[Upload your data](#)

 Create folder


 Base de datos - Magis ...


 Distribucio.de.la.rend ...

 Ethical Tech Job Resc ...

 Ethical Tech Job Resc ...

 Radios precensales 2 ...

 Wrong data type.csv ...

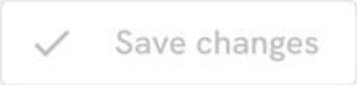
 magistrados-justicia-f ...

 matriculas_por_escol ...

 User guide

 Report an issue

	col1	col2
2	1	2
3	3	4
4	5	6
5	7	8
6	9	10
7	11	12
8	13	14
9	15	16
10	17	18
11	19	20
12	21	bad



Type Error 1

The value does not match the schema type and format for this field.

Row Number	Col1	Col2
12	21	bad
13		



Upload your data

Create folder

Base de datos - Magis

Distribucio.de.la.rend

Ethical Tech Job Resc

Ethical Tech Job Resc

Radios precensales 2

Wrong data type.csv

magistrados-justicia-f

matriculas_por_escol

User guide

Report an issue

	justicia_federal_o_...	camara	organo_tipo	organo_nombre	cargo_tipo	cargo_detalle	n
2	Nacional	Cámara Nacional d...	Cámara	Sala C	Juez	Juez de la Cámara ...	B
3	Nacional	Cámara Nacional d...	Juzgado	Juzgado Nacional e...	Juez	Juez Nacional de P...	P
4	Nacional	Cámara Nacional d...	Juzgado	Juzgado Nacional d...	Juez	Juez Nacional de P...	F
5	Federal	Cámara Nacional d...	Cámara	Sala I	Juez	Juez de la Cámara ...	L
11	Nacional	Cámara Nacional d...	Cámara	Sala I	Juez	Juez de la Cámara ...	A
12	Federal	Cámara Nacional d...	Juzgado	Juzgado Nacional d...	Juez	Juez Nacional de P...	C
13	Federal	Cámara Federal de...	Cámara	Sala A	Juez	Juez de la Cámara ...	B
14	Nacional	Cámara Nacional d...	Cámara	Sala A	Juez	Juez de la Cámara ...	L
15	Nacional	Cámara Nacional d...	Cámara	Sala A	Juez	Juez de la Cámara ...	C

AI Assistant



If you proceed, the Open Data Editor will only share the names of the columns in your table to suggest improvements to the titles and descriptions associated with them. Do you want to proceed?

Cancel

Confirm

	justicia_federal_o_...	camara	organo_tipo	organo_nombre	cargo_tipo	cargo_detalle	magistr
2	Nacional	Cámara Nacional d...	Cámara	Sala C	Juez	Juez de la Cámara ...	BELLUS
3	Nacional	Cámara Nacional d...	Juzgado	Juzgado Nacional e...	Juez	Juez Nacional de P...	PERES,
4	Nacional	Cámara Nacional d...	Juzgado	Juzgado Nacional d...	Juez	Juez Nacional de P...	FERNAI
12	Federal	Cámara Nacional d...	Juzgado	Juzgado Nacional d...	Juez	Juez Nacional de P...	OCHOA
13	Federal	Cámara Federal de...	Cámara	Sala A	Juez	Juez de la Cámara ...	BAIGOF
14	Nacional	Cámara Nacional d...	Cámara	Sala A	Juez	Juez de la Cámara ...	LLANO
15	Nacional	Cámara Nacional d...	Cámara	Sala A	Juez	Juez de la Cámara ...	GALME

Upload your data

Create folder

Base de datos - Magis

Distribucio.de.la.rend

Ethical Tech Job Resc

Ethical Tech Job Resc

Radios precensales 2

Wrong data type.csv

magistrados-justicia-f

matriculas_por_escol

User guide

Report an issue

AI Assistant



Please enter your OpenAI API key:

OpenAI API Key

Click [here](#) to learn how to find your key. You can also check OpenAI terms and policies [here](#).

Cancel

Confirm

Here are the suggested improvements to the column names along with descriptions for each:

1. **jurisdiction_type**

- **Description:** Indicates whether the jurisdiction is federal or national.

2. **court_chamber**

- **Description:** The specific division or chamber of the court.

3. **organ_type**

- **Description:** The type of the entity or governmental body involved.

4. **organ_name**

- **Description:** The name of the specific entity or governmental body.

5. **position_type**

- **Description:** The classification or category of the position held.

6. **position_details**

OK



The approach of the Open Data Charter

*Natalia Carfi
CEO, Open Data Charter*



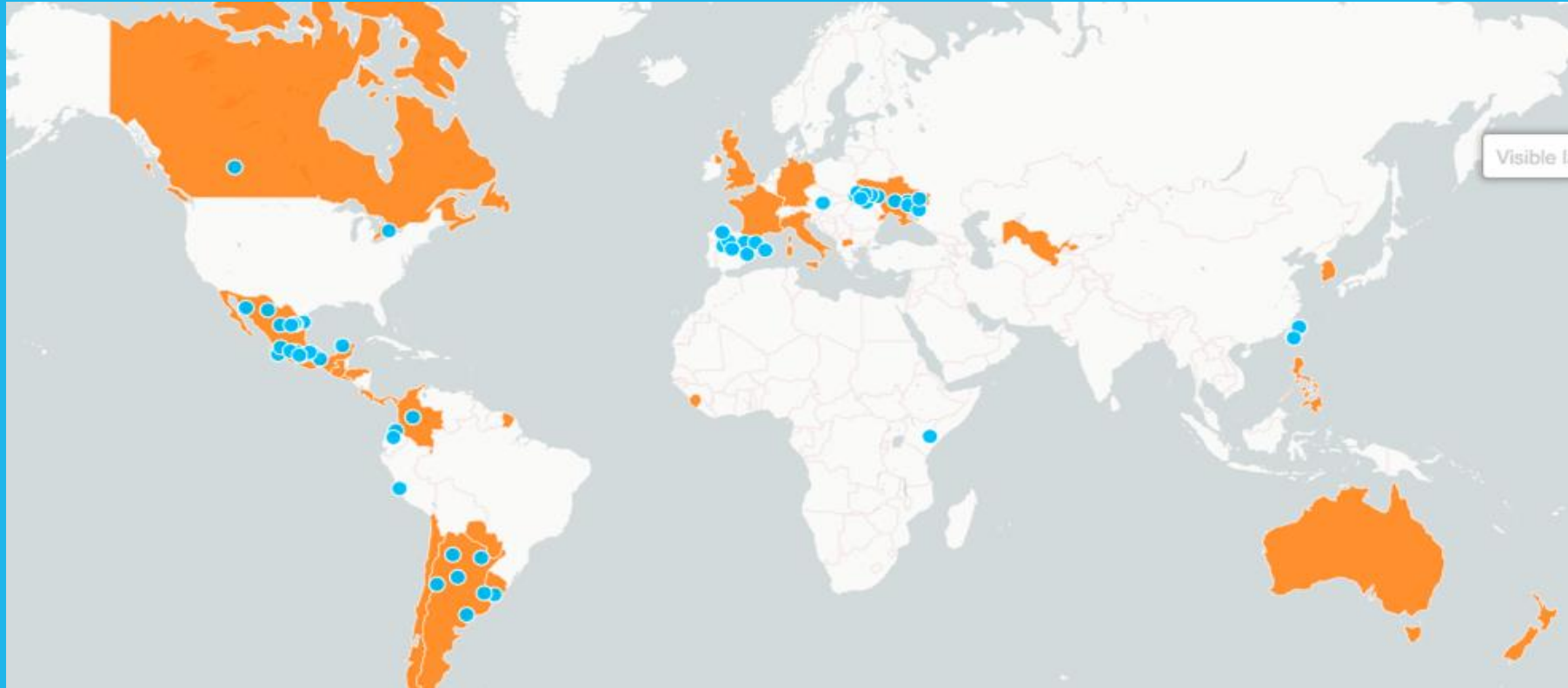


Open Data Charter Principles

“From data to metadata: enhancing quality across borders”

March | 2025

Open data is digital data that is made available with the technical and legal characteristics necessary for it to be **freely used, reused, and redistributed by anyone, anytime, anywhere.**



The Open Data Charter is a collaboration between +200 governments and experts working to open up data. It was founded in 2015 around six principles for how governments should be publishing information.



6 global principles



Open by default



Comparable and interoperable



Timely and comprehensive



For Improved Governance and Citizen Engagement



Accessible and usable



For Inclusive Development and Innovation



1. Open by Default

1. Data held by national, regional, local & city governments, international government bodies, wider public sector, & data created for governments/public
2. Government data is of value of society & economy
3. Global development & adoption of resources, standards, & policies for the creation, use, exchange, & harmonization of open data
4. Will not compromise right to privacy, citizens have right to influence collection & use of personal data



2. Timely and Comprehensive

1. Human & technical resources to identify data for release/publication
2. Importance of consultation to prioritize data for release/improvement
 - including with citizens, government, civil society, private sector
1. To be of value data must be comprehensive, accurate, and high quality



3. Accessible and Useable

1. Opening up data enables governments, citizens, civil society, and private sector organizations to make more informed decisions
2. Open data should be discoverable and accessible, without bureaucratic or administrative barriers



4. Comparable and Interoperable

1. Data should be easy to compare within and between sectors, location, and time
2. Data should be presented in structured and standardized formats to support interoperability, traceability, and effective reuse



5. For Improved Governance and Citizen Engagement

1. Release of open data improves transparency, strengthens government, rule of law, and trust in public institutions
2. Improved civic engagement and better development, implementation, and assessment of programs and policies
3. Citizens, civil society, and private sector can help to highlight what data should be prioritized
4. City and local governments have a crucial role in citizen engagement



6. For Inclusive Development and Innovation

1. Importance of openness in stimulating creativity and innovation
2. Identify social and economic challenges, facilitating sustainable development
3. Open data is an equitable resource but can be limited by global digital divide
4. Government's role does not end with the release of data, it must take active role in understanding and reuse of data

Adoption Mechanism

- Aims to find a balance between achieving **high level political support/commitments**; while **avoiding burdensome entry barriers** that that could hinder adoption.
- The proposal is to ask for a **high-level public statement** (*Head of State, Secretary, Minister, Governor, Mayor, etc.*) **with the following characteristics**:
 1. **Statement of support towards the Open Data Charter** and its principles;
 2. **Appointment of key areas or individuals** responsible for following the work around the Open Data Charter; and
 3. **Outlining the mechanism(s) through which the principles of the Open Data Charter will be operationalized** by the country/city/local government (*ex. OGP National Action Plan, Open Data Action Plans, national/local Open Data implementation policies, etc.*)



Open Up Guides



- Key datasets
- How data could be collected, stored, shared and published
- Good data policies and frameworks, including metadata, standards and governance frameworks if available;
- Existing gaps or challenges in the policies and frameworks; and
- Use cases from real-life examples

Photo by John Schnobrich on Unsplash



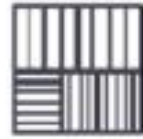
High value datasets

CO₂

Emissions
Related Data



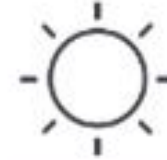
Agricultural Data



Land Use Land
Use Change and
Forestry Data



Electricity Data



Stationary
Energy Data



Transport Data



Waste Data



Natural Hazards
and Impacts Data



Socioeconomic
Data



Climate
Finance Data

[Complete Airtable](#)



Data quality- lessons learned from implementations

1

When possible, use open data standards from the start of the implementation

2

Disaggregation into multiple institutions at the time of data generation is one of the main challenges.

3

Data interoperability projects take a long time to be implemented

4

Measure and keep measuring-continuous data quality monitoring



Data quality- lessons learned from implementations

5

Establish metrics BEFORE measuring

6

Data quality dimensions: completeness, uniqueness, timeliness, validity, accuracy, consistency

7

Data interoperability + PEOPLE interoperability

8

Thematic implementations: expert discussions tend to go deeper faster



Thanks!!

Natalia Carfi

nati@opendatacharter.org

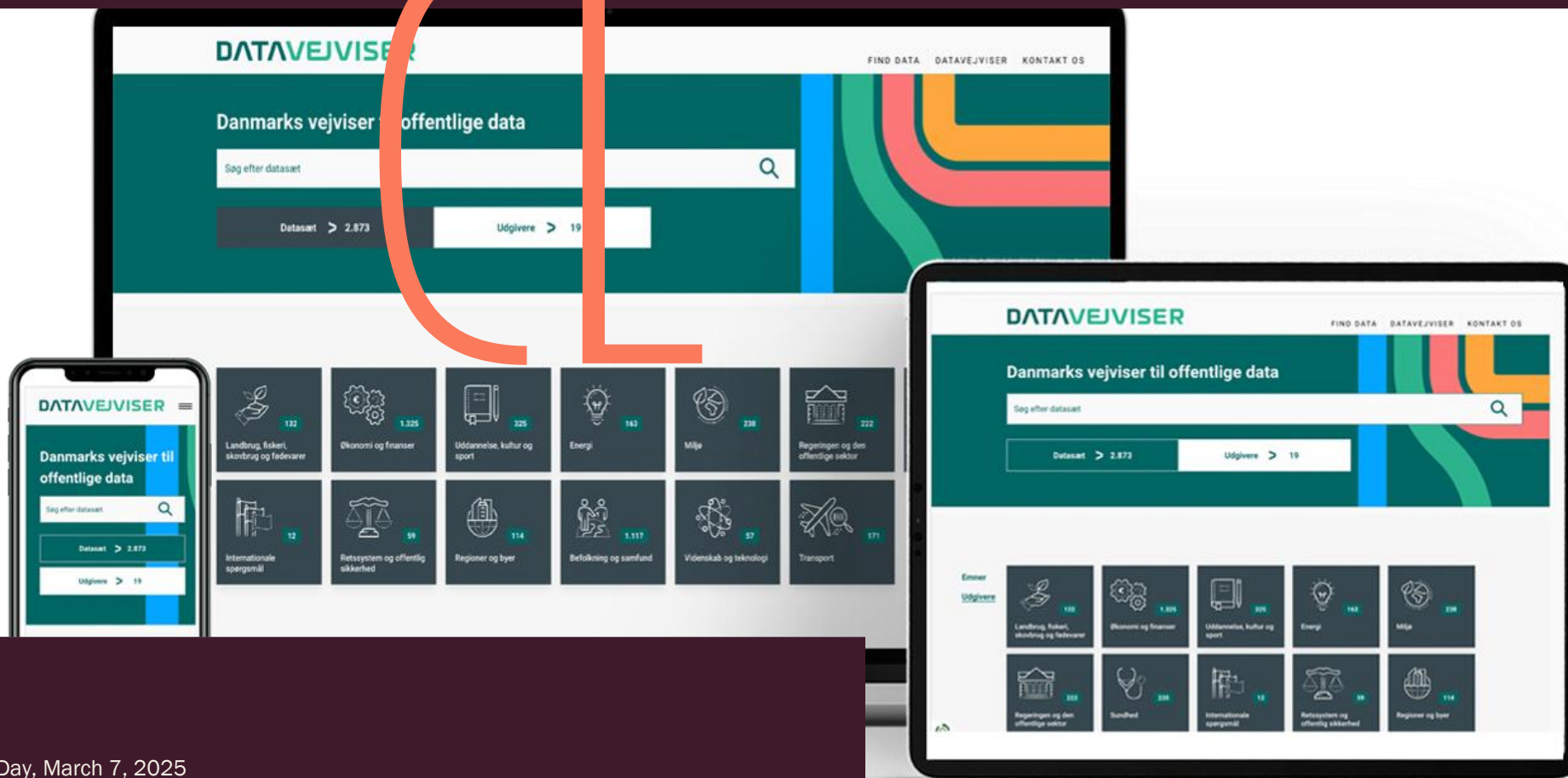
opendatacharter.com

@opendatacharter

Improving metadata quality on the Danish Data Portal

*Christian Nicolai Larsen
Policy Officer, Danish Agency for
Digital Government*



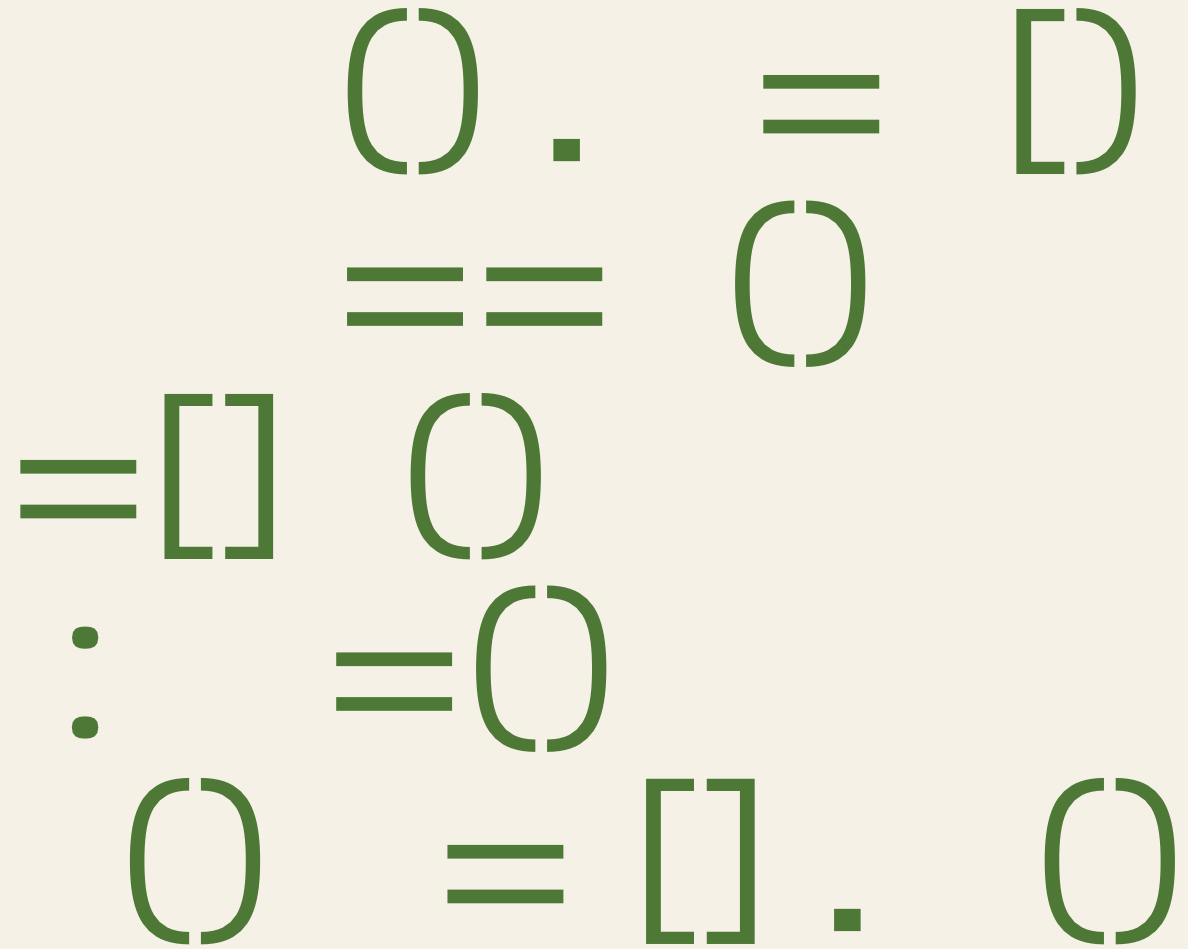


Open Data Day, March 7, 2025

Improving metadata quality on the Danish Data Portal

Agenda

- The relationship between data quality and metadata quality
- Introduction to Datavejviser, the Danish national data portal
- Metadata quality: How do we measure it, why should we worry about it, and how do we improve it?
- One way to make it easier for European public authorities to create and maintain high quality metadata



The relationship between data quality and metadata quality

- Is not additive:
 - Utility of data sharing = Data quality + metadata quality

The relationship between data quality and metadata quality

- ~~Is not additive:~~

- ~~Utility of data sharing = Data quality + metadata quality~~

- But multiplicative:

- Utility of data sharing = Data quality × metadata quality

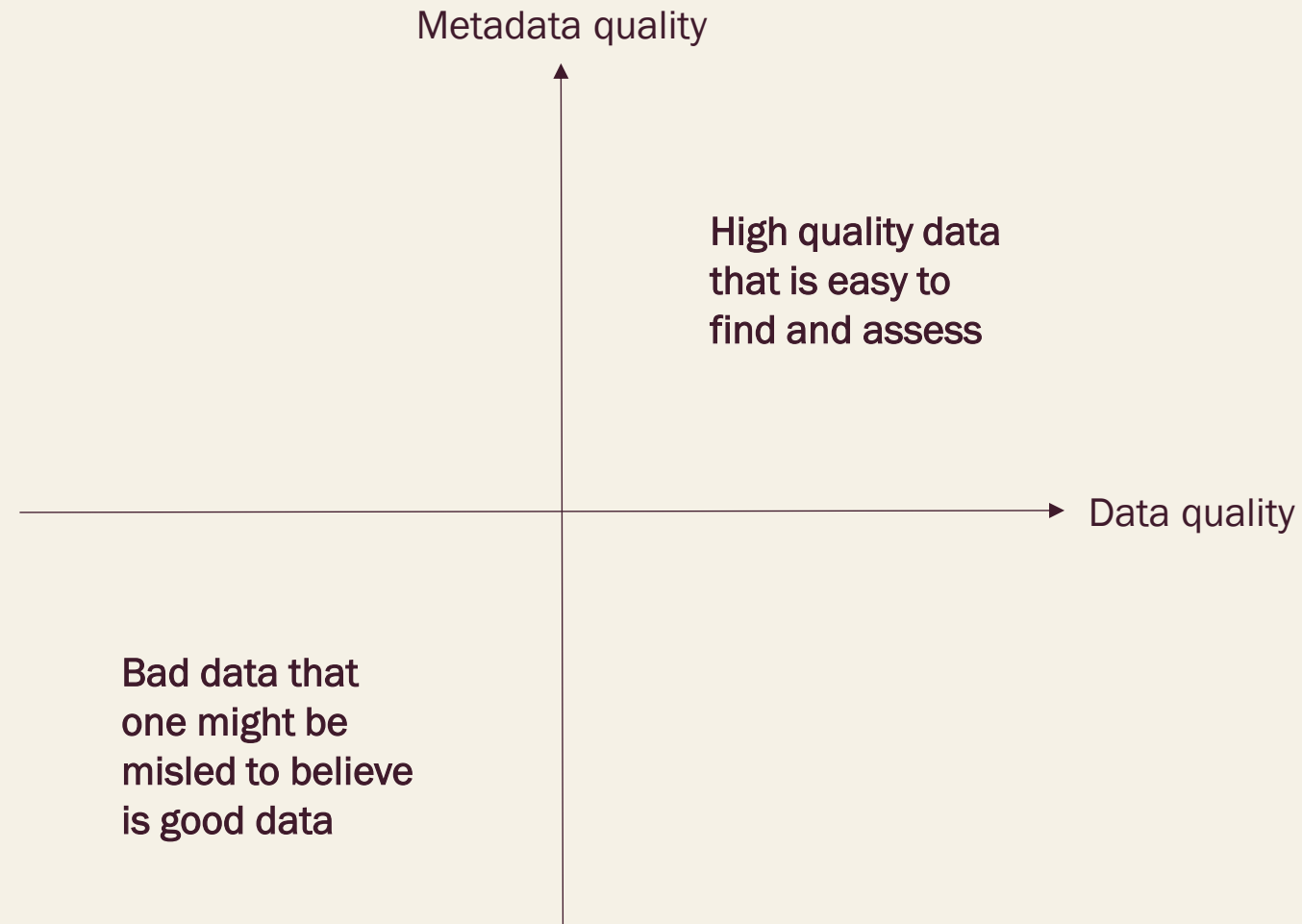
The relationship between data quality and metadata quality



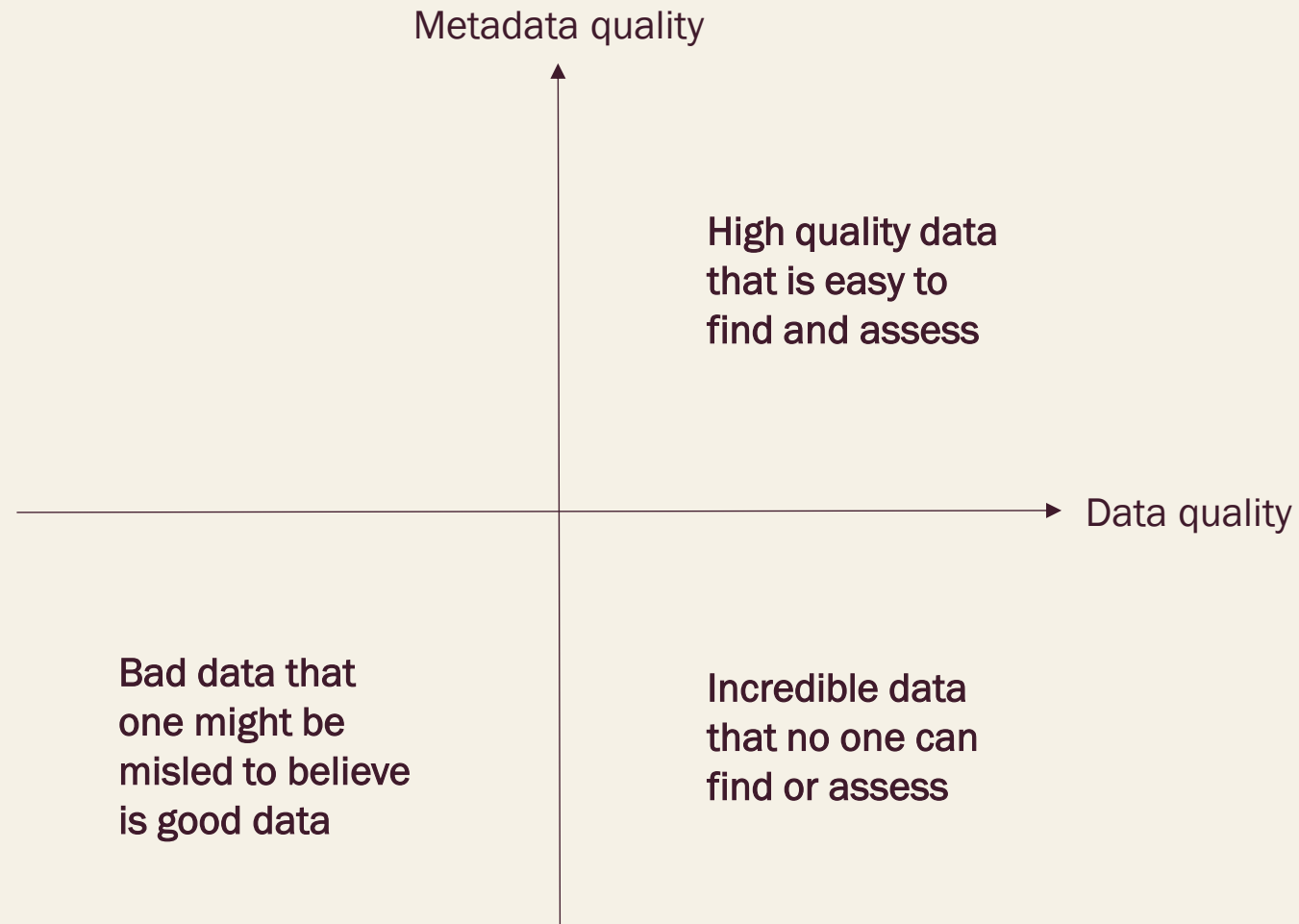
The relationship between data quality and metadata quality



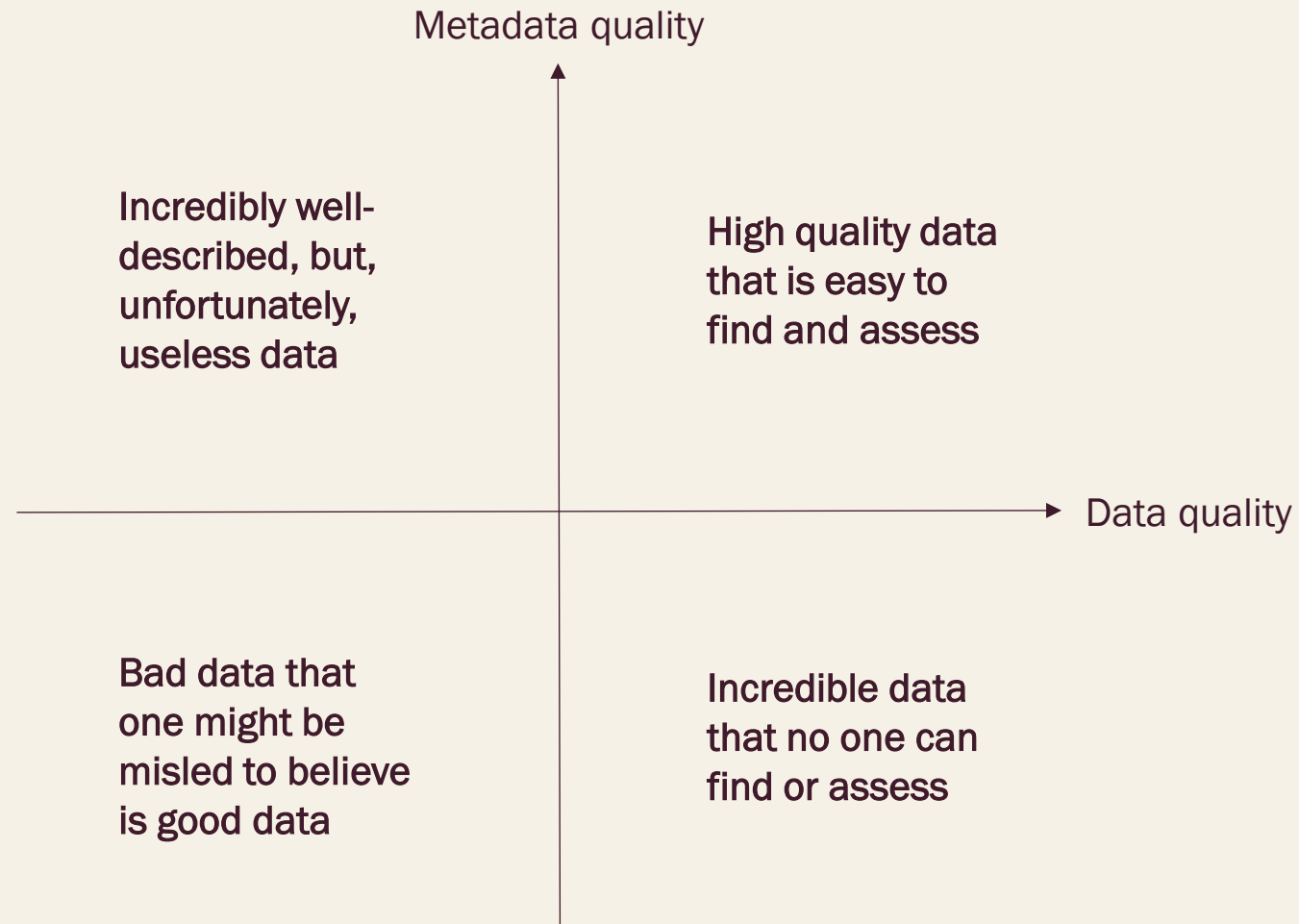
The relationship between data quality and metadata quality



The relationship between data quality and metadata quality




The relationship between data quality and metadata quality





Datavejviser is the single point of access for Danish public sector data


The screenshot shows the Datavejviser website interface. At the top left is the logo "DATAVEJVISER" in green. To the right are navigation links: "FIND DATA", "DATAVEJVISER", and "KONTAKT OS". The main heading is "Danmarks vejviser til offentlige data". Below this is a search bar with the placeholder text "Søg efter datasæt" and a magnifying glass icon. At the bottom of the search area, there are two filters: "Datasæt > 3.189" and "Udgivere > 52". The background features a dark teal color with a decorative graphic of overlapping curved lines in blue, green, red, and orange.


Why does Datavejviser focus more on metadata quality than data quality?


 We *do* also care about data quality and support efforts across the public sector that deal with this matter

 A number of recent projects in the Danish public sector have improved data quality and availability considerably

 The EU legislation that we are charged with implementing specifically stipulates the provision of metadata

 A good metadata catalogue is useful as an inventory to identify where further data quality improvements are needed

 Easier to generalise good practice than for data quality, which is very domain-specific and context-dependent

 To make matters more meta: We are also very interested in metadata *about* data quality, especially in light of the AI Act, EHDS act etc., which emphasize such metadata

CASE: Beierholm working with tax authority API



Accounting and consulting firm with ~2,000 employees on 42 locations around Denmark



Discovered API access to tax authority data using Datavejviser

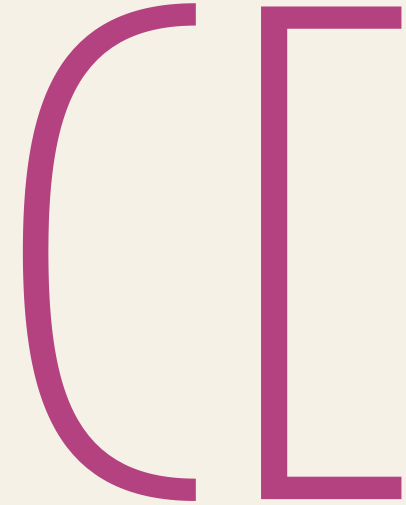
The APIs enable automated data acquisition vis-à-vis the status quo where customers' tax data used for accounting purposes were imported manually



Time saved: Expected to free up 8,200 employee-hours per year when fully implemented



The data quality of the product was improved by using automatically collected rather than manually transferred data



Many data providers and catalogues contribute metadata to Datavejviser, which is then forwarded to DEU and others



Datavejviser basics and metadata sources

- Metadata catalogue – no storage or distribution of actual data
- Data model: DCAT-AP-DK
- 3,000+ datasets from 50+ public sector publishers
- Catalogue for ODD and DGA data (and supports federation of metadata from other legislation, e.g. ITS/NAP, EHDS, etc.)
- CKAN backend (harvesting, metadata storage, rdf endpoint), React frontend
- Once-only entry of metadata for publishers
- Automatic harvest and re-use of metadata with as few manual/proprietary fixes as possible

Datavejviser basics and metadata sources

- Metadata catalogue – no storage or distribution of actual data
- Data model: DCAT-AP-DK
- 3,000+ datasets from 50+ public sector publishers
- Catalogue for ODD and DGA data (and supports federation of metadata from other legislation, e.g. ITS/NAP, EHDS, etc.)
- CKAN backend (harvesting, metadata storage, rdf endpoint), React frontend
- Once-only entry of metadata for publishers
- Automatic harvest and re-use of metadata with as few manual/proprietary fixes as possible

1. Multi-publisher domain catalogues (e.g. language data, local/regional data, geodata)

 Sprogteknologi.dk



 OPEN DATA DK

Datavejviser basics and metadata sources

- Metadata catalogue – no storage or distribution of actual data
- Data model: DCAT-AP-DK
- 3,000+ datasets from 50+ public sector publishers
- Catalogue for ODD and DGA data (and supports federation of metadata from other legislation, e.g. ITS/NAP, EHDS, etc.)
- CKAN backend (harvesting, metadata storage, rdf endpoint), React frontend
- Once-only entry of metadata for publishers
- Automatic harvest and re-use of metadata with as few manual/proprietary fixes as possible

1. Multi-publisher domain catalogues (e.g. language data, local/regional data, geodata)

 Sprogteknologi.dk



 OPEN DATA DK

2. Single-publisher endpoints (e.g. employment data, national archives, energy data)

Jobindsats



Rigsarkivet

Datavejviser basics and metadata sources

- Metadata catalogue – no storage or distribution of actual data
- Data model: DCAT-AP-DK
- 3,000+ datasets from 50+ public sector publishers
- Catalogue for ODD and DGA data (and supports federation of metadata from other legislation, e.g. ITS/NAP, EHDS, etc.)
- CKAN backend (harvesting, metadata storage, rdf endpoint), React frontend
- Once-only entry of metadata for publishers
- Automatic harvest and re-use of metadata with as few manual/proprietary fixes as possible

3. Single-publisher metadata entered into dedicated metadata entry portal (misc.)



Klimarådet.

VisitDenmark 

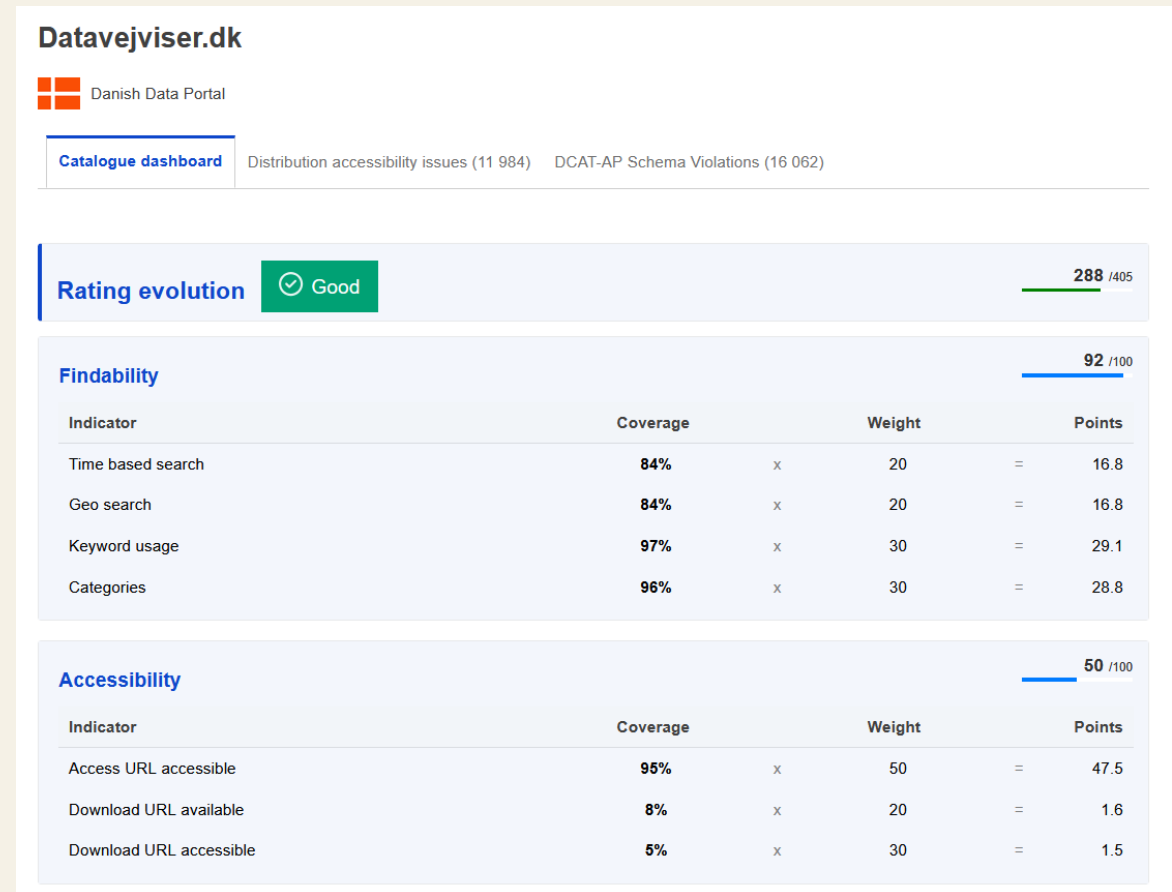


DANSKE
UNIVERSITETER

UNIVERSITIES DENMARK

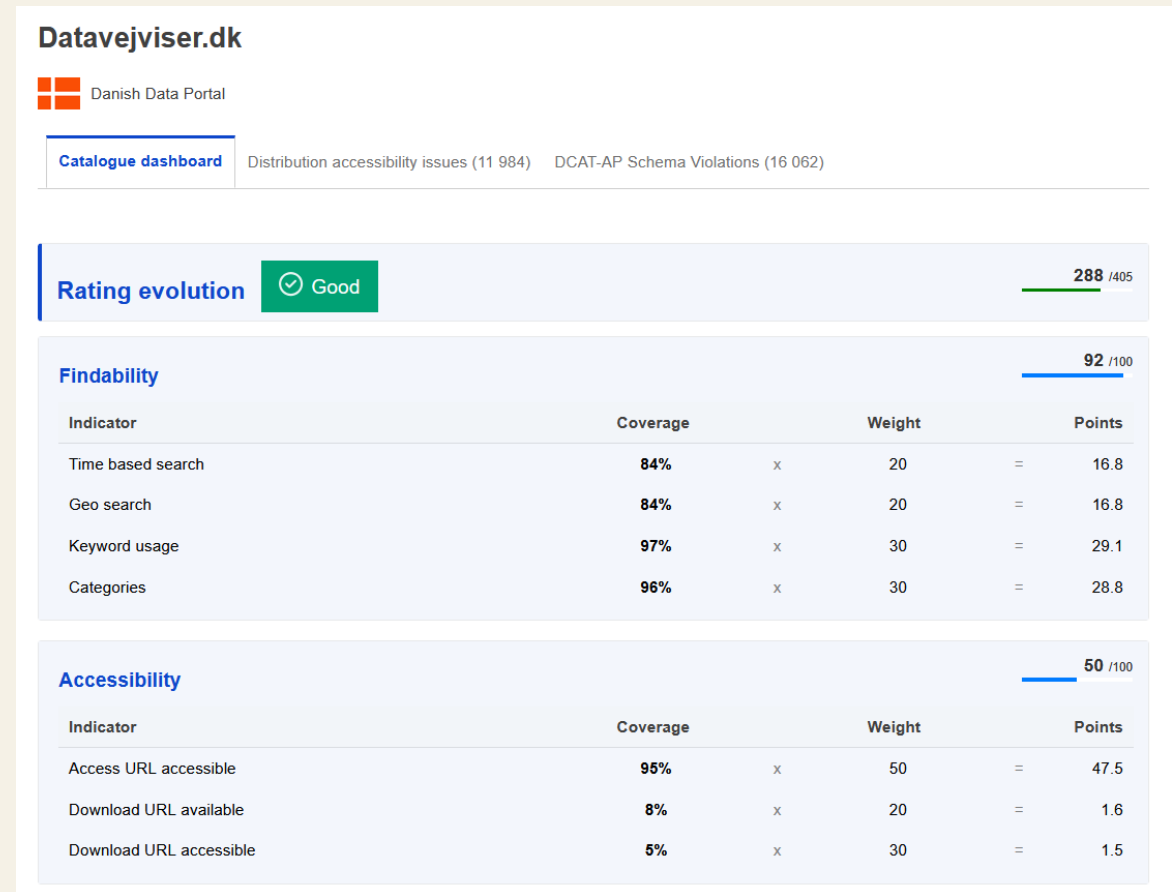
Metadata quality: What is it, and how do we measure it?

- "Quantitative" criteria and fixes
 - Automated checks for dead or inaccessible links
 - DCAT-AP conformance as measured through SHACL validation, DEU metadata quality dashboard (MQA), etc.
- Advantages:
 - Makes it transparent for publishers what they are being scored on
 - Incentivises actual improvement of the metadata at the national portals and data.europa.eu, which is the whole point



Metadata quality: What is it, and how do we measure it?

- "Quantitative" criteria and fixes
 - Automated checks for dead or inaccessible links
 - DCAT-AP conformance as measured through SHACL validation, DEU metadata quality dashboard (MQA), etc.
- Limitations:
 - Maximalist approach: Measures whether a dataset has, e.g., an associated keyword, but not whether that keyword is helpful for the user.
 - All datasets are weighted the same, although key registries may only amount to a few sets of metadata, the improvement of which then doesn't count for much



Metadata quality: What is it, and how do we measure it?

Storage Expected Availability

The expected and actual availability of injection and withdrawal capacity shown as percentage of the total technical capacity.

The expected injection and withdrawal capacity is calculated one month in advance.
The actual capacity is calculated every morning for the current gasday...

Mere ▾

Destinationsside ⓘ
<https://www.energidataservice.dk/gas-storage-denmark/StorageAvailablePct>

Udgivelsesdato ⓘ 2020-01-01	Ændringsdato ⓘ 2024-10-17	Opdateringsfrekvens ⓘ Dagligt	Dækningsperiode ⓘ 2015-05-01 -
---------------------------------------	-------------------------------------	---	--

Distribution ⓘ

API	JSON	Adgangs-URL	+
Filudtræk	CSV	Download-URL	+
Filudtræk	JSON	Download-URL	+
Websidevisning	HTML	Adgangs-URL	+

Udgiver ⓘ **ENERGINET**

Energinet
<https://energinet.dk/>

På Energi Data Service udstiller Energinet offentligt tilgængelige energidata til fri afbenyttelse. Energinet er en selvstændig offentlig virksomhed under Klima-, Energi- og Forsyningsministeriet. Energinet ejer og udvikler el- og gasnet i Danmark for at indpasse mere vedvarende energi, opretholde forsyningsikkerhed og sikre lige markedsadgang til nettene.

Se alle udgivers datasæt

Datasætkontakter

Kontaktpunkt
Energinet
✉ energidata@energinet.dk

Skaber ⓘ
Energinet

- "Qualitative" criteria and fixes – a necessary supplement
- Are the textual metadata elements, e.g. descriptions and provenance statements, well-written and reflective of the actual content of the data?
- Is the user experience/journey intuitive and helpful?
- Are the listed publisher name and description up to date? Political context entails frequent reorganization of the public sector
- Limitations: Time-consuming, somewhat arbitrary

What do we do to improve metadata quality on Datavejviser?

- Help authorities navigate DCAT-AP and its dialects (MobilityDCAT-AP, GeoDCAT-AP, etc.).
- Make the most of opportunities presented by funding rounds, user demand, alignment with strategies for digitisation, and legal requirements to carry out wider metadata improvements in collaboration with publishers
- Make metadata DCAT-AP compliant as close to the source as possible – help and fund authorities building endpoints, offer metadata entry portal which is DCAT-AP compliant by design
- Highlight datasets that contain identifiers/keys that make them interoperable with other datasets (not yet fully implemented)
- Translate the CC license suite into Danish, work the licenses into the OP authority lists, reach a consensus across Danish government to recommend their use for open data

What do we do to improve metadata quality on Datavejviser?

- Help authorities navigate DCAT-AP and its dialects (MobilityDCAT-AP, GeoDCAT-AP, etc.).
- Make the most of opportunities presented by funding rounds, user demand, alignment with strategies for digitisation, and legal requirements to carry out wider metadata improvements in collaboration with publishers
- Make metadata DCAT-AP compliant as close to the source as possible – help and fund authorities building endpoints, offer metadata entry portal which is DCAT-AP compliant by design
- Highlight datasets that contain identifiers/keys that make them interoperable with other datasets (not yet fully implemented)
- Translate the CC license suite into Danish, work the licenses into the OP authority lists, reach a consensus across Danish government to recommend their use for open data

What do we do to improve metadata quality on Datavejviser?

- Help authorities navigate DCAT-AP and its dialects (MobilityDCAT-AP, GeoDCAT-AP, etc.).

mobilityDCAT-AP - Version 1.1.0

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

NAPCORE Recommendation 17 January 2025



GeoDCAT-AP 3.0.0

04 October 2024

▼ More details about this document

Latest published version:

<https://semiceu.github.io/GeoDCAT-AP/releases/>



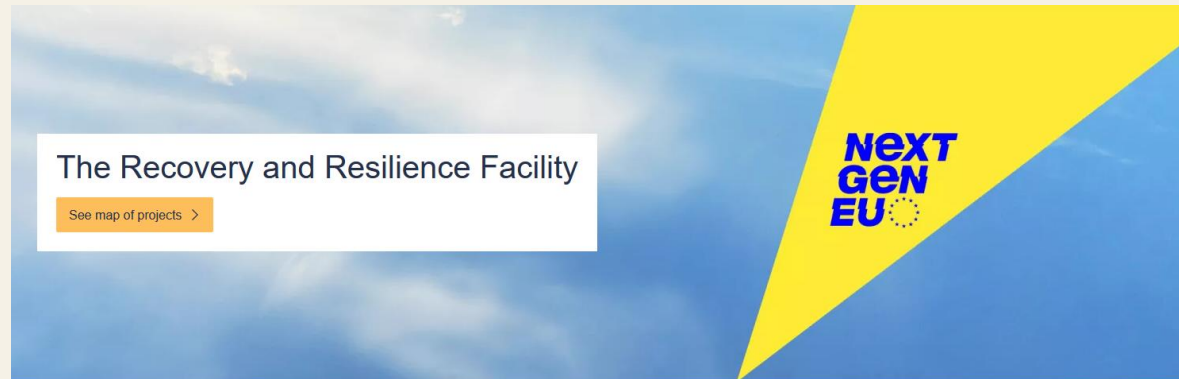
HealthDCAT-AP

Unofficial Draft 22 December 2023



What do we do to improve metadata quality on Datavejviser?

- Make the most of opportunities presented by funding rounds, user demand, alignment with strategies for digitisation, and legal requirements to carry out wider metadata improvements in collaboration with publishers



What do we do to improve metadata quality on Datavejviser?

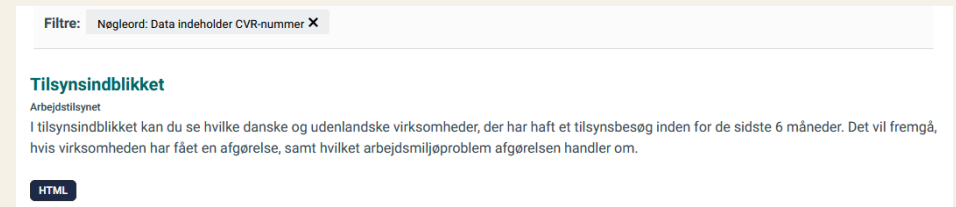
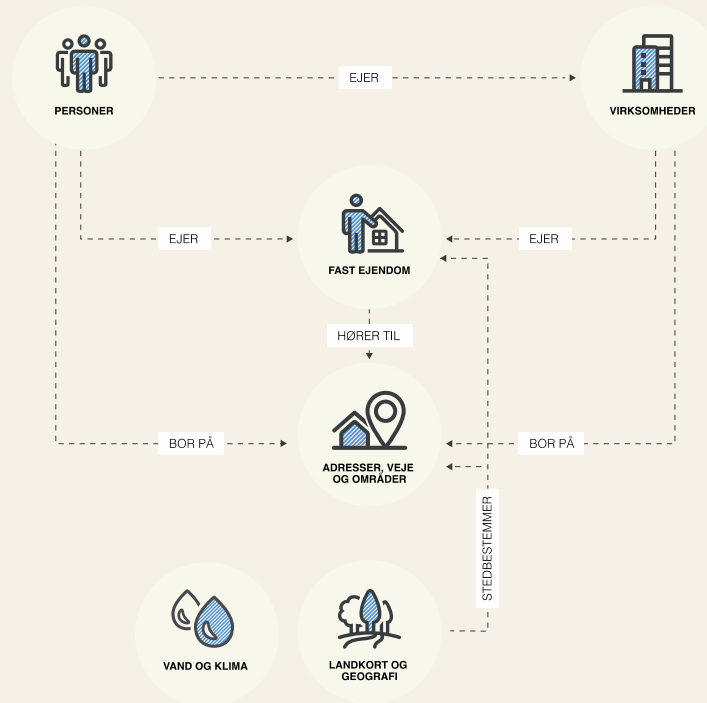
- Make metadata as DCAT-AP compliant as close to the source as possible – help and fund authorities building endpoints, offer metadata entry portal which is DCAT-AP compliant by design

The screenshot shows the 'Redigér' (Edit) page for the 'DEU test dataset'. The breadcrumb trail is 'Udgivere / Test / DEU test dataset / Redigér'. At the top, there are tabs for 'Redigér metadata' (selected) and 'Ressourcer', and a 'Se datasæt' button. A progress bar below the tabs has six steps: 1. Basisinformation (highlighted in green), 2. Tidslige karakteristika, 3. Emner & nøgleord, 4. Akterer, 5. Villkår, and 6. Øvrige. The 'Basisinformation' section contains the following text: 'Datasættets basisinformation er titel, beskrivelse, destinationsside og udgiver. Alle er esentielle for at anvendere kan finde og vurdere datasættet og understøtter brugerrejsen videre fra Datavejviser. Det er obligatorisk at angive titel, beskrivelse samt udgiver og anbefalet at angive en destinationsside.' Below this is a form for the title, with the current value 'DEU test dataset' and a Danish flag icon. A tooltip explains: 'En god titel giver en bred gruppe anvendere præcis information om datasættets indhold og formål, og den er specifik, så det er nemt at skelne datasættet fra andre datasæt med lignende emner.' There is a '+ Tilføj engelsk titel' link and a 'Læs mere' button. The URL field shows 'URL: datavejviser-indtastning.digst.govcloud.dk/dataset/deu-test-dataset' with a 'Redigér' button. The description field has a note: 'Du kan bruge Markdown formatering her' and a text area containing 'A description of the dataset'.

What do we do to improve metadata quality on Datavejviser?

- Highlight datasets that contain identifiers/keys that make them interoperable with other datasets (not yet fully implemented)

- Shared data model for basic data ensures interoperability between base registries
- Other datasets using the same shared keys (e.g. Central Business Register number) allows data to be combined – if users know this possibility exists. Thus, data and metadata quality complement each other.



- Full implementation would require use of a formalised catalogue of terms/keys.

What do we do to improve metadata quality on Datavejviser?

- Translate the CC license suite into Danish, work the translations into the OP authority lists, reach a consensus across Danish government to recommend their use for open data

```
<skos:definition xml:lang="fr">
  CC BY 4.0 permet aux autres de distribuer, remix, adapter l'œuvre de l'auteur,
  diffusion et l'utilisation maximales d'œuvres licenciées sous CC.
</skos:definition>
<skos:definition xml:lang="da">
  Denne licens tillader andre at videredistribuere, omarbejde, tilpasse og bygge videre på materialet,
  videreformidling og brug af det licenserede materiale.
</skos:definition>
<skos:definition xml:lang="en">
  CC BY 4.0 lets others distribute, remix, tweak, and build upon the author's work, even
  </skos:definition>
<skos:exactMatch rdf:resource="https://spdx.org/licenses/CC-BY-4.0.html"/>
<skos:exactMatch rdf:resource="http://creativecommons.org/licenses/by/4.0"/>
<dc:identifier>CC_BY_4_0</dc:identifier>
```

Licens og brugsvilkår for offentlige data

Digitaliseringsstyrelsen anbefaler, at danske myndigheder tilknytter deres åbne data en standardlicens fra Creative Commons (CC), fortrinsvis CC0, medmindre særlige forhold taler imod det.

Her kan du læse om, hvad det betyder, og hvordan du i praksis bærer dig ad med at følge vores anbefaling.

One way to make it easier to create and maintain high quality metadata: Tying together legal, semantic, and technical aspects



Etc.

One way to make it easier to create and maintain high quality metadata: Tying together legal, semantic, and technical aspects




GeoDCAT-AP 3.0.0
04 October 2024



▼ More details about this document
Latest published version:
<https://semiceu.github.io/GeoDCAT-AP/releases/>




mobilityDCAT-AP - Version 1.1.0
A mobility extension for the DCAT application profile for data portals in Europe



NAPCORE Recommendation 17 January 2025



HealthDCAT-AP
Unofficial Draft 22 December 2023



Etc.

One way to make it easier to create and maintain high quality metadata: Tying together legal, semantic, and technical aspects




GeoDCAT-AP 3.0.0
04 October 2024



▼ More details about this document
Latest published version:
<https://semiceu.github.io/GeoDCAT-AP/releases/>




mobilityDCAT-AP - Version 1.1.0
A mobility extension for the DCAT application profile for data portals in Europe



NAPCORE Recommendation 17 January 2025



HealthDCAT-AP
Unofficial Draft 22 December 2023



Etc.




One way to make it easier to create and maintain high quality metadata: Tying together legal, semantic, and technical aspects




Etc.

GeoDCAT-AP 3.0.0
04 October 2024




▼ More details about this document
Latest published version:
<https://semiceu.github.io/GeoDCAT-AP/releases/>

mobilityDCAT-AP - Version 1.1.0
A mobility extension for the DCAT application profile for data portals in Europe



NAPCORE Recommendation 17 January 2025

HealthDCAT-AP
Unofficial Draft 22 December 2023



Thank you for listening!

DATAVEJVISER

A searchable catalogue of **metadata** describing available data from the public sector in Denmark. Datavejviser supplements and supports existing data distributions and makes it easier to find and validate data from various public authorities.

Datavejviser aims to:

- Increase findability of public data
- Gather and analyze demands for more data
- Contribute to enhance the availability of public data

Datavejviser is operated by the Danish Agency for Digital Government and can be visited at datavejviser.dk

Contact us at kontakt@datavejviser.dk



Finansieret af
Den Europæiske Union
NextGenerationEU



Digitaliseringsstyrelsen

Data and metadata quality by the data.europa.eu



Data quality guidelines

Metadata quality assurance



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

Home Data EU Open Data Days Academy Community Publications Documentation

Home > Metadata Quality Dashboard > Dashboard

Metadata quality

Dashboard Catalogues Methodology Download as report

At the moment we are experiencing problems with the calculation of the metadata quality.

Overview

Catalogues: Top 12

Dimensions

The Metadata Quality Assurance is intended to help data providers and data portals to check which metrics we use for indicator measurements, please have a look at our methodology

This page provides a general overview of all metadata available to data.europa.eu. For a more detailed view, please visit our Catalogues page and select the desired catalogue.

Rating evolution Sufficient

Findability

Country	Name	Findability 100 Points	Accessibility 100 Points	Interoperability 110 Points	Reusability 75 Points	Contextuality 20 Points	Rating 405 Points
	SALTED Project (IOR)	100 / 100	100 / 100	80 / 110	65 / 75	15 / 20	360 / 405 Excellent
	Czech National Open Data Portal (CZE)	97 / 100	92 / 100	101 / 110	62 / 75	0 / 20	352 / 405 Excellent
	Executive Agency for Small and Medium-sized Enterprises (EUROPE)	100 / 100	100 / 100	60 / 110	75 / 75	15 / 20	350 / 405 Good
	European Political Strategy Centre (EU)	80 / 100	100 / 100	80 / 110	75 / 75	5 / 20	340 / 405 Good
	dati.gov.it (ITA)	100 / 100	41 / 100	99 / 110	75 / 75	20 / 20	335 / 405 Good

Q&A



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



Renata Ávila
CEO,
Open Knowledge
Foundation



Romina Colman
Product Owner,
Open Data Editor



Natalia Carfi
CEO,
Open Data Charter



Christian Nicolai Larsen
Policy Officer,
Danish Agency for Digital
Government

Stay up-to-date on our
2025 activities!

The logo for Data Europa Academy is located in the bottom-left corner. It consists of a large orange circle containing a smaller dark blue circle. Inside the dark blue circle, the text "data.europa academy" is written in white, lowercase letters. The word "data" is on the top line, "europa" is on the second line, and "academy" is on the third line. Small orange dots are placed above the 'a' in "data" and above the 'o' in "europa".

data.
europa
academy

Register now for our next webinar!

WEBINAR

Smart cities and digital twin technology: the case of Dublin

data. europa academy

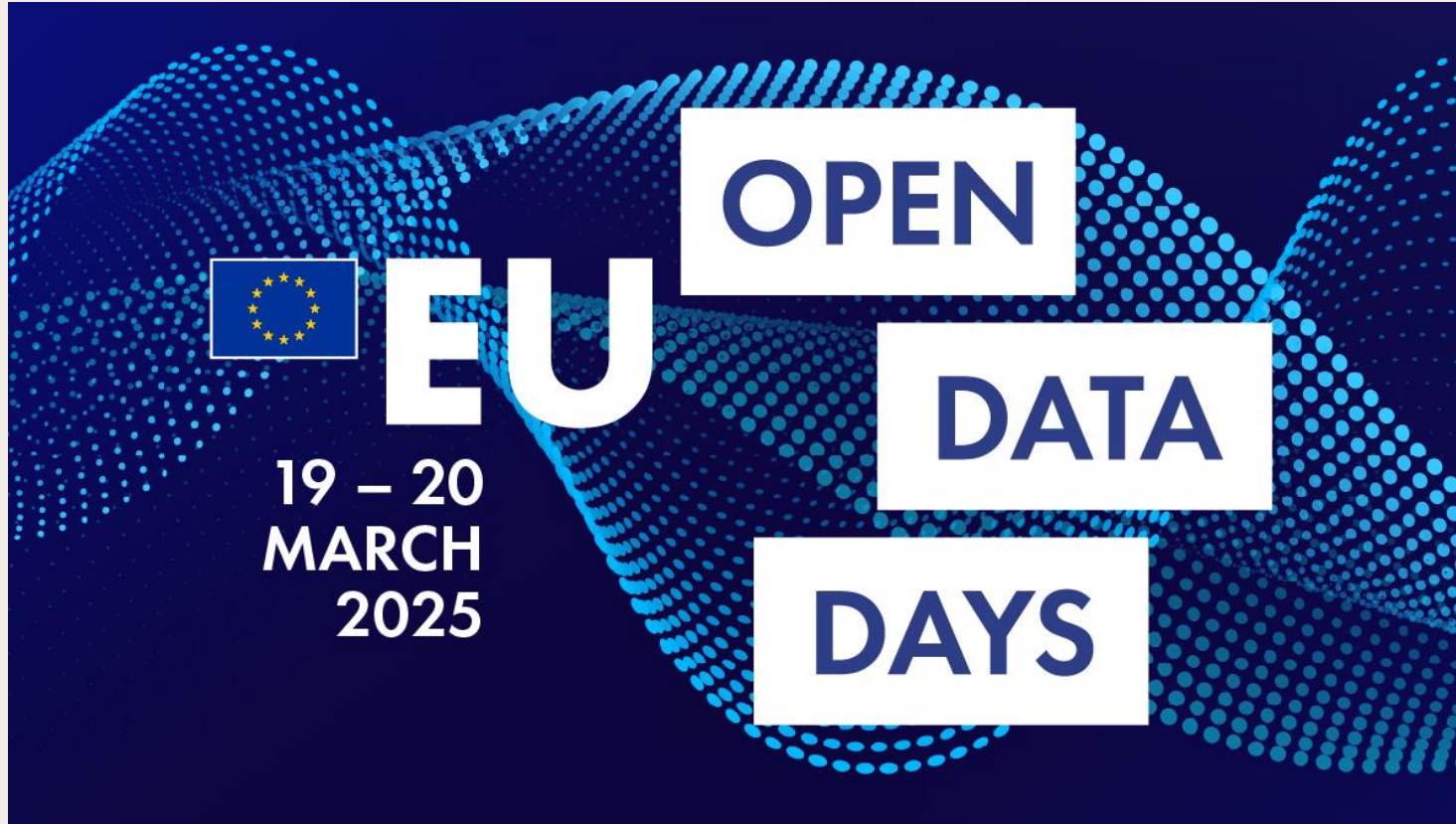
14 March 2025
10.00 – 11.00 CET

A graphic for a webinar poster. It features a large red circle on the right side. Inside this circle is a smaller white circle, which contains a colorful pie chart with segments in green, red, blue, and yellow. Two hands are shown interacting with the pie chart: one hand is pointing at a segment, and the other is holding a red slice. The background of the poster is dark blue.

data. europa academy



Join us online for the EU Open Data Days!



data.
europa
academy



Your opinion is important to us!



Thank you!

