WEBINAR

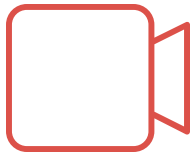
How to use open data for your research?

data.europa academy

19 October 2023

10.00 – 12.00 CET

Rules of the game



The workshop will be recorded

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

For questions, please use the ClickMeeting chat

data europa academy

Introduction

Louise van ‘t Erve
Communication and data.europa academy

Leon Mook
Data science expert

Dieuwertje Bloemen
University of Leuven

Leon van Wissen
University of Amsterdam

Maria Pia Di Buono
University of Naples L’Orientale

data europa academy

Agenda

10.00 – 10.05	Introduction to the workshop and rules of the game – *Louise van ‘t Erve*
10.05 – 10:15	Search data like a pro: how to use data.europa.eu– *Louise van ‘t Erve*
10.15 – 10.30	Data publication and re-use – *Dieuwertje Bloemen*
10.30 – 10.45	Cinema Context in (linked) data driven research – *Leon van Wissen*
10.45 – 11.00	From Data to Knowledge: Strategies for Research and Real-World Applications – *Maria Pia Di Buono*
11.00 – 11.05	Break
11.05 – 11.35	Case with open data – *Leon Mook*
11.35 – 12.00	Wrap up case and Q&A – *Louise van ‘t Erve*

data europa academy

Search data like a pro: how to use data.europa.eu

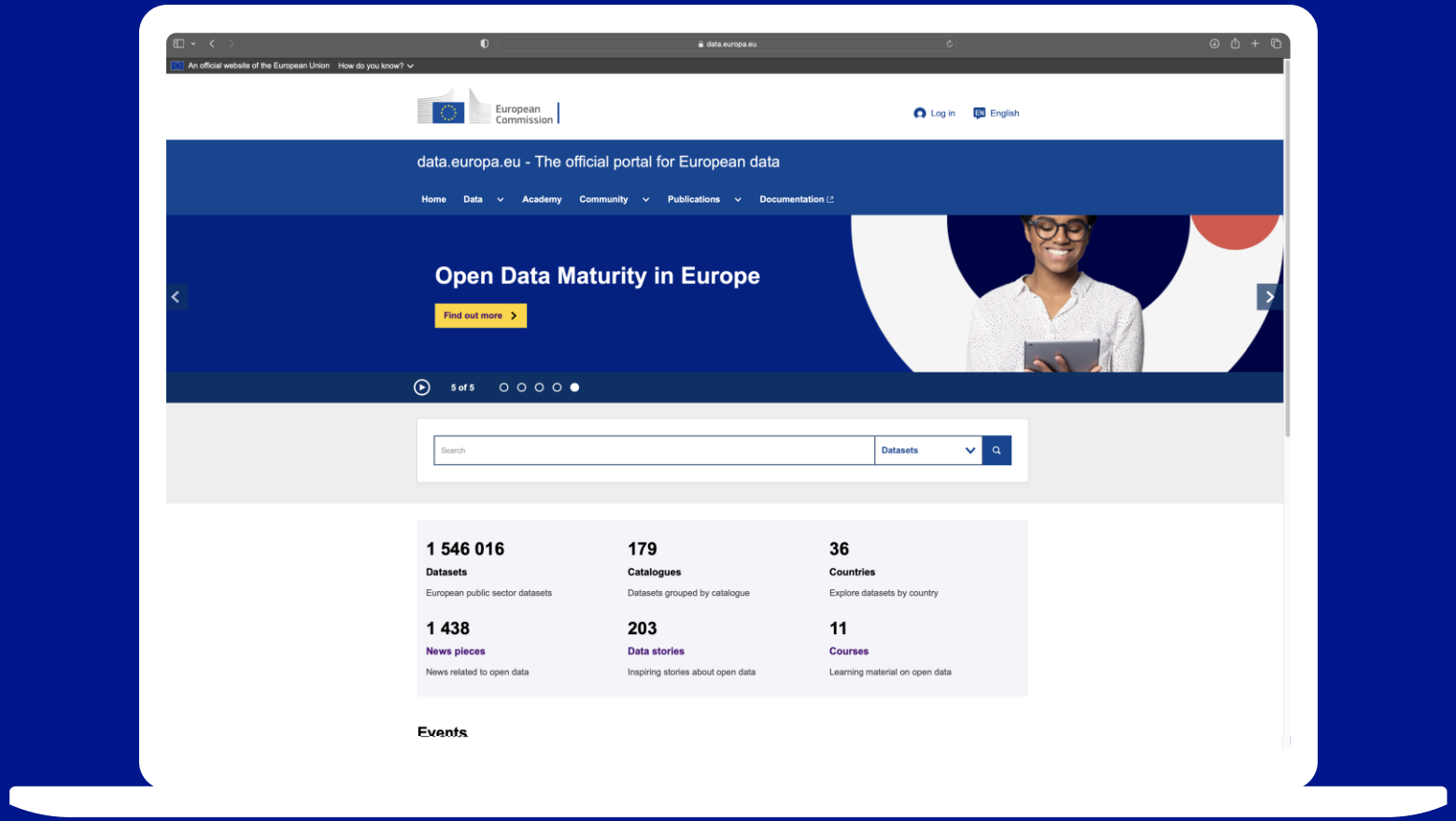
Louise van ‘t Erve

data europa academy

data.europa.eu

What is data.europa.eu?

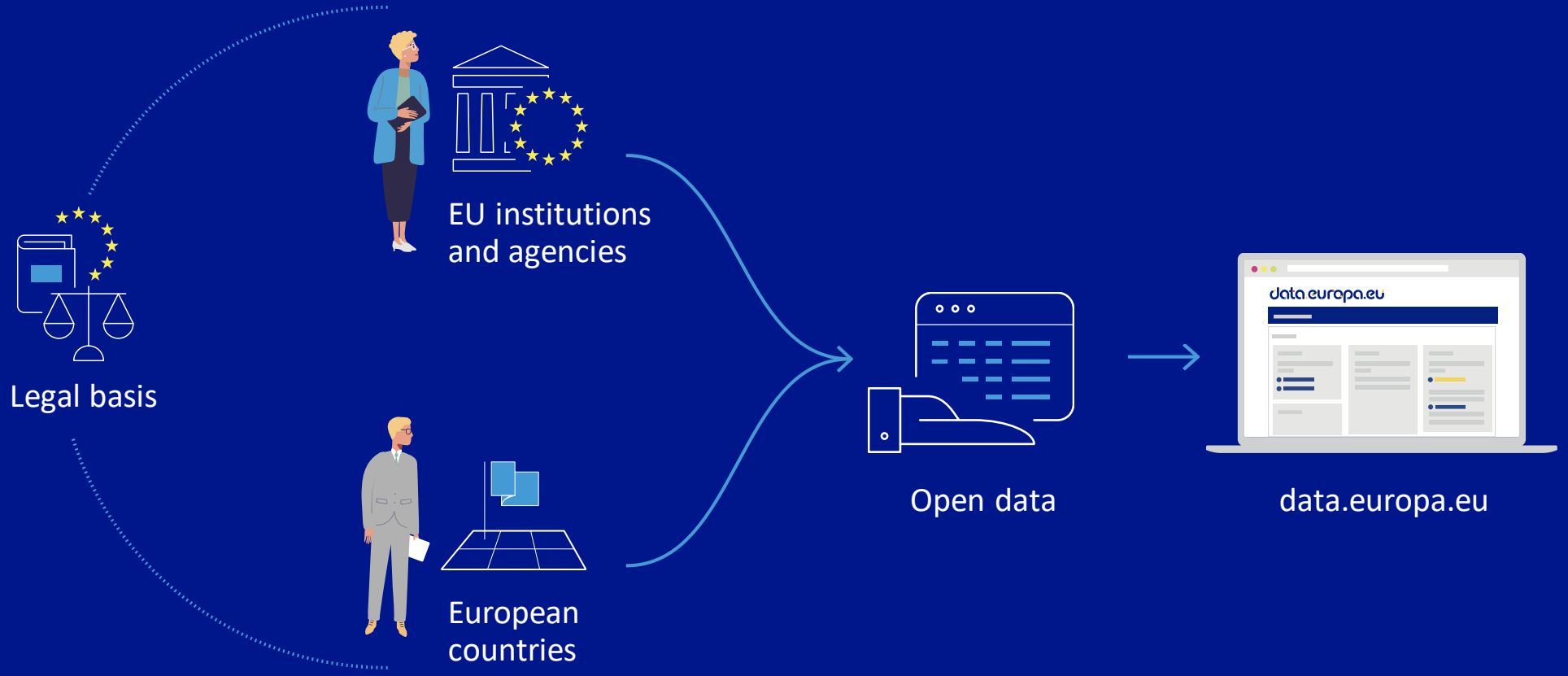
A platform providing central point of access to **open data** from international European Union, national, regional, local and geodata portals.



data.europa.eu

Who are publishers and why do they publish on the portal?

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.



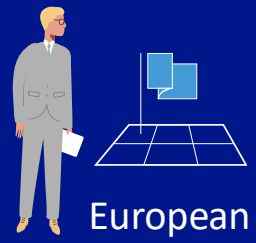
data.europa.eu

The portal is a bridge between the data providers and data users

Data providers

EU institutions and agencies

Geoportals



European countries

data.europa.eu

Data users

Developers

Researchers

Public administrations

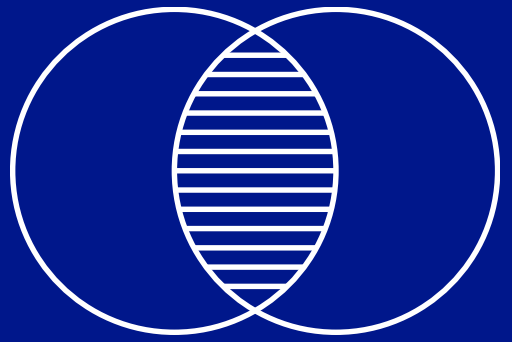
General public

Others

Private companies

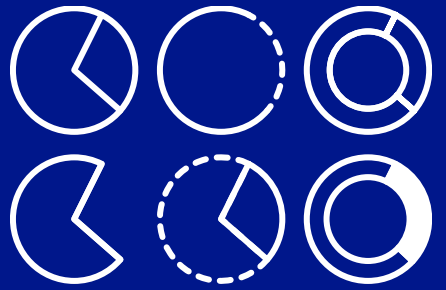
Fact journalism

data.europa.eu



Increasing transparency

Improving open data access



Boosting (re)use

data.europa.eu

Our services in a nutshell

Data

Providing access to free public data resources across Europe via a single platform (the portal).

Academy

Supporting the EU institutions, EU Member States and European countries to publish metadata of their open data in a harmonised manner.

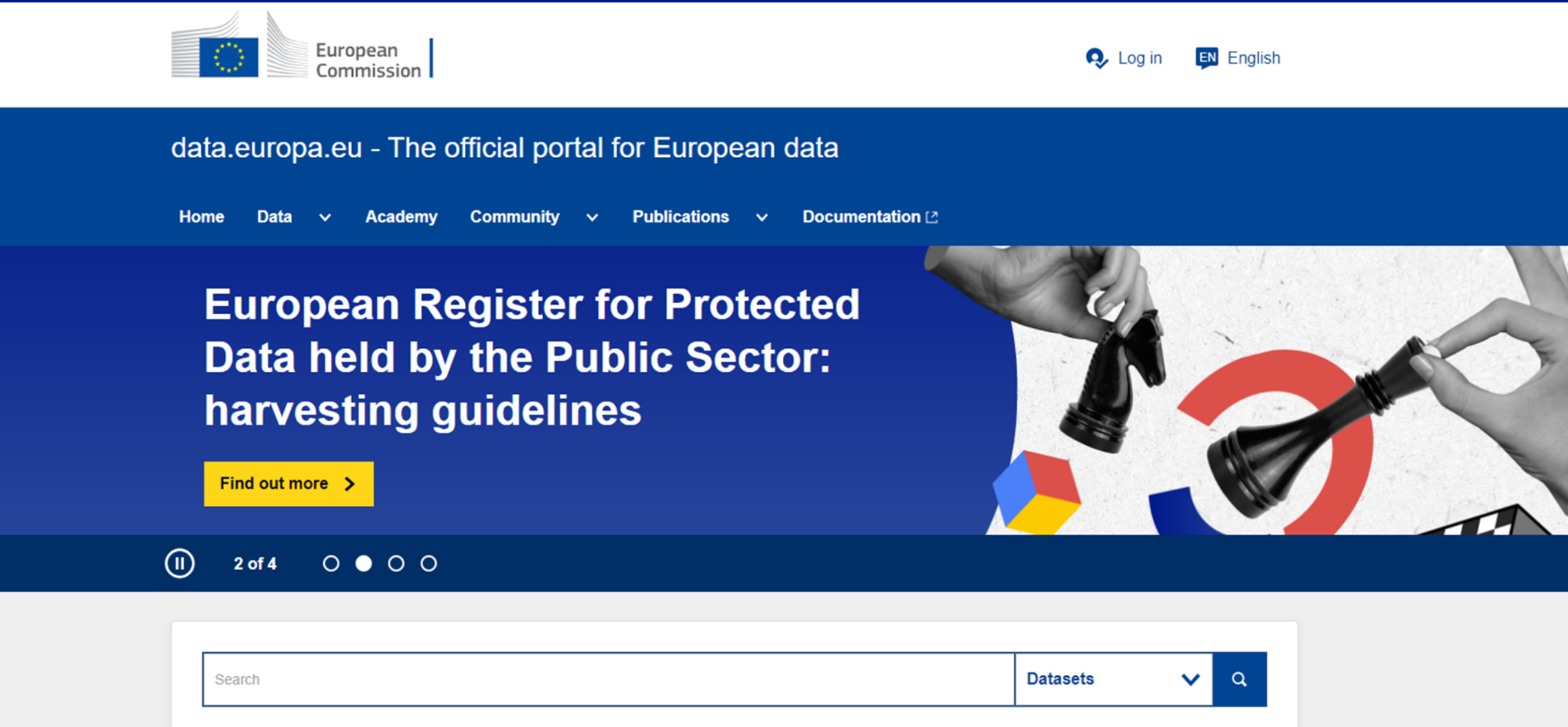
Community

Organising open data competitions and conferences; communicating via social media and newsletters.

Publications

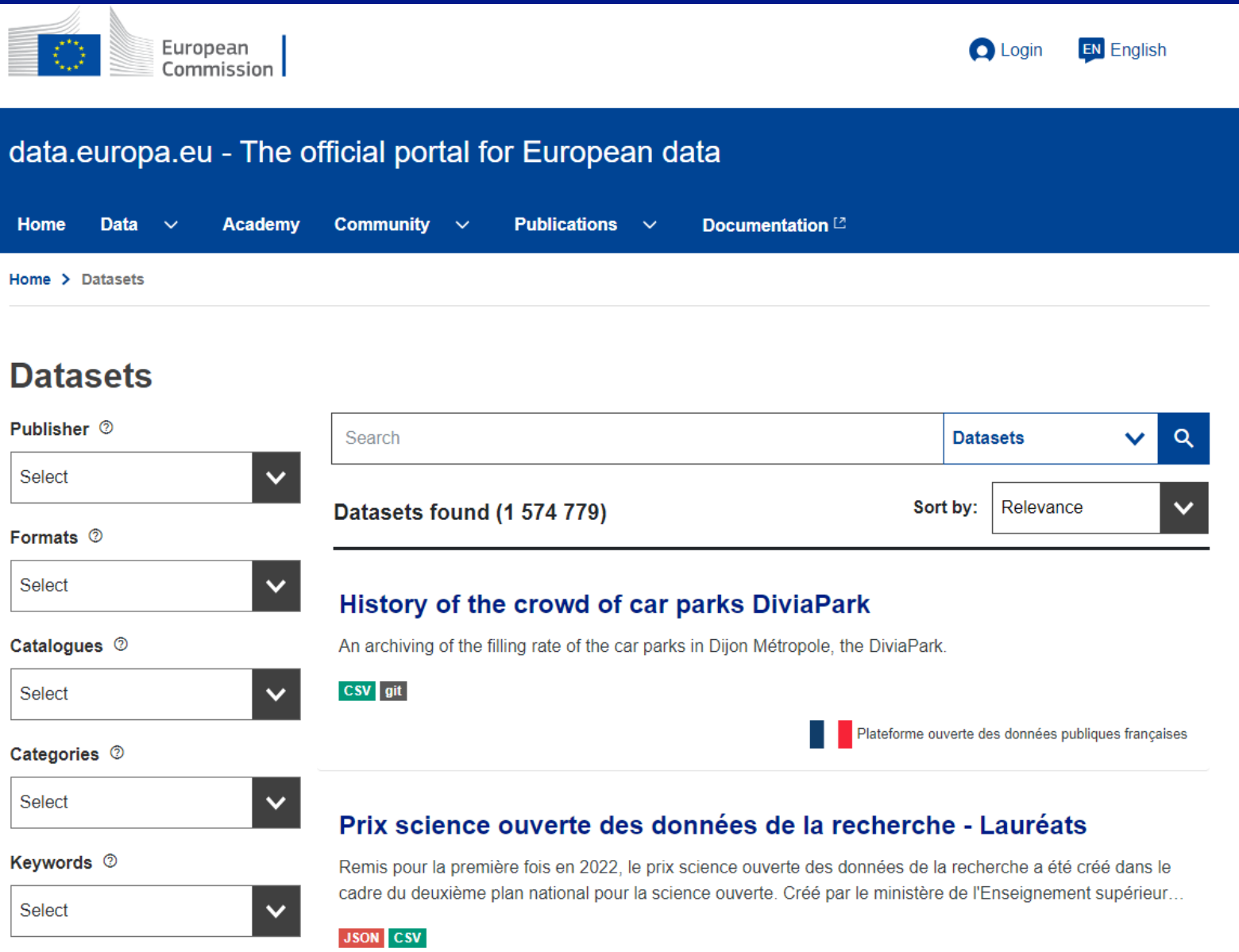
Assessing open data maturity in Europe; providing reports, studies and training via data.europa academy.

The website


European Commission
Log in
EN English
data.europa.eu - The official portal for European data
Home
Data
Academy
Community
Publications
Documentation
European Register for Protected Data held by the Public Sector: harvesting guidelines
Find out more
2 of 4
Search
Datasets

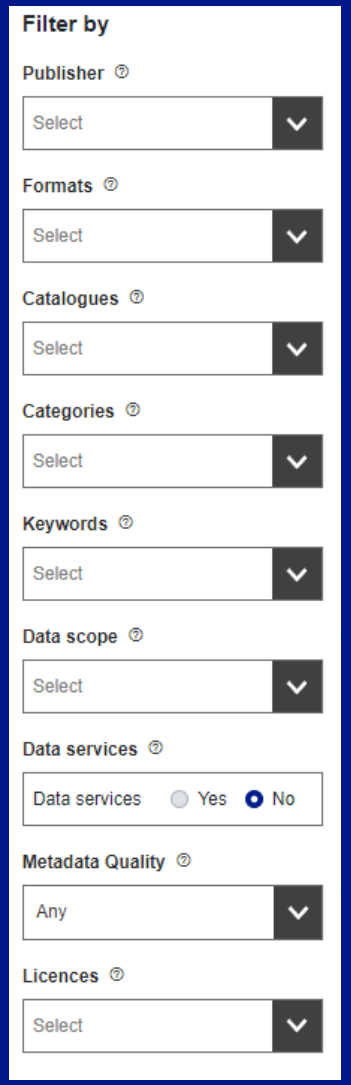
Where to find datasets

Where to find datasets



How to search - Filtering

These are tools that allow filtering of datasets according to various characteristics.



data europa academy

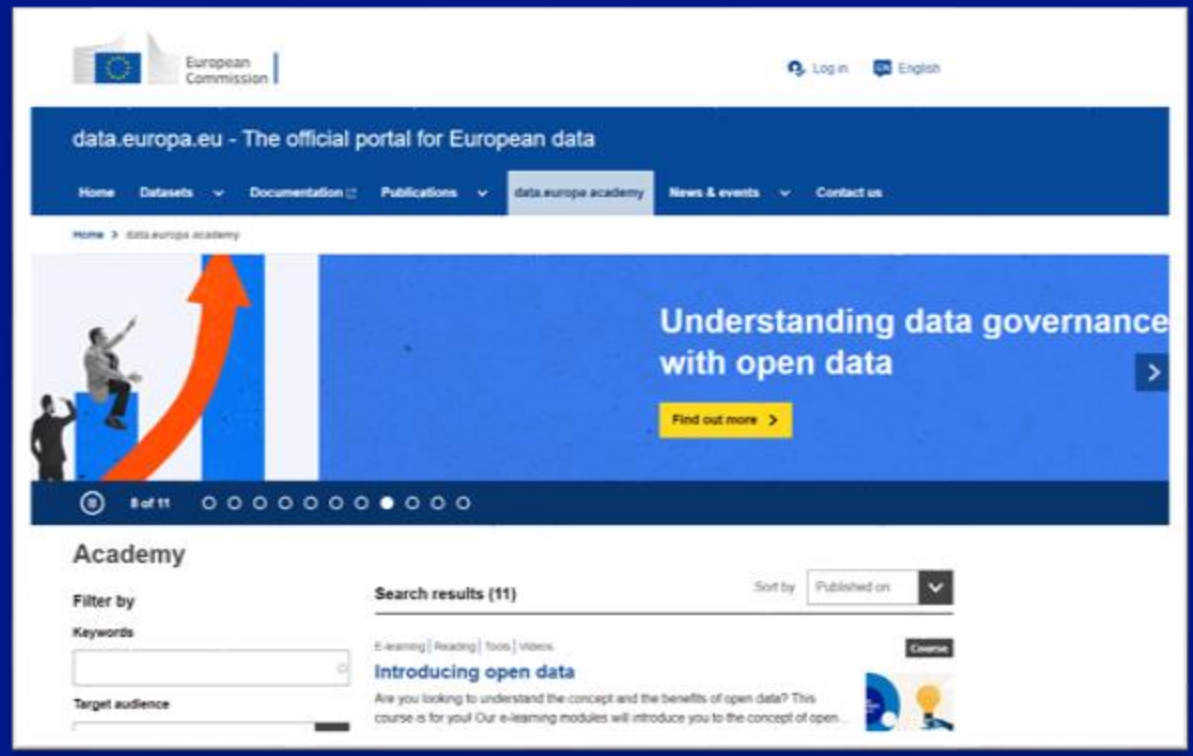
data.europa academy – your open data knowledge hub

Aim: improve **data literacy**

Content:

- Grouped in 11 thematic courses (legal, technical, business, data visualisation, …)
- Articles, videos, slide shows, reports, ...
- Designed according to **users' needs**

Constantly **updated and improved**



data.europa academy

Data publication and re-use

Dieuwertje Bloemen

data europa academy

KU LEUVEN LIBRARIES

LIBIS INTO INFO

Data publication & reuse

The cycle of finding, accessing, using and publishing data.

Data.Europa.eu webinar, 19/10/2023

Dieuwertje Bloemen, Productmanager RDR & Lirias (Dieuwertje.Bloemen@kuleuven.be)

About KU Leuven RDR

KU Leuven's institutional research data repository for the publication of 'finished' research data

Part of the KU Leuven RDM infrastructure and data life cycle

Tool for FAIR publication of research data produced by KU Leuven researchers and research

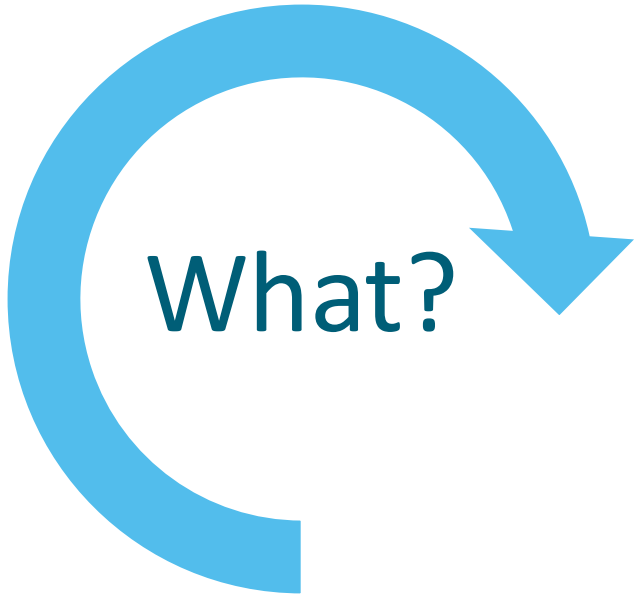
Managed by LIBIS (part of the KU Leuven Libraries)

Content

- What is data publication
- Finding data
- Data reuse

Data publication

What is data publication/data sharing



- Share research data **after** a research project ends

  At minimum share the data that is necessary to **replicate your research**

- Publish data to **support a publication**

- Publish data as a **stand-alone research output**

CC-BY-4.0

Where is data typically published/shared by researchers

Institutional data repositories

KU LEUVEN RESEARCH DATA REPOSITORY

DataverseNO

Domain-specific data repositories

EGA
European Genome-phenome Archive

PANGAEA.
Data Publisher for Earth & Environmental Science

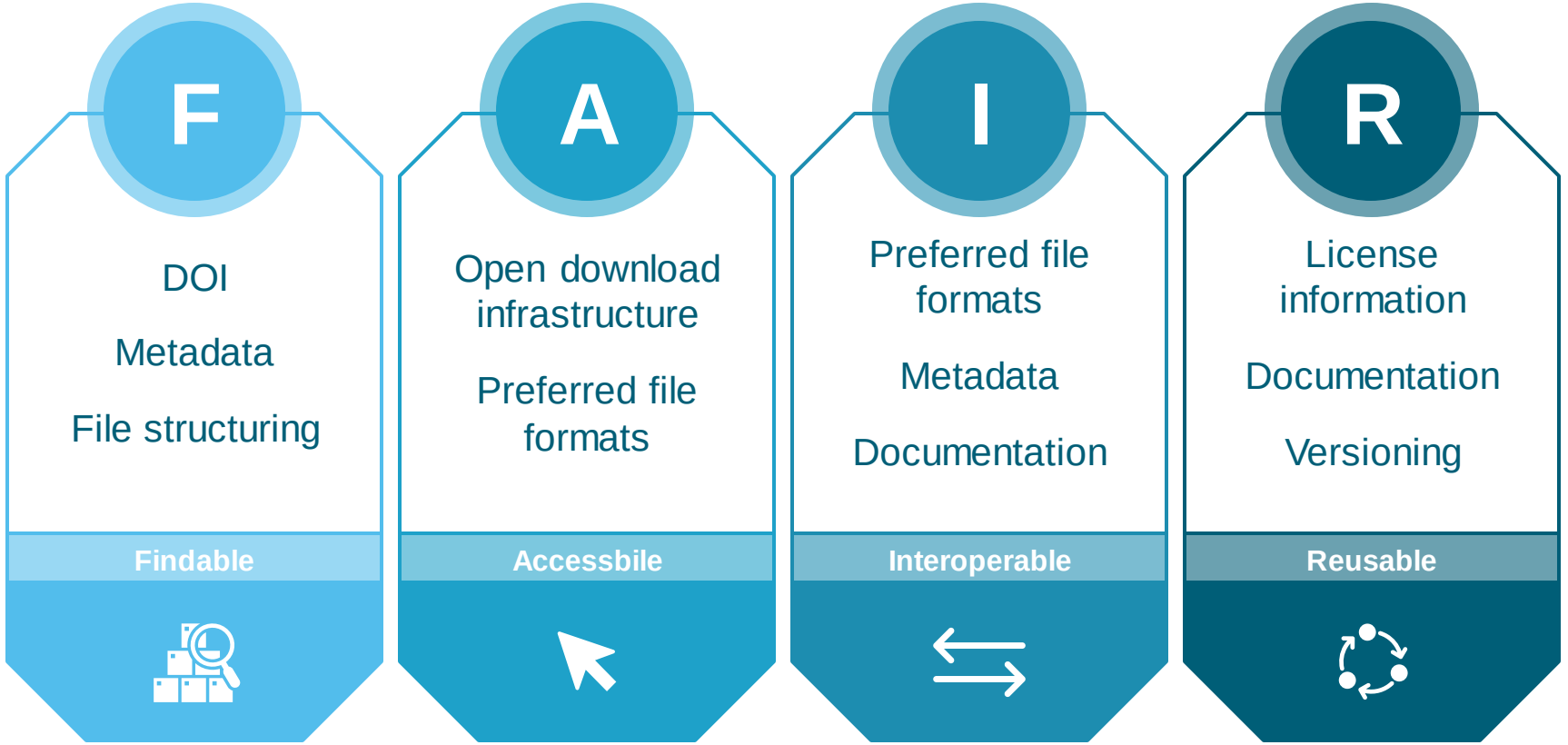
data.europa.eu

Generalist data repositories

zenodo

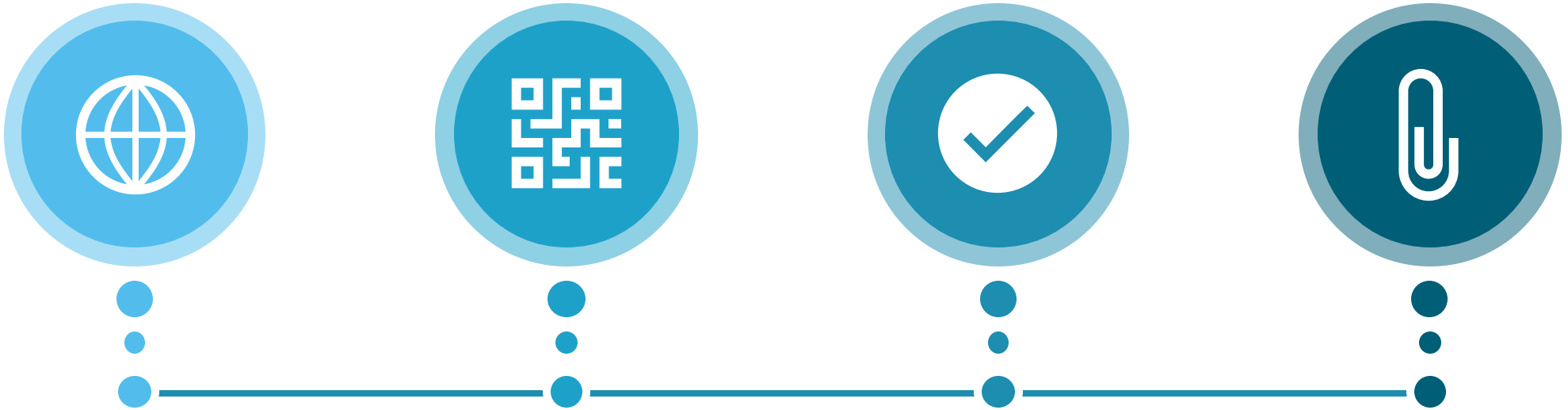
DRYAD

Why would researchers publish their data?
Repositories are FAIR-enabling (e.g., RDR)



CC-BY-4.0

Why would researchers publish their data?
Elements of a published dataset in RDR



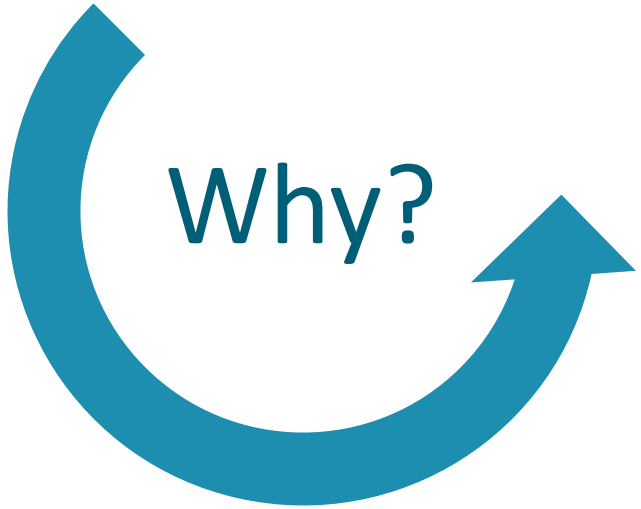
Persistent identifier

Standardized metadata

License or Terms of Use

Files & Documentation

Why would researchers publish their data?
Reuse of research data



- Makes research **verifiable and replicable**
- Increases the **reach and impact** of research
  - A dataset as its own citable research output
  - Articles that link to data in a repository have up to 25.36% higher citation impact [1]
- Also **lowers cost** of future research because of possible reuse

1. Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B (2020) The citation advantage of linking publications to research data. PLoS ONE 15(4): e0230416. https://doi.org/10.1371/journal.pone.0230416

CC-BY-4.0

Finding data

LIBIS INTO INFO
KU LEUVEN LIBRARIES

Where to find published data

Where to find published data

data.europa.eu

OpenAIRE

DataCite Commons

Where to find published data

data.europa.eu

OpenAIRE

DataCite Commons

DataverseNO

KU LEUVEN RESEARCH DATA REPOSITORY

EGA
European Genome-phenome Archive

DRYAD

zenodo

PANGAEA.
Data Publisher for Earth & Environmental Science

How can the data be accessed

- How can the data be accessed?
  - Case-by-case evaluation
- Access point:
  - Via portal like data.europa.eu
  - Via repository
  - On request
- Before access
  - Read Terms of Use/Licensing information

How can the data be accessed

Example

KU LEUVEN RESEARCH DATA REPOSITORY

https://rdr.kuleuven.be/dataset.xhtml?persistentId=doi:10.48804/YVZDUY (https://doi.org/10.48804/YVZDUY)

- Files
- Metadata
- Terms
- Versions

Data reuse

What to avoid

- Unlawful reuse
  - Licensing/terms of use
    - no license = no permission to access, use and share
  - DTA (data transfer agreements) if applicable
    - protect internal data access to adhere to DTA
    - tip: save the DTA alongside your data

What to do

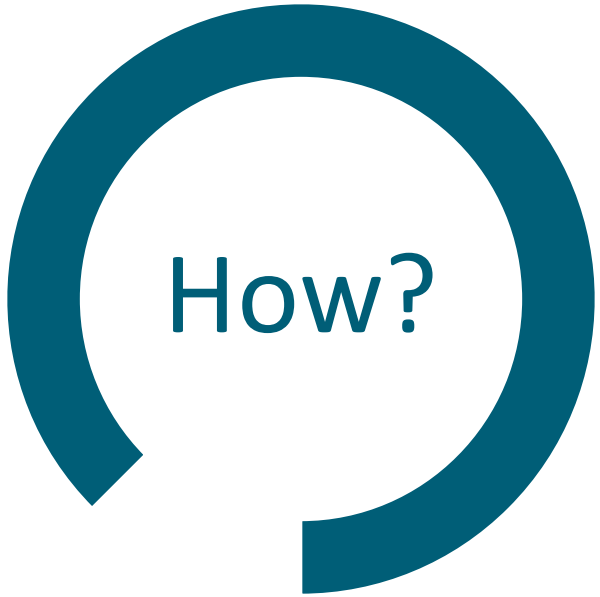
- Download data
- Export and save metadata alongside files
  - (including version & license info and accessed date)
- Ensure internal sharing also adheres to lawful reuse & users are aware of stipulations

What to do next

- Data used for your research
  - Either derived from pre-existing data or new data generated

- Publish your data
  - Avoid data duplication. Refer instead to pre-existing data where possible
  - Use trusted repositories/platforms
  - Always refer/cite the original dataset if reused in any way

How to share your data



Selection of data

At minimum share the data that is necessary to **replicate your research**

Any restrictions? (3rd party ownership, existing licensing, privacy issues, IP…)

Protect your data (anonymization & pseudonymisation)

As open as possible, **as closed as necessary**

In a trusted research data repository

CC-BY-4.0

Cinema context in (linked) data driven research

Leon van Wissen

data europa academy

Cinema Context

in (linked) data driven research

Leon van Wissen, l.vanwissen@uva.nl

How to use open data for your research - Workshop data.europa.eu - 19 October 2023

University of Amsterdam
Faculty of Humanities

creative Amsterdam
an E-Humanities perspective

Introduction

Leon van Wissen
Data Engineer, University of Amsterdam
l.vanwissen@uva.nl

https://www.create.humanities.uva.nl/
https://aihr.uva.nl/humanities-labs/humanities-labs.html

University of Amsterdam
Data Science Centre

creative Amsterdam
an E-Humanities perspective

Dataset

Cinema Context

www.cinemacontext.nl

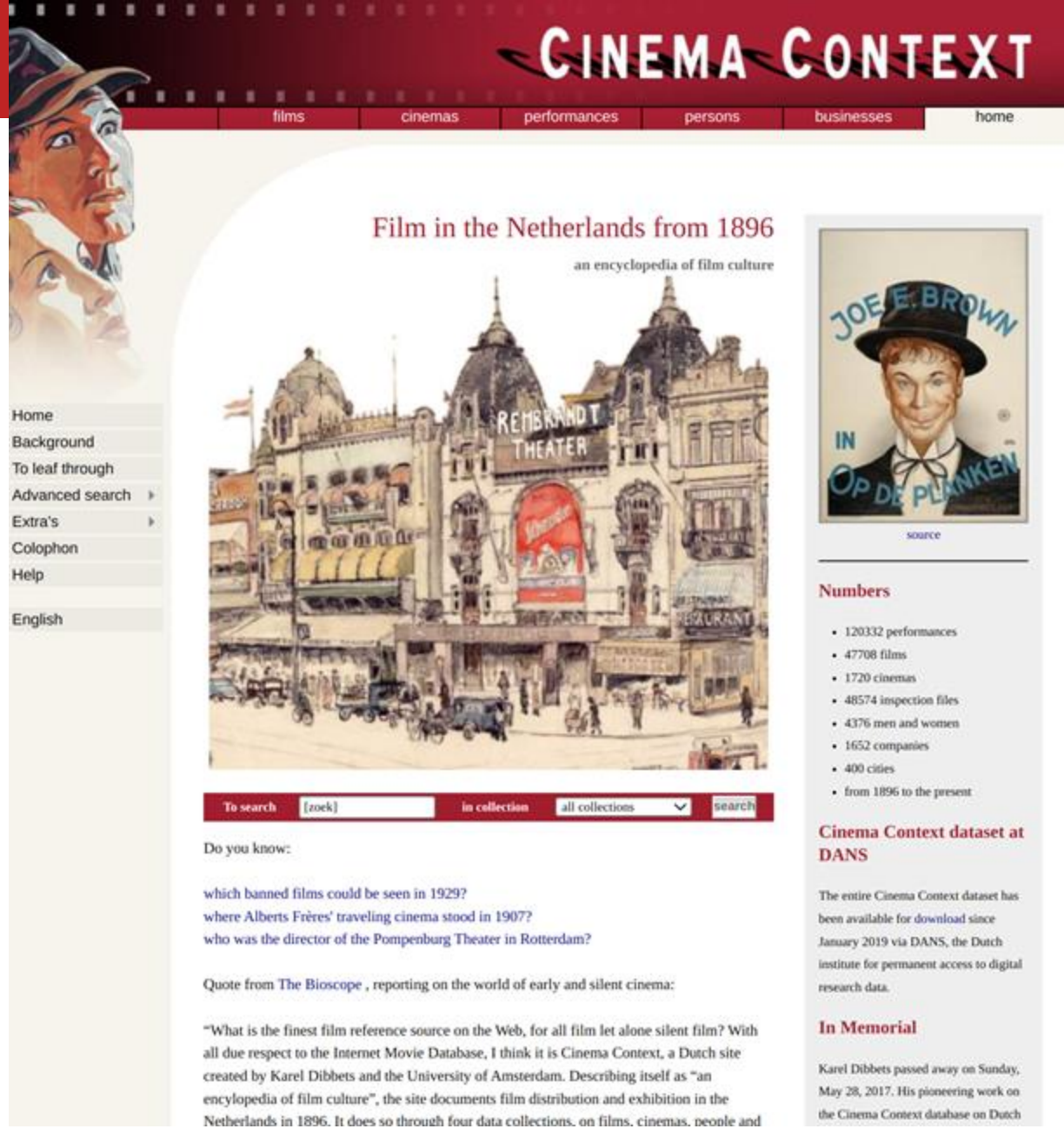
Cinema Context

About

- Encyclopedia of Dutch film culture
- History of cinema exhibition and distribution
  - 'New Cinema History': moviegoing as a social and cultural phenomenon

Content

- Films, Venues, Screenings, Persons, Companies
- Outgoing links to other datasets (IMDb)

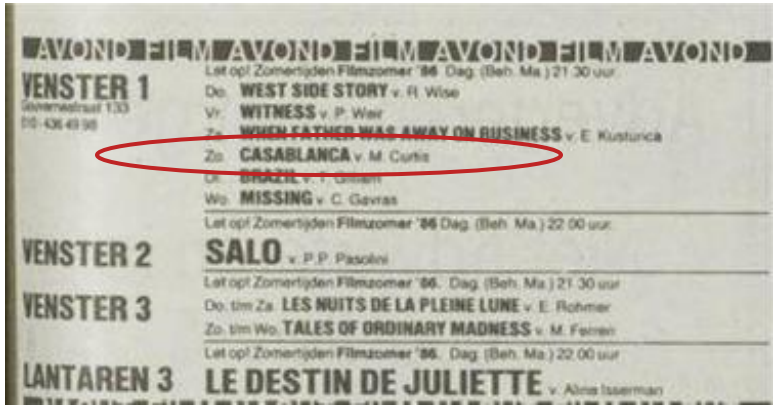


Van Oort, T., & Noordegraaf, J. (2020). The Cinema Context Database on Film Exhibition and Distribution in the Netherlands: A Critical Guide, *Research Data Journal for the Humanities and Social Sciences*, *5*(2), 91-108 https://doi.org/10.1163/24523666-00502008

Adding new data

- From archival sources
  - Newspapers, magazines
- Delpher (NL) - https://www.delpher.nl/
  - Digitized corpus of printed media
- Automatic extraction of screening data
  - OCR + NER
  - Crowdsourcing (https://rotterdamspubliek.nl/)
  - After curation, stored in Cinema Context database

Delpher



Het vrije volk: democratisch-socialistisch dagblad (1986-08-06)
https://resolver.kb.nl/resolve?urn=ddd:010948988:mpeg21:p029

Conversion to Linked Open Data (LOD)

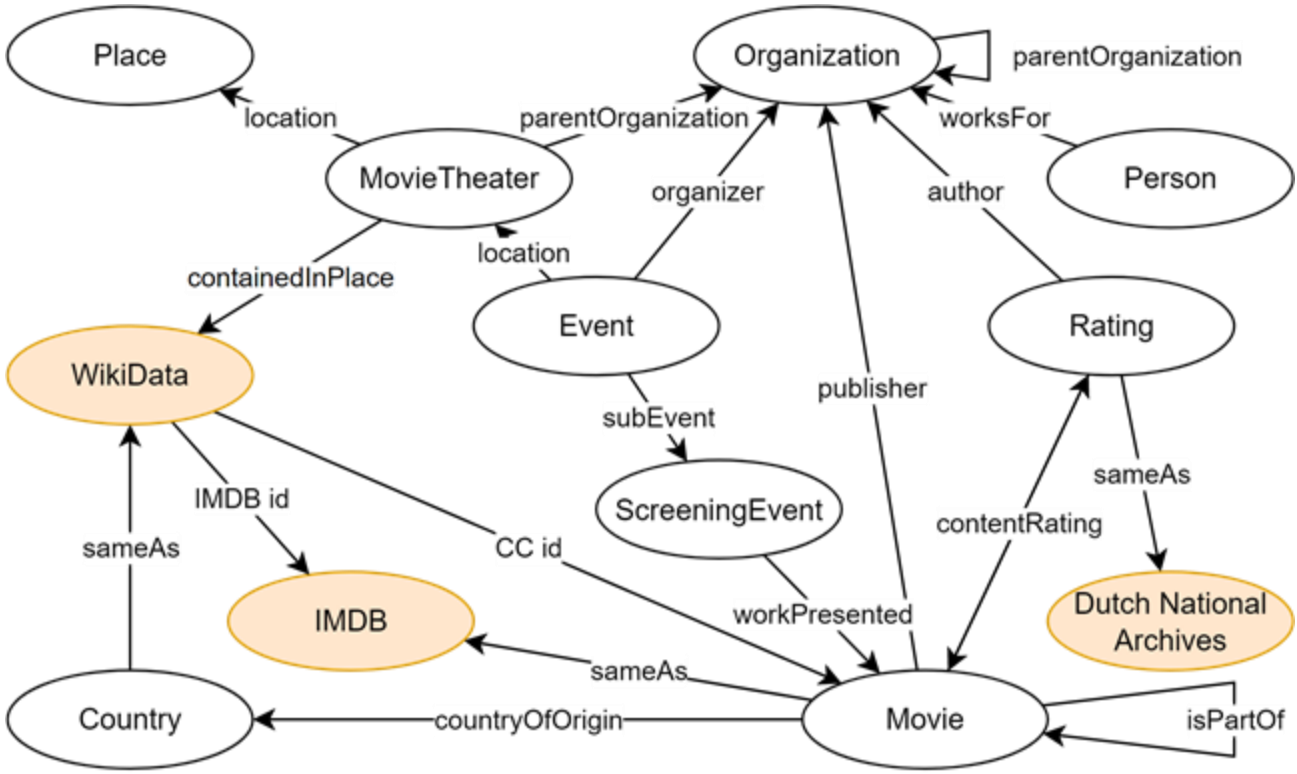
van Wissen, L., van Oort, T., Noordegraaf, J., & Kisjes, I. (2021). Cinema Context as Linked Open Data. *SEMANTICS*, 5. https://ceur-ws.org/Vol-2941/paper10.pdf

Cinema Context in RDF

Linked Open Data (LOD)
Gaining momentum in digital cultural heritage

- Each *thing* has a URI as unique identifier
  - We use its permalink (e.g. http://www.cinemacontext.nl/id/F020802)
- Statements about these *things* are stored in 'triples' in a graph database
- External relations

With these URIs we (**and others!**) can use the web to describe our data



e.g. your country's film archive!

LOD - example in RDF

- Resource Description Framework
- Triples
- Uses URLs/URIs as identifiers
- 'Linked' by pointing to other (online) datasets

Subject – predicate – object

film1 – is of the type – Film

film1 – has title – Casablanca

film1 – is directed by – Michael Curtiz

film1 – is located in - Morocco

person1 – is of the type – Person

person1 – has name – Michael Curtiz

Casablanca – is directed by – Michael Curtiz

<http://www.wikidata.org/entity/Q132689> <http://www.wikidata.org/prop/direct/P57> <http://www.wikidata.org/entity/Q51491>

Casablanca – is located in - Morocco

<http://www.wikidata.org/entity/Q7903> <http://www.wikidata.org/prop/direct/P17> <http://www.wikidata.org/entity/Q1028>

Linked Data on the web

Our *specialised* Cinema Context dataset can add to this:

- **Programmes and Events**
- Reviews and censorships
- Companies and Theaters

Casablanca

1942 · Romance/Drama · 1h 42m

Play trailer on YouTube

Watch now
From €2.99

Watched it?

Watchlist

All watch options

8,5/10	99%	100%
IMDb	Rotten Tomatoes	Metacritic

86% liked this film

Google users

During WWII, Rick, a nightclub owner in Casablanca, agrees to help his former lover Ilsa and her husband. Soon, Ilsa's feelings for Rick resurface and she finds herself renewing her love for him.

Release date: April 4, 1947 (Netherlands)

Director: Michael Curtiz

Featured song: As Time Goes By

Awards: Academy Award for Best Picture, MORE

Languages: French, German, Italian, English

Cast

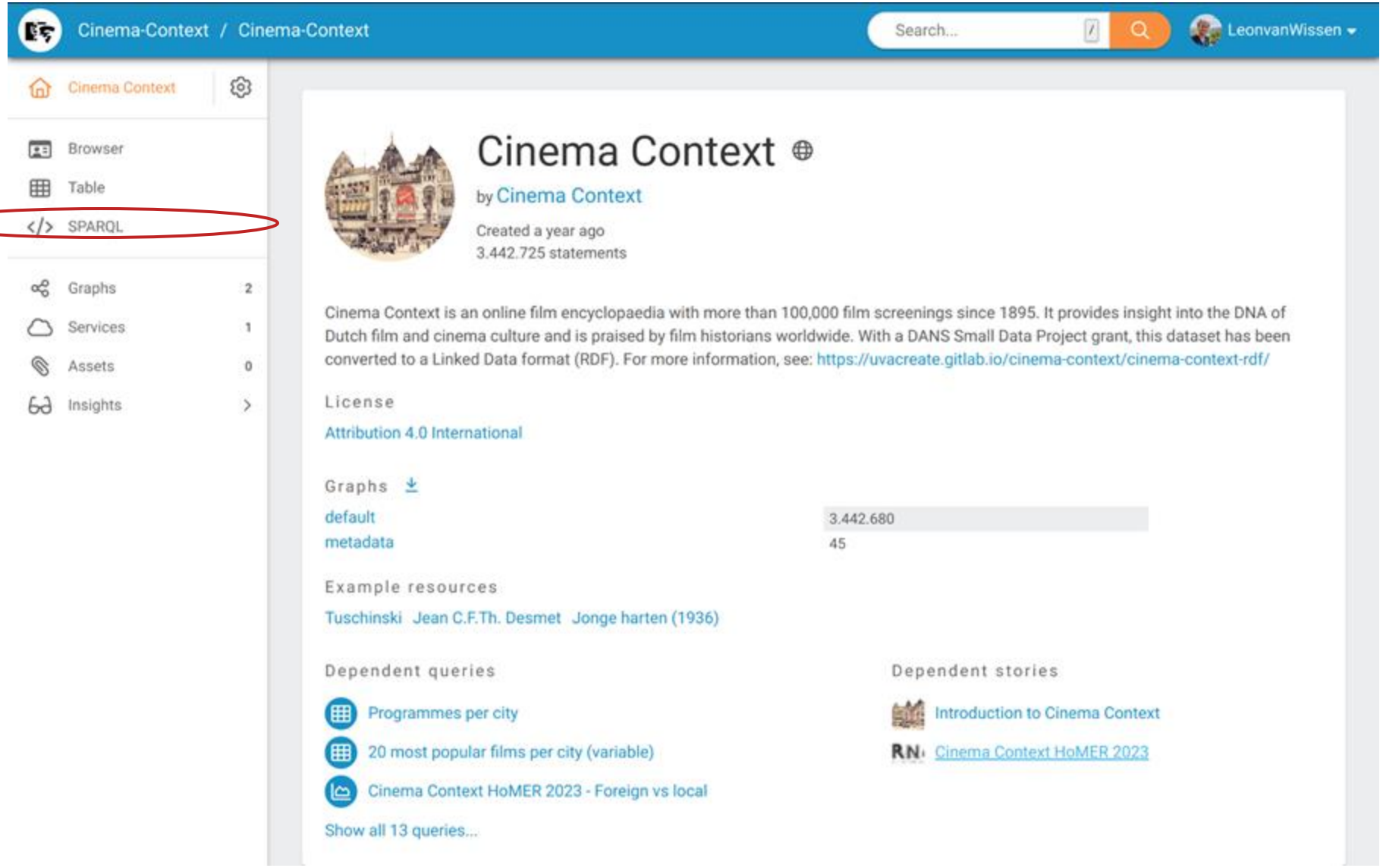
View 15+ more

Humphrey Bogart
Rick Blaine

Ingrid Bergman
Ilsa Lund

Claude Rains
Captain Louis Renault

Paul Henreid
Victor Laszlo



Hosted by the UvA University Library: https://lod.uba.uva.nl/Cinema-Context/Cinema-Context

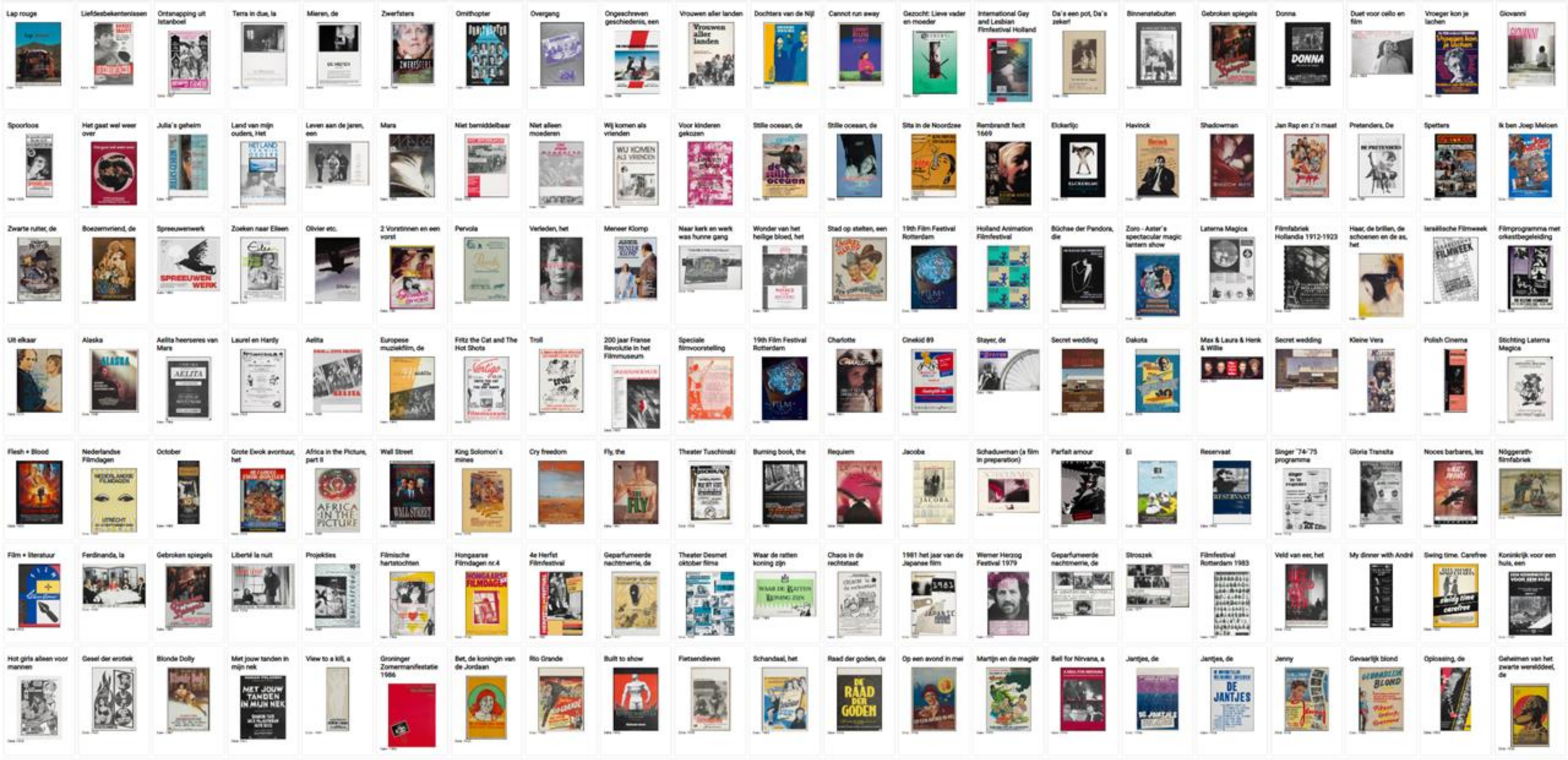
Linking to other datasets

Poster collection

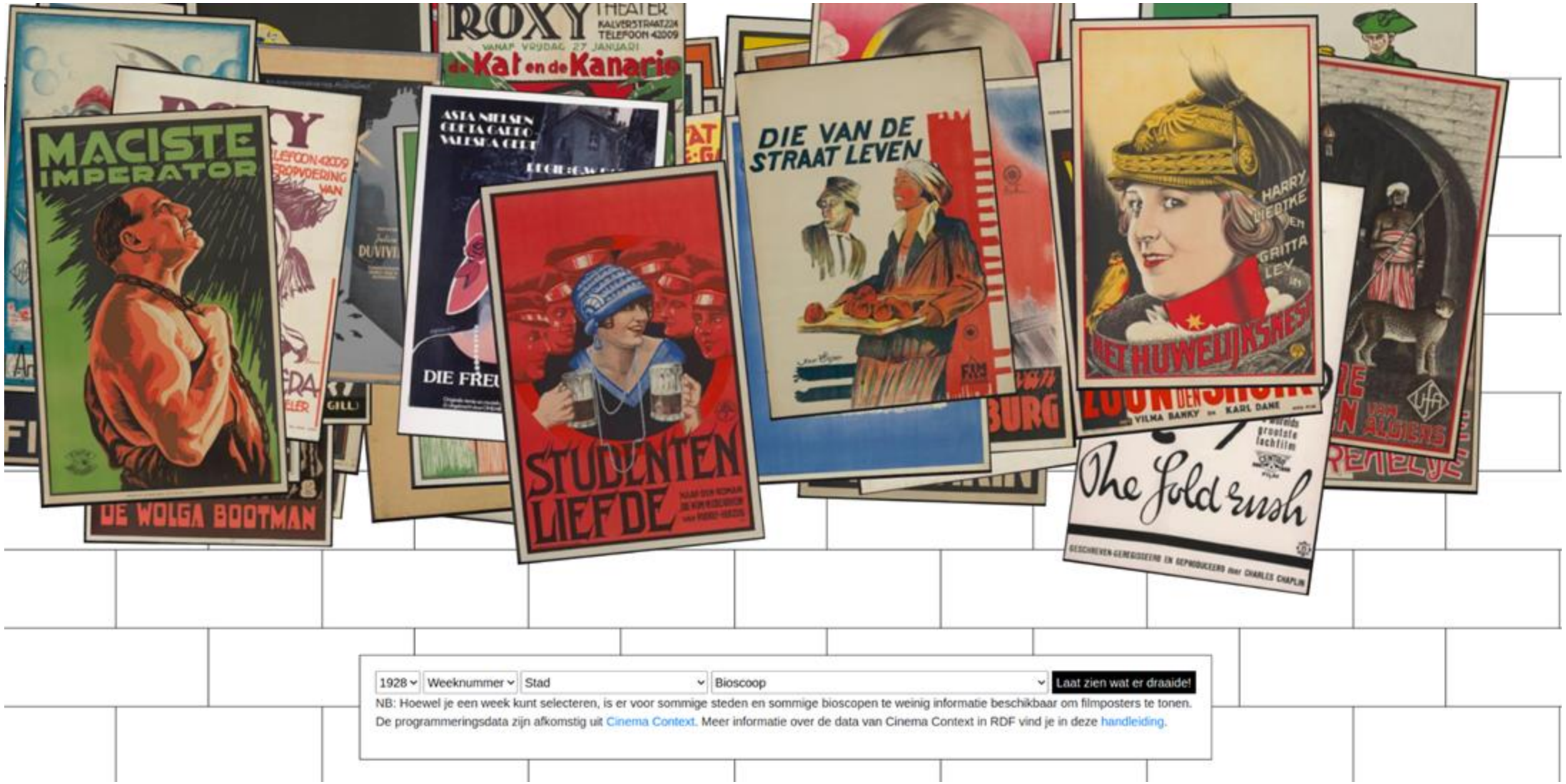
The Memory Database (https://geheugen.delpher.nl/)
Online heritage collections – paintings, drawings, stamps, posters and photographs

- EYE Museum Collection
  - Poster Project: https://geheugen.delpher.nl/en/geheugen/pages/collectie/Afficheproject/Collectie+EYE
  - Turned into LOD during the Hack-a-LOD 2020 https://hackalod.com/

https://resolver.kb.nl/resolve?urn=urn:gvn:NAG02:EYE-A41790



Query result: All posters from this collection with basic metadata



Query result: Combining screening data from Cinema Context with images

Interactive demo: https://www.hicsuntleones.nl/plakmuur/eye/

Thanks to Thunnis van Oort (RU) and Menno den Engelse (http://islandsofmeaning.nl/) for building this together.

Data stories

A new way of documenting your research

Data stories

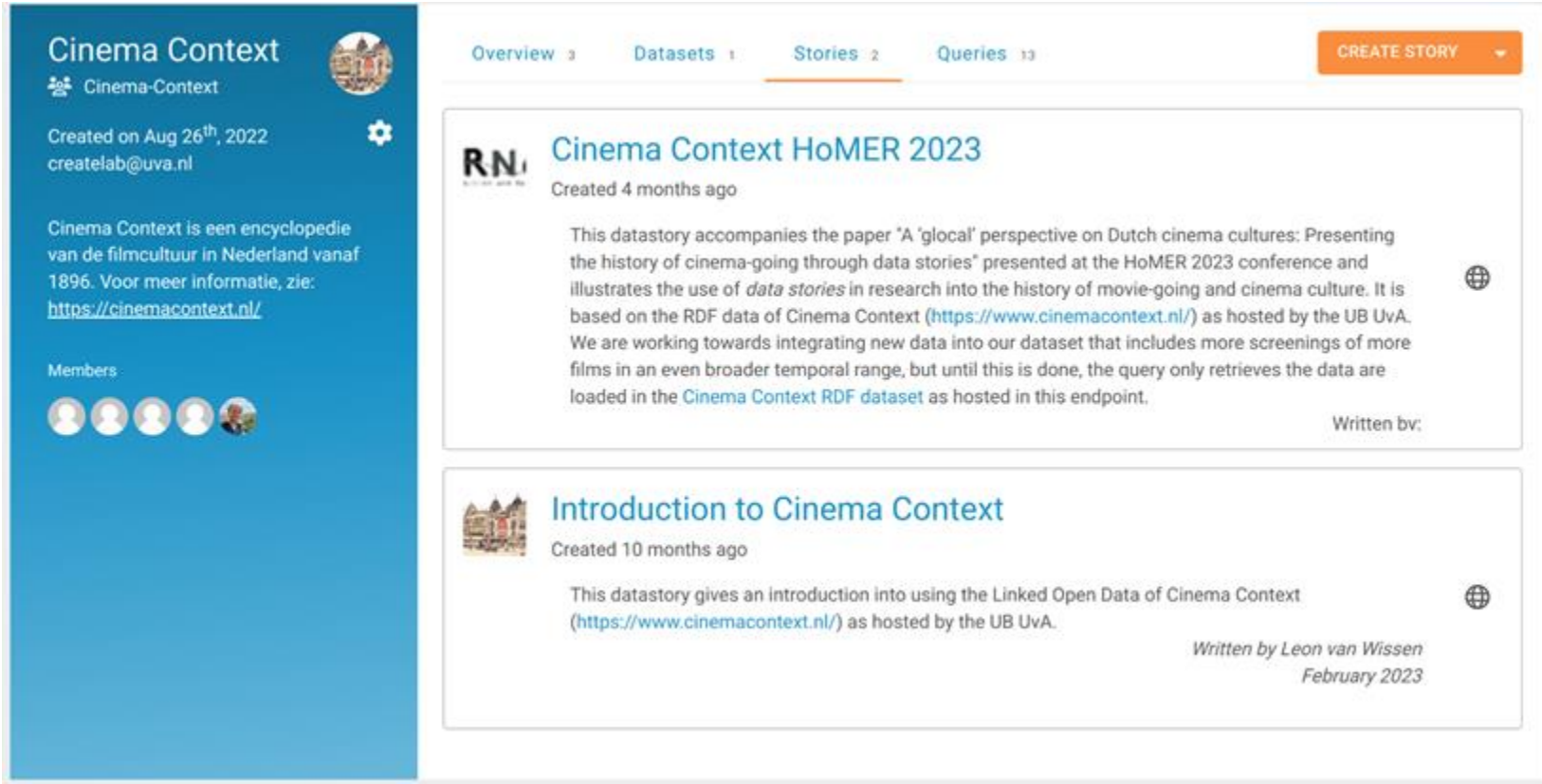
- Narratives around data-driven research
  - Research question + explanation
  - Code (queries, programming)
  - Results + visualizations
- Improve digital literacy
  - Explain to your audience what you do and make them repeat it
- 'Data publication' for your research paper

Example:

1. Data extraction documented and explained in **markdown** a repository on Github
2. Idem, in an interactive **Jupyter notebook**
3. Or, via **SPARQL** in an interactive Linked Open Data story

Cinema Context Data Stories

- Examples of static documentation and Jupyter notebooks:
  - https://uvacreate.gitlab.io/cinema-context/cinema-context-rdf/
- Linked Data Examples in the UvA's LOD platform:
  - https://lod.uba.uva.nl/Cinema-Context/-/stories



Go to dataset Try this query yourself

name *

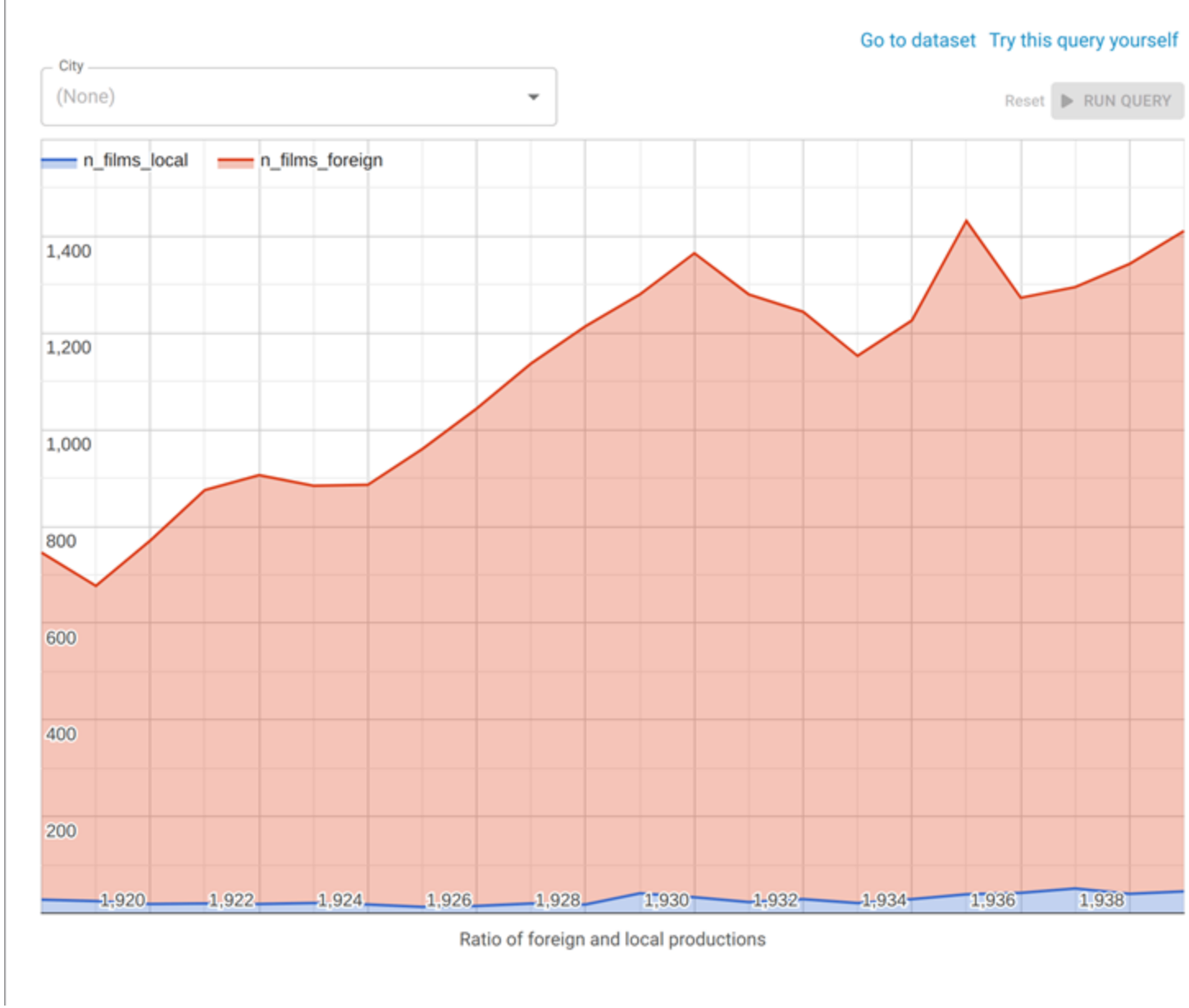
city

Reset ▶ RUN QUERY

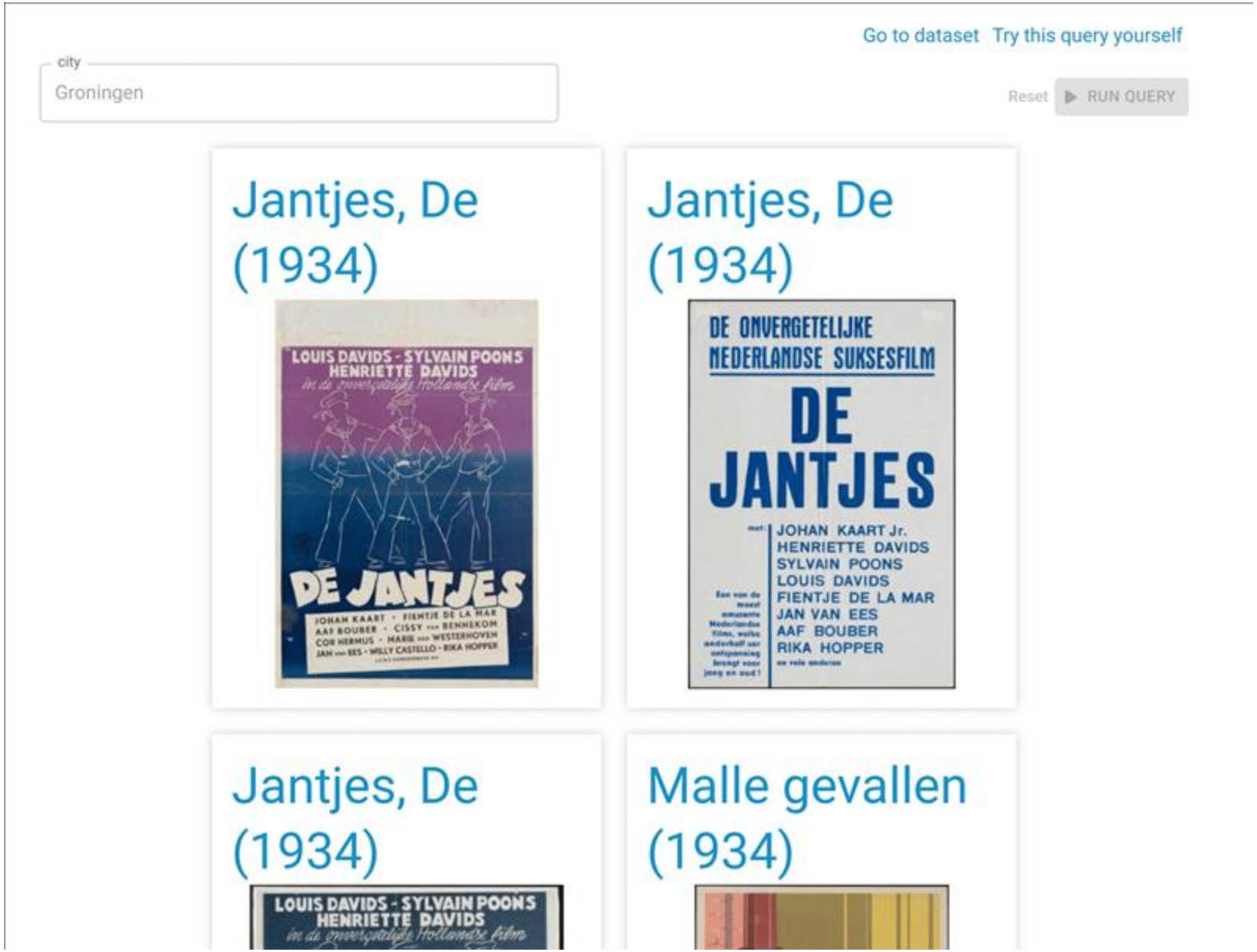
	film	name	programme	venueName	city	date
1	<http://www.cinemacontext.n...	Walzertraum, Ein (1925)	<http://www.cinemacontext.nl/i...	National Theater	Amster...	1928-08...
2	<http://www.cinemacontext.n...	Tiger Rose (1923)	<http://www.cinemacontext.nl/i...	Scala (Kruiskade)	Rotterd...	1929-08...
3	<http://www.cinemacontext.n...	Faust (1926)	<http://www.cinemacontext.nl/i...	Luxor	Rotterd...	1934-03...
4	<http://www.cinemacontext.n...	Jantjes, De (1934)	<http://www.cinemacontext.nl/i...	Olympia	Rotterd...	1934-04...
5	<http://www.cinemacontext.n...	Frisco Jenny (1932)	<http://www.cinemacontext.nl/i...	Ooster Theater	Rotterd...	1934-08...
6	<http://www.cinemacontext.n...	Two gun men, The (1931)	<http://www.cinemacontext.nl/i...	Ooster Theater	Rotterd...	1934-08...
7	<http://www.cinemacontext.n...	Hygiene der Ehe (1922)	<http://www.cinemacontext.nl/i...	Scala (Kruiskade)	Rotterd...	1937-10...
8	<http://www.cinemacontext.n...	Hygiene der Ehe (1922)	<http://www.cinemacontext.nl/i...	Scala (Kruiskade)	Rotterd...	1937-11...
9	<http://www.cinemacontext.n...	Slim (1937)	<http://www.cinemacontext.nl/i...	Harmonie	Rotterd...	1937-12...
10	<http://www.cinemacontext.n...	Seventh Heaven (1937)	<http://www.cinemacontext.nl/i...	Harmonie	Rotterd...	1937-12...
11	<http://www.cinemacontext.n...	Glücklichste Ehe der Welt...	<http://www.cinemacontext.nl/i...	Corso	Den Haag	1939-01...
12	<http://www.cinemacontext.n...	Hurricane (1937)	<http://www.cinemacontext.nl/i...	Corso	Den Haag	1939-01...
13	<http://www.cinemacontext.n...	Under two flags (1936)	<http://www.cinemacontext.nl/i...	Rex (Lange Poten)	Den Haag	1939-03...
14	<http://www.cinemacontext.n...	Dir gehört mein Herz (19...	<http://www.cinemacontext.nl/i...	Colosseum	Rotterd...	1940-01...
15	<http://www.cinemacontext.n...	Shopworn angel (1938)	<http://www.cinemacontext.nl/i...	Colosseum	Rotterd...	1940-01...
16	<http://www.cinemacontext.n...	Gibraltar (1938)	<http://www.cinemacontext.nl/i...	Centraal (Binnen...	Rotterd...	1940-01...
17	<http://www.cinemacontext.n...	Pfarrer von Kirchfeld, Der...	<http://www.cinemacontext.nl/i...	Roxy	Den Haag	1940-01...
18	<http://www.cinemacontext.n...	Boys town (1938)	<http://www.cinemacontext.nl/i...	Roxy	Den Haag	1940-01...
19	<http://www.cinemacontext.n...	Indianapolis speedway (1...	<http://www.cinemacontext.nl/i...	Harmonie	Rotterd...	1940-02...
20	<http://www.cinemacontext.n...	Club de femmes (1936)	<http://www.cinemacontext.nl/i...	Harmonie	Rotterd...	1940-02...

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <https://schema.org/>

SELECT ?film ?name ?programme ?venueName ?city ?date {

   ?film a schema:Movie ;
         schema:name ?name .

   ?programme a schema:Event ;
            schema:location ?venue ;
            schema:startDate ?date ;
            schema:subEvent [ a schema:ScreeningEvent ;
                              schema:workPresented ?film ] .

   ?venue a schema:MovieTheater ;
          schema:name ?venueName ;
          schema:location ?place .

   ?place a schema:Place ;
            schema:address [ schema:addressLocality ?city ] .

} GROUP BY ?film ?name ORDER BY ?date LIMIT 20
```

Query: All screenings for one film in a city (+ venue and date)



Query: Number of foreign and non-foreign films per city (based on country of production)



Query: Federating between datasets: Cinema Context and Poster collection (additionally: Wikidata and others)

Contact

Leon van Wissen
University of Amsterdam
l.vanwissen@uva.nl

With special thanks to the University Library of the University of Amsterdam, for hosting this data in their Linked Open Data platform!
https://uba.uva.nl/en/home

From Data to Knowledge: Strategies for Research and Real-World Applications

Maria Pia Di Buono

data europa academy

How to use open data for your research?
19 October 2023

From Data to Knowledge

Strategies for Research and Real-World Applications

Maria Pia di Buono
University of Naples “L’Orientale” - Italy

Use Cases

Across Europe with MAGGIE

NEAT - Named Entities in Archaeological Texts

data.europa.eu

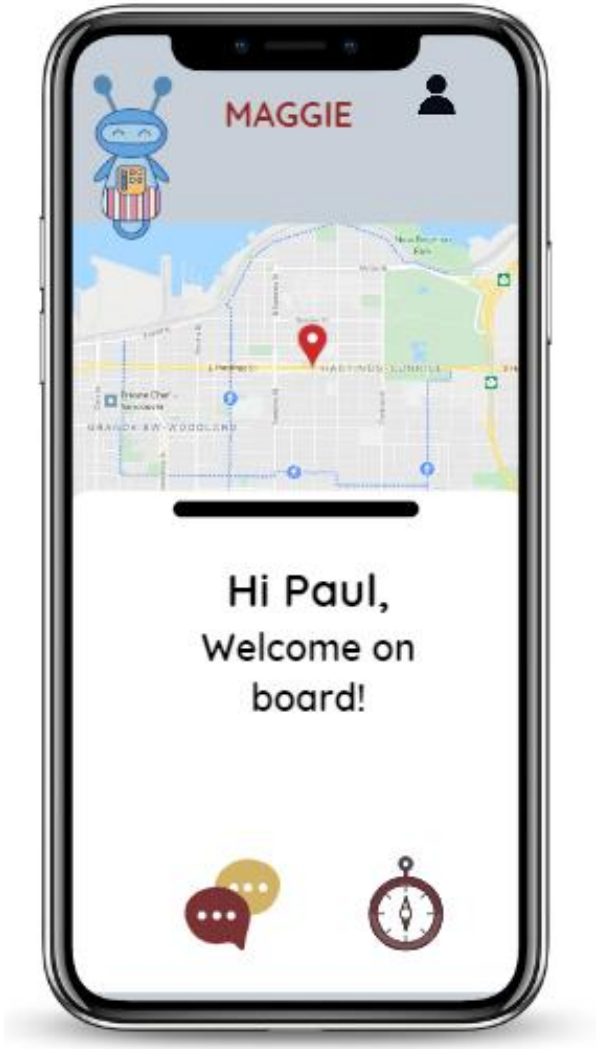
europeana

Across Europe with MAGGIE 1

3rd place: UNIOR NLP

Your guide to cultural content across Europe (Italy)

Across Europe with MAGGIE 2

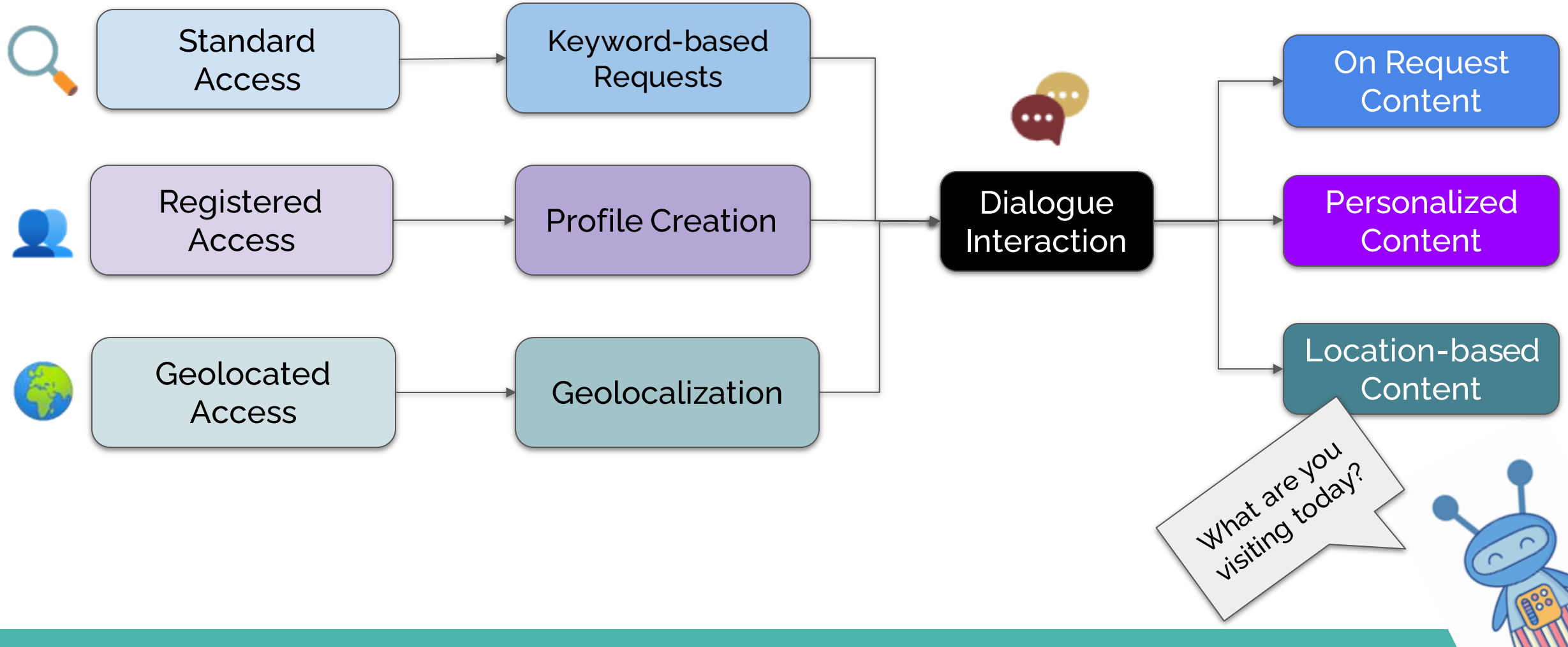


You
Hi Maggie!
Which paintings by Tiziano can I find in the Uffizi Gallery?

Maggie
Hi Paul! You can admire:
1) Venus and Cupid;
2) Venus of Urbino;
3) Flora;
4) Resurrection of Lazarus.
and others! Are you interested in any of them? Or do you want to see the full list?

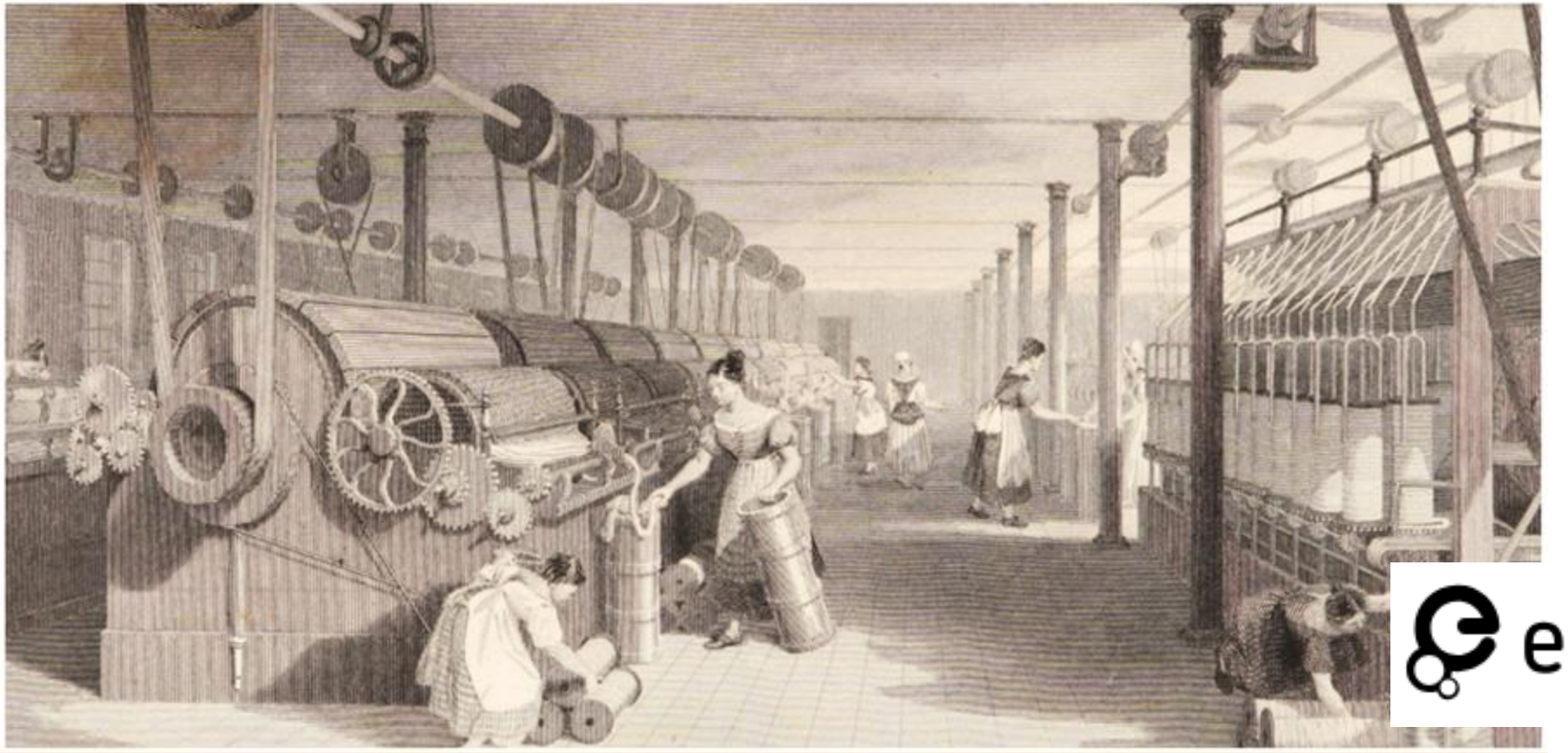
Let's explore Europe!

Across Europe with MAGGIE 3



NEAT - Named Entities in Archaeological Texts 1

EuropeanaTech Challenge for Europeana AI/ML Datasets



europeana pro

NEAT - Named Entities in Archaeological Texts 2

What

A dataset for Named Entity Recognition (NER) and Term Extraction in Italian and English

How

Europeana Open data + external vocabularies and models (i.e., Getty and CIDOC CRM)

Why

Development and evaluation of AI/ML based technologies

Type	Total (Unique)
E2—Temporal	–
E4—Period	94 (12)
E5—Event	385 (90)
E18—Physical Thing	17,656 (4,240)
E28—Conceptual Thing	233 (68)
E39—Actor	15,654 (2,626)
E41—Appellation	111 (35)
E52—Time Span	–
E53—Place	4,648 (595)
E54—Dimension	4 (2)
E57—Material	2,217 (137)
Total	41,002 (7,805)

Four Strategies

Reusing

Enriching

Developing

Sharing

Reusing

- Data Collection
- Data Cleaning
- Data Normalization

- *Data Normalization*

Creator: "Jan Vladimir Hrasky (architect) ,
Jurij Stare (photographer)"

→ Jan Vladimir Hrasky

→ Jurij Stare

- *Data Collection from several sources:*

data.europa.eu

europeana

Dataset Europeana

Examples of GeoDataset about GLAM and Museums:

Namur – Museen

Museums (data.gov.be)

Museums (dados.gov.pt)

Users' Preferences
Location, Interests, [...]

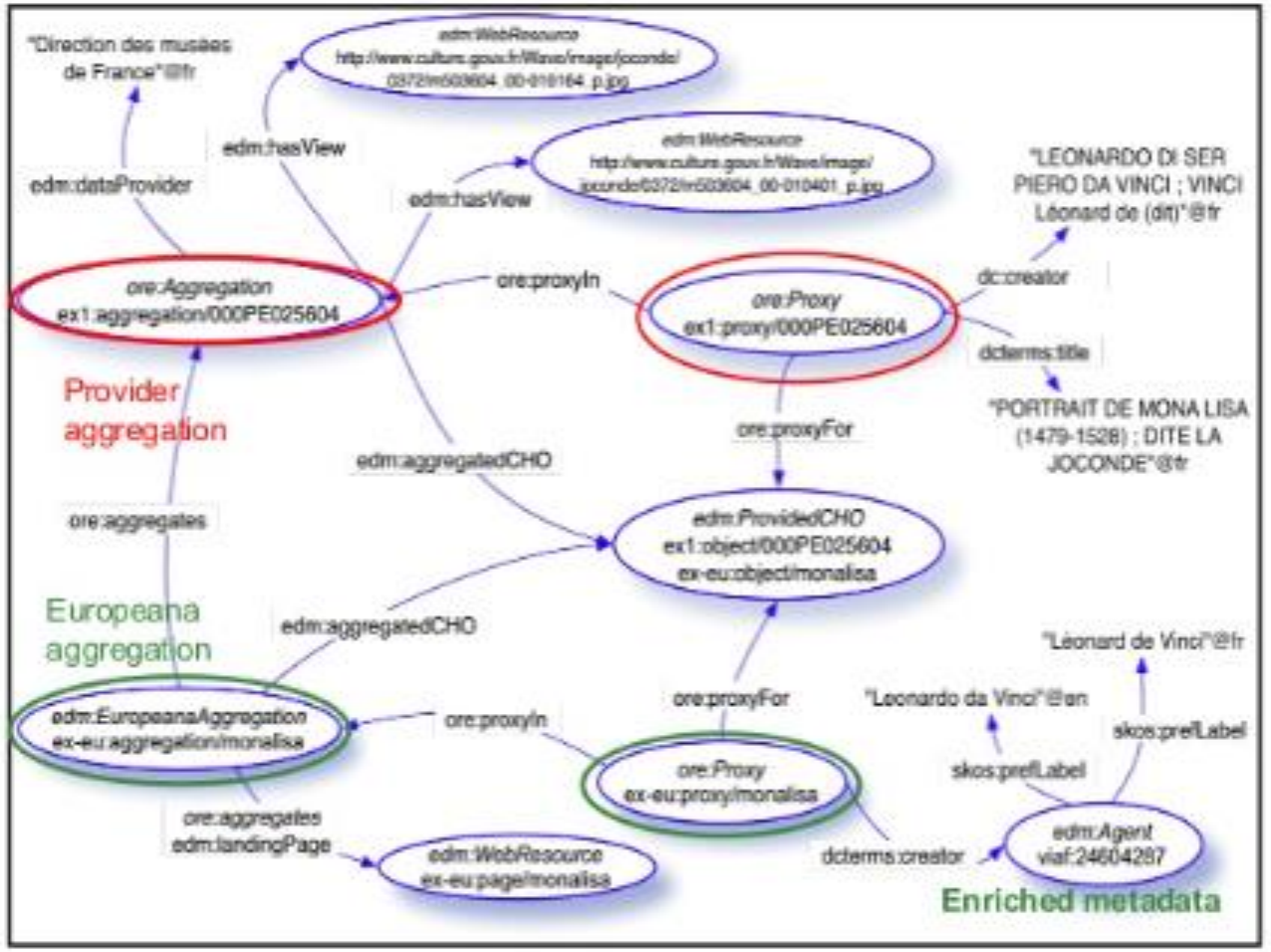
- *Terminology Representation*
- *Multilingual Terminology Management*
- *Discourse Representation*

Enriching

- Data Integration
- Data Alignment & Harmonization
- Data Linking

- *Data Alignment*

Jan Vladimir Hrasky ⟶ Q3499090



Developing

- Output definition
- Data test and validation
- Data visualization

Natural Language Processing

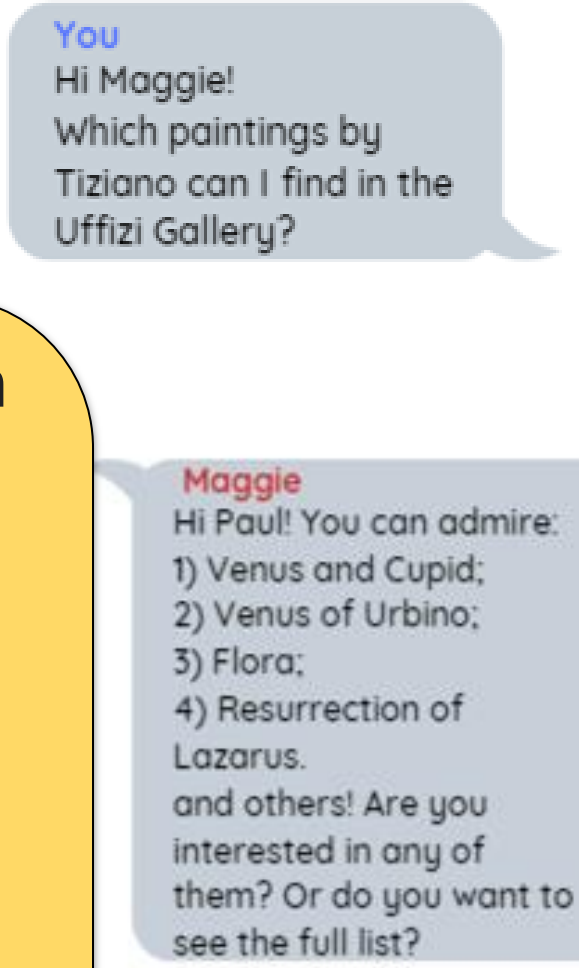
- Terminology Extraction + Named Entity Extraction + Keyword Extraction
- Entity Linking
- Semantic Parsing
- Intent Classification
- Relation Extraction

Paintings	`created-by`	**Tiziano**
Paintings	`located-in`	**Uffizi Gallery**

Knowledge Extraction

- *Information Extraction*
- *Data Aggregation*
  - Venus and Cupid
  - Venus of Urbino
  - Flora
  - Resurrection of Lazarus
  - Rape of Europe
  - …
- *Output Generation*

"Hi, `[user_name]`*! You can admire: …"*



Sharing

- Metadata enrichment
- Documentation
- Publication
- Maintenance

Thank you!
Questions?

Break
11.05 - 11.10

data.europa academy

Case with open data

Leon Mook

data europa academy

Exercise for the workshop 'How to use open data for your research

Case study

30 minutes

Fill in the answers in this exercise

If you're stuck, move on to the next question and see how far you can get

Get creative, you're allowed to use all tools you want

Feel free to ask your questions in the chat

Questions & Answers

data europa academy

Please provide your feedback!

Stay up-to-date on our **activities!**

data.europa academy

Join our next webinar!

data europa academy

Sign up for the newsletter: data.europa.eu/newsletter

Follow us on social media:

EU_opendata

Publications Office of the European Union

data.europa.eu

data europa academy

Thank you

data.europa academy