Multilingual challenges and ongoing work to tackle them at Europeana

Antoine Isaac
with slides from Hugo Manguinhas, Valentine Charles, Juliane Stiller, Mónica Marrero

CLUBS final workshop

Saarbrücken , 7 June 2019

Quadruplettes à Buffalo
Agence de presse Meurisse
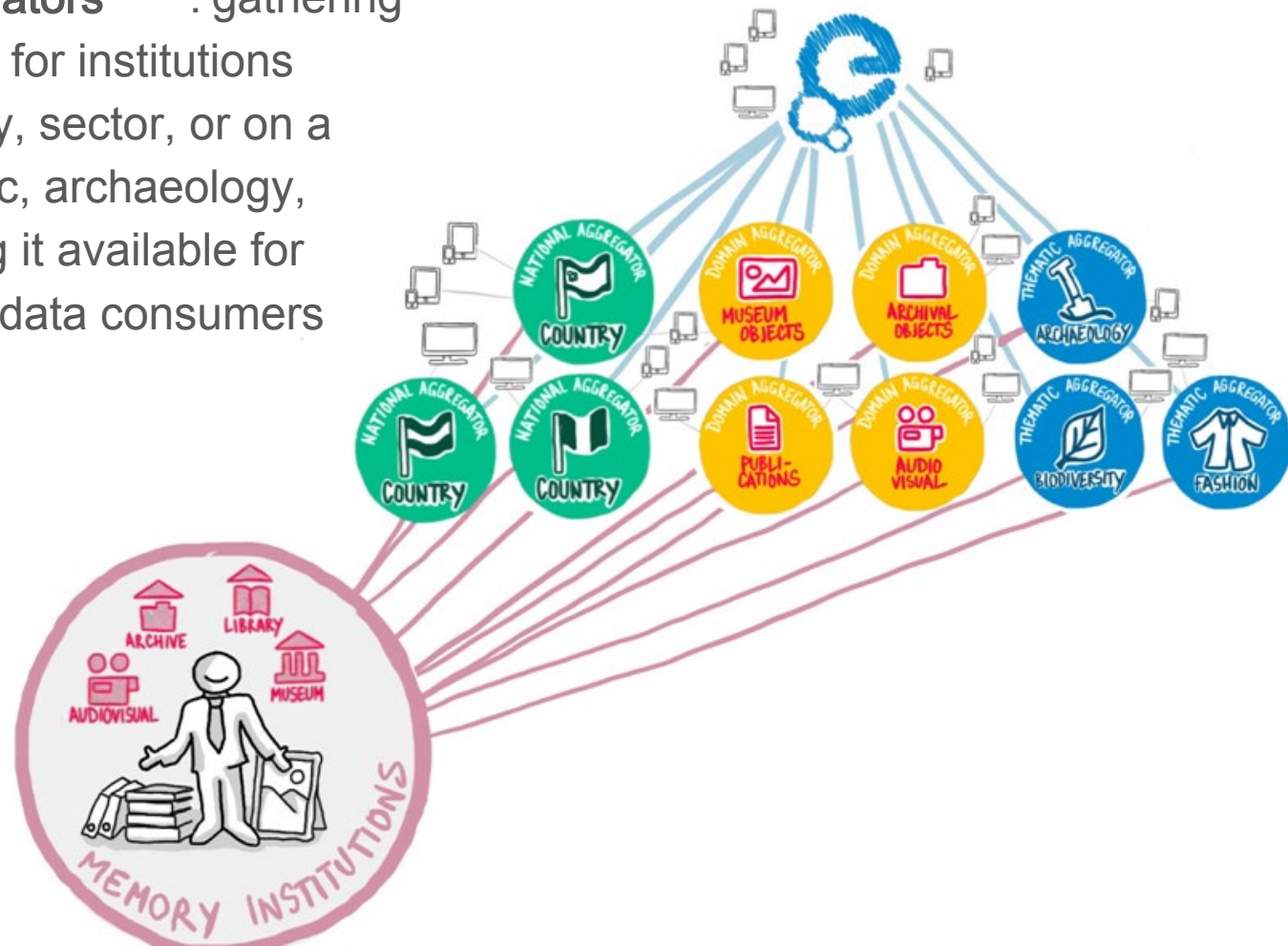1909, National Library of France
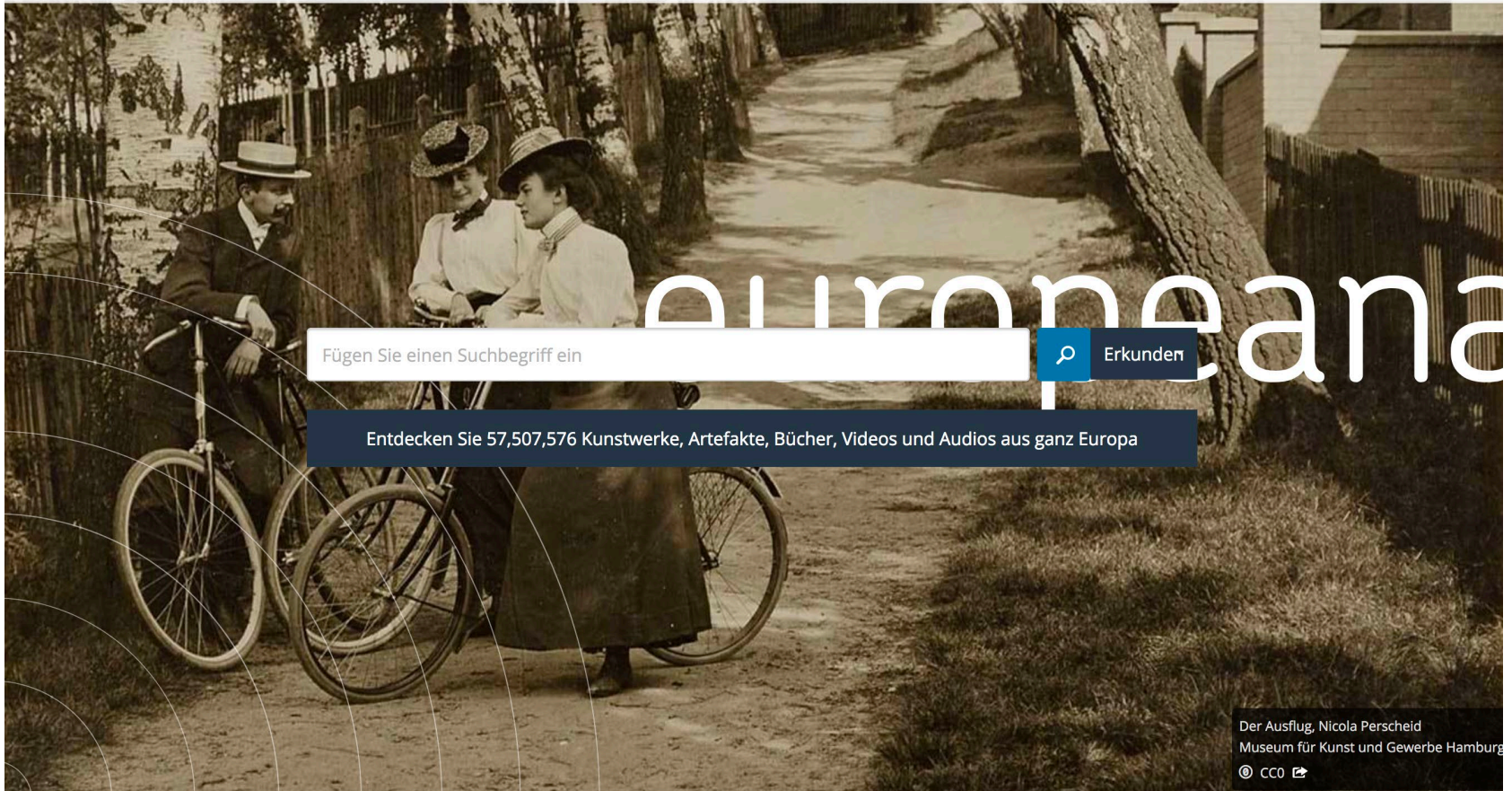France, Public Domain

What is Europeana?

- Europeana is the European Commission's digital platform for cultural heritage.

- We provide access to digital collections from libraries, museums and archives around Europe

A network of data partners

- **Data providers:** Cultural heritage institutions providing content and metadata to Europeana

- **"Intermediate" Aggregators**: gathering metadata and content for institutions from a specific country, sector, or on a specific domain (music, archaeology, theater…) and making it available for Europeana and other data consumers

europeana

57 million digitized objects, from 3,700 institutions in 44 countries

What's multilingual in there?

We working for a multilingual audience, in Europe and elsewhere

- Dynamic content   - user interface

- Static content such as curated virtual exhibitions

- Search - returning cultural object for users' queries

europeana

Multilinguality in Europeana search

Users from all around the world querying in their own language

Metadata in 37 languages

europeana collections Results

Collections ▾ Explore ▾ Exhibitions ▾ Blog ▾   OUR SITES   🄰 LANGUAGE ▾

Saarbrücken ✕   Add a search term   🔍

REFINE YOUR SEARCH   1 - 12 of 2,577 results   Per page: 12   ▦ GRID   ☰ LIST

COLLECTIONS   ▾

MEDIA   ▾
☐ Only items with links to media

CAN I USE IT? ❓   ▾

PROVIDING COUNTRY   ▾

LANGUAGE   ▲

☐ German (2,310)
☐ Dutch; Flemish (70)
☐ English (53)
☐ Catalan; Valencian (52)
☐ Multilingual (34)

☐ French (18)
☐ Italian (18)
☐ Spanish; Castilian (11)
☐ Czech (3)
☐ Croatian (2)

Paramythes, imitées d'Herder (1794) | Ludwig Franz von Bilderbeck

View at Ghent University Library

📖 Text

Saarbrücken | Möbius, Walter (Fotograf)

Saarbrücken Saarbrücken

View at Deutsche Fotothek

🖼 Image

Ausstellung Jugoslawische Volkskunst Grafik lebender jugoslawischer Kunstler (1933)

Katalog etnografske izložbe i grafičke izložbe jugoslavenskih umjetnika, održane 1933. u Saarbruckenu, sadrži popis izlagača, popratne tekstove i reprodukcije.

View at Croatian Academy of Sciences and Arts
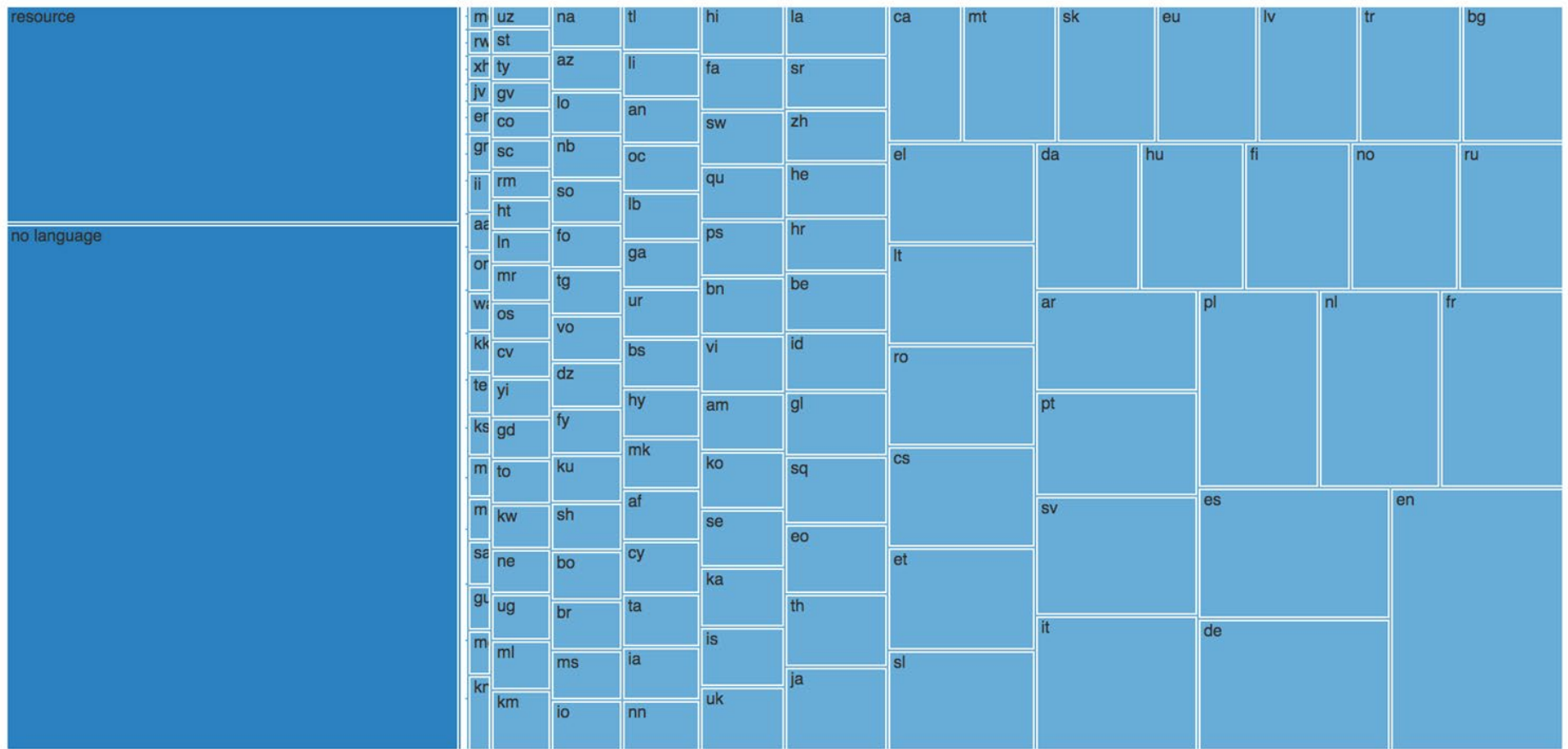
📖 Text

europeana

The data Europeana holds about cultural objects

- **Descriptive and technical   metadata**
  - Creator, title, rights…

- Some content for specific projects
  - newspapers text and   images

Massively multilingual metadata

57 million digitized objects, *from 3,700 institutions in 44 countries*

- Officially we get metadata in   37 languages

- But there are more languages used in individual metadata fields

resource												
m	uz	na	tl	hi	la	ca	mt	sk	eu	lv	tr	bg
rw	st											
xh	ty	az	li	fa	sr							
jv	gv	lo	an	sw	zh							
er	co	nb	oc			el	da	hu	fi	no	ru	
gr	sc			qu	he							
ii	rm	so	lb									
ht				ps	hr							
aa	ln	fo	ga			lt						
or	mr	tg		bn	be							
wa	os		ur				ar	pl	nl	fr		
kk	cv	vo	bs	vi	id	ro						
te	yi	dz										
ks	gd	fy	hy	am	gl		pt					
			mk									
m	to	ku		ko	sq	cs						
m	kw	sh	af	se			sv	es	en			
sa	ne	bo	cy		eo	et						
gu	ug	br	ta	ka	th		it	de				
m	ml			is		sl						
kr	km	ms	ia	uk	ja							
		io	nn									

| no language | | | | | | | | | | | | |

All fields aggregated

1. **no field instance**: 4 272 165 045
2. **no language**: 770 143 229
3. **resource**: 320 367 666
4. **en**: 145 021 537
5. **de**: 82 024 654
6. **es**: 80 571 570
7. **fr**: 79 175 871
8. **nl**: 77 276 373
9. **pl**: 76 621 843
10. **it**: 70 659 189

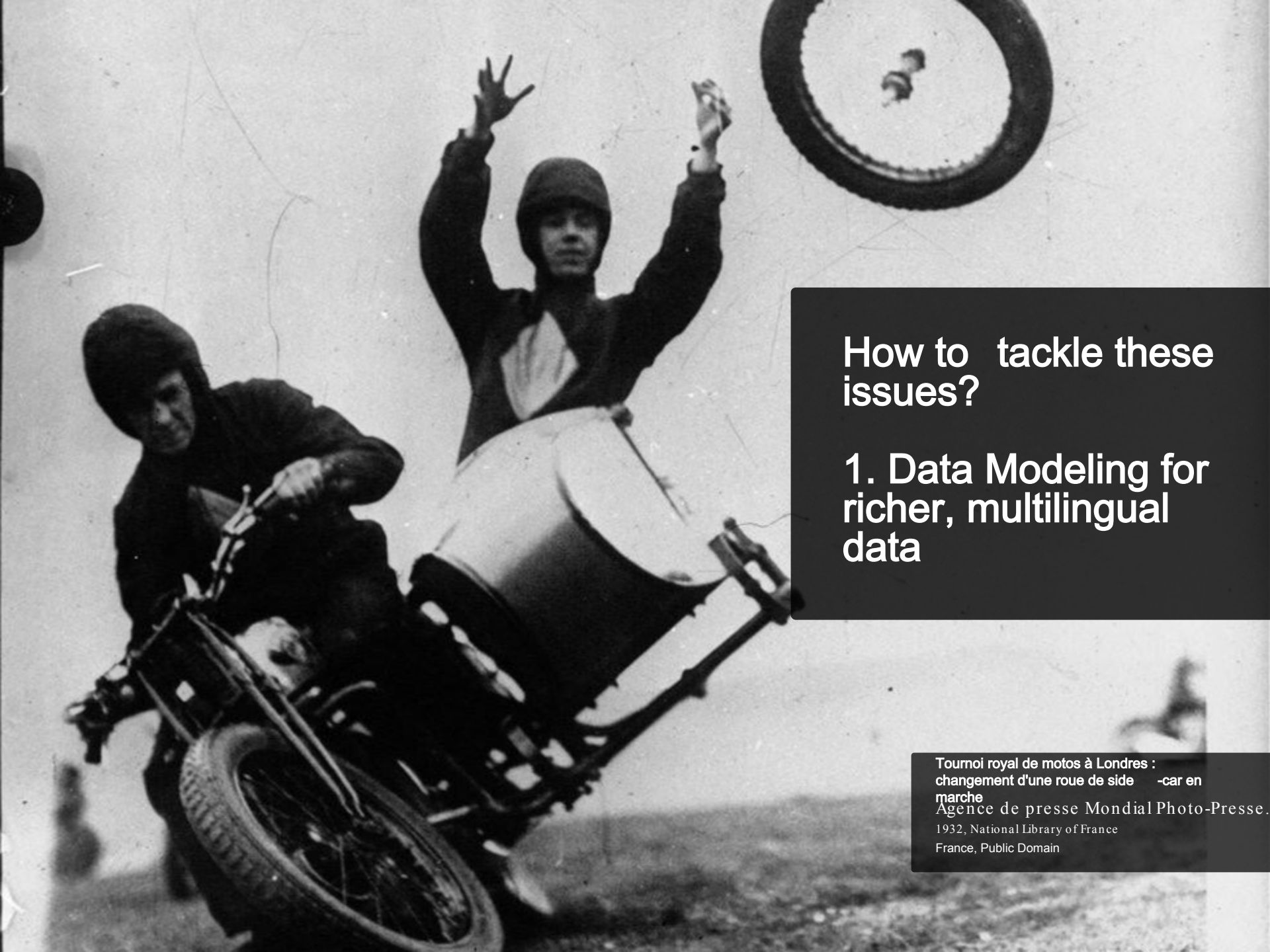Work by Péter Kiraly (Göttingen Research alliance)

europeana

388. **gez**: 12
389. **Spa**: 12
390. **efi**: 8
391. **ar-IL**: 7
392. **ar-LB**: 7
393. **la-x-medieval**: 6
394. **x-highgerm**: 6
395. **zh-latn**: 6
396. **x-aramaic-latn**: 6
397. **jut**: 5
398. **tyv**: 4
399. **RU**: 4
400. **cyr**: 4
401. **spa**: 4
402. **x-local**: 3
403. **dut**: 3
404. **ji**: 3
405. **iw**: 2
406. **x-stdgerm**: 2
407. **alb**: 2
408. **e**: 2
409. **uk-latn**: 2
410. **RS**: 2
411. **swb**: 2
412. **tr-latn**: 2
413. **gml**: 2
414. **fa-latn**: 2
415. **grc-latn**: 2
416. **tir**: 1
417. **fa-latn-x-middle**: 1
418. **dan**: 1
419. **olo**: 1
420. **phn**: 1
421. **ell**: 1
422. **src**: 1

Data analysis by Péter Kiraly (Göttingen Research alliance)

Massively multilingual metadata

- Officially we get metadata in   37 languages

- But there are more languages used in individual metadata fields

  - Over 400 language codes
    - E.g., 6 values in x-aramaic-latn - not a valid code by the way

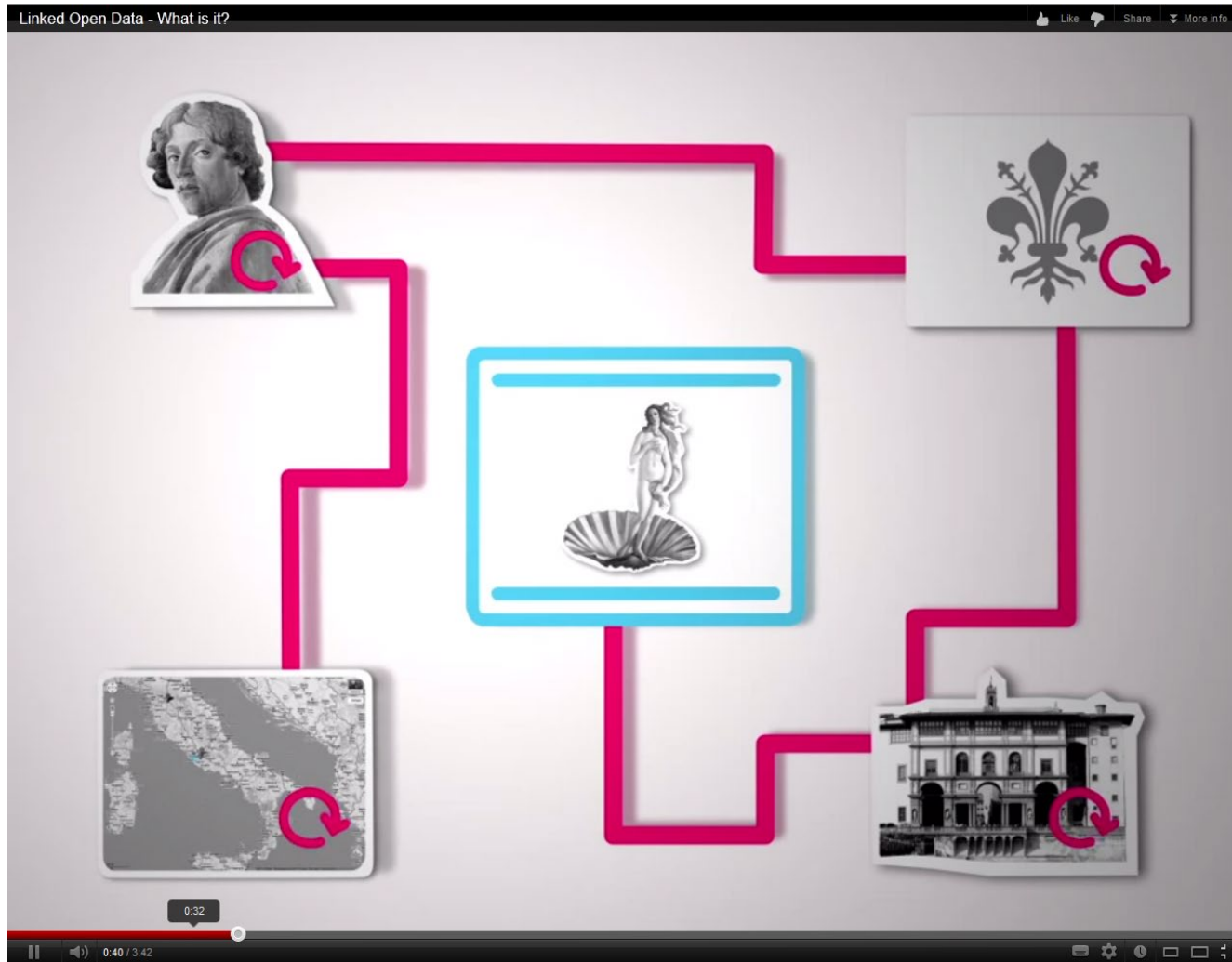 - But the most common case is lack of language information!

How to tackle these issues?

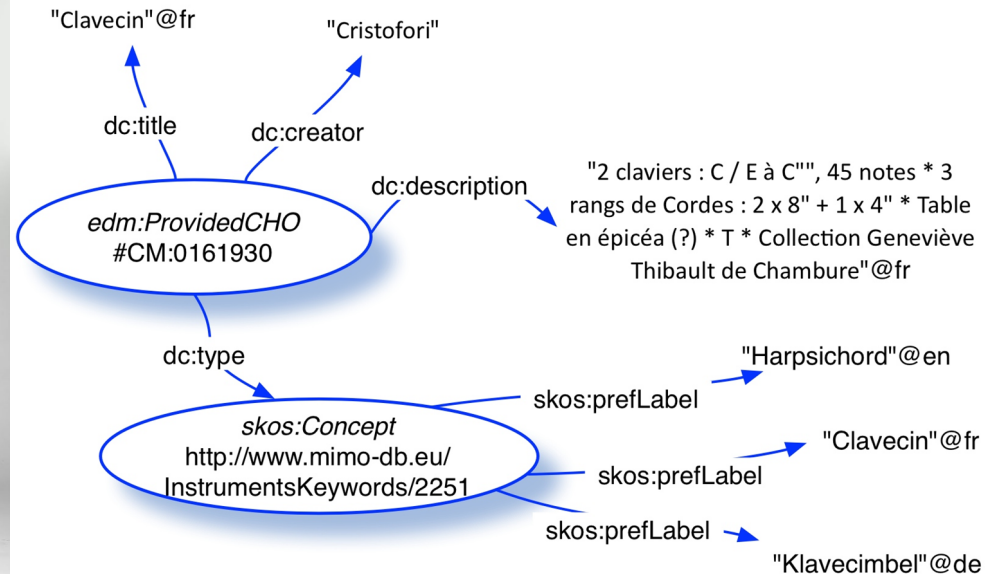# 1. Data Modeling for richer, multilingual data

Following the Linked Open Data principles

europeana

http://vimeo.com/36752317

The Europana Data Model as handle for multilingual data

Here, from the Musical Instrument Museums Online (MIMO) project

Clavecin , Bartolomeo Cristofori
Cite de la Musique,
MIMO - Musical Instruments Museums Online | CC BY-NC-SA

Europeana Data Model example

How to tackle these issues?

2. Enriching metadata

A Linked Data Strategy for Europeana

- The Europeana Data Model (EDM) offers a base for linking metadata
  - Aims at providing data as resources (with URIs!), not only strings
  - Enables the development of a multilingual data environment

- We encourage data providers to contribute their own links to vocabularies

europeana

Related to a general   effort on quality

We have set up a Data Quality Committee   working on
recommendations for the   Europeana community   on:

- ○ Mandatory metadata   elements
- ○ Metadata normalization
- ○ Multilingualism

…

europeana

A general effort on quality

Europeana Publishing Framework: Metadata

Here you see how the different metadata tiers of the Publishing Framework work.

TIER	LANGUAGE	ENABLING ELEMENTS*	CONTEXTUAL CLASSES**	BENEFITS (CUMULATIVE)
A Europeana as a basic search platform. *'I want to find a specific item that I'm looking for.'*	At least 25% of the provided EDM metadata fields that are relevant have at least 1 language qualified value	At least one enabling element taken from one of the 'Discovery scenario' groups	None	**Findability** - material is available on Europeana Collections, indexed by search engines, and linked data technology **Web traffic** - click-throughs to your site
B Europeana as an exploration platform. *'I want to browse and explore Europeana even if I'm not sure what I'm looking for.'*	At least 50% of the provided EDM metadata fields that are relevant have at least 1 language qualified value	At least three distinct elements taken from two of the 'Discovery scenario' groups	At least one contextual class with all minimum required elements, OR link to LOD vocabulary	Greater potential for: **Use in thematic collections** - providing context and relation to other collections **More marketing** through Europeana
C Europeana as a knowledge platform. *'I want to search and browse in a more precise way, by named authors, specific subjects or topics.'*	At least 75% of the provided EDM metadata fields that are relevant have at least 1 language qualified value	At least four distinct elements taken from two of the 'Discovery scenario' groups	At least two distinct contextual classes with all minimum required elements, OR links to LOD vocabularies	Greater potential for: **Use in Europeana's existing partnerships and projects**, in e.g. education and research **Use in apps and services** **Use by creative industry**

https://pro.europeana.eu/post/publishing-framework

europeana

LOD Vocabularies currently recognized by Europeana in providers' metadata

Vocabulary	URL
MIMO Concepts	http://www.mimo-db.eu/
MIMO Instrument makers	http://www.mimo-db.eu/
The Getty - Art & Architecture Thesaurus (AAT)	http://vocab.getty.edu/
The Getty - Union List of Artist Names (ULAN)	http://vocab.getty.edu/
Virtual International Authority File (VIAF)	http://viaf.org/viaf/
Geonames	http://sws.geonames.org/
IconClass	http://iconclass.org/
Gemeinsame Normdatei (GND)	http://d-nb.info/gnd
Israel Museum Jerusalem Concepts	http://www.imj.org.il/imagine/thesaurus/objects/
Partage Plus concepts	http://partage.vocnet.org/
data.europeana.eu WWI Concepts from Library of Congress Subject Headings (LCSH)	http://data.europeana.eu/concept/loc
Europeana Sounds Genres	http://data.europeana.eu/concept/soundgenres/
EAGLE Material & Object Type	http://www.eagle-network.eu/voc/
DISMARC Formats & Genres	http://purl.org/dismarc/ns/
UDC	http://udcdata.info/rdf/
UNESCO Thesaurus	http://vocabularies.unesco.org/thesaurus/

NB: providers can send data from their local vocabularies and authorities too — if they embed it in the EDM metadata they send to us!

europeana

A Linked Data Strategy for Europeana

- The Europeana Data Model (EDM) offers a base for linking metadata

  - Aims at providing data as resources (with URIs!), not only strings

  - Enables the development of a multilingual data environment

- We encourage data providers to contribute their own links to vocabularies

- In parallel, we apply automatic enrichment to link object metadata to reference datasets

A Linked Data Strategy for Europeana – Contextual Entities

We are building an "Entity Collection"

- Centralized point of reference and access to data about contextual entities: places, agents (persons and organizations), concepts...

- Caching and curating data from the wider Linked Open Data cloud
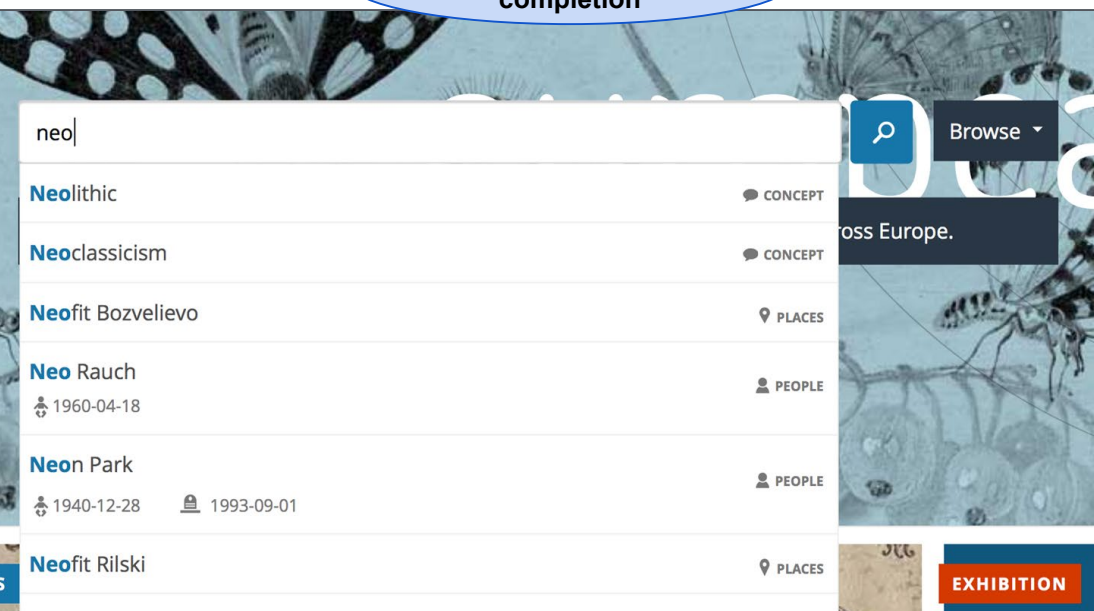
- A sort of Europeana "knowledge   graph"

Uses cases for the Entity　Collection

Improve user experience on Europeana services

- *Findability*: users can search with and for people, places and subjects, not only objects. In many more languages, and with less ambiguity

- *Contextualization*: users see contextual information about cultural heritage objects. Entity Pages group and present all assertions about an entity

…

Semantic auto-completion

Entity Pages

Data currently in the Entity Collection

216,302 resources

- *Places*
  *a subset of **Geonames**, corresponding to places which are part of European countries and of some specific feature classes.*

165,005 resources

- *Agents*
  *a subset of **DBpedia** corresponding to most of the instances of dbp:Artist with some exceptions, and integrated from 49 DBpedia language editions.*

1,572 resources

- *Concepts*
  *a subset of **DBpedia** corresponding to a selection concepts matching the needs from Europeana Collections (e.g., WWI battles).*
  *Europeana Sounds music genres (obtained from **Wikidata**)*
  Photo Consortium's photography vocabulary

1,077 resources

- *Organizations*
  *Extracted from Europeana's CRM and aligned to **Wikidata** when possible*

Selecting data sources

- **Availability and access** : open license, published on the web as linked data

- **Granularity, size and coverage** : multilingual data, helping to answer key user needs for Europeana's CH collections. Too generic or large datasets can create too much ambiguity for the simple processes we have (e.g. enrichment)

- **Quality** : intrinsic aspects like correctness of representation (data structures)

- **Connectivity** : good data sources are well -connected internally and externally to other datasets

An example
DBpedia resource for "Mozart" in our data

```
{
    "@context": http://data.europeana.eu/entity/context.json,
    id: http://data.europeana.eu/agent/base/146951,
    type: "Agent",
  ▼ prefLabel: {
        ar: "فولفغانغ أماديوس موتسارت",
        az: "Volfqanq Amadey Mosart",
        be: "Вольфганг Амадэй Моцарт",
        bg: "Волфганг Амадеус Моцарт",
        bs: "Wolfgang Amadeus Mozart",
        ca: "Wolfgang Amadeus Mozart",
        cs: "Wolfgang Amadeus Mozart",
        cy: "Wolfgang Amadeus Mozart",
        da: "Wolfgang Amadeus Mozart",
        de: "Wolfgang Amadeus Mozart",
        el: "Βόλφγκανγκ Αμαντέους Μότσαρτ",
        en: "Wolfgang Amadeus Mozart",
        es: "Wolfgang Amadeus Mozart",
        et: "Wolfgang Amadeus Mozart",
        eu: "Wolfgang Amadeus Mozart",
        fi: "Wolfgang Amadeus Mozart",
        fr: "Wolfgang Amadeus Mozart",
        ga: "Wolfgang Amadeus Mozart",
        gd: "Wolfgang Amadeus Mozart",
        gl: "Wolfgang Amadeus Mozart",
        he: "וולפגנג אמדאוס מוצרט",
        hi: "वोल्फ़गांक आमदेयुस मोसार्ट",
        hr: "Wolfgang Amadeus Mozart",
        hu: "Wolfgang Amadeus Mozart",
        hy: "Վոլֆգանգ Ամադեուս Մոցարտ",
        is: "Wolfgang Amadeus Mozart",
        it: "Wolfgang Amadeus Mozart",
        ja: "ヴォルフガング・アマデウス・モーツァルト",
        ka: "ვოლფგანგ ამადეუს მოცარტი",
        ko: "볼프강 아마데우스 모차르트",
        lt: "Volfgangas Amadėjus Mocartas",
        lv: "Volfgangs Amadejs Mocarts",
        mk: "Волфганг Амадеус Моцарт",
        nl: "Wolfgang Amadeus Mozart",
        no: "Wolfgang Amadeus Mozart",
        pl: "Wolfgang Amadeusz Mozart",
        pt: "Wolfgang Amadeus Mozart",
        ro: "Wolfgang Amadeus Mozart",
        ru: "Моцарт, Вольфганг Амадей",
        sk: "Wolfgang Amadeus Mozart",
        sl: "Wolfgang Amadeus Mozart",
        sq: "Wolfgang Amadeus Mozart",
        sr: "Волфганг Амадеус Моцарт",
        sv: "Wolfgang Amadeus Mozart",
        tr: "Wolfgang Amadeus Mozart",
        uk: "Вольфганг Амадей Моцарт",
        yi: "וואלפגאנג אמאדעוס מאצארט",
        zh: "沃尔夫冈·阿马德乌斯·莫扎特"
    },
  ▼ altLabel: {
        en: "Mozart, Johann Chrysostom Wolfgang Amadeus",
        "en.en": "Mozart, Wolfgang Amadeus"
    },
  ▶ biographicalInformation: […],
  ▼ dateOfBirth: [
        "1756-01-27"
    ],
  ▼ dateOfDeath: [
        "1791-12-05"
    ],
  ▼ end: [
        "1791-12-05"
    ],
  ▼ identifier: [
        "32197206"
    ],
  ▼ placeOfBirth: [
      ▼ {
            "@id": http://dbpedia.org/resource/Salzburg
        },
      ▼ {
            "@id": http://dbpedia.org/resource/Austria
        }
    ],
  ▼ placeOfDeath: [
      ▼ {
            "@id": http://dbpedia.org/resource
        },
      ▼ {
            "@id": http://dbpedia.org/resou
        }
    ],
  ▼ sameAs: [
        http://sw.dbpedia.org/resource/Wolfgang_Amadeus_Mozart,
        http://ro.dbpedia.org/resource/Wolfgang_Amadeus_Mozart,
        http://bg.dbpedia.org/resource/Волфганг_Амадеус_Моцарт,
        http://nds.dbpedia.org/resource/Wolfgang_Amadeus_Mozart,
        http://pa.dbpedia.org/resource/ਵੋਲਫ਼ਗਾਂਗ_ਆਮਾਡੇਅਸ_ਮੋਜ਼ਾਰਟ,
        http://oc.dbpedia.org/resource/Wolfgang_Amadeus_Mozart,
```

Preferred labels for 48 languages

Inter-linking information

Coreference links to 6 other datasets
(e.g. Freebase, Wikidata)

europeana

An enrichment example

Search by keyword here 🔍

We want your feedback on our new item page, use our feedback button to leave your comments.

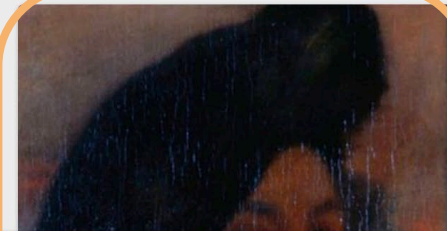## Zittende jongeman bij ton met het hoofd naar links

Drawing

Drawing is a form of visual art that makes use of any number of drawing instruments to mark a two-dimensional medium. Instruments used include graphite pencils, pen and ink, inked brushes, wax...

View more

Type of this object

Links to contextual entities

Created by

Johannes Vermeer

Johannes Vermeer

Johannes, Jan or Johan Vermeer (Dutch: [joˈɦɑnəs jɑn vərˈmeːr]; 1632 – December 1675) was a Dutch painter who specialized in domestic interior scenes of middle-class life. Vermeer was a moderately...

Anoniem (17e eeuw), Vermeer, Johannes (1632 - 1675), tekenaar, oude toeschrijving tekenaar

FEEDBACK

And what it allows

🏠 Return to Home / **Item**

We want your feedback on our new item page, use our feedback button to leave your comments.

Zittende jongeman bij ton met het hoofd naar links

Drawing

Drawing is a form of visual art that makes use of any number of drawing instruments to mark a two-dimensional medium. Instruments used include graphite pencils, pen and ink, inked brushes, wax...

View more

Type of this object

Johannes Vermeer

Johannes, Jan or Johan Vermeer (Dutch: [joˈɦɑnəs jɑn vərˈmeːr]; 1632 – December 1675) was a Dutch painter who specialized in domestic interior scenes of middle-class life. Vermeer was a moderately...

Created by

Johannes Vermeer

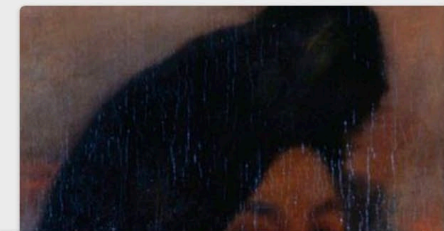Anoniem (17e eeuw), Vermeer, Johannes (1632 - 1675), tekenaar, oude toeschrijving tekenaar

FEEDBACK

And what it allows

We want your feedback on our new item page, use our feedback button to leave your comments.

Zittende jongeman bij ton met het hoofd naar links

Zeichnung (Kunst)

Eine Zeichnung ist ein Bild, das ein Motiv in unterschiedlich stark vereinfachender Weise mit Linien und Strichen darstellt. Das unterscheidet die Zeichnung von der Malerei, die ein Motiv durch den...
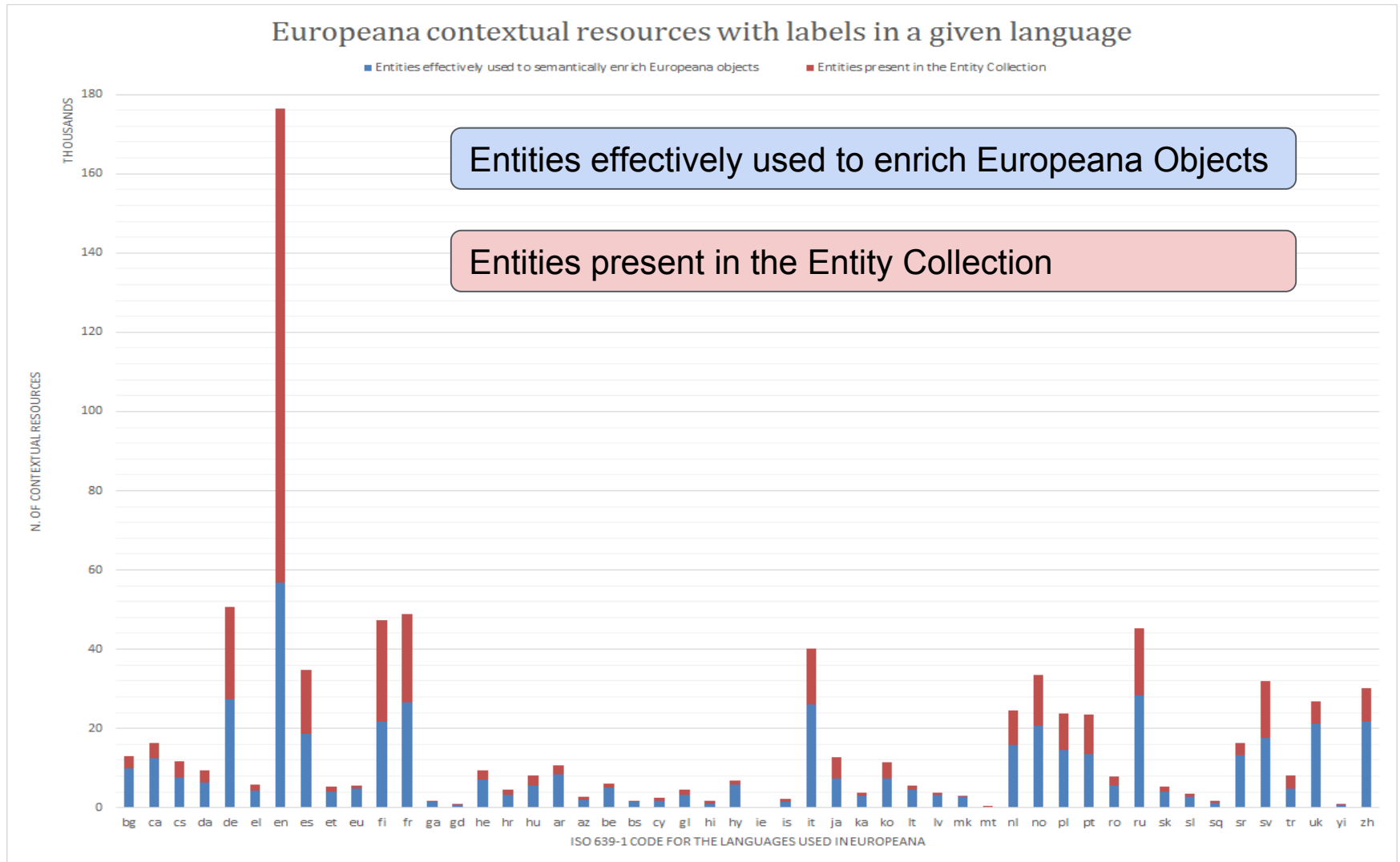
Mehr anzeigen

Type of this object

Created by

Jan Vermeer

Jan Vermeer

Jan Vermeer van Delft (getauft 31. Oktober 1632 in Delft; begraben 15. Dezember 1675 in Delft; zeitgenössisch: Joannis ver Meer, Joannis van der Meer) ist einer der bekanntesten holländischen Maler...

Anoniem (17e eeuw), Vermeer, Johannes (1632 - 1675), tekenaar, oude toeschrijving tekenaar

FEEDBACK

The Entity Collection's contribution to multilingual coverage

Europeana contextual resources with labels in a given language

■ Entities effectively used to semantically enrich Europeana objects   ■ Entities present in the Entity Collection

Entities effectively used to enrich Europeana Objects

Entities present in the Entity Collection

ISO 639-1 CODE FOR THE LANGUAGES USED IN EUROPEANA

Multilingual   enrichment is not   easy!

Inde
Lire | Modifier | Plus ▾ | Rechercher 🔍

🔖 *Pour les articles homonymes, voir Inde (homonymie),
Sous-continent indien et Plaque indienne.*

L'**Inde**, en forme
longue la **République
de l'Inde** (en hindi
भारत (*Bharat*) et भारत
गणराज्य (*Bharat
Gaṇarajya*), en anglais
India et *Republic of
India*) est un pays du
sud de l'Asie qui
occupe la majeure
partie du
sous-continent indien.

L'Inde est le
deuxième pays le plus
peuplé et le septième
pays le plus grand du
monde. Le littoral
indien s'étend sur plus
de sept mille
kilomètres. Le pays a
des frontières
communes avec le

République de l'Inde
भारत गणराज्य (**hi**) 🔊 Écouter
Republic of India (**en**)

Drapeau de l'Inde

सत्यमेव जयते
Emblème de l'Inde

Inde
Skatīt | Labot | Vairāk ▾ | Meklēt 🔍

Vikipēdijas lapa

Inde ir viela vai vielu maisījums,
kas jau mazās devās nāvējoši
iedarbojas uz dzīvu organismu.

🧪 *Šis ar ķīmiju saistītais
raksts ir nepilnīgs. Jūs
varat dot savu ieguldījumu
Vikipēdijā, papildinot to.*

🦋 *Šis ar bioloģiju saistītais
raksts ir nepilnīgs. Jūs
varat dot savu ieguldījumu
Vikipēdijā, papildinot to.*

ES standartos noteiktais apzīmējums
toksiskām vielām

Kategorijas: Nepabeigti raksti par ķīmiju | Nepabeigti raksti par bioloģiju
| Toksikoloģija

*Poisonous India or the Importance of a Semantic and Multilingual
Enrichment Strategy*
Marlies Olensky, Juliane Stiller, Evelyn Dröge, MTSR 2012
http://link.springer.com/chapter/10.1007%2F978-3-642-35233-
1_25

We're not really finished!

Concours de cycles nautiques sur le lac d'Enghien : Berregent piloté par Austerling

Agence de presse Meurisse

1914, National Library of France
France, Public Domain

Work in progress – the Entity Collection

- We've made enough progress to release a first version of the Entity Collection used in Europeana's production services.

- But there are still challenges:
  - Expand data coverage with new data sources
    - For existing types of contextual entities
    - For new types, e.g., events
  - Better enrich Europeana object metadata

Work in progress - UI

UNSERE SEITEN

🔤 SPRACHE ▾

Hier nach Stichwort suchen 🔍

We want your feedback on our new item page, use our feedback button to leave your comments.

Zittende jongeman bij ton met het hoofd naar links

Zeichnung (Kunst)

Eine Zeichnung ist ein Bild, das ein Motiv in unterschiedlich stark vereinfachender Weise mit Linien und Strichen darstellt. Das unterscheidet die Zeichnung von der Malerei, die ein Motiv durch den...

Mehr anzeigen

Type of this object

Created by

Jan Vermeer

Anoniem (17e eeuw), Vermeer, Johannes (1632 - 1675), tekenaar, oude toeschrijving tekenaar

Jan Vermeer

Jan Vermeer van Delft (getauft 31. Oktober 1632 in Delft; begraben 15. Dezember 1675 in Delft; zeitgenössisch: Joannis ver Meer, Joannis van der Meer) ist einer der bekanntesten holländischen Maler...

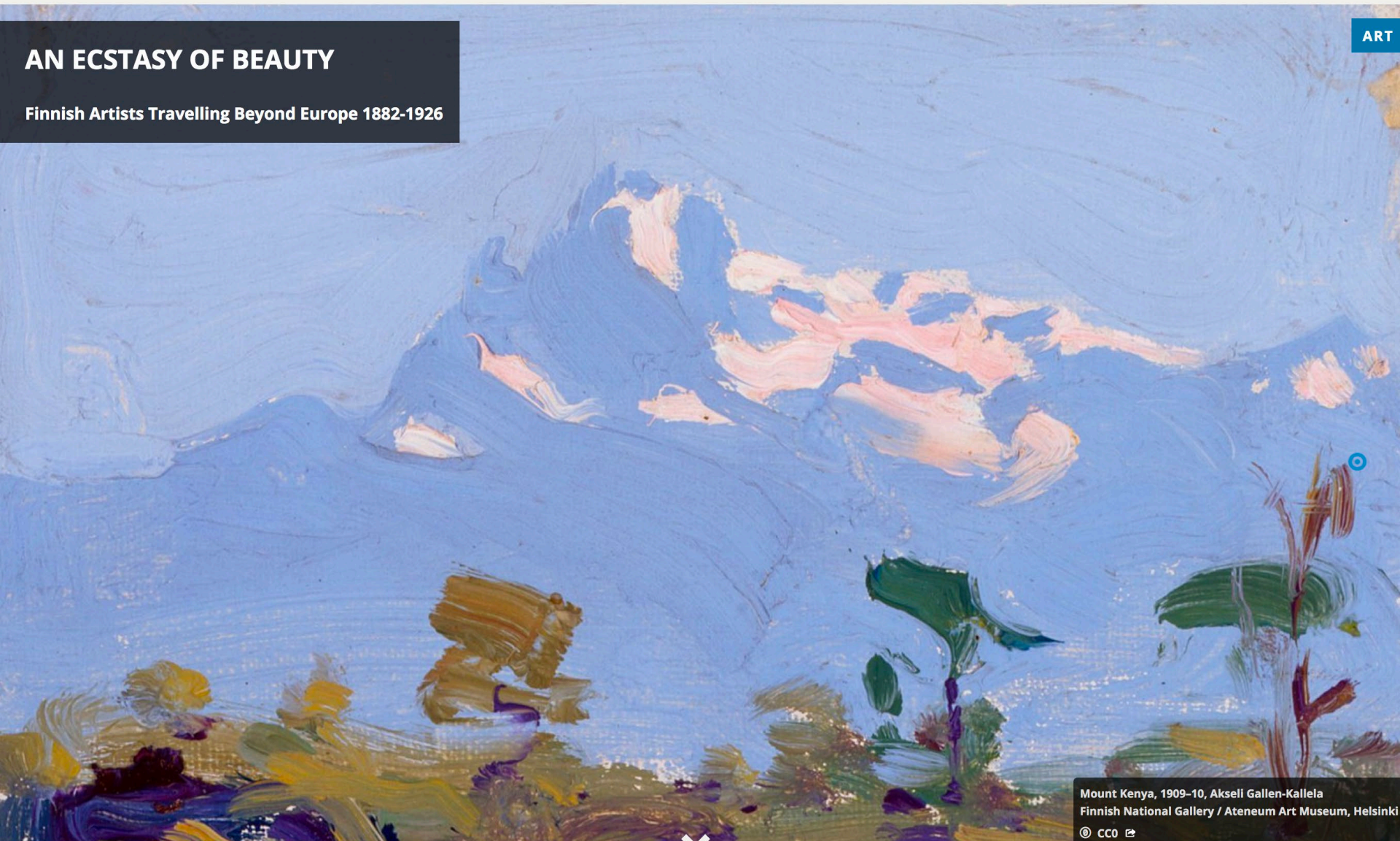FEEDBACK

Work in progress – Static content

OUR SITES

LANGUAGE ▾

Suomi

ART

AN ECSTASY OF BEAUTY

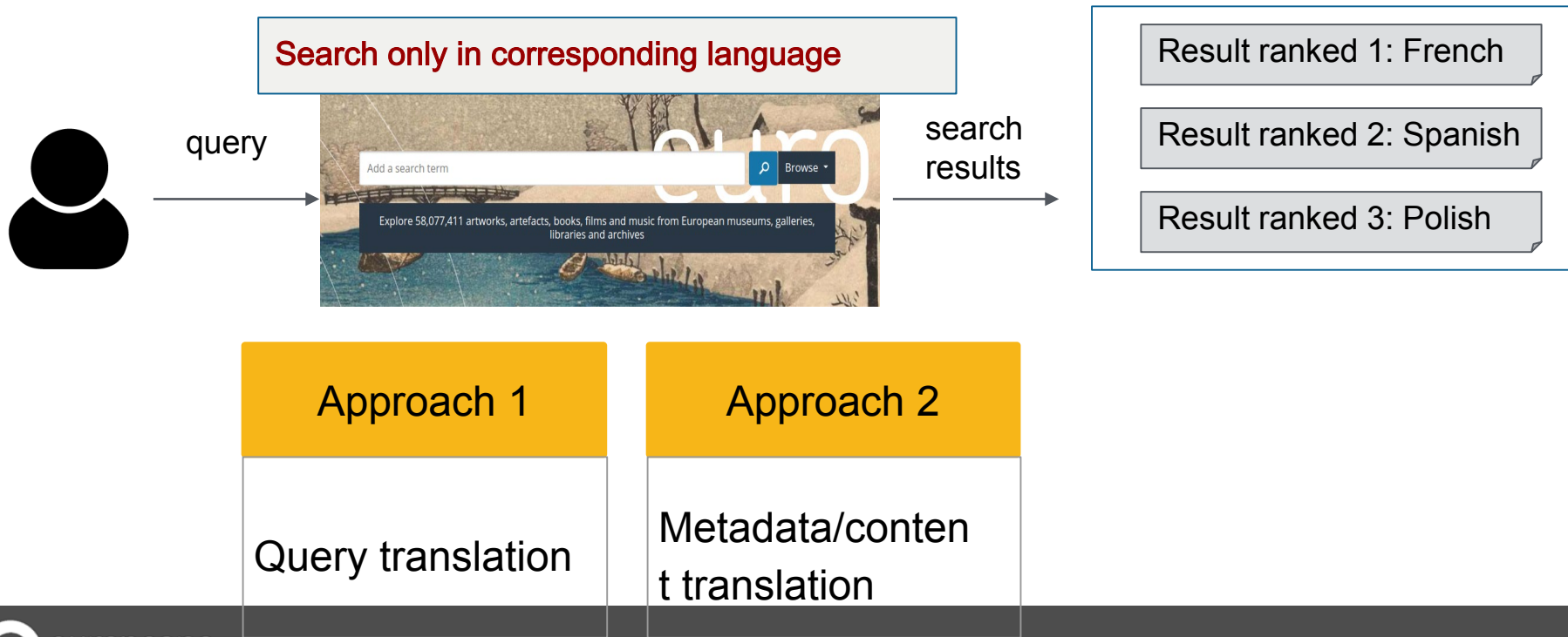**Finnish Artists Travelling Beyond Europe 1882-1926**

Mount Kenya, 1909–10, Akseli Gallen-Kallela
Finnish National Gallery / Ateneum Art Museum, Helsinki

Ⓜ CC0

Automatic translation?

- First experiments with the European Commission's eTranslation service, on curated (static) content
- So far it needs to be better quality, or applied to other Europeana areas, like search and object display

Search only in corresponding language

query

search results

Add a search term

Explore 58,077,411 artworks, artefacts, books, films and music from European museums, galleries, libraries and archives

Result ranked 1: French

Result ranked 2: Spanish

Result ranked 3: Polish

Approach 1	Approach 2
Query translation	Metadata/content translation

europeana

Flowers
Anonymous
1700-1799, Rijksmuseum
Netherlands, Public Domain

europeana
#AllezCulture

antoine.isaac@europeana.eu
@antoine_isaac

europeana.eu   @EuropeanaEU

Co-financed by the European Union
Connecting Europe Facility