



europeana
think culture

Multilingual challenges for accessing digitized culture online

Antoine Isaac

Multilingual Digital Single Market Riga Summit, 28 April 2015



Europe's platform to access cultural heritage

Home My Europeana Choose a language

Search All fields Search Help

New exhibition : Leaving Europe
Why did 30 million Europeans emigrate to America in the 19th/20th centuries? This exhibition of rarely seen pictures tells their story.
[Open Exhibition](#)

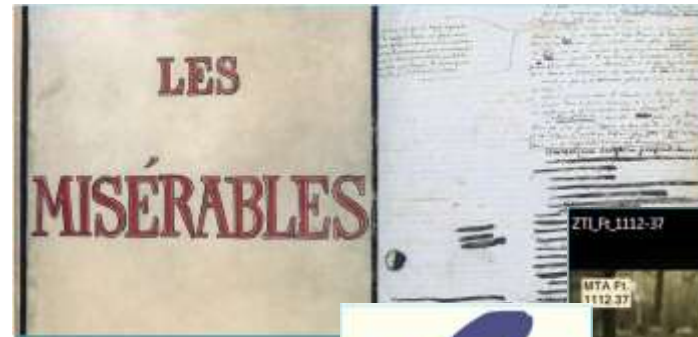
From the blog

Marseille, the Capital of Culture and a Giraffe
This weekend, Marseille celebrates its status as a European Capital of Culture. There are parties and events all weekend, so if you're in the south

Poisonous Nature
Did you know that potatoes are poisonous? Or holy? Or centipedes? 'Poisonous Nature' is an online exhibition from a Europeana partner.

Featured Item
Church of the Assumption
The church of Virgin Mary's Assumption is located at the district of Panagia (Virgin Mary), in the old city of Kavala, between the

Featured Partner
Hellenic Educational Television
Hellenic Educational Television aims to provide educators and students with the opportunity of using available video content. Through Europeana explore over 600 educational videos from Hellenic

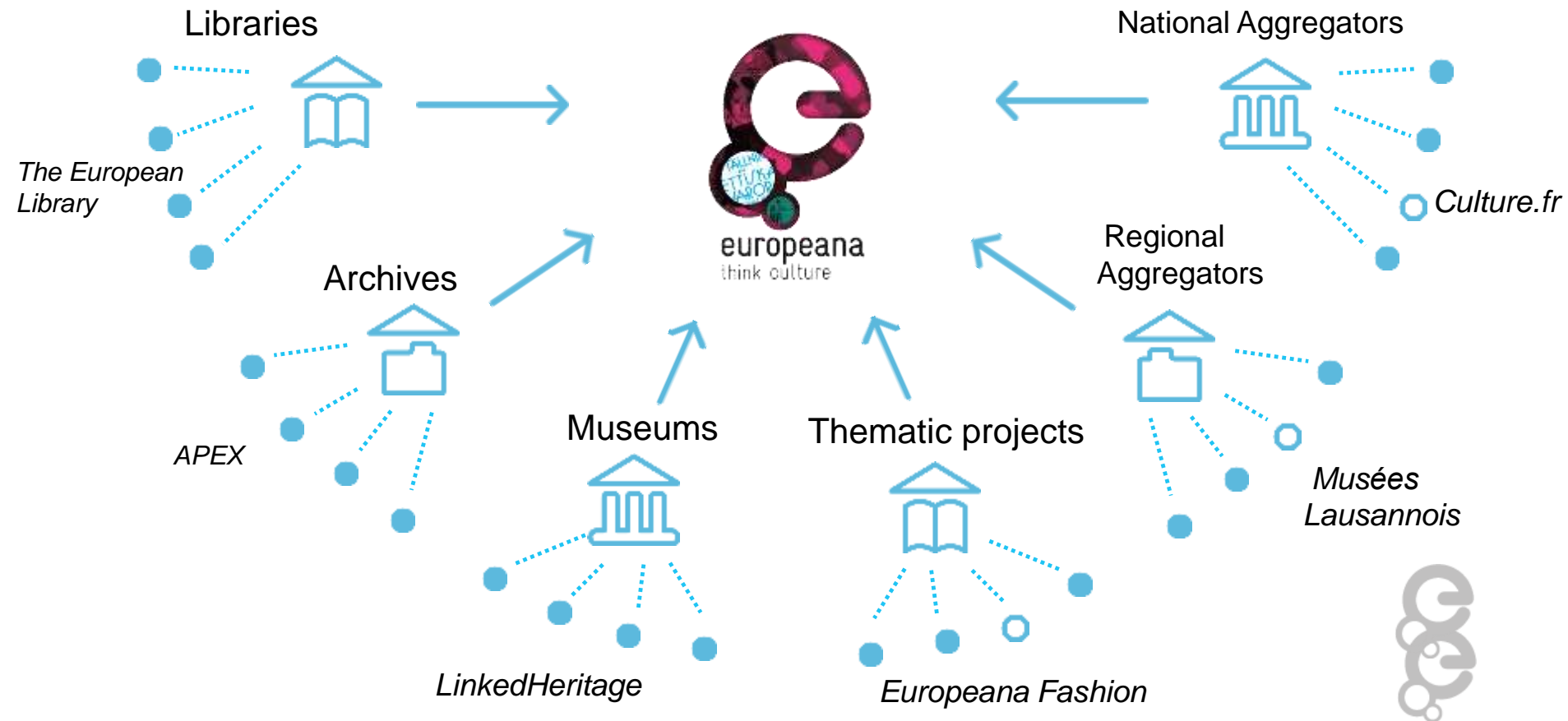


42M objects and counting

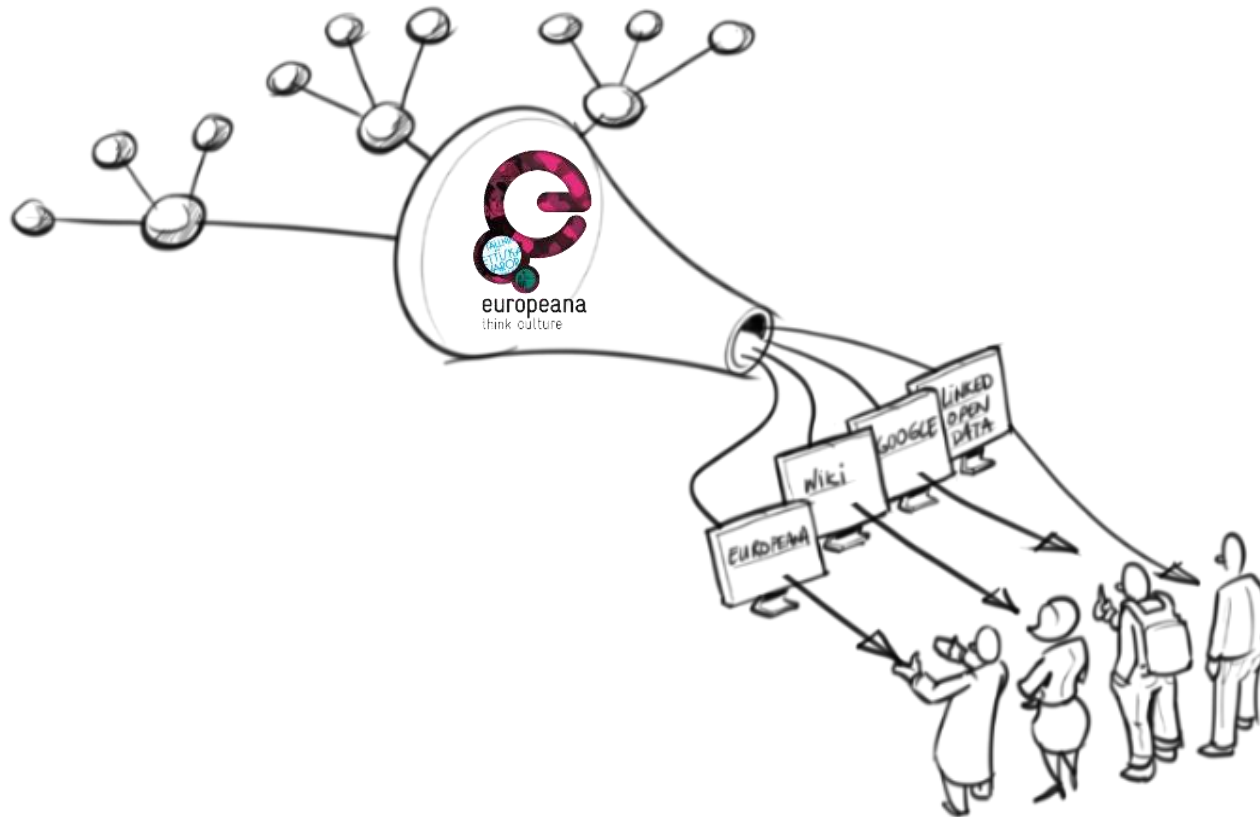


Built on metadata from a broad network

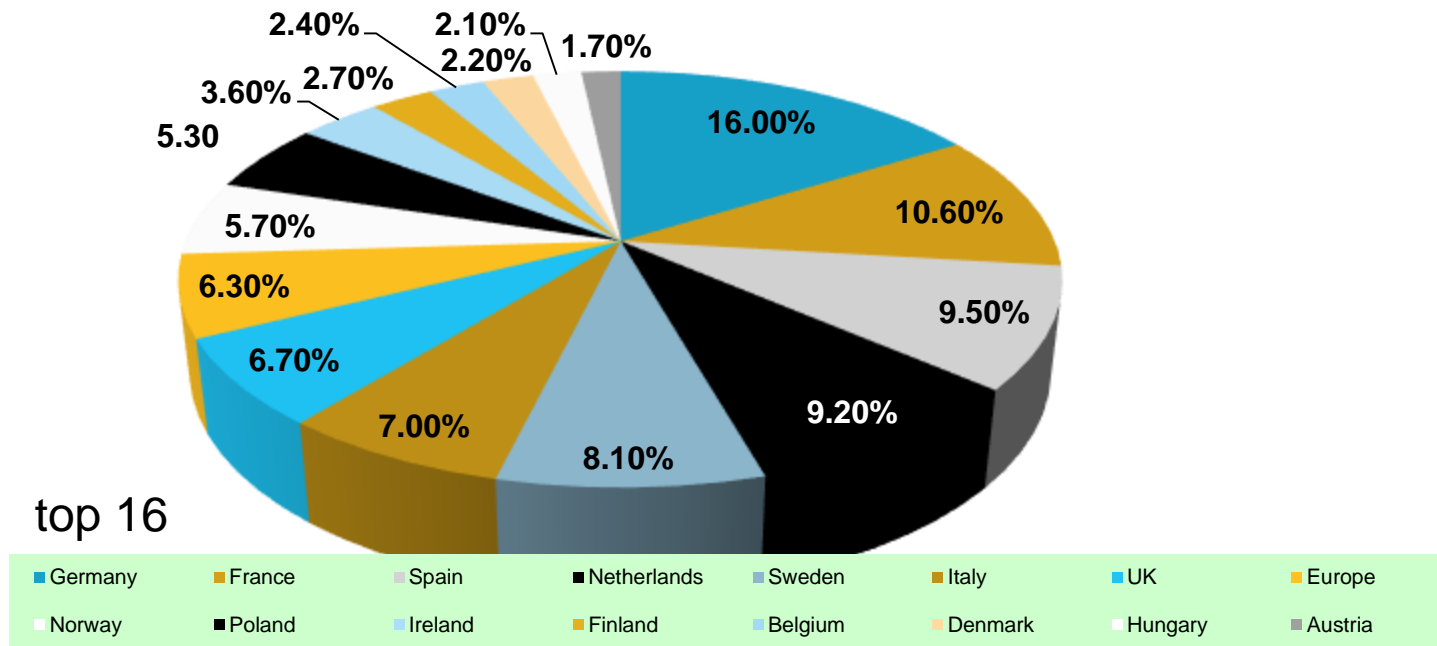
2,300 museums, archives and libraries



A platform for access and re-use



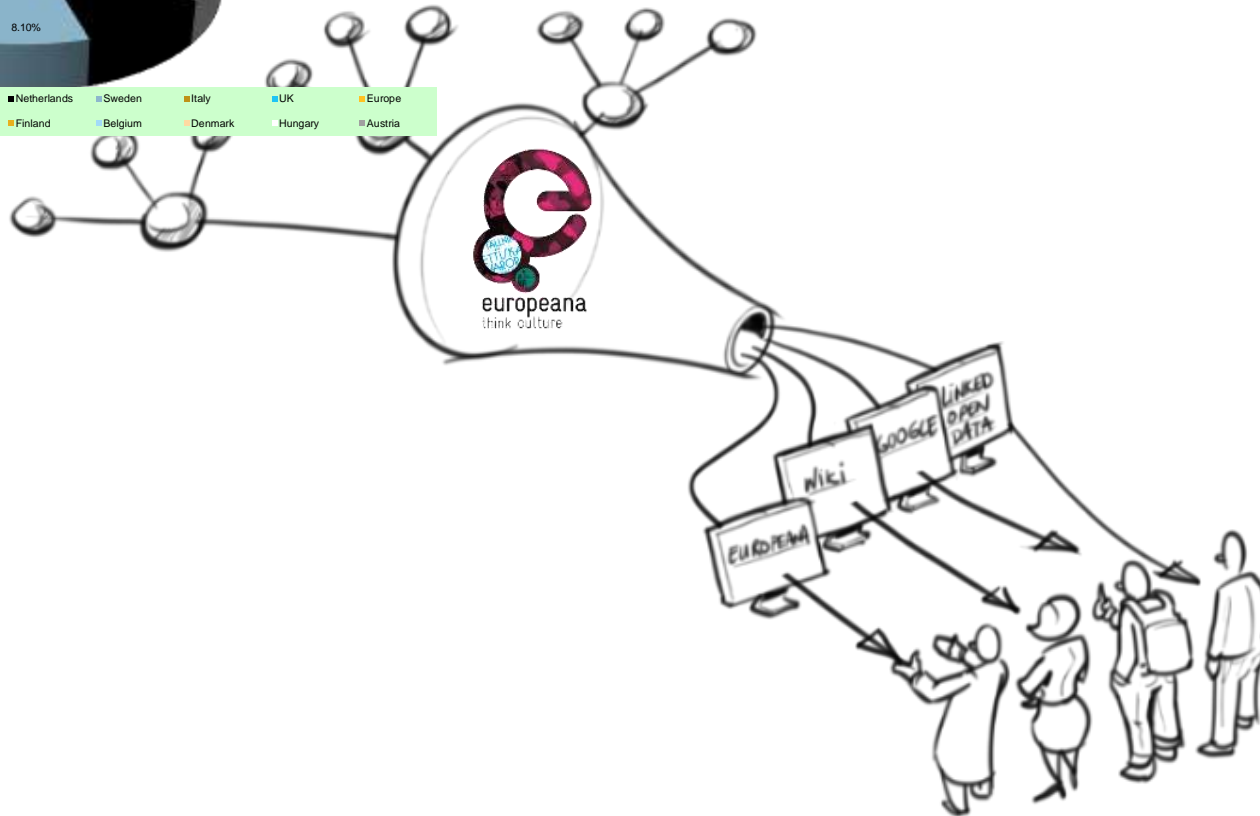
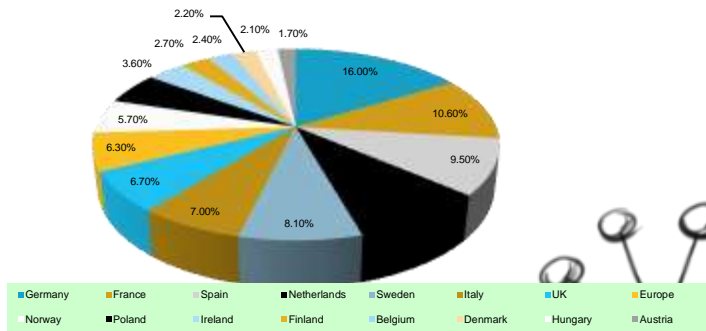
Data from all over EU



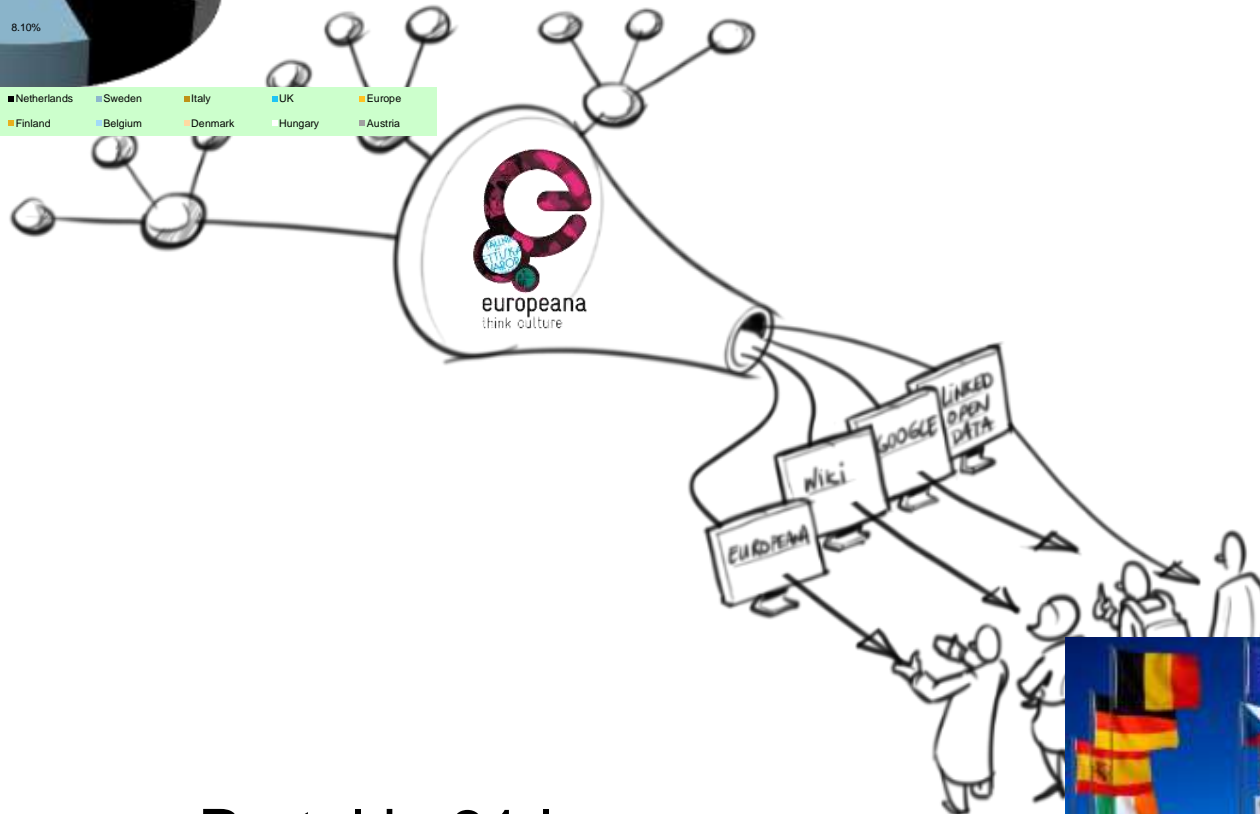
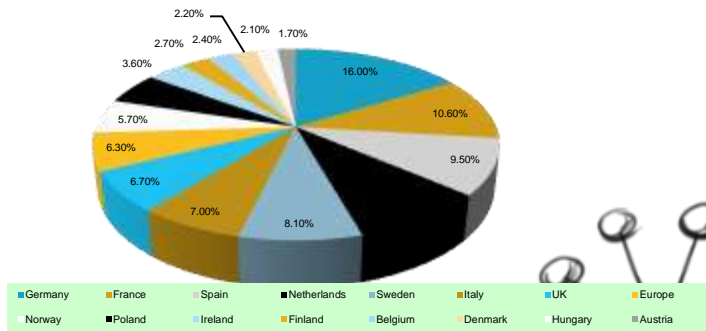
Items from 36 countries
Metadata in 33 languages



Data from all over EU



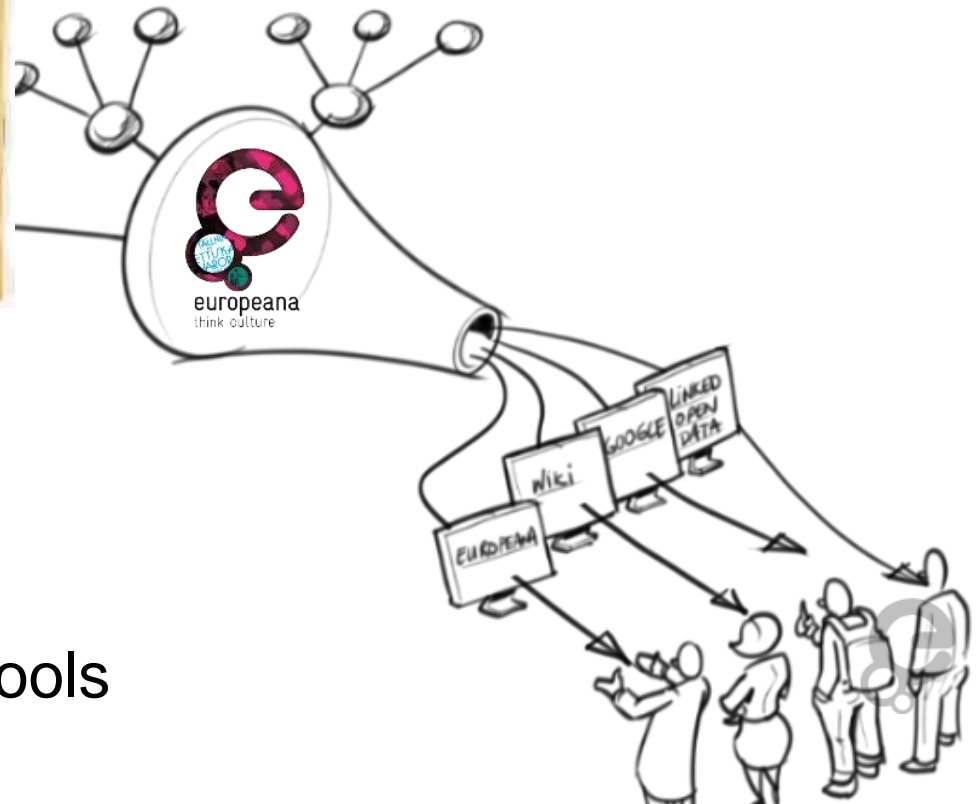
(Re-)users from all over EU



Portal in 31 languages
Data access from anywhere

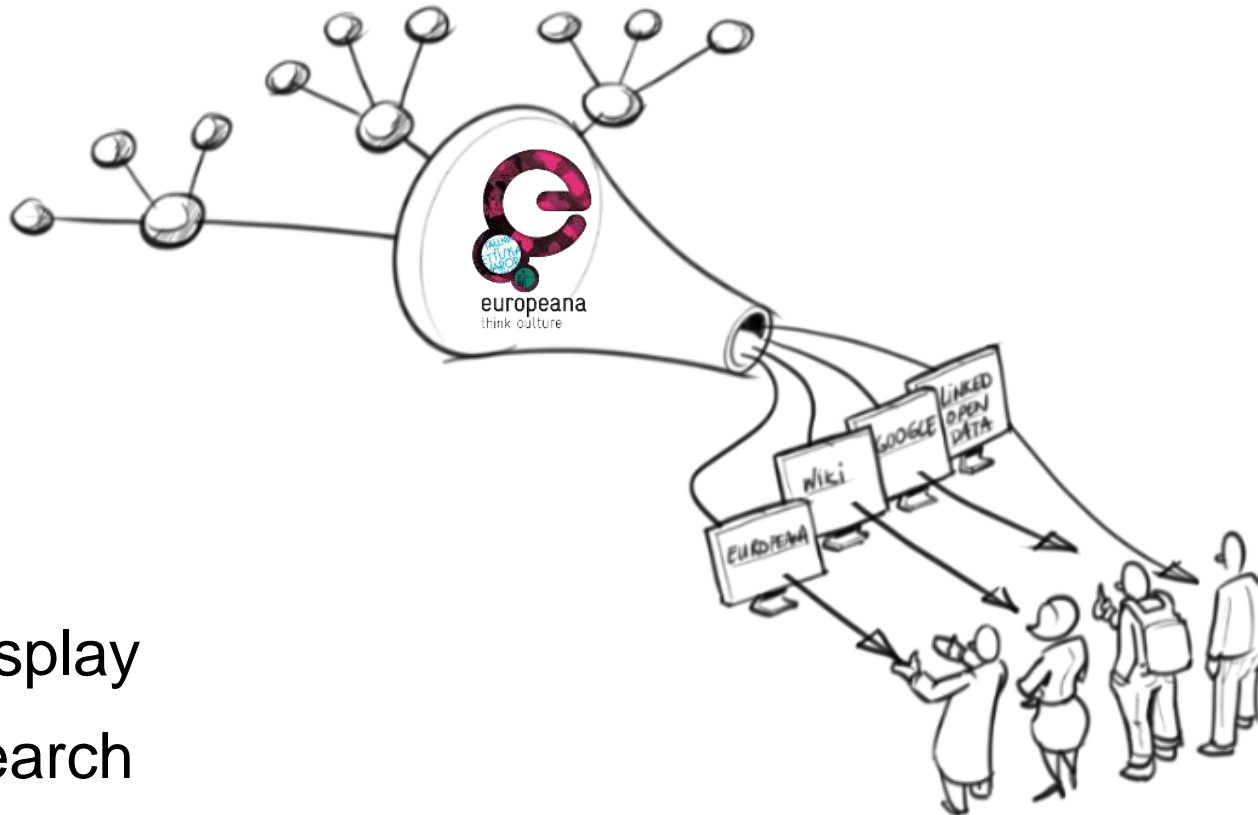


Diversity of domains makes it hard



- Data is heterogeneous
- Makes it harder to train tools

Diversity of applications makes it harder



- Display
- Search
- Data services

.....




Some issues for applying language technology for online culture

- Typical queries are very **short**
- **Identification of query language** is not easy, even manually
- Plenty of **named entities** (persons, places...)



Progress and issues – data enrichment

Return to search results < Previous Next >



© Free access - no re-use

View item at
[National Library of the Netherlands - Koninklijke Bibliotheek](#)

Share

Cite

Translate

Select language

Powered by

De koppelaarster

Description: Litt. : Christopher Braider: Refiguring the real. Picture and modernity in word and image. 1400-1700. Princeton, 1993

Creator: Johannes Vermeer ;
http://dbpedia.org/resource/Johannes_Vermeer

Contributor: Johannes Vermeer

Date of creation: 16XX

Type: StillImage

Format: schilderij ; image/jpeg

Subject: koppelen

Identifier:
<http://resolver.kb.nl/resolve?urn=urn:gvn:NCRD01:107484099>

Providing country: Netherlands

Search also for:

Title
[De koppelaarster](#) (7)

Who
[Johannes Vermeer](#) (8)
http://dbpedia.org/resource/Johannes_Vermeer (3)
[jan vermeer](#) (5)
[扬·弗美尔](#) (3)
[jan vermeer](#) (5)
[johannes vermeer](#) (8)
[jan vermeer](#) (5)
[johannes vermeer](#) (8)
[johannes vermeer](#) (8)
[johannes vermeer](#) (8)
[вермеер, ян](#) (3)
[johannes vermeer](#) (8)
[johannes vermeer](#) (8)

What

[Auto-generated tags](#) ▾

[Who](#) ▾

Agent Term: http://dbpedia.org/resource/Johannes_Vermeer

Agent Label: [jan vermeer] (de) ; [扬·弗美尔] (zh) ; [jan vermeer] (it) ; [johannes vermeer] (pt) ; [jan vermeer] (pl) ; [johannes vermeer] (sv) ; [johannes vermeer] (fr) ; [johannes vermeer] (en) ; [вермеер, ян] (ru) ; [johannes vermeer] (es) ; [johannes vermeer] (nl)

[Auto-generated tags](#) ▾

[Who](#) ▾

Problems of multilingual enrichment

Art

Read

View source

More ▾

Search



From Wikipedia, the free encyclopedia

This article is about the general concept of art. For the group of creative disciplines, see [The arts](#). For other uses, see [Art \(disambiguation\)](#).

Art is a diverse range of human activities and the products of those activities, usually involving imaginative or technical skill. In their most general form these activities include the production of works of art, the criticism of art, the study of the history of art, and the aesthetic dissemination of art. This article focuses



Clockwise from upper left: a self-portrait by Vincent van Gogh; a female ancestor figure by a Chokwe artist; detail from the *Birth of Venus* by Sandro Botticelli; and a Japanese Shisa lion.

Art

Minn Wikipedija, l-enciklopedija l-hielsa

*Dan l-artiklu huwa dwar il-pjaneta **Art** . Jekk qed tfittex informazzjoni dwar Art f'sinifikati ohra, mur [Art \(Disambigwitatiet\)](#).*

Art



Karatteristiċi fiżiċi

Raġġ tal-Ekwatur	6.378,14 km
Raġġ polari	6.356,78 km
Ċirkonferenza Ekwatorjali	40.075 km
Volum	$1,0832 \times 10^{12} \text{ km}^3$
Massa	$5,9737 \times 10^{24} \text{ kg}$
Densità	$5,515 \text{ g/cm}^3$
Area tas-Superfiċi	510.065.700 km ²
Gravità tal-Ekwatur	$9,766 \text{ m/s}^2$ (=1 g)
Velocità tal-Harba	11.180 m/s

Problems of multilingual enrichment

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 Gaṇarājya*), 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 dzīvū *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%2F978-3-642-3523>

1_25

Exploiting multilingual data from partners



Instrument Keywords / Hornbostel & Sachs Classification

Harpsichord

Translations

- German : [Cembalo](#)
- Swedish : [Cembalo](#)
- French : [Clavecin](#)
- Italian : [Clavicembalo](#)
- English : [Harpsichord](#)
- Dutch : [Klavecimbel](#)



Problems need pooling resources

- From cultural heritage sector
- From language technology industry

- EU level: Horizon2020 projects, DSIs
 - MT@EU
- National level
 - Cooperation Tilde & Latvian Ministry of Culture



Being open

- Open source and open data, as much as possible!
- Sharing the word about problems and solutions
EuropeanaTech Insight first issue on multilingualism





europaena
think culture

Thank you!

Antoine Isaac

antoine.isaac@europaena.eu

[@EuropeanaEU](#)

