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

42M objects and counting

Built on metadata from a broad network

2,300 museums, archives and libraries
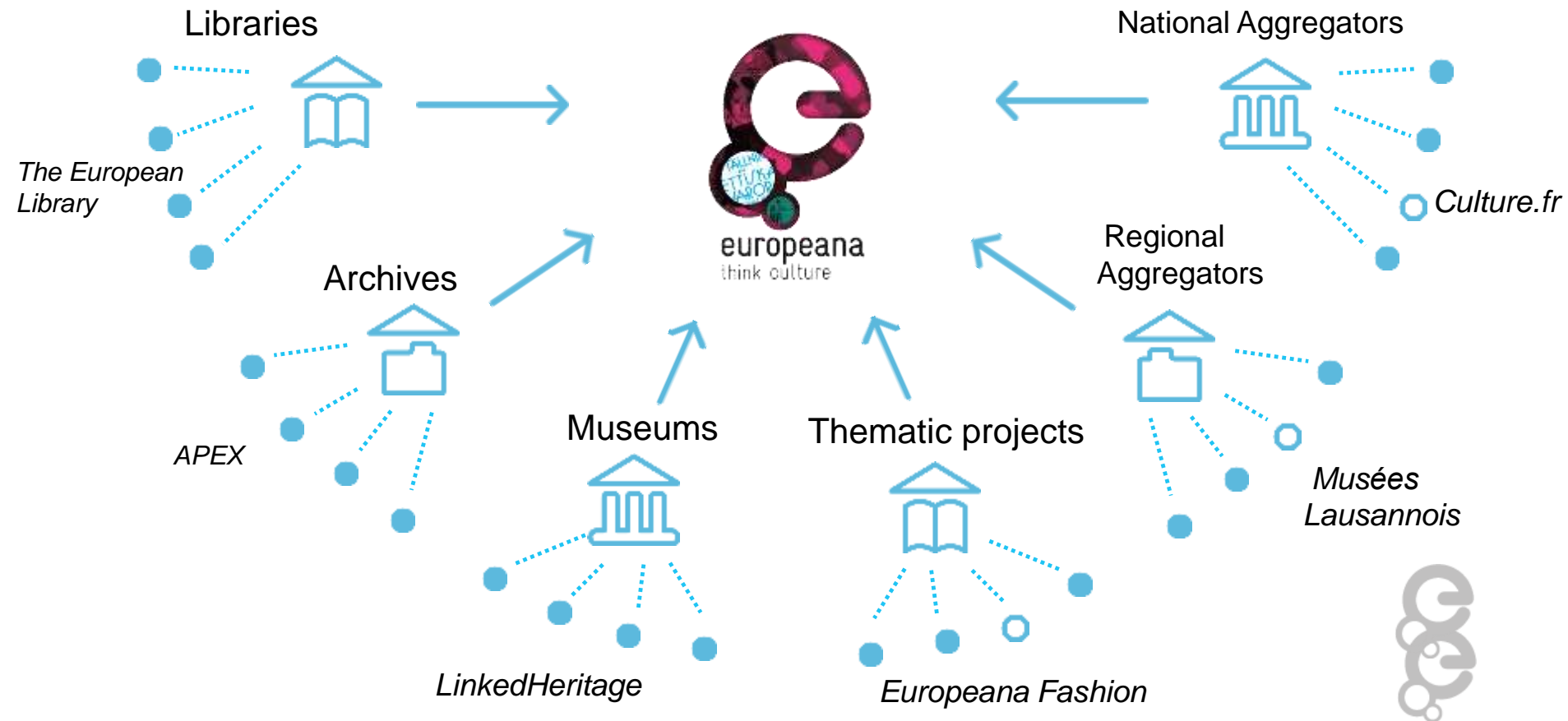
A platform for access and re-use

Data from all over EU

top 16

■ Germany	■ France	■ Spain	■ Netherlands	■ Sweden	■ Italy	■ UK	■ Europe
■ Norway	■ Poland	■ Ireland	■ Finland	■ Belgium	■ Denmark	■ Hungary	■ Austria
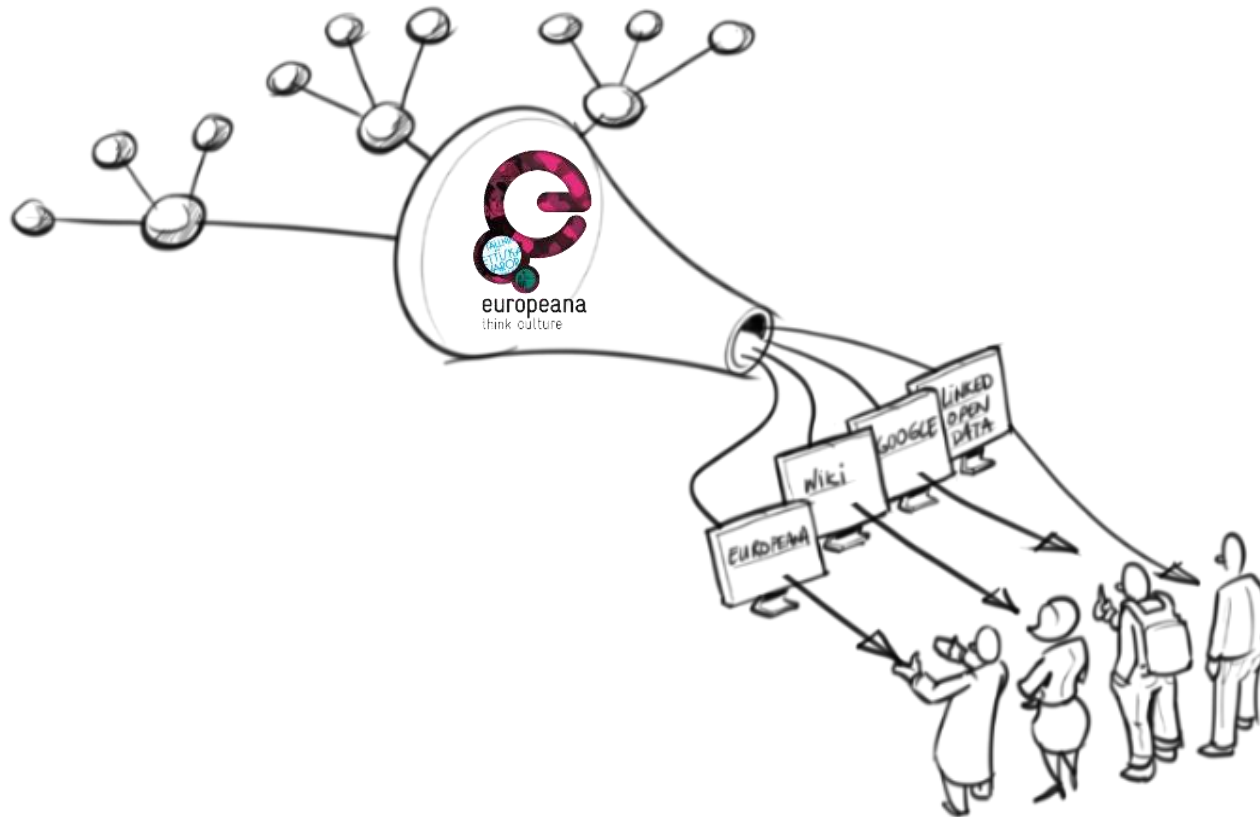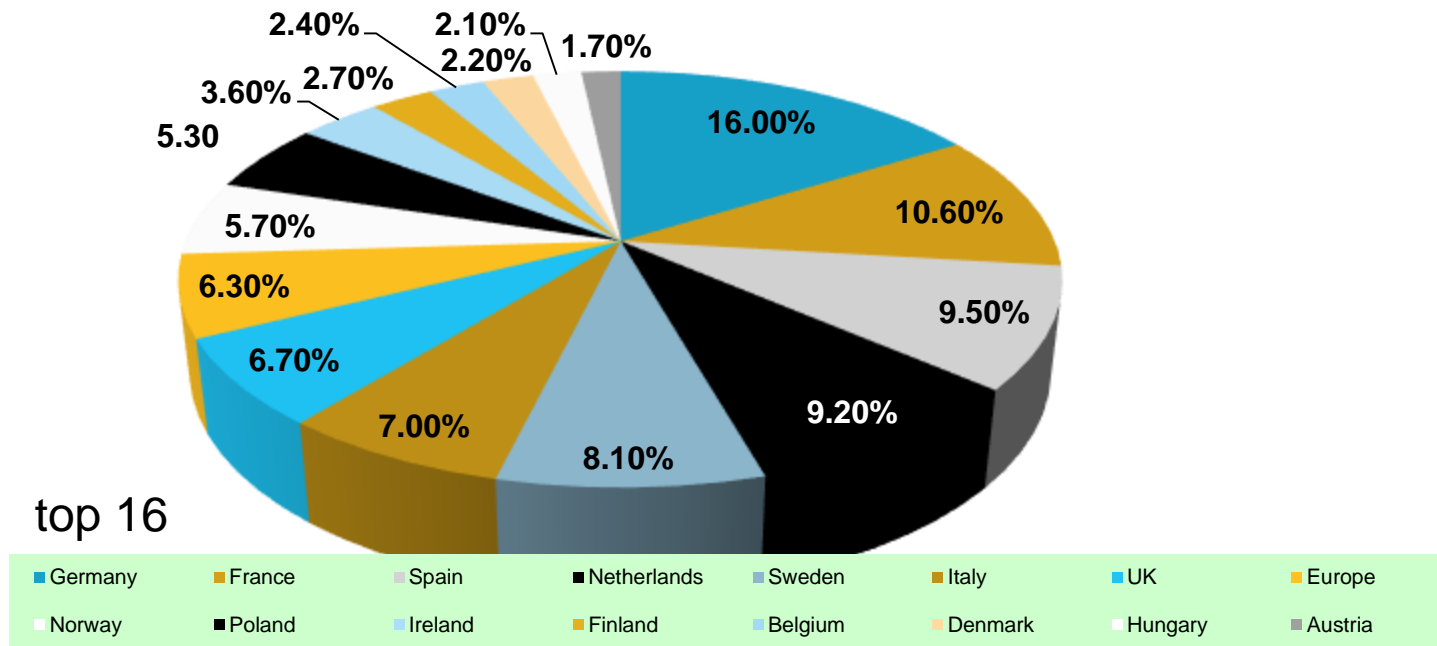
Items from 36 countries
Metadata in 33 languages

Data from all over EU

Pie chart percentages: 16.00%, 10.60%, 9.50%, 8.10%, 7.00%, 6.70%, 6.30%, 5.70%, 3.60%, 2.70%, 2.40%, 2.20%, 2.10%, 1.70%

Legend: Germany, France, Spain, Netherlands, Sweden, Italy, UK, Europe, Norway, Poland, Ireland, Finland, Belgium, Denmark, Hungary, Austria

(Re-)users from all over EU

Germany | France | Spain | Netherlands | Sweden | Italy | UK | Europe
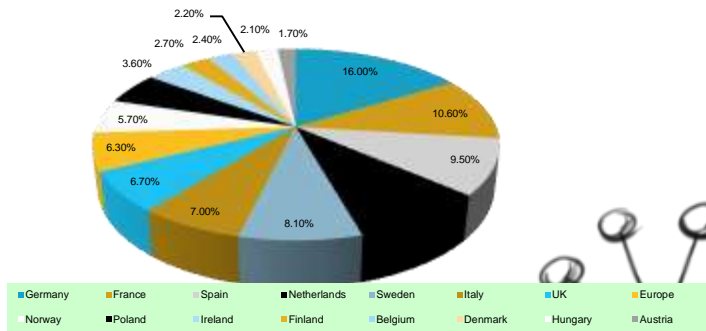Norway | Poland | Ireland | Finland | Belgium | Denmark | Hungary | Austria

16.00% 10.60% 9.50% 8.10% 7.00% 6.70% 6.30% 5.70% 3.60% 2.70% 2.40% 2.20% 2.10% 1.70%
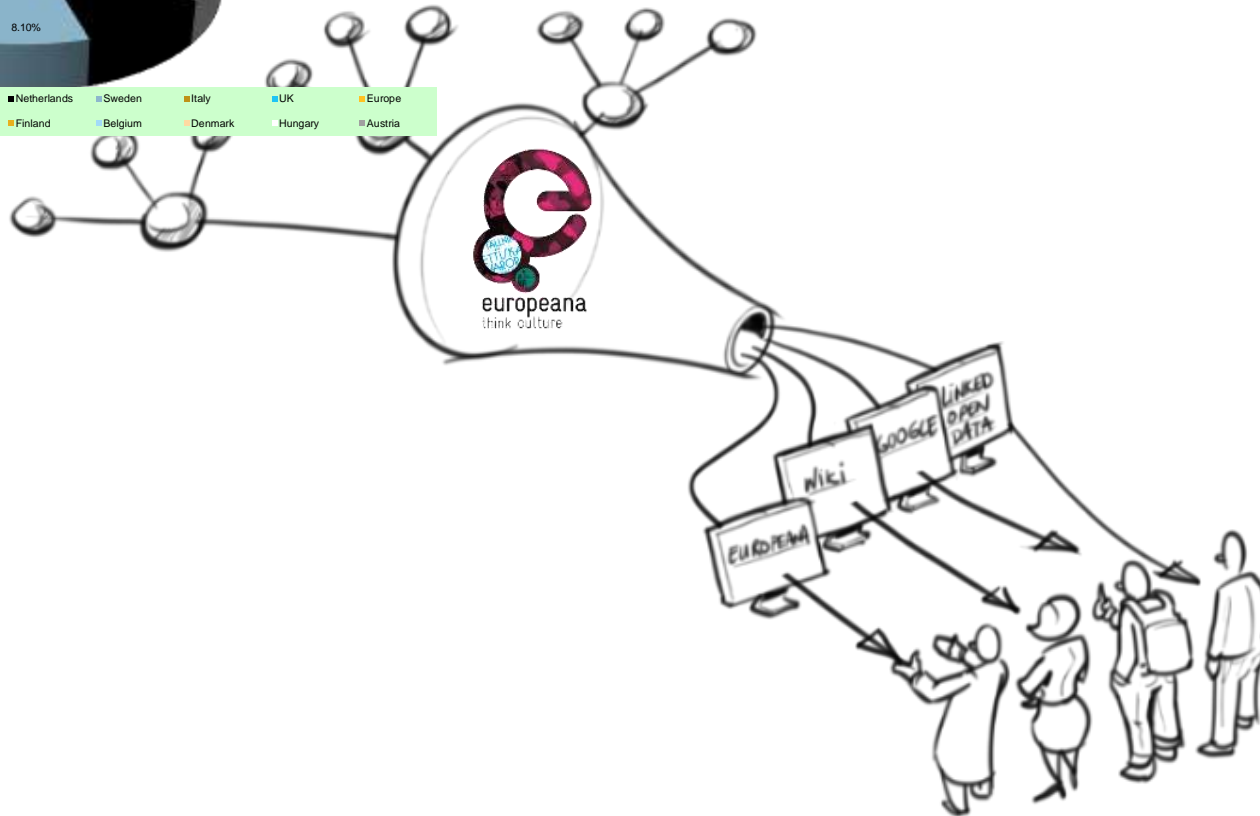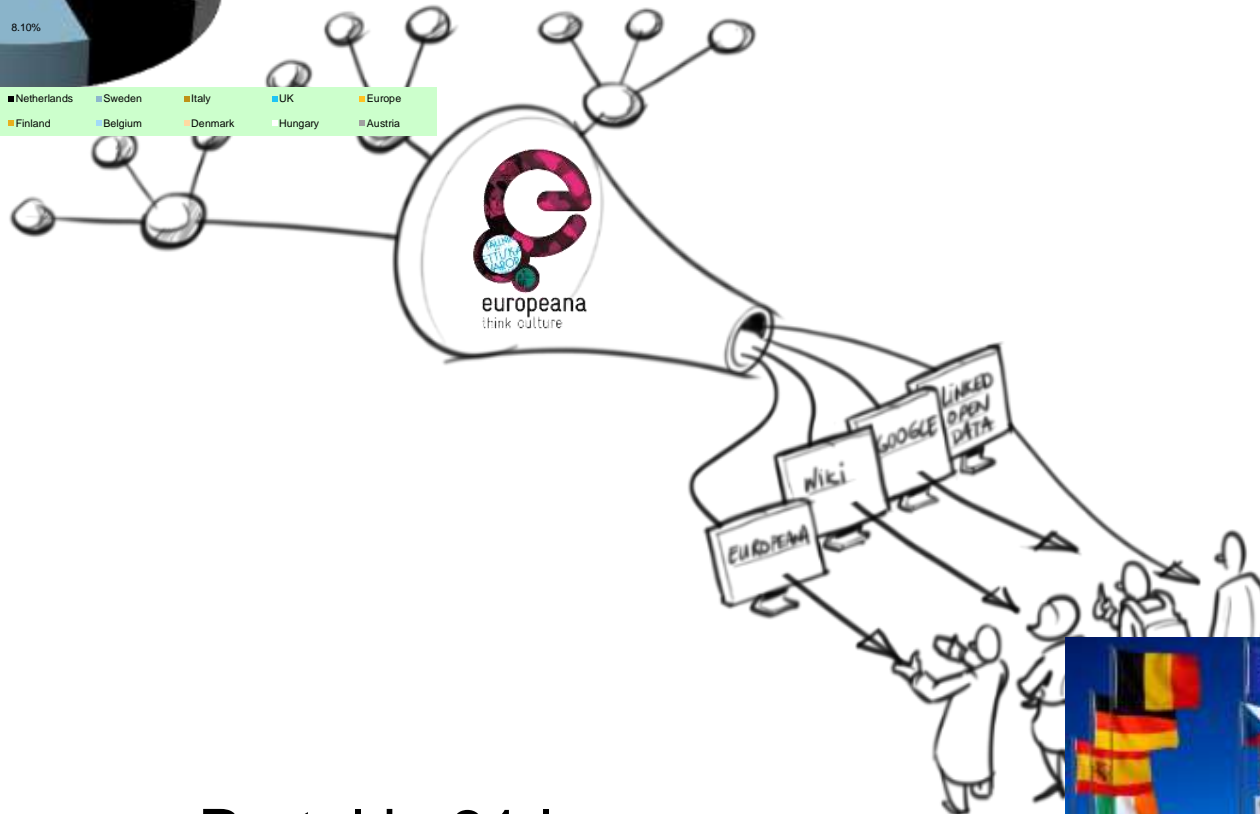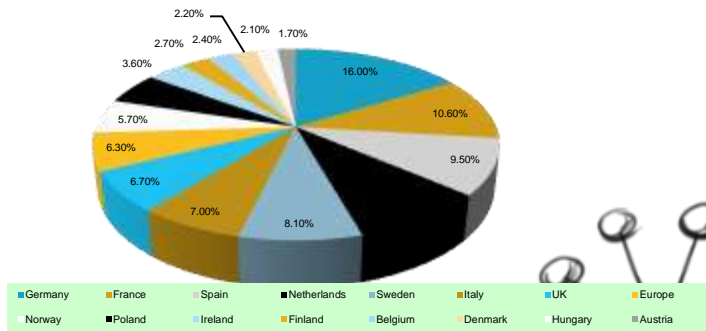
Portal in 31 languages
Data access from anywhere

Diversity of domains makes it hard

→ Data is heterogeneous

→ Makes it harder to train tools

Diversity of applications makes it harder
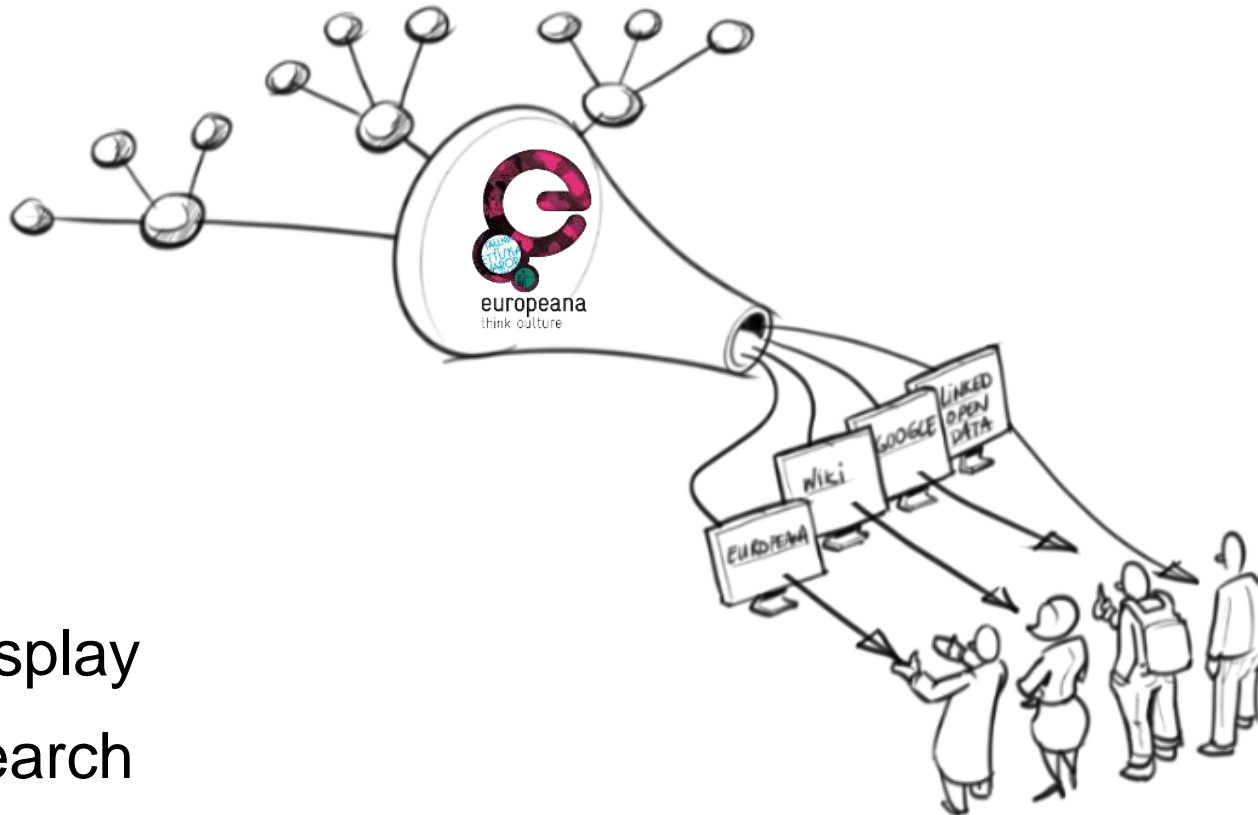
→ Display

→ Search

→ Data services

….

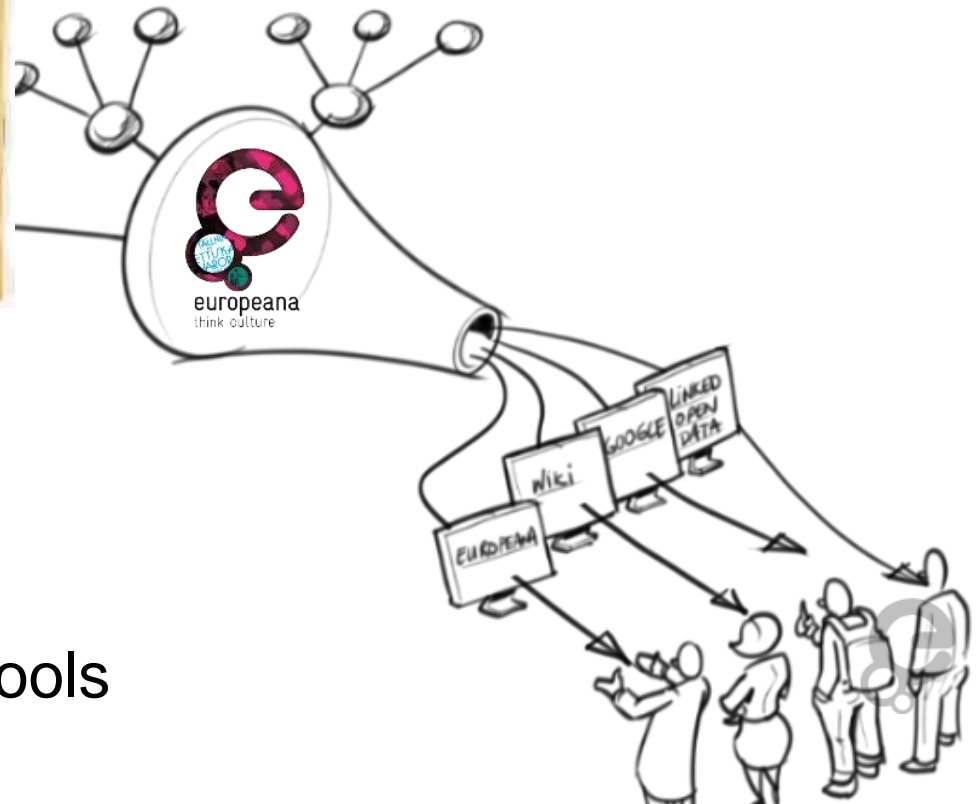# Some issues for applying language technology for online culture

→ 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

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-enċiklopedija l-ħielsa

Dan l-artiklu huwa dwar **il-pjaneta Art** . Jekk qed tfittex informazzjoni dwar Art f'sinifikati oħra, mur Art (Disambigwitajiet).

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

Exploiting multilingual data from partners

MIMO
musical instrument museums online

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

Thank you!

Antoine Isaac

antoine.isaac@europeana.eu

@EuropeanaEU