#### Délégation générale à la langue française et aux langues de France

Riga Summit 2015 MultilingualWeb

Building a Multilingual Website with No Translation Resources

Thibault Grouas April 29 2015



# 1. Technology and digital policies for languages in France

web services and applications





Delegation for french language and languages of France is responsible for the national linguistic policies It is part of the french Ministry of Culture and Communication Covers all languages spoken in France including French, regional languages and foreign languages. There are about 75 languages historically spoken in France, called « Languages of France » (most are regional languages) ☐ We actively promote **multilingualism**, which is the language of Europe The Delegation has a team dedicated to digital environment Digital and technologies appear to be a major issue for language policies today **Linked Open Data** is a new opportunity for providing multilingual



#### 2. The Sémanticpédia partnership



On November 2012, a major partnership was set-up between french public authorities, technology experts and a knowledge community:

#### French Ministry of Culture, INRIA and Wikimédia France





## 3. Sémanticpédia and DBpédia.fr



- Sémanticpedia links together research (Inria), Culture (Ministry of Culture) and Web 2.0 (Wikimédia France)
  - Inria brings its technical expertise of the web and linked data
  - Ministry of Culture has a unique expertise on Culture
  - Wikimedia France is the largest provider of cultural content on the web, with more than 25 million visitors per month in France
- Objectives of the partnership : to mutualise expertise and research to imagine new connections and technologies for Culture
- The partnership will be **reinforced** in late 2015 to integrate public cultural institutions like *Musée du Louvre*, *Monuments historiques* or *Bibliothèque nationale de France*
- □ The first project was **Dbpédia.fr**, a Wikipedia extraction conforming to linked open data standards based on the Dbpedia.org project, freely available through a **SPARQL endpoint** since 2013.



#### 3. Experimenting with DBpédia.fr



- Ministry of Culture conducted several experiments with DBpédia.fr to demonstrate the potential of linked data with multilingualism
- ☐ Firsts experiments :
  - **HDA-Lab** (2013): hdalab.iri-research.org. 3000 articles related to Culture available in 6 languages
  - Museosphere.net (2013) a museum search web service available in 9 languages
- We organised a competition on 2014 to help developing cultural projects based on Dbpédia.fr.
  - Winner for « linguistic » award : CROTOS Lab http://zone47.com/crotos/, browse and tag famous paintings in 33 languages.



#### 4. The Joconde**Lab** project

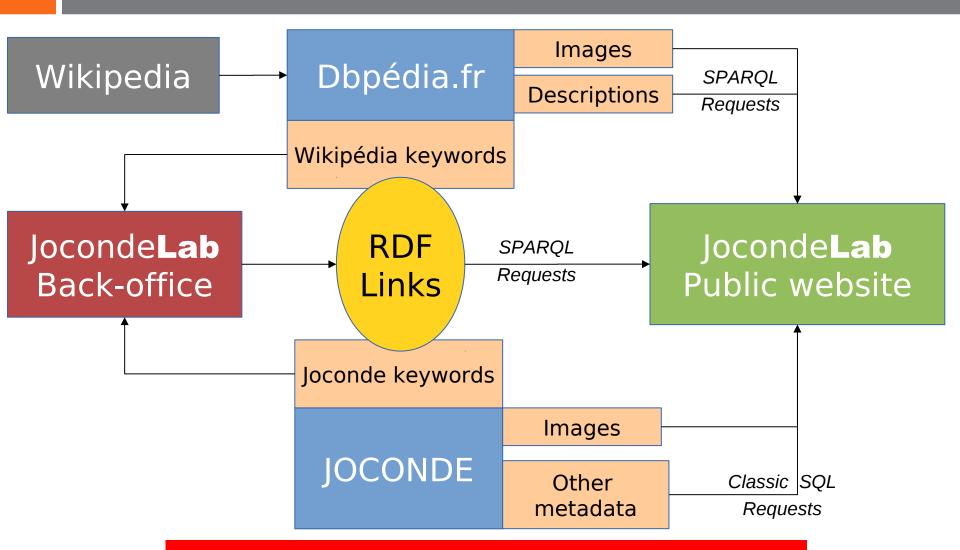


- ☐ The experiment took place in 2013-2014, with **JOCONDE database** form Ministry of Culture.
- □ JOCONDE contains about 300,000 images of various works of art displayed in french museums.
- □ The objective was to allow full navigation in the database with 14 languages (those directly integrated into Dbpédia.fr), with no translation, including 4 regional languages.
- □ Project started June 2013 with a partnership with *Institut de recherche et d'innovation du Centre pompidou*
- **8 months later**, The JocondeLab website opened with 300,000 illustrated files of works of art available in 14 languages
- □ Only **half a page** of interface texts was translated for the project (menus, buttons...)
- The project won a « Data Intelligence Award » 3 months later
- It is a major step towards linguistic diversity



#### 5. How does the magic work?





#### 6. Demonstration





# Joconde**Lab**

#### Online demo

http://jocondelab.iri-research.org/jocondelab

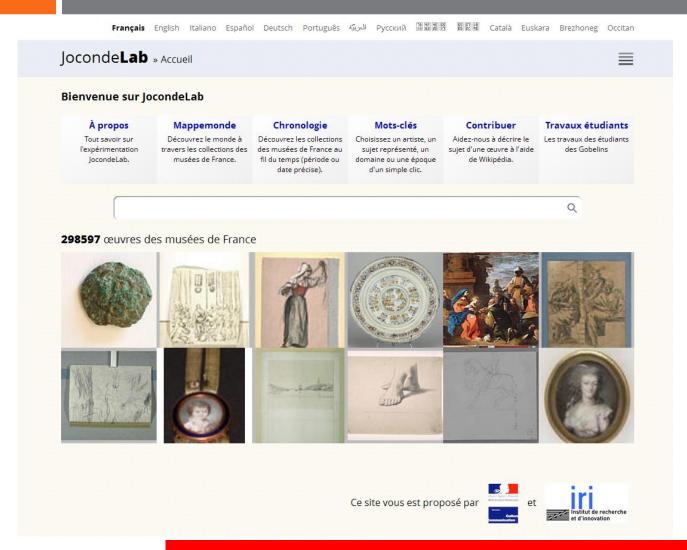
Offline demo

file:///E:/Anim\_Jocondelab/jocondelab.htm





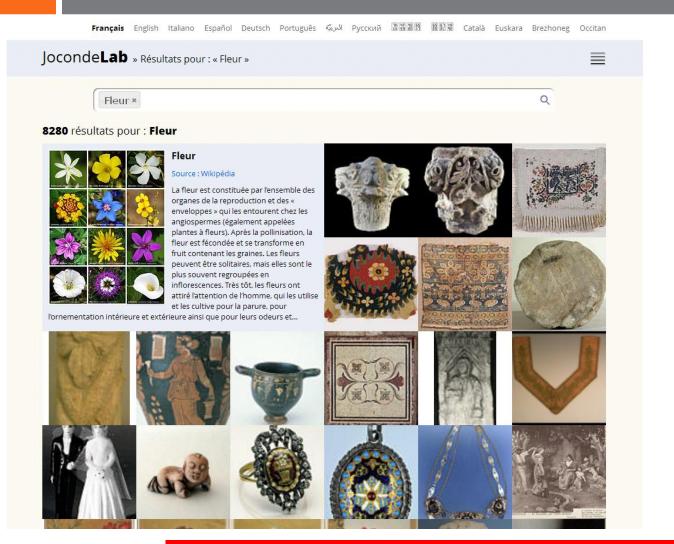




Homepage (french)







Search : « Fleur » (french)





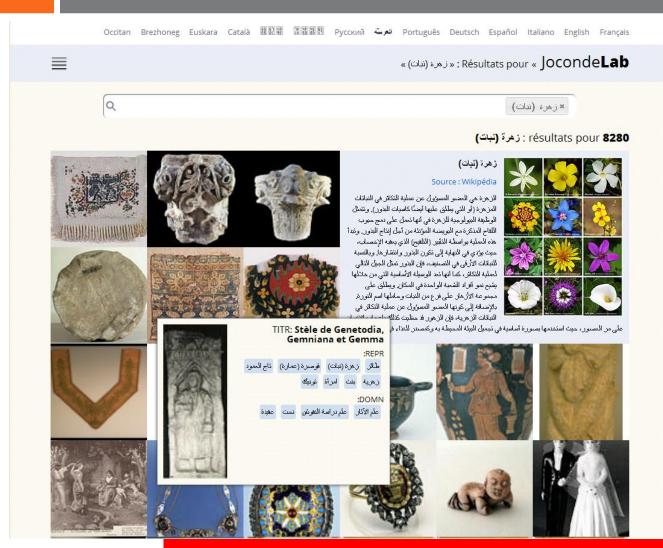


Responsive design optimised for mobile use.







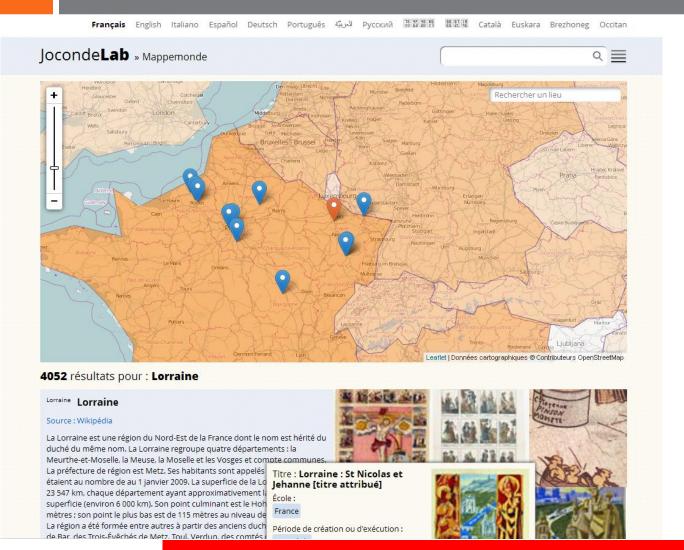


Search results for « Fleur » (arabic).

Note that texts are aligned to the right



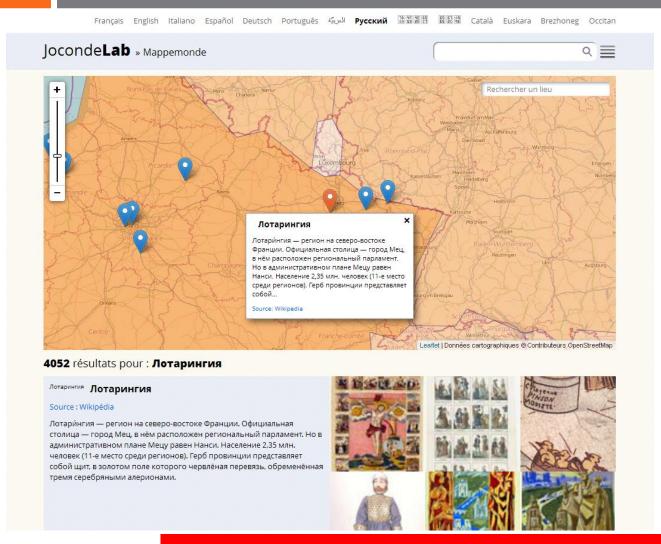




Navigation with the map tool : results for « Lorraine » (french)







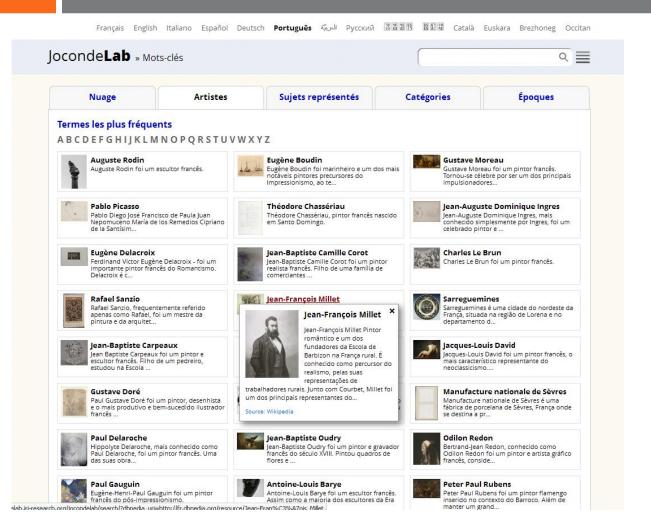
Navigation with the map tool : results for « Lorraine » (russian)

Descriptive texts come from Wikipedia through DBpédia.fr







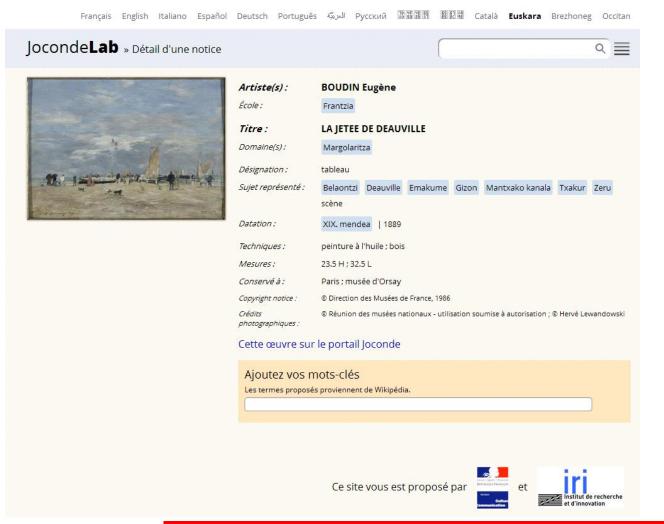


Navigation by keywords (artists) in portuguese

Images and texts come from Wikipedia through DBpédia.fr







Detail of a file from the Joconde Database (basque)

Linked tags appear in blue and are automatically translated

Field names were manually translated

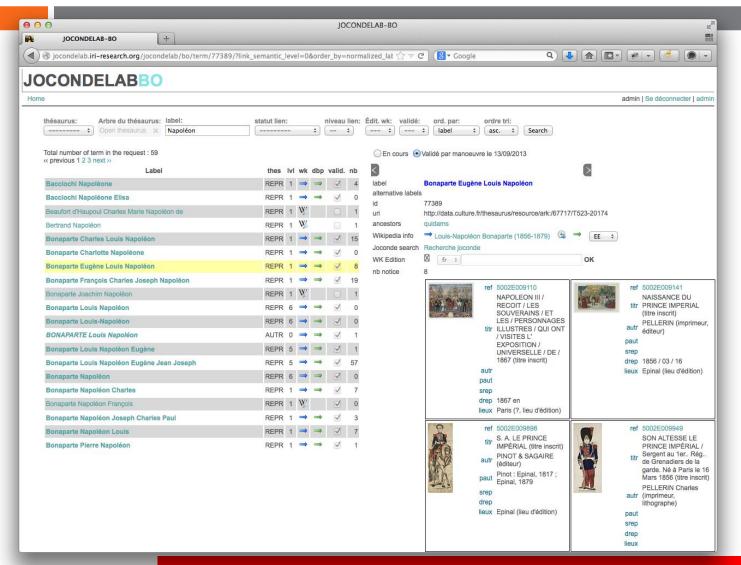
It is possible to add new keywords with the multilingual contribution tool below.



#### Joconde**Lab** back-office







The back-office is a simple tool to facilitate and help the linking process between keywords from DBpédia and JOCONDE.

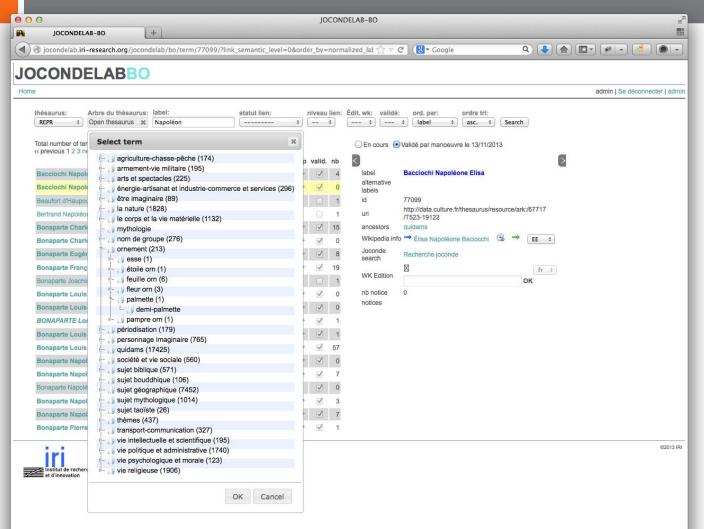
All proposed matches are manually validated by Ministry of Culture



#### Joconde**Lab** back-office



Communication



Detail of the Joconde thesaurus

87 000 keywords
were linked
during the project
by 2 persons
during 3 months
(total 6 months
work).

95 % of Joconde keywords were found in DBpédia.fr which is pretty good









# Thibault Grouas thibault.grouas@culture.gouv.fr @tgrouas





