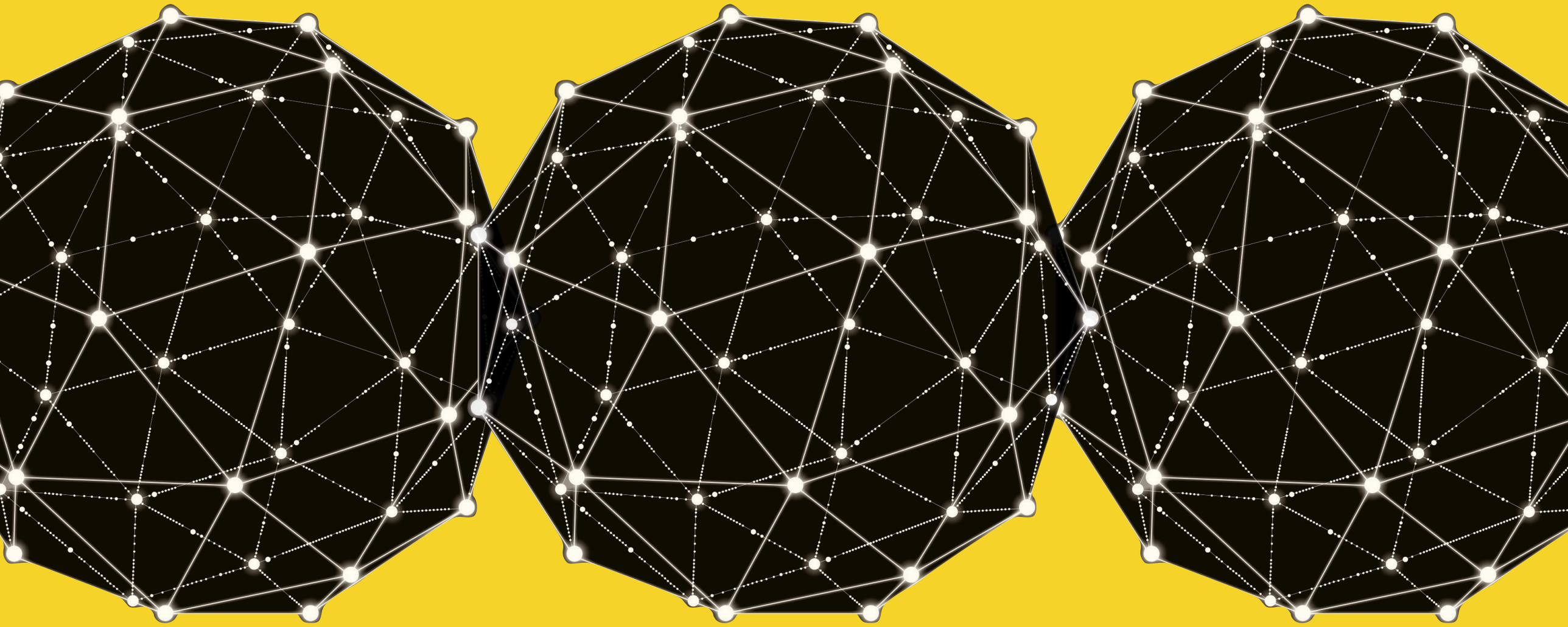


# Métadonnées, Protocoles et Modèles

## Infovisualisation, mashup et datawrangling appliquées aux données culturelles

Antoine Courtin
responsable de la Cellule d'ingénierie documentaire
Département des études et de la recherche à l'INHA

Master Histoire de l'art - Université de Bourgogne (Dijon)
29 mars 2018

# 2 séance

- 4 avril matin

  > enjeux et problématiques autour de la réutilisation des données culturelles

  > panorama Infoviz dans les GLAMs

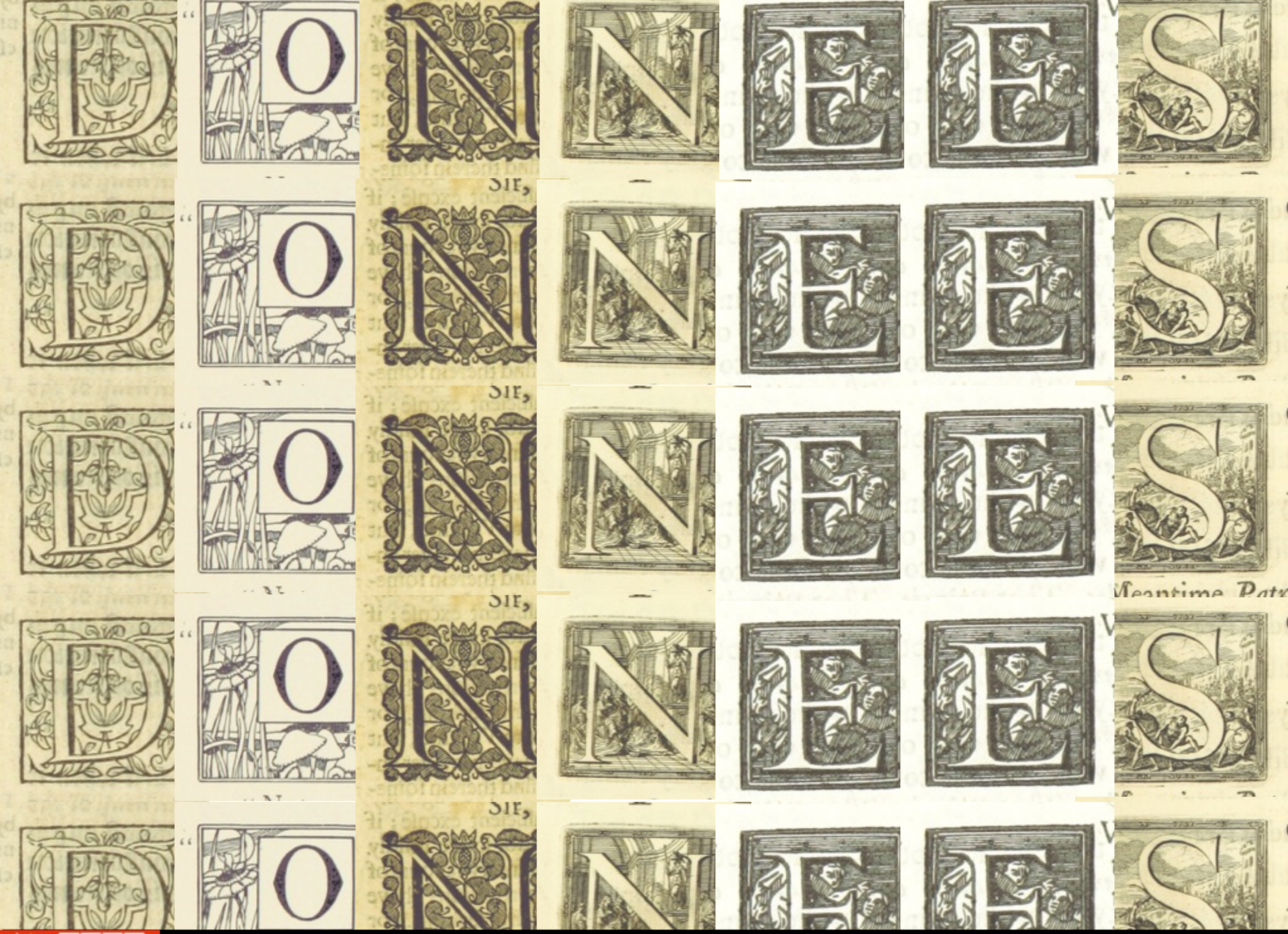
  > exercices pratiques

- 5 avril matin

  > traitement de données : kesako ?

  > Exercices pratiques et initiation à OpenRefine

# Contexte


DONNEES
SIR,
Meantime Patr

Open Data

Open Content

Linked Open Data

Big Data

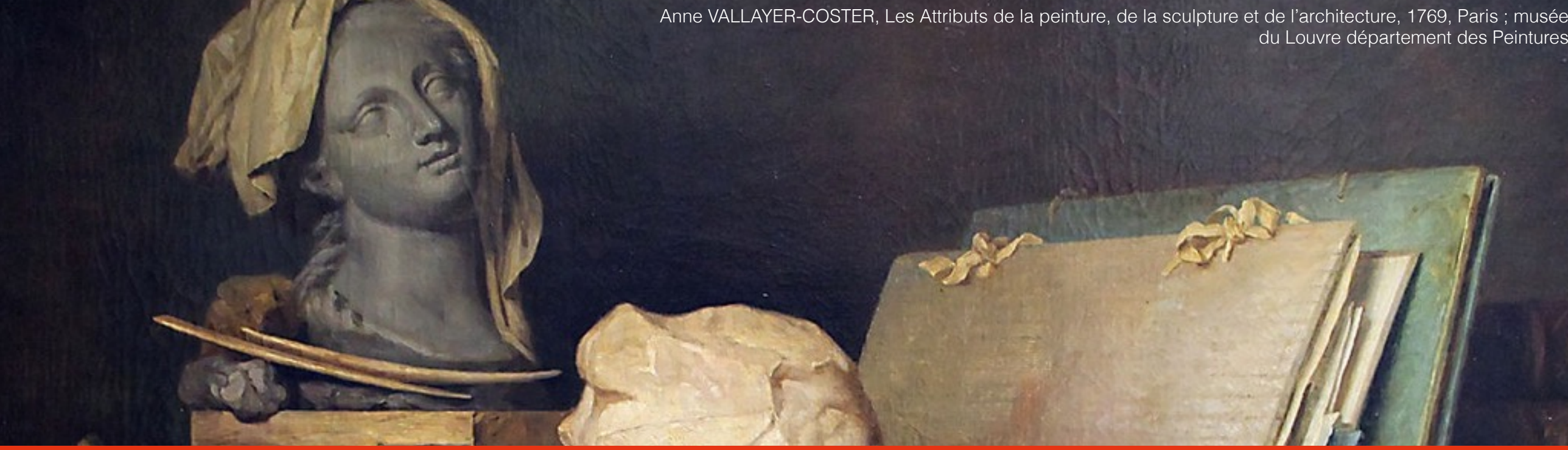
Open Source

Open Standards

Linked Entreprise Data

Open Knowledge

# The old problem was access, the new problem is analysis



Anne VALLAYER-COSTER, Les Attributs de la peinture, de la sculpture et de l'architecture, 1769, Paris ; musée du Louvre département des Peintures

# Pour suivre auj. > 1 URL à retenir !

## https://github.com/antoinecourtin/EbD_2018

# [apartés]

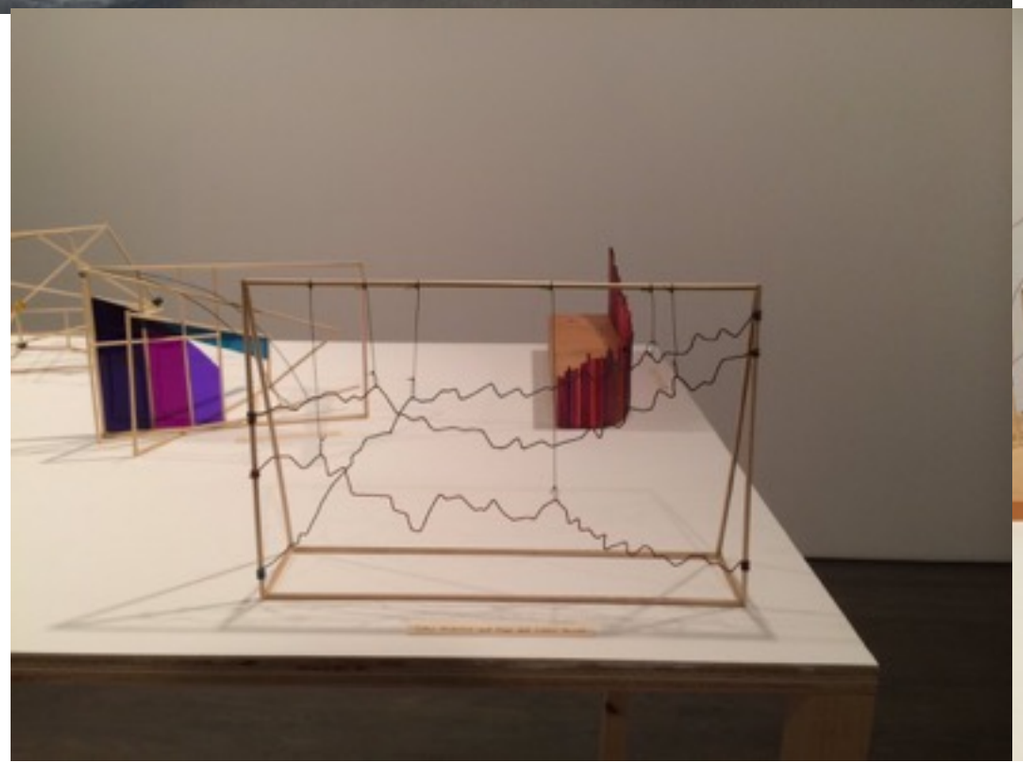
## les *data* comme *matériaux* à la création plastique

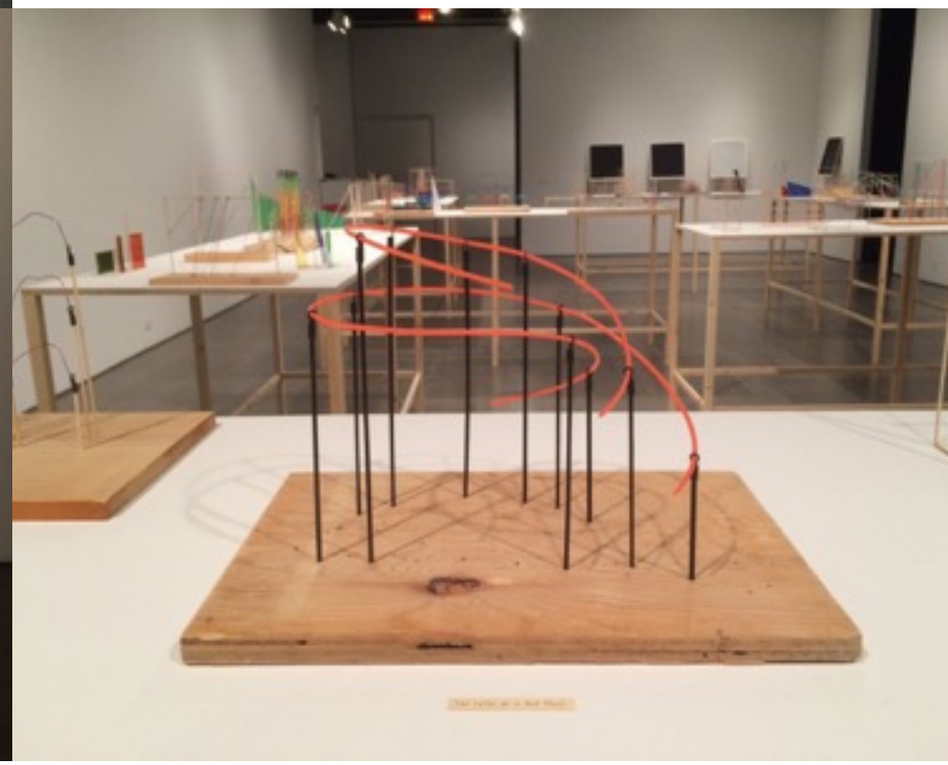
# DATA DADA

http://www.data-dada.net/

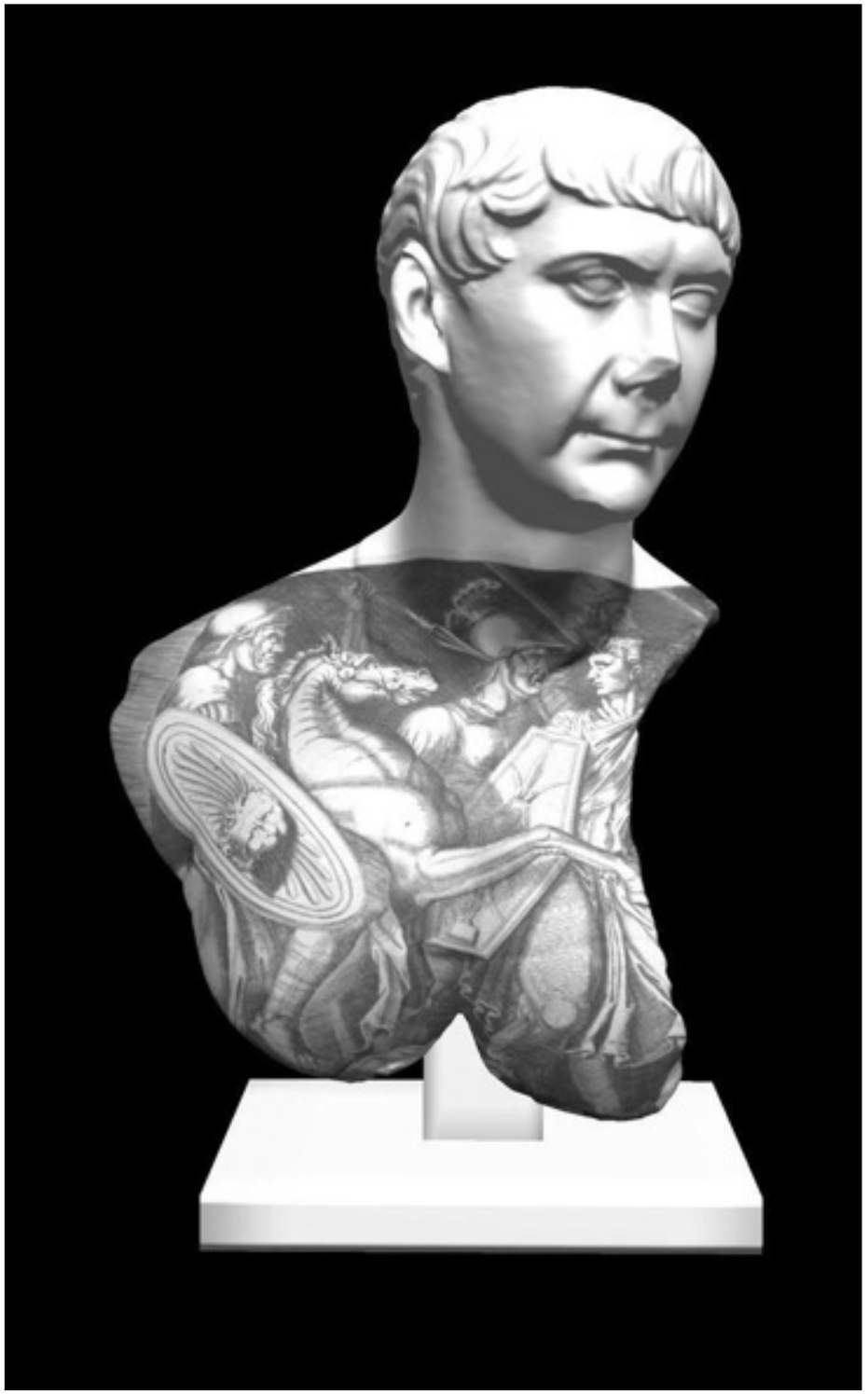
https://www.franceculture.fr/emissions/le-numerique-et-nous/donnees-voir-regards-dartistes-contemporains#

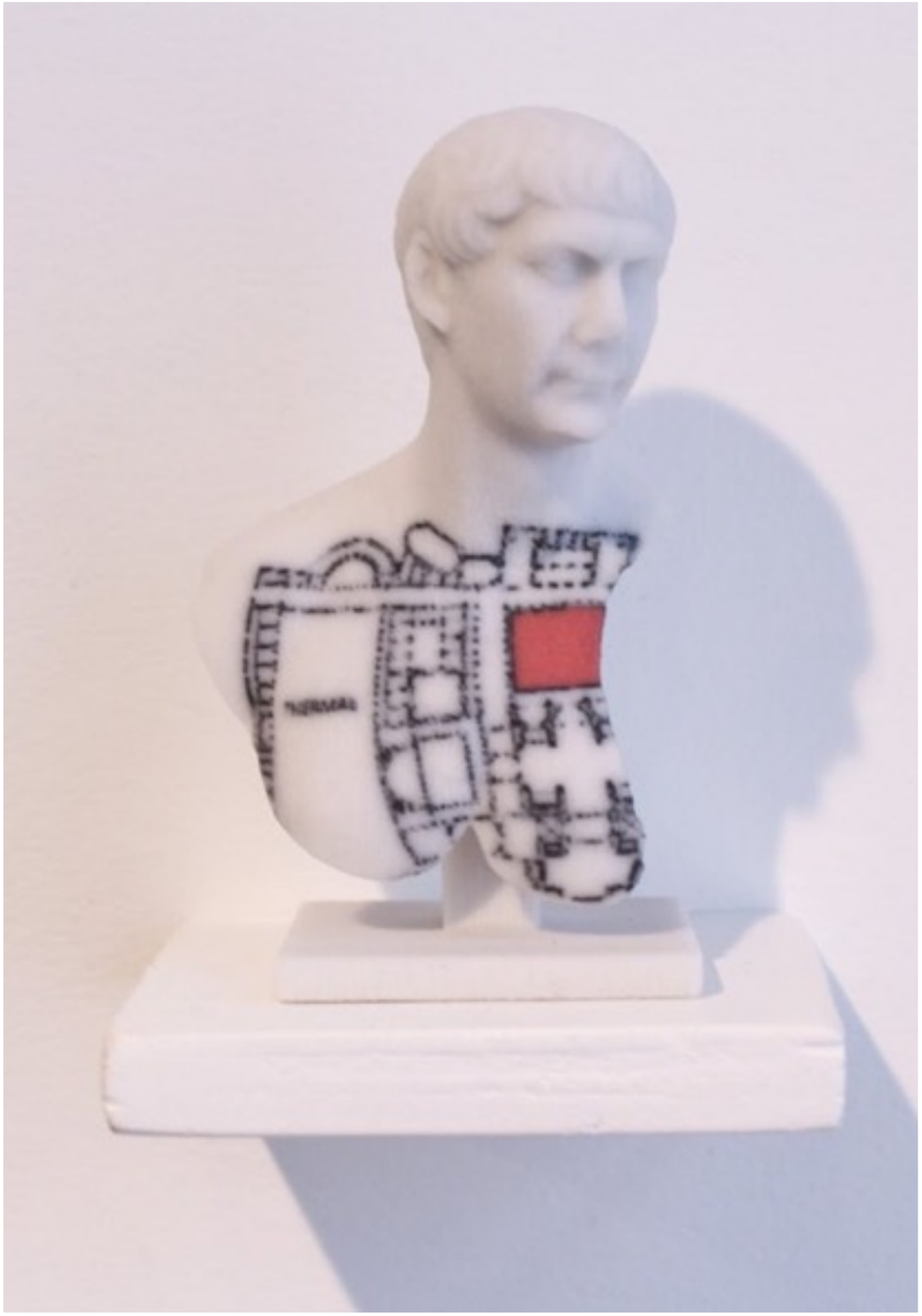
*Each Number Equals One Inhalations and One Exhalation,* 2016





https://twitter.com/PDCutup



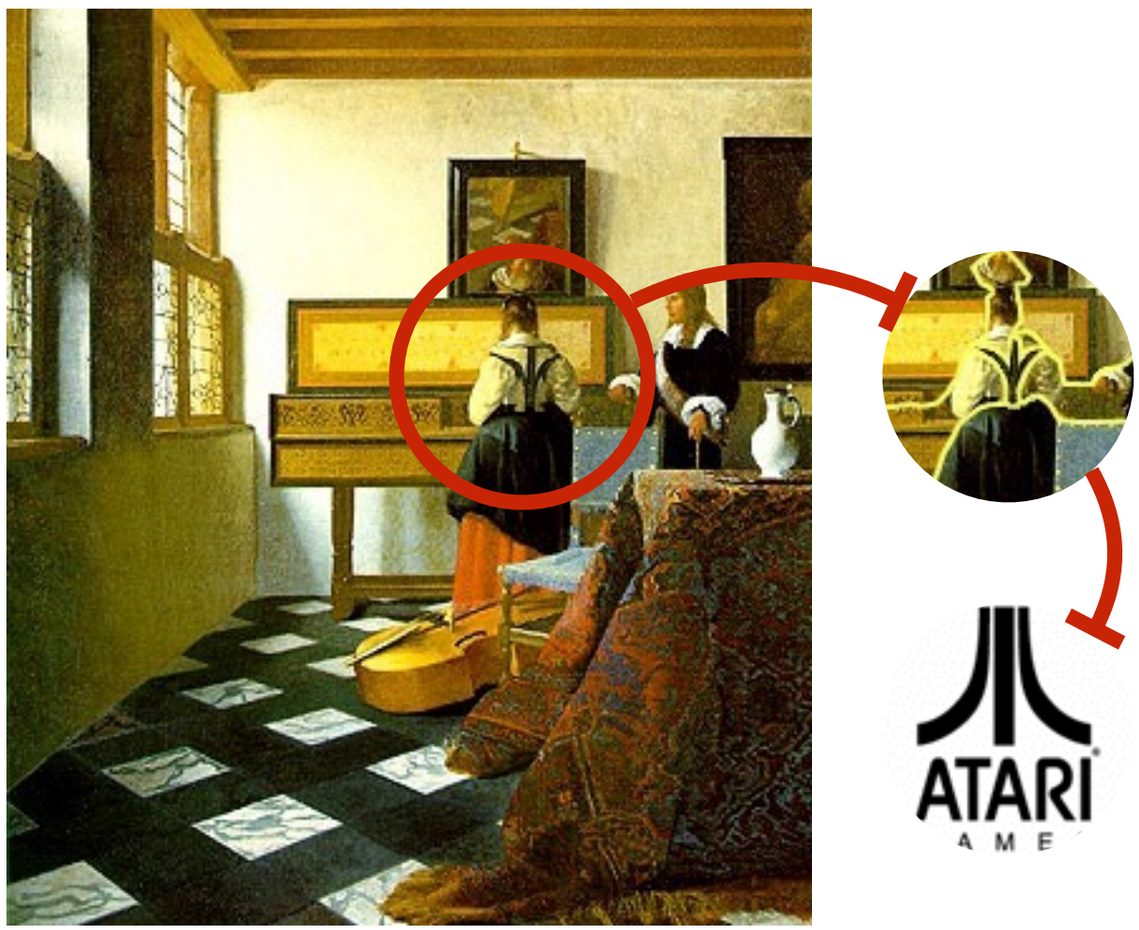


http://cargocollective.com/alicecmartin

@NicaoGr

# Logo.Hallucination par l'artiste Christophe Bruno

The software based on neural network image recognition was exhibited at the Rencontres Internationales Paris Berlin in November 2006.




Jan Vermeer, The Music Lesson
c. 1662-1665. Oil on canvas, 74.6 x 64.1 cm (Royal Collection, St. James' Palace, London).

Giovanni Toscani, Trittico con Madonna col Bambino, S. Girolamo e S. Caterina (Firenze, Museo dello Spedale degli Innocenti).

MOTOROLA

https://goo.gl/GVJFwj

# [apartté]

# Ressources et histoire de l’art - une histoire déjà ancienne

**1975-1979** Création par le ministère de la culture d'une base de données des peintures des musées français.

**1982** Inauguration de la banque d’images consacrée aux miniatures médiévale de la bibliothèque Sainte-Geneviève

**1986** Poste de consultation numérique au musée d’Orsay

Vidéodisque de 15 000 images du Musée d’art moderne, des fonds nationaux et régionaux d’art contemporain

**1992** La base Joconde du Ministère passe sur le Minitel

**1995** Ouverture du site internet du Louvre

Mise en ligne des bases Joconde et Mérimée

**1997** Création de la base Archidoc du MCC

Inauguration de la 1er version de GALLICA

**1998** Mise en ligne de la base Palissy (création en 1989)

**2000** Mise en ligne de la seconde version de Gallica

**2003** Première mise en ligne d’état-civil numérisé en archives départementales

**2004** Lancement Google Book Search
Mise en ligne de Photo-RMN

**2007** Publication du moteur « Collection » sur le site du MCC

**2008** Première version du portail Europeana

**2010** 94 services d’archives ont mis des documents numérisés en ligne

**2011** Version bêta du projet APEnet (Portail Européen des archives)

**2013** Publication des notices objets mobiliers du service de l’inventaire du patrimoine de la Région Alsace sur la plateforme OpenData

**2014** Mise en ligne du projet JocondeLab

**2015** Nouvelle version de Gallica (de la bibliothèque nationale de France)

**2016** Lancement du portail http://parismuseescollections.paris.fr/

# Où sont les données ?

bases métiers

poste local - dossiers de session

OPEN DATA

des fichiers mis à dispo
sur plateforme OpenData

site web - CMS

WIKIDATA

des plateformes avec
des données structurées

Data.culture

FranceArchives

Data.gouv

bases métiers

des fichiers mis à dispo
sur plateforme OpenData

des plateformes avec
des données structurées

poste local - dossiers de session

site web - CMS

Data.culture

FranceArchives

Data.gouv

## Plateforme OpenData national

Plateforme data.gouv.fr

## Plateforme nationale thématique

Plateforme open data du
Ministère de la Culture

## Plateforme d'opérateur

Bibliothèque du Congrès de
Washington sur FlickR

## Plateforme ad-hoc développé par l'institution

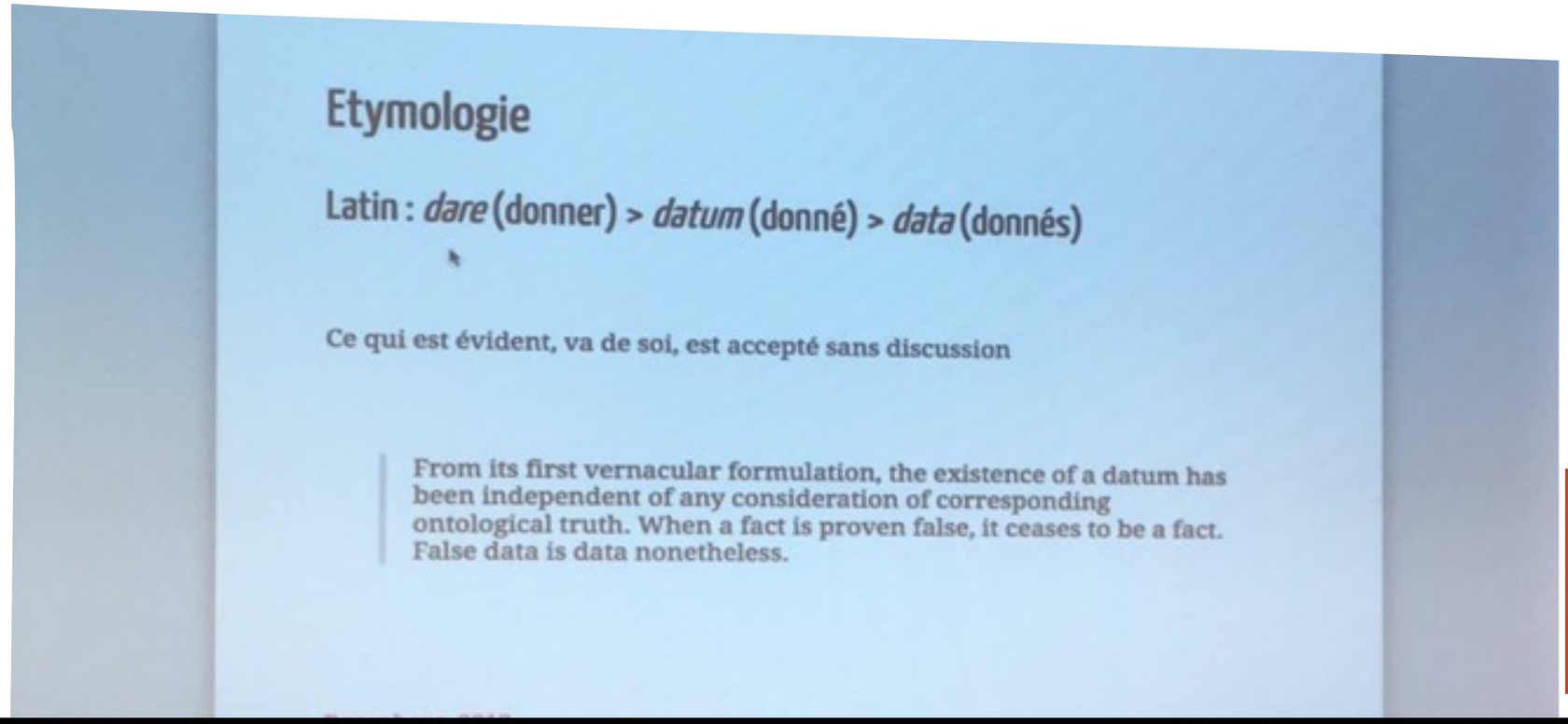
Rijksmuseum d'Amsterdam

## Plateforme de la sphère wikimedia

Bunersarchiv sur Wikimedia Commons

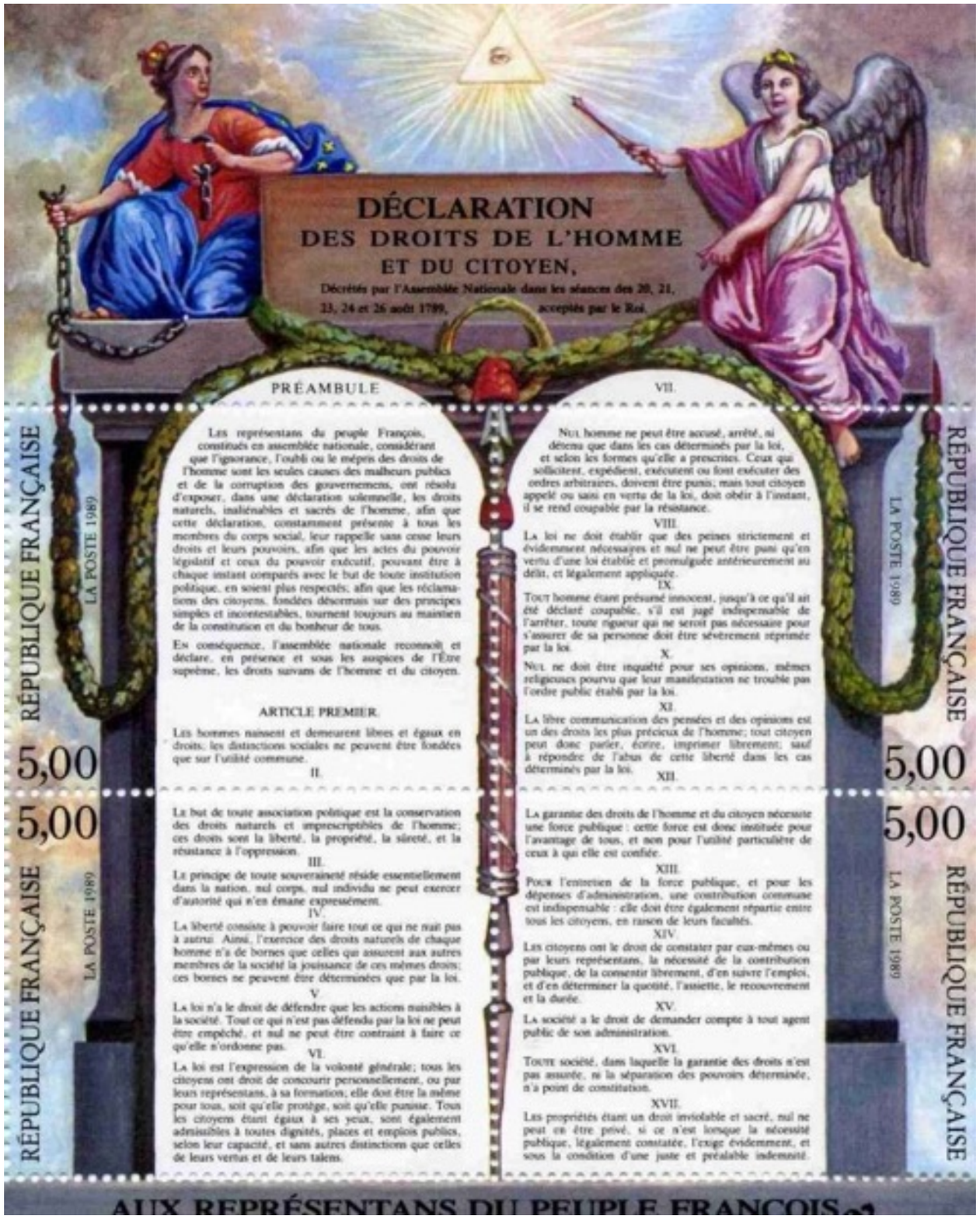
# L'Open Data : qu'est-ce que c'est ?

En Français, l'open data ou données ouvertes **désigne à la fois un ensemble d'initiatives et un ensemble de principes** appliqués à des jeux de données. La volonté étant de **mettre à disposition des données** numériques - **d'origine publiques ou privées** - en **libre accès, gratuitement, réutilisables par tous et à toutes fins** (y compris commerciales), **sans contraintes techniques**.




https://twitter.com/datactivi_st/status/957913780503089152

Bien qu'apparue dans une agence environnementale américaine en 1995 puis reprise en 2008 par l'administration d'Obama sous l'angle de *l'OpenGovernment* avant d'être popularisée dans le monde, cette volonté de transparence et d'accès à l'information est présente dès la Déclaration des Droits de l'Homme et du Citoyen de 1789 avec son article 15 qui stipule clairement que « La société a le droit de demander compte à tout agent public de son administration ». On retrouve ensuite cette notion dans la loi - dite CADA - du 17 juillet 1978 ayant pour objectif de faciliter et de contrôler l'accès des particuliers aux documents administratifs.



Les données publiques dites ouvertes, sont finalement, l'ensemble des données dont on décide l'ouverture et que les administrations utilisent pour gérer le territoire et mettre en œuvre leurs politiques. **Il existe cependant des restrictions liées aux données dites sensibles pouvant toucher à la sécurité nationale, à la protection des données personnelles, à des droits de propriété intellectuelle de tiers ou encore celles étant couvertes par les différents secrets légaux.**

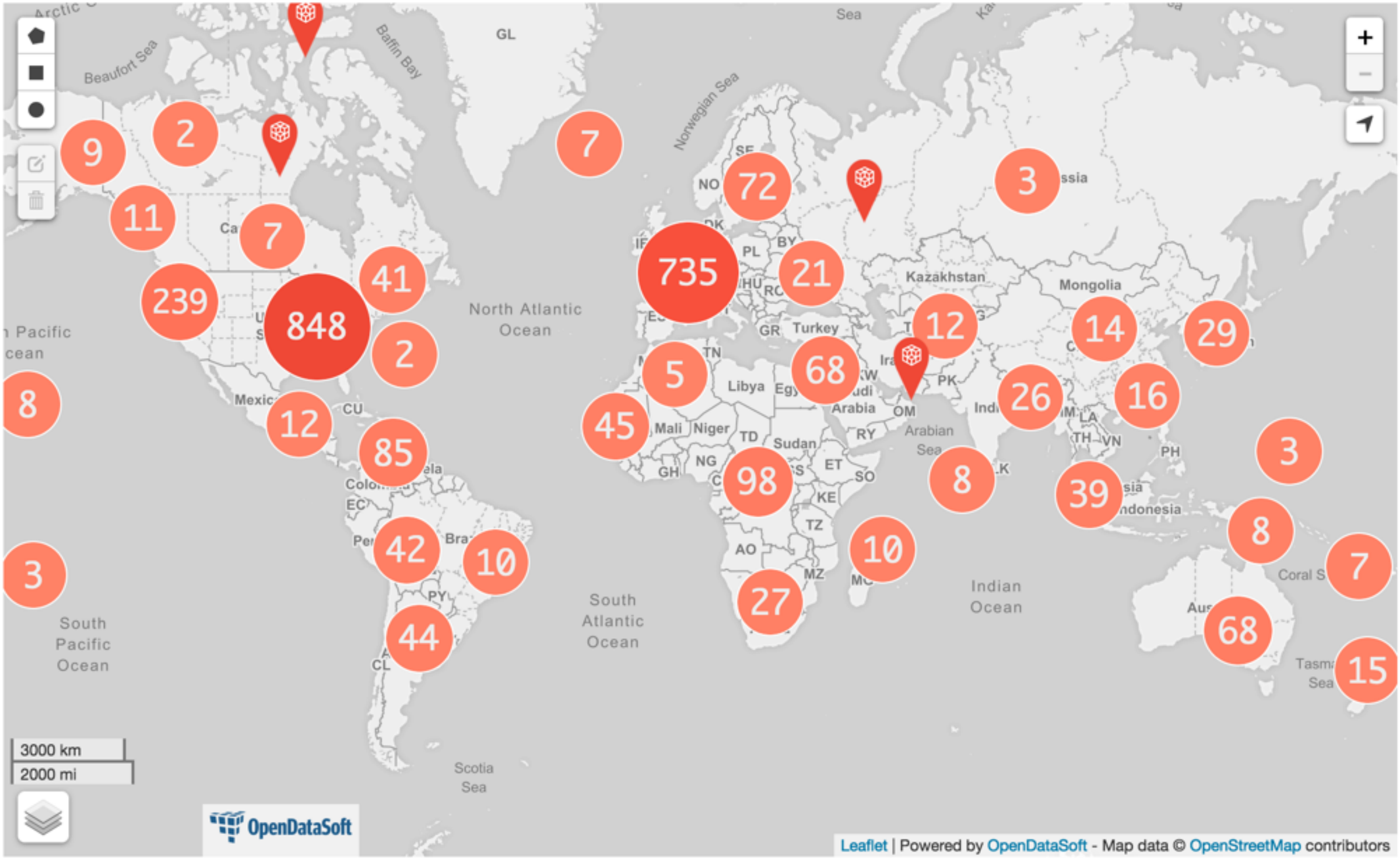
Références


- Branislav M. et al. (2015), Livre blanc sur les données ouvertes, HAL Id: halshs-01162692, disponible en ligne: https://halshs.archives-ouvertes.fr/halshs-01162692
- Domange, C. (2013), Guide Data Culture : Pour une stratégie numérique de diffusion et de réutilisation des données publiques numériques du secteur culturel. Paris : ministère de la Culture et de la Communication, 2013. 50 p. Disponible en ligne : http://www.culture.fr/content/download/9175
- Chignard, S. (2013), Une brève histoire de l'Open Data, ParisTech Review, disponible en ligne : http://www.paristechreview.com/2013/03/29/origines-open-data/
- Maurel, (2013), Ouverture des données culturelles : The Times, They Are A Changing ?. – S.I.Lex – Carnet de veille et de réflexion d'un juriste et bibliothécaireDisponible en ligne : https://scinfolex.com/2013/07/31/ouverture-des-donnees-culturelles-the-times-they-are-changing/
- Le Coz, L. (2017), Patrimoine numérisé et Open Content : Quelle place pour le domaine public dans les bibliothèques numériques patrimoniales ? sous la direction de Sylvain Machefert, Mémoire de l'ENSSIB. Disponible en ligne : http://www.enssib.fr/bibliotheque-numerique/notices/67307-patrimoine-numerise-et-open-content-quelle-place-pour-le-domaine-public-dans-les-bibliotheques-numeriques-patrimoniales


Textes officiels


- Loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public. Disponible en ligne : https://www.legifrance.gouv.fr/eli/loi/2015/12/28/2015-1779/jo/texte
- Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique. Disponible en ligne : https://www.legifrance.gouv.fr/eli/loi/2016/10/7/2016-1321/jo/texte

# La liste des portails #OpenData dans le monde



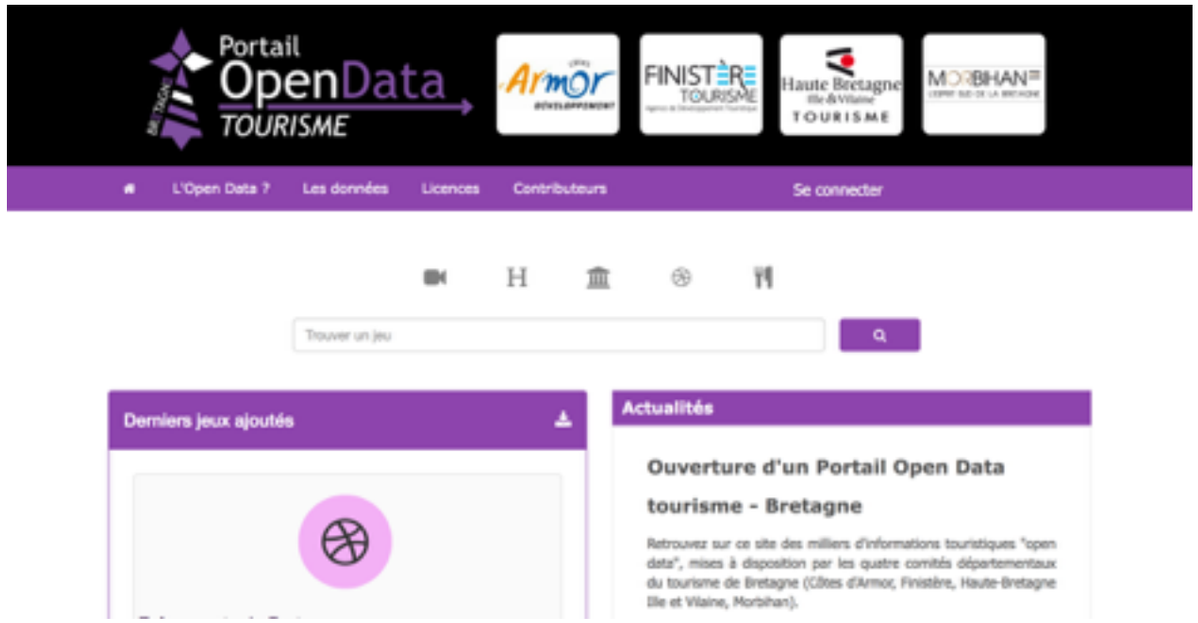

https://opendatainception.io /

## Plateforme nationale

## Plateforme thématique

## Portail de ville

## Portail dédié à une région



# La jungle des licences…

# Plusieurs jalons législatifs importants…

- Loi du 7 octobre 2016 pour une République numérique (Loi Lemaire) ;
- Loi du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public (Loi Valter) ;
- Loi du 7 juillet 2016 relative à la liberté de la création, à l'architecture et au patrimoine.

## Loi CADA de juillet 1978

Principe : oblige les autorités à organiser un accès aux documents administratif

Exception (régime dérogatoire des données culturelles) : pour l'ESR et les établissements, organismes ou service culturels

## Propositions CNNUM - juin 2015

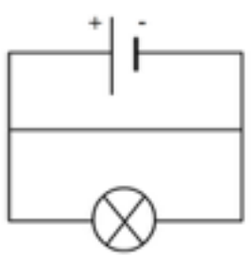
La numérisation des contenus culturels du domaine public, détenus par les établissements publics, ne peut conduire à l'imposition de nouvelles limitations à l'usage des œuvres (*cf. copyfraud*[184]). Il s'agit d'un travail porté par la puissance publique et non d'un acte créatif.

Par ailleurs, **les données publiques produites par les institutions culturelles** (ex : métadonnées sur les œuvres, statistiques de fréquentation...) **devraient rejoindre le régime général des données publiques.**

## Projet de loi pour une République numérique + consultation citoyenne

Le projet initial du gouvernement instaure un principe d'Open Data par défaut pou les données produites par les administrations.

**Le texte ne prévoit aucune dérogation possible pour les institutions culturelles**



## Loi Valter - décembre 2015

Intervient pour transposer une directive européenne sur la réutilisation des informations du secteurs public. Cours-circuite l'élan par le projet de loi  République numérique car confiée aux Secrétariat en charge de la Réforme de l'Etat (et non au Secrétariat à l'économie numérique).

Loi Valter - décembre 2015

-> **Instauration d'un principe de gratuité pour la réutilisation des informations publiques**
-> **L'article 3 précise que l'article 11 (de la loi de 1978)** est abrogé, la régime dérogatoire pour les GLAMs semble supprimé

# En réalité, l'exception est maintenue pour les données issues de la numérisation des collections

Article 5. II : « La réutilisation peut également donner lieu au versement d'une redevance lorsqu'elle porte sur des informations **issues des opérations de numérisation des fonds et des collections des bibliothèques, y compris des bibliothèques universitaires, des musées et des archives** et, le cas échéant, sur des informations qui y sont associées lorsque ces dernières sont commercialisées conjointement.
Le produit total du montant de cette redevance, évalué sur une période comptable appropriée, ne dépasse pas le montant total des coûts de collecte, de production, de mise à disposition ou de diffusion, de conservation de leurs informations et d'acquisition des droits de propriété intellectuelle. »

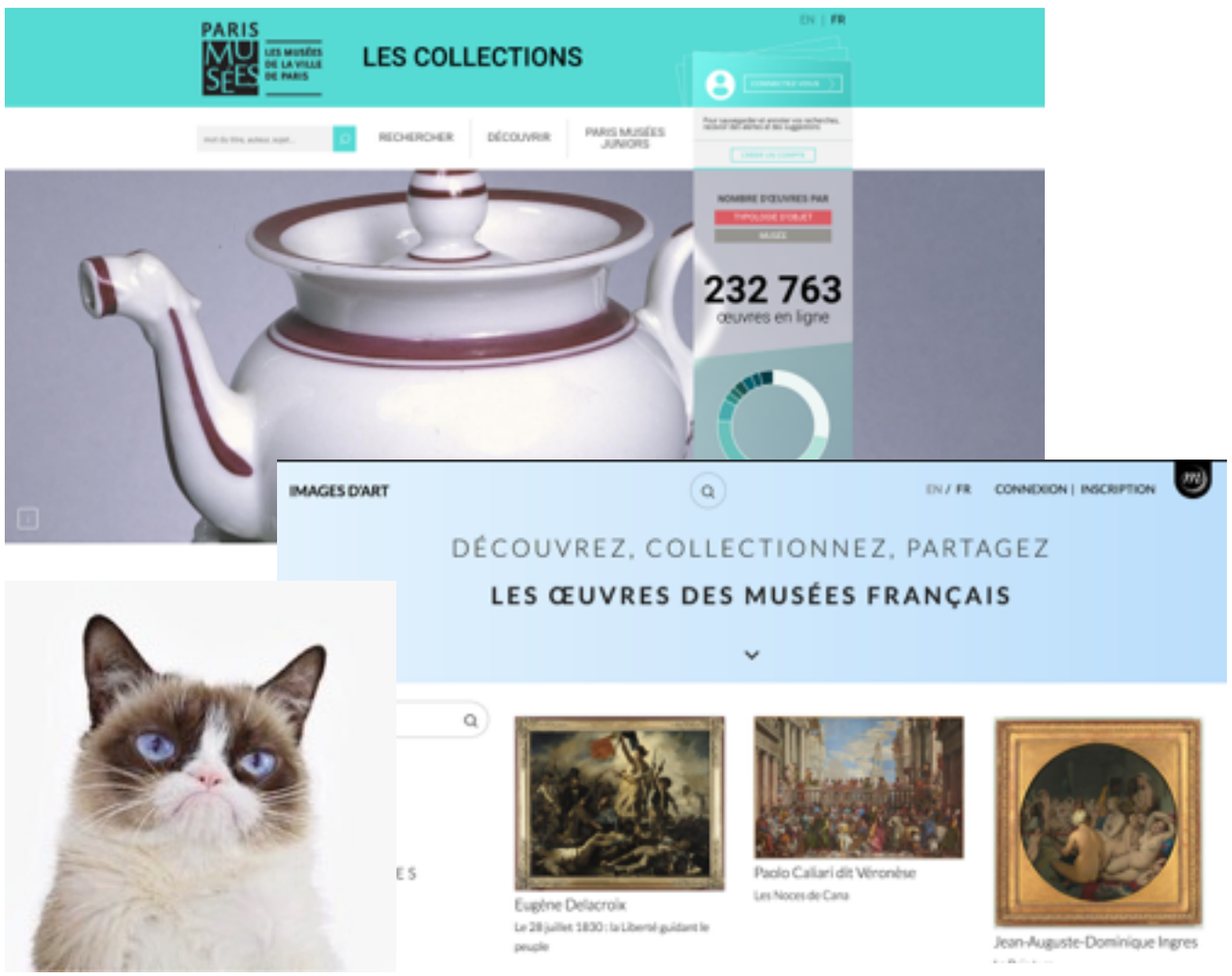
(l'article 5 précise que les administrations (qui dégagent au moins 25% selon un décret) de ressources propres, peuvent établir une redevance de réutilisation lorsqu'elles sont tenues de couvrir pour des recettes propres une part substantielle des coûts liés à l'accomplissement de leurs missions de service publics).

D'après les slides de Lionel Maurel https://goo.gl/EXHtQ2

Loi Lemaire - 7 octobre 2016

**Au final, l'Open Data par défaut est consacré (Article 6)...**

# Mais les institutions culturelles restent libres de décider d'ouvrir ou non (avec ou sans redevance)



# Des initiatives (très) récentes et (très) encourageantes

ARCHIVES NATIONALES
Fontainebleau - Paris - Pierrefitte-sur-Seine

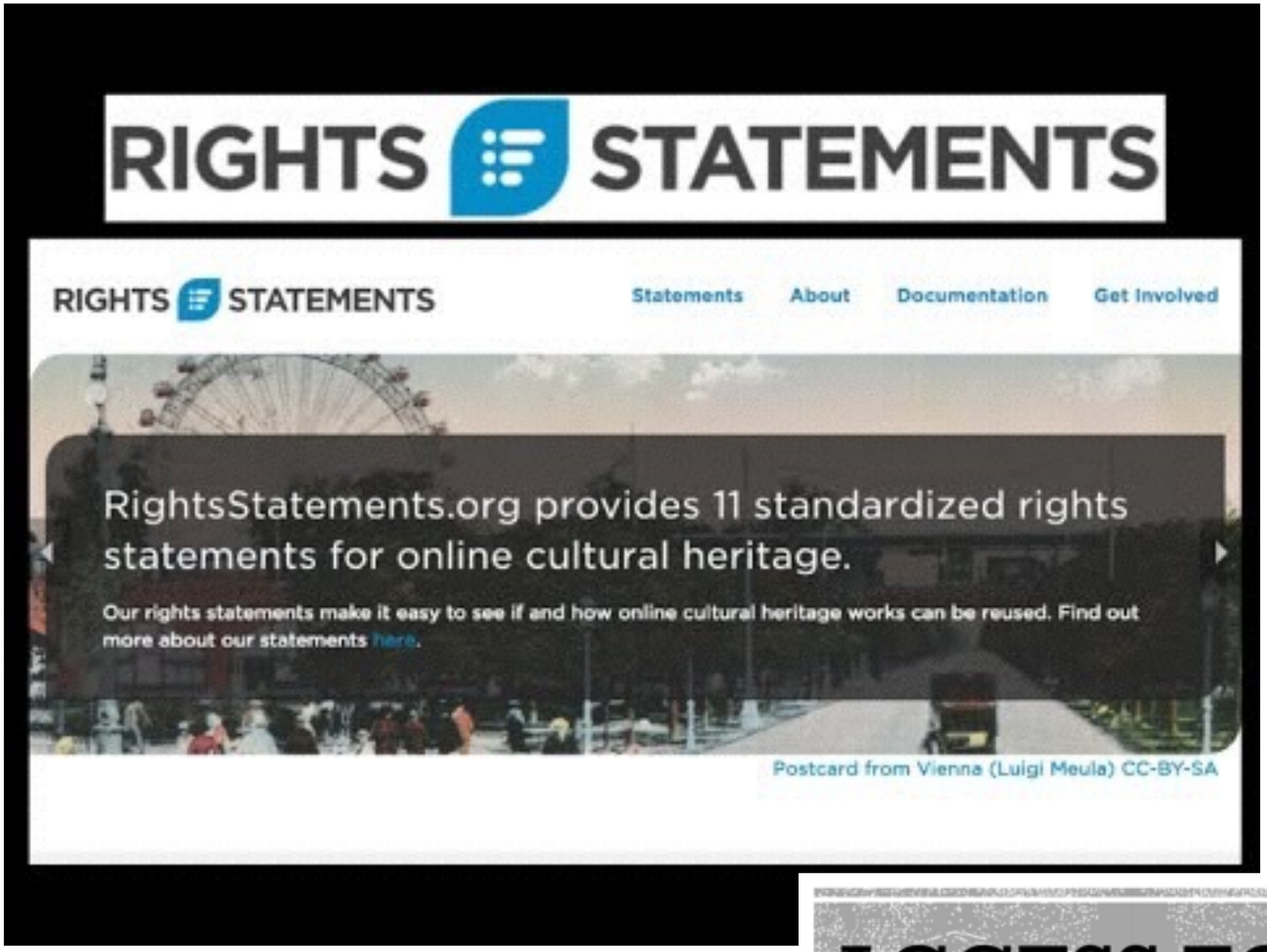
**Juillet 2017 : adoption d'un nouveau règlement pour appliquer par défaut un principe de gratuité valable pour « les documents librement communicables à tous et sur lesquels des tiers ne détiennent pas de droits de propriétés intellectuelles ».**

Inscription au barcamp du 9 décembre 2017:
https://goo.gl/MwHJXW

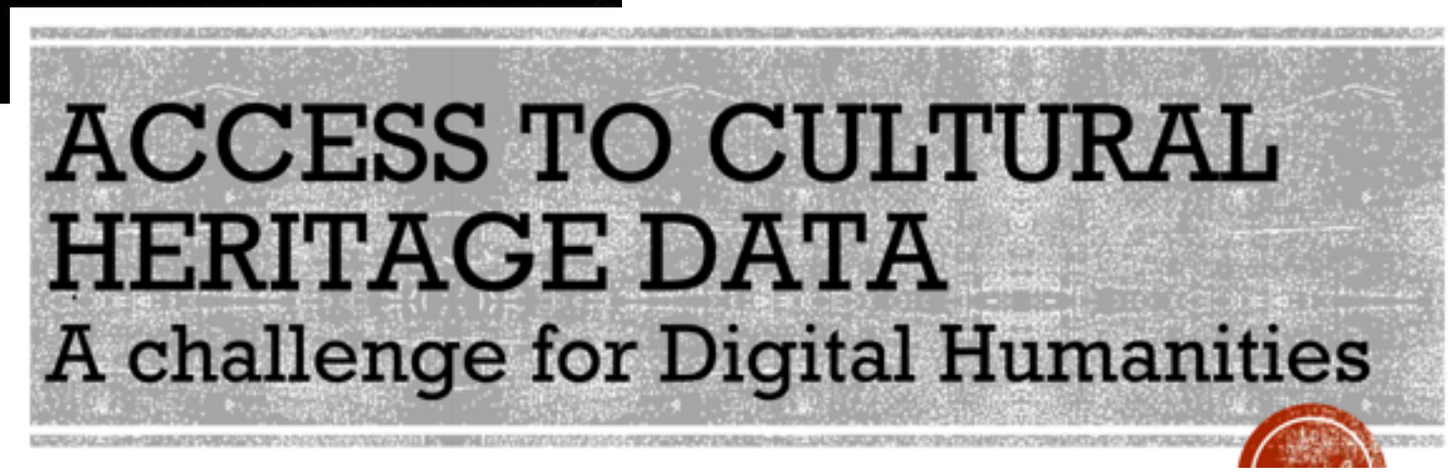
Lancement lors des JEP 2017, du portail des colletions du musée de Bretagne (des Champs-Libres).

- Intitulé **Des collections en partage**
- respect de bonnes pratiques (ARK, etc.)
- collaboration avec la communauté wikipedia et OpenStreetMap

**collections.musee-bretagne.fr/**




http://rightsstatements.org/en/




Rapport d'Anne Baillot, Marie Puren, Charles Riondet, Dorian Seillier and Laurent Romary https://hal.archives-ouvertes.fr/hal-01582176/file/dh2027_CHDRC.pptx.pdf

« la problématique de l'ouverture des contenus culturels ou open content doit être dissociée de la démarche d'ouverture et de partage des données publiques culturelles. ».

Camille Domange, *Guide DataCulture,* 2013

Les œuvres du domaine public numérisées ne sont en effet pas ce qu'on a généralement en vue quand on parle de données publiques, et il semble que cette notion soit inappropriée pour désigner de tels objets. On a plutôt affaire à des contenus culturels, et le terme de données s'appliquerait avec plus d'à-propos aux métadonnées qui les accompagnent.

Laura Le Coz, *Patrimoine numérisé et Open Content : Quelle place pour le domaine public dans les bibliothèques numériques patrimoniales ?,* Mémoire de l'ENSSIB, 2017

**Courtin Antoine**
@seeksanusername

SI t'es pas geek, tu as du 1024x1480 pour des manuscrits du XIV contre du 4686 × 6635 si tu sais jouer avec l'URL #FractureNumériqueRempante

23:03 - 21 juil. 2016

**6** Retweets **7** J'aime

2 6 7

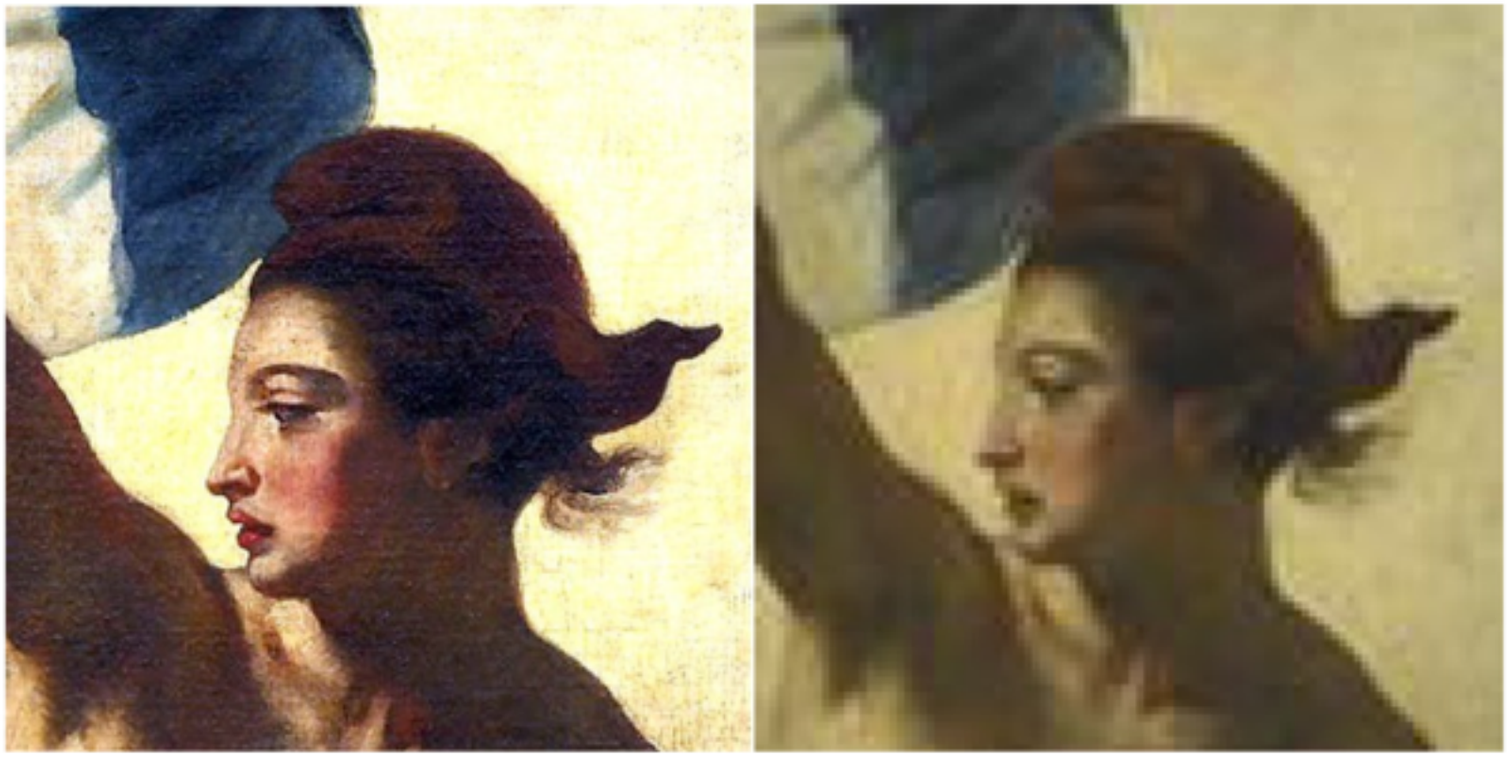
**Courtin Antoine**
@seeksanusername

De l'importance de la #HD en histoire de l'art #fracturenumérique #OpenGLAM

19:59 - 29 mars 2017

**3** J'aime

3



À gauche la version disponible sur Wikimedia Commons, à droite celle d'Images d'art.

D'après l'article de Sylvain Machefert : https://goo.gl/oscZsk

- Les images, les « oubliées » de l'effort d'ouverture d'interopérabilité ?

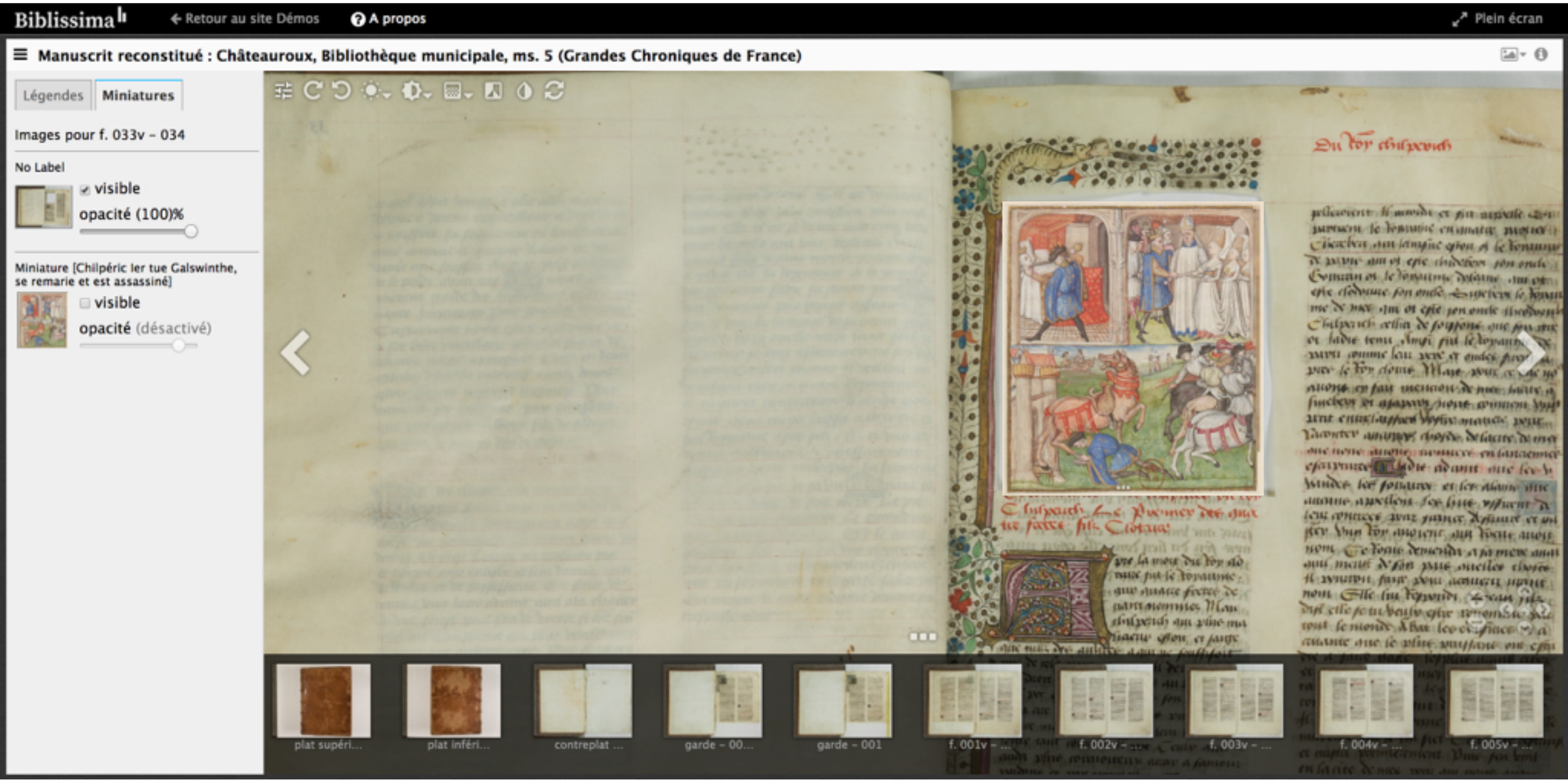
International
Image
Interoperability
Framework

Le but de IIIF est de créer un cadre technique commun grâce auquel les bibliothèques numériques peuvent délivrer leurs contenus de manière standardisée sur le Web afin de les rendre consultables, manipulables et annotables par n'importe quelle application ou logiciel compatible.

Régis Robineau, https://insula.univ-lille3.fr/2016/11/comprendre-iiif-interoperabilite-bibliotheques-numeriques/

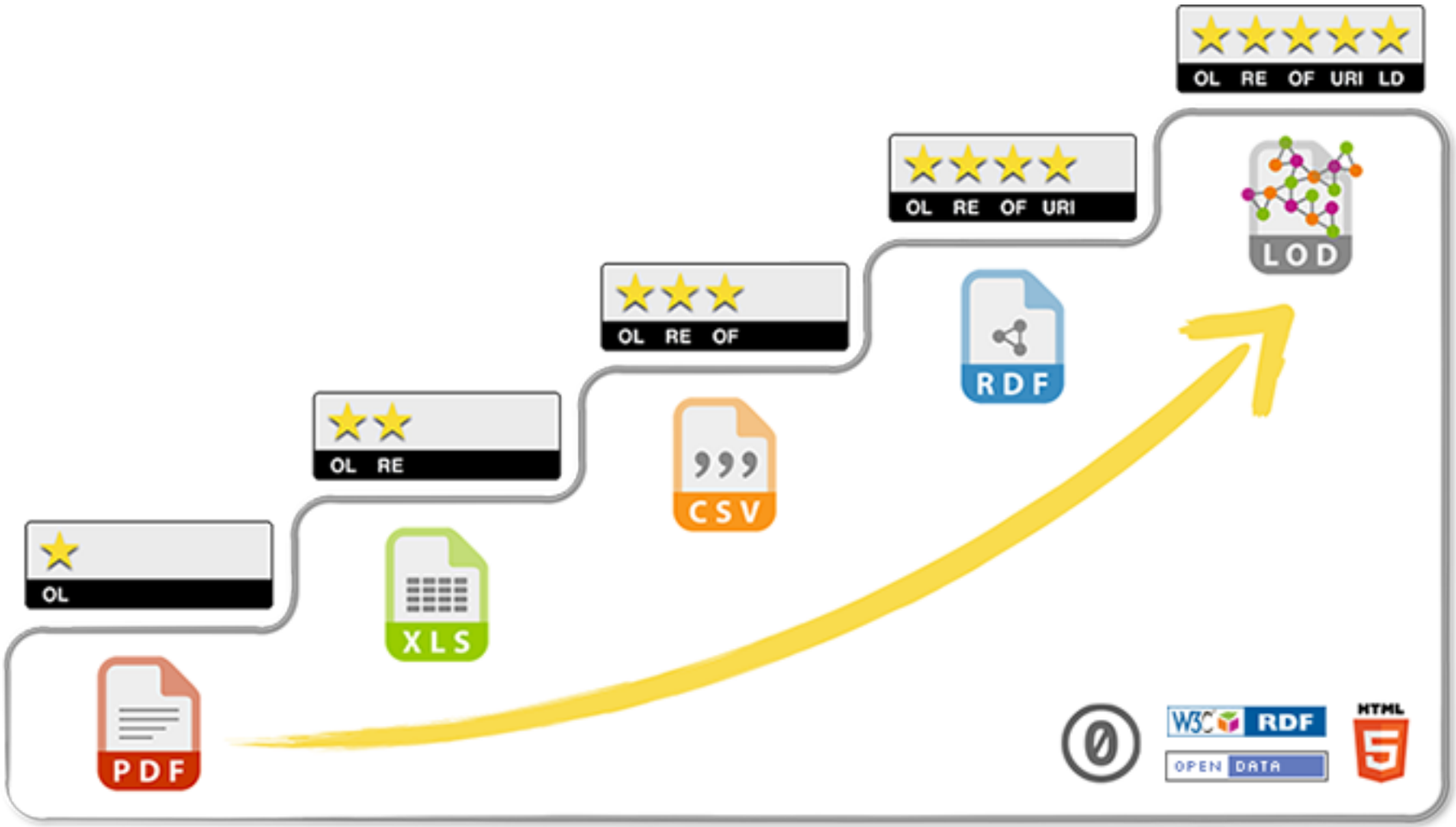
Ce cadre technique est évolutif se compose:

- d'un modèle de données : Shared Canvas (http://iiif.io/model/shared-canvas/1.0/)
- de 4 APIs fonctionnant de manière conjointe et complémentaire :
  - API Image 2.1 : http://iiif.io/api/image
  - API Presentation 2.1 : http://iiif.io/api/presentation
  - API « Content Search » 1.0 : http://iiif.io/api/search
  - API Authentification 1.0 : http://iiif.io/api/auth/1.0/




http://demos.biblissima-condorcet.fr/chateauroux/

La numérisation des vignettes d'un côté, du manuscrit de l'autre, permet un repositionnement virtuel, facilité par certaines avancées technologiques en matière de visualisation et d'interopérabilité des images.

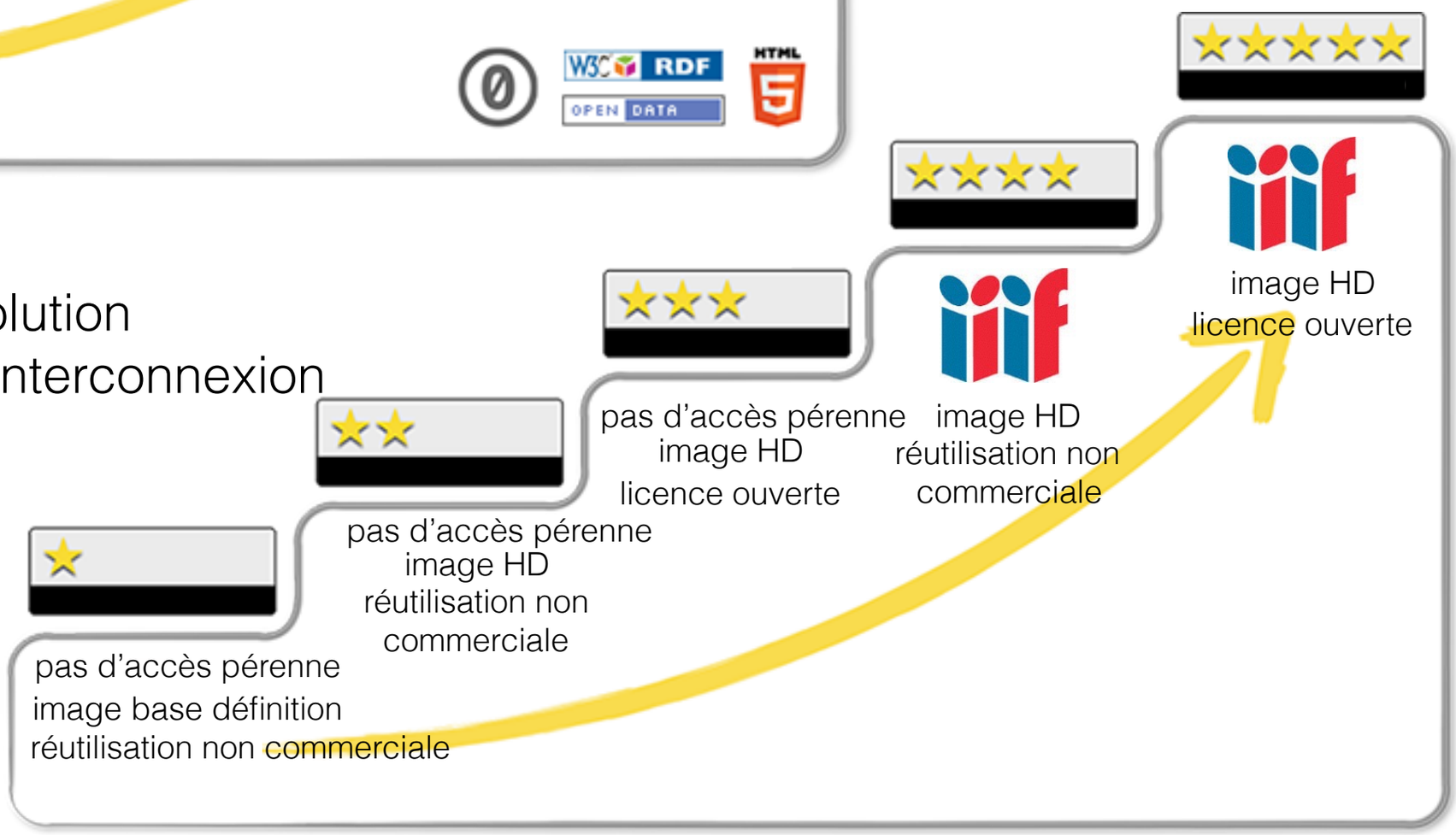



Tim Berners-Lee, the inventor of the Web and Linked Data initiator, suggested a 5-star deployment scheme for Open Data. Here, we give examples for each step of the stars and explain costs and benefits that come along with it.

http://5stardata.info/en/

3 dimensions d'analyse :

- le format / qualité / résolution
- accès / manipulation / interconnexion
- licence / réutilisation



- **Télécharger la version HD d’une numérisation dans Gallica**

http://gallica.bnf.fr/ark:/12148/btv1b53084829z/

gallica.bnf.fr/**iiif**/ark:/12148/btv1b53084829z/**f1/full/full/0/native.jpg**

https://medium.com/@seeksanusername/astuce1-r%C3%A9cup%C3%A9rer-de-la-hd-sur-gallica-bef0a6cc7f89

# IIIF-Download

Extension for Firefox and Chrome browsers offering user to download a high-def version of images available in digital libraries supporting the IIIF protocol (only Gallica for the moment).

## Téléchargement

pour Firefox

pour Chrome

## Historique

**0.1 (2017-11-28)**

- Première version disponible. Téléchargement des images depuis Gallica.

http://www.geobib.fr/tool/iiif/

- actionpyllicalabs.py
- actionpyllicalabsjpg.py
- actionpyllicalabsjpgpress.py
- actionpyllicalabspdf.py
- pyllica.py
- pyllicalabs.py
- pyllicalabsjpg.py
- pyllicalabsjpgpress.py
- pyllicalabspdf.py

fourchette définie par l'utilisateur;

- Pyllicalabspdf permet d'extraire les pdf de numéros de périodiques;
- Pyllicalabsjpg permet d'extraire les pages d'un document sous forme de fichier jpg, tif ou png, avec une qualité définie par l'utilisateur;
- Pyllicalabsjpgpress permet de faire la même opération pour des numéros de périodiques.

Pour chacun des outils, il y a deux fichiers: un fichier contenant le programme, et un fichier d'instruction (commençant par le mot « action »). Pour utiliser un outil, il suffit de placer les deux fichiers dans un nouveau dossier et de modifier le fichier d'instruction en fonction des livres ou périodiques que l'on veut télécharger. Sous Mac, s'il y a des problèmes d'encodage du texte au moment de l'utilisation des outils, il ne faut pas cliquer sur le fichier d'instruction pour le lancer mais démarrer IDLE depuis depuis le Terminal ($ idle3) et faire Fichier > Ouvrir pour lancer l'action.

### Pyllicalabs

On modifie le contenu du fichier actionpyllicalabs:

```
textpress(url="http://gallicalabs.bnf.fr/ark:/12148/cb32817642h/date",
        title="lemoderniste",
        year=1889,
        month=5,
        day=25,
        item=52,
        rate=7,
        lastpage=11)
```

La fonction comprend les commandes suivantes :
**url**: on indique l'adresse sur Gallica de la page du périodique indiquant toutes les années disponibles.
**title**: on choisit un titre qui sera indiqué dans le nom du fichier.

https://github.com/Dorialexander/Pyllica

Un site du ministère de la Culture

**data.culturecommunication.gouv.fr**

la plate-forme de données ouvertes du ministère de la Culture

Les données | La démarche | Créer vos cartes | API | Contact

589 277

**Collections des musées de France : extrait de la base Joconde**

**Camille SC** @CamilleSC2 · 4 h

En réponse à @shona_gon

oh, y'a une carte ! ah, crotte, les 182 851 oeuvres de Paris sont toutes à la station Châtelet...

1

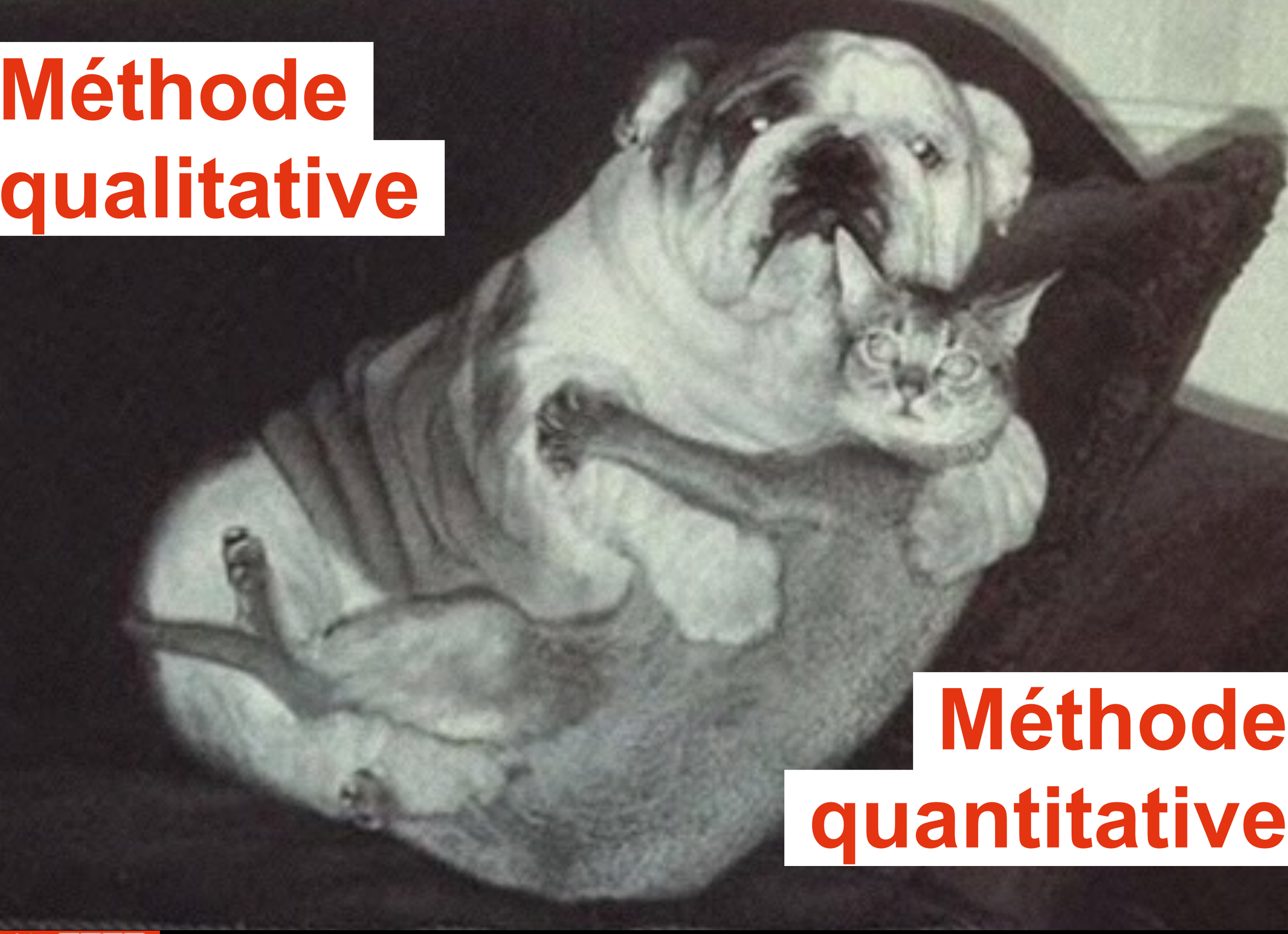
› Plus

| Dénomination | |
|---|---|
| album | 68 469 |
| tableau | 64 624 |

| | |
|---|---|
| Langue | Français |
| Modifié | 3 avril 2018 09:40 |
| Producteur | Ministère de la Culture |
| Dernier traitement | 3 avril 2018 17:01 (métadonnées)<br>3 avril 2018 17:01 (données) |

https://data.culturecommunication.gouv.fr/explore/dataset/base-joconde-extrait/information/

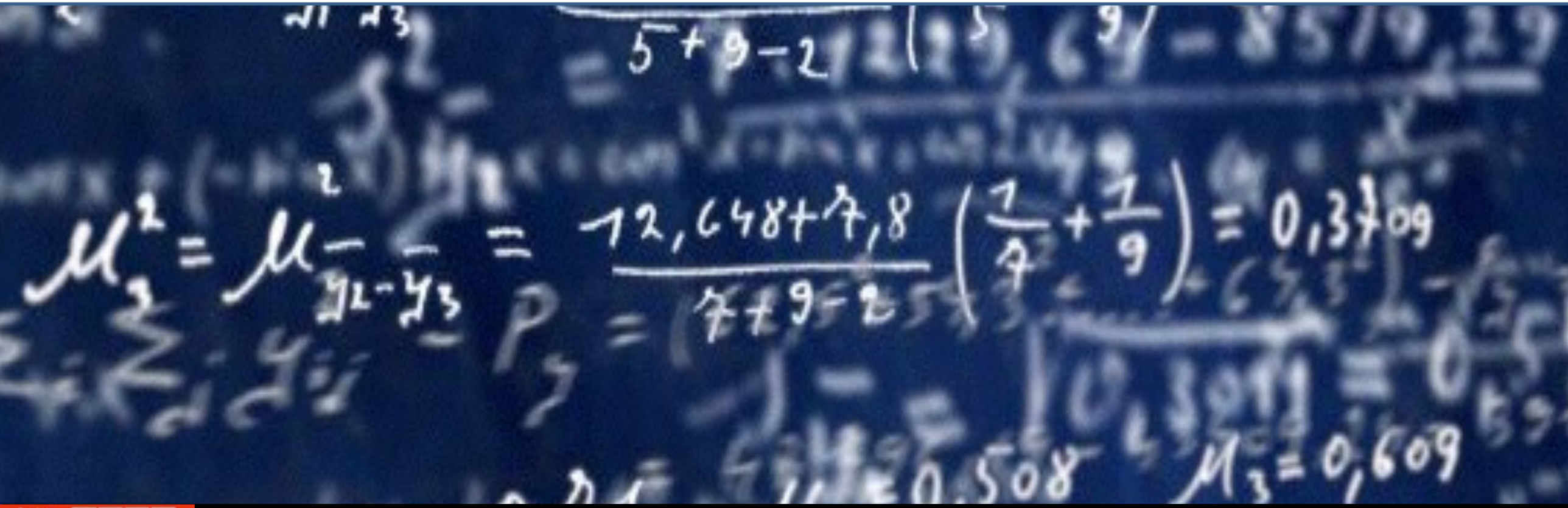
# Données brutes ?



# Méthode qualitative

# Méthode quantitative

# Boite noire ? / Algorithmes



# Mahsup / InfoVizualisation

Un mashup est une œuvre/ création originale mise au point à partir de l'assemblage/ mise en relation d'œuvres/ créations existantes.

Le mot mashup est le plus souvent utilisé pour la musique mais on le retrouve dans les autres arts ainsi que dans le monde numérique lorsqu'on parle de mashup de données.

définition reprise de Gautier Poupeau

C’est l’ensemble des techniques de représentation graphique et d’exploration visuelle de données quantitative et/ou statistiques permettant de traduire un ensemble de données brutes en information décisive

# Une/Des terminologies

- Dataviz/datavision
- Visualisation de données
- Design d’information
- Infographie
- etc,

# Infovisualisation comme…

… outil d’aide à la décision / contrôle qualité

… outil d’accès et d’exploration de ressources

… outil d’aide à l’interprétation

# Mettre en image/graphiques des informations/ données

- Faire « parler » des données (brutes) ?
- Traduire un important corpus de manière visuelle
- Synthétiser les enjeux essentiels d’un corpus

# Visualisation de données ou infographie ?

| Une infographie… c'est plutôt : | Une dataviz… c'est plutôt : |
| --- | --- |
| ■ favoriser la compréhension d'une information que l'on connait déjà en la représentant sous forme graphique | ■ faire émerger une information que l'on ignorait en analysant des données présentées sous forme graphique |
| ■ un volume modeste de données | ■ un gros volume de données |
| ■ une approche didactique tournée vers autrui | ■ d'abord un outil de connaissance pour soi-même |
| ■ une aide à la compréhension | ■ une aide à la décision |

# A quoi sert la visualisation de données

## Ce qu'en dit Jacques Bertin.

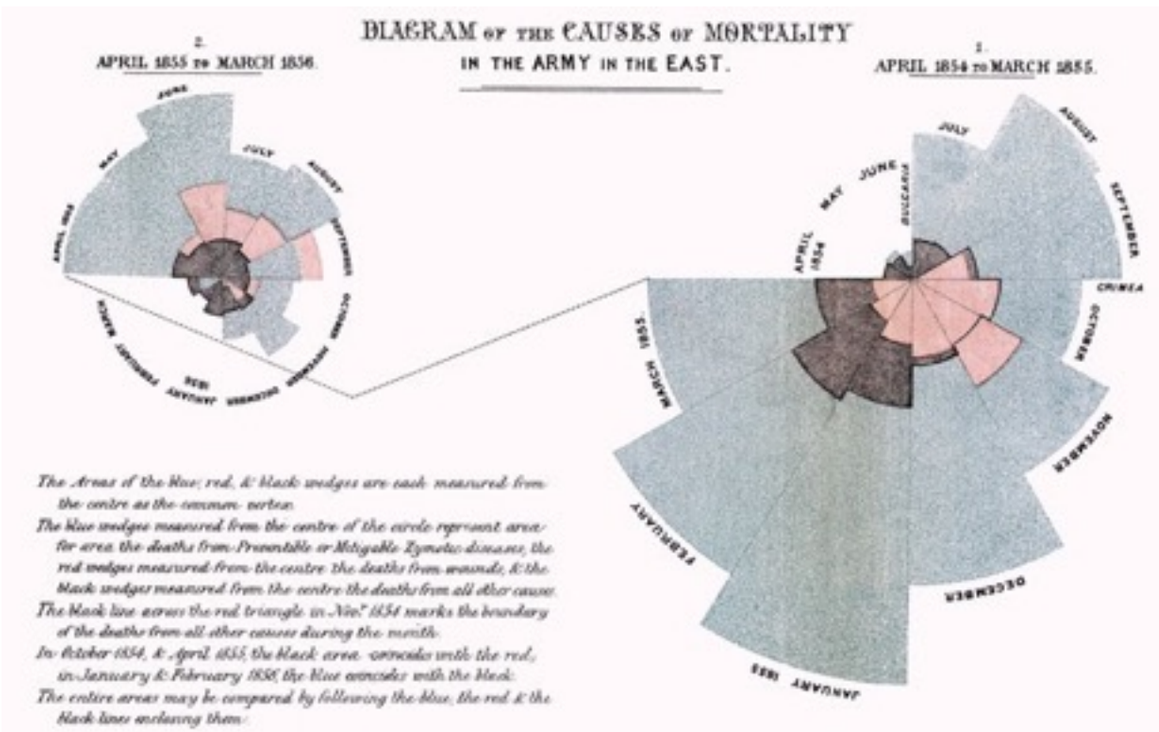
1. A traiter des données pour comprendre et et en tirer de l'information
   → Idée de la fouille de données
   → Traiter des données implique l'exhaustivité.

2. A communiquer cette information
   → Il faut quelques heures pour construire un graphique ; quelques secondes pour le lire.
   → Communiquer implique (souvent) la simplification.

source Christophe Bontemps

En fait de calculs et de proportions, le plus sûr moyen de frapper l'esprit, est de parler aux yeux

William Playfair, 1780



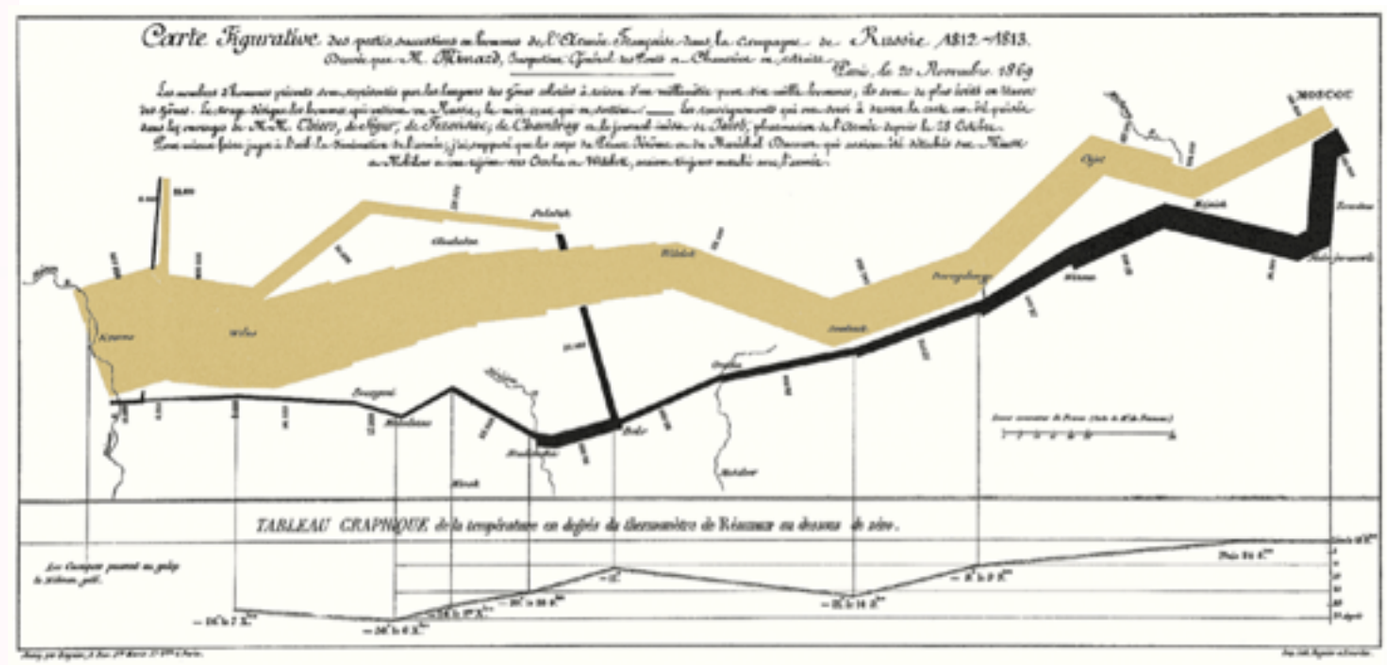
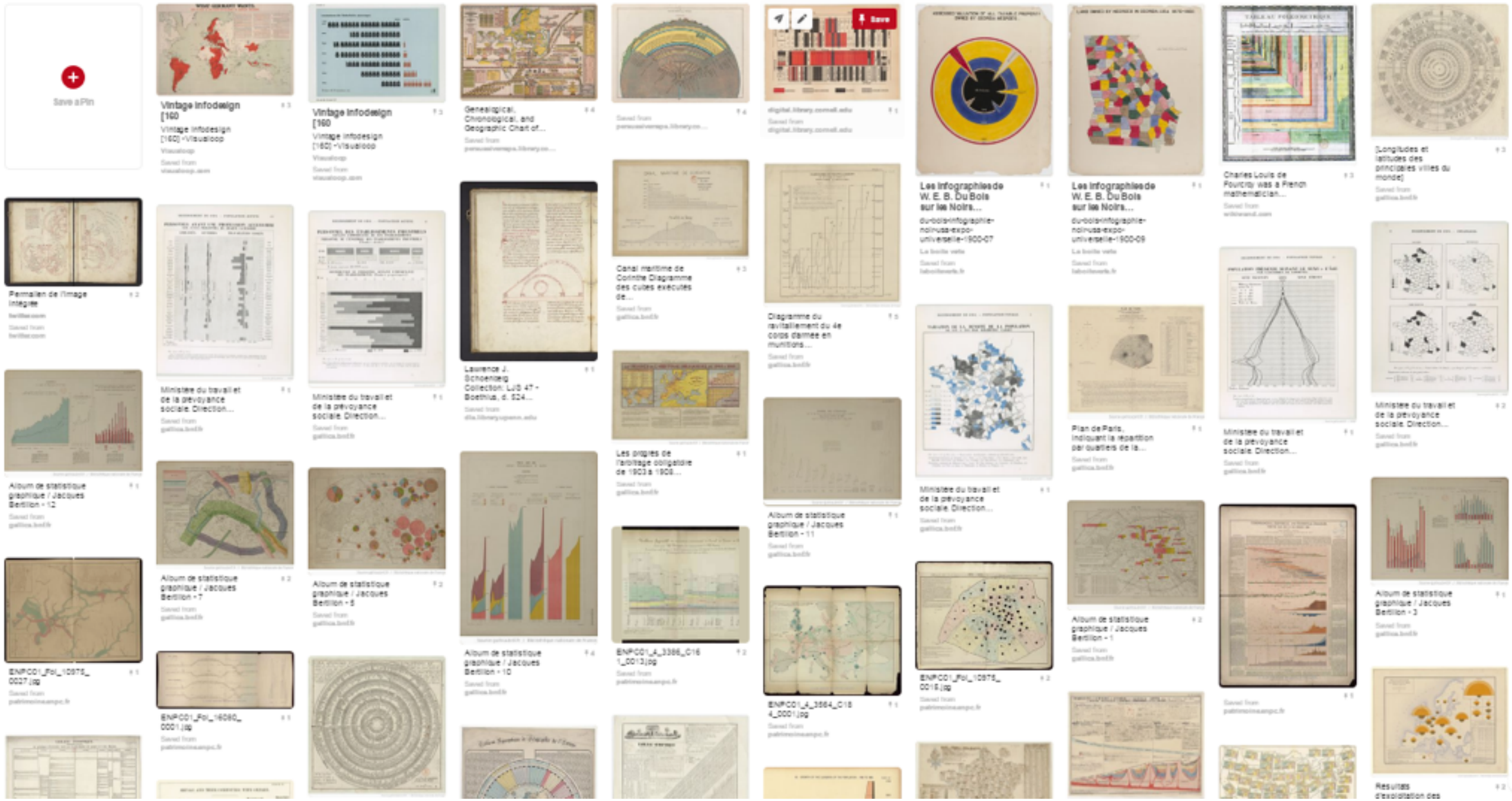




*Diagramme des causes de mortalité au sein de l'armée en Orient* par Florence Nightingale.

Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie 1812-1813 par Charles-Joseph Minard.

# InfoViz Vintage

88 Pins
106 Followers
Sélection personnelle et subjective d'infoviz/dataviz datant des siècles derniers, issues des bibliothèques numériques telles que Gallica, DPLA etc.




https://fr.pinterest.com/antoinecourtin/infoviz-vintage

# Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization

An illustrated chronology of innovations by **Michael Friendly** and **Daniel J. Denis**



## Pre-1600: Early maps and diagrams

The earliest seeds of visualization arose in geometric diagrams, in tables of the positions of stars and other celestial bodies, and in the making of maps to aid in navigation and exploration. We list only a few of these here to provide some early context against which later developments can be viewed.

In the 16th century, techniques and instruments for precise observation and measurement of physical quantities were well-developed. As well, we see initial ideas for capturing images directly, and recording mathematical functions in tables. These early steps comprise the beginnings of the husbandry of visualization.

Jump to Milestone...

### 6200 BC Unknown — Oldest known map?

Added: 2005-10-01

**The oldest known map? (There are several claimants for this honor.)**

Town map, with an errupting volcano (Hasan Daö?) and the Konya plain

An extended description of the most ancient maps

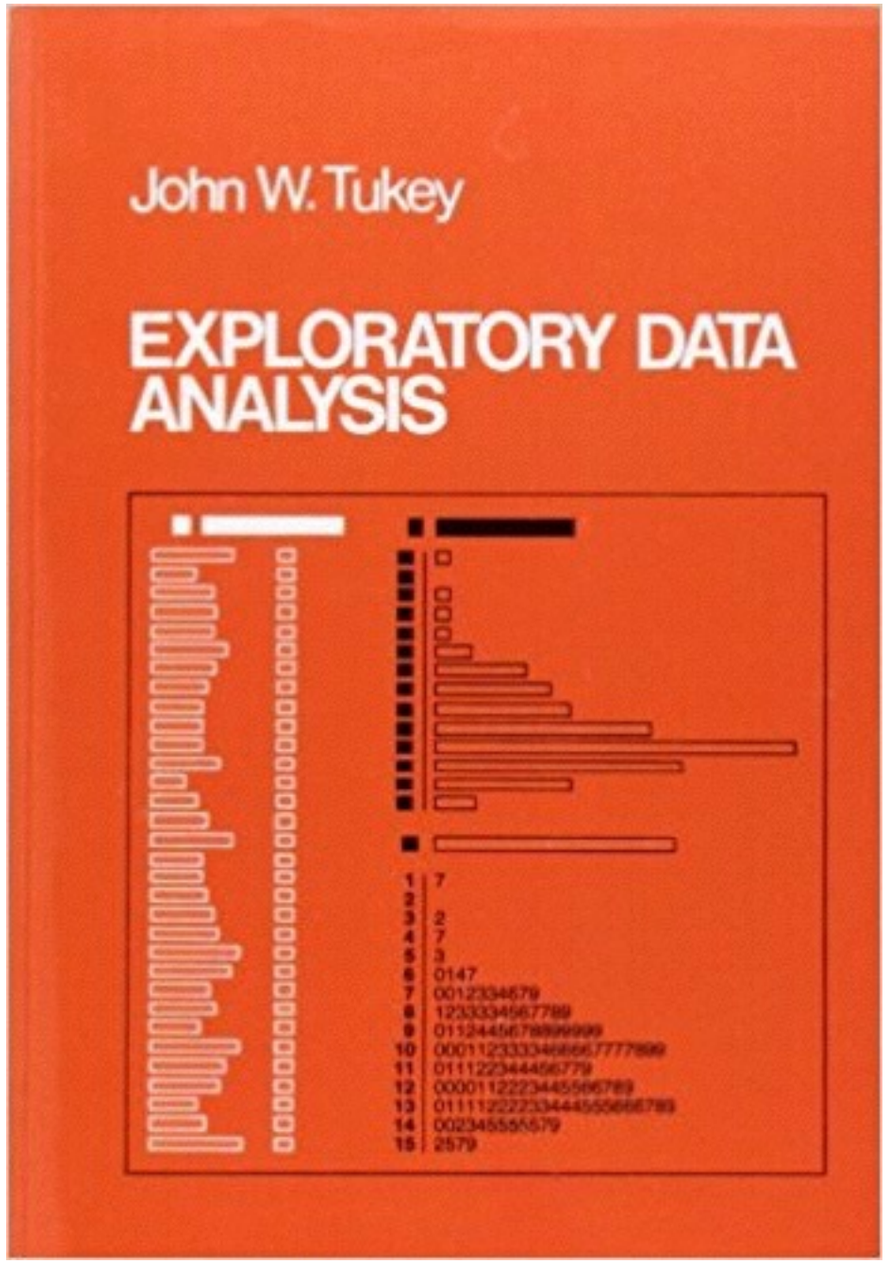
Konya town map

**References:**

*none*

Mickaël Friendly, Milestones in the history of thematic cartography,statistical graphics, and data visualization, http://www.datavis.ca/

- Les visualisations profitent des caractéristiques du système de traitement cognitif humain. D'après Colin Ware, les visualisations offrent de nombreux avantages dont les principaux sont les suivants :
  - Les visualisations graphiques ont le pouvoir de permettre à l'homme de manipuler des structures bien plus complexes représentées par une visualisation (représentations externes) que dans la mémoire de travail visuelle et verbale (représentations internes).
  - Elles permettent aussi de percevoir l'émergence de propriétés dans les données cartographiées qui ne sont pas anticipées.
  - Les visualisations permettent de mettre en évidence des problèmes dans les données, dans leur collecte. Avec une visualisation appropriée, les erreurs dans les données sont rapidement perceptibles.
  - La visualisation permet de percevoir simultanément des propriétés à grande et à petite échelle sur les données.
  - La visualisation facilite la formation d'hypothèses sur les données.



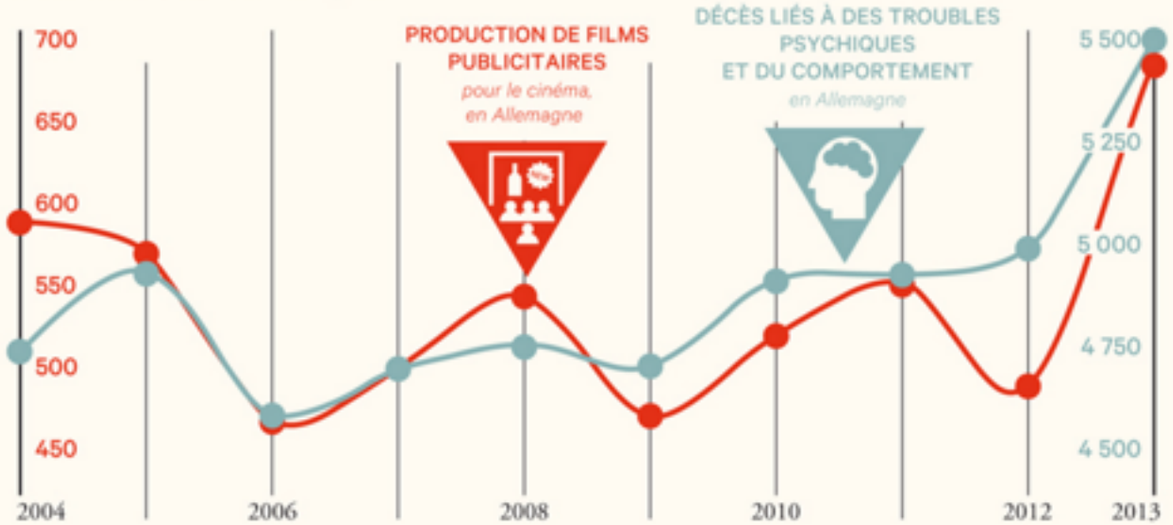

"Pour faire des découvertes surprenantes dans les données, le meilleur outil dont nous disposons est l'oeil, et sa faculté de lecture des images.

" L'approche exploratoire des données est un type d'analyse des données pour en résumer les principales caractéristiques d'une façon claire et compréhensible, souvent en utilisant des graphiques, sans utiliser de modèles statistiques, et en n'ayant pas non plus formulé d'hypothèse."

WARNING

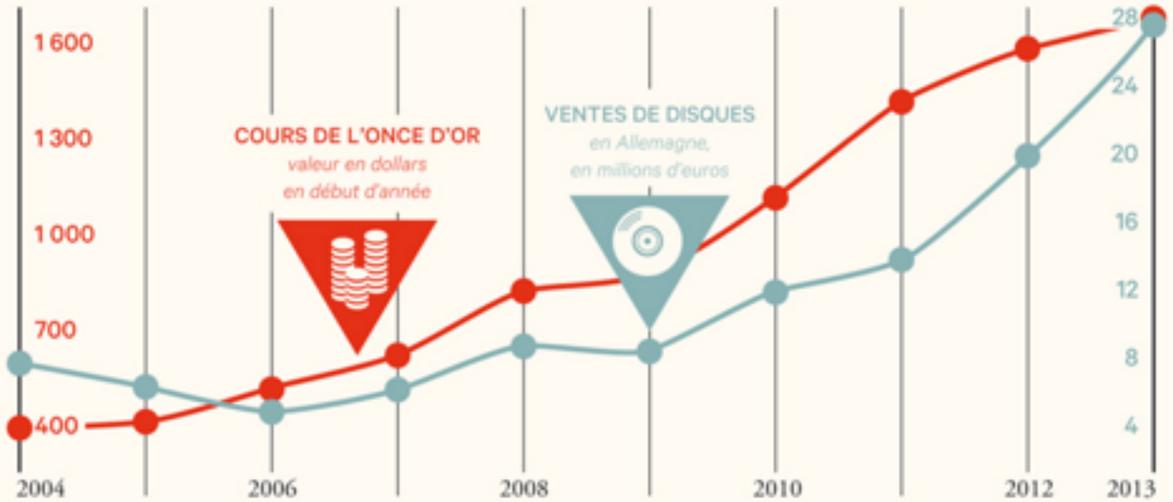
## PUBLICITÉ MENSONGÈRE

Les publicités au cinéma ont-elles une incidence sur le nombre de décès dans la population ?
Coefficient de corrélation : 0,792



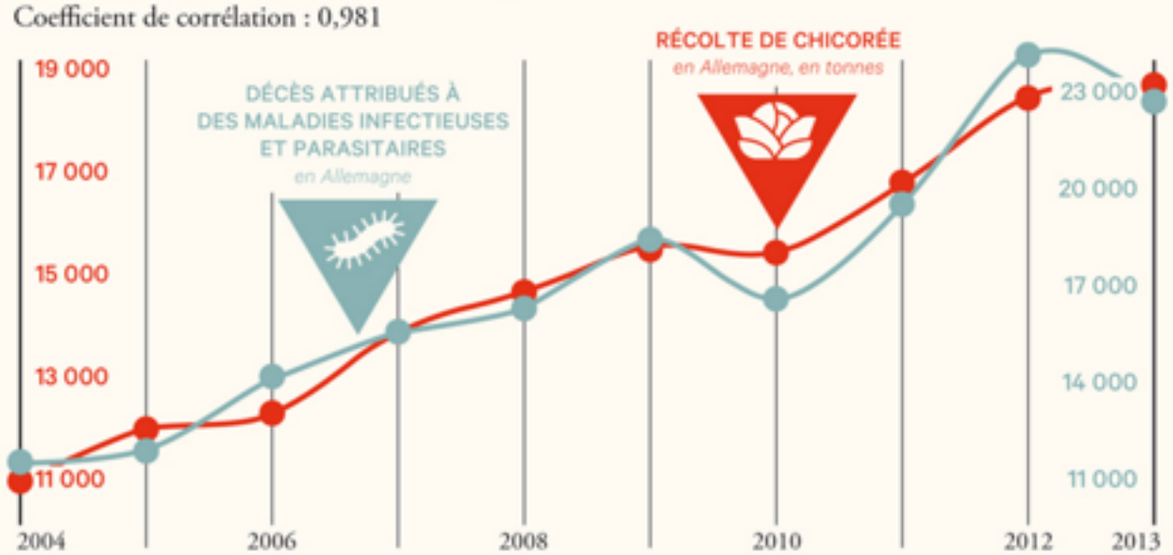

## DISQUE D'OR

Les ventes de disques vinyle font-elles monter le cours de l'or ?
Coefficient de corrélation : 0,905




## POISON VERT

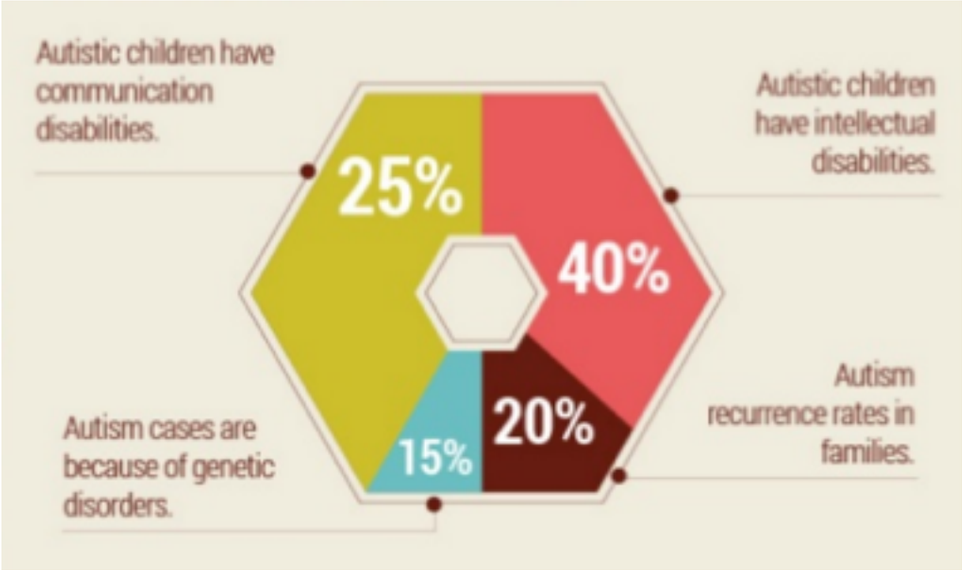
La chicorée consommée en salade est-elle responsable du nombre de maladies infectieuses ?
Coefficient de corrélation : 0,981

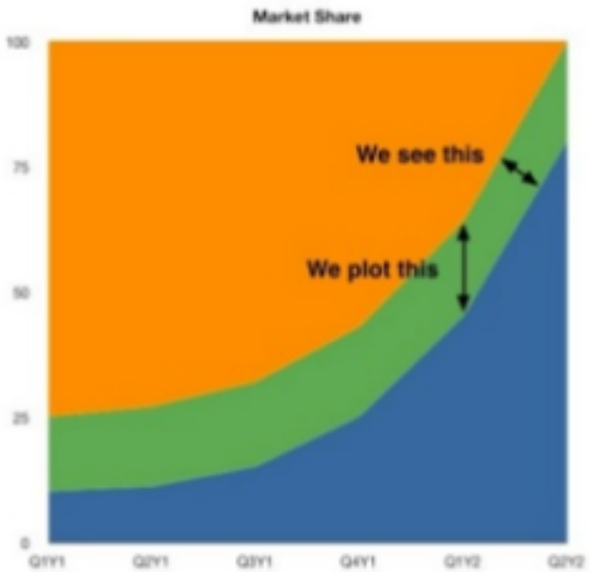



## ON ACHÈVE BIEN LES CHEVAUX

Les abattoirs de chevaux augmentent-ils leur activité en fonction du nombre de fonctionnaires ?
Coefficient de corrélation : 0,821

Statistiques. Les corrélations de l'absurde, http://goo.gl/9mM7zl



Martin Grandjean. Introduction `a la visualisation de donn´ees : l'analyse de r´eseau en histoire. Geschichte und Informatik, Chronos, 2015, pp.109-128.

## *The Rules:*

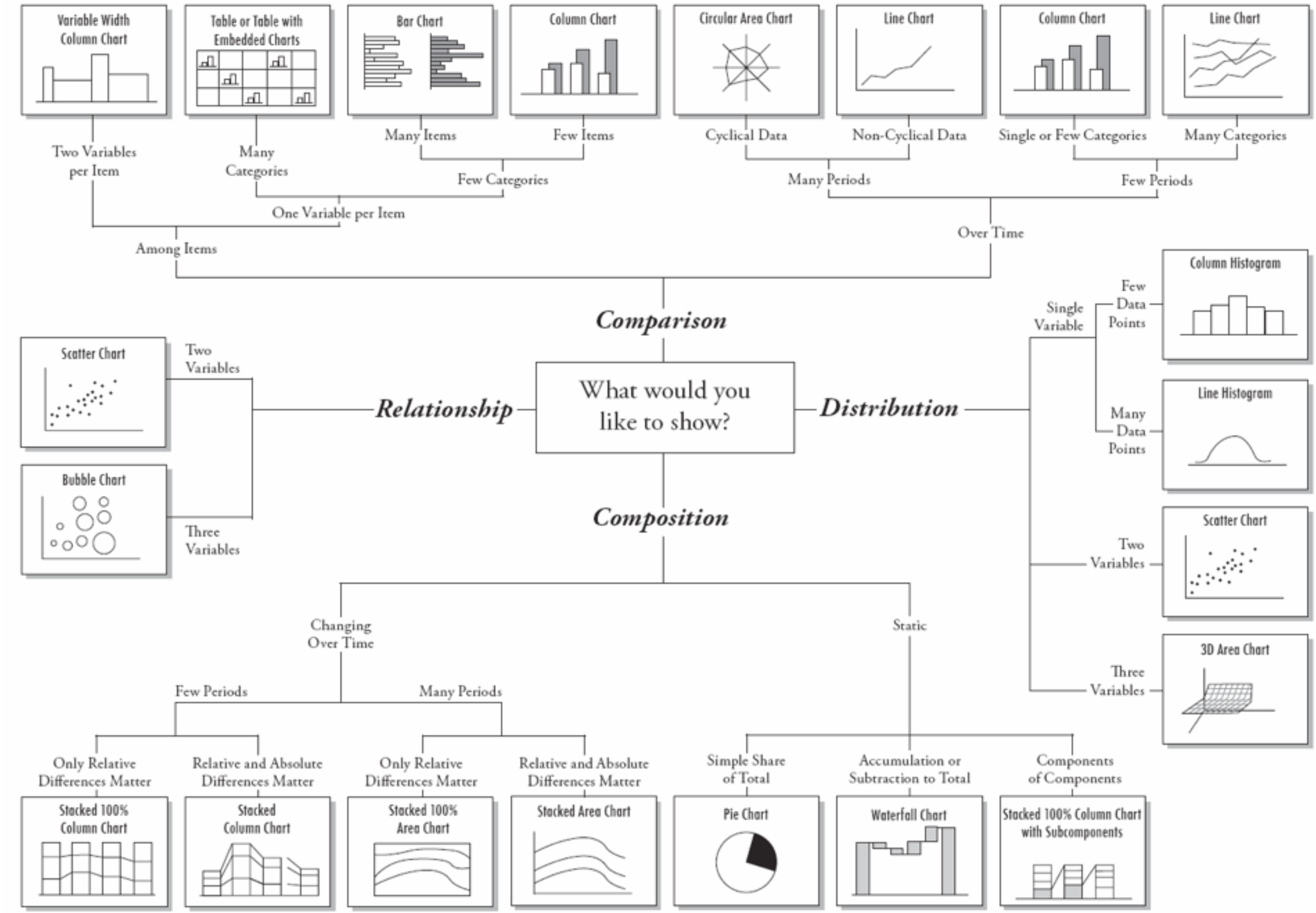
1. Show Your Data
2. Use Graphics
3. Avoid Chartjunk
4. Utilize Data-ink
5. Use Labels
6. Utilize Micro/Macro
7. Separate Layers
8. Use Multiples
9. Utilize Color
10. Understand Narrative

http://www.sealthreinhold.com/school/tuftes-rules/

Dans la vision académique d'Edward Tufte, l'excellence graphique peut être estimée au travers de 3 critères fondateurs :

- **data-ink ratio**
  - Le ratio entre informations restituées/encre utilisée ().
- **lie-factor**
  - distorsion de la restitution par rapport à la réalité des données sources).
- **chartjunk**
  - L'absence de scories graphiques (élément décoratif n'apportant aucune plus-value informative)

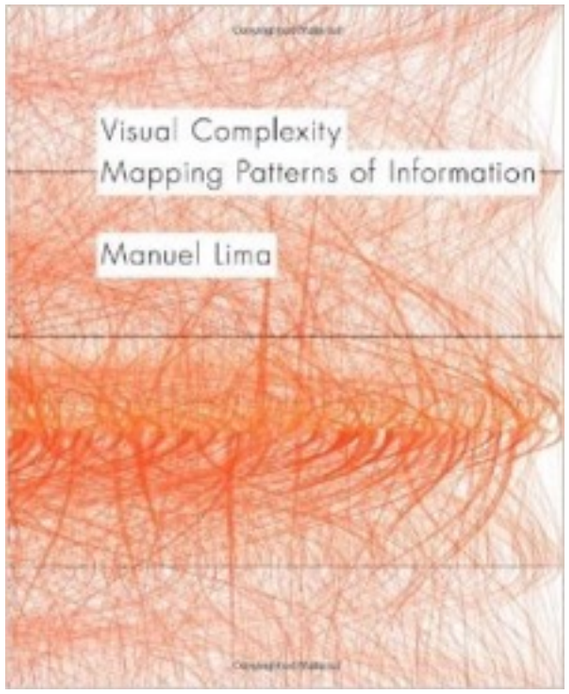
| | |
|---|---|
| Visualisation de données temporelles | Visualisation de flux |
| Visualisation de la relation entre plusieurs variables | Visualisation de données géolocalisées |
| Visualisation d’une arborescence | Visualisation des proportions |
| | Visualisation de réseaux |

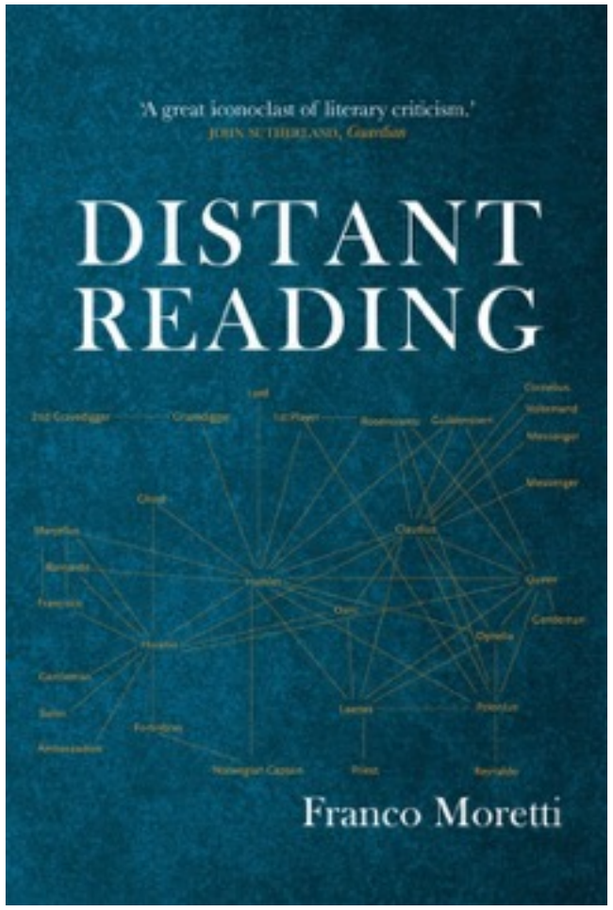

What would you like to show?
Comparison
Relationship
Distribution
Composition
Variable Width Column Chart
Table or Table with Embedded Charts
Bar Chart
Column Chart
Circular Area Chart
Line Chart
Column Chart
Line Chart
Two Variables per Item
Many Categories
Many Items
Few Items
Cyclical Data
Non-Cyclical Data
Single or Few Categories
Many Categories
Few Categories
Many Periods
Few Periods
One Variable per Item
Over Time
Among Items
Scatter Chart
Bubble Chart
Two Variables
Three Variables
Single Variable
Few Data Points
Many Data Points
Column Histogram
Line Histogram
Two Variables
Scatter Chart
Three Variables
3D Area Chart
Changing Over Time
Static
Few Periods
Many Periods
Only Relative Differences Matter
Relative and Absolute Differences Matter
Only Relative Differences Matter
Relative and Absolute Differences Matter
Simple Share of Total
Accumulation or Subtraction to Total
Components of Components
Stacked 100% Column Chart
Stacked Column Chart
Stacked 100% Area Chart
Stacked Area Chart
Pie Chart
Waterfall Chart
Stacked 100% Column Chart with Subcomponents

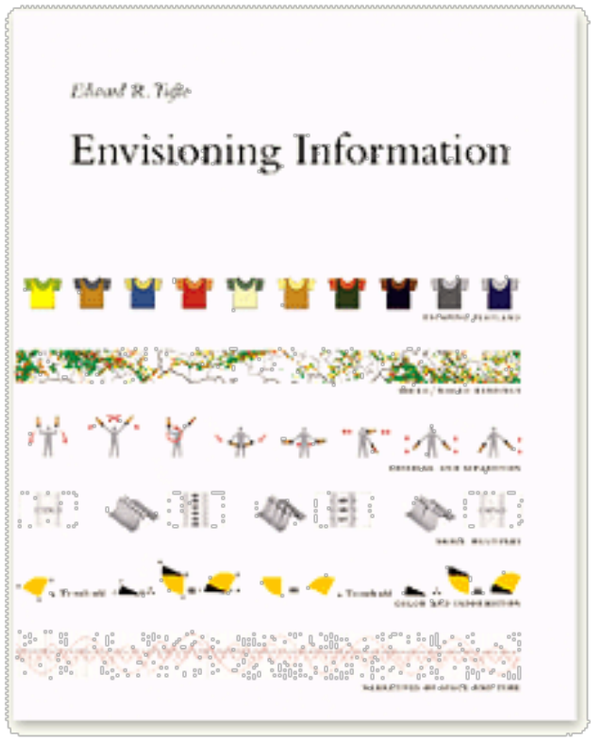
# Conception et source d’inspiration

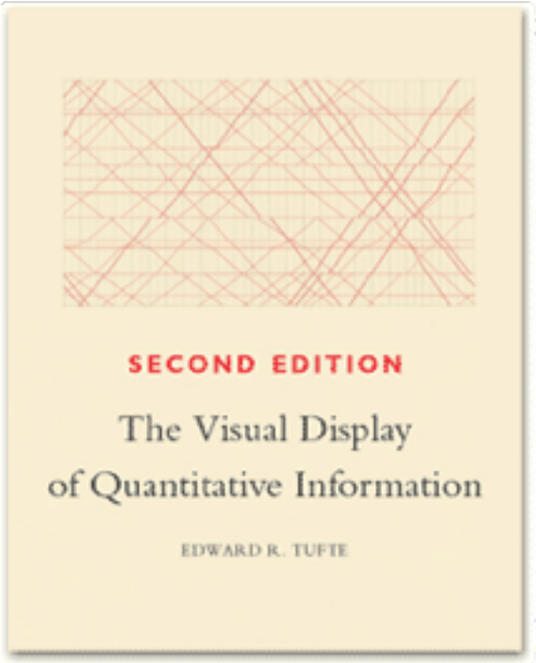
http://www.datavizcatalogue.com

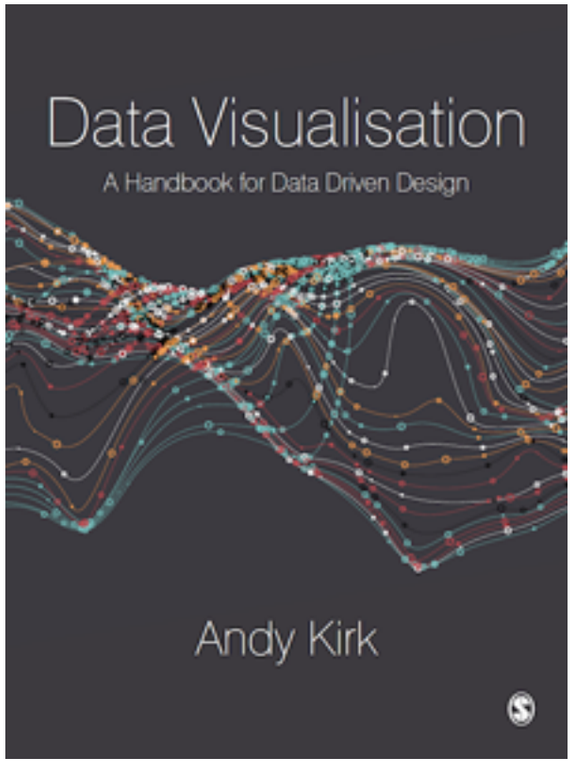
http://www.visualcomplexity.com/vc/

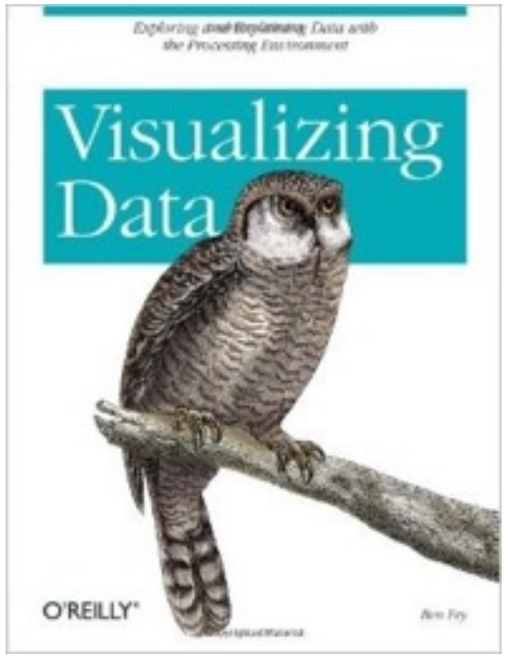

Visual Complexity
Mapping Patterns of Information
Manuel Lima

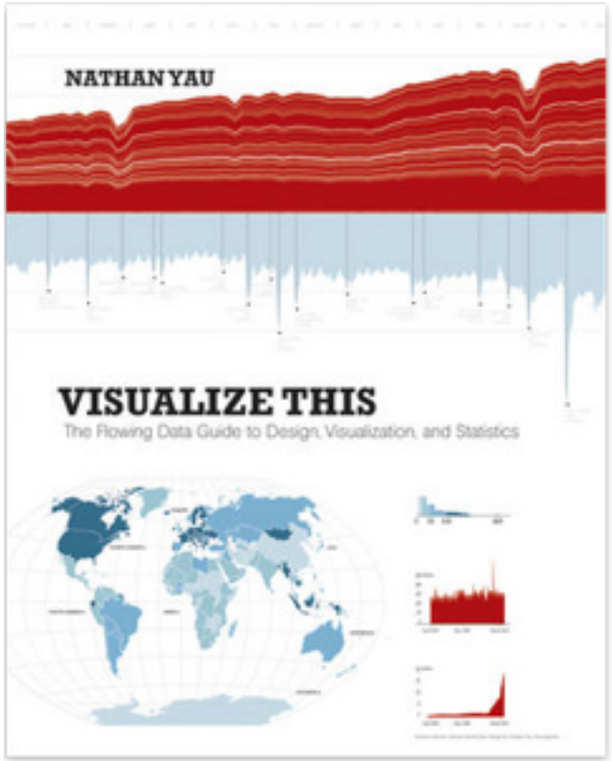

'A great iconoclast of literary criticism.'
DISTANT READING
Franco Moretti

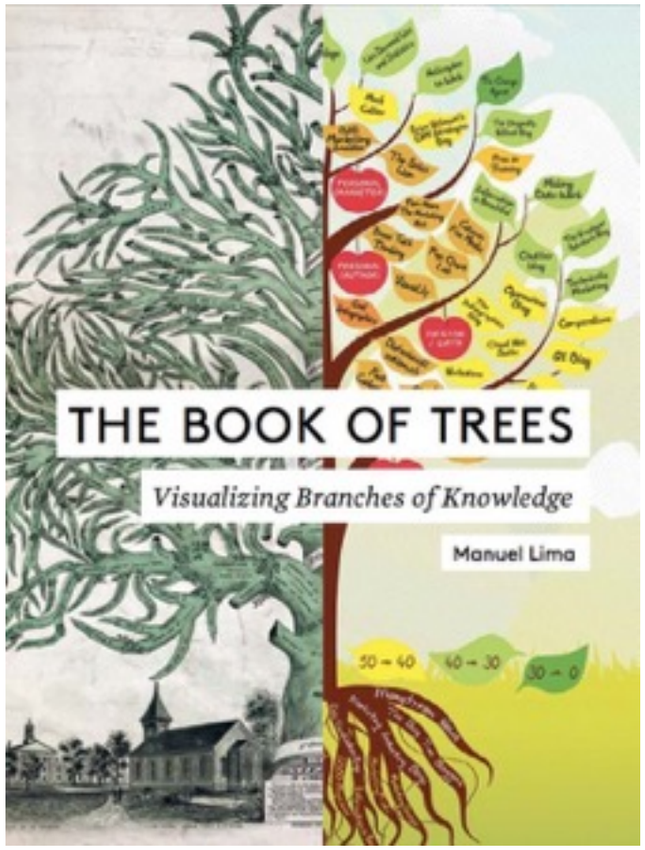

Edward R. Tufte
Envisioning Information


SECOND EDITION
The Visual Display of Quantitative Information
EDWARD R. TUFTE


Data Visualisation
A Handbook for Data Driven Design
Andy Kirk


Visualizing Data
O'REILLY


NATHAN YAU
VISUALIZE THIS
The Flowing Data Guide to Design, Visualization, and Statistics


THE BOOK OF TREES
Visualizing Branches of Knowledge
Manuel Lima

# Exemples

case study

antidot

Monuments historiques

Recherche

› Naviguer

» Type de monument
» Région / Département / Ville
» Période historique

› Explications

L'ouverture du site data.gouv.fr le 5 décembre 2011, aussitôt suivie d'autres initiatives, a marqué une accélération du mouvement Open Data en France.
Nous avons voulu apporter notre pierre à l'édifice, en réalisant une démonstration qui met en avant le grand intérêt qu'il y a à pouvoir mailler des données issues de différentes sources grâce aux standards du web sémantique, et la capacité de notre solution Antidot Information Factory à le faire rapidement et simplement, dans une approche industrielle.
Et parce que la France est la première destination touristique mondiale et que nos territoires regorgent de trésors architecturaux et patrimoniaux, nous avons choisi de réaliser une application de recherche qui vous permet de partir à la découverte près de 44.000 monuments historiques français !

>> En savoir plus

› Exemples de monuments

Mis à disposition sur Wikimedia Commons

## Anciennes Grandes Boucheries

**Type** : Monument
**Adresse** : 2 rue du Vieux-Marché-aux-Poissons
**Localité** : Strasbourg (Bas-Rhin)
**Liens externes** : Base Mérimée - Wikimedia Commons

***Description de l'ensemble inscrit ou classé*** *: Grandes Boucheries (anciennes) : classement par arrêté du 6 juillet 1928*

## Les musées en France

Service de recherche basé sur un ensemble de données culturelles liées et publiquement accessibles

< Retourner sur la page de résultats

### Musée des Augustins

Le musée des Augustins est le musée des beaux-arts de Toulouse en Haute-Garonne. Créé par décision du 23 décembre 1793 et ouvert en 1795, il est abrité dans l'ancien couvent des Augustins de Toulouse. Il rassemble des collections importantes de peinture et de sculpture.

**Direction**

Aucune information disponible

**Informations pratiques**

- Horaires
- Ouverture nocturne
- Téléphone
- Tarifs

  Collections permanentes : 3 € Réduit : 1,50 € Gratuit pour les moins de 18 ans, les étudiants et pour tous le premier dimanche du mois
- Date d'ouverture
- Equipements
- Services

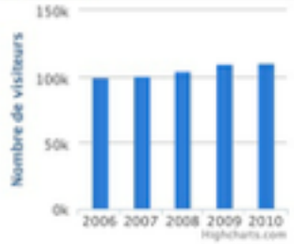
**Liens**

**Réseaux sociaux**

**Localisation**

21, rue de Metz
31000 Toulouse
Midi-Pyrénées
Haute-Garonne

**Fréquentation**




**Tweets** — Suivre

Musée du Louvre @MuseeLouvre 23h
Profitez des #vacances pour vous inscrire aux #cours avec Alain Schnapp #AuditoriumLouvre dès la semaine prochaine bit.ly/1ewvL7v
Étendre

Musée du Louvre @MuseeLouvre 26 Févr
Aujourd'hui à Paris, Le Coup de soleil par Jacob Isaackszoon Van Ruisdael #Louvreweather bit.ly/1mBH2Kv

Tweeter à @MuseeLouvre

Oeuvres (7) | Artistes (5) | Bibliographie (14) | Images de Wikimedia Commons (1230) | Images des réseaux sociaux (18) | À proximité (33)

Affiner votre recherche : Par mots-clés | Par filtres

7 oeuvres correspondent à ce musée

Beauté

Jeune Fille dans un parc

Saint Jean l'Évangéliste et Saint Augustin

Nymphes et satyres dans un paysage

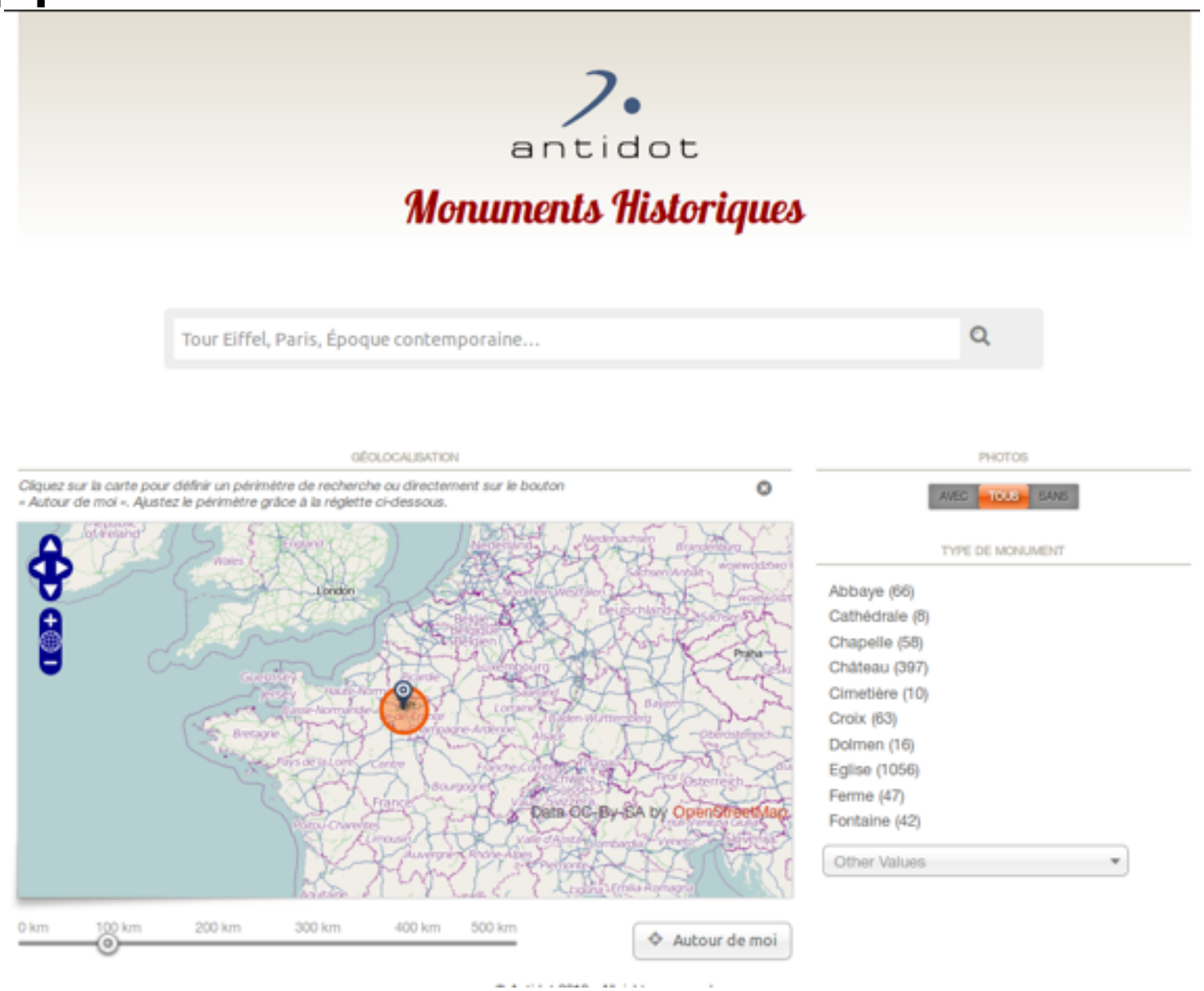
Portrait de la Baronne de Crussol

Gargouilles de l'ancienne église des Cordeliers

Chapiteau de l'histoire de Job de la basilique de la Daurade (Toulouse)

# Création des IHMs

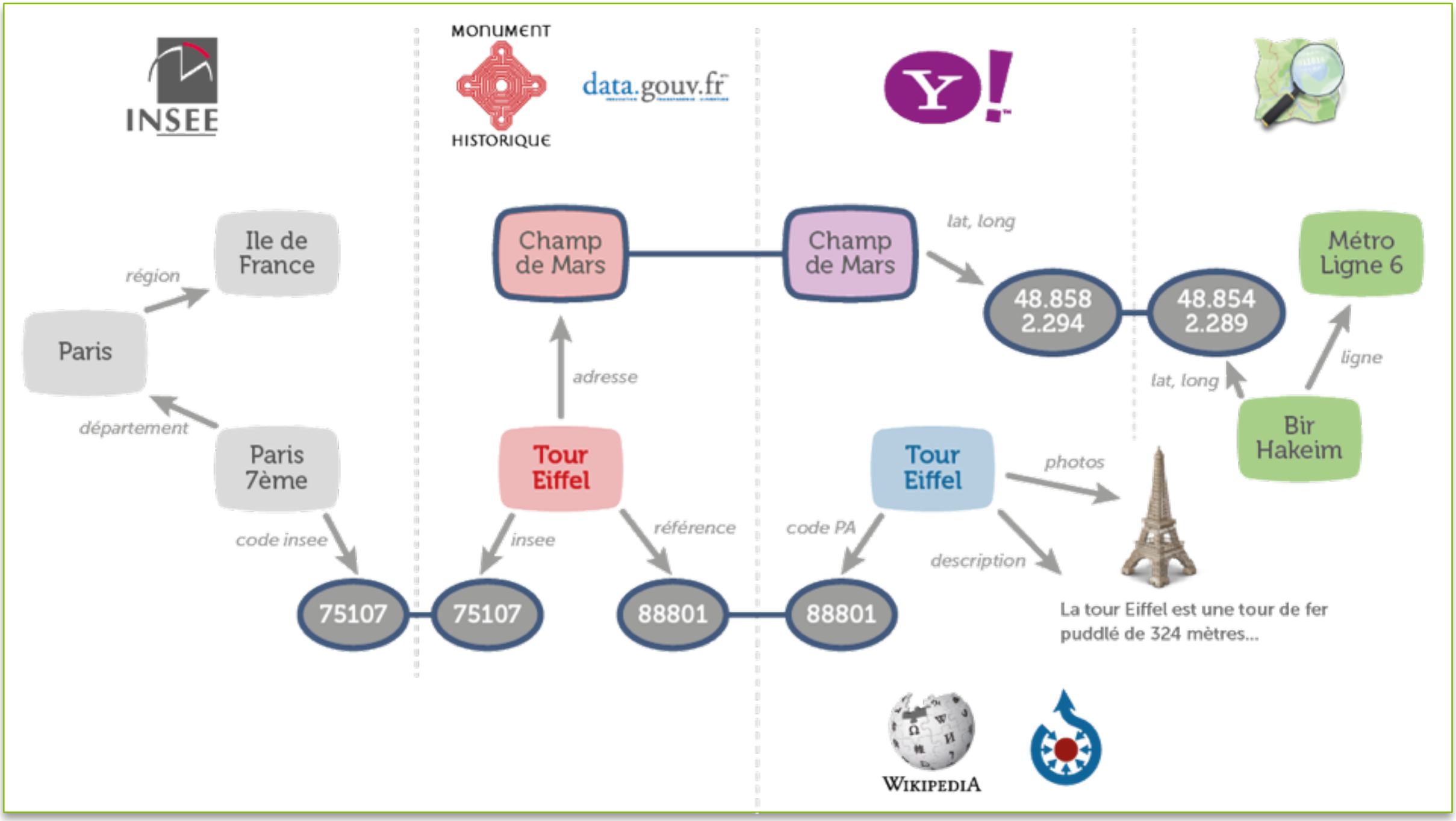
## Le moteur de recherche pour créer simplement des applications riches et faciles d'accès




http://labs.antidot.net/widgets/monuments

slide volé à Gautier Poupeau

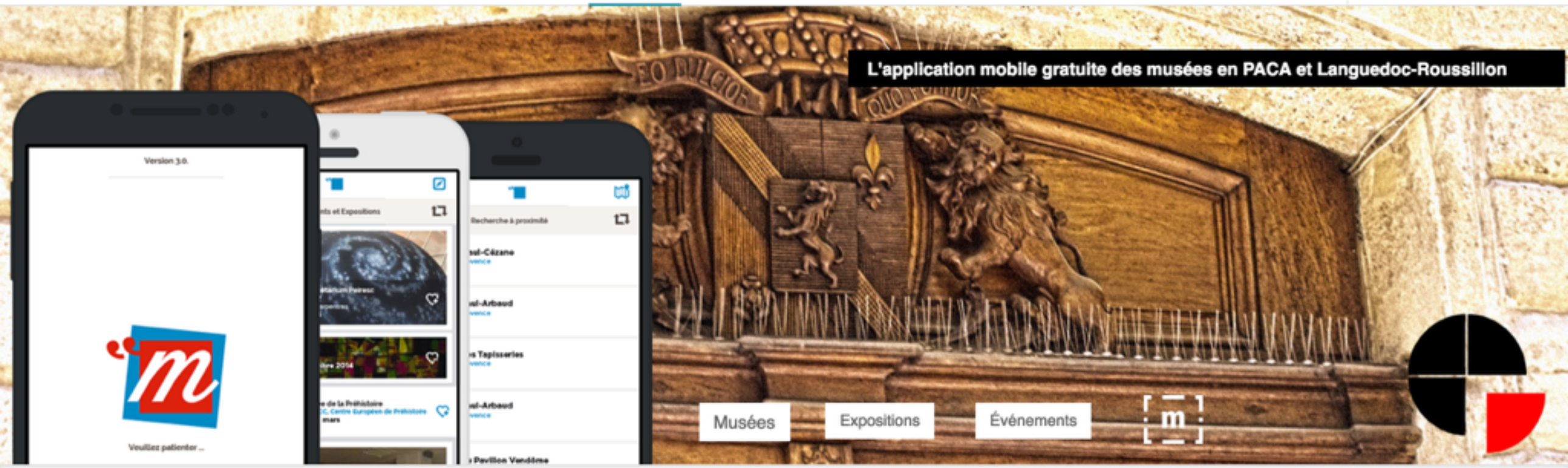
# Modèle de données




slide volé à Gautier Poupeau

Musambule

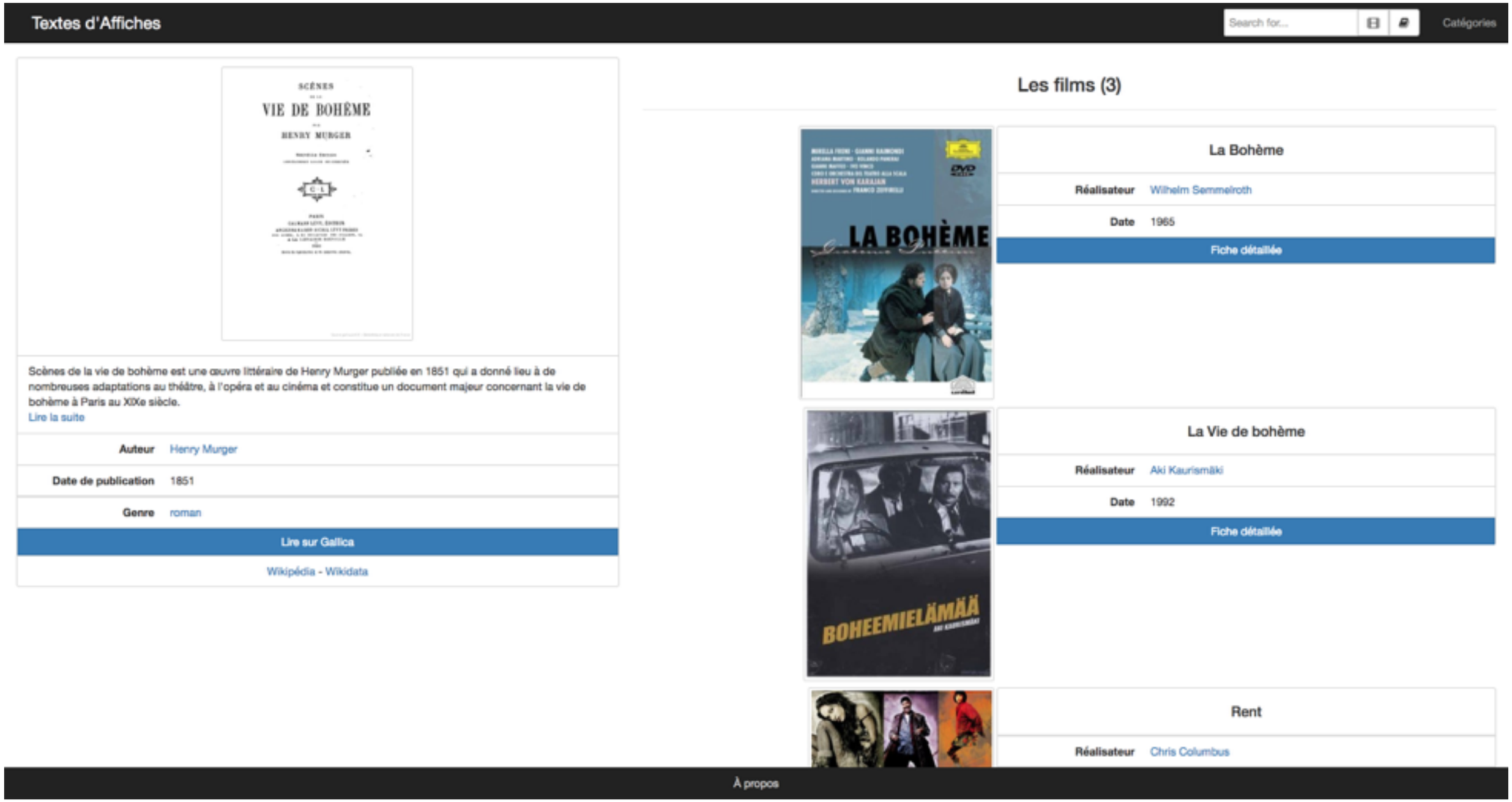
Accueil | Voconces | Actualité | Contact | OpenData / API | Agccpf-Paca




L'application mobile gratuite pour découvrir les **musées** des régions **Provence Alpes Côte-d'Azur** et **Languedoc Roussillon**.

interface qui met en relation directe les films avec les œuvres littéraires adaptées, accessibles sur Gallica. Inversement à partir de livres accessibles sur Gallica on peut en retrouver les adaptations cinématographiques.

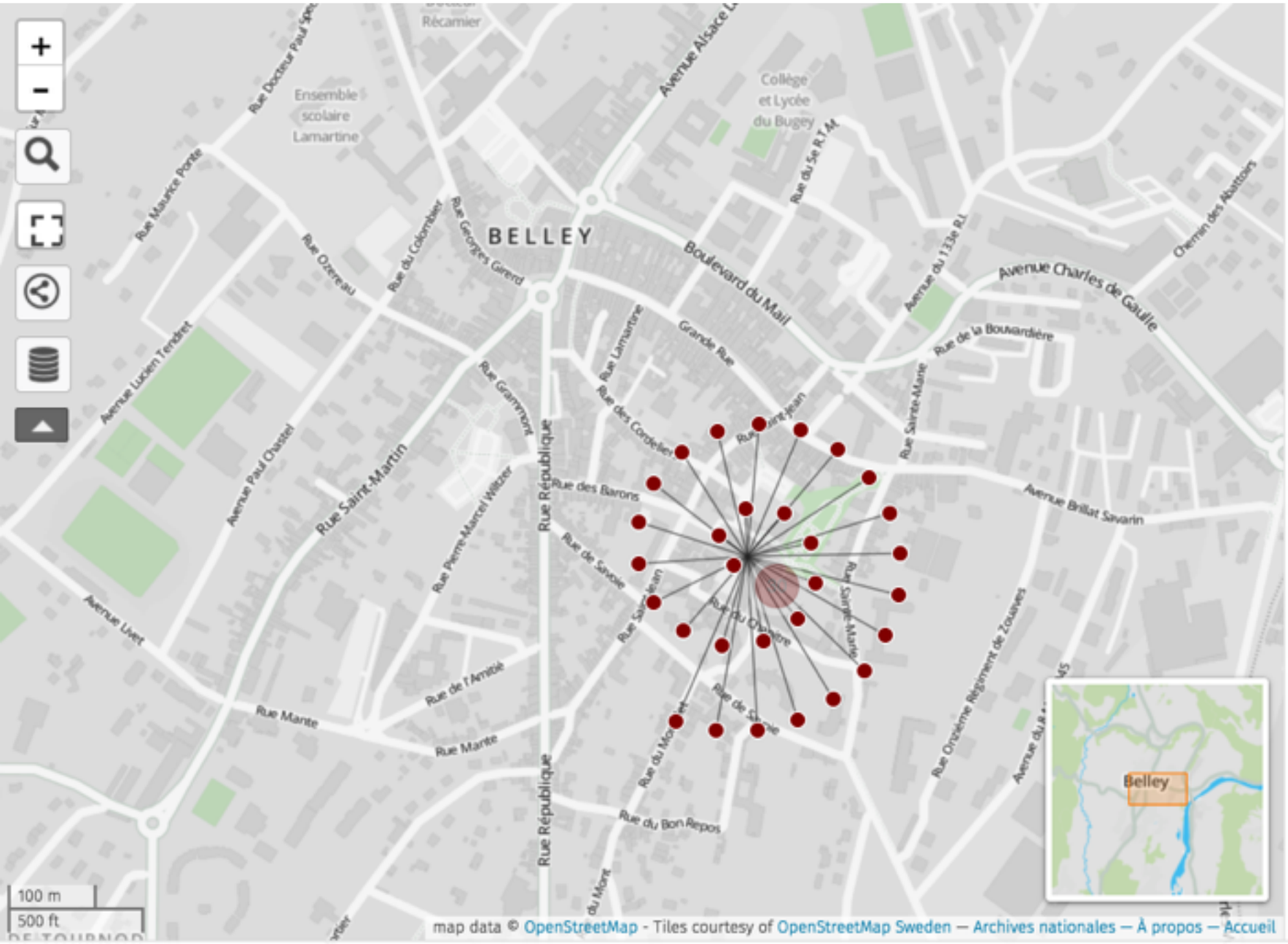
Propulsé par Wikidata, imdb, omdb, data.bnf.fr, Gallica




https://textesdaffiches.histoiredelart.fr/

Documents issus des cotes F/19/7577-F/19/7926. Répertoire numérique détaillé.
Documents figurés des travaux de restauration dans les cathédrales (1802-1912)
Environ 25 mètres linéaires soit 2906 plans, 324 photographies et 1 daguerréotype.

http://www.siv.archives-nationales.culture.gouv.fr/siv/rechercheconsultation/consultation/ir/consultationIR.action?irId=FRAN_IR_054873

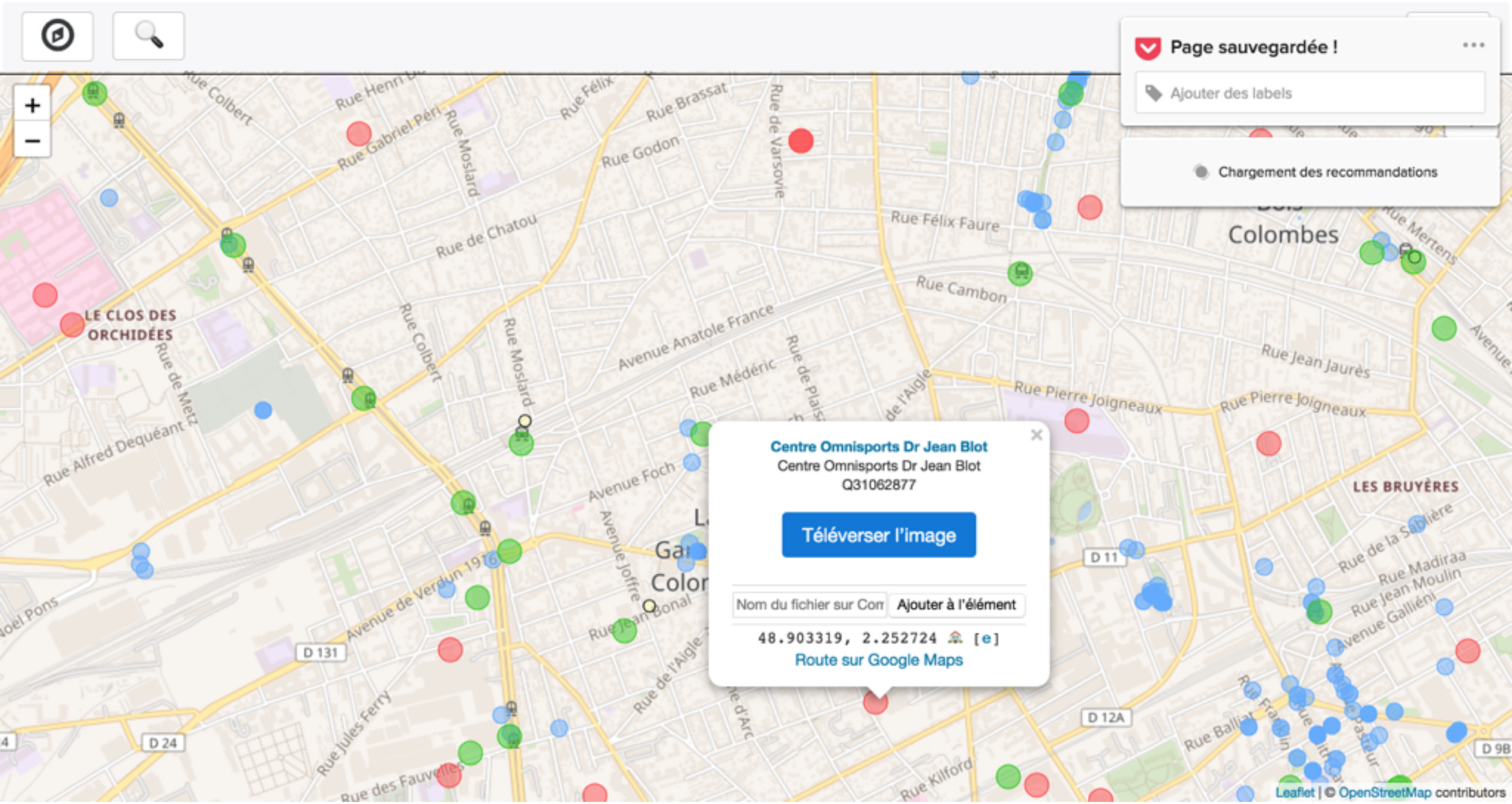



Tout voir | Fermer

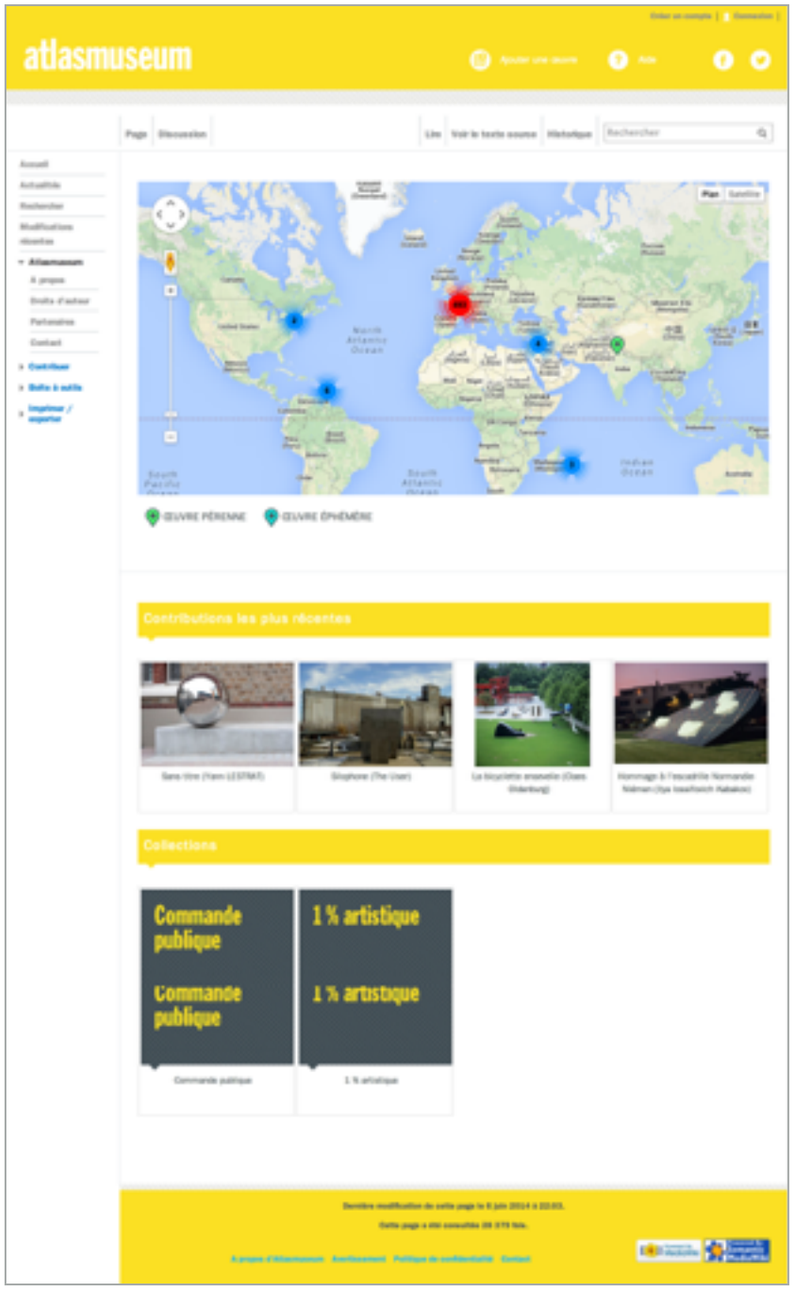
Image issue de Commons (restauration et numérisation des plans à venir)

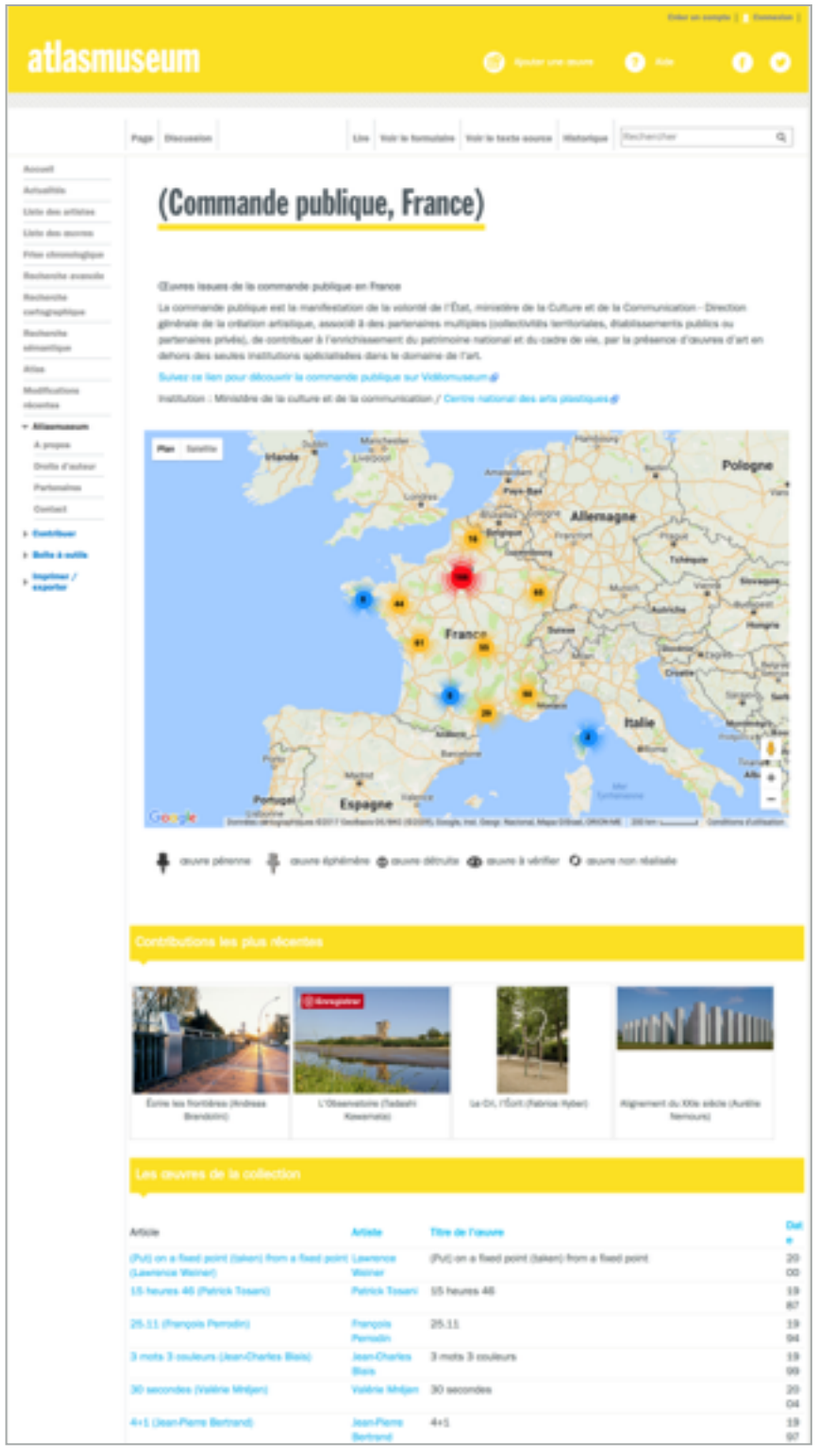
**Belley : cathédrale Saint-Jean**
**Cote** : CP/F/19/7640
**Description** : Plan général de ses abords" : acquisition du terrain Taborin. Par Antoine Marie Chenavard, architecte.
1 plan ; signé ; légende.
**Date** : 15 février 1851
**Echelle** : 7,5 cm / 50 m
**Dimensions** : 39 cm x 29 cm

**Cathédrales (1802-1912) : plans, dessins et photographies - Archives nationales** — À propos | Visualiser les données




https://tools.wmflabs.org/wikishootme/index.html



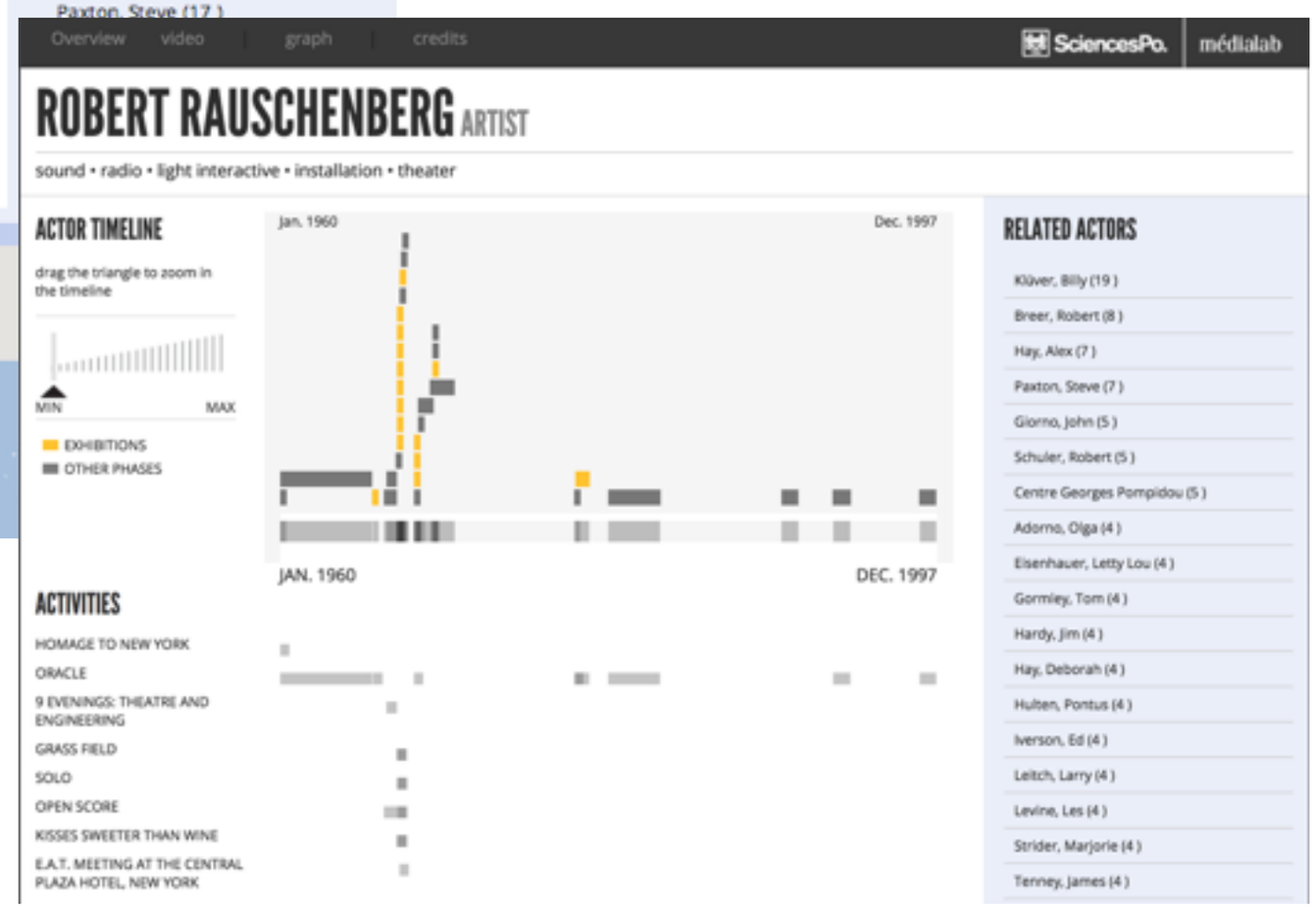



http://publicartmuseum.net/wiki/Accueil

# E.A.T. DATASCAPE 1962-TODAY

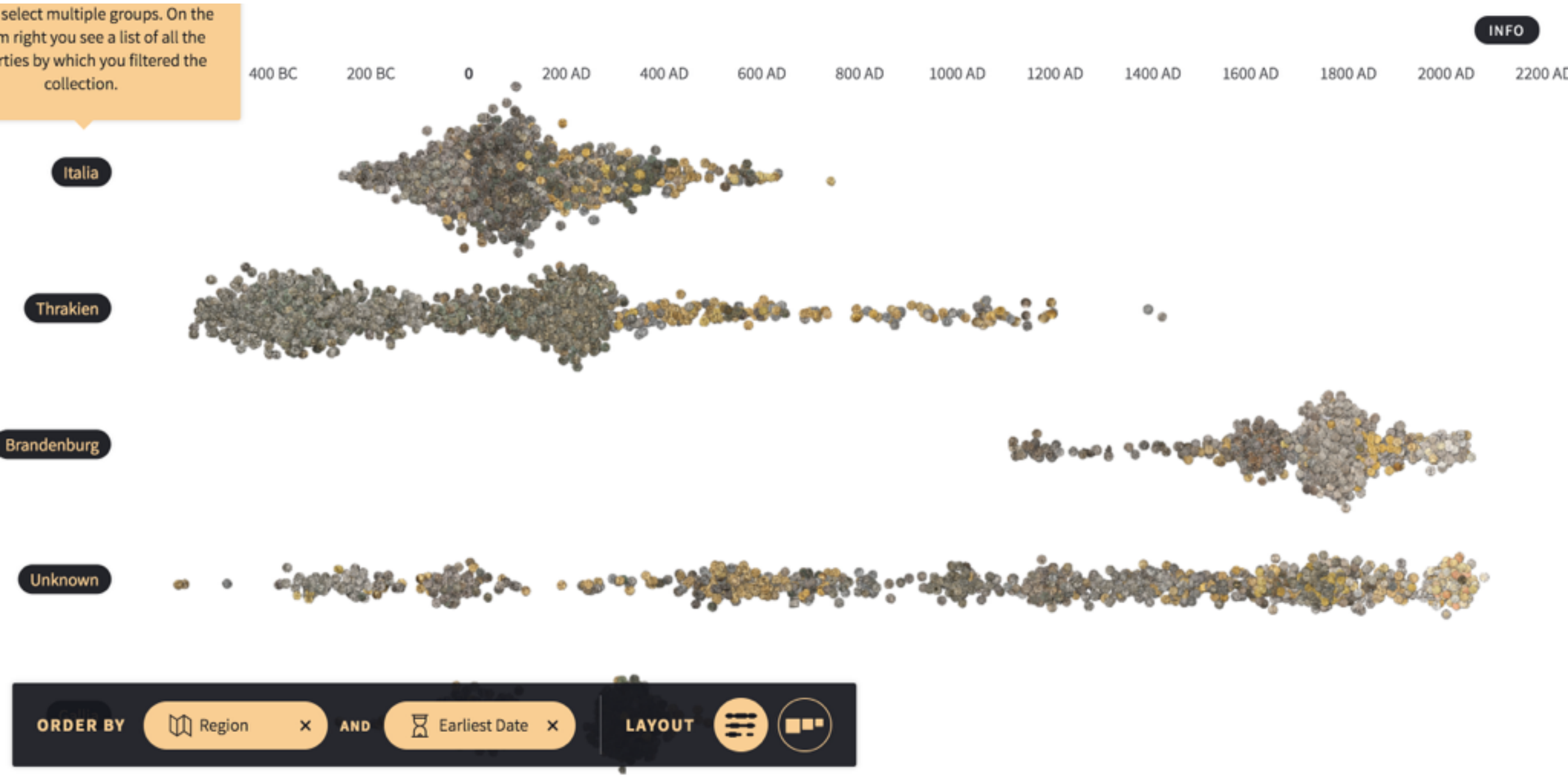
An Exploration of the Experiments in Art and Technology Archives

E.A.T. est une organisation fondée en 1966 dans le but de favoriser la collaboration entre artistes et ingénieurs, en vue de réaliser des projets artistiques ou pas.

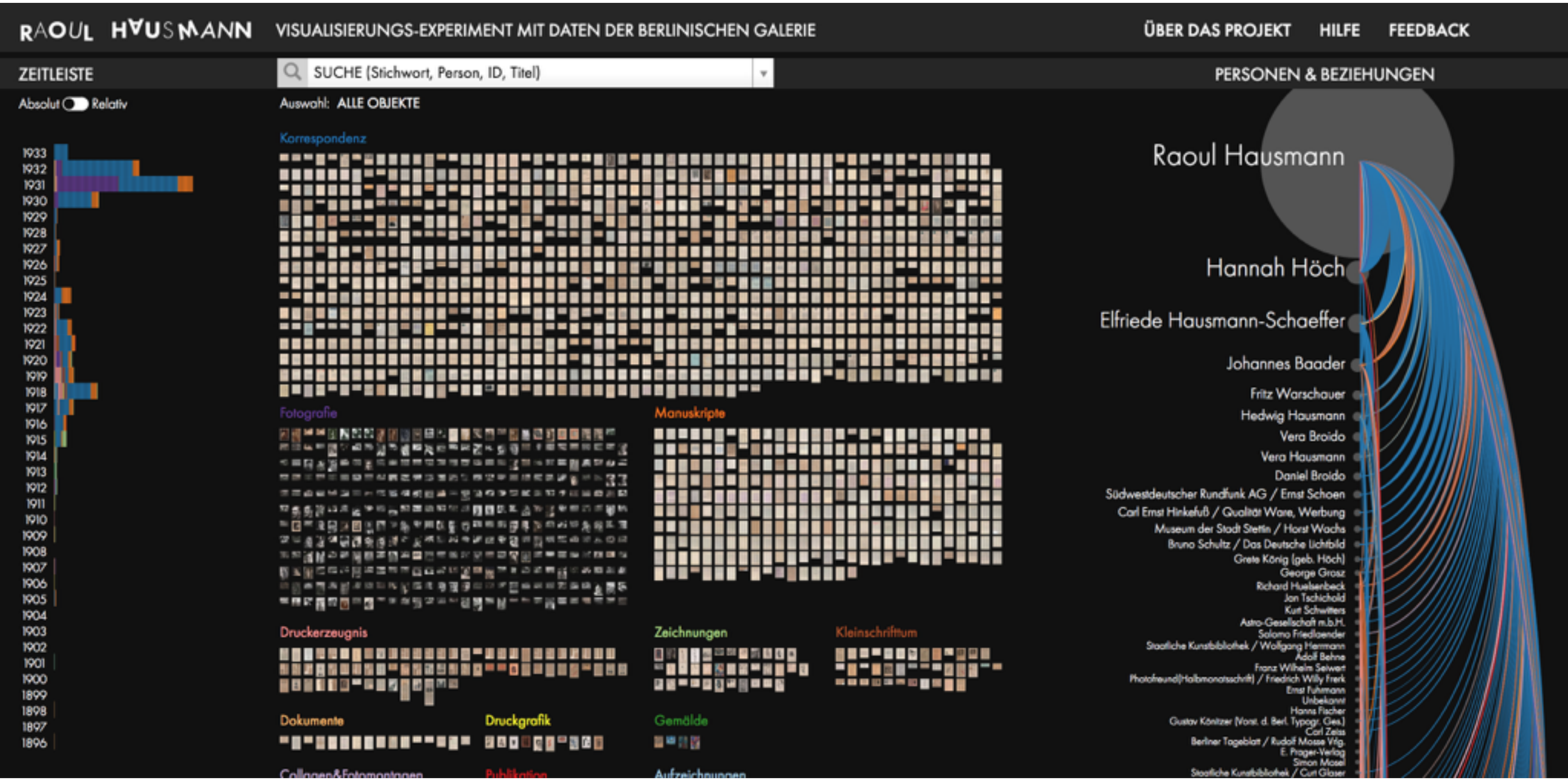



Il s'agit d'une recherche en histoire sociale de l'art qui s'appuie sur un instrument des humanités numériques – le
data-scape (paysage de données), à la fois outil et méthode d'analyse, de visualisation et d'exploration d'archives – pour traiter du cas particulier de l'organisation Experiments in Art and Technology (E.A.T.).

http://jiminy.medialab.sciences-po.fr/eat_datascape




https://uclab.fh-potsdam.de/coins/




https://uclab.fh-potsdam.de/hausmann/

## From *Panel* to *Canvas* and back again

### Traditionally moveable images were painted on panel.

**Panels are made of one or more wooden planks joined together, either horizontally or vertically.**

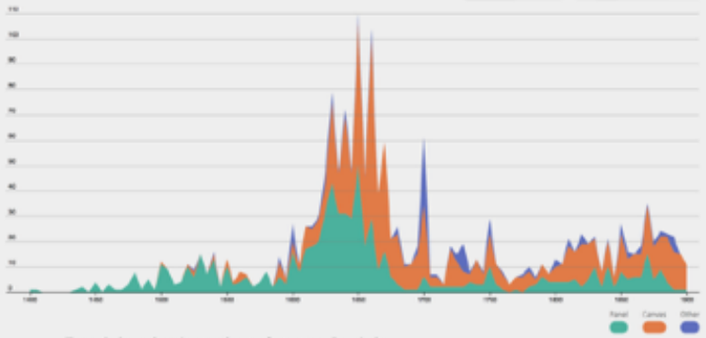
In northern Europe oak was the wood of choice, whereas in the south white poplar was the most popular. The thickness of the planks ranged from 5 to 15 mm depending on the surface of the entire work, which was limited due to the weight of its construction.

Around 1500 canvas was introduced as a support medium in Italy, more precisely in Venice. This city was renowned for its high quality canvas made for sailing. The advantages of canvas are the reduced costs and increased portability. Canvas could be transported rolled up to be mounted on a frame after arriving at its destination.

Painting on panel, however, never went completely out of fashion. In the nineteenth century we even see a revival of the use of wooden panels.

Canvas vs. Panel | **Stacked area chart**

Absolute | Relative | 1 | 5 | 25 | 50 | Year




The stacked area chart shows us the use of support mediums in the Rijksmuseum dataset. We can switch between a relative and absolute view. The more years we group together, for example 50, the clearer the graph becomes. This allows us to see the introduction of canvas around 1500, the increase in usage until about 1700, and the comeback of panel around the 1850s.

## The devil is in the detail

### Over time, panel warps, canvas relaxes, and paint loses its elasticity.

**These processes ultimately lead to a fine pattern of cracks.**

Paint craquelure can not only be used to determine the age of paintings, but also to detect art forgery. Scientific analysis can equally tell us a great deal about the origin of the materials used. Nevertheless, the eye can certainly be deceived.

1 — 2 — 3 — 4

### Do the test

Art historians are trained to look at the details of an image to discover its truths and mysteries. Try it yourself and take the challenge by answering these questions.

1. Which of these paintings do you think is the oldest?
2. Two are made around 1300. Which one is the forgery made 600 years later?
3. **Use your mouse or finger and drag the images in their correct order from old to new.** Need a hint

NEXT

## Migration of artists

### When describing a work of art we usually only mention its current location.

**So for all paintings in the Rijksmuseum digital collection this would be Amsterdam.**

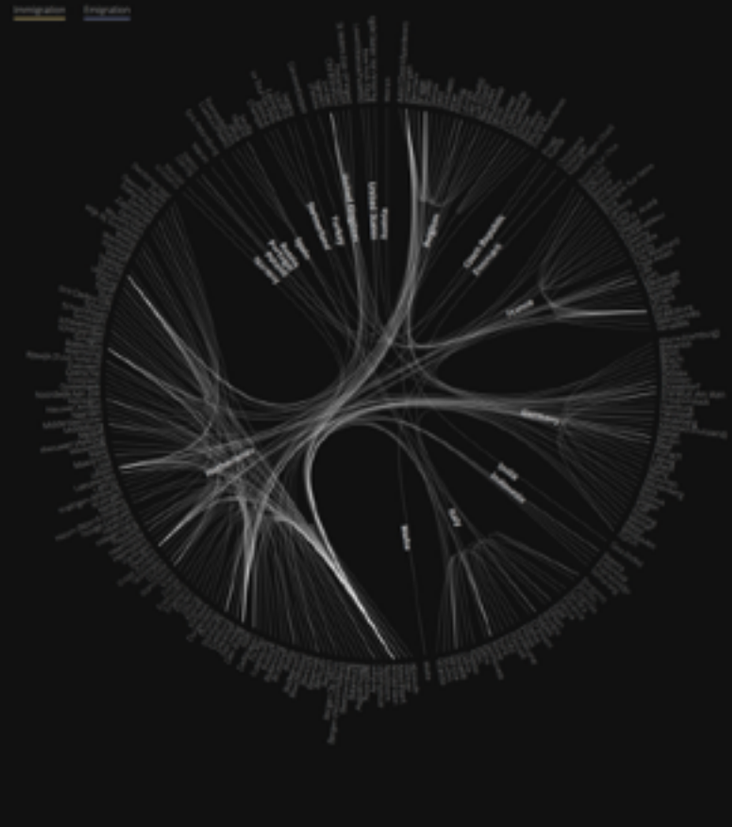
The place of manufacture, however, is far more relevant from an art historical perspective. Although we often do not know where exactly a painting was made, the place of birth and death of the artist can give away valuable clues. At the very least these places give an indication of the geographic area where a painter was active during his lifetime.

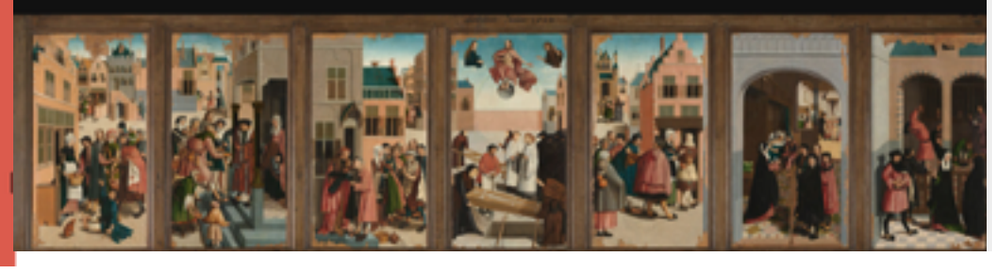
In the visualization below hierarchical edge bundling is used to display all places of birth and death associated with the painters in the Rijksmuseum dataset. Blue lines show patterns of emigration, whereas yellow or red lines show patterns of immigration. If you hover over the name of a country (inner circle) or city (outer circles), all artist migrations related to that specific country or city light up.

The majority of the artists are from the Netherlands, with the city of Amsterdam as the biggest creative hub. What also stands out is that Italy attracted painters from various nationalities, while most Italians moved only within their own country.

Artist Migrations | **Hierarchical Edge Bundling**

Immigration | Emigration





## Deliberately Damaged

### For some of the most famous artworks, we do not know who the maker is.

***The Seven Works of Mercy* was commissioned by the regents of the Holy Spirit Almshouse in Alkmaar.**

It is a polyptych consisting of seven panels held together by its original frame. Despite the fact that the artwork encourages merciful action, it was viciously attacked at the time of the Iconoclastic Fury. Each panel was slashed with sharp objects, especially the faces. The middle panel represents the climax of the violence that was done to the medieval cartoon.

On this panel two scenes have been combined. The upper half depicts the Last Judgment. On Judgment Day, Christ will distinguish between the righteous and the unjust, who will either go to heaven or burn forever in hell. The observance of the Seven Works of Mercy during life can be decisive. The vision of Christ flanked by Mary and John the Baptist remained remarkably unharmed.

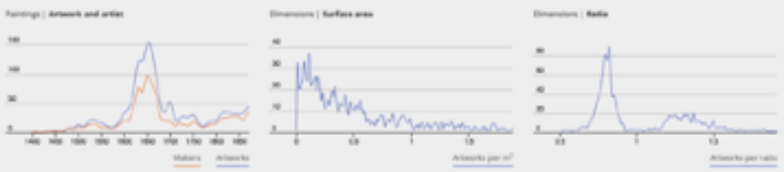
The lower half represents the seventh work of mercy, to bury the dead. Monks carry a coffin to a pit. The grave digger and two priests stand beside the grave, and the ceremony is attended by clergymen dressed in black. From 1971 to 1975 all panels of *The Seven Works of Mercy* underwent large-scale restoration, and were stripped of their later retouches. This revealed the purposeful mutilation of especially the middle panel.

When sliding from the image in color to the recording from the 70s in black and white, the damage becomes clearly visible. All human figures connected to the Roman Catholic Church were scratched or even scraped off completely. The main criticism was their practices of providing indulgences in exchange for money. In the eyes of the perpetrators, part of the Protestant Reformation, this was a foul trade.

The faces where no paint is left now show underdrawings. Nevertheless, the conservation team of the Rijksmuseum decided to only partly reconstruct and retouch the frame and panels of *The Seven Works of Mercy*. As a result the images can be read without much distraction, while some of the history of the artwork remains visible as well.

**Master of Alkmaar, *The Seven Works of Mercy*, 1504**

## Explore by keyword



http://visualizingvisions.com/

ExpeDicoArchi

# Interface d'expérimentation autour du *Dictionnaire des élèves architectes à l'Ecole des Beaux-Arts (1800-1968)*

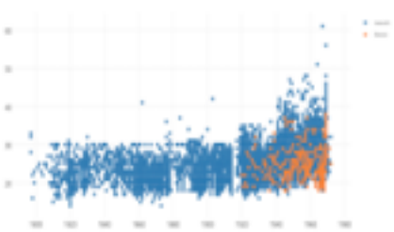
**La base est publiée initialement sur AGORHA, base de donnée réalisée par l'INHA**

4 infovisualisations ont été crées: cartographie, scatterplot, réseaux

Le corpus ci-dessous provient d'un 1er matching entre le *Dictionnaire* et la base *Mérimée* à partir duquel chaque architecte a été réaligné sur Wikidata

Réalisé pour la journée d'étude du 13 avril 2016 par la *Cellule d'ingénierie documentaire*

## Infovisualisation



## Réalignement

Légende

Wikidata (Reasonator) ULAN du GETTY BnF data.bnf.fr Fichier d'autorité international virtuel

| Architectes | Biographie | Edifices | Autres référentiels | Degré fiabilité réalignement |
|---|---|---|---|---|
| Abella, Charles | AGORHA | | | |
| Abraham, Pol | AGORHA | | BnF | |
| Alleman, Jacques | AGORHA | | | |
| Arfvidson, Andre | AGORHA | | | |
| Armand, Alfred | AGORHA | | BnF | |
| Arretche, Louis | AGORHA | | | |
| Aubert, Alberic | AGORHA | | | |

encA | à propos | charte de participation | rejoindre l'encA | mode d'emploi | contact | version pilote | connexion

ACCUEIL | PERSONNALITÉS | COLLECTIFS | INSTITUTIONS | PUBLICATIONS | THÉMATIQUES | ŒUVRES | ÉVÉNEMENTS | CONTRIBUTEURS

# BONNIER Louis

145/1096

BONNIER Louis (H)

architecte, enseignant et urbaniste
de nationalité française

né le 14 juin 1856 à Templeuve (France)
décédé en 1946 à Paris (France)
période d'activité : v.1892-
établi à Paris (France)

formation :
École nationale supérieure des beaux-arts de Paris (1877-1886)

Dictionnaire des élèves architectes de l'École des beaux-arts (1800-1968)

matricule : 3094 (masculin)
autre nom : Bonnier, Louis Bernard

naissance : le 14 juin 1856 (Templeuve)
décès : le 16 sep 1946 (Paris)

présence au concours d'admission : le 16 août 1877
dernière mention de présence à l'École : le 20 déc 1886

Concours

Prix Rougevin 1885
classement : 2ème Prix

Prix Godeboeuf 30 déc 1885
classement : 1ère médaille

Ateliers

- atelier de Moyaux, Constant
- atelier d'André, Jules

© INHA

→ voir sur AGORHA

Données d'AGORHA

PURL

postes officiels :
Architecte-voyer de la Ville de Paris

Directeur des services d'architecture et des promenades et

Louis Bernard Bonnier est né le 14 juin 1856 à Templeuve (Nord) et mort à Paris en 1946.

Il est le père de Jacques Bonnier, également architecte.

Il entre à l'École des beaux-arts de Paris en 1877, où il est élève d'André Moyaux et de Louis-Jules André. Il obtient son diplôme en 1886.

Comme Henri Sauvage, il est l'un des principaux acteurs du passage de l'Art nouveau à l'architecture moderne. Il reste fidèle au rationalisme constructif, base de l'Art nouveau, qui en France s'ancre dans la pensée d'Eugène Emmanuel Viollet-le-Duc pour conduire au style international.
Les quatre villas qu'il construit à Ambleteuse (Nord) en 1892, la mairie de Templeuve (Nord) en 1893, puis la maison Flé à Ambleteuse en 1894, inaugurent ce goût pour la rationalité, l'économie de moyens, l'assujettissement du décor à la structure.

En 1895, Bonnier est l'architecte du fameux Salon de l'Art nouveau, aménagé par le marchand d'objets orientaux Samuel Bing rue de Provence à Paris.

Pour l'Exposition universelle de 1900, il conçoit un globe terrestre imaginé par le géographe anarchiste Élisée Reclus (non réalisé), ainsi que le très impressionnant pavillon Schneider (Forges du Creusot) quai Branly.

À côté de son œuvre privé, Bonnier développe une importante activité administrative. Dès 1884, non encore diplômé, il devient architecte-voyer de la Ville de Paris. Il sera encore directeur des services d'architecture et des promenades et plantations vers 1910, inspecteur général des services techniques d'architecture et d'esthétique de la Seine, architecte en chef des bâtiments civils et palais nationaux (chargé du palais de l'Élysée de 1901 à 1911), etc.

Rapporteur de la commission chargée de réviser les règlements de voirie de Paris, il prépare le nouveau décret de 1902 sur les hauteurs et saillies qui, autorisant l'ajout de corniches et de combles aux formes libres, permet l'éclosion de l'Art nouveau, et qui modifie le gabarit des constructions.

À la même époque, Louis Bonnier participe également à la mise en place du cadre institutionnel et réglementaire du logement social et de la planification urbaine (que l'on n'appelle pas encore l' « urbanisme »).
En 1912, il supervise le premier concours d'habitations à bon marché (HBM) de la Ville de Paris et réalise le premier plan d'extension de Paris avec Marcel Poëte. En 1916, il propose la création du Casier archéologique et artistique de la préfecture de la Seine.

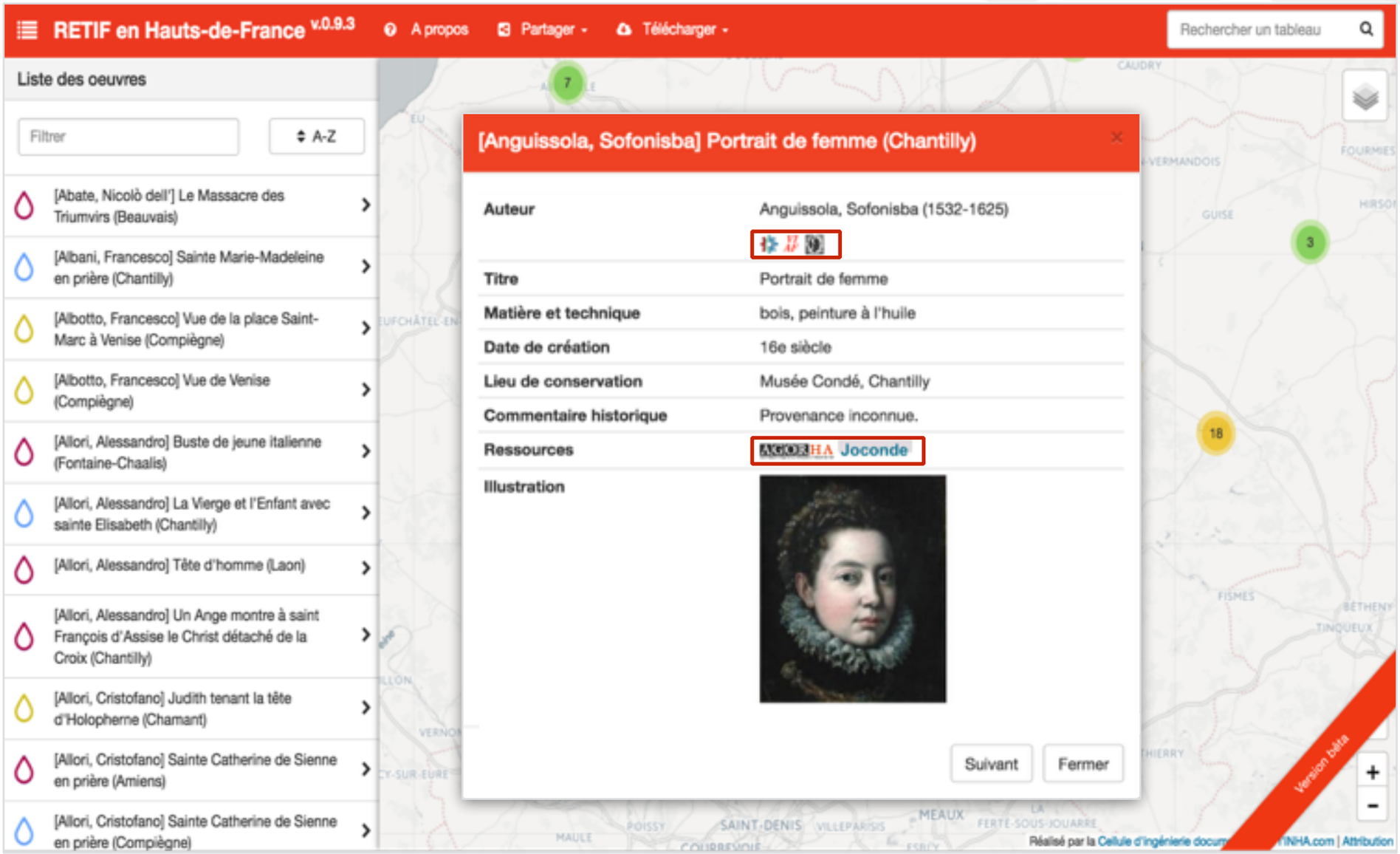
En 1917, il fonde l'École supérieure d'art public. Dans la foulée de la loi Cornudet sur les plans d'aménagement, d'embellissement et d'extension, cette école devient, en 1919, l'École des hautes études

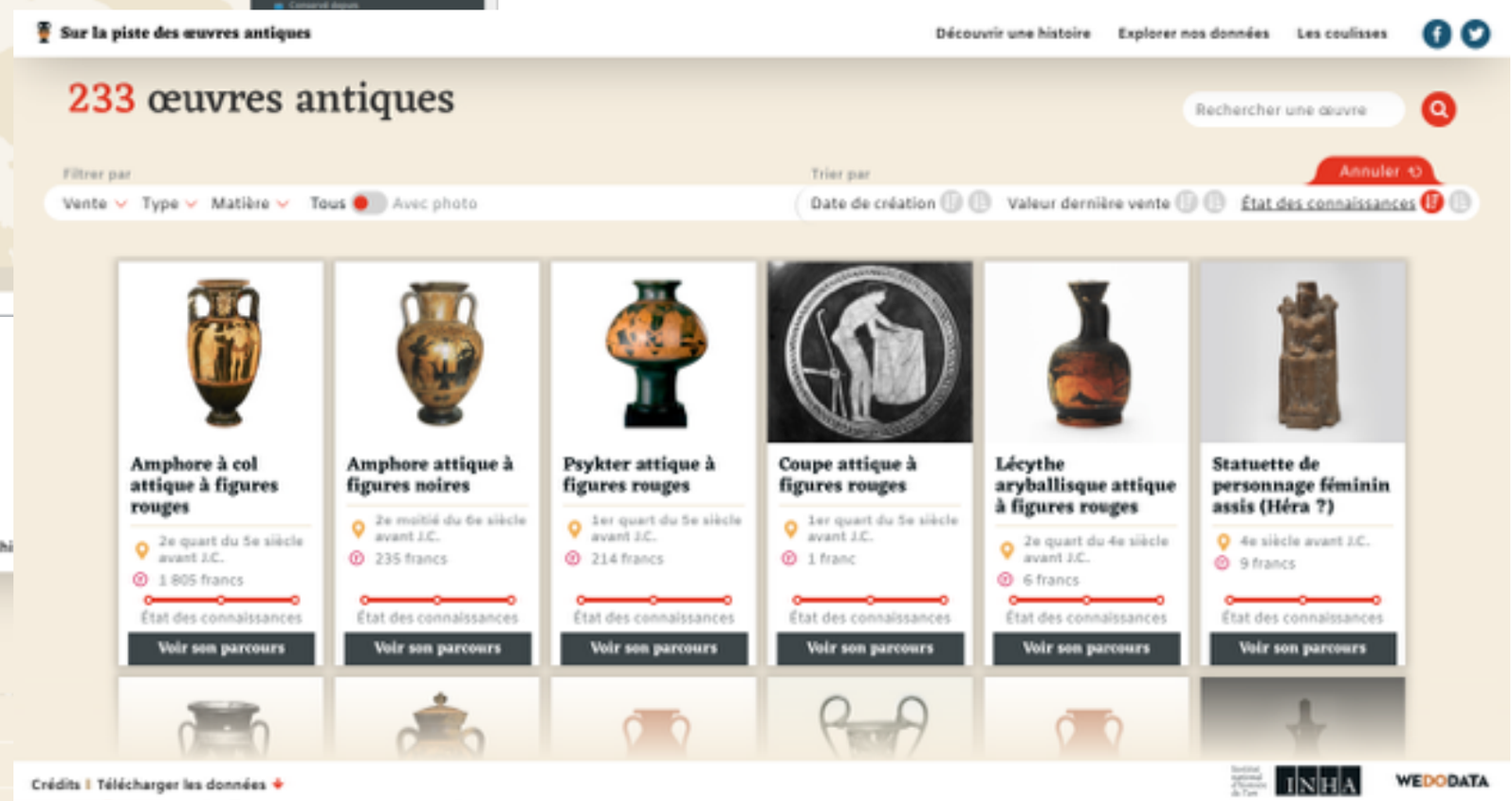
Archiwebture
Cité de l'architecture et du patrimoine
Archives d'architecture du XXe siècle
publié le 15 déc 2015
licence CC BY NC ND 4.0

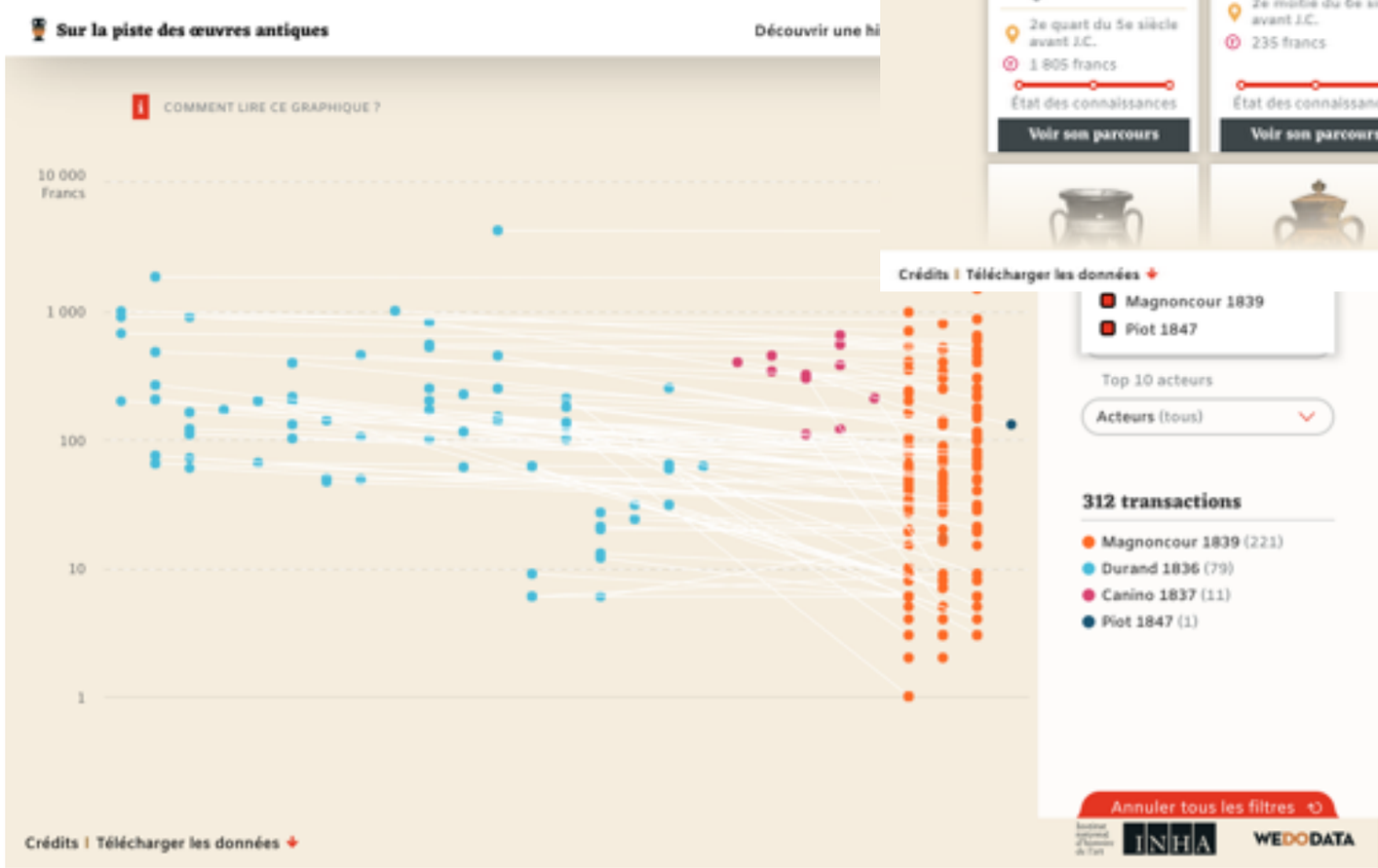
encA, encyclopédie numérique collaborative, Cité de l'architecture et du patrimoine

# RETIF en Haut-de-France

Retrouver les peintures italiennes en Haut-de-France dans le cadre de l'exposition Les "Heures Italiennes"






ventesdantiques.inha.fr

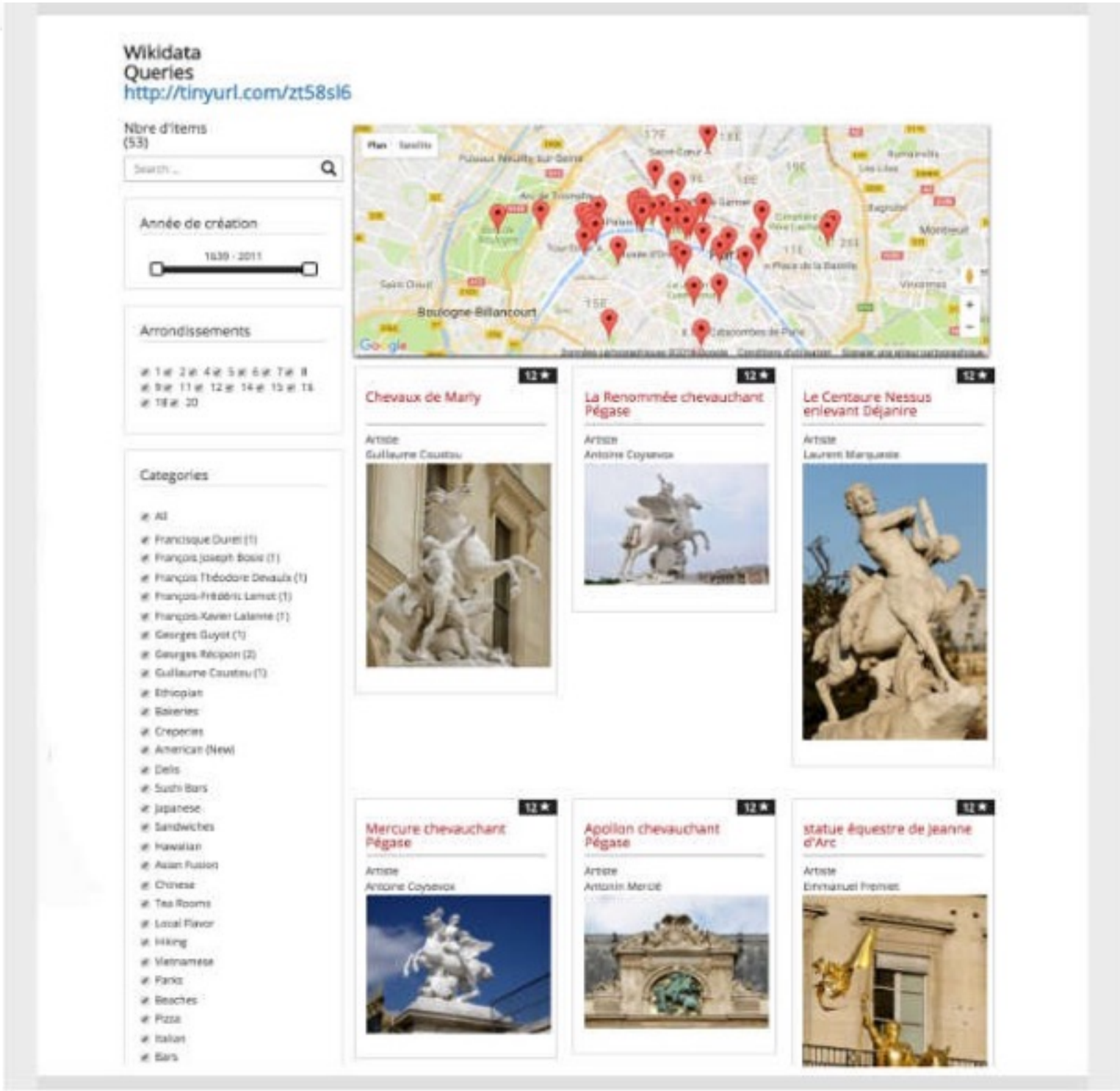
# L’importance du dynamisme/ interactivité

## L’apport de la publication web

>

Wikidata Query | Examples | Prefixes | Tools | Help

```
SELECT DISTINCT
  ?item
  (SAMPLE (?titleL) AS ?title)
  (group_concat(distinct ?creatorL ; separator = ", ") as ?creator)
  (group_concat(distinct ?placeL ; separator = ", ") as ?place)
  (group_concat(distinct ?arr ; separator = ", ") as ?arrondissement)
  (SAMPLE (?img) AS ?image)
  (SAMPLE (?coord) AS ?coordinates)
(SAMPLE (?date) AS ?d)
{
  SELECT ?item ?date ?titleL ?creatorL ?placeL ?arr ?img ?coord
  WHERE {
    # selecting items...
    {
      # items in Paris...
      ?item wdt:P131 wd:Q90 .
    } UNION {
      # ...or in an arrondissement of Paris
      ?item wdt:P131/wdt:P131* wd:Q90 .
    }
    # genre: public art

    # depictions of horses, centaurs, donkeys...
    ?item wdt:P136 wd:Q557141 .
    { ?item wdt:P136 wd:Q659396 . } UNION
    { ?item wdt:P180/wdt:P279* wd:Q726. } UNION
    { ?item wdt:P180/wdt:P279* wd:Q83203 } UNION
    { ?item wdt:P180/wdt:P279* wd:Q27022 } UNION
    { ?item wdt:P180/wdt:P31 wd:Q83203 } UNION
    { ?item wdt:P180/wdt:P279* wd:Q2962925 }

    # creators
    OPTIONAL {
      ?item wdt:P170 ?c.
      ?c rdfs:label ?creatorL filter (lang(?creatorL) = "fr") .
    }
    OPTIONAL
    {
      ?item wdt:P571 ?date.
    }

    # place
    OPTIONAL {
      ?item wdt:P276 ?p .
      ?p rdfs:label ?placeL filter (lang(?placeL) = "fr") .
    }

    # arrondissement
    OPTIONAL {
      ?item wdt:P131 ?a .
      ?a wdt:P131 wd:Q90 .
      ?a rdfs:label ?arrL filter (lang(?arrL) = "fr").
      BIND(REPLACE(?arrL, '^([0-9]+).*$', "$1", "ai") AS ?arr)
    }

    # image
    OPTIONAL {?item wdt:P18  ?img.}
    # coordinates
    OPTIONAL {?item wdt:P625 ?coord.}
    # title
    OPTIONAL {?item rdfs:label ?titleL filter (lang(?titleL) = "fr") .}
  } ORDER BY xsd:integer(?arr) ?titleL ?genreL ?createurL ?lieuL
}
GROUP BY ?item
ORDER BY xsd:integer(REPLACE(?arrondissement,'^([0-9]+),.*$', "$1", "ai")) ?place ?title
```




http://tinyurl.com/zt58sl6

# Chapitre 1

# *La der des ders...*

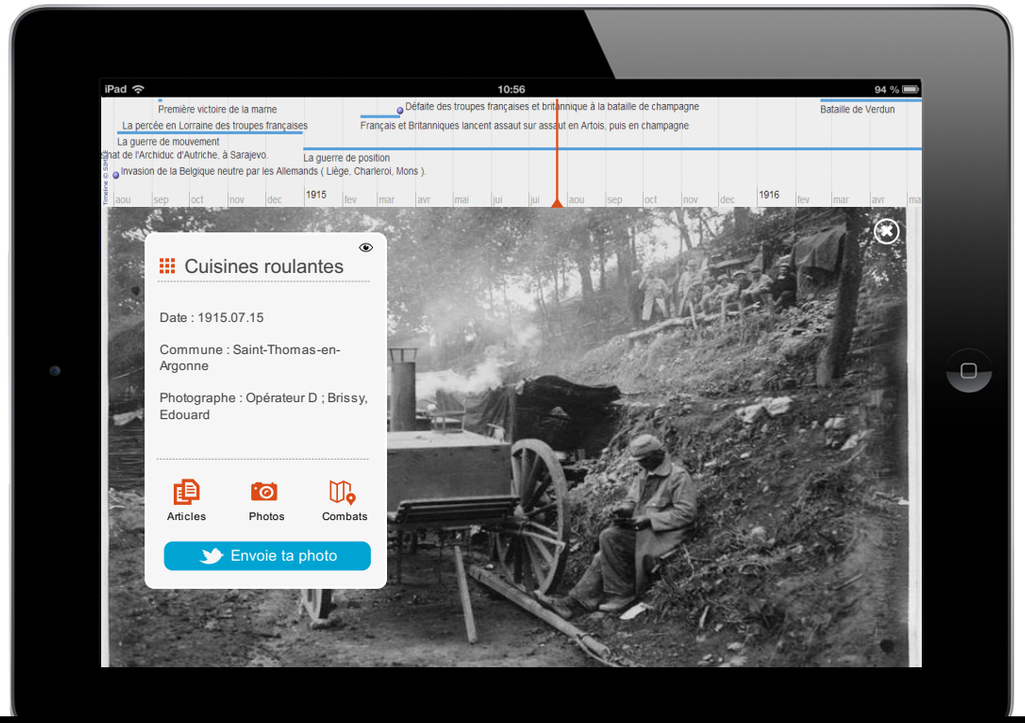
Carnets de guerre, circuit de mémoire

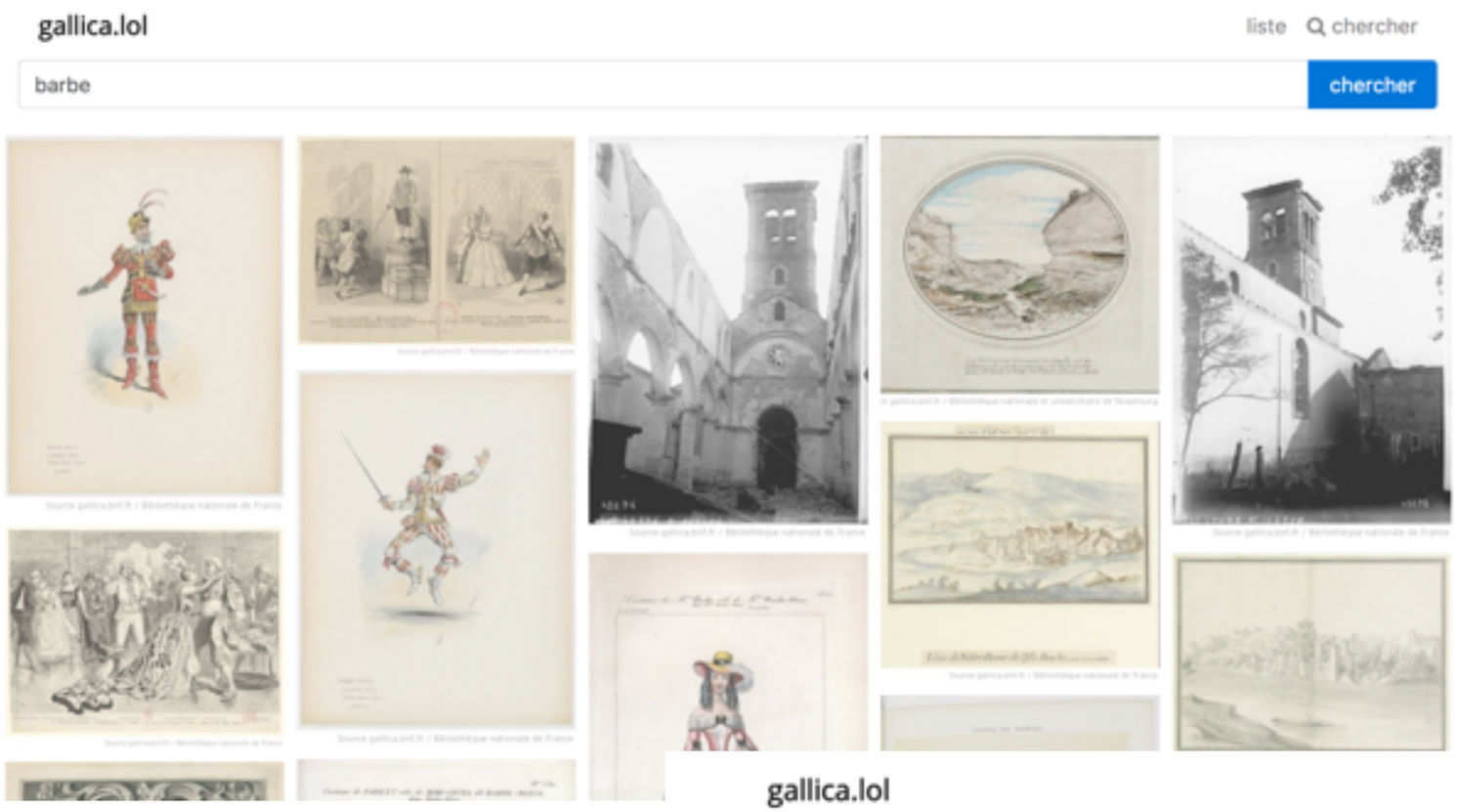
**Contenus principaux**

| Base Mémoire | carto14-18 |
|---|---|
| Notices + Images | Fonds de cartes |
| CSV | png/html |
| http://www.data.gouv.fr/DataSet/576006?xtmc=m%C3%A9moire&xtcr=1 | http://www.carto1418.fr/ |

**Enrichissement de données**

| Gallica | Europeana | Isidore | DBpédia |
|---|---|---|---|
| Ouvrages en rapport | Ressources icono | Articles OpenAccess | historique lieux, évènements |
| OAI-PMH | API | SPARQL | SPARQL |
| http://oai.bnf.fr/oai2/OAIHandler?verb=ListRecords&metadataPrefix=oai_dc&set=gallica:corpus:1418 | http://europeana.eu/portal/api/registration.html | http://www.rechercheisidore.fr/sparql/ | http://dbpedia.org/sparql |

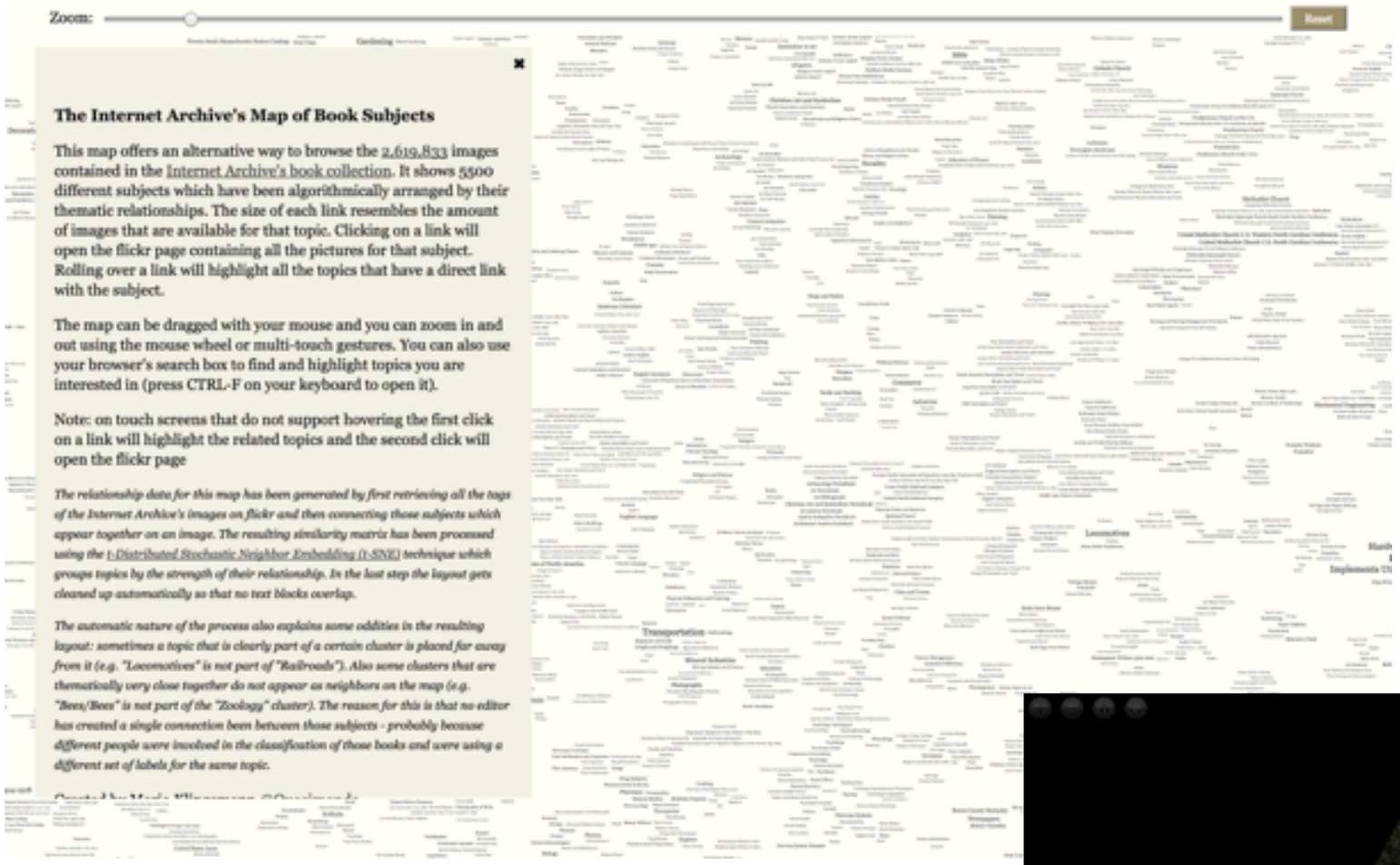




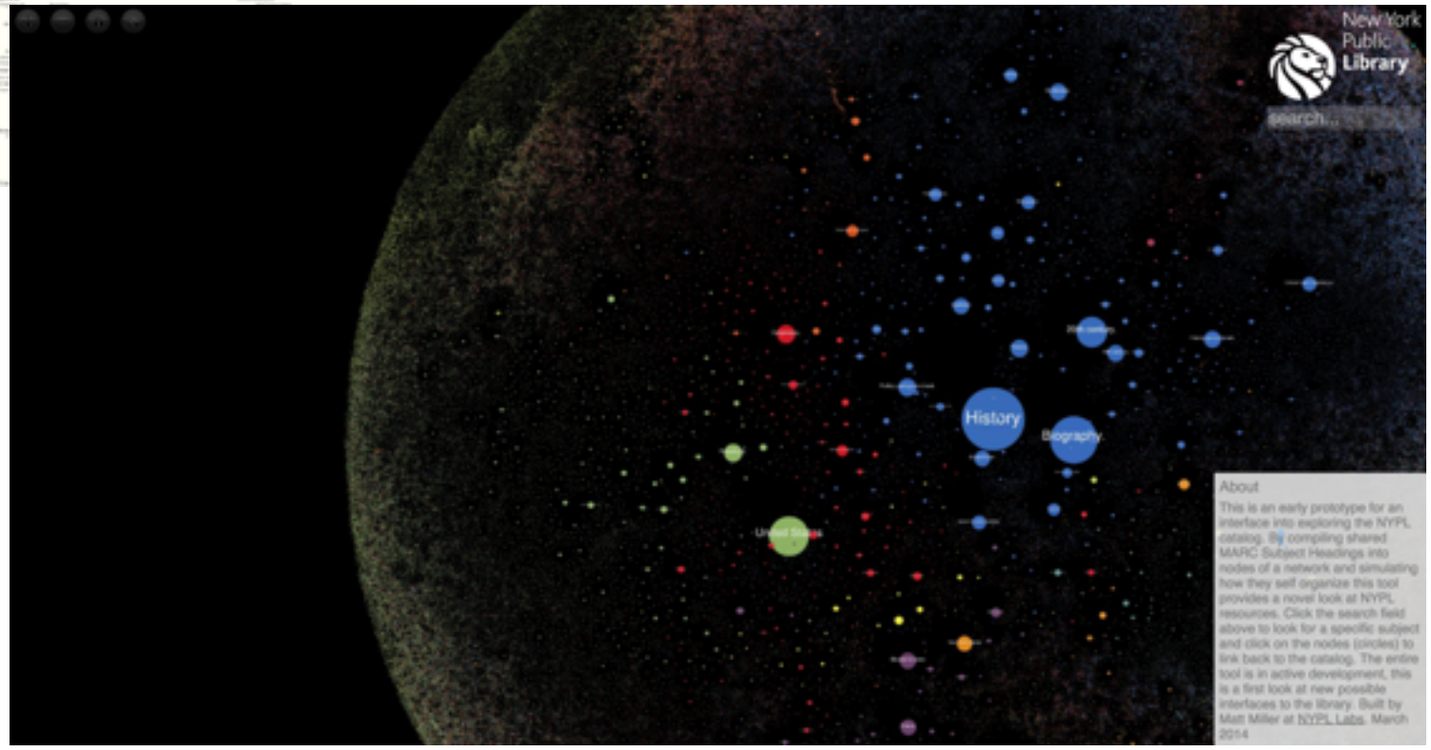
https://gallica.lol/

# Et dans les bibliothèques ?
# [et dans les GLAMs en général]

# Explorer les métadonnées des catalogues




http://incubator.quasimondo.com/internetarchive/InternetArchiveBookSubjectsMap.html?utm_content=buffer0d234&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer




http://catalog-network.s3-website-us-east-1.amazonaws.com/

# Explorer les métadonnées des catalogues

DPLA Visual Search Prototype -- @schuyberg, 2014 About

portrait Go!

124001 results found.

114602 images

Subjects

- Portraits -- 42480
- Photographs -- 16513
- Portrait -- 14233
- Portrait photographs -- 8025
- Men -- 6892
- People -- 6754
- Group portraits -- 6525
- Native Americans -- 6484
- Legislators -- 5775
- Portrait photography -- 5594
- Children--Portraits -- 5591
- Clothing & Apparel -- 5582
- Dress accessories -- 4924
- Women -- 4901
- Dress Accessory -- 4675
- Women--Portraits -- 4365
- Individuals -- 4197
- Politics -- 4027
- Government -- 4001
- Politics and Government -- 3980
- Portrait m -- 3729
- Studio -- 3535
- Double portraits -- 3348
- Minnesota Legislator -- 3318
- Minnesota Legislature -- 3316
- State governments--Officials and employees -- 3314
- Men--Portraits -- 3193

Date range

5,000
0
1860 1880 1900 1920 1940 1960 1980

[Portrait.]
["Ribera, Jusepe (1590-1652)"]
The New York Public Library

Portrait
1821-12-08
A
The New York Public Library

Portrait
1850-1850
The New York Public Library

http://schuyberg.com/dplaviz/

# Explorer les métadonnées des catalogues

## Epoques et sujets

L'aperçu chronologique visualise les plus courantes des mots clés, lieux, personnes et organisations par période.

## Mots clés

Les mots-clés les plus courants sont représentés selon leur taille. Fréquent est visualisée en cliquant sur.

## Lieux et divisions

Vue d'ensemble des villes avec le plus d'entrées. Les anneaux sont divisés selon les catégories représentées.

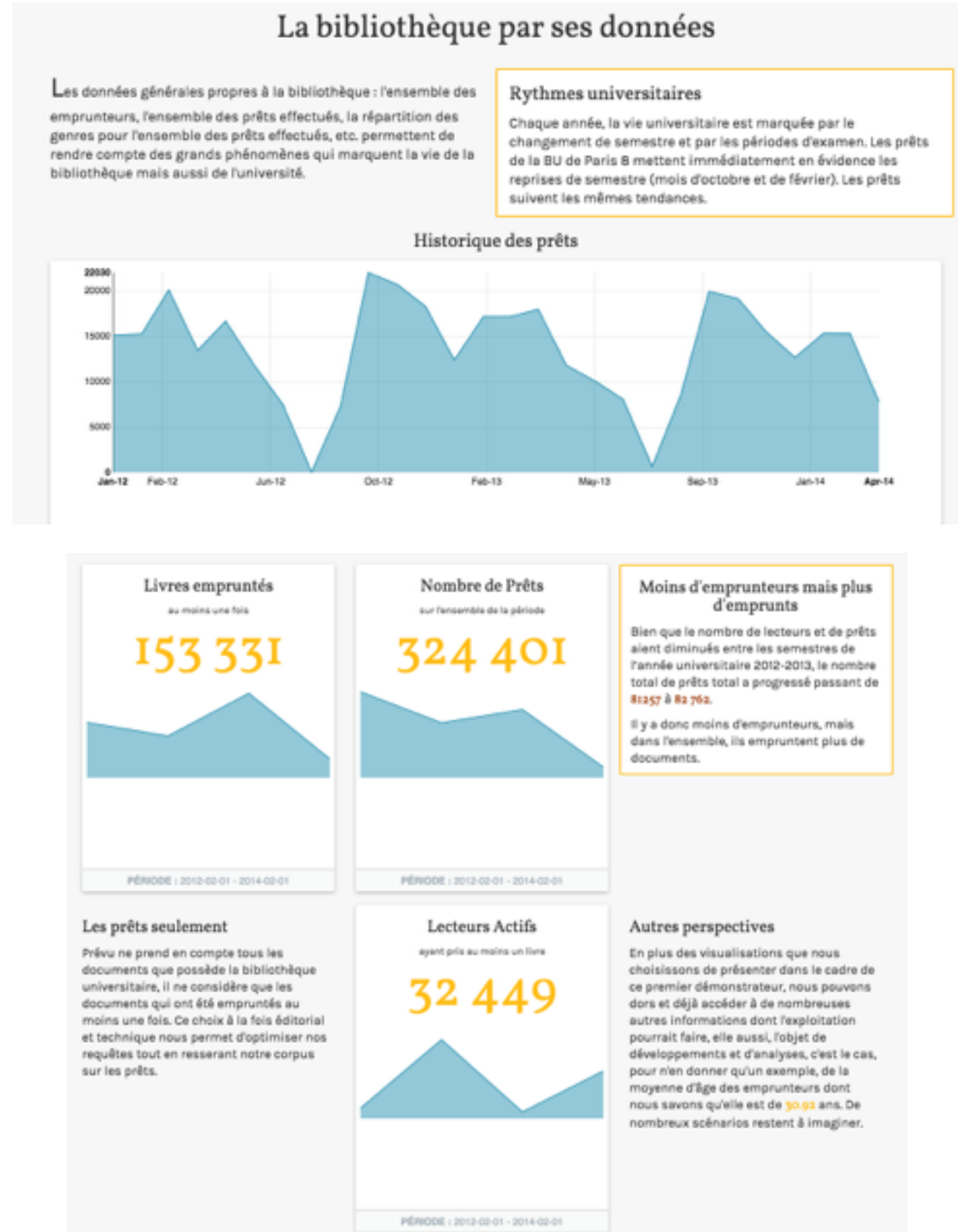
## People & organisations

Réseaux de relations entre les personnes et les organisations dans le stock le plus actif de périodes au choix.

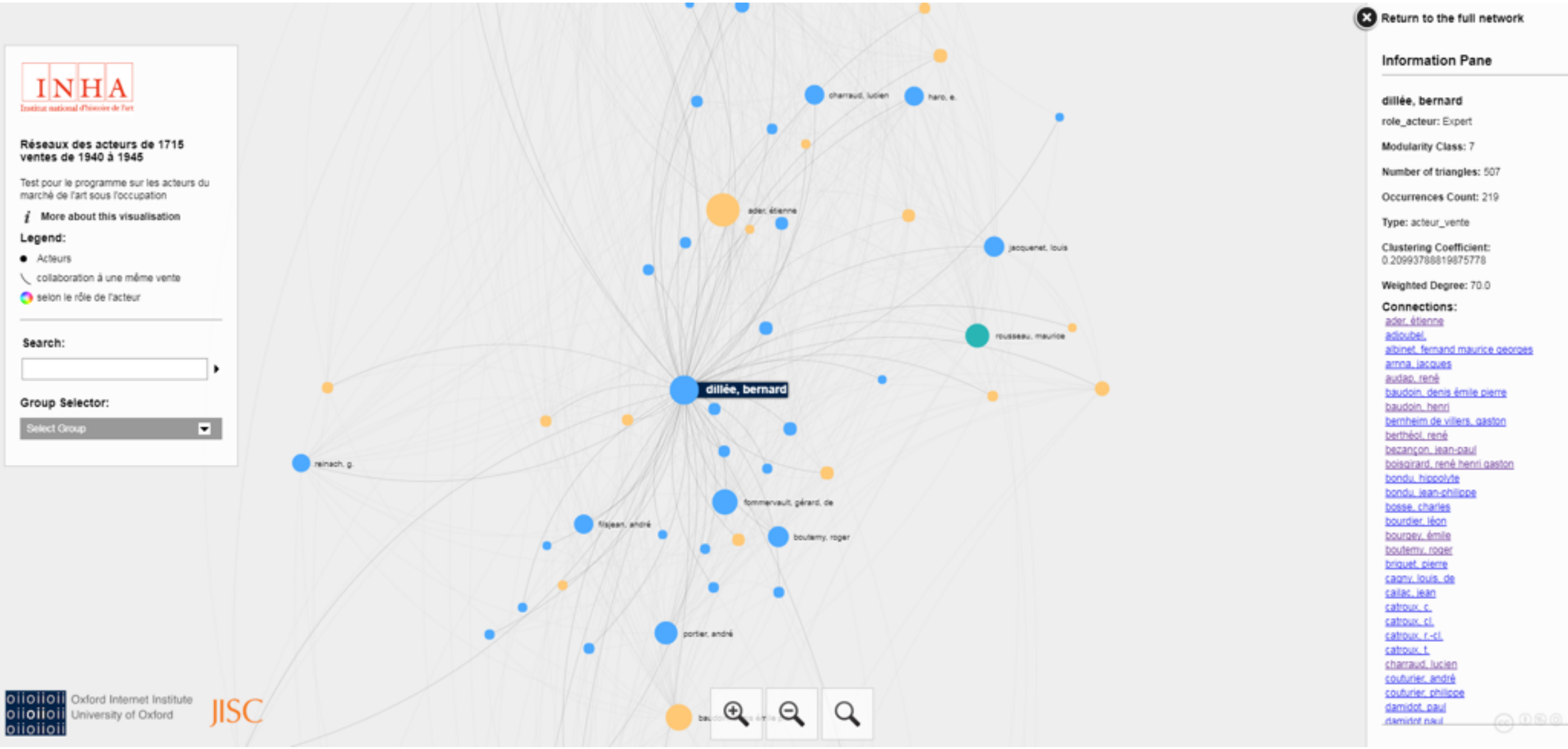
http://infovis.fh-potsdam.de/ddb/

# Evaluer les emprunts des bibliothèques

http://prevu.fr/



# Visualiser les données issue d'une numérisation de fonds à la bibliothèque de l'INHA



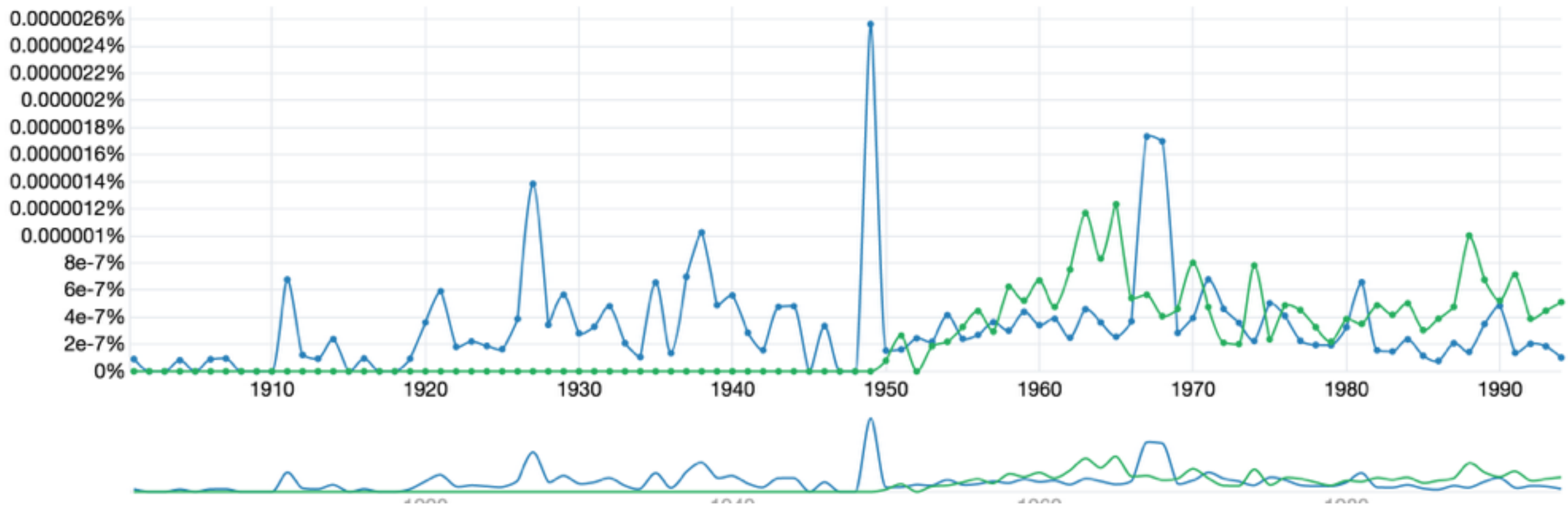

https://skylab.inha.fr/Acteurs_Marche_FrancoAllemand/network/

# Historische Kranten
## *ngramviewer*

Typ één of meer frases

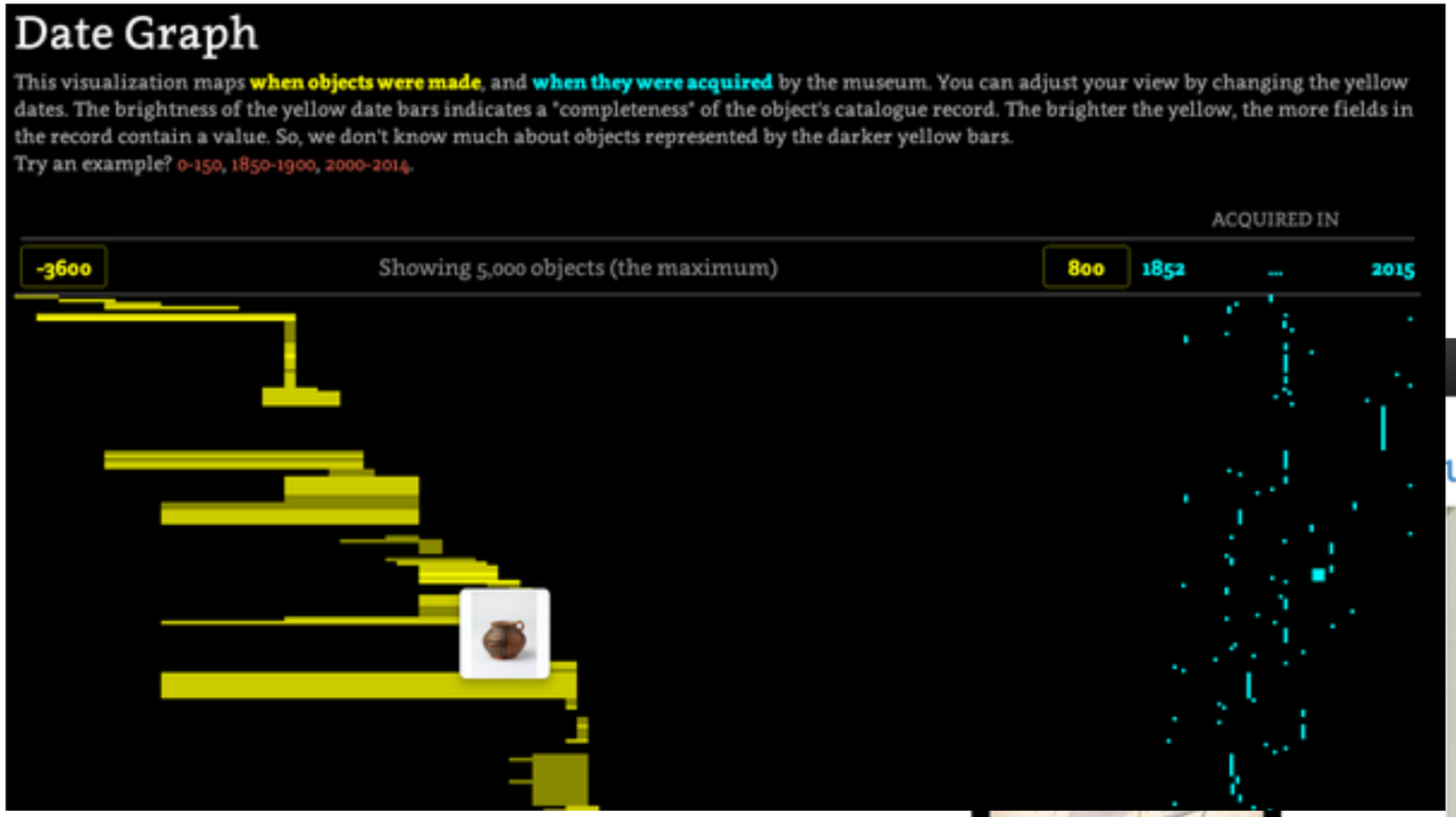
× kees van dongen × karel appel

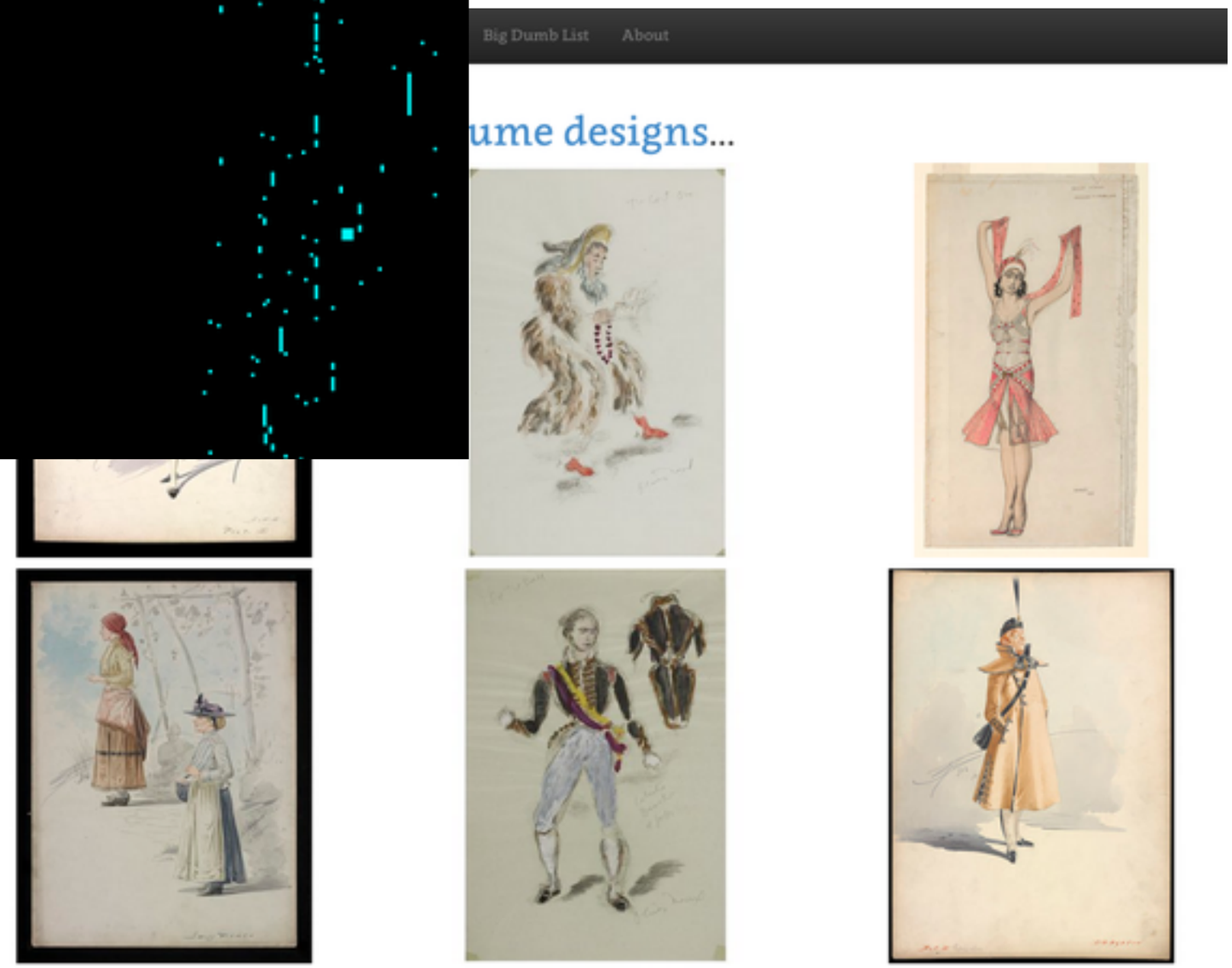
Share!




http://lab.kb.nl/

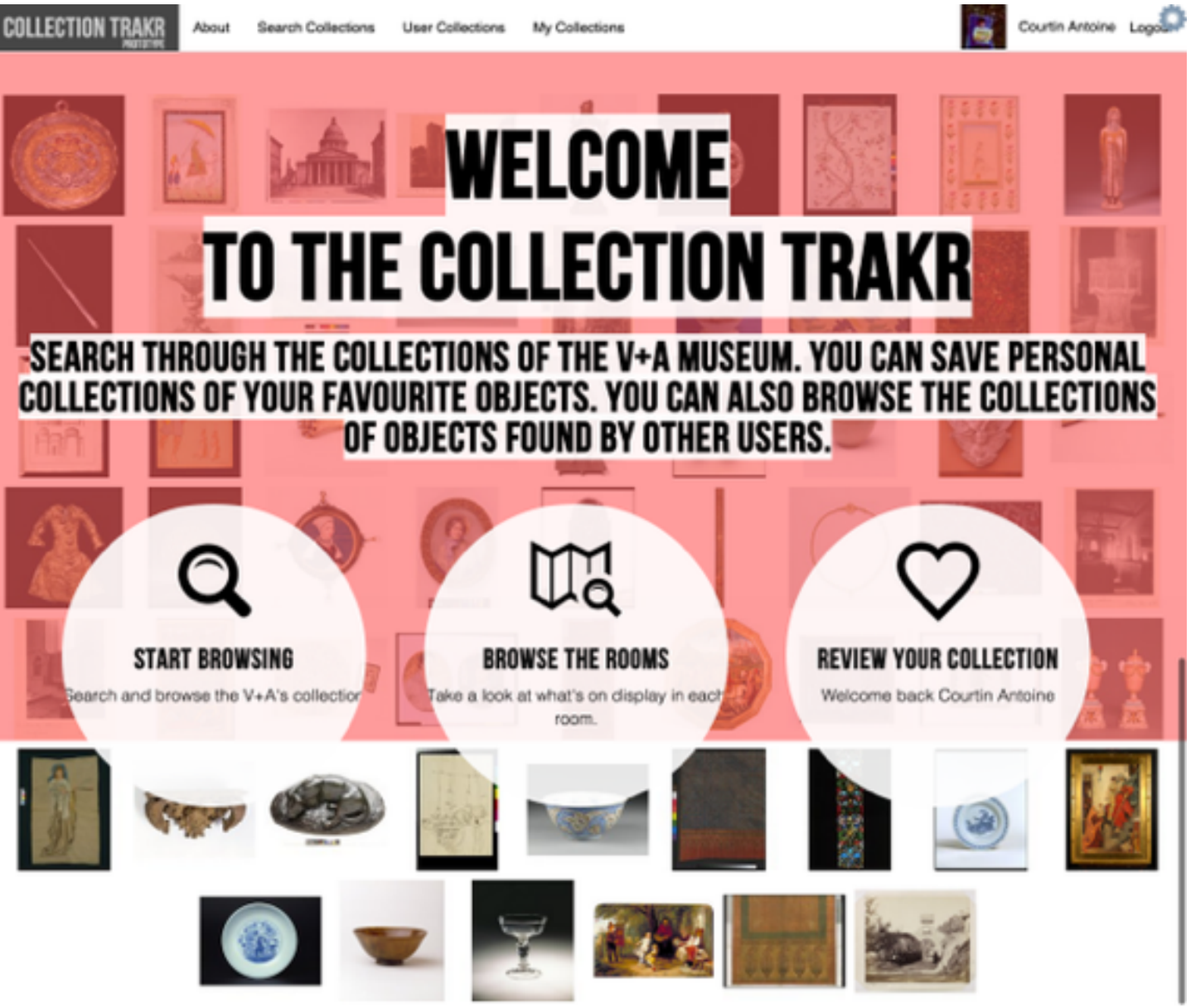
# V&V Spelunkeer - Cas d’utilisation de l’API du V&V

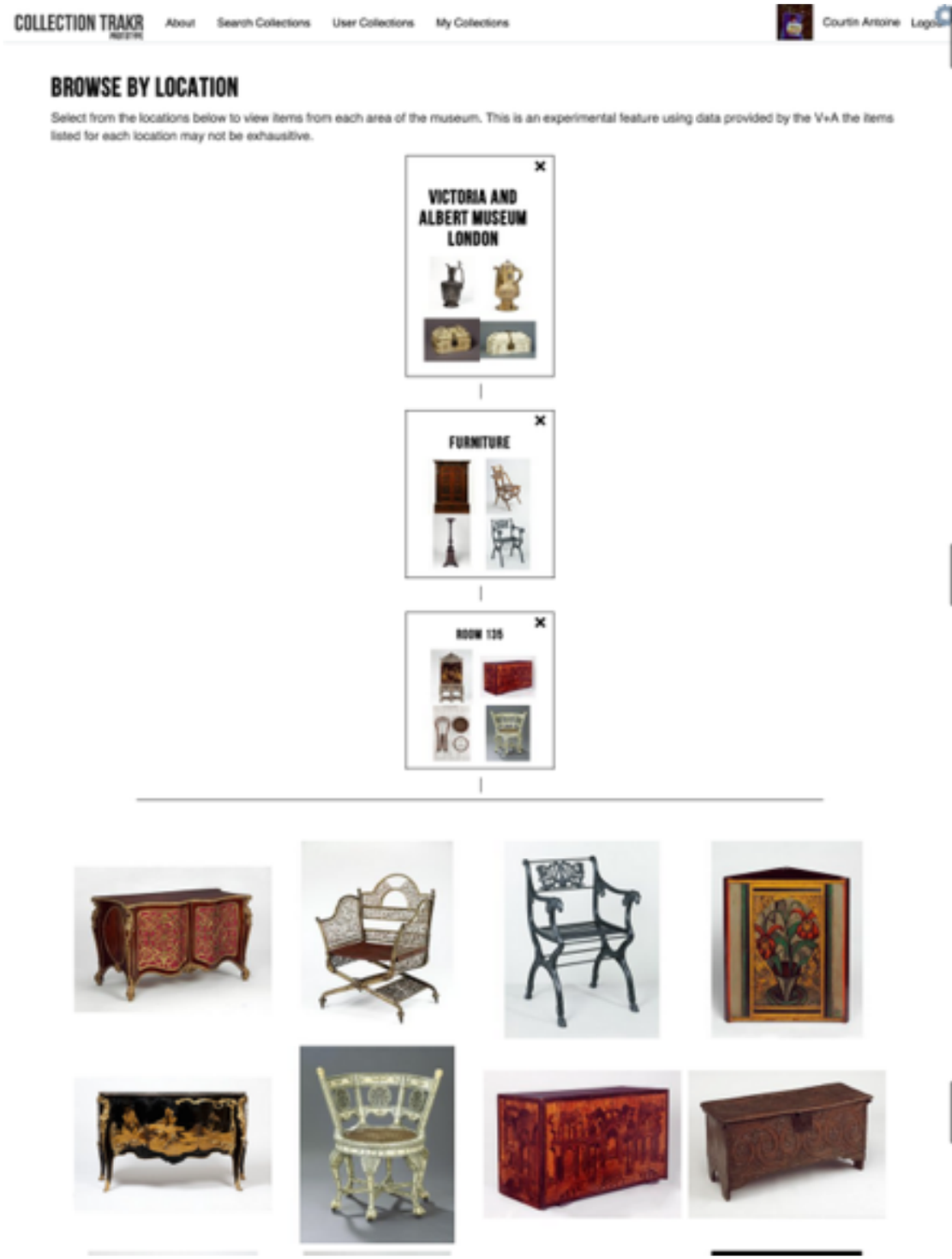


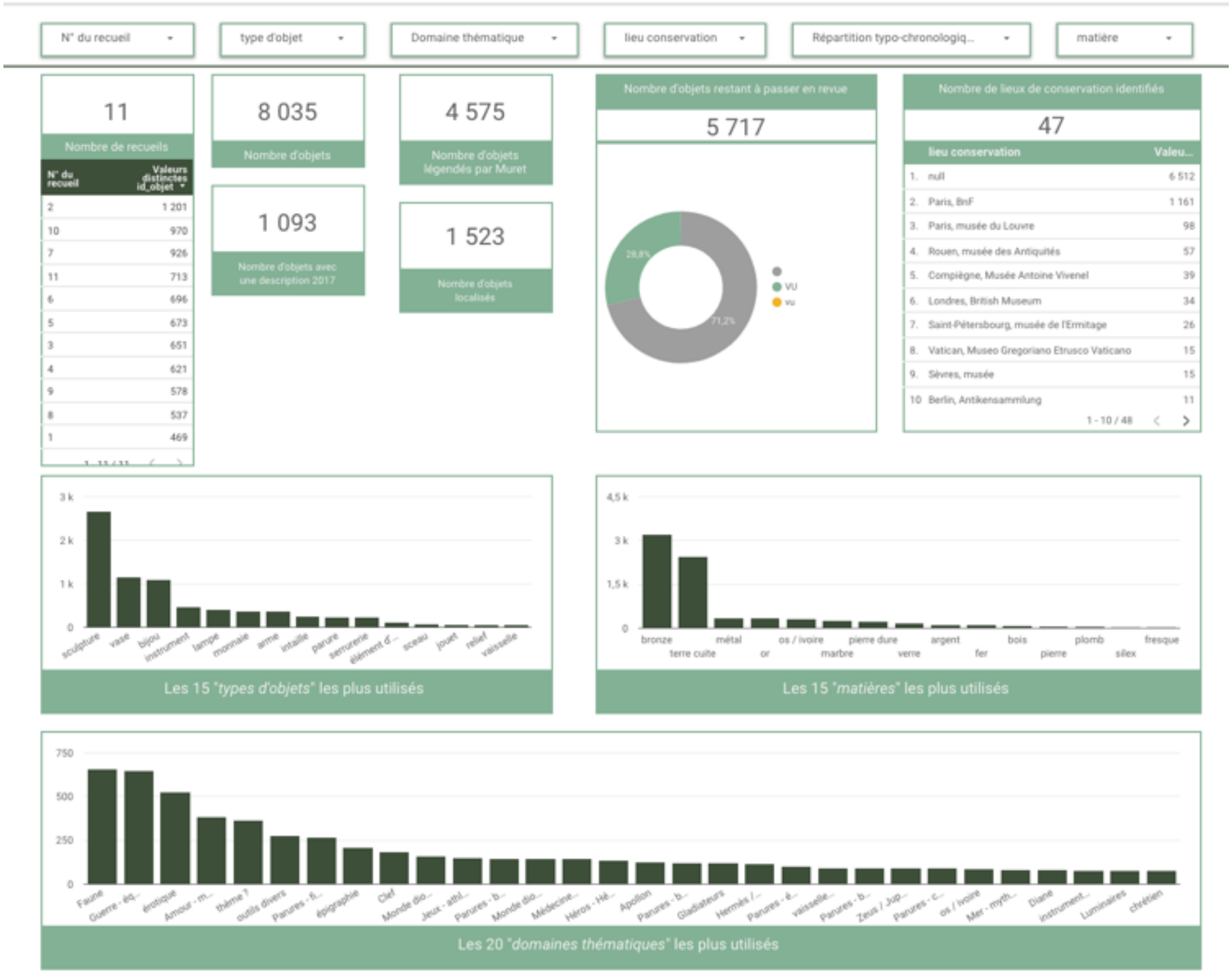




http://va.goodformandspectacle.com/

# V&V Spelunkeer - Cas d’utilisation de l’API du V&V






- Montrer en live par de la data, la recherche *work in progress* via l'utilisation d'un dashboard dynamique

https://goo.gl/kWXLoF

# TOOLS

OpenRefine

VOYANT
see through your text

Gephi

WIKIDATA

Datawrapper
Create high-quality charts & maps in minutes.

# import.io

Site web: https://import.io/

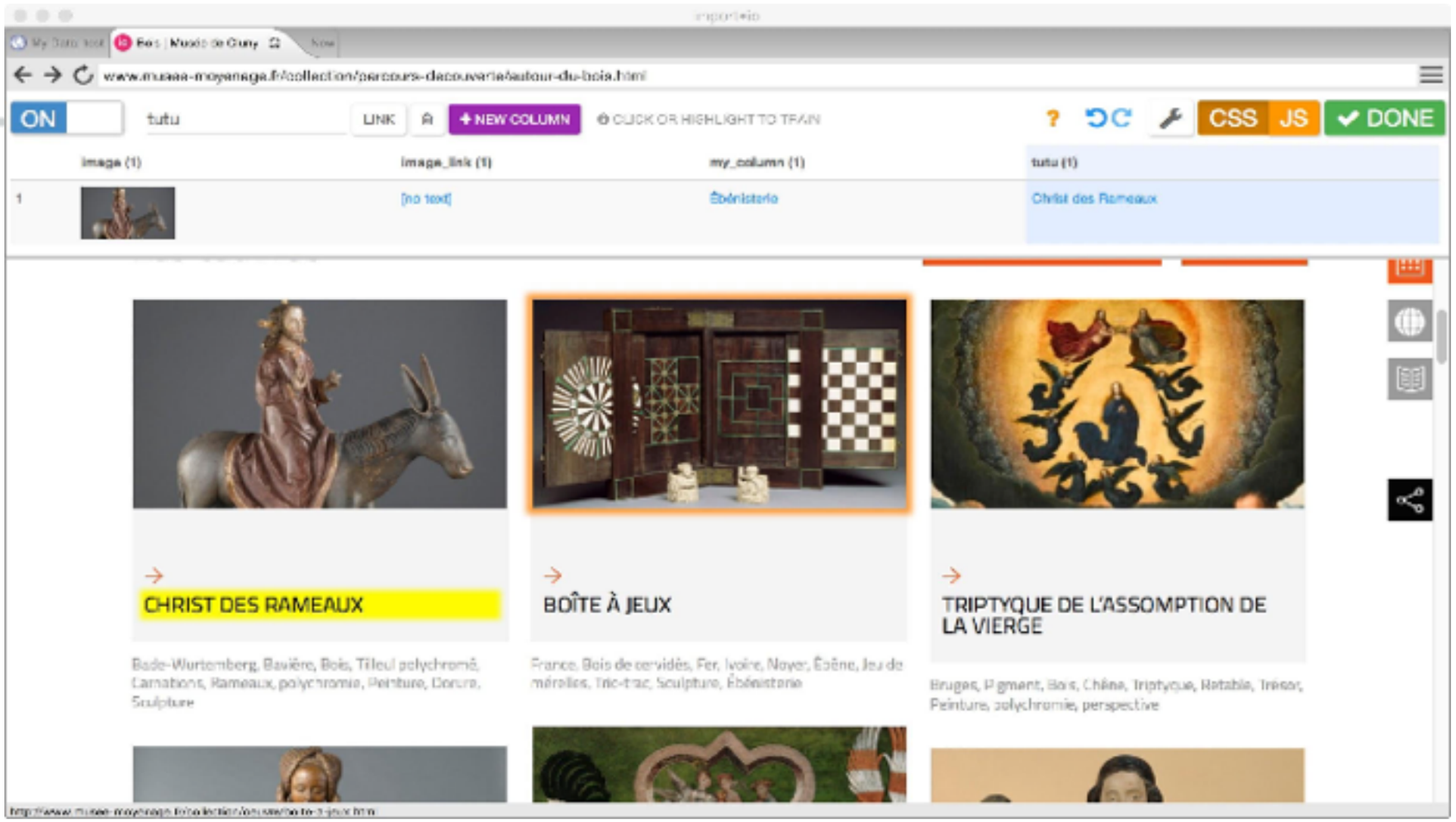
Ressources: http://goo.gl/VzlCGb

Pour faire quoi ?

- Acquérir et formater des données

Les avantages

- Prise en main aisée et courbe de progression rapide
- Fait pour les non développeurs
- Développements très actifs
- intègre pot.ly pour visualiser des données sous forme de graphe
- permet de créer une api



# app.raw

Site web: http://app.raw.densitydesign.org/

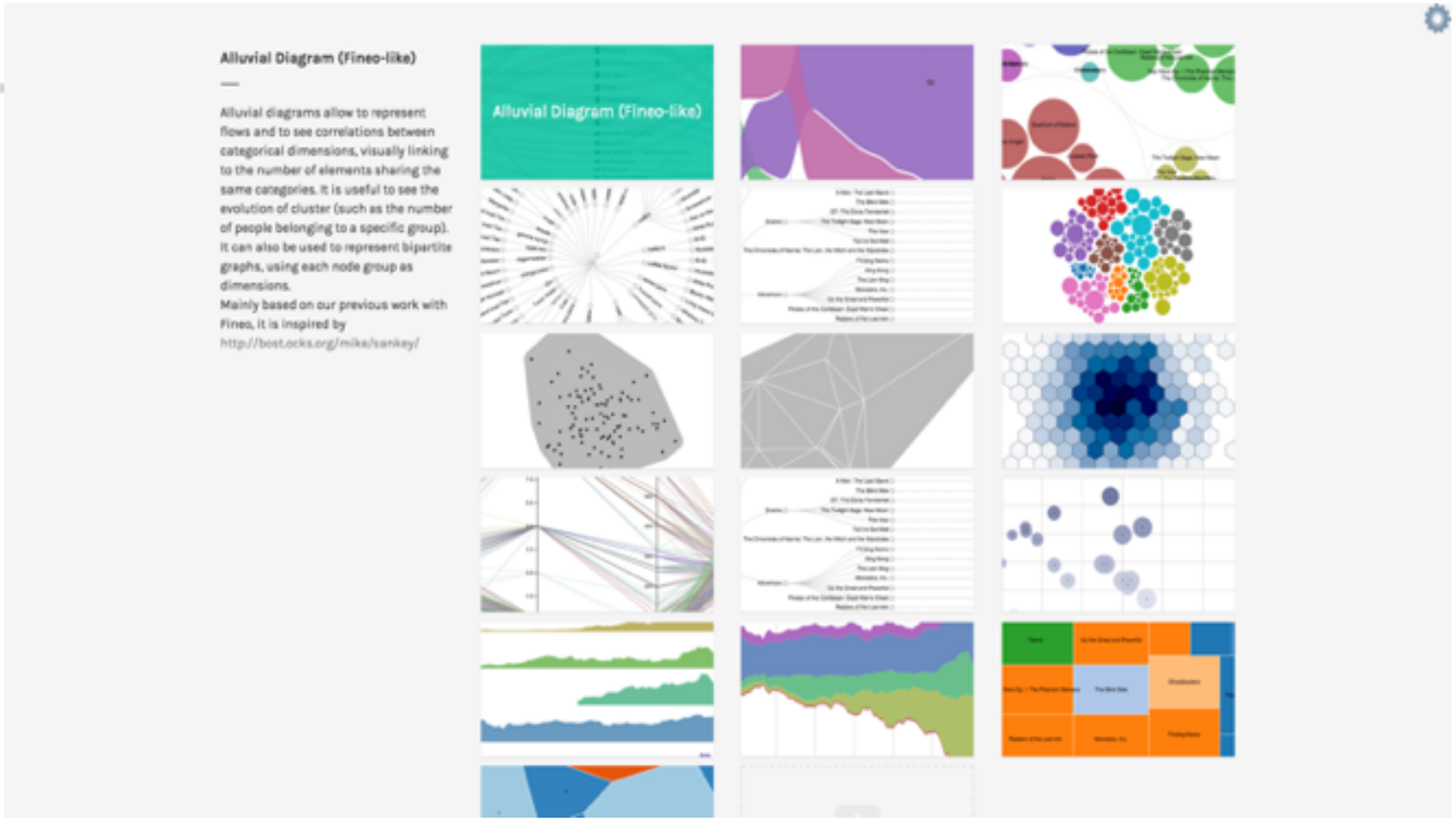
Ressources: http://goo.gl/qu5rCK

Pour faire quoi ?

- Visualiser des données facilement avec données en CSV en utilisant la librairie D3.js

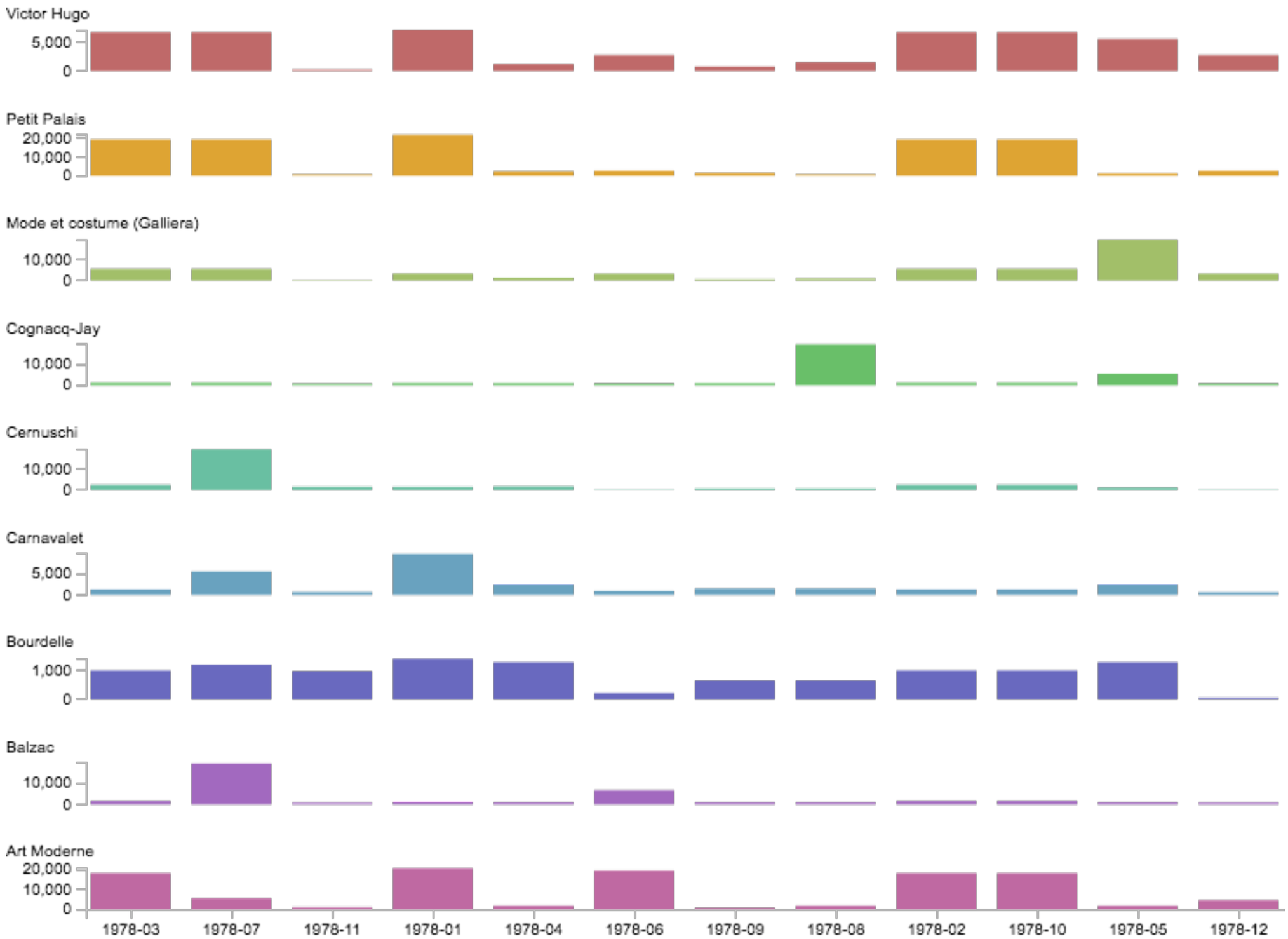
Les avantages

- Très facile à prendre en main
- Fonctionne avec du simple csv ou données tabulées
- Quelques options selon la forme du graphique
- Exporte les visualisation en SVG, PNG et les données en json (pr ensuite, aller + loin)



# A vous de jouer !

- Télécharger un jeu de données
  - datas - AppRaw1.csv
- Tester différents graphes et poser vous des questions ?




- Que pouvons-nous faire comme critiques ?

# gephi

Site web: http://gephi.github.io/

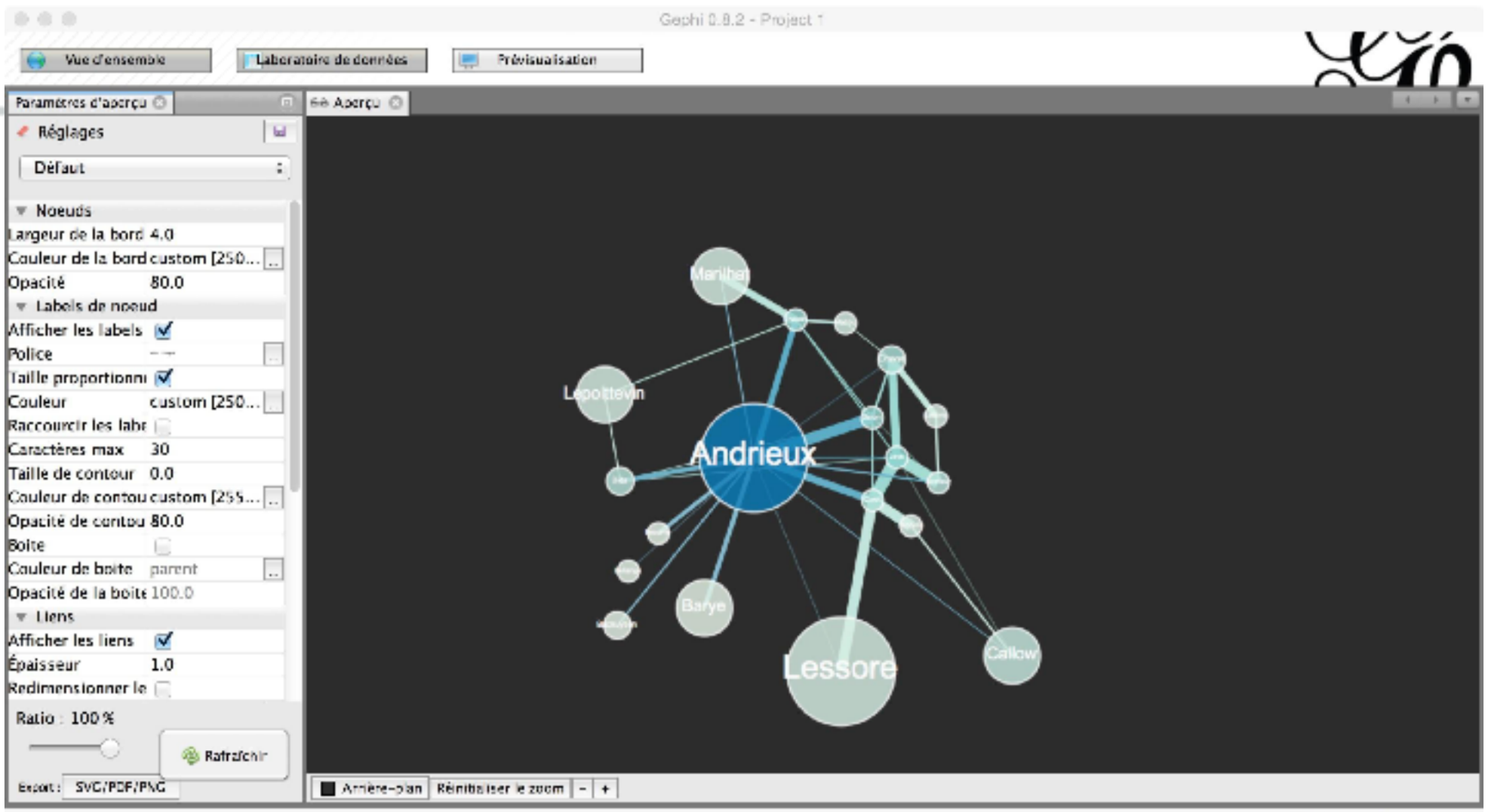
Ressources: http://goo.gl/i48zJd

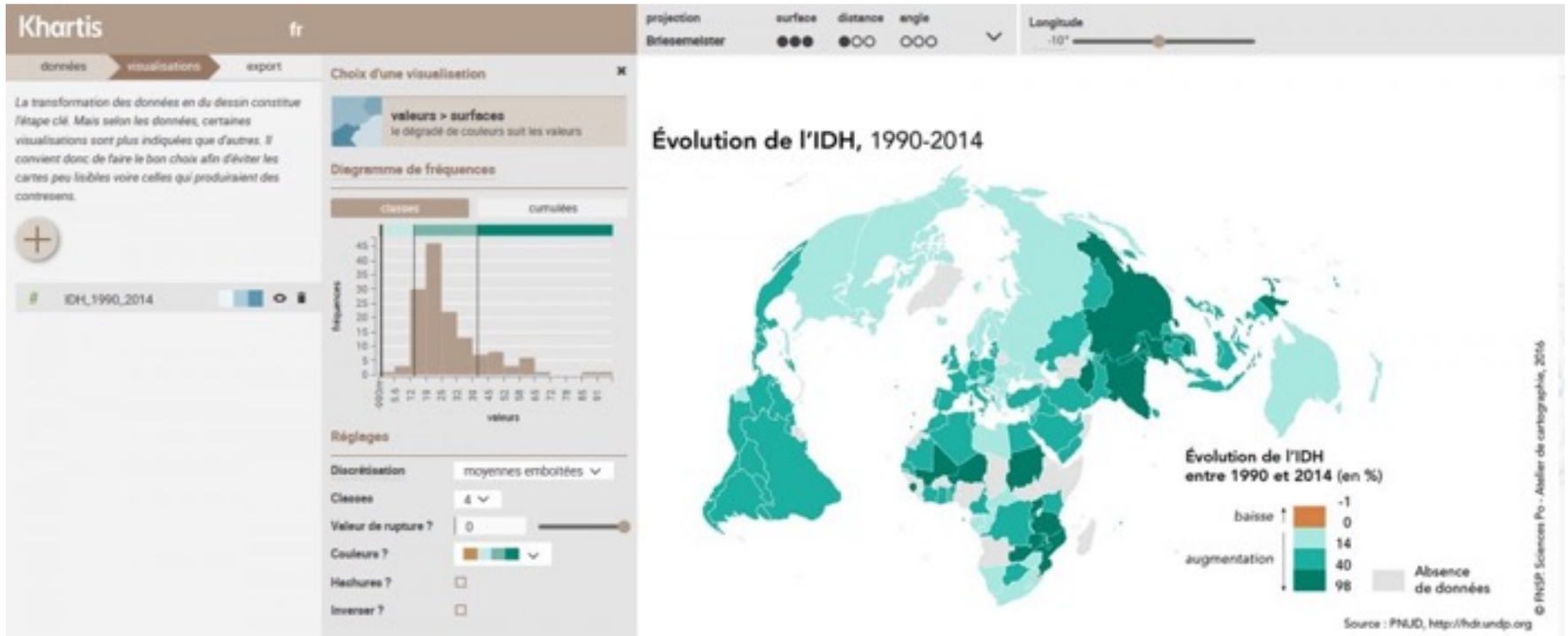
Pour faire quoi ?

- Visualiser les données sous forme de réseaux.

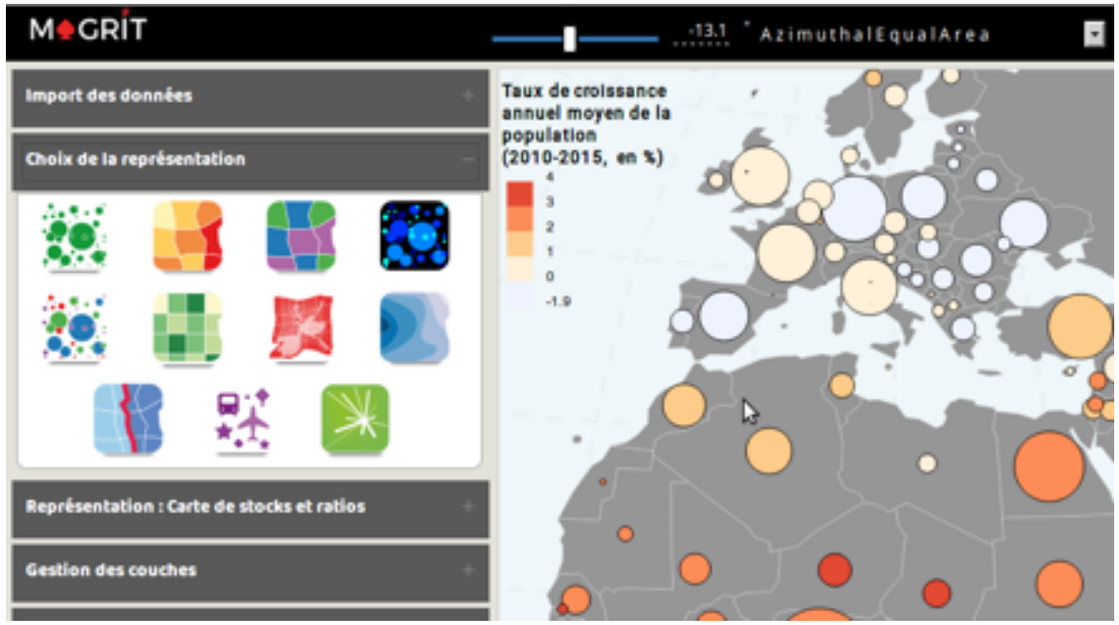
Les avantages

- Logiciel très complet en open source
- De nombreux plugins
- Personnalisation des visualisations
- Nombreuses sorties possibles: format réseaux, svg, png, pdf, etc.

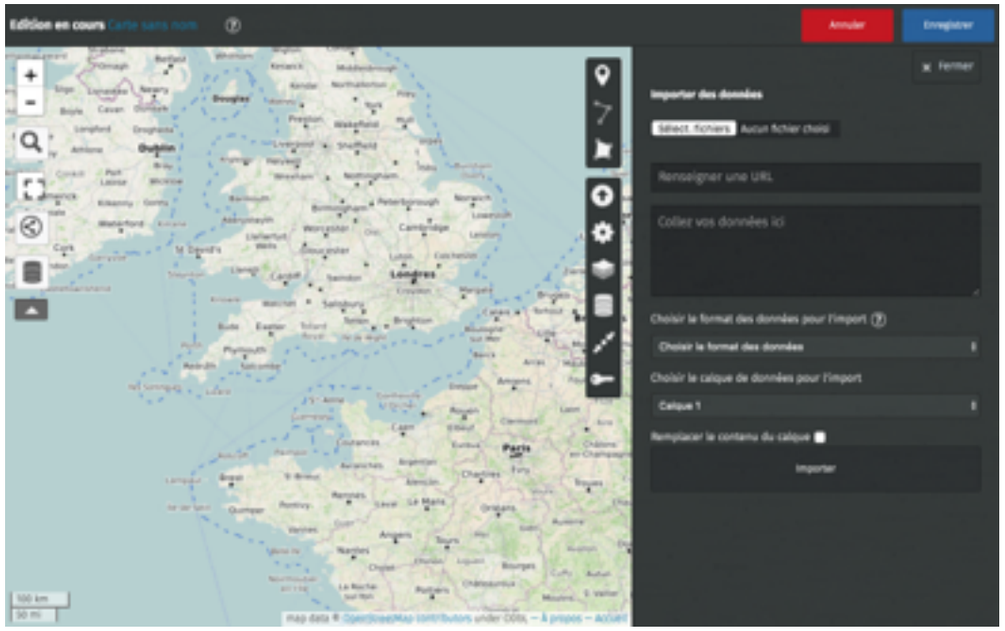





http://www.sciencespo.fr/cartographie/khartis/

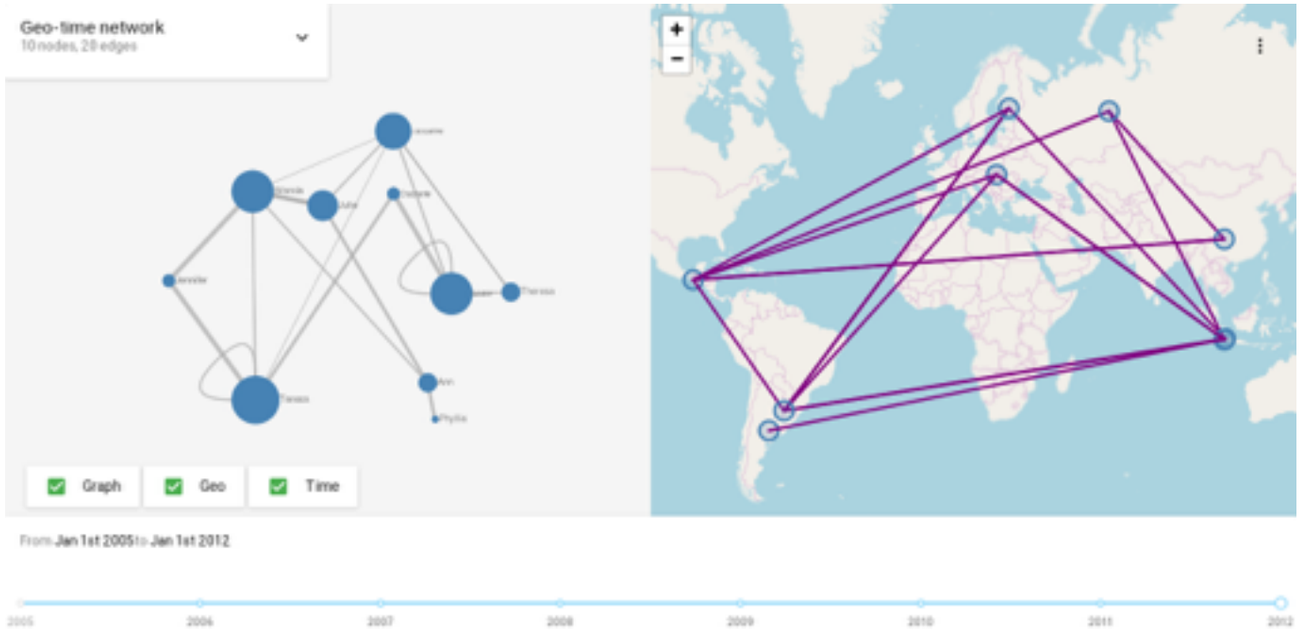



http://magrit.cnrs.fr/



https://umap.openstreetmap.fr

https://goo.gl/btXWbj

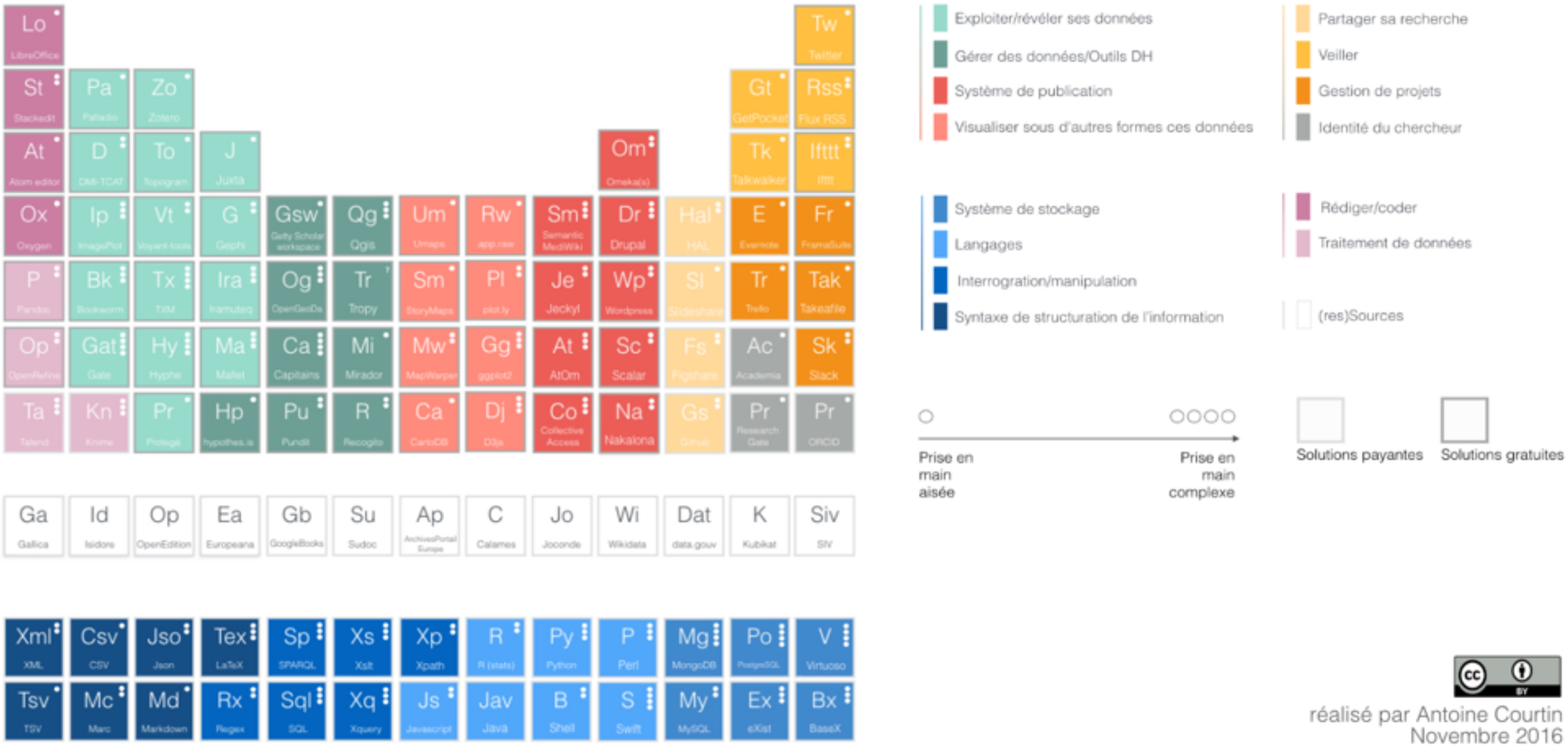



http://topogram.io/#/?id=topogram

Weviz

Démarche de création collaborative

https://goo.gl/rwoHNu




réalisé par Antoine Courtin
Novembre 2016

https://goo.gl/Ll9nVj

# Pour choisir l’outil, se poser des questions :

- Qu’est-ce que je souhaite faire apparaître ?
- Quels sont les formats de mes données ?
- Quel est le type de mes données (temporelle, cartographique, statistiques, etc.) ?
- Est-ce pour explorer un jeux de données par moi même ou l’objectif est de partager le résultat ?
- Est-ce pour du *print* ou pour une version en ligne ? Et si oui, est-il nécessaire d’avoir une version dynamique ?

# Conclusion

- Visualisation très efficace mais à condition de suivre des règles strictes
- Tout ne s’y prête pas forcément
- Garder un esprit critique faces aux visualisations
- Les visualisations efficaces nécessitent un apprentissage

“On ne lit pas un graphique, on lui pose des questions”

Jacques Bertin, 1981

# Palladio

# palladio

Site web: http://palladio.designhumanities.org/

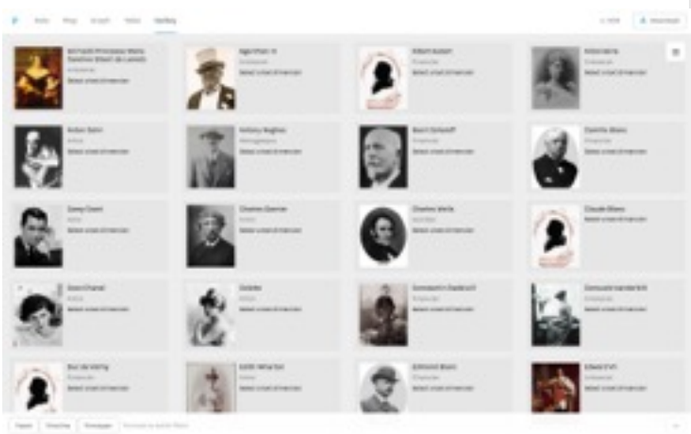
Ressources: http://goo.gl/9hZmek

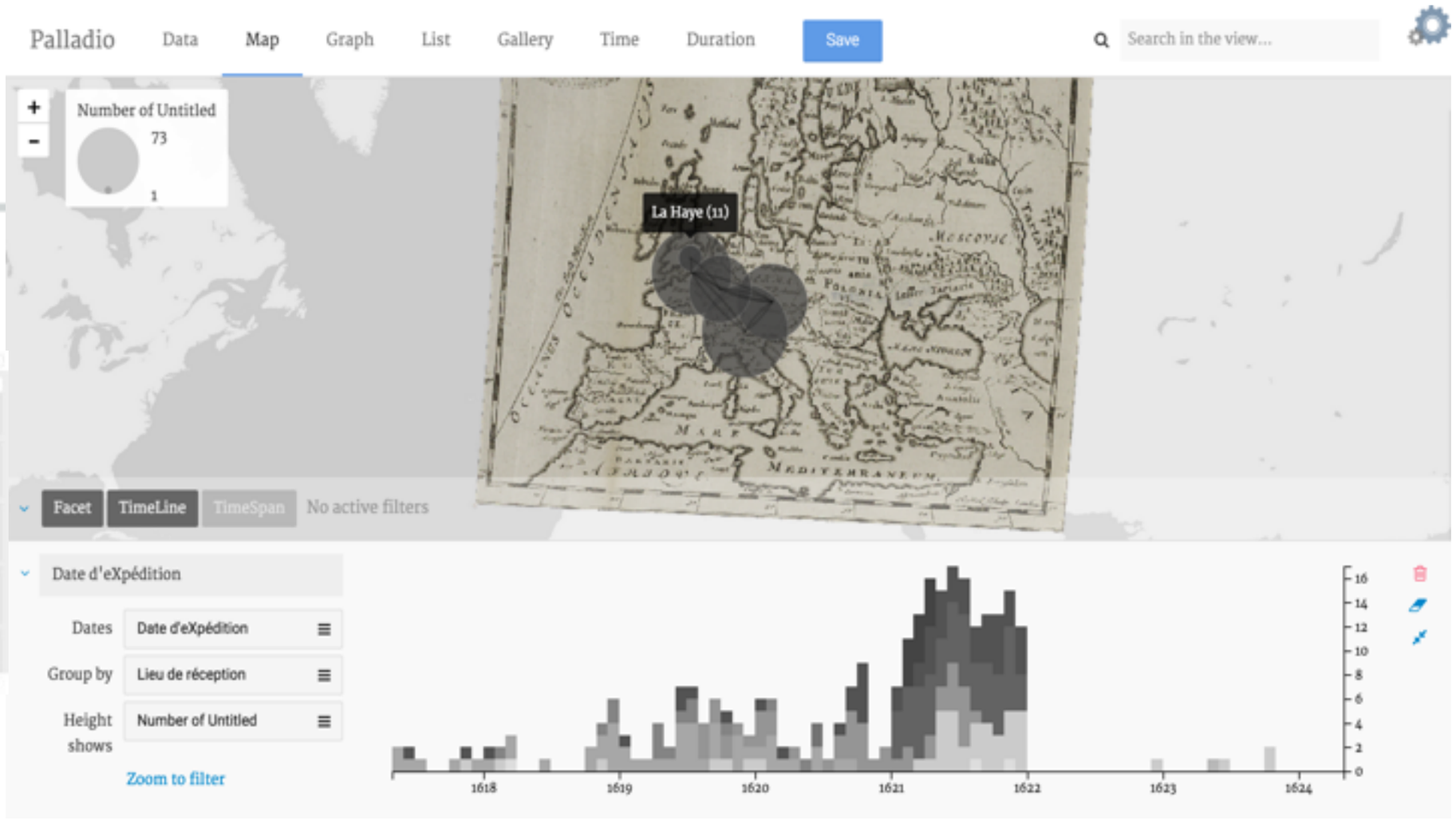
Pour faire quoi ?

- Explorer ses données sous forme temporelle, géographique, de réseaux, tableaux, etc.

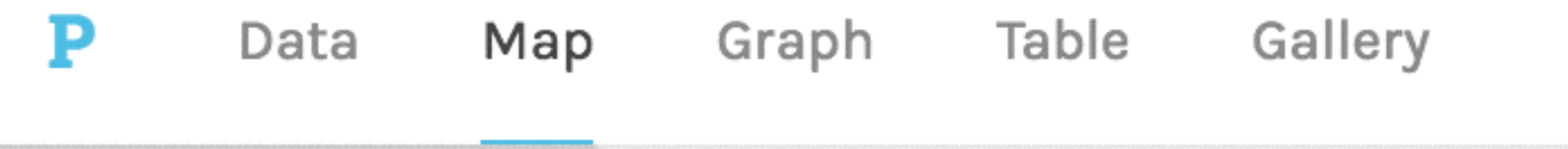
Les avantages

- Très simple d'utilisation
- Nombreux filtres dynamiques
- export

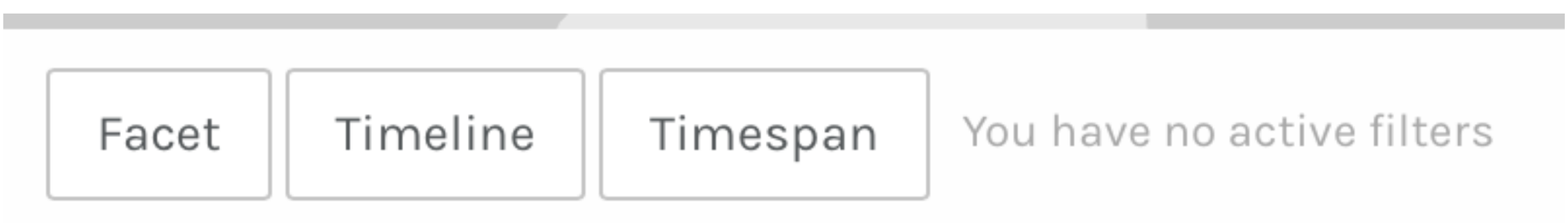


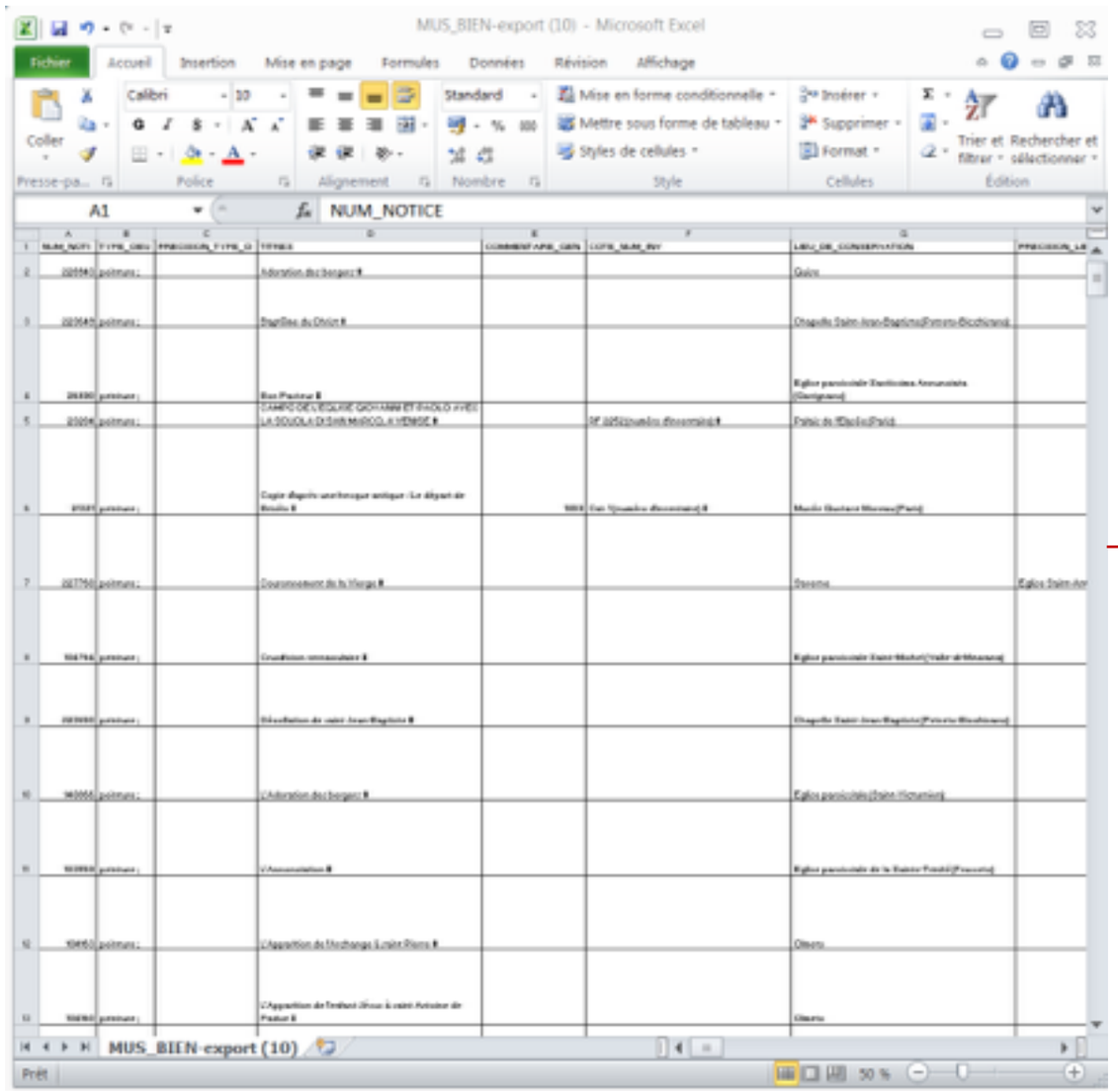


Palladio est divisé en 5 grandes rubriques:



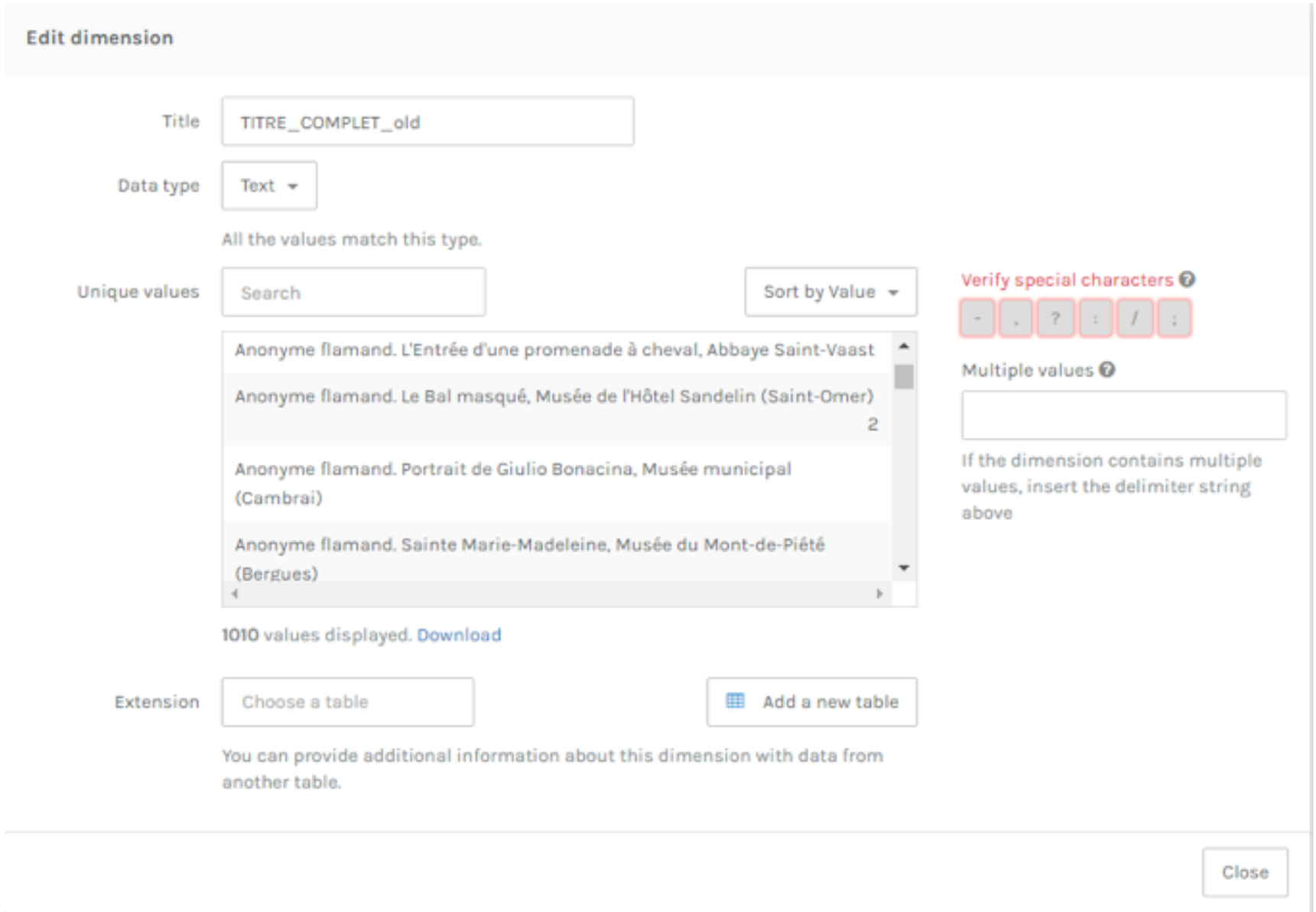

Et des Filtres de différents types !




MUS_BIEN-export (10) - Microsoft Excel
Fichier
Accueil
Insertion
Mise en page
Formules
Données
Révision
Affichage
A1
NUM_NOTICE
MUS_BIEN-export (10)
Prêt

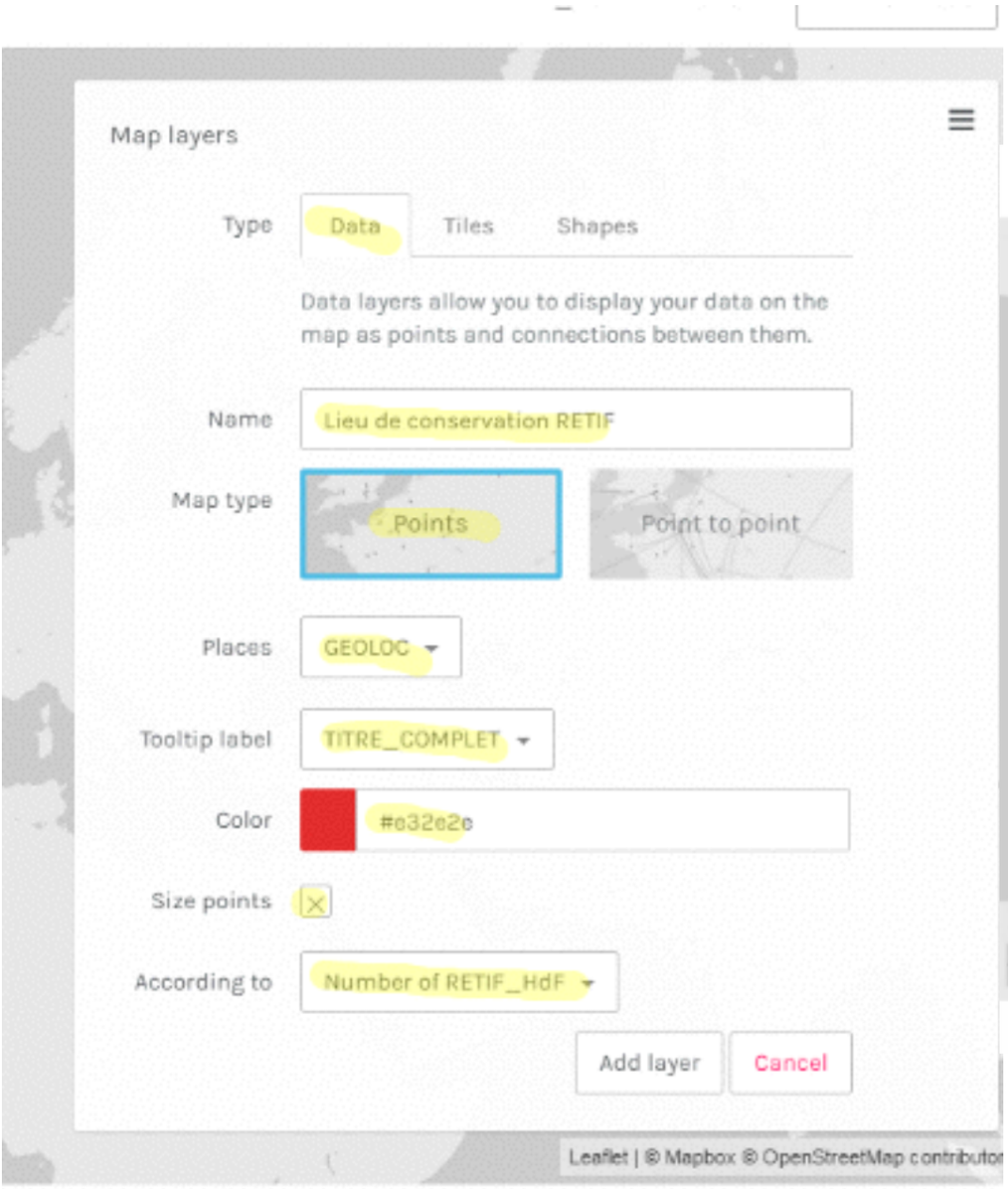
# DATA

Permet de faire le tour de ses data, d'en changer le type, de vérifier les éventuels erreurs, de rajouter à posteriori, des tables intermédiaires, etc.



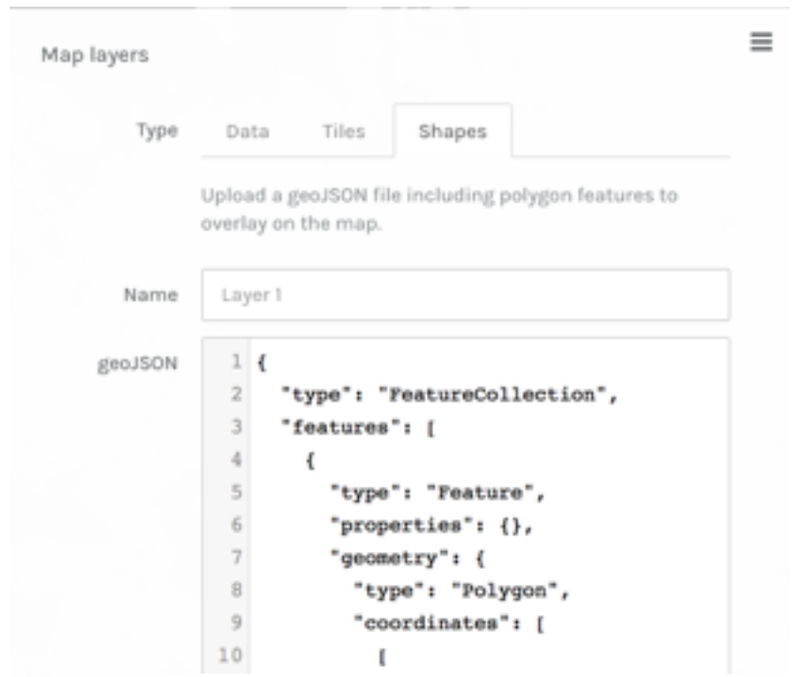
# MAP

Permet de créer des cartes. La rubrique la plus sympa de Palladio car il est possible d'ajouter ses data (préalablement géolocalisées) mais aussi jouer sur des fonds de carte, ajouter des shapes, etc.



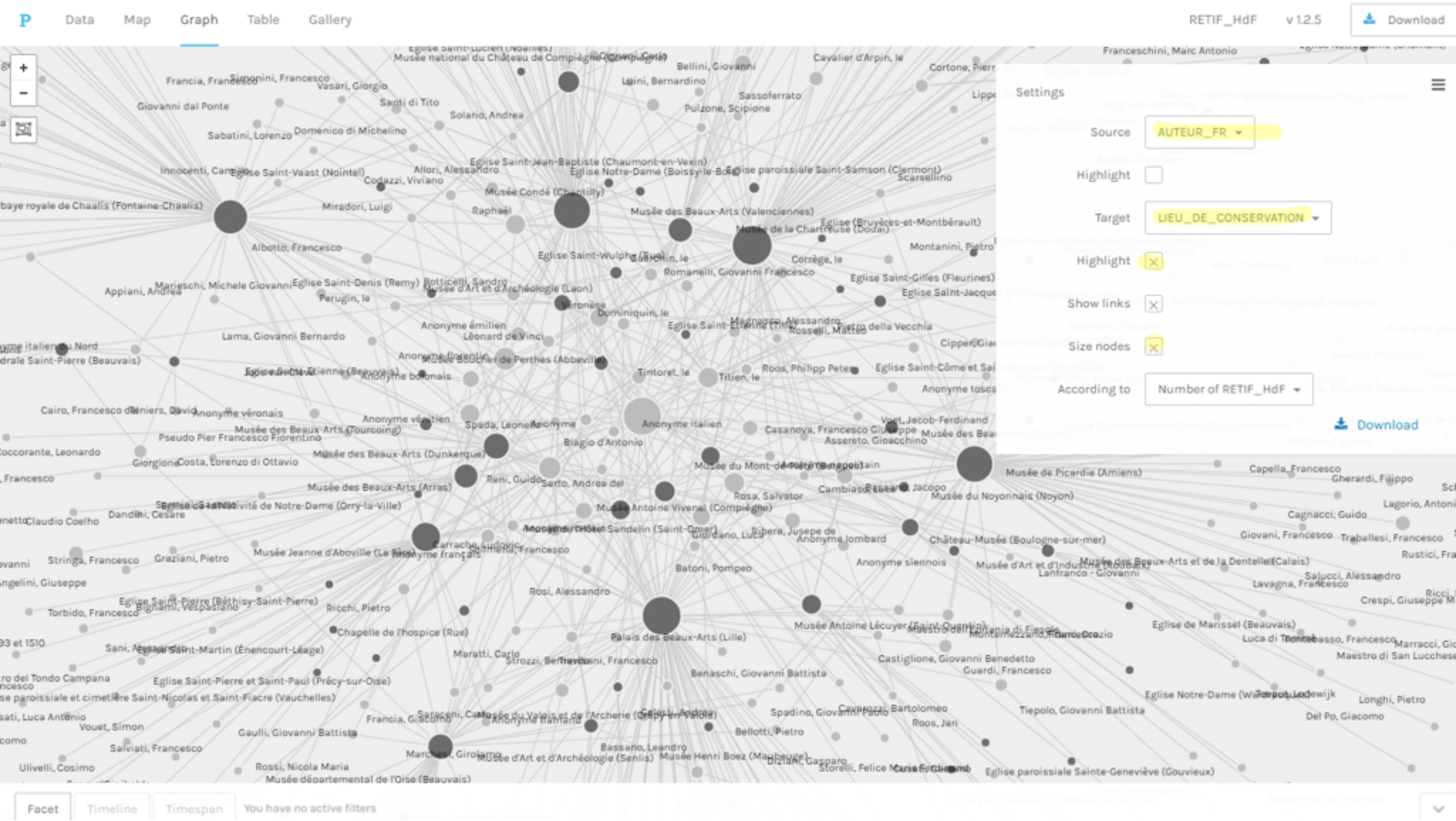
# MAP

Permet de créer des cartes. La rubrique la plus sympa de Palladio car il est possible d'ajouter ses data (préalablement géolocalisées) mais aussi jouer sur des fonds de carte, ajouter des shapes, etc.



# GRAPH

Permet de visualiser sous la forme d'un graph de réseau, les relations entre 2 champs. Il suffit de choisir un champ "source" et un champ "cible" et Palladio créé un réseau simple.



# TABLE

Espace permettant de générer des listes en réalisant des regroupements logique à partir des données.

P Data Map Graph Table Gallery RETIF_HdF v 1.2.5 Download

| | | | |
|---|---|---|---|
| Vaccaro, Andrea | attribué à | Palais des Beaux-Arts (Lille) | |
| Ulivelli, Cosimo | attribué à | Musée Jeanne d'Aboville (La Fère) | |
| Troncossi, Giuseppe Francesco | de | Musée des Beaux-Arts (Valenciennes) | |
| Trevisani, Francesco | genre de, de, copié d'après | Eglise (Méru), Musée départemental de l'Oise (Beauvais), Musée Henri Boez (Maubeuge), Musée de la Chartreuse (Douai) | Beauvais, Le Quadrilatère (du 27 avril au 17 septembre 2017) |
| Traballesi, Francesco | de | Musée de Picardie (Amiens) | |
| Toscani Giovanni di Francesco | de | Eglise paroissiale Saint Sulpice (Pierrefonds) | Amiens, musée de Picardie (du 10 mars au 2 juillet 2017) |
| Torbido, Francesco | attribué à | Musée des Beaux-Arts (Arras) | |
| Toeput, Lodewijk | de | Musée de Picardie (Amiens) | Chantilly, musée Condé (du 24 mars au 2 juillet 2017) |
| Tiziano | atelier de | Musée Condé (Chantilly) | Chantilly, musée Condé (du 24 mars au 2 juillet 2017) |
| Tito, Tiberio di | attribué à, copié d'après | Musée de la Chartreuse (Douai) | |
| Titien, le | copié d'après, attribué à, d'après | Eglise Saint-Martin (Longueil-Sainte-Marie), Musée de Picardie (Amiens), Eglise paroissiale Saint-Samson (Clermont), Eglise Saint-Côme-et-Saint-Damien (Rosoy), Musée des Beaux-Arts (Valenciennes), Eglise Saint-Nicolas et Saint-Blaise (Oursel-Maison), Eglise Saint-Jean-Baptiste (Chaumont-en-Vexin), Eglise Saint-Jacques (Compiègne), Eglise Sainte-Maxence (Pont-Sainte-Maxence), Eglise Notre-Dame (Boissy-le-Bois), Musée des Beaux-Arts (Dunkerque), Palais des Beaux-Arts (Lille), Musée de l'Hôtel Sandelin (Saint-Omer), Musée d'Art et d'Archéologie (Laon), Musée Condé (Chantilly), Eglise Saint-Wulphy (Rue), Musée Jean Calvin (Noyon) | Chantilly, musée Condé (du 24 mars au 2 juillet 2017) |

Settings

Row dimension: AUTEUR_FR

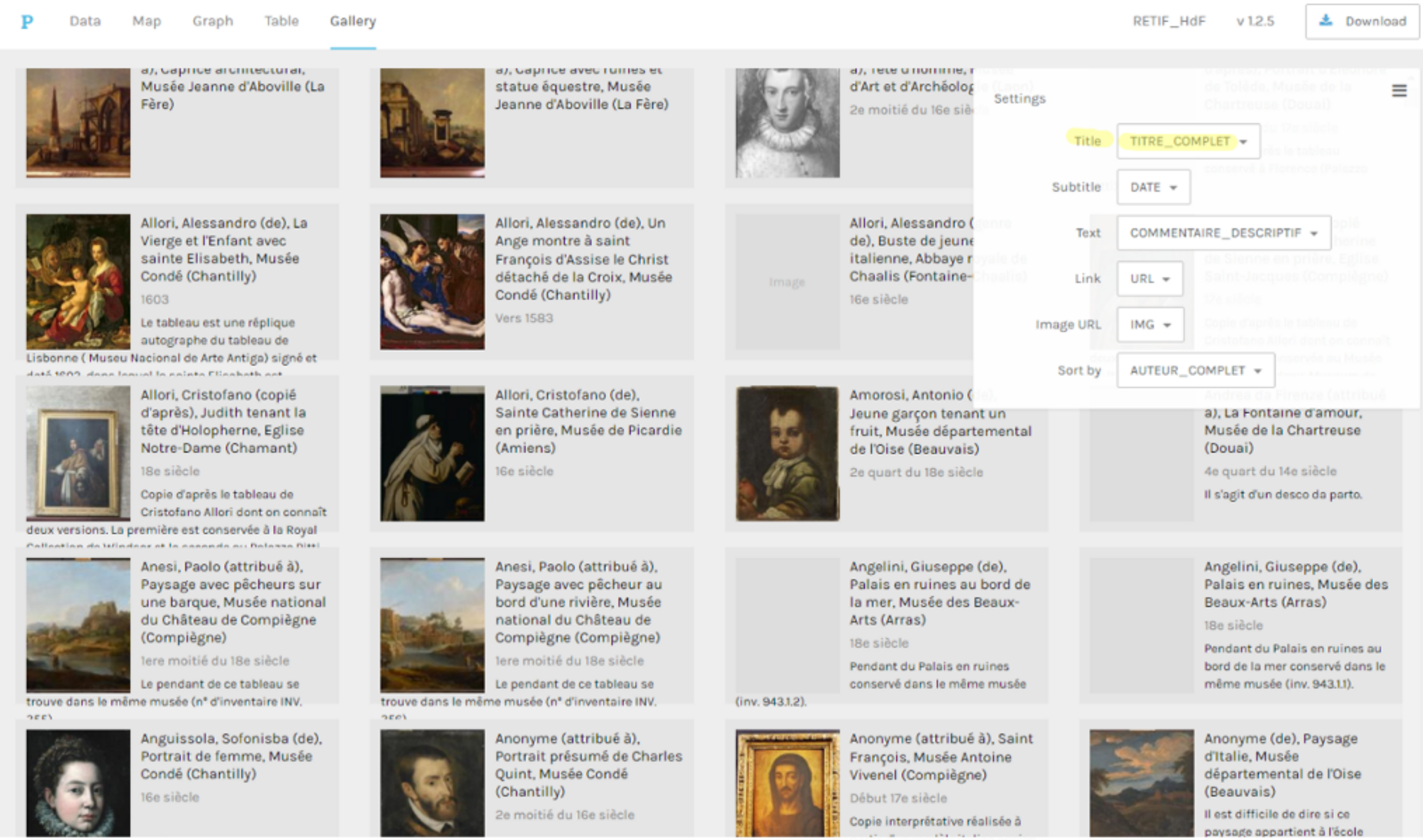
At least one row per value in this dimension. Multiple values will be displayed as lists in each cell.

Dimensions: AUTEUR_QUALIF, LIEU_DE_CONSERVATION, LIE...

Download

# GALLERY

Grâce à la récupération des URL des vignettes, il est possible de générer cette galerie très facilement et de choisir les éléments qui en constituent les cartels. La source de l'information n'est pas perdue car il est possible d'associer une URL à chaque enregistrement



# LES FILTRES

Facet | Timeline | Timespan | You have no active filters

| auteur (33) | |
|---|---|
| Albani, Francesco (1578 - 1... | 20 / 20 |
| Baglione, Giovanni (Vers 15... | 13 / 13 |
| Guercino (1591 - 1666) | 12 / 12 |
| Benaschi, Giovanni Battist... | 6 / 6 |
| Gaulli, Giovanni Battista (1... | 6 / 6 |
| Assereto, Gioacchino (1600... | 4 / 4 |
| Belvedere, Andrea (1642 - 1... | 3 / 3 |

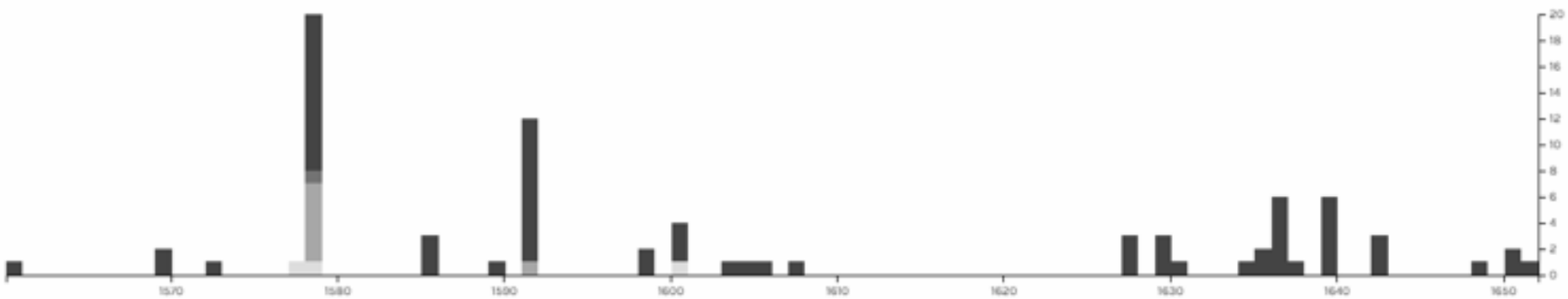
| date (26) | |
|---|---|
| 17e siècle | 48 / 48 |
| 1620 | 9 / 9 |
| 1ere moitié du 17e siècle | 8 / 8 |
| 2e moitié du 17e siècle | 5 / 5 |
| 1er quart du 17e siècle | 4 / 4 |
| 1640 | 2 / 2 |
| 1630 | 2 / 2 |

| titre (96) | |
|---|---|
| L'Annonciation | 3 / 3 |
| Moïse sauvé des eaux | 2 / 2 |
| Loth et ses filles | 2 / 2 |
| Apparition de Notre Dame ... | 1 / 1 |
| Clio | 1 / 1 |
| Erato | 1 / 1 |
| Circé | 1 / 1 |

| technique (4) | |
|---|---|
| toile | 86 / 86 |
| cuivre | 9 / 9 |
| bois | 4 / 4 |
| marbre | 1 / 1 |

| Date_Naissance_pientre (27) | |
|---|---|
| 1578 | 20 / 20 |
| | 19 / 19 |
| 1591 | 12 / 12 |
| 1639 | 6 / 6 |
| 1636 | 6 / 6 |
| 1600 | 4 / 4 |
| 1585 | 3 / 3 |

Facet | Timeline | Timespan | You have no active filters




Timeline

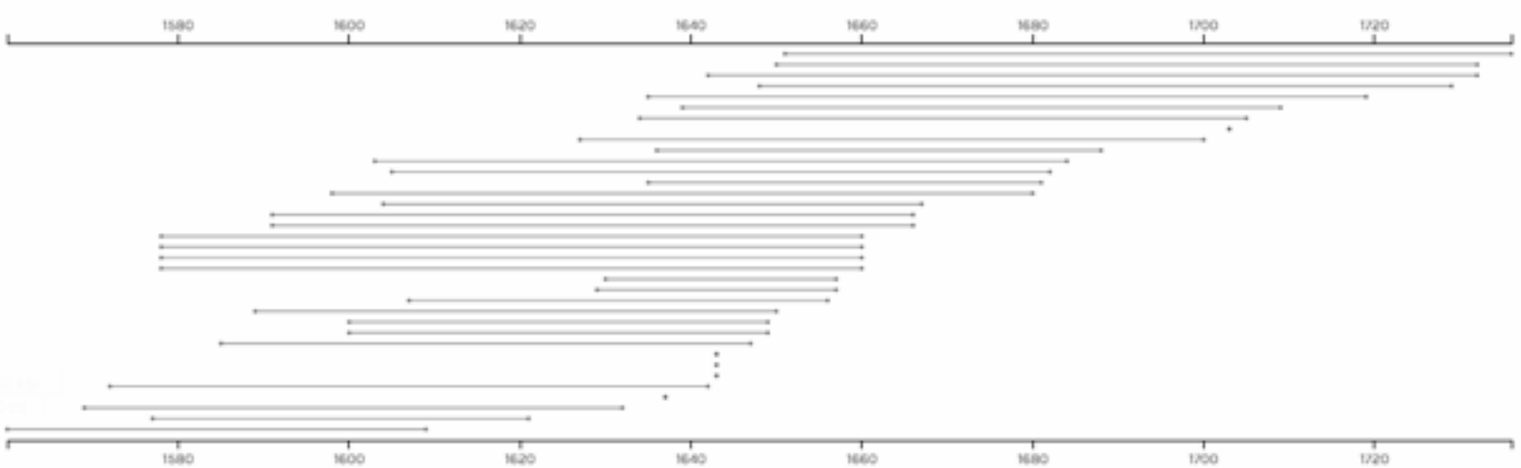
Description: Date_Naissance_pientre

Dates: Date_Naissance_pientre

Height: Number of Untitled

Group by: technique

Clear




Timespan

Description: Time Span Filter

Layout: Bars

Start date: date_mort_auteur

End date: Date_Naissance_pientre

Label: technique

Group: technique

Clear

# A vous de jouer !

- Télécharger un jeu de données test déjà formaté pour Palladio
  - Importer le fichier et créer une carte pour visualiser les lieux de conservation https://hdlab.stanford.edu/palladio/
  - créer une galerie d’image
  - créer une table avec les dimensions d’analyse de votre choix

http://hdlab.stanford.edu/palladio-app/#/upload

P

v 1.2

**Create a new project**

Load an existing project

Try with sample data

Create a new Palladio project by uploading your data from a spreadsheet or flat-file, or load data from a SPARQL endpoint. Not sure how Palladio works?

### Load .csv or spreadsheet

Copy and paste out of your spreadsheets, drag-and-drop to upload tabular data (e.g. .csv, .tab, .tsv), or link to a file in a public Dropbox folder to create a new Palladio project. Not sure how to structure your data?

1

Load

**More than one table? No problem!** If you have more than one table, start by uploading your primary table. The primary table should contain the main entities you want to visualize. It could be a collection of objects, like persons or letters, or more abstract concepts, like relationships or flows. Once you have uploaded your primary table, you will be able to extend it with additional information from other tables.

Load data from a SPARQL endpoint (beta)

**SPARQL endpoint (beta)**

You can load data from a SPARQL endpoint by providing both the endpoint URL and a valid SPARQL query. After you run the SPARQL query you will have the opportunity to validate your data and re-run the query if necessary before loading the data into Palladio.

SPARQL endpoint

https://query.wikidata.org/bigdata/namespace/wdq/sparql

1

La requête SPARQL

Run query

Create a new Palladio project by uploading your data from a spreadsheet or flat-file, or load data from a SPARQL endpoint. Not sure how Palladio works?

**Load .csv or spreadsheet**

Copy and paste out of your spreadsheets, drag-and-drop to upload tabular data (e.g. .csv, .tab, .tsv), or link to a file in a public Dropbox folder to create a new Palladio project. Not sure how to structure your data?

```
1 creatorLabel,datebirth,datedeath,placebirthLabel,coord,lat_long,placedeathLabel,coord_2,lat_long
  _2
2 Friedrich Sustris,1540-01-01T00:00:00Z,1599-01-01T00:00:00Z,Italie,Point(12.0 43.0),"43,
  12",Munich,,"48.133333333333, 11.566666666667"
3 Pieter Brueghel l'Ancien,1520-01-01T00:00:00Z,1569-09-19T00:00:00Z,Brée,Point(5.583333333
  51.133333333),"51.133333333333, 5.5833333333333",Bruxelles,,"50.8467, 4.3547"
4 Johannes Vermeer,1632-10-31T00:00:00Z,1675-12-15T00:00:00Z,Delft,Point(4.359444444
  52.011944444),"52.011944444444, 4.3594444444444",Delft,,"52.011944444444, 4.3594444444444"
5 Barbara Longhi,1552-10-11T00:00:00Z,1638-12-23T00:00:00Z,Ravenne,Point(12.199444444
  44.417777777),"44.417777777778, 12.199444444444",Ravenne,,"44.417777777778, 12.199444444444"
6 Arent de Gelder,1645-10-26T00:00:00Z,1727-08-27T00:00:00Z,Dordrecht,Point(4.67 51.81),"51.81,
  4.67",Dordrecht,,"51.81, 4.67"
7 Giovanni Benedetto Castiglione,1609-01-01T00:00:00Z,1665-01-01T00:00:00Z,Gênes,Point(8.933983333
  44.407186111),"44.407186111111, 8.9339833333333",Mantoue,,"45.156388888889, 10.791111111111"
8 Pieter Brueghel l'Ancien,1525-01-01T00:00:00Z,1569-09-19T00:00:00Z,Bréda,Point(4.383333333
```

Load

**More than one table? No problem!** If you have more than one table, start by uploading your primary table. The primary table should contain the main entities you want to visualize. It could be a collection of objects, like persons or letters, or more abstract concepts, like relationships or flows. Once you have uploaded your primary table, you will be able to extend it with additional information from other tables.

**SPARQL endpoint (beta)**

You can load data from a SPARQL endpoint by providing both the endpoint URL and a valid SPARQL query. After you run the SPARQL query you will have the opportunity to validate your data and re-run the query if necessary before loading the data into Palladio.

SPARQL endpoint

https://query.wikidata.org

```
14   BIND(YEAR(?date) AS ?year).
15 ?creator wdt:P19 ?placebirth. # lieu de naissance
16 ?creator wdt:P569 ?datebirth.
17 ?creator wdt:P570 ?datedeath.
18 ?placebirth wdt:P625 ?coord.
19 ?placebirth p:P625 ?declaration.
20 ?declaration psv:P625 ?coord_geo.
21 ?coord_geo wikibase:geoLatitude ?lat.
22 ?coord_geo wikibase:geoLongitude ?long.
23 ?creator wdt:P20? ?placedeath.
24 ?placedeath wdt:P625 ?coord_death.
25 ?placedeath p:P625 ?declaration2.
26 ?declaration2 psv:P625 ?coord_geo_2.
27 ?coord_geo_2 wikibase:geoLatitude ?lat_2.
28 ?coord_geo_2 wikibase:geoLongitude ?long_2.
```

Run query

Search: Show 10 entries

| | creatorLabel | datebirth | datedeath | placebirthLabel | coord | lat_long | plac |
|---|---|---|---|---|---|---|---|
| 1 | Friedrich Sustris | 1540-01-01T00:00:00Z | 1599-01-01T00:00:00Z | Italie | Point(12.0 43.0) | 43, 12 | Muni |
| 2 | Pieter Brueghel l'Ancien | 1520-01-01T00:00:00Z | 1569-09-19T00:00:00Z | Brée | Point(5.583333333 51.133333333) | 51.133333333333, 5.5833333333333 | Bruxe |
| 3 | Johannes Vermeer | 1632-10-31T00:00:00Z | 1675-12-15T00:00:00Z | Delft | Point(4.359444444 52.011944444) | 52.011944444444, 4.3594444444444 | Delft |
| 4 | Barbara Longhi | 1552-10-11T00:00:00Z | 1638-12-23T00:00:00Z | Ravenne | Point(12.199444444 44.417777777) | 44.417777777778, 12.199444444444 | Rave |
| 5 | Arent de Gelder | 1645-10- | 1727-08- | | | | |

Les data sont « uploadable »

Pour la requête : https://goo.gl/PjN9uk

Résultat de la requête

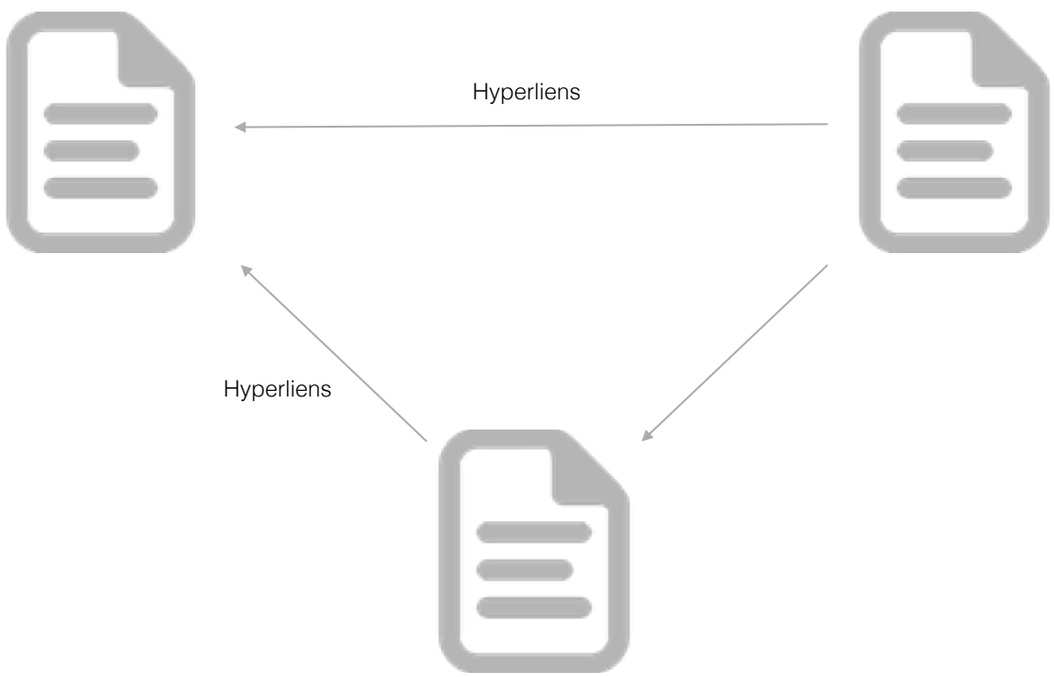
# A vous de jouer !

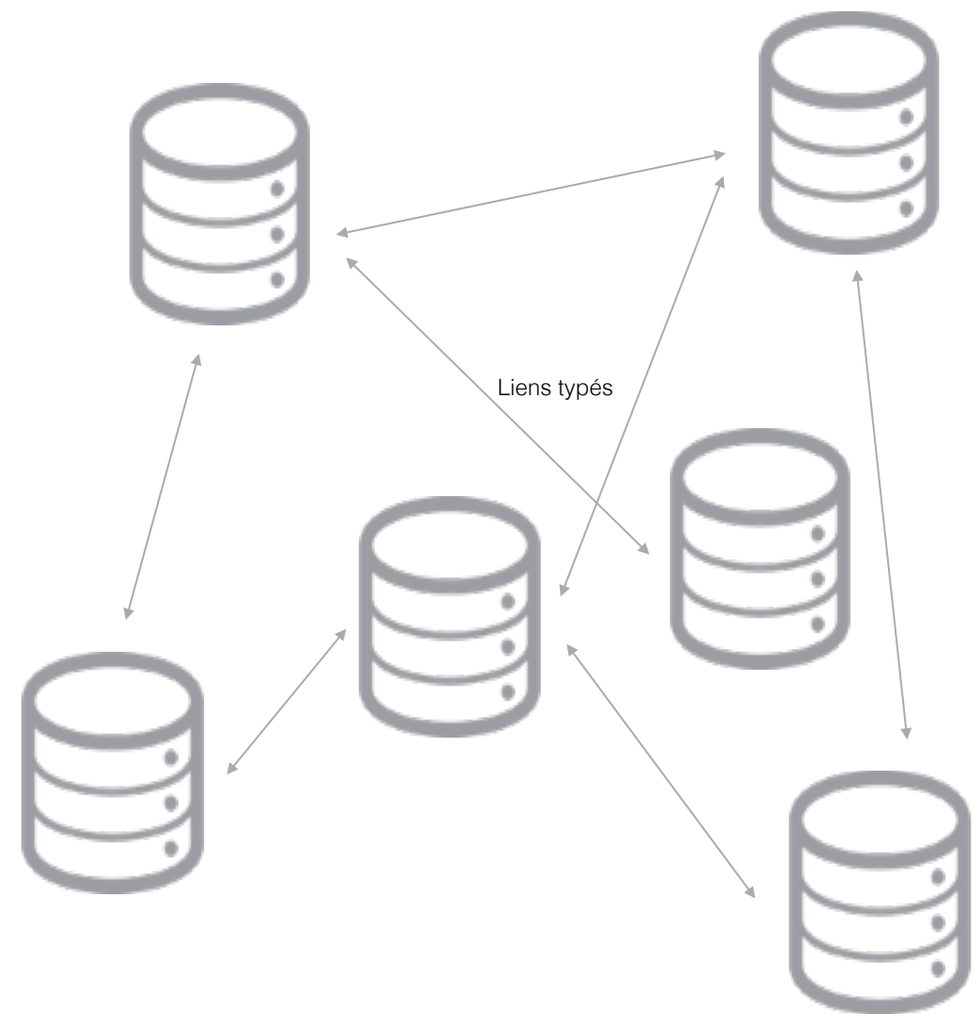
- Récupérer le code de la requête SPARQL tinyurl.com/ybrcyhb8 et importer les données dans Palladio
- Créer une carte avec lieu de naissance et lieux de mort
  - Jouer avec les filtres timespan
- Retourner dans wikidata pour
  - Modifier la requête pour non plus avoir les peintres mais les sculpteurs au XVIIe siècle et récupérer les portraits de ces personnes
  - recharger les données dans Palladio puis créer une galerie.

# Rappel !

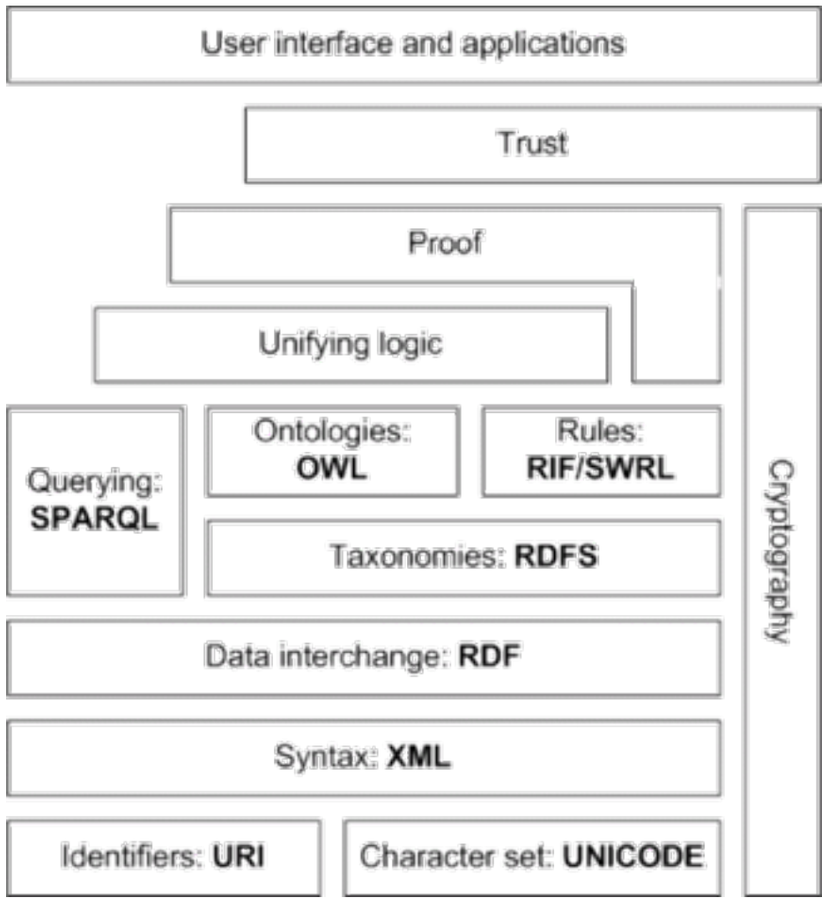
# Un web de données liées (1)

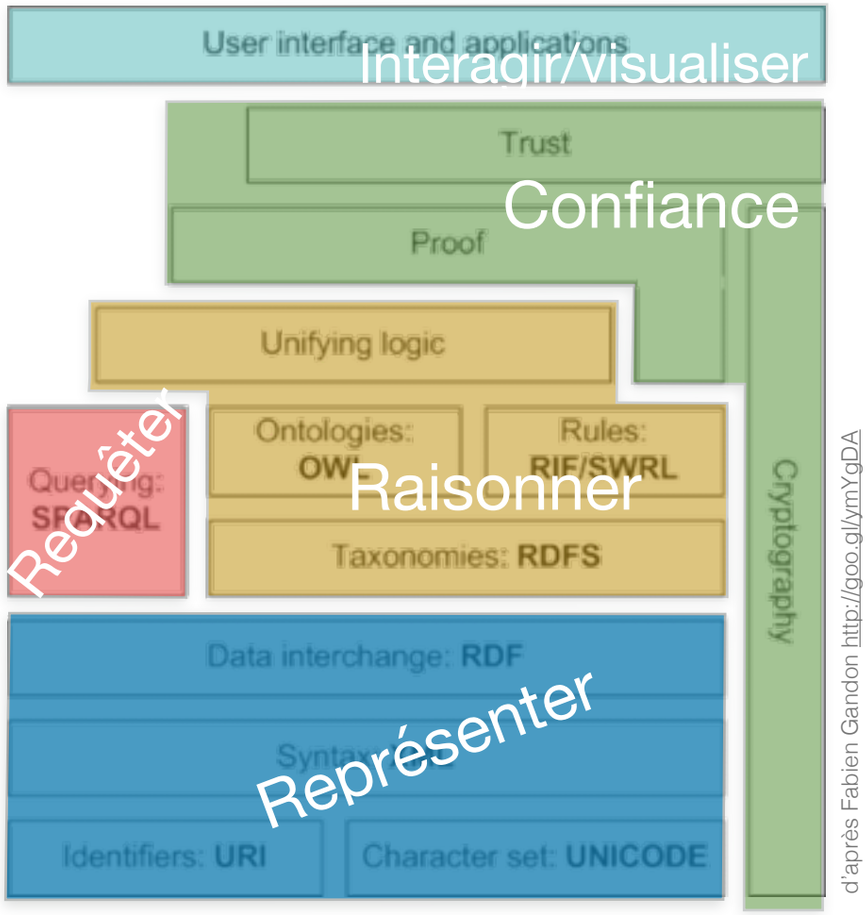
Web de document → Web de données



# Un web de données liées (2)




d'après Fabien Gandon http://goo.gl/ymYgDA

Le Web de données souvent dénommé linked data vise donc à favoriser la publication de données structurées, en reliant ces données entres elles pour constituer une gigantesque base de données (Bizer et Heat, 2001) et non sous la forme de silos de données isolés. Pour ce faire, plusieurs briques technologiques ont été mises en place.

# Un web de données liées (3)

Système d'identifiants fiables et pérennes: **URI**

Grammaire RDF et ses syntaxes **RDF/XML, N3, Turtle, N-triple, RDFa**

Protocole, langage de requête pour le RDF **SPARQL**

Ontologies exprimées **RDFs, OWL**

# Un web de données liées (4)

## Grammaire RDF (Ressource Description Format)

Modèle où toute connaissance respecte la forme d'un triplet selon la syntaxe suivante

http://dbpedia.org/ressource/Bruxelles

http://dbpedia.org/ontology/birthPlace

http://dbpedia.org/ressource/ClaudeLevyStrauss

# Un web de données liées (6)

*est* → sculpture

*date* → 1565-1585

*représente*

*s'appelle* → Marc Aurèle

*a sa biographie ici*

*est* → site web

Énonciation

Triplet RDF

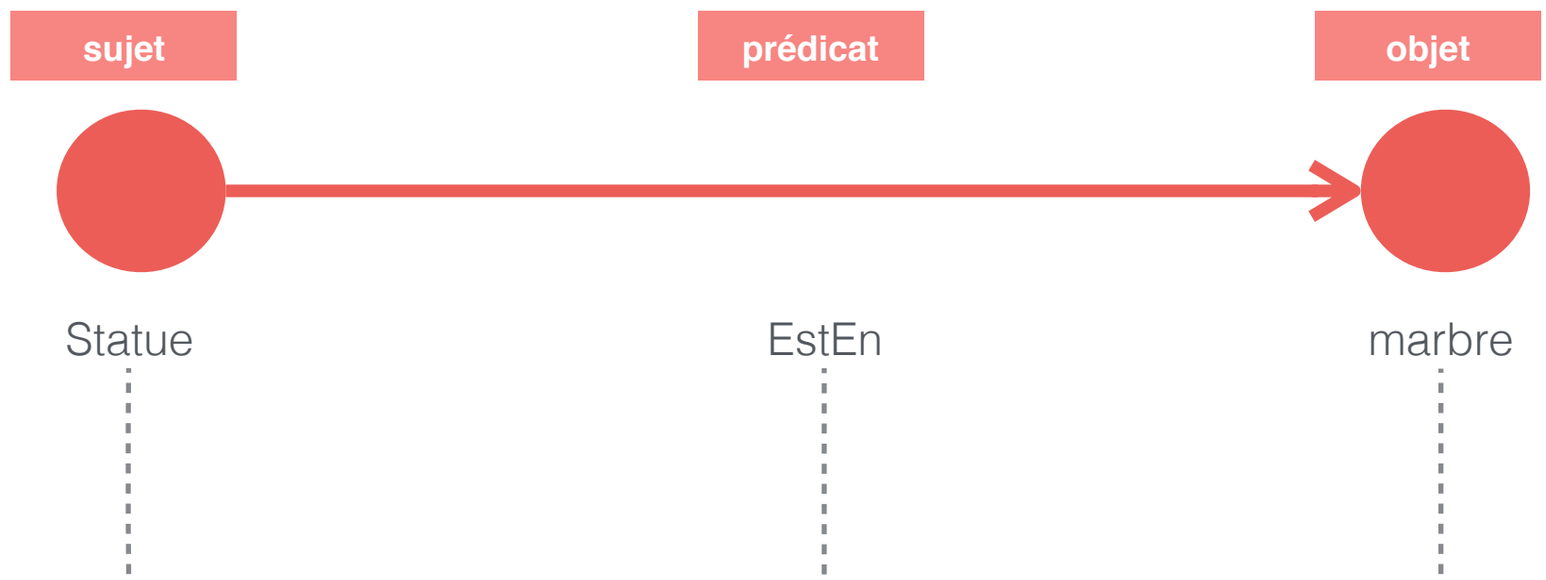
Graph

RDFs

CIDO-CCRM

Sérialisation RDF/XML

# “ La statue est en marbre ”




**E24_Physical_Man-Made_Think**
= *Statue*

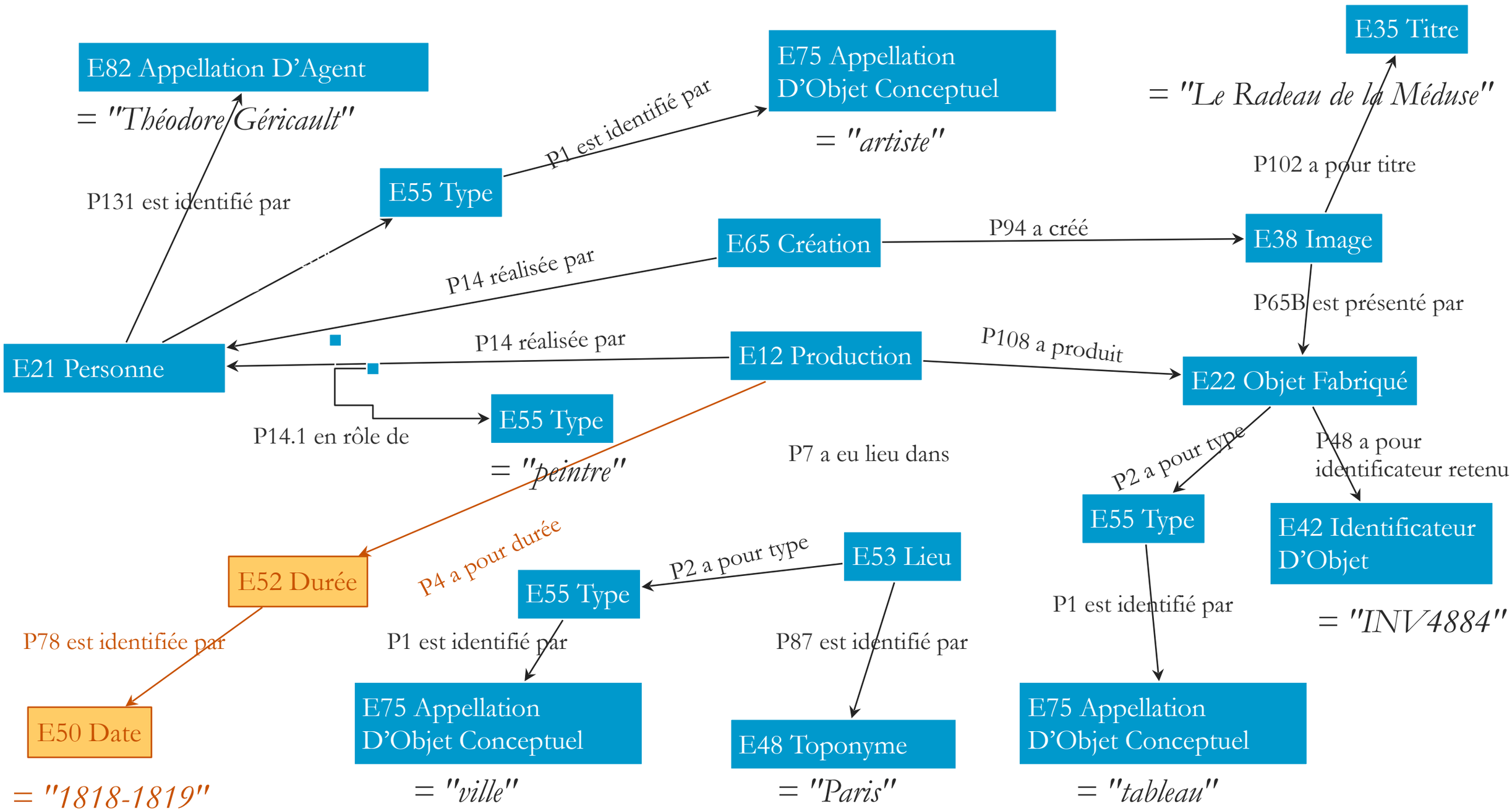
**P45_consists_of**

**E57_Material**
= *marbre*

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
    xmlns:bmo="http://collection.britishmuseum.org/id/ontology/"
    xmlns:crm="http://erlangen-crm.org/current/"
    xmlns:idThes="http://collection.britishmuseum.org/id/thesauri/">
    <crm:E24_Physical_Man-Made_Think rdf:about="http://collection.britishmuseum.org/id/object/EOC3130">
        <crm:P45_consists_of rdf:resource="http://collection.britishmuseum.org/id/thesauri/x11794"/>
    </crm:E24_Physical_Man-Made_Think>
</rdf:RDF>
```

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF>
<crm:E57_Material rdf:about="http://collection.britishmuseum.org/id/thesauri/x11794">
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:scopeNote>blabla</skos:scopeNote>
    <skos:inScheme rdf:resource="http://collection.britishmuseum.org/id/thesauri/material"/>
    <skos:prefLabel>marble</skos:prefLabel>
</crm:E57_Material>
</rdf:RDF>
```

- « L'artiste Théodore Géricault a peint son tableau *Le Radeau de la Méduse* (INV4884) dans la ville de Paris en 1818-1819. »

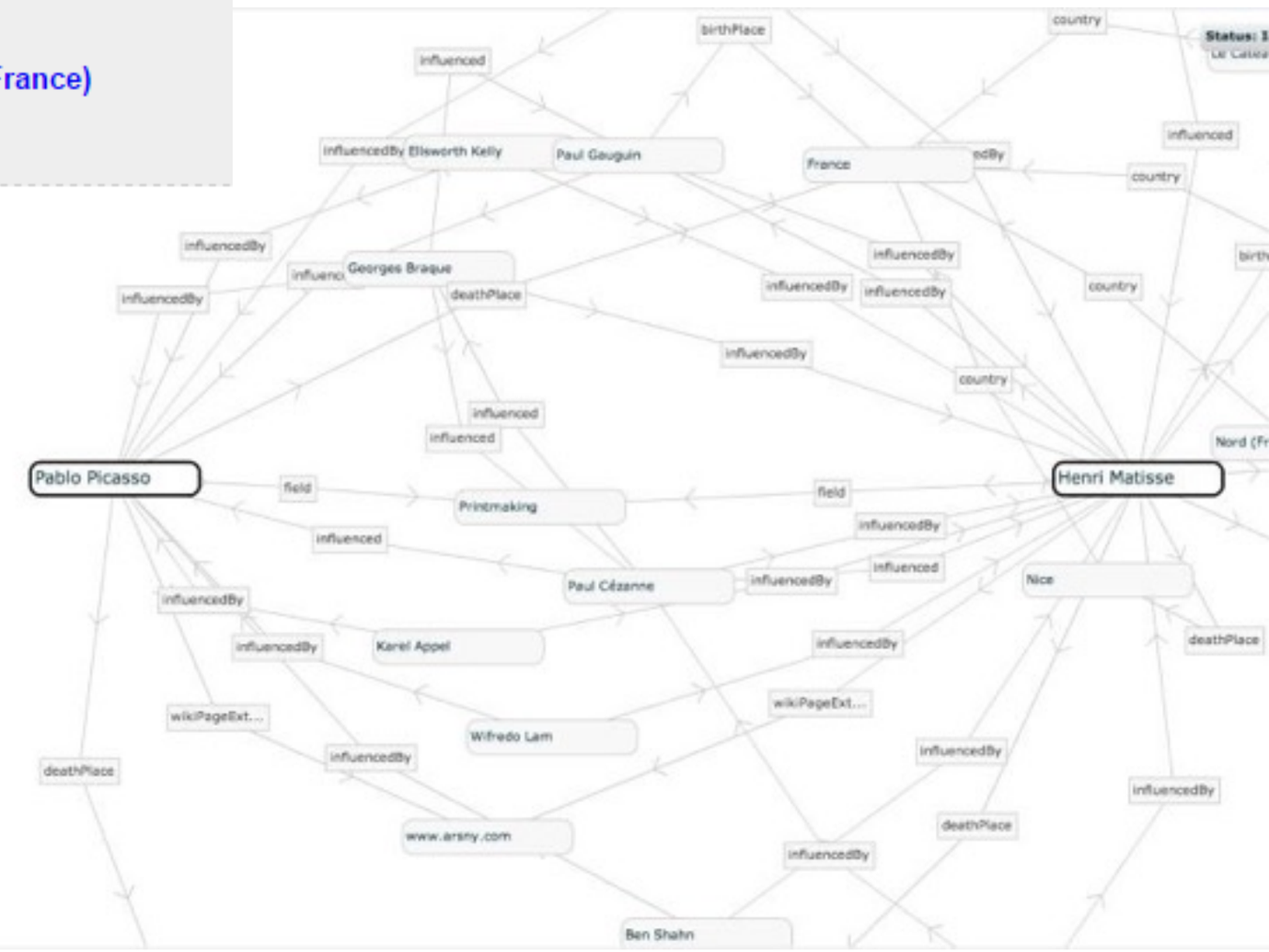


**Person 1** Picabia, Francis (ranskalainen taiteilija, 1879-1953)

**Person 2** Monet, Claude (ranskalainen taiteilija, 1840-1926)

**Picabia, Francis (French painter and writer, 1879-1953)**
*friend of*
**Duchamp, Marcel (French artist, 1887-1968)**
*member was*
**Society of Independent Artists (American and European artists' society, active 1916-1944)**
*member of*
**Sloan, John (American painter, printmaker, 1871-1951)**
*student of*
**Anshutz, Thomas Pollock (American painter, 1851-1912)**
*student of*
**Eakins, Thomas (American painter, photographer, and sculptor, 1844-1916)**
*student of*
**Bonnat, Léon (French painter, teacher, and collector, 1833-1922)**
*teacher of*
**Wentzel, Gustav (Norwegian painter, 1859-1927)**
*student of*
**Thaulow, Fritz (Norwegian painter and engraver, 1847-1906, active also in France)**
*friend of*
**Monet, Claude (French painter, 1840-1926)**

http://www.kulttuurisampo.fi/ff.shtml




http://www.visualdataweb.org/relfinder.ph

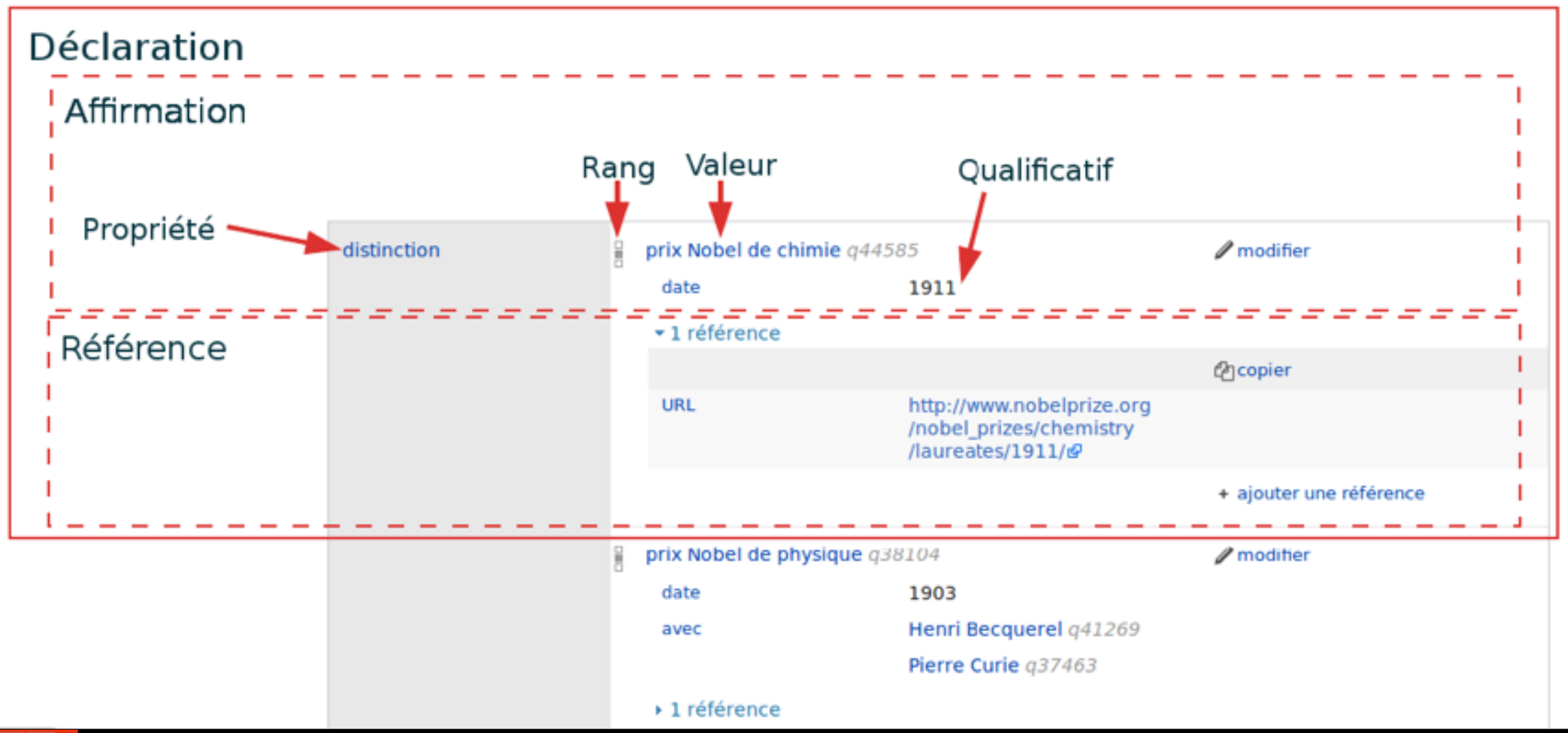
# Galaxie des projets Wikimedia

# Élément Wikidata

- Un identifiant de la forme **Qxxxxxxx**
- Un **libellé**, une **description** et des **alias** par langue
- Des **déclarations propriété/valeur**
- Des liens vers des **pages Wikimedia**
- **La base : propriété / valeur**

# Propriété Wikidata

- Un **identifiant** de la forme **Pxxxx**
- Une page dédiée : https://www.wikidata.org/wiki/Property:P1639
- Un seul **type de valeur**, strict à l'édition
- Des **contraintes**, souples à l'édition (voir la page de discussion de la propriété)
- Des propriétés qui **peuvent évoluer** à l'usage
- De **nouvelles propriétés** régulièrement créées



# A vous de jouer !

- Créer une requêtee Wikidata de votre choix
  - ex: les peintures de Monet et leurs lieu de conservation

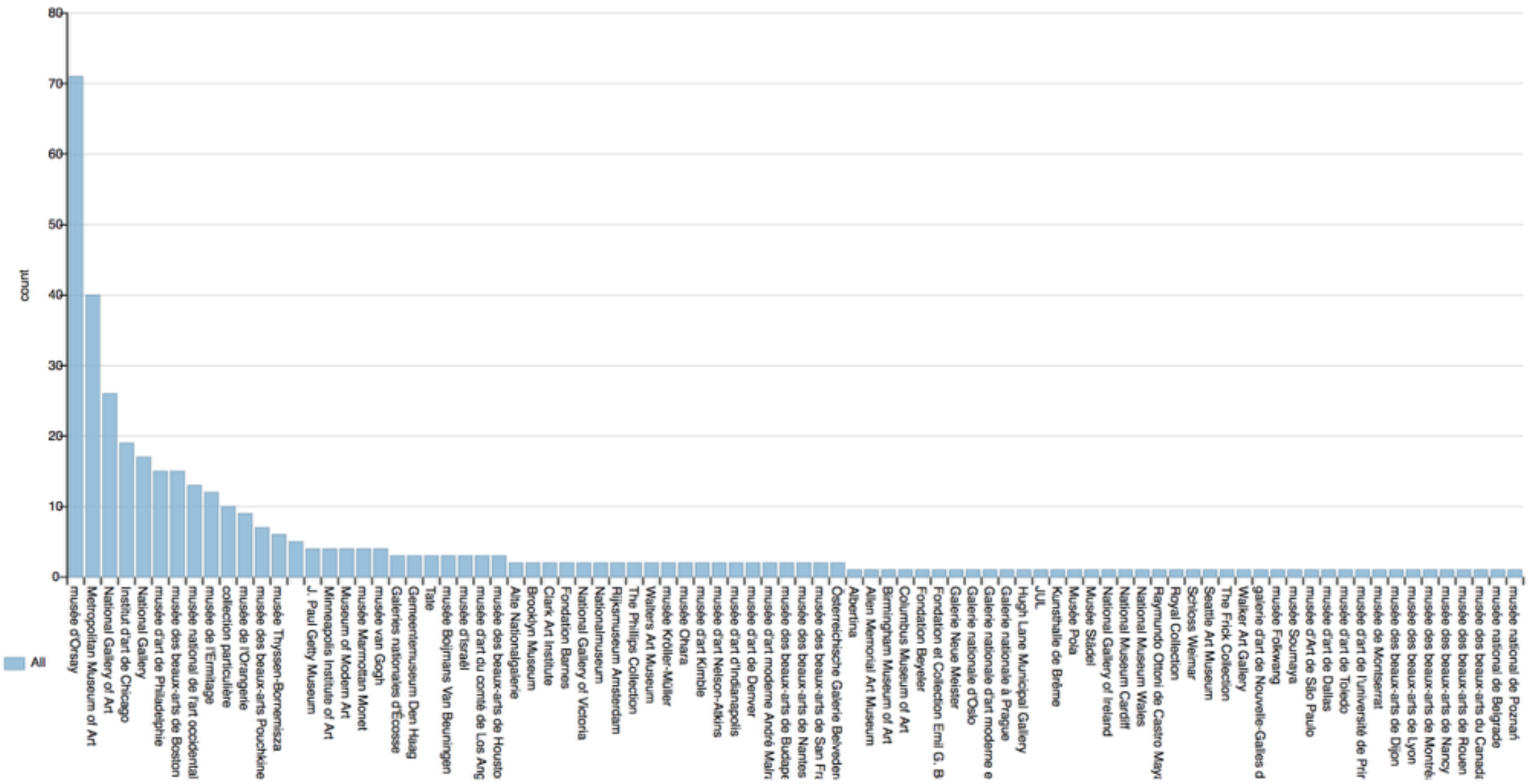
**TRUCS ET ASTUCES**

Le répertoire GitHub et le fichier

helpful_tips_wikidata.md

# exercice : les peintures de Monet et leurs lieu de conservation

## traduction perso : visualiser le nombre de Monet par lieu de conservation




https://gist.github.com/antoinecourtin/149401410f486372ffb1958ead9f0264

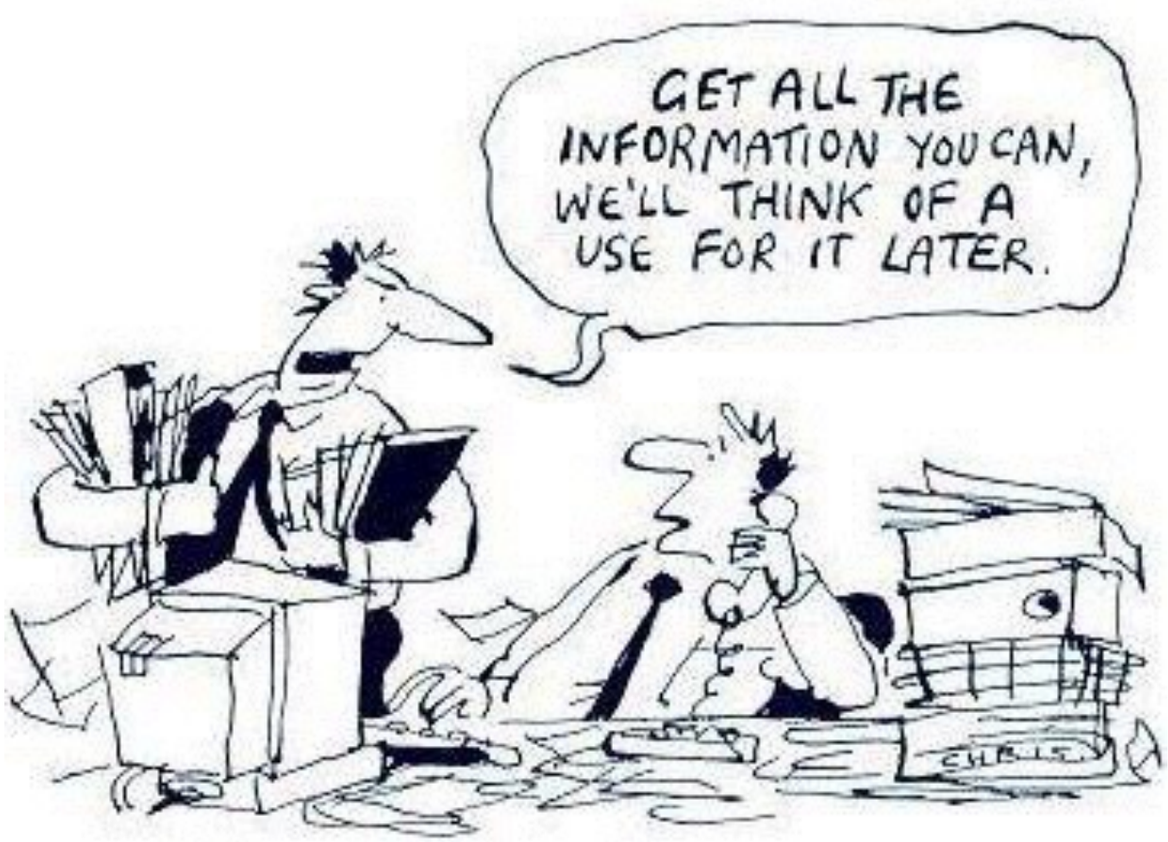
# Et Maintenant ?

Dirty Data – SparkPlugs, http://www.slideshare.net/SSMM_JoshMartin/dirty-data-sparkplugsslides

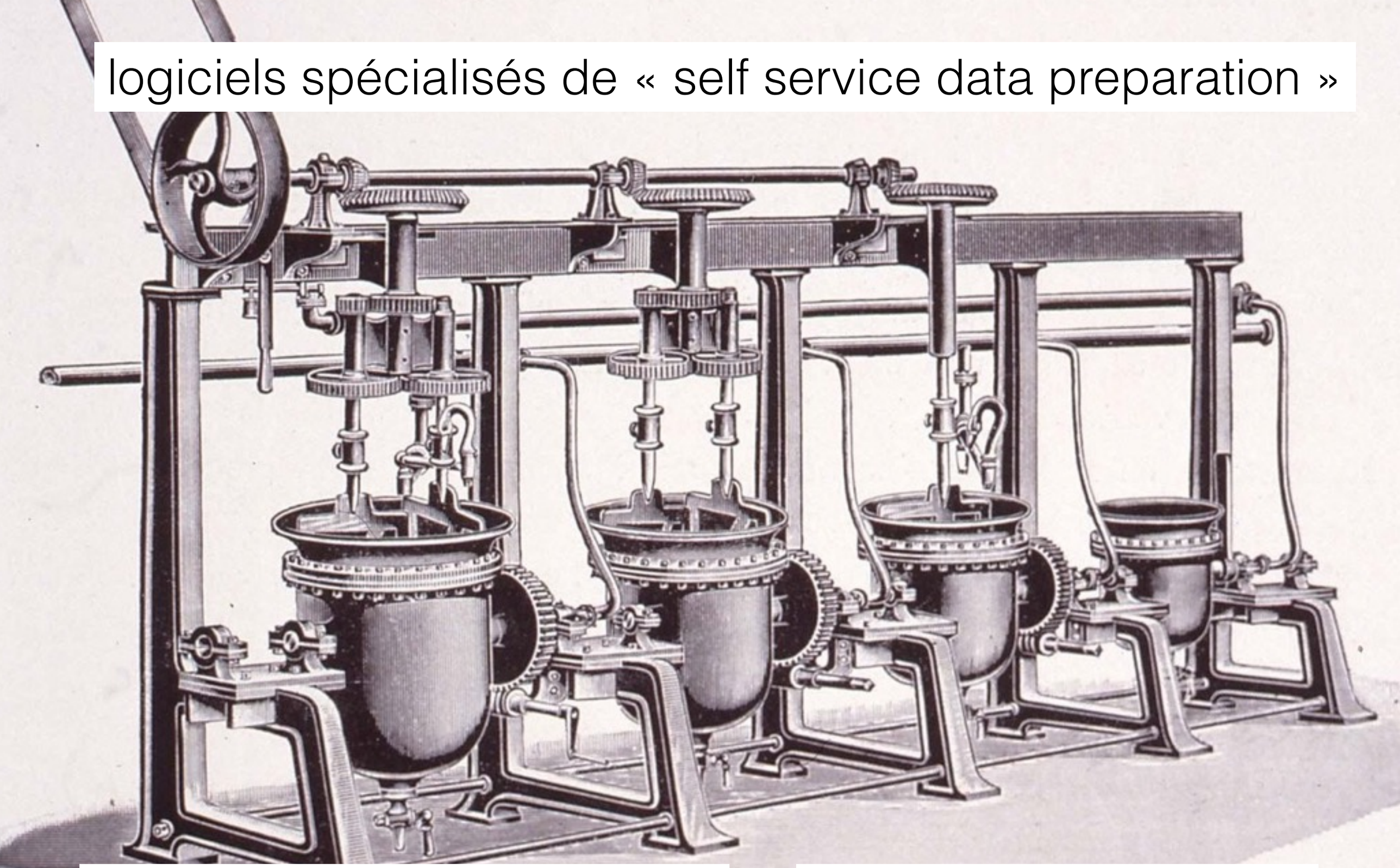
https://media.licdn.com/mpr/mpr/AAEAAQAAAAAAAALAAAAAJGU3YjAwY2I5LTg1ODMtNDg5Ny04MTg2LTNhMWUwYTdkZDk2Zg.pn

**Data wrangling** préparation préalable des données

Un principe essentiel régit les technologies de data wrangling : ce sont ceux qui connaissent le mieux les données qui doivent les étudier et les préparer



# logiciels spécialisés de « self service data preparation »



- Trifacta wrangler
- Open Refine (ex Google refine)
- Talend data preparation
- Dataiku DSS

**Acquisition des données**

- saisie des données
- collecte de données
- enrichissement des données tierces

**Structurer les données**

- nettoyer les données
- optimiser les formats
- compléter les données

**Visualiser l'information**

- choisir des informations à révéler
- tester les formes de visualisation

**Editorialiser la visualisation de données**

- apporter un discours
- légender

Du rêve

# Nettoyer et préparer des données

à la réalité

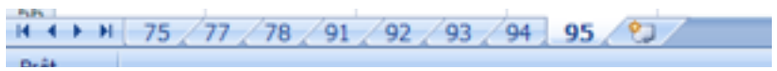
d'après une slide de Mathieu Saby.

PREVENTION
IS BETTER
THAN CURE

- Une feuille = un jeu de données
  - ➔ Un tableau par feuille

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | ID | Année | Budget | | Directions | 1er semestre 2012 | 2e trimestre 2012 |
| 2 | CP-13 | Année 2013 | 295 562 € | | Service A | 25 368 € | 16 357 € |
| 3 | CP-12 | Année 2012 | 183 687 € | | Service B | [illegible] | 19 963 € |
| 4 | CP-11 | Année 2011 | 255 665 € | | Service C | [illegible] | 8 350 € |
| 5 | CP-10 | Année 2010 | 199 355 € | | | | |
| 6 | CP-9 | Année 2009 | 222 887 € | | Directions | 1er [illegible]estre | [illegible]e trimestre 2011 |
| 7 | CP-8 | Année 2008 | 231 300 € | | Service A | [illegible] | 16 357 € |
| 8 | | | | | Service B | [illegible] | 19 963 € |
| 9 | | | | | Service C | 14 555 € | 8 350 € |
| 10 | | | | | | | |
| 11 | | | | | | | |

- 1 onglet = un jeu de données
  - ➔ Ou 1 jeu = fusion des onglets




  - ➔ Exemples
    - → Recensement des équipement sportifs = 1 fichier redécoupé en 8 jeux de données (1 par département)
    - → Domaines d'intérêt majeur (DIM) : équipements mi-lourds financés en 2012 = 1 jeu de données reprenant l'ensemble des onglets

- En-têtes sur la 1ère ligne (= titres de colonnes)

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Unité communication | | Service innovation numérique | | |
| 2 | | | | | |
| 3 | | ID | Année | Budget | |
| 4 | | CP-13 | Année 2013 | 295 562 € | |
| 5 | | CP-12 | Année 2012 | 183 687 € | |
| 6 | | CP-11 | Année 2011 | 255 665 € | |
| 7 | | CP-10 | Année 2010 | 199 355 € | |
| 8 | | CP-9 | Année 2009 | 222 887 € | |
| 9 | | CP-8 | Année 2008 | 231 300 € | |
| 10 | | | | | |
| 11 | | | | | |

- Pas de cellule vide dans les titres de colonnes

| | A | B | C | D |
|---|---|---|---|---|
| 1 | ID | Année | | Budget |
| 2 | CP-13 | Année 2013 | 25 368 € | 295 562 € |
| 3 | CP-12 | Année 2012 | 35 987 € | 183 687 € |
| 4 | CP-11 | Année 2011 | 14 555 € | 255 665 € |
| 5 | CP-10 | Année 2010 | 16 357 € | 199 355 € |
| 6 | CP-9 | Année 2009 | 19 963 € | 222 887 € |
| 7 | CP-8 | Année 2008 | 8 350 € | 231 300 € |

- Pas de cellule fusionnée (titres et contenu)

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Secteurs | Services | 1er semestre 2012 | |
| 2 | Secteur exemple 1 | Service A | 25 368 € | 16 357 € |
| 3 | | Service B | 35 987 € | 19 963 € |
| 4 | | Service C | 14 555 € | 8 350 € |
| 5 | | Service D | 14 490 € | 6 883 € |
| 6 | Secteur exemple 2 | Service E | 9 084 € | 2 880 € |
| 7 | | Service F | 3 677 € | 1 124 € |
| 8 | | Service G | 21 729 € | |
| 9 | | Service H | 7 136 € | 9 131 € |

- Attention aux lignes masquées !
  - ➔ elles s'affichent en CSV

- Éviter les lignes ou colonnes vides

- Attention aux données « orphelines » !

| | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | Directions | 1er semestre 2012 | 2e trimestre 2012 | | |
| 2 | Service A | 25 368 € | 16 357 € | | 26,03% |
| 3 | Service B | 35 987 € | 19 963 € | | 33,70% |
| 4 | Service C | 14 555 € | 8 350 € | | |
| 5 | | | | | |
| 6 | NB. Suppression le 3/03/12 du dispositif... | | | | |
| 7 | | | | | |

- Pas d'information transmise par la couleur

| | | |
|---|---|---|
| 7753301 | MUSEE DEPARTEMENTAL STEPHANE MALLARME | YVELINES-SUR-SEINE |
| 7817201 | MUSEE DE LA BATELLERIE | CONFLANS-STE- |
| 7832201 | MUSEE DE LA TOILE DE JOUY | JOUY EN JOSAS |
| 9406801 | MUSEE DE SAINT-MAUR - VILLA MEDICIS | LA VARENNE-SAINT- |
| 9407901 | MUSEE EMILE JEAN | VILLIERS SUR MARNE |
| 9408101 | MUSEE D'ART CONTEMPORAIN DU VAL-DE-MARNE | VITRY-SUR-SEINE |
| 9501801 | MUSEE D'ART ET D'HISTOIRE | ARGENTEUIL |
| 9520501 | MUSEE NATIONAL DE LA RENAISSANCE | ECOUEN |
| 9529501 | MUSEE ARCHEOLOGIQUE DU VAL D'OISE | GUIRY EN VEXIN |
| 9531301 | MUSEE LOUIS SENLECQ | L'ISLE ADAM |
| 9535101 | MUSEE INTERCOMMUNAL D'HISTOIRE ET D'ARCHEOLOG | LOUVRES |
| 9542801 | MUSEE JEAN-JACQUES ROUSSEAU | MONTMORENCY |

http://www.data.gouv.fr/DataSet/30382387?xtmc=frequentation+des+musees+de+france&xtcr=2

Galeries nationales du Grand Palais

Musées nationaux
Musées de la ville de Paris - appellation en 2004
Musées du Muséum National d'Histoire Naturelle
Appellation M de F en 2004
Appellation M de F en 2006
Appellation M de F en 2007
Appellation M de F en 2009

Données confidentielles - Contacter le chef d'établissemen
Fermeture du musée Picasso en septembre 2009 pour travaux pour
Chiffres d'entrées pour les expositions temporaires
Fermeture du musée le 10 janvier 2010 pour travaux. Réouverture p
Fermeture du musée de l'Homme (muséum) en mars 2009 pour trava
Collections permanentes gratuites depuis le 11 septembre 2007 à ca

➔ Dans le format CSV, ces données sont supprimées !

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 7510103 | MUSEE DE L'ORANGERIE DES T... | PARIS | Musée national | 447 093 | 153 417 | 599 762 | 140 729 |
| 7510106 | MUSEE DU LOUVRE | PARIS | Musée national | 8 314 000 | 2 647 000 | 8 224 843 | 2 703 40 |
| 7510602 | MUSEE EUGENE DELACROIX | PARIS | Musée national | 34 044 | 9 890 | 38 425 | 9 916 |
| 7510305 | MUSEE NATIONAL PICASSO | PARIS | Musée national | 501 060 | 210 779 | 470 500 | 151 867 |
| 7510501 | MUSEE NATIONAL DU MOYEN-AG... | PARIS | Musée national | 299 958 | 124 049 | 293 975 | 129 570 |

# Quelques exemples

## Modifier l'organisation des données

| *Nom* | *Dupont* | *Martin* | *Durand* |
|---|---|---|---|
| Code postal | 06100 | 18000 | 06000 |
| Ville | Nice | Bourges | Nice |

→

| *Nom* | *Code postal* | *Ville* |
|---|---|---|
| Dupont | 06100 | Nice |
| Martin | 18000 | Bourges |
| Durand | 06000 | Nice |

## Harmoniser des valeurs textuelles

Nice
France, Nice
Nice
NICE
[Nice, Fr]

→ **Nice (France)**

## Corriger des valeurs textuelles

Nietzsche
Nietsche
Nieztsche
Nietzshe

→ **Nietzsche**

## Harmoniser des dates

15/5/2017
5/15/2017
2017-05-15
15 mai 2017
May 15th 2017

→ **2017-05-15**

## Extraire des valeurs

Né en 1960
Born on 1978
1958
Naissance: 1985

→ **1960**
**1978**
**1958**
**1985**

d'après une slide de Mathieu Saby.

# Quelques exemples

| *Nom* | *Code postal* |
|---|---|
| Dupont | 06100 |
| Martin | 18000 |
| David | |
| Durand | 06000 |

+

| *Code postal* | *Ville* |
|---|---|
| 06100 | Nice |
| 06000 | Nice |
| 18000 | Bourges |
| 34000 | Montpellier |

→

| *Nom* | *Code postal* | *Ville* |
|---|---|---|
| Dupont | 06100 | Nice |
| Martin | 18000 | Bourges |
| David | | |
| Durand | 06000 | Nice |

| *Nom* | *Age* | *Profession* |
|---|---|---|
| Dupont | **80** | **Lycéen** |
| Martin | **18** | **Retraité** |
| David | 18 | Lycéen |
| Durand | 65 | Retraité |

→

Actions possibles:
- supprimer la ligne
- remplacer par « N/A »
- chercher à corriger

d'après une slide de Mathieu Saby.

# OPEN Refine

# openRefine

Site web: http://openrefine.org/

Ressources: https://github.com/OpenRefine/OpenRefine/wiki/External-Resources

Pour faire quoi ?

- Acquérir des données
- Nettoyer des données
- Lier des données

Les avantages

- Prise en main aisée et courbe de progression rapide
- Importante communauté open source
- Scalabilité
- Pour de nombreux profils différents
- Beaucoup de format en entrée et en sortie
- Permet de créer des process réutilisables

## Installation d'OpenRefine :

http://openrefine.org/download.html

L'outil OpenRefine est historiquement lié à l'existence de Freebase, l'un des tous premiers projets collaboratifs d'entrepôts de données sémantiques sur le web

**Historique :**

2007 mars : Metaweb lance Freebase
2010 : Freebase est proposée avec ***l'outil de nettoyage et de traitement des données Gridworks*** (v1.0)
2010 juillet : Google rachète Metaweb
2010 nov : Gridworks devient Google Refine
2012 oct : Arrêt de Google Refine dont le code est libéré et qui devient **OpenRefine**
————

*[*
*2014 déc : Google annonce la fermeture de Freebase*
*2015:Fermeture progressive de Freebase et des services associés*
*2016 mai: Fermeture définitive de Freebase*
*]*

Versions
Dernière version bêta : 2.7rc2 (2017). **À installer de préférence à la 2.5**
Dernière version officielle : 2.5 (2011, développée par Google)

# Aspect technique & Installation

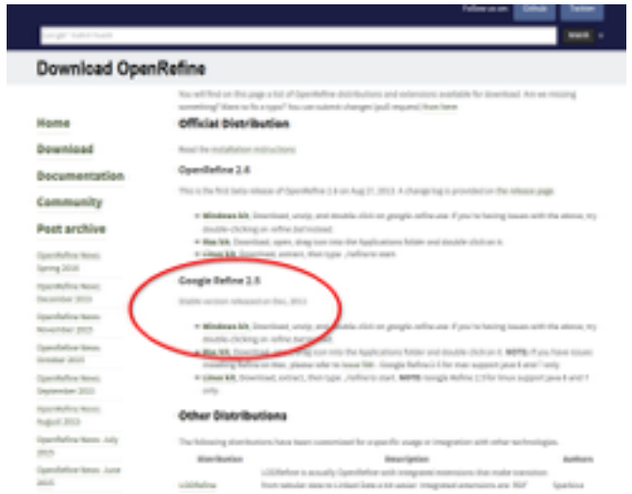
- Écrit en langage Java (peut compliquer l'installation sous Mac)
- Installation mono-poste, sous PC, Mac et Linux
- Interface accessible via un navigateur internet (adresse http://127.0.0.1:3333/ )
- Les données et le logiciel restent sur le PC (pas besoin de connexion Internet)

## Les plugins/extensions les + utiles

- Exporter les données en RDF (à partir d'un squelette issue d'une ontologie par ex.)
  - https://github.com/fadmaa/grefine-rdf-extension/releases
  - (ATTENTION: renommer le dossier dezippé en « rdf-extension » avant de le copier dans le répertoire / webapp/extensions/)
- Extraction d'entités nommées
  - https://github.com/RubenVerborgh/Refine-NER-Extension
- VIB-BITS: 3 plugins (gestions des actions/historique; comparaison entre 2 textes; gestions des facettes)
  - https://www.bits.vib.be/index.php/software-overview/openrefine
- Réaliser des statistiques
  - https://github.com/sparkica/refine-stats

## Installer les plugins

- Télécharger l'extension souhaitée
- Dézipper le dossier
- Quitter OpenRefine s'il est ouvert
- Dans le répertoire d'OpenRefine
  - /.local/share/google/refine/
  - Créer un sous-dossier « extension s»
  - Placer le dossier dézippé de l'extension dans le répertoire courant
- Relancer OpenRefine



A quoi ça ressemble ?

# Interface générale

Facettes et filtres

Historique

Lien vers le projet

Contenu du fichier
Système par colonnes

Nouveau projet

Export

Refine OPEN doaj article sample Permalien

Ouvrir... Exporter Aide

Facette / Filtre | Défaire / Refaire 0

**Utiliser les facettes et les filtres**

Utiliser les facettes et les filtres pour sélectionner les sous-ensembles de données à traiter. Choisir les méthodes de facette et de filtre dans les menus situés dans les entêtes de colonne.

Vous ne savez pas par où commencer ?
**Regarder ces tutoriels vidéos**

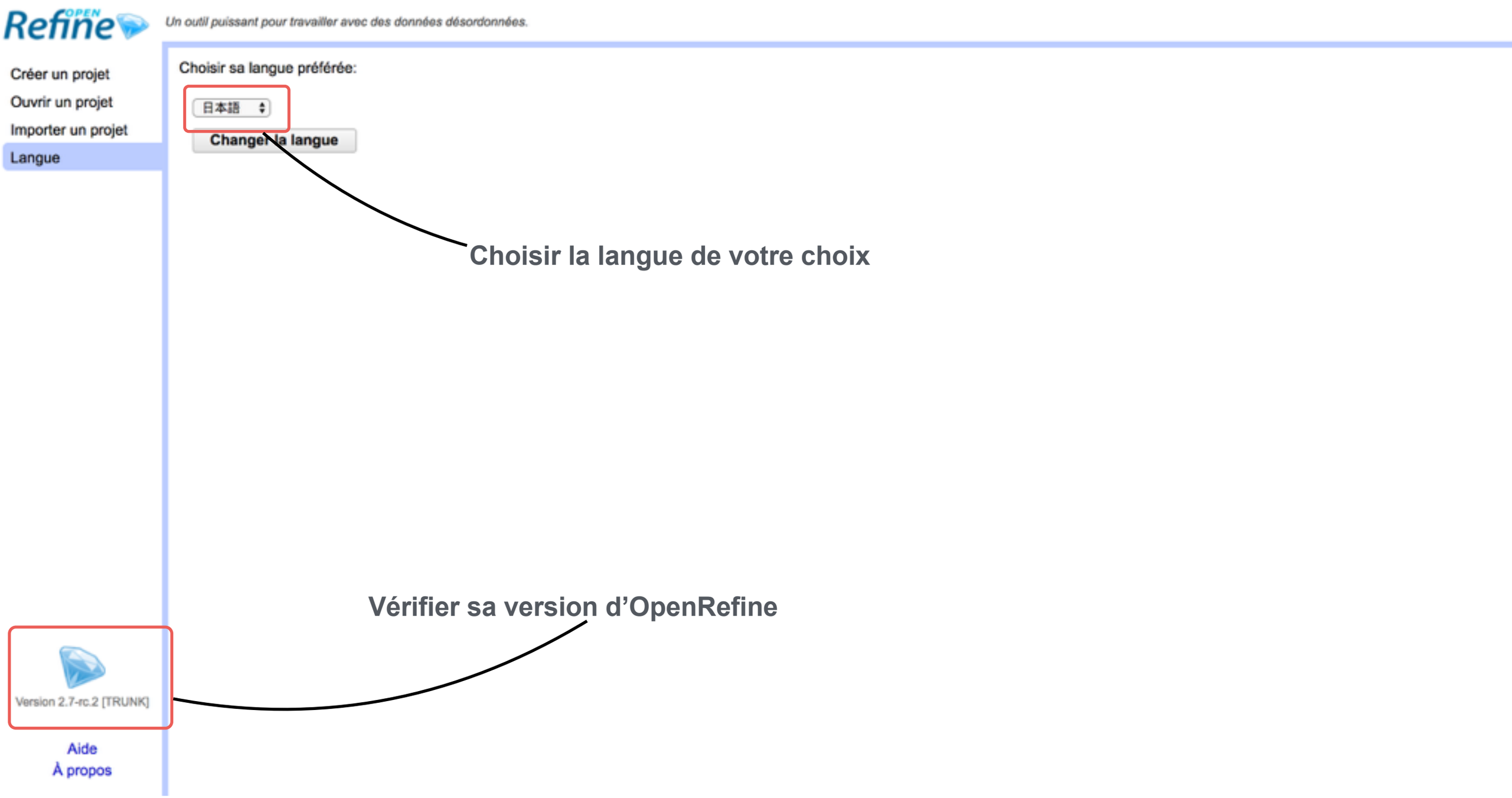
**1001 lignes** Extensions:

Voir en: **lignes** entrées Afficher: 5 **10** 25 50 lignes « première ‹ précédente **1 - 10** suivante › dernière »

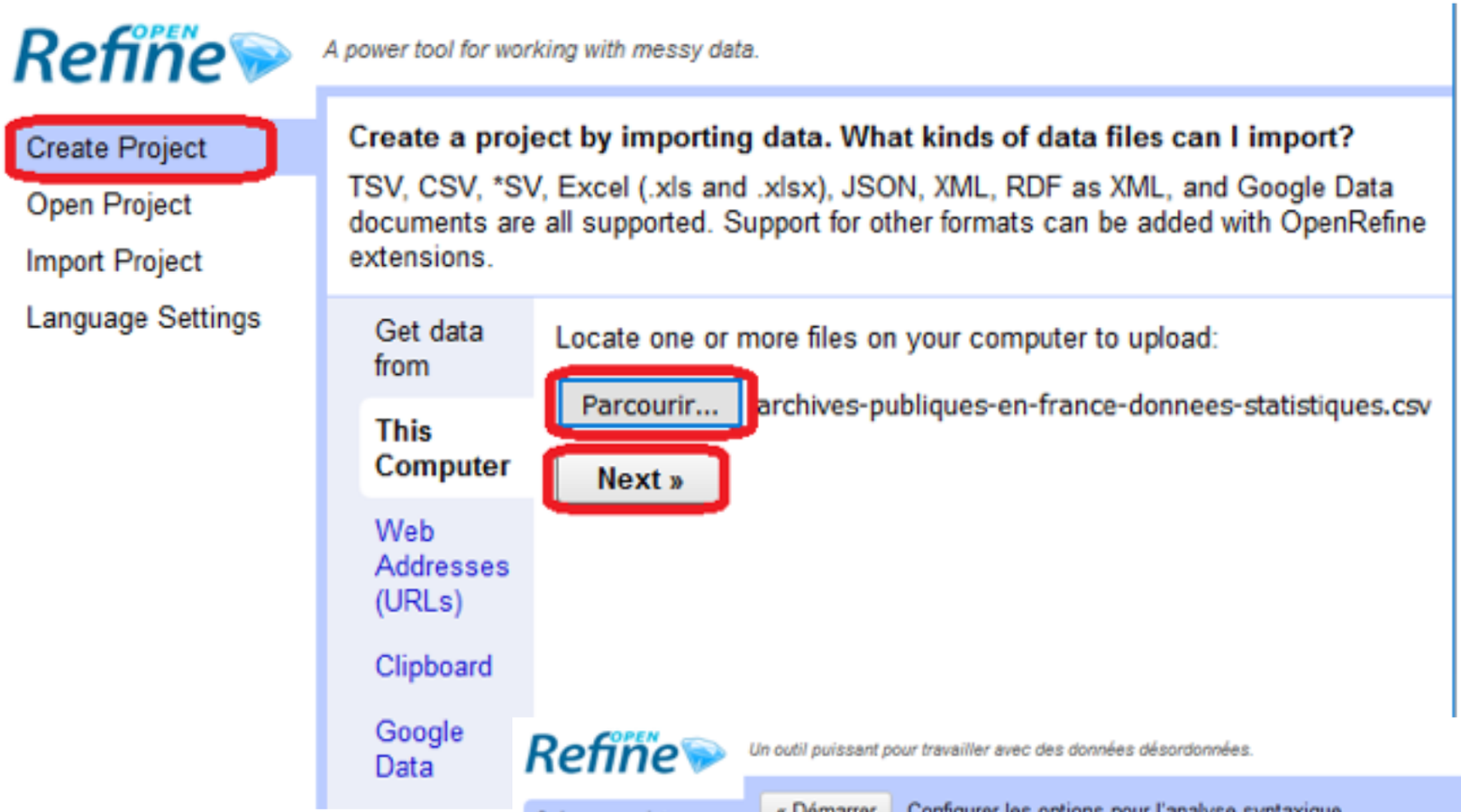
| Toutes | Title | Authors | DOI | URL | Date | Language | Subjects |
|---|---|---|---|---|---|---|---|
| 1. | The Fisher Thermodynamics of Quasi-Probabilities | Flavia Pennini\|Angelo Plastino | 10.3390/e17127853 | https://doaj.org/article/b75e8d5cca3f46cbbd63e91be5b32412 | 01/11/2015 | English | Fisher information\|quasi-probabilities\|complementarity\|Physics\|QC |
| 2. | Aflatoxin Contamination of the Milk Supply: A Pakistan Perspective | Naveed Aslam\|Peter C. Wynn | 10.3390/agriculture5041172 | https://doaj.org/article/0edc5af6672641c0bd45608812a34f9e | 01/11/2015 | English | aflatoxins\|AFM1\|AFB1\|milk marketing cha carcinoma\|Agriculture (General)\|S1-972\| |
| 3. | Metagenomic Analysis of Upwelling-Affected Brazilian Coastal Seawater Reveals Sequence Domains of Type I PKS and Modular NRPS | Rafael R. C. Cuadrat\|Juliano C. Cury\|Alberto M. R. DÃ¡vila | 10.3390/ijms161226101 | https://doaj.org/article/d9fe469f75a0442382b84ba4f50007ee | 01/11/2015 | English | PKS\|NRPS\|metagenomics\|environmental genomics\|upwelling\|coastal environment\|Chemistry\|QD1-999\|Science |
| 4. | Synthesis and Reactivity of a Cerium(III) Scorpionate Complex Containing a Redox Non-Innocent 2,2â²-Bipyridine Ligand | Fabrizio Ortu\|Hao Zhu\|Marie-Emmanuelle Boulon\|David P. Mills | 10.3390/inorganics3040534 | https://doaj.org/article/95606ed39deb4f43b96f7e6308ad15d3 | 01/11/2015 | EN | lanthanide\|cerium\|scorpionate\|tris(pyrazo non-innocent\|Inorganic chemistry\|QD146 |
| 5. | Performance and Uncertainty Evaluation of Snow Models on Snowmelt Flow Simulations over a Nordic Catchment (Mistassibi, Canada) | Magali Troin\|Richard Arsenault\|FranÃ§ois Brissette | 10.3390/hydrology2040289 | https://doaj.org/article/18b1d70730d44573ab5c264b7f93e041 | 01/11/2015 | EN | snow models\|hydrological models\|snowmelt\|uncertainty\|Canada\|Sci |
| 6. | Dihydrochalcone Compounds Isolated from Crabapple Leaves Showed Anticancer Effects on Human Cancer Cell Lines | Xiaoxiao Qin\|Yun Feng Xing\|Zhiqin Zhou\|Yuncong Yao | 10.3390/molecules201219754 | https://doaj.org/article/5765b418183c4b70bb0b774f4627beb5 | 01/11/2015 | English | Malus crabapples\|leaves\|dihydrochalcon compounds\|anticancer activity\|human ca chemistry\|QD241-441\|Chemistry\|QD1-99 |
| 7. | Ionic Liquids as Carbene Catalyst Precursors in the | Anton Axelsson\|Linda Ta\|Henrik SundÃ©n | 10.3390/catal5042052 | https://doaj.org/article/d1d39464834447c8bd9c269710e1e17d | 01/11/2015 | English | ionic liquid\|NHC\|OTHO\|multicomponent reaction\|carbene\|organocatalysis\|organo ketoester\|Physical and theoretical |

d'après une slide de Mathieu Saby.

# 你喜欢什么样的语言？


Refine
Un outil puissant pour travailler avec des données désordonnées.
Créer un projet
Ouvrir un projet
Importer un projet
Langue
Choisir sa langue préférée:
日本語
Changer la langue
Choisir la langue de votre choix
Vérifier sa version d'OpenRefine
Version 2.7-rc.2 [TRUNK]
Aide
À propos

# Les modes d'imports



# Les filtres

Facette / Filtre | Défaire / Refaire 0

Rafraîchir | Tout réinitialiser | Tout supprimer

**Titre**

octroi

☐ sensible à la casse ☐ expression rationnelle

**970 matching lignes** (39703 total)

Voir en: **lignes** entrées  Afficher: 5 10 25 **50** lignes

| Toutes | N_Ordre | Date_Delib | Titre |
|---|---|---|---|
| 3. | 1909/40 | 2009-01-14 | apports. Remises aux employés d'octroi. Crédit p. 40 |
| 7. | 1909/57 | 2009-01-14 | apports. Réduction de la taxe d'octroi sur la glace à rafraîchir p. 44 |
| 11. | 1909/95 | [illegible] | tions. Service de l'octroi. Achat d'une machine à écrire et d'un duplicateur p |
| 14. | 1909/110 | 2009-02-22 | tions. Liquidation de la pension de retraite Marignères (octroi) p. 67 |
| 15. | 1909/112 | 2009-02-22 | tions. Liquidation de la pension de retraite Auzeing (octroi) p. 67 |
| 49. | 1909/357 | 2009-05-04 | troductions. Pension de retraite revenant à Marie Cazaux veuve de M Paiss |
| 62. | 1909/413 | 2009-05-19 | troductions. Pension de retraite veuve Audoubert (octroi) p. 175 |
| 63. | 1909/419 | 2009-05-19 | troductions. Pension de retraite veuve Cep (octroi) p. 175 |
| 92. | 1909/532 | 2009-06-18 | troductions. Pension de retraite veuve Bonnefoy (octroi) p. 213 |
| 105. | 1924/595 | 2024-11-29 | ent sur la caisse des retraites des employés d'octroi. Modifications p. 385 |
| 139. | 1925/92 | 2025-12-03 | Session extraordinaire - Liquidation de la pension de retraite revenant à Marie Noël en qualité de veu Toulouse p. 64 |

Facette
Filtrer le texte
Éditer les cellules
Éditer la colonne
Transposer
Trier...
Aperçu
Réconcilier

Refine OPEN — AM Toulouse délib  Permalien

Facette / Filtre | Défaire / Refaire 5

Rafraîchir | Tout réinitialiser | Tout supprimer

**Date_Delib** changer

742 choices Trier par: **nom** compte  Groupe

1965-03-20 29
1965-03-23 1
1965-06-28 138
1965-10-05 185
1965-10-25 161
1966-01-24 126
1966-02-28 151
1966-05-12 181
1966-06-06 119
1966-07-19 1
1966-09-06 1

**39703 lignes**

Voir en: **lignes** entrées  Afficher: 5 **10** 25 50 lignes  « première  ‹ précédent

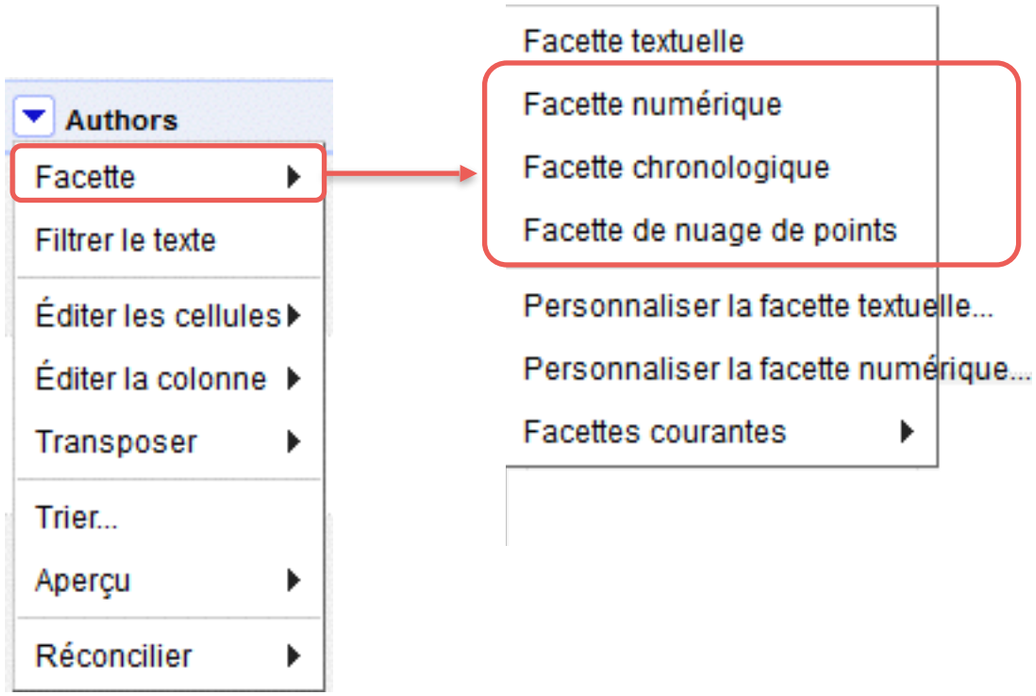
| Toutes | N_Ordre | Date_Delib | Titre | Titre 1 |
|---|---|---|---|---|
| 1. | 1909/31 | | | |
| 2. | 1909/34 | | | |
| 3. | 1909/40 | | | |
| 4. | 1909/45 | | | |
| 5. | 1909/47 | 2009-01-14 | Session extraordinaire. Rapports. Liquidation de la pension de retraite revenant à M Cuénin Louis brigadier de police p. 41 | Session extra Rapports. Liq de la pension retraite revena Cuénin Louis de police |

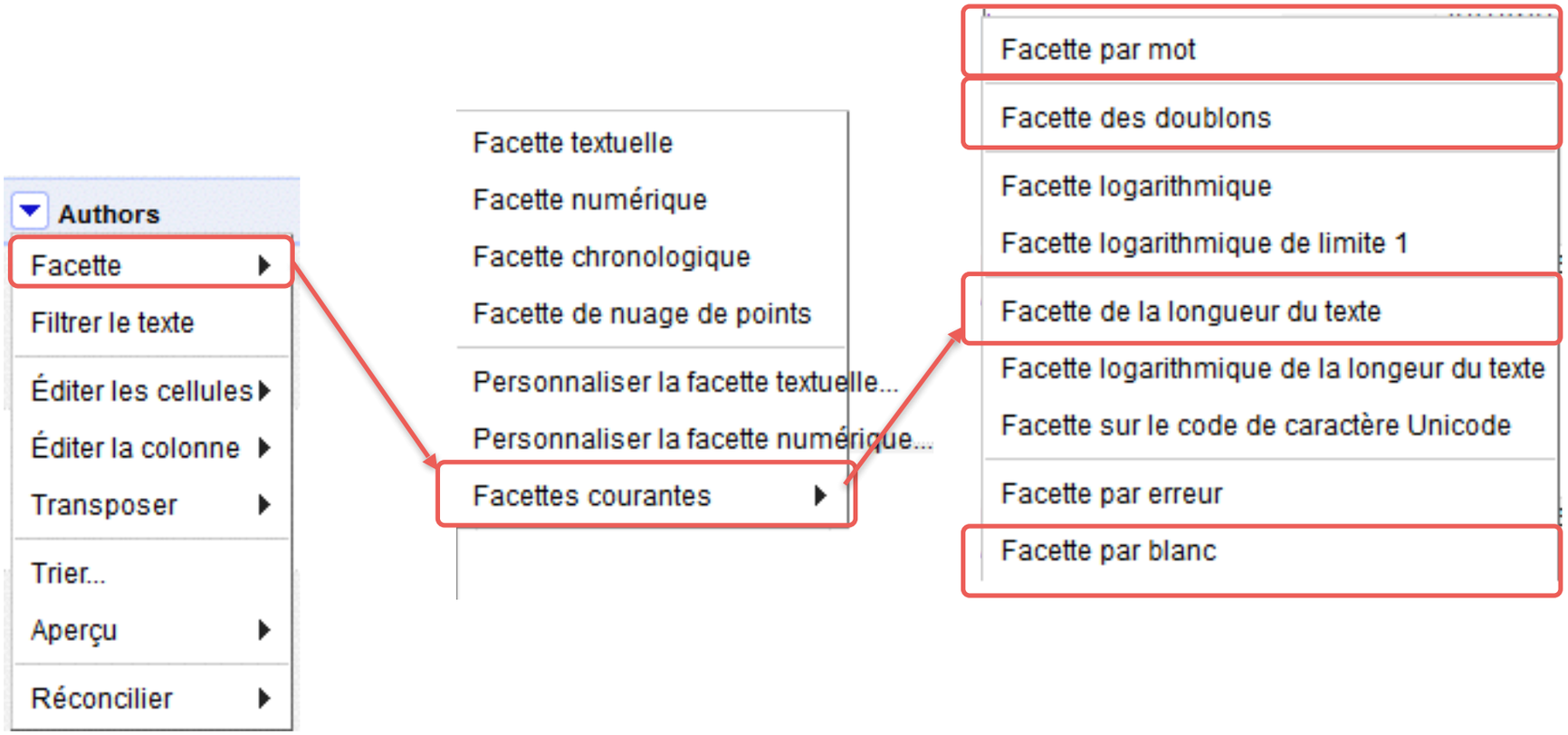
Facette
Filtrer le texte
Éditer les cellules
Éditer la colonne
Transposer
Trier...
Aperçu
Réconcilier

Facette textuelle
Facette numérique
Facette chronologique
Facette de nuage de points
Personnaliser la facette textuelle...
Personnaliser la facette numérique
Facettes courantes

d'après des captures de Maïwenn Bourdic.

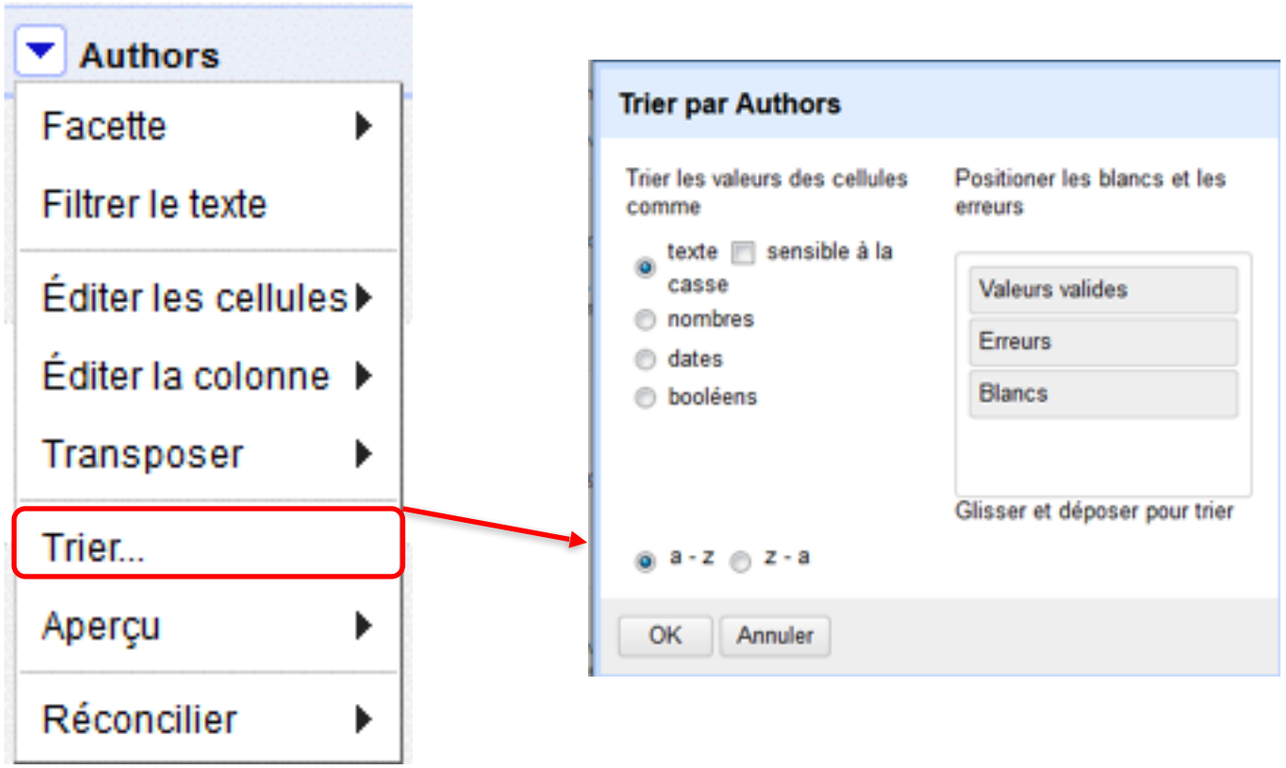
# Les filtres



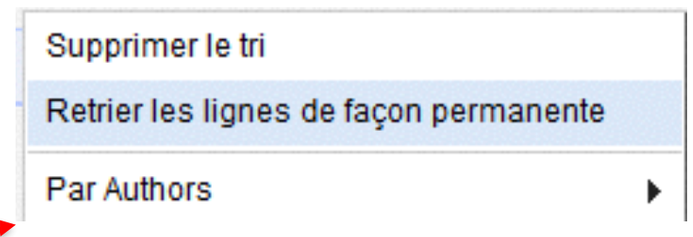



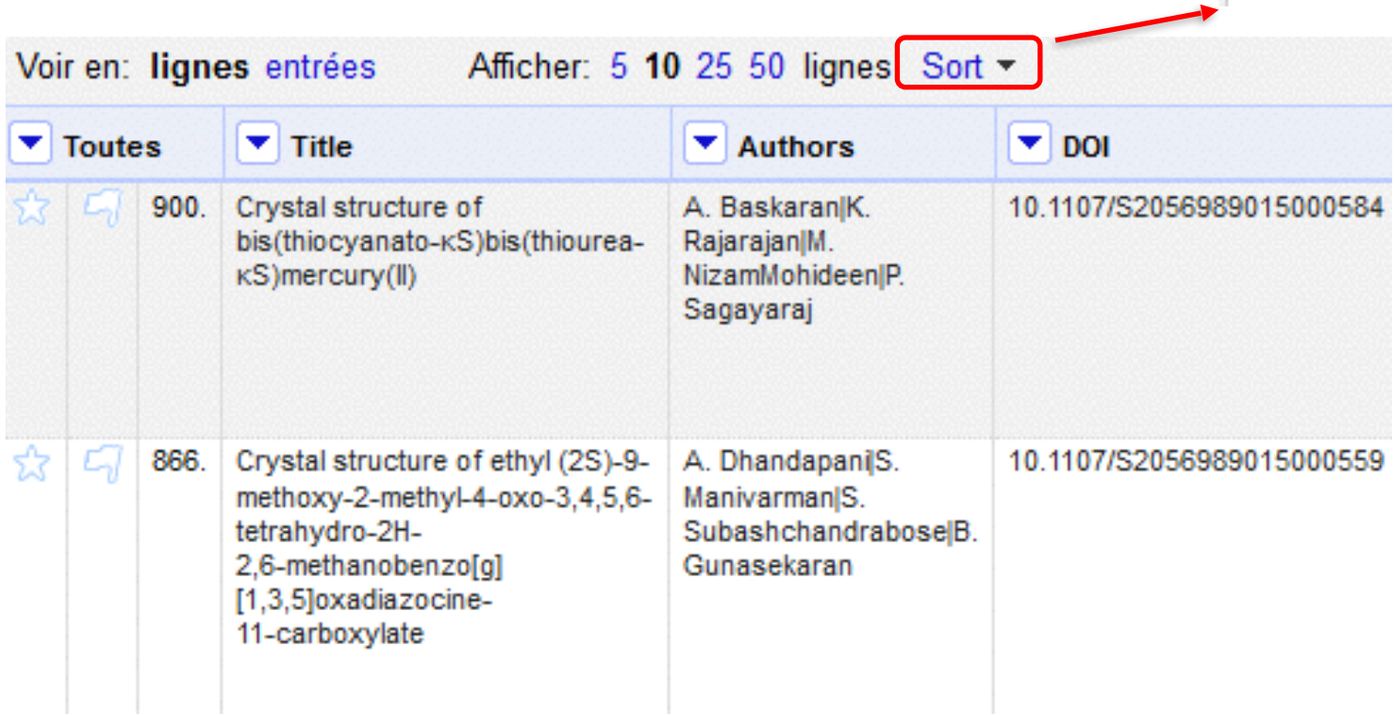

d'après des captures de Maïwenn Bourdic.

# Les tris




Nouveau menu



# Cluster & merging

Facette / Filtre | Défaire / Refaire 0

Rafraîchir | Tout réinitialiser | Tout supprimer

**Anciennes régions** changer

34 choices Trier par: **nom** compte | Groupe

- Bretagne 3066
- Centre 2786
- Champagne-Ardenne 1446
- Corse 304
- Franche-comté 1 éditer include
- Franche-Comté 1345
- Guadeloupe 104
- Guyane 83
- Haute-Normandie 1155
- haute-Normandie 1
- Ile-de-France 3842
- Languedoc-Roussillon 2071
- Limousin 1024
- Lorraine 1657
- Martinique 101

Cet outil vous aide à identifier des groupes de cellules ayant des valeurs différentes mais qui peuvent correspondre à des représentations alternatives de la même valeur. Par exemple, les deux chaînes "New York" et "new york" n'ont qu'une différence de casse et font très certainement référence à la même ville. "Gödel" et "Godel" se réfèrent probablement à la même personne. En trouver davantage...

Méthode collision de clés | Fonction de codage empreinte | **300** clusters trouvé

| Taille du groupe | Nombre de lignes | Valeurs dans le groupe | Fusionner ? | Nouvelle valeur pour la cellule |
|---|---|---|---|---|
| 6 | 25 | • Ancien évêché (14 rows)<br>• Evêché (ancien) (5 rows)<br>• Ancien Evêché (3 rows)<br>• Ancien évêché (1 rows)<br>• Evêché (ancien) (1 rows)<br>• ancien évêché (1 rows) | ☐ | Ancien évêché |
| 5 | 8 | • Ancien grand séminaire (4 rows)<br>• Ancien Grand Séminaire (1 rows)<br>• Ancien Grand séminaire (1 rows)<br>• Ancien grand Séminaire (1 rows)<br>• Séminaire (ancien grand) (1 rows) | ☐ | Ancien grand séminaire |
| 5 | 14 | • Ancienne porte de ville (8 rows)<br>• Porte de Ville (ancienne) (3 rows)<br>• Ancienne Porte de ville (1 rows)<br>• Ancienne porte de Ville (1 rows)<br>• Porte de ville (ancienne) (1 rows) | ☐ | Ancienne porte de ville |
| 4 | 11 | • Vieux château (6 rows)<br>• Château (vieux) (3 rows)<br>• Château Vieux (1 rows) | ☐ | Vieux château |

**# Choix dans le groupe**

2 — 6

**# Lignes dans le groupe**

0 — 1200

**Longueur moyenne des choix**

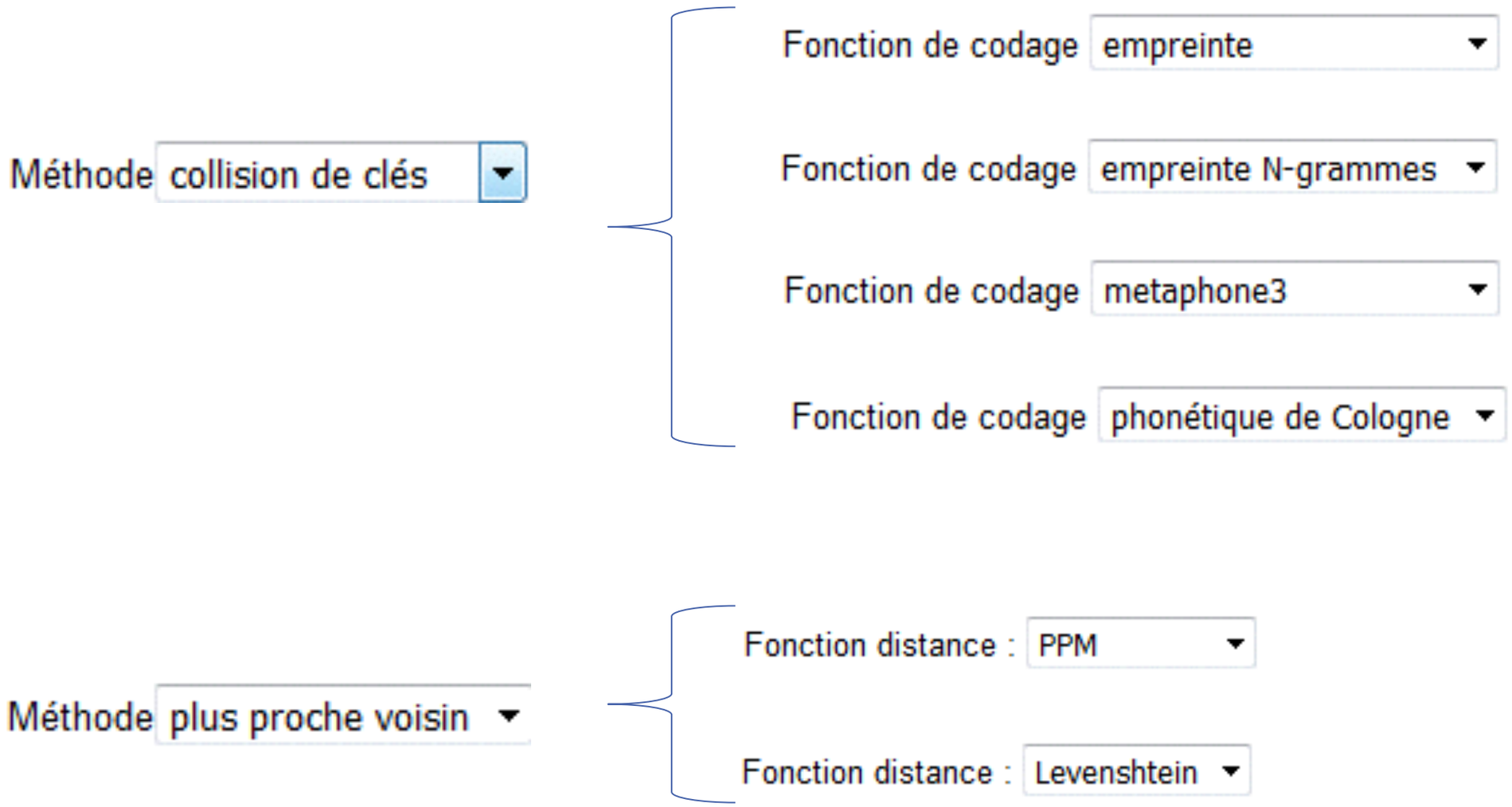
6 — 87

**Variabilité moyenne des choix**

0 — 3.56

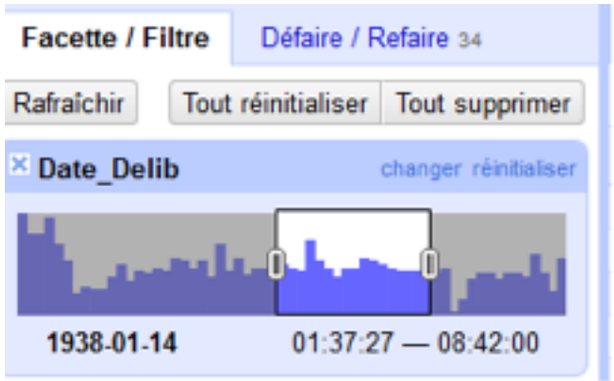
Tout sélectionner | Tout déselectionner | Exporter les groupes | **Fusionner la sélection & regrouper** | Fusionner la sélection & fermer | Fermer

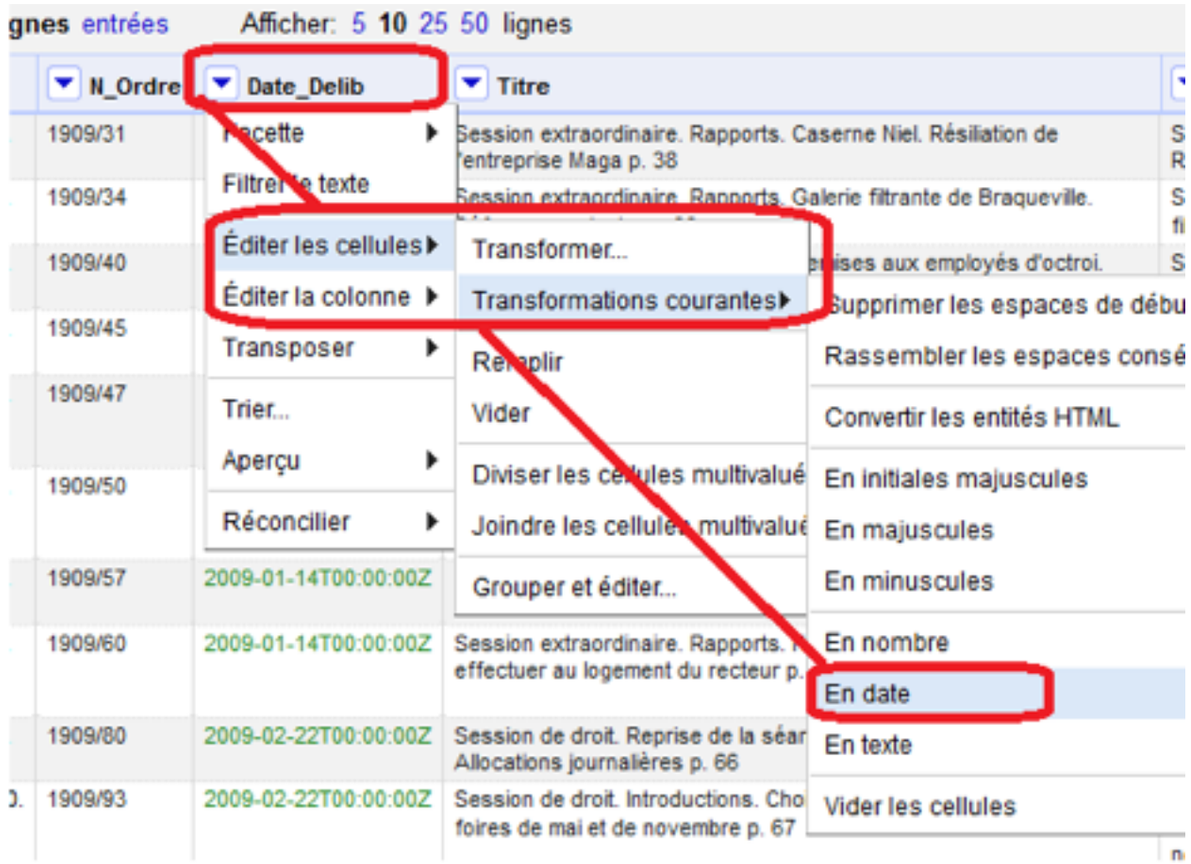
Plusieurs algorithmes

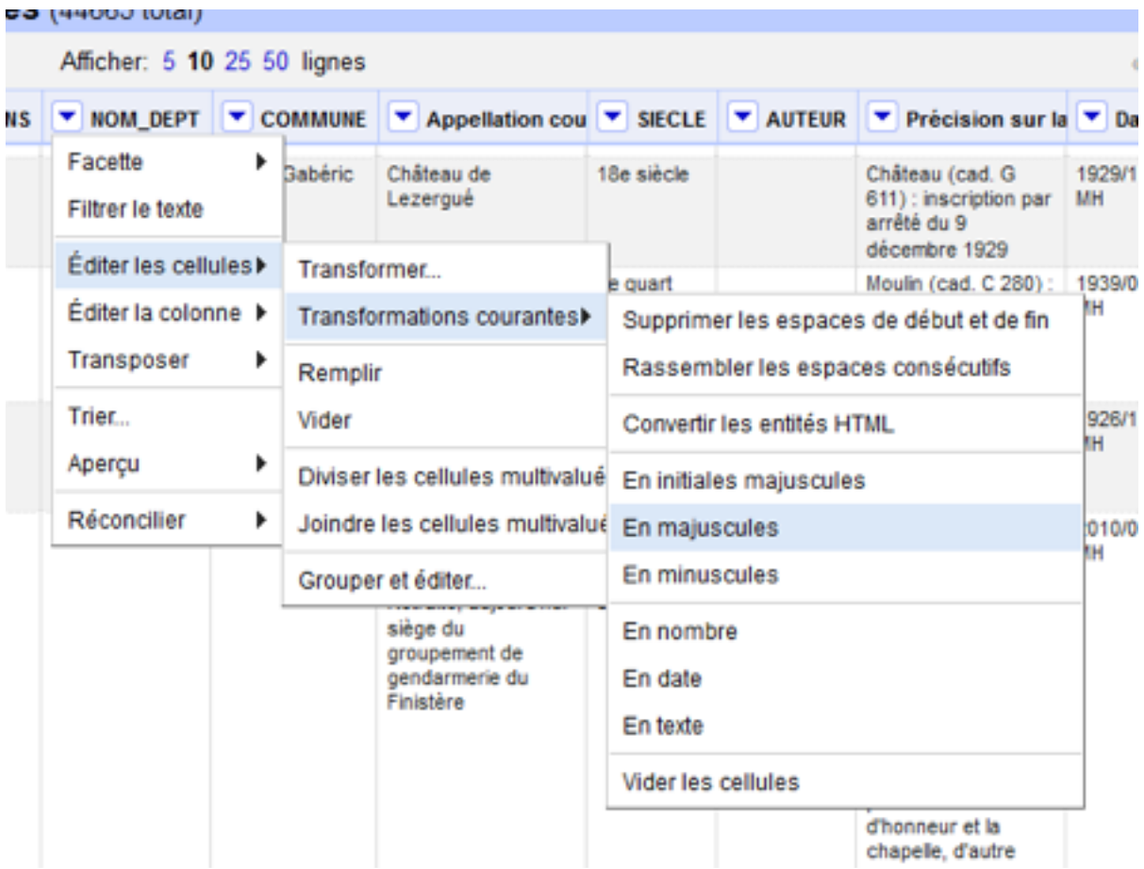


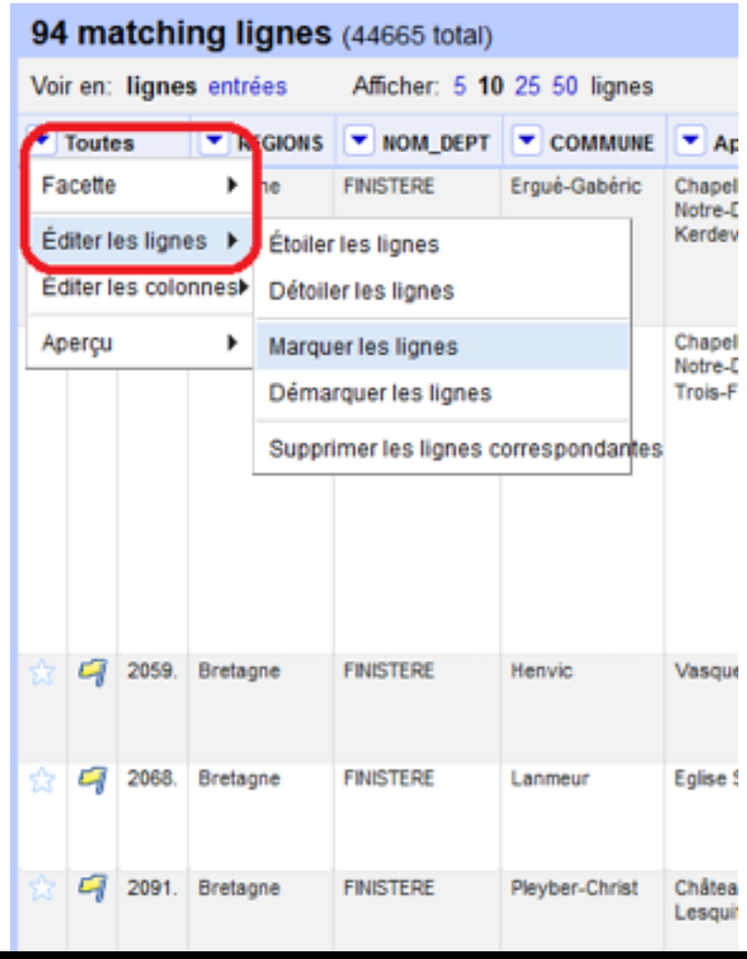

d'après une slide de Mathieu Saby.

# Les transformations de bases (déjà paramètrées



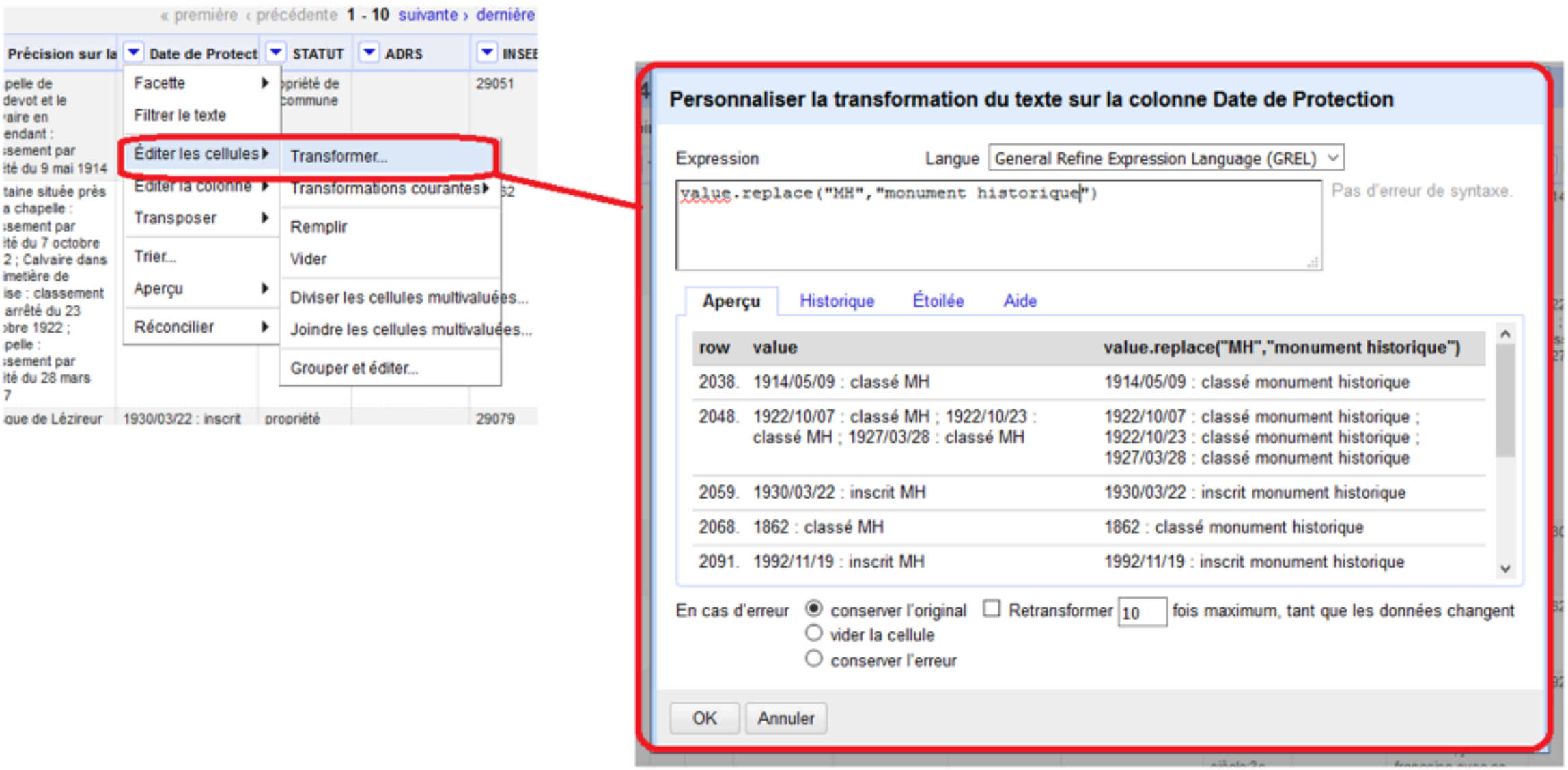



# Les transformations + complexes

Exemples :
Remplacer les "espace tiret espace" par un "tiret" : value.replace(" - ","-")
Des choses plus poussées (avec regex), comme:
"Enlever tout sauf les chiffres" : replace(value,/[[a-z],[A-Z],(é|è|à|ù),\,\;\:\.\?\/\!\=\+\"\'\-\(\)\[\]]/,"")
Extraire une année, soit une suite de 4 chiffres qui se suivent : value.match(/.*(\d{4}).*/)[0]



Pendant quelques secondes, une information s'affiche après une modification des données réalisée via le menu:

Ex:

Text transform on 5 cells in column Title: value.trim()

Text transform on 3 cells in column Title:
value.replace(/\s+/,' ') Défaire

Elle indique la **formule** utilisée par OpenRefine.

Ici :

| | |
|---|---|
| *value.trim()* | supprime les espaces initiaux et finaux |
| *value.replace(/\s+/,'')* | simplifie les espaces répétés |

Ces formules se retrouvent aussi dans l'historique des traitements.

## Deux menus utiles: aide et historique

Aide précieuse!
Utiliser Crtrl+F pour rechercher une commande

Aperçu Historique Étoilée **Aide**

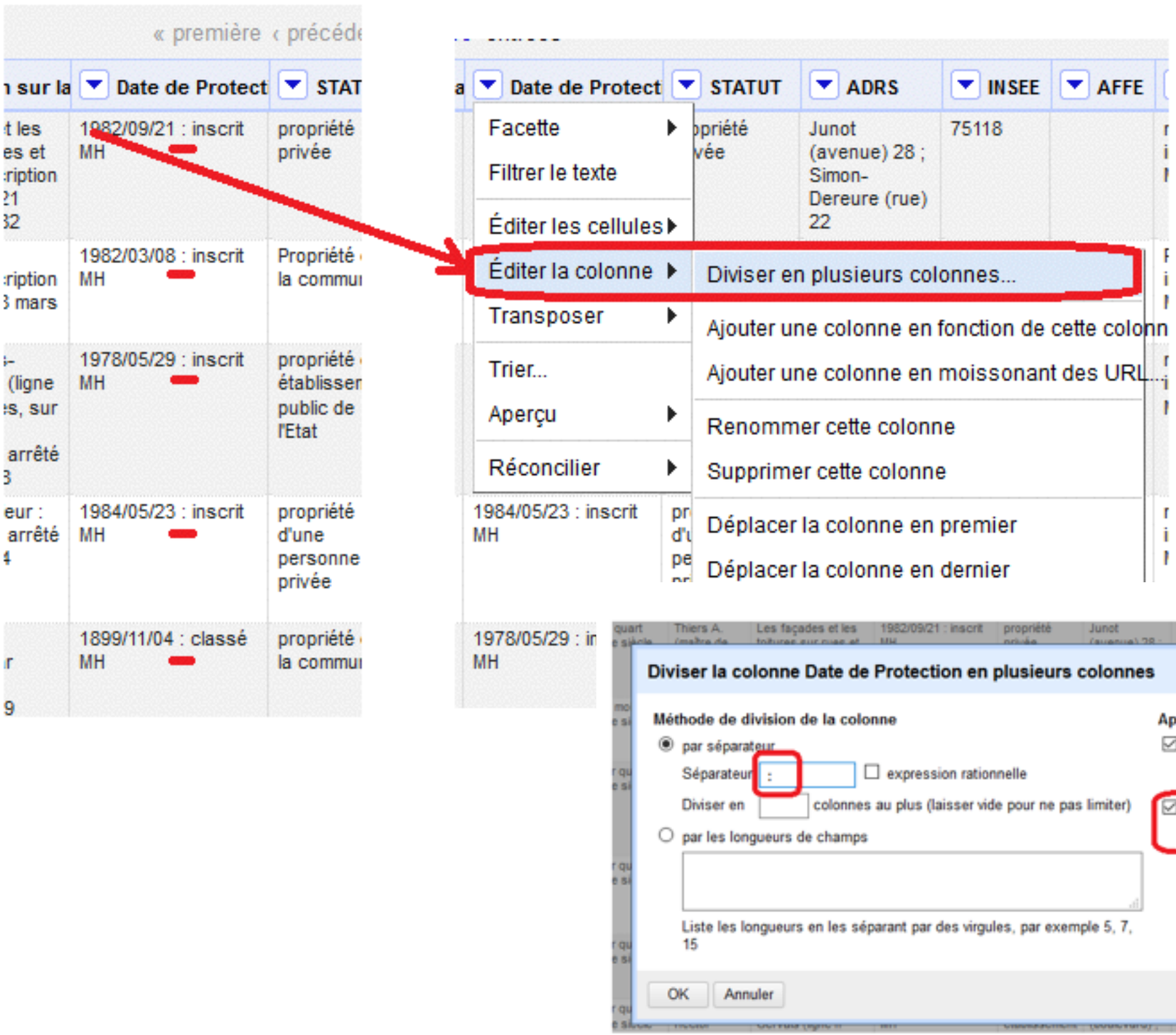
**Variables**

| | |
|---|---|
| **cell** | La cellule en cours. Elle a deux champs : 'value' et 'recon'. |
| **value** | La valeur de la cellule en cours. C'est un alias de 'cell.value'. |
| **row** | La ligne en cours. Elle a cinq champs : 'flagged', 'starred', 'index', 'cells' et 'record'. |
| **cells** | Les cellules de la ligne en cours. C'est un alias de 'row.cells'. Une cellule spécifique peut être retrouvée avec 'cells.<nom de la colonne>' lorsque <nom de la colonne > est un mot simple et avec 'cells.["<nom de la colonne>"]' dans les autres cas. |
| **rowIndex** | L'index de la ligne en cours. C'est un alias de 'row.index'. |

d'après des slides de Mathieu Saby.

# Spliter / diviser des informations / extraire

Pour ce faire : Clic sur la colonne / Éditer la colonne / Diviser en plusieurs colonnes > définir le séparateur commun



# Historique des modifications et automatisation des traitements

**My Address Book**

| friend | address |
|---|---|
| john | 120 Main St. |
| mary | 50 Broadway Ave. |
| anne | 17 Morning Crescent |

**Christmas Gifts**

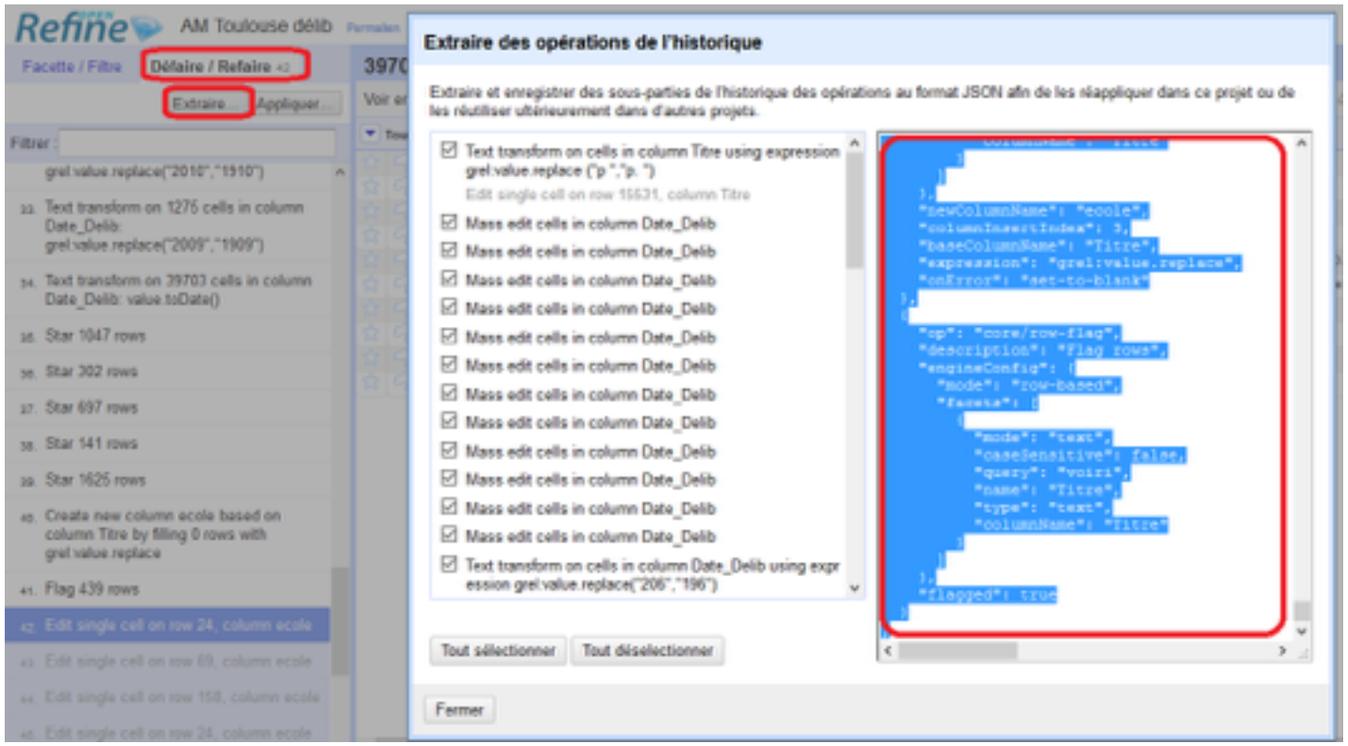
| gift | recipient |
|---|---|
| lamp | mary |
| clock | john |

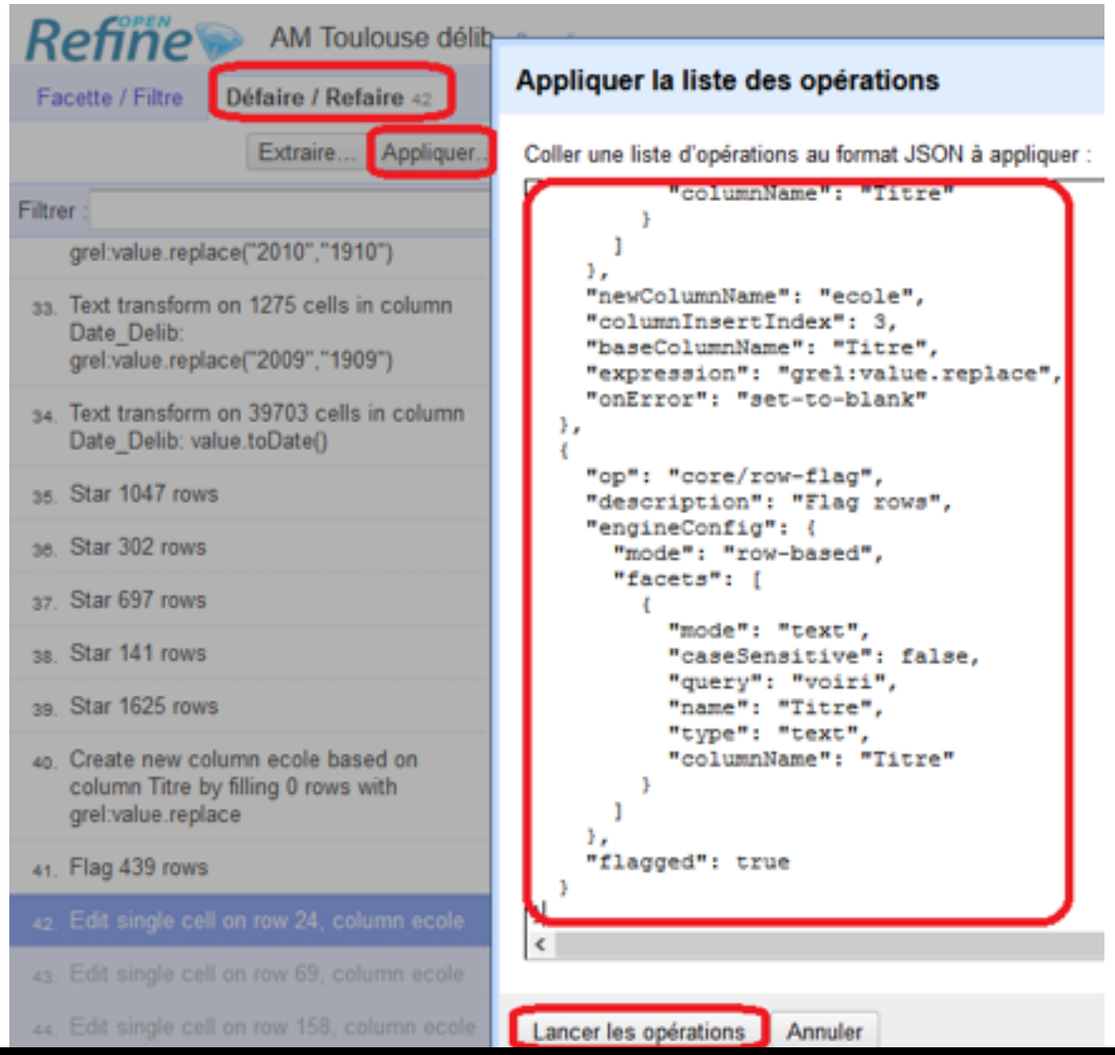
En se plaçant dans le projet « Christmas Gifts » on veut rapatrier l'adresse

```
cell.cross("My Address Book", "friend")[0].cells["address"].value
```

# Historique des modifications et automatisation des traitements

Importer / exporter le script
Il est possible de récupérer partie ou totalité du script de modifications, pour par exemple le réappliquer sur des fichiers semblables.
Dans le menu gauche, Défaire / Refaire puis Extraire.



# Export / templating

Open... | Export | Help

14807 rows — Extensions: VIB-Bits — RDF

Show as: **rows** records — Show: 5 10 25 50 rows — » last »

| All | | idCritique | nom | prenom | typeSignature | titreCritique | critic | | | ecise |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. | 5472 | APOLLINAIRE | Guillaume | patronyme | Gazette d'art. Le Pergamon à Berlin | | | | | 15 |
| 2. | 5468 | APOLLINAIRE | Guillaume | patronyme | L'Exposition de Düsseldorf | | | | | 11 |
| 3. | 5469 | APOLLINAIRE | Guillaume | patronyme | Le musée germanique de Nuremberg | | | | | 13 |
| 4. | 5471 | APOLLINAIRE | Guillaume | patronyme | Questions franco-allemandes. L'Exposition de Düsseldorf | | | | | |
| 5. | 5474 | APOLLINAIRE | Guillaume | patronyme | Gazette d'Art. Des Faux | | | | | 01 |
| 6. | 5476 | APOLLINAIRE | Guillaume | patronyme | Le cinématographe | | | | | 21 |
| 7. | 5478 | APOLLINAIRE | Guillaume | patronyme | Les jeunes : Picasso, peintre | | | | | 15 |
| 8. | 5480 | APOLLINAIRE | Guillaume | patronyme | Picasso, peintre et dessinateur (Galeries Serrurier) | | | | | |
| 9. | 5483 | APOLLINAIRE | Guillaume | patronyme | Bernheim-Wagram | | | | | 25 |
| 10. | 5492 | APOLLINAIRE | Guillaume | patronyme | Henri Matisse | | La Phalange | 1907 | 1907-12-15 | |

- Export project
- Tab-separated value
- Comma-separated value
- HTML table
- Excel
- ODF spreadsheet
- Triple loader
- MQLWrite
- Custom tabular exporter...
- Templating...
- RDF as RDF/XML
- RDF as Turtle

Refine — Archi_US_dataset_ticetlart_20160609 csv — Open... | Export | Help

Facet / Filter | Undo / Redo | Extensions:

Refresh | Reset All | « first « previous 1 - 1 next » last »

Nomc complet: judgment — 0 choices — Sort by: name count — Facet by choice counts

Nomc complet: best candidate's score — 100.00 — 1.00

Nomc complet: judgment — 0 choices — Sort by: name count — Facet by choice counts

Nomc complet: best candidate's score — 100.00 — 1.00

## Templating Export

Prefix

```
{
  "rows" : [
```

Row Template

```
    {
      "UNIQUE_KEY" : {{jsonize(cells["UNIQUE_KEY"].value)}},
      "Nom_usuel" : {{jsonize(cells["Nom_usuel"].value)}},
      "Nomc complet" : {{jsonize(cells["Nomc complet"].value)}},
      "id wikidata" : {{jsonize(cells["id wikidata"].value)}},
      "Prenom_usuel" : {{jsonize(cells["Prenom_usuel"].value)}},
      "Sexe" : {{jsonize(cells["Sexe"].value)}},
      "Date_naissance_recherche" : {{jsonize(cells["Date_naissanc
      "Date_naissance_texte" : {{jsonize(cells["Date_naissance_te
      "Lieu_naissance 1" : {{jsonize(cells["Lieu_naissance 1"].va
      "geoloc" : {{jsonize(cells["geoloc"].value)}},
      "long" : {{jsonize(cells["long"].value)}},
      "lat" : {{jsonize(cells["lat"].value)}},
      "Date_deces_recherche" : {{jsonize(cells["Date_deces_recher
      "Date_deces_texte" : {{jsonize(cells["Date_deces_texte"].va
      "Lieu_deces" : {{jsonize(cells["Lieu_deces"].value)}},
      "Formation" : {{jsonize(cells["Formation"].value)}},
      "Atelier" : {{jsonize(cells["Atelier"].value)}},
      "Commentaire biographique" : {{jsonize(cells["Commentaire b
      "Notice_MLC" : {{jsonize(cells["Notice_MLC"].value)}},
```

Row Separator

```
,
```

Suffix

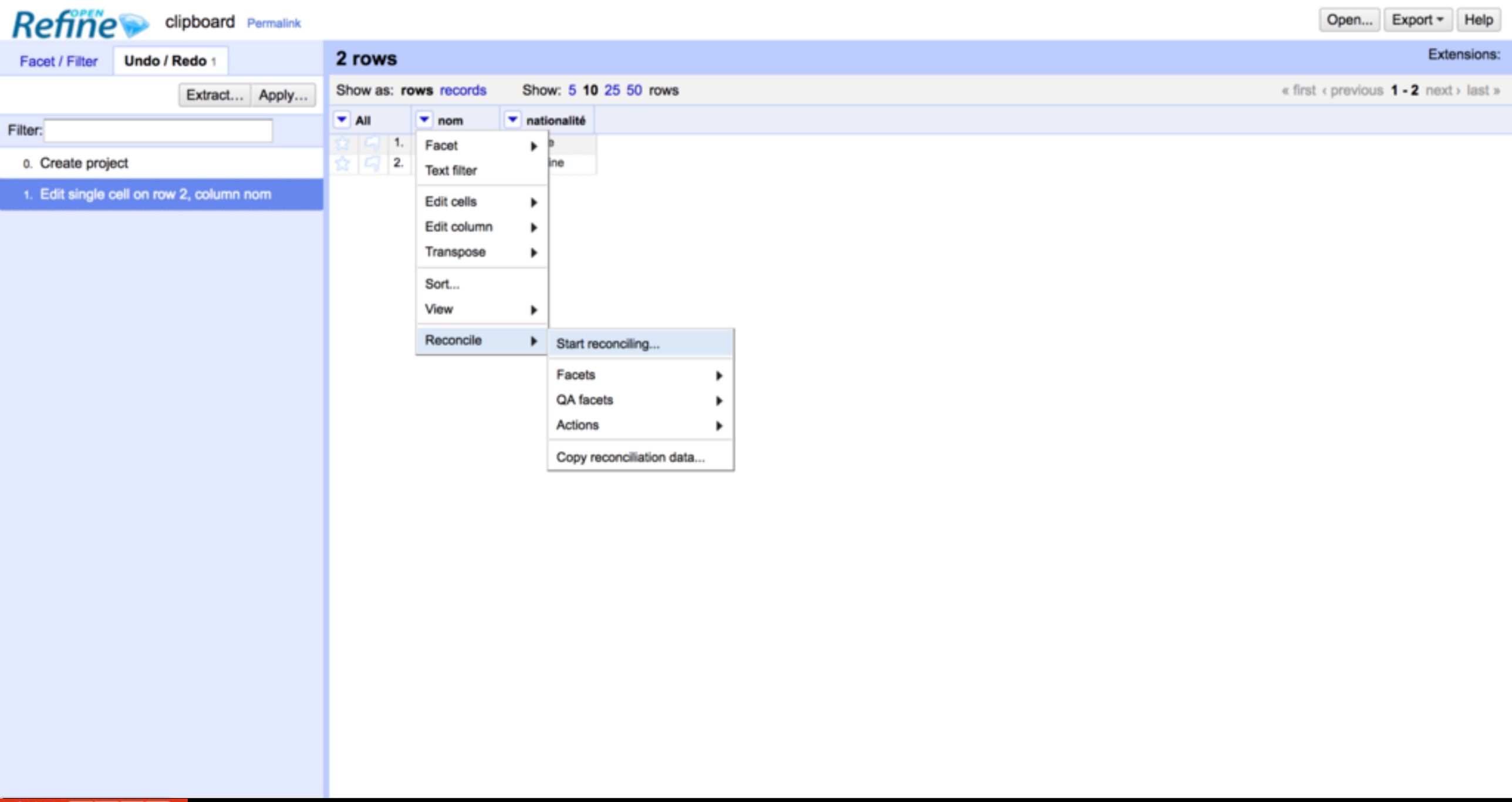
```
  ]
}
```

```
{
  "rows" : [
    {
      "UNIQUE_KEY" : "75113",
      "Nom_usuel" : "Adams",
      "Nomc complet" : "Adams, Lewis Greenleaf",
      "id wikidata" : "Q6536618",
      "Prenom_usuel" : "Lewis Greenleaf",
      "Sexe" : "masculin",
      "Date_naissance_recherche" : "23 novembre 1897",
      "Date_naissance_texte" : "1897-11-23",
      "Lieu_naissance 1" : "Lenox, Etat-Unis",
      "geoloc" : "{\n   \"results\" : [\n      {\n         \"addr
      "long" : -73.2848762,
      "lat" : 42.35646089999999,
      "Date_deces_recherche" : "29-nov-77",
      "Date_deces_texte" : "29/11/1977",
      "Lieu_deces" : "New York",
      "Formation" : "formation-scolarité à l'École des beaux-arts
      "Atelier" : null,
      "Commentaire biographique" : null,
      "Notice_MLC" : null,
      "Matricule" : null
    }
  ]
}
```

Reset Template | Export | Cancel

# Reconcilier des jeux de données par des entités nommées

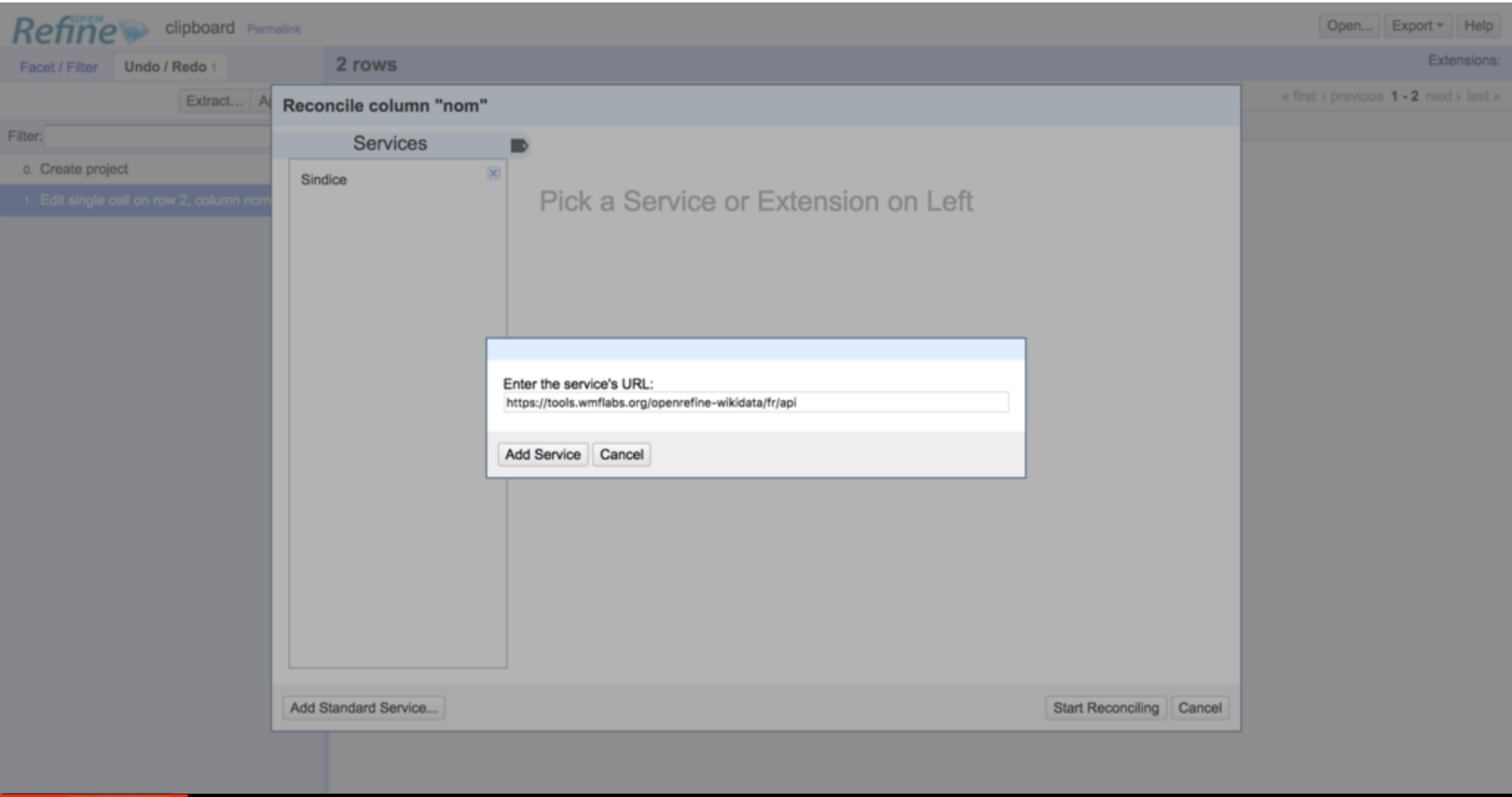
Après l'import de votre fichier de données, choisir une colonne de votre choix, puis le menu Reconcile et Start reconciling.



# Reconcilier des jeux de données par des entités nommées

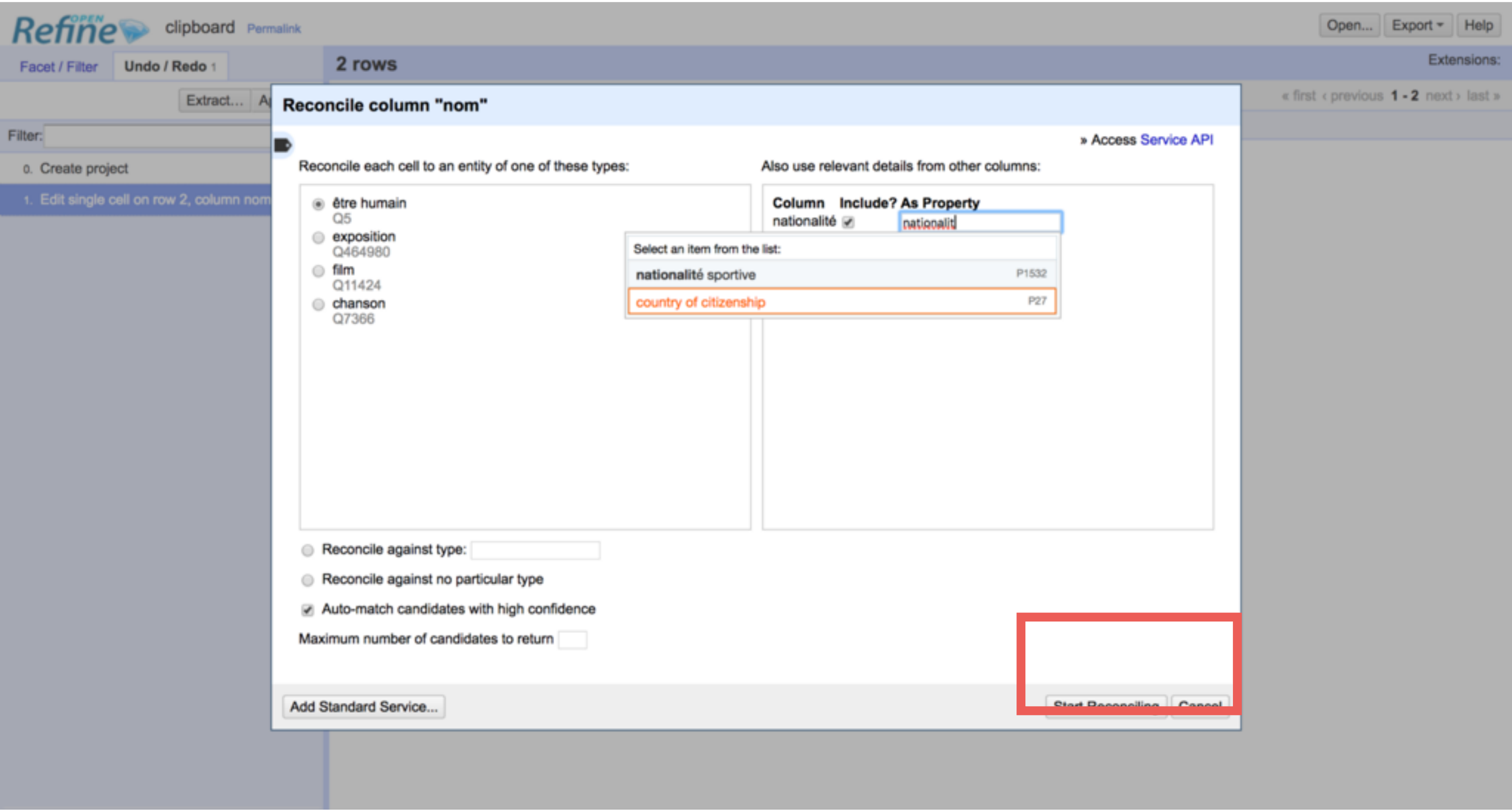
2. Une fenêtre s'ouvre, cliquer sur le bouton en bas à gauche Add Standard Service

3. Une popup s'ouvre, ajouter l'URL suivante : **https://tools.wmflabs.org/openrefine-wikidata/fr/api**.

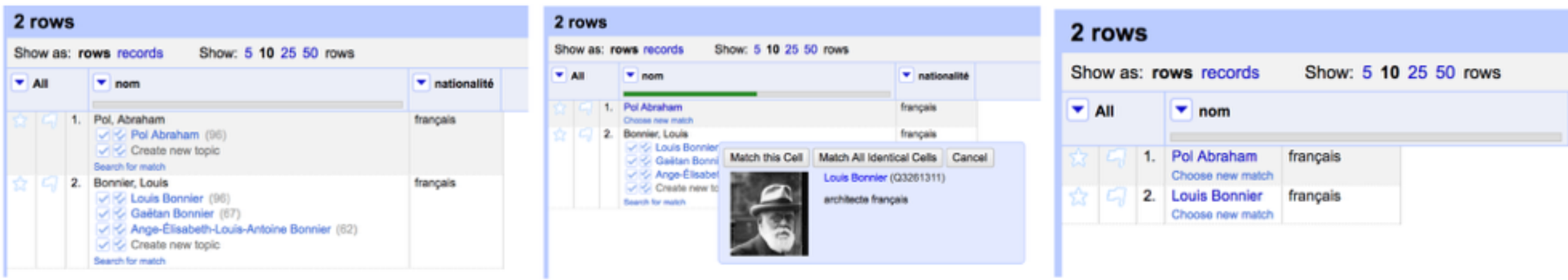


# Reconcilier des jeux de données par des entités nommées

Une fois cette URL rentrée, la colonne de gauche propose différents “types” (sous le forme de Q????? dans wikidata) susceptibles de correspondre à votre type de données. Dans notre cas, il identifie (à raison) le Q5 -> être humain. [capture n°3]



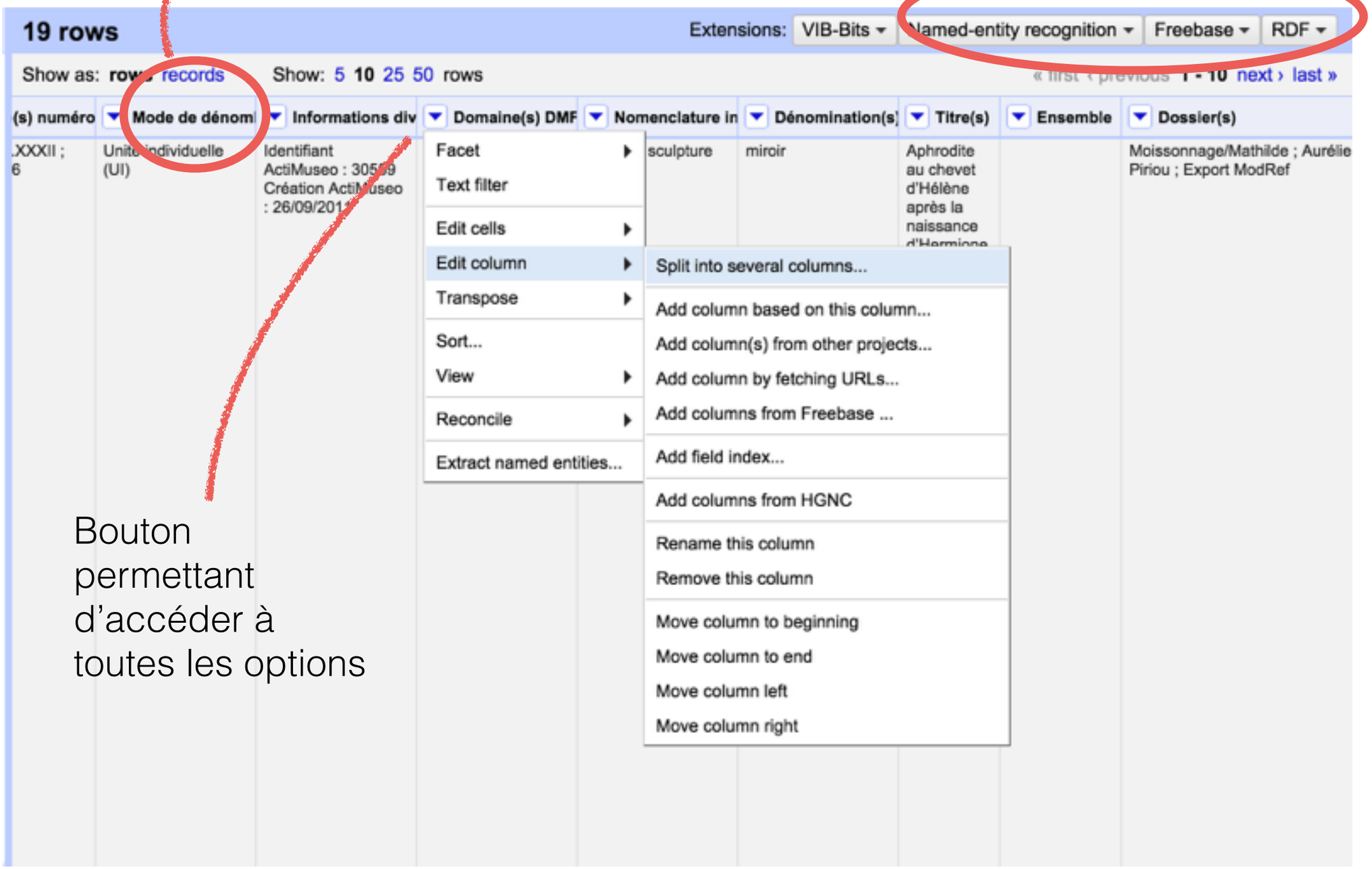
Edit Column > Add column base on this column [capture n°7]. Une fenêtre s'affiche, ajouter dans la zone "Expression", l'élément suivant: cell.recon.match.id




[capture n°4–5–6] 3 étapes de validation

Plugins RDF

Champs de description




Bouton permettant d'accéder à toutes les options

créer le squelette d'export RDF

Google refine export d taill 2015 03 24 09_32 csv Permalink

Open... Export ▾ Help

Facet / Filter | Undo / Redo 0

Refresh | Reset All | Remove All

Matières et techniques change

14 choices Sort by: name count Cluster

argile (peint (revêtement et décor)) 4

améthyste (taille en intaille) 2

20 rows

Extensions: VIB-Bits ▾ Named entity recognition ▾ Freebase ▾ RDF ▾

- Edit RDF Skeleton...
- Reset RDF Skeleton...
- Add reconciliation service ▸

Show as: rows records Show: 5 10 25 50 rows

| techn | Mesures | Avec manche | Avec monture | Column3 | Avec socle | Column4 | C |
|---|---|---|---|---|---|---|---|
| | | Hauteur | Largeur | Hauteur | Longueur | Largeur | Haute |
| ert | D. 16 cm (sans manche) ; H. 23,8 cm | 23,8 | | | | | |

Google refine repfall Permalink

Open... Export ▾ Help

Facet / Filter | Undo / R

Freebase ▾ RDF ▾

Using facets and fi

Use facets and filters to of your data to act on. filter methods from the of each data column.

Not sure how to get sta Watch these screenca

revious 1 - 10 next › last »

geoloc | PRECISION_LIEU_C

## RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

**Base URI:** http://localhost:3333/ edit

RDF Skeleton | RDF Preview

Available Prefixes: rdfs foaf xsd owl rdf + add prefix manage prefixes

(row index) URI — add rdf:type

- property? → NUM_NOTICE cell
- property? → TYPE_OEUVRE cell
- property? → PRECISION_TYPE_OEUVRE cell
- property? → TITRES cell
- property? → COMMENTAIRE_GENERAL cell
- property? → COTE_NUM_INV cell
- property? → LIEU_DE_CONSERVATION cell
- property? → geoloc cell
- property? → PRECISION_LIEU_DE_CONSERVATION cell
- property? → INFORMATIONS_PRATIQUES cell
- property? → COMMENTAIRE_DESCRIPTIF cell

Add another root node | Save

OK | Cancel

## RDF Schema Alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

**Base URI:** http://localhost:3333/ edit

RDF Skeleton | RDF Preview

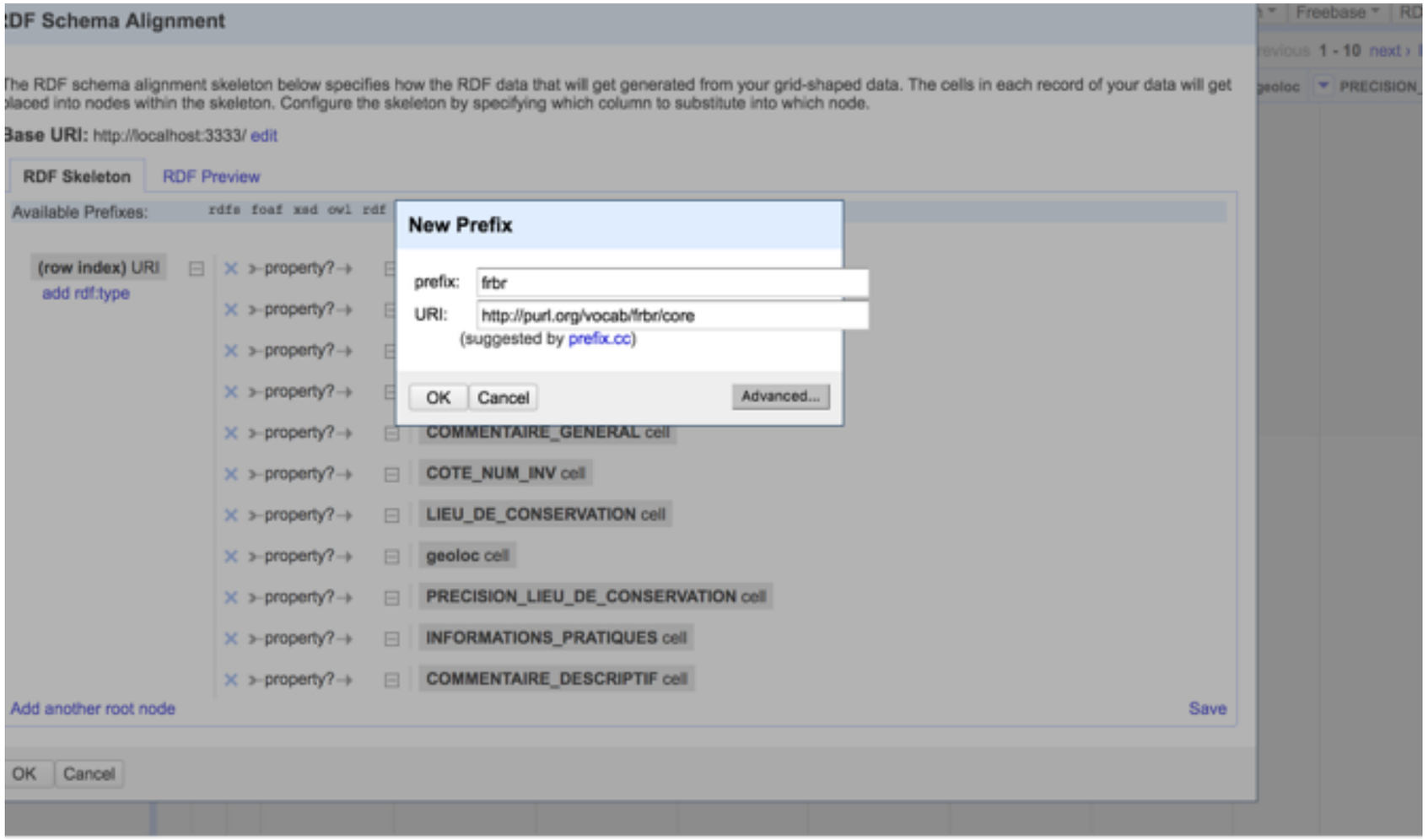
Available Prefixes: rdfs foaf xsd owl rdf crm add prefix manage prefixes

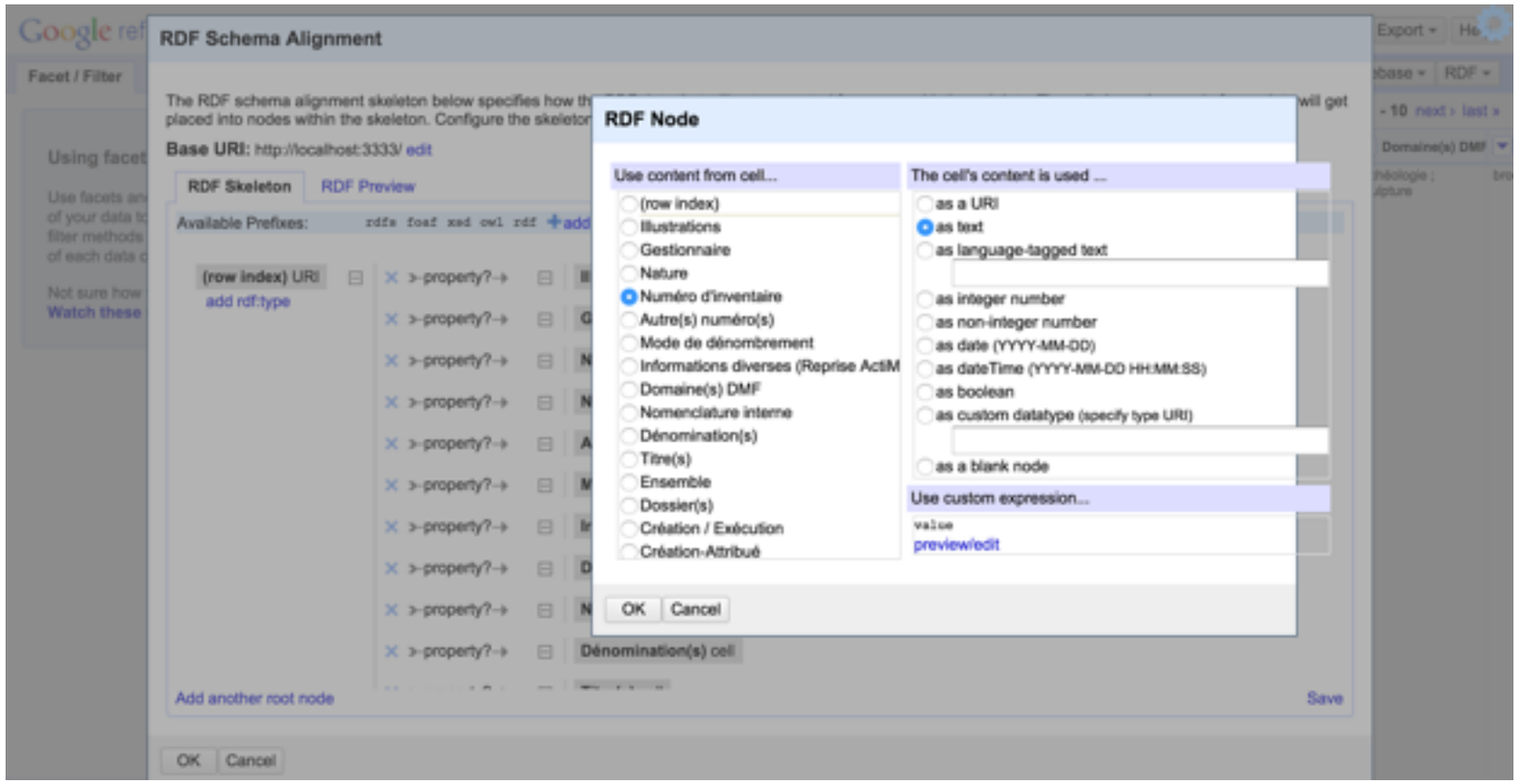
ajouter un namespace càd une ontologie que l'on pourrait réutiliser dans le squelette

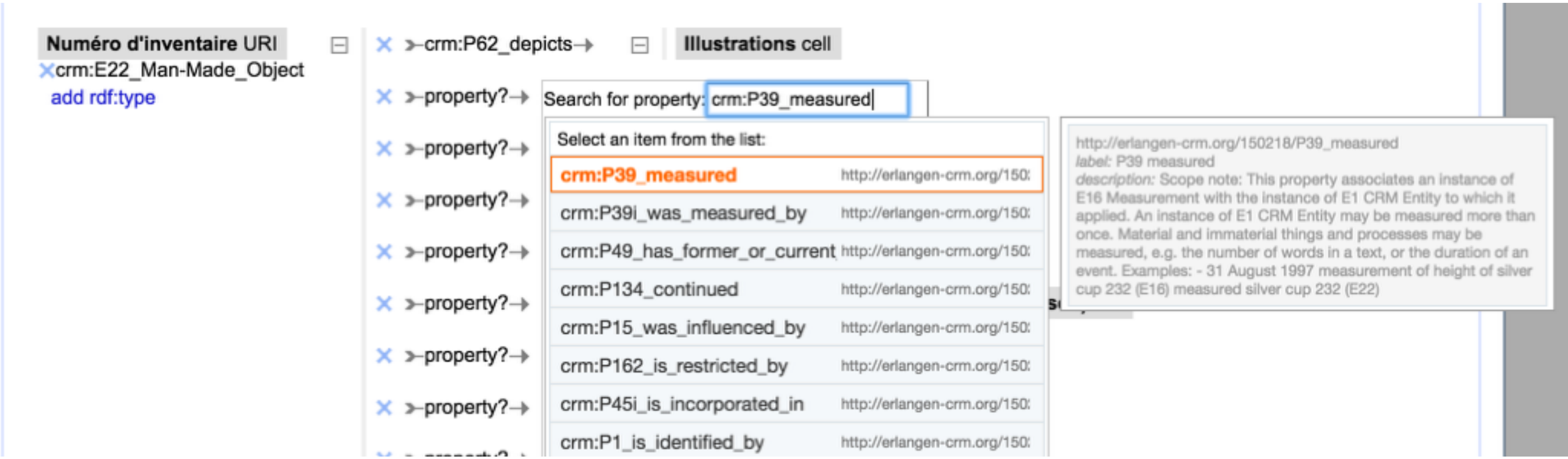
**Numéro d'inventaire** URI
crm:E22_Man-Made_Object
add rdf:type

crm:P62_depicts → **Illustrations** cell

property? → **Gestionnaire** cell




RDF Schema Alignment
Facet / Filter
Using facet
Base URI: http://localhost:3333/ edit
RDF Skeleton
RDF Preview
Available Prefixes: rdfs foaf xsd owl rdf
(row index) URI
add rdf:type
Dénomination(s) cell
Add another root node
Save
OK
Cancel
RDF Node
Use content from cell...
(row index)
Illustrations
Gestionnaire
Nature
Numéro d'inventaire
Autre(s) numéro(s)
Mode de dénombrement
Informations diverses (Reprise ActiM
Domaine(s) DMF
Nomenclature interne
Dénomination(s)
Titre(s)
Ensemble
Dossier(s)
Création / Exécution
Création-Attribué
The cell's content is used ...
as a URI
as text
as language-tagged text
as integer number
as non-integer number
as date (YYYY-MM-DD)
as dateTime (YYYY-MM-DD HH:MM:SS)
as boolean
as custom datatype (specify type URI)
as a blank node
Use custom expression...
value
preview/edit
OK
Cancel


Numéro d'inventaire URI
crm:E22_Man-Made_Object
add rdf:type
crm:P62_depicts
Illustrations cell
property?
Search for property: crm:P39_measured
Select an item from the list:
crm:P39_measured http://erlangen-crm.org/150
crm:P39i_was_measured_by http://erlangen-crm.org/150
crm:P49_has_former_or_current http://erlangen-crm.org/150
crm:P134_continued http://erlangen-crm.org/150
crm:P15_was_influenced_by http://erlangen-crm.org/150
crm:P162_is_restricted_by http://erlangen-crm.org/150
crm:P45i_is_incorporated_in http://erlangen-crm.org/150
crm:P1_is_identified_by http://erlangen-crm.org/150
http://erlangen-crm.org/150218/P39_measured
label: P39 measured
description: Scope note: This property associates an instance of E16 Measurement with the instance of E1 CRM Entity to which it applied. An instance of E1 CRM Entity may be measured more than once. Material and immaterial things and processes may be measured, e.g. the number of words in a text, or the duration of an event. Examples: - 31 August 1997 measurement of height of silver cup 232 (E16) measured silver cup 232 (E22)

# A vous de jouer !

- Dans le dossier sur GitHub : test_EBD_1.csv
  - importer le fichier dans OpenRefine et vérifier les paramètres d’imports
  - Supprimer les colonnes vides
  - Vérifier s’il y a des doublons dans les intitulés
    - Jouer avec les facettes (plusieurs façon de faire)
  - Supprimer les doublons identifiés
  - Extraire l’année du champs « Date_debut » sous forme aaaa
  - extraire la ville du lieu d’exposition
    - soit en « splitant » soit en extrayant les parenthèses
    - supprimer les parenthèses le cas échéant et contrôler grâce aux filtres

- Toujours dans le même répertoire télécharger sur GitHub
  - Archi_US_dataset_ticetlart_20160609.csv
  - A partir des identifiants UK, créer une URL cliquable amenant vers la notice AGORHA
  - Ajouter à la colonne Lieu_naissance 1, le terme de Etats-Unis
  - Lancer la géolocalisation via « by fetching URL »
    - "http://maps.google.com/maps/api/geocode/json?sensor=false&address=" + escape(value, "url")
  - Extraite les coordonnées géographiques dans une autre colonne
  - Créer une colonne Nom_complet rassemblant le nom et le prénom
  - Reconcilier le nom des architectes avec Wikidata puis extraire l'identifiant wikidata.
  - Dans wikidata, faire une requête pour exporter en CSV tous les architectes américains avec les informations de lieu de naissance et de décès.
  - Réaliser une jointure à partir de l'id wikidata pour récupérer les lieux dans le fichier d'origine.

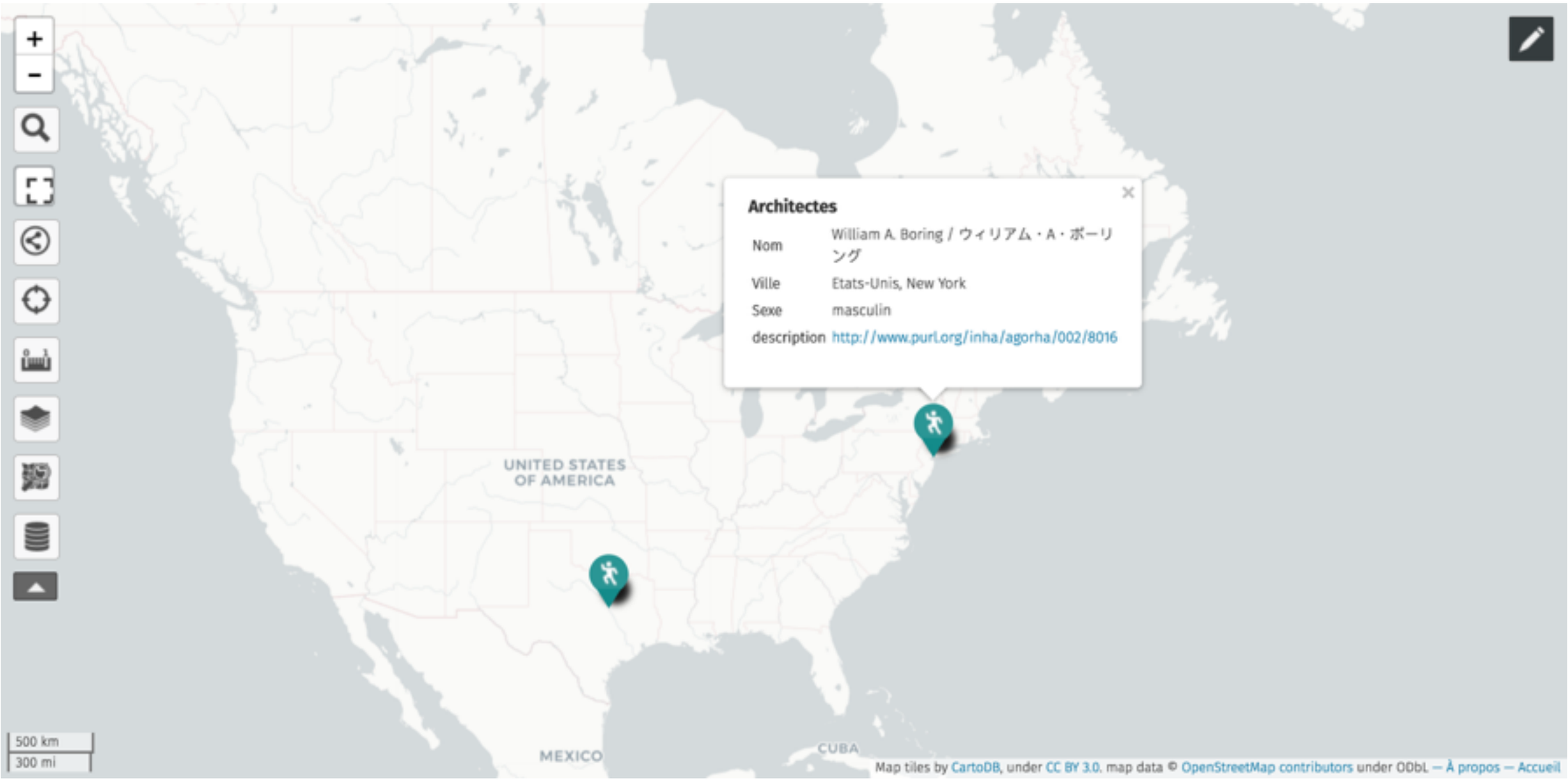
- Avec le fichier exo1.csv
  - harmoniser la colonne *ville*
  - dans la colonne *adresse,* pour la rue Jean Jaurès, quel est le soucis ?
- Avec le fichier exo2.csv
  - voir la différence entre « lignes » et « entrées »
  - dans la colonne *ville,* remplir les champs vides puis de nouveau retirer les valeurs répétées

- Dans le dossier téléchargé de githud : test_cnam_1.csv
  - importer le fichier dans OpenRefine et vérifier les paramètres d’imports
  - Supprimer les colonnes vides
  - Vérifier s’il y a des doublons dans les intitulés
    - Jouer avec les facettes (plusieurs façon de faire)
  - Supprimer les doublons identifiés
  - Extraire l’année du champs « Date_debut » sous forme aaaa
  - extraire la ville du lieu d’exposition
    - soit en « splitant » soit en extrayant les parenthèses
    - supprimer les parenthèses le cas échéant et contrôler grâce aux filtres

- Toujours dans le même répertoire télécharger sur GitHub
  - ouvrir le dossier export_XML_EAD
  - utilizer le fichier clipboard.csv
  - créer un fichier xml en utilisant la template expliquée dans le fichier template_export_IR.md

- Toujours dans le même répertoire télécharger sur GitHub
  - Importer le fichier Archi_US_dataset_CNAM_2018 csv
  - A partir des identifiants (colonne UNIQUE_Key), créer une URL cliquable amenant vers la notice AGORHA (en sachant que l'URL de base est > http://www.purl.org/inha/agorha/002/
  - Dans la colonne « Profession », vérifier si elle est bien vide. Et après, rajouter l'info « architecte » à la 1er ligne et répéter l'info sur toutes les lignes suivantes.
  - Ajouter à la colonne « Lieu_naissance », le terme de « Etats-Unis, »
  - Lancer la géolocalisation via « by fetching URL »
    - "http://maps.google.com/maps/api/geocode/json?sensor=false&address=" + escape(value, « url")"
  - Extraite les coordonnées géographiques dans une autre colonne sous la forme « latitutde, longitude ».
  - créer un nom complet en additionnant « Nom usuel » + « Prenom »
  - Reconcilier le nom des architectes avec Wikidata en cochant « correspondance automatique »
    - Qu'observez-vous pour « Hastings, Thomas ». Comment pourrons-nous améliorer le matching ?
    - Rejouer la reconciliation.
  - puis extraire l'identifiant wikidata et le label wikidata
  - Dans wikidata, faire une requête pour exporter en CSV tous les architectes américains avec les informations de lieu de naissance et sa version en russe et en japonais (http://tinyurl.com/y73gz2ub)
  - Réaliser une jointure à partir de l'id wikidata pour récupérer les lieux dans le fichier d'origine.

- Générer un fichier geojson qui récupère les infos du fichiers
- Visualiser-le dans umaps d’OpenStreetMap




https://umap.openstreetmap.fr/fr/

Merci !