



Observatoire des cultures écrites  
de l'argile à l'imprimé

# Workshop: “Openrefine Reconciliation with Biblissima”

Kevin Bois (EPHE, Paris France)  
Eduard Frunzeanu (EPHE, Paris France)  
Emmanuelle Morlock (CNRS, Lyon, France) (**facilitator**)  
Régis Robineau (EPHE, Paris France)

DH 2024, Pre-conference workshop, 5 August 2024, Washington DC

Ce travail a été réalisé grâce à une aide (ou : la consultation) de Biblissima+, qui bénéficie d'une aide de l'Etat gérée par l'ANR au titre du Programme d'investissements d'avenir intégré à France 2030, portant la référence ANR-21-ESRE-0005.

# Presentation on the DH2024 website

<https://dh2024.adho.org/program/workshops/>

Aligner, vérifier et enrichir des jeux de données sur le patrimoine écrit ancien: découvrir et pratiquer le web service de réconciliation pour OpenRefine de Biblissima

If you have never found the time to experiment with the super easy data reconciliation feature of OpenRefine, this half-day workshop will give you the opportunity! It will also introduce the Biblissima endpoint for authoritative data about written cultures (shelfmarks for manuscripts or ancient prints, people, places, iconographic descriptors, etc.), which will be used for the exercises. While the process is straightforward, this session may lead to interesting discussions among the participants on how to balance the need for automation and critical review of the results obtained. See <https://data.biblissima.fr/w/Accueil/en>. The workshop will be given in English and/or French depending on the speaking language of the audience

# Agenda

- Introduction to
  - ◆ Biblissima
  - ◆ Reconciliation
  - ◆ OpenRefine
- Quick DemoTime
  - ◆ Openrefine's interface
  - ◆ Standard reconciliation with Wikidata
- Hands-on exercices
  - ◆ Reconcile with data.biblissima
  - ◆ Improve results
  - ◆ Align and enrich a TEI file and visualize the result with TEI Publisher

# Downloads

Software:

<https://openrefine.org/download>

Exercises and handouts:

<https://github.com/emmamorlock/workshop>

# **Introduction**

~ 30 min

# A Digital Infrastructure for Ancient Written Heritage

<https://projet.biblissima.fr/en>

Funded by the French government since 2012.

Two phases of funding:

- Biblissima 2012-2020
- Biblissima+ 2021-2029

PI: Anne-Marie Turcan Verkerk  
(EPHE)

Supporting institution: Campus Condorcet

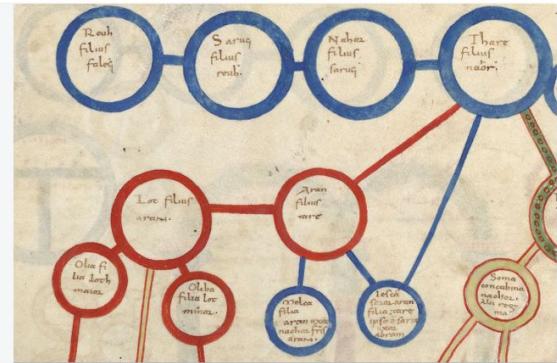


Biblissima+ is a multi-site digital infrastructure for research and service dedicated to the history of the transmission of ancient texts, from Antiquity to the Renaissance, in the West and in the East. [Project news](#)

## From Biblissima to Biblissima+

The Biblissima+ infrastructure (2021-2029) is structured around **7 areas of expertise or "clusters"** bringing together researchers, curators and engineers. They form centres of expertise on a national scale around the study of ancient texts and their transmission through History...

[Read more](#)

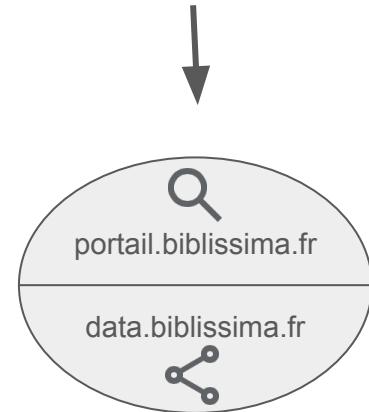


# Biblissima: Building interoperability

2012-2020

Building a virtual library of libraries offering a single access point to the scholarly documentation on the history of the transmission of texts (8th-18th centuries)

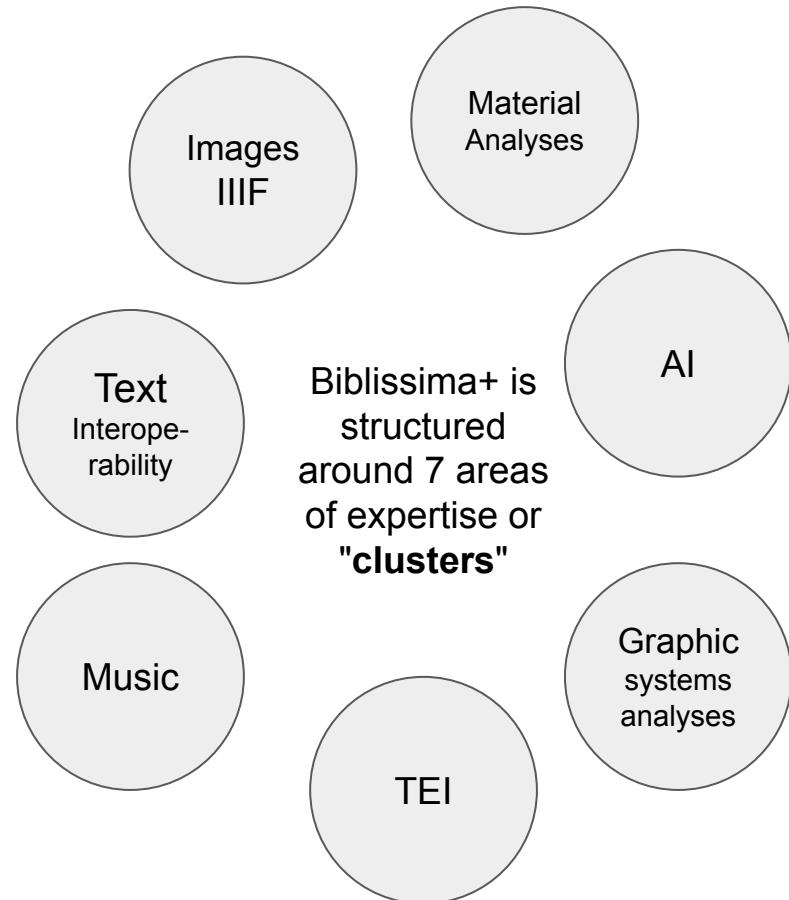
- Data Cluster
- Web portal ([portail.biblissima](http://portail.biblissima))
- Authorities hub ([data.biblissima](http://data.biblissima))
- Community (call for projects, summer schools)
- Cutting edge tools
  - e.g. IIIF Collections (search engine)

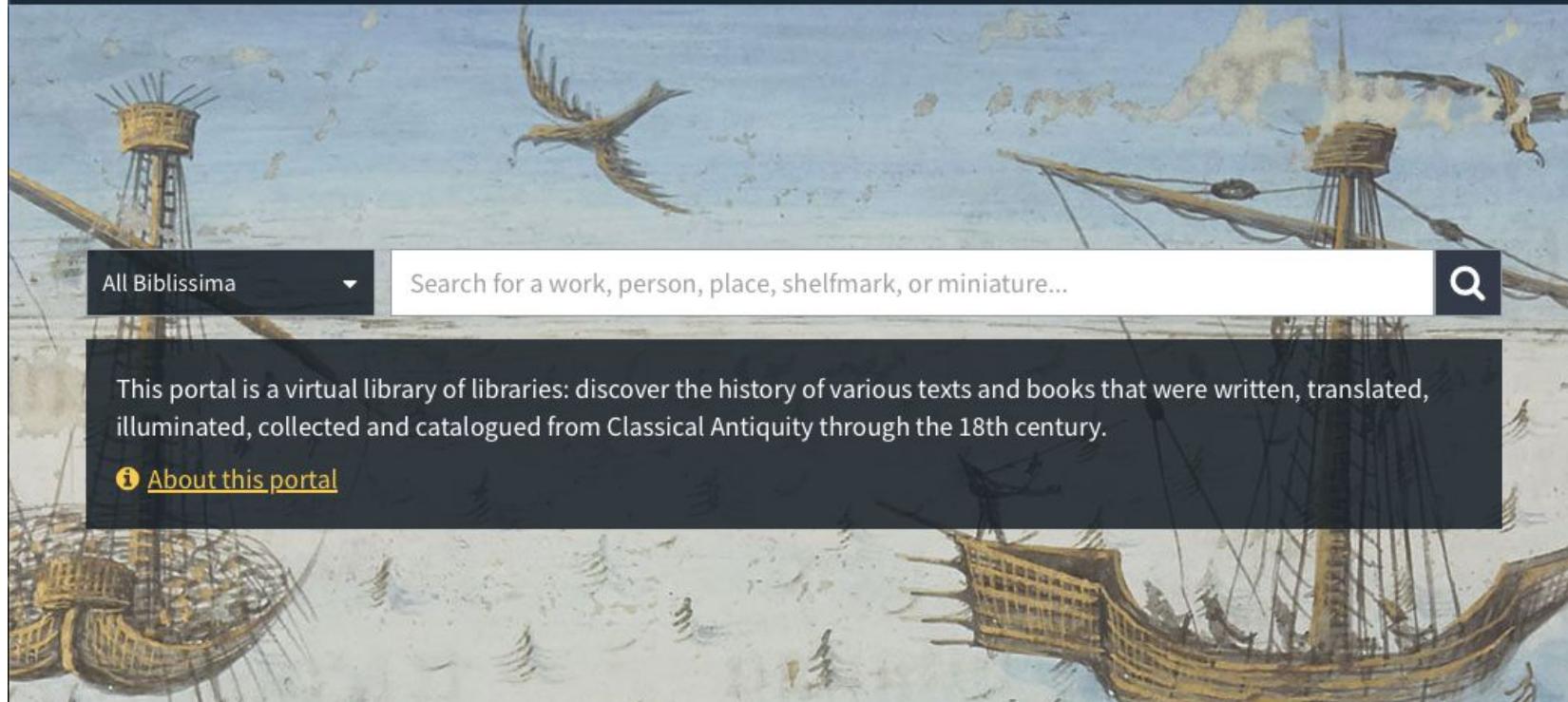


# Biblissima+: Broadening the scope

2021-2029

- Extension to antiquity
- Beyond the book: archeological resources (seals, inscriptions, coins), music, material analysis...
- Numerous scientific challenges:
  - Fonts without unicode
  - Handling manuscripts in all ancient languages (Armenian, Greek, Spanish, Ethiopian)
  - Provide access to full-text secondary bibliography





All Biblissima ▾

Search for a work, person, place, shelfmark, or miniature...

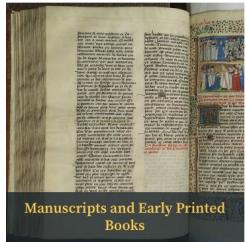


This portal is a virtual library of libraries: discover the history of various texts and books that were written, translated, illuminated, collected and catalogued from Classical Antiquity through the 18th century.

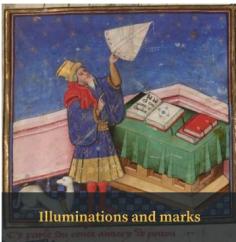
 [About this portal](#)

Biblissima Web Portal : ~776 000 pages – <https://portail.biblissima.fr/en>

# Resources in the Biblissima portal



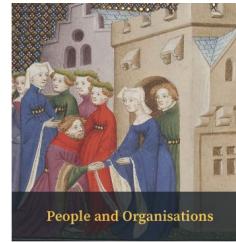
Manuscripts and Early Printed Books



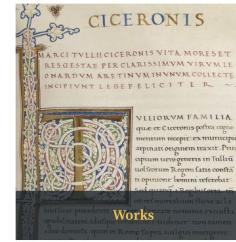
Illuminations and marks



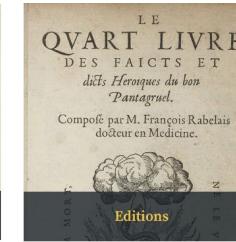
Places



People and Organisations



Works



Editions



Historical Collections

27601 [Descriptors](#)

306346 [Illuminations and decoration](#)

6774 [Places](#)

46596 [People](#)

25046 [Works](#)

6936 [Marks](#)

163 [Topics](#)

4720 [Multi-volume manuscripts](#)

7778 [Editions](#)

5938 [Organisations](#)

231734 [Manuscripts](#)

67562 [Texts](#)

25316 [Early Printed Books](#)

654 [Inventories](#)

13597 [Historical Collections](#)

All Biblissima

Search for a work, person, place, shelfmark, or miniature...



## Example: WORK

Presentation

Related texts

Work witnesses

Life cycle of this work

Data sources

Exports

## Le chevalier au lion

Biblissima authority file: <https://data.biblissima.fr/entity/Q286568>

**Other Forms :** Chrétien de Troyes (113.-1185?) | Le chevalier au lion  
Chrétien, de Troyes, active 12th century. Chevalier au lyon  
chrétien de troyes, chevalier au lion  
chrétien de troyes, chevalier au lion (vers. pierre sala)  
Der Löwenritter  
Le chevalier au lion ou Le roman d'Yvain  
Le chevalier au lion, Yvain  
Le Roman du chevalier au Lyon  
Le Roman du Chevalier au Lyon, par Maistre Gasse  
Le roman d'Yvain  
The knight of the lion  
The knight with the lion  
Yvain  
Yvain, le chevalier au lion  
Yvain or The knight with the lion  
Yvain ou Le chevalier au lion  
Yvain, the knight of the lion  
Ywain and Gawain

**Author :** Chrétien de Troyes (113.-1185?)

**Languages :** French, Old (842-ca.1400)

**Notes :** À employer également pour les adaptations anonymes en faisant suivre la vedette de la subdivision de forme Adaptations suivie de Histoire et critique dans le cas d'études critiques (Source : BnF)  
Roman, sans doute écrit vers 1177-1181, d'environ 6800 octosyllabes. - Adaptations en allemand par Hartmann von Aue sous le titre "Iwein", en norrois au XIIIe s. sous le titre "Ivens saga", en anglais au XIVe s. sous le titre "Ywain and Gawain", en 4032 vers (Source : BnF)  
Mentionné dans: La Croix du Maine (E1v, p. 50) « Le Roman du chevalier au Lyon. » (Source : Bibliothèques françoises)

## Related texts

---

Chrétien de Troyes (113.-1185?) : Le chevalier au lion [Unknown language].

Chrétien de Troyes (113.-1185?) : Le chevalier au lion [Unknown language]. Enlumineur Guillaume Leroy II (14..-15..)

## Work witnesses

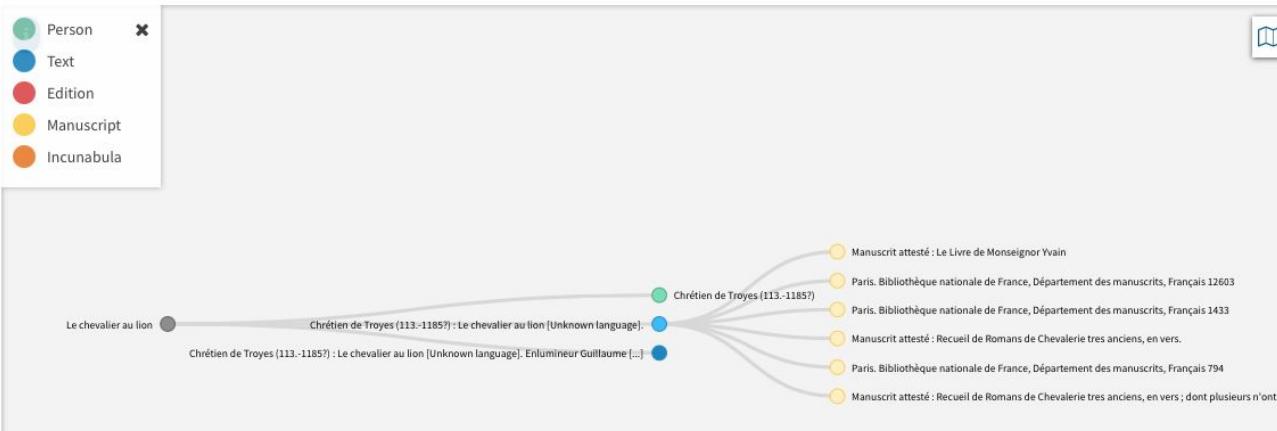
---

### Manuscripts

- > Manuscrit attesté : Le Livre de Monseignor Yvain
- > Manuscrit attesté : Recueil de Romans de Chevalerie tres anciens, en vers.
- > Manuscrit attesté : Recueil de Romans de Chevalerie tres anciens, en vers ; dont plusieurs n'ont jamais été imprimez.
- >  Paris. Bibliothèque nationale de France, Département des manuscrits, Français 794
- >  Paris. Bibliothèque nationale de France, Département des manuscrits, Français 1433
- >  Paris. Bibliothèque nationale de France, Département des manuscrits, Français 1638
- >  Paris. Bibliothèque nationale de France, Département des manuscrits, Français 12603

## Life cycle of this work

---



## Data sources

---

- > Bibliothèques Françoises : [http://bibfr.bvh.univ-tours.fr/biblio/lcdm-bibl\\_07153h](http://bibfr.bvh.univ-tours.fr/biblio/lcdm-bibl_07153h) (More information about this resource)
- > BnF Archives et manuscrits
- > Corpus Fabliaux
- > Esprit des livres
- > Jonas
- > Mandragore

All Biblissima

Search for a work, person, place, shelfmark, or miniature



## Example: Historical collection

[Presentation](#)   [Other former owners](#)   [Currently held at\\*](#)   [Data sources](#)   [Exports](#)

## Bibliothèque de Jean duc de Berry

Type : Collection privée / laïc

Former Owner : Jean (duc de Berry, 1340-1416)

**BIBLIOTHÈQUE DE JEAN DUC DE BERRY**

**COLLECTION**  
Bibliothèque de Jean duc de Berry

IIIIF Collection provided by Biblissima.

**100 manifestes**

Paris. Bibliothèque nationale de France, Département des manuscrits, Latin 10484

Paris. Bibliothèque nationale de France, Département des manuscrits, Français 606

Paris. Bibliothèque nationale de France, Département des manuscrits, Français 12420

Angers. Bibliothèque municipale, Ms. 162

**FERMER**

IIIF collection  
<https://portail.biblissima.fr/iiif/collection/ark:/43093/coldat>

**CHANTILLY, Bibliothèque du château, 0065 (1284) (Reproduction intégrale)**

plat supérieur      plat inférieur      contreplat et garde      contreplat - f. 001      f. 001v - 002

f. 002v - 003      f. 003v - 004      f. 004v - 005      f. 005v - 006

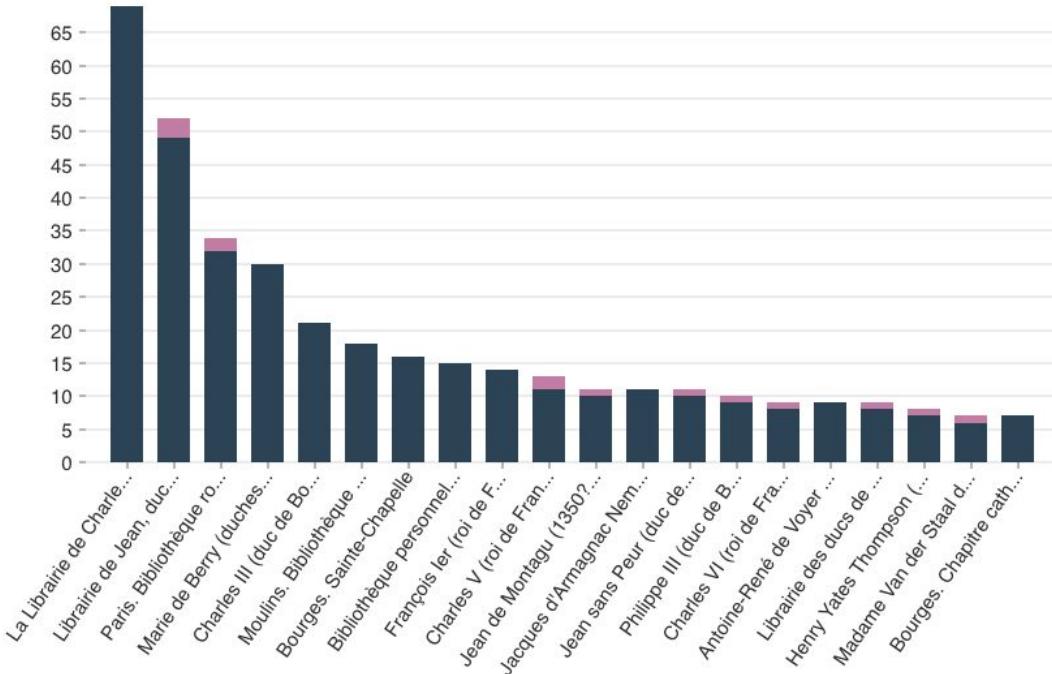
A large red arrow points from the left panel to the right panel, indicating a connection between the two collections.

## Other former owners for this collection

Only the 20 most frequent co-occurring collections are displayed. To see them all, choose the CSV export:

Export data (csv)

Click on a bar to display more information.



## Currently held at

---

Angers. Bibliothèque municipale

- >  [Angers. Bibliothèque municipale, Ms. 162](#)

Bourges. Bibliothèque municipale

- >  [Bourges. Bibliothèque municipale, Ms. 124](#)
- >  [Bourges. Bibliothèque municipale, Ms. 33](#)
- >  [Bourges. Bibliothèque municipale, Ms. 335](#)
- >  [Bourges. Bibliothèque municipale, Ms. 34](#)
- >  [Bourges. Bibliothèque municipale, Ms. 35](#)
- >  [Bourges. Bibliothèque municipale, Ms. 36](#)
- >  [Bourges. Bibliothèque municipale, Ms. 48](#)
- >  [Bourges. Bibliothèque municipale, Ms. 67](#)

Bruxelles. Bibliothèque royale de Belgique

- > [Bruxelles. Bibliothèque royale de Belgique, Ms. 9542](#)
- > [Bruxelles. Bibliothèque royale de Belgique, Ms. 9553](#)
- > [Bruxelles. Bibliothèque royale de Belgique, Ms. 9554](#)
- > [Bruxelles. Bibliothèque royale de Belgique, Ms. 9555-58](#)

Chantilly. Bibliothèque et Archives du Château

- >  [Chantilly. Bibliothèque et Archives du Château, Ms. 122](#)
- >  [Chantilly. Bibliothèque et Archives du Château, Ms. 277](#)
- >  [Chantilly. Bibliothèque et Archives du Château, Ms. 339](#)
- > [Chantilly. Bibliothèque et Archives du Château, Ms. 487](#)
- >  [Chantilly. Bibliothèque et Archives du Château, Ms. 65](#)
- >  [Chantilly. Bibliothèque et Archives du Château, Ms. 757](#)

(...)

## Turin. Archivio di Stato

- > [Turin. Archivio di Stato, Biblioteca Antica J.b.II.18](#)

## Valence. Universitat de València, Biblioteca històrica

- >  [Valence. Universitat de València, Biblioteca històrica, BH Ms. 387](#)

## Unknown holding institution

- > [Baltimore. Walters Art Museum, W.125 | Baltimore. Walters Art Museum, W.126](#)
- >  [Baltimore. Walters Art Museum, W.140 | Paris. Bibliothèque nationale de France, Département des manuscrits, Français 316](#)
- > [Hambourg. Kunsthalle Kupferstichkabinett, Cod. fr. 1 | Paris. Bibliothèque de l'Arsenal, Ms-5212 réserve](#)
- > [Londres. British Library, Add MS 6416 | Paris. Bibliothèque nationale de France, Département des manuscrits, NAF 15939-15944](#)
- > [Paris. Bibliothèque nationale de France, Département des manuscrits, Français 835 | Paris. Bibliothèque nationale de France, Département des manuscrits, Français 606 | Paris. Bibliothèque nationale de France, Département des manuscrits, Français 836 | Paris. Bibliothèque nationale de France, Département des manuscrits, Français 605 | Paris. Bibliothèque nationale de France, Département des manuscrits, Français 607](#)
- > [Vente, London, Sotheby's, 1981, 8 décembre, no.94](#)

## Data sources

---

- > Bibale : <https://bibale.irht.cnrs.fr/collection/1443> (More information about this resource)
- > Bibale : <https://bibale.irht.cnrs.fr/collection/607> (More information about this resource)
- > Alexandrine
- > BnF Archives et manuscrits
- > Esprit des livres
- > Europeana regia
- > Initiale
- > Jonas
- > Persée - DER-IRHT, LXXXIV



Accueil  
Rechercher  
Types d'entités  
Modifications récentes

Wikibase  
Créer un item  
Créer une propriété  
Liste des propriétés  
Tous les items et propriétés  
  
Outils  
Pages liées  
Suivi des pages liées  
Pages spéciales  
Version imprimable  
Lien permanent  
Informations sur la page

Biblissima  
Portail Biblissima  
IIIF Collections  
  
Dans d'autres langues  
Ajouter des liens

## Home

Cette page est une version traduite de la page Accueil et la traduction est terminée à 100 %.

Autres langues: English · français

Sommaire [masquer]

- 1 Biblissima Authority Files
  - 1.1 Available Authority Files
  - 1.2 In progress
  - 1.3 Data access, API
  - 1.4 Data Licence
  - 1.5 Presentations

## Biblissima Authority Files

Total number of items in the authority files: 406,905

This platform provides Biblissima's authority file in the form of structured data. Its realization is part of the Biblissima+ programme, Observatory of Written Cultures, from Clay to Print (2021-2029). It follows Biblissima, Observatory of the written heritage of the Middle Ages and the Renaissance from the 8th to the 18th century (2012-2021).

These authority files are based on data collected and processed within the framework of the Biblissima+ programme.

They form the backbone of the Biblissima portal and the prototype IIIF Collections of Manuscripts and Rare Books.

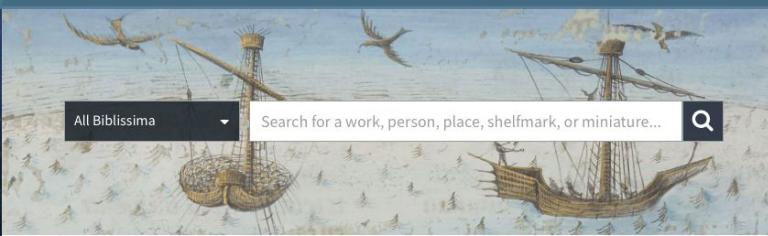
They have been set up in order to identify, disambiguate, harmonize and align named entities from a wide range of data sources (digital libraries, catalogues and specialized databases on manuscripts and early printed books). This effort has been ongoing since the beginning of Biblissima in 2013.

For any request, or if you want to contribute, you can contact us at [team@biblissima-condorcet.fr](mailto:team@biblissima-condorcet.fr)

**Disclaimer:** In its first phase, this platform is designed as a tool for managing, publishing and exposing the Biblissima authority files. Its primary goal is to create a hub of identifiers and structured data, that is accessible, interoperable and reusable for other projects.

This hub plays a central role in the interoperability of the resources processed within the framework of Biblissima, while allowing other projects to benefit from the work of reconciliation and harmonisation carried out in this direction. The data is indeed accessible in an automated way by various mechanisms intended primarily for machines (see below [Data access and API](#)). Nevertheless, developments are planned to improve the user interface, the full-text search, and to offer easier browsing and querying modes.

Hub for semantic authorities: ~315 000 entities – <https://data.biblissima.fr>



All Biblissima

Search for a work, person, place, shelfmark, or miniature...


[Presentation](#) [Associated works and texts](#) [Associated documents](#) [In this library](#) [Data sources](#) [Exports](#)

## Cassiodore (0490?-0580?)



Other Forms :

Cassiodore (v. 485-v. 580)  
 Cassiodoro, ca 487-ca 580  
 Cassiodoro, ca 487-ca 580>, co-autor  
 Aurelius Cassiodorus (0490?-0580?)  
 Flavio Magno Aurelio Cassiodoro (0490?-0580?)

Birth :

490 (Uncertain data)

Death :

580 (Uncertain data)

Notes :

Dizionario Biografico degli Italiani, 21, 1978: Cassiodoro (Flavius Magnus Aurelius Cassiodorus Senator; più spesso detto Cassiodorus Senator o semplicemente Senator, che è nome proprio e non designazione di dignità) (Source : Biblissima)  
 A. Molinier, Les Sources de l'histoire de France - Des origines aux guerres d'Italie (1494): 224. Magnus Aurelius Cassiodorus Senator (Source : Biblissima)  
 Préfet du prétoire et historiographe de Théodoric. - Fondateur du monastère de Vivarium (Source : Bnf)  
 Date : (Source : Initiale)

## Associated works and texts

This section lists the works attributed to the person as well as his contributions related to textual units.

Item Discussion

Lire Voir l'historique

Rechercher sur Référentiels

## Cassiodore (0490?-0580?) (Q2785)

Aucune description fournie

Cassiodore (0490?-0580?) | Cassiodore 0490?-0580? | Cassiodore | Cassiodore (v. 485-v. 580) | Cassiodore (v. 485-v. 580)

Plus de langues

Configurer

Langue	Libellé	Description	Également connu comme
français	Cassiodore (0490?-0580?)	Aucune description fournie	Cassiodore (0490?-0580?) Cassiodore 0490?-0580? Cassiodore Cassiodore (v. 485-v. 580) Cassiodore (v. 485-v. 580)

Toutes les langues saisies

## Déclarations

nature de l'élément	être humain	modifier
	▼ 0 référence	+ ajouter une référence
		+ ajouter une valeur

date de naissance	490 Grégorien	modifier
	qualité de l'information	prisé
	▼ 0 référence	+ ajouter une référence
		+ ajouter une valeur

date de mort	580 Grégorien	modifier
	qualité de l'information	prisé
	▼ 0 référence	+ ajouter une référence
		+ ajouter une valeur

Biblissima authority file: <https://data.biblissima.fr/entity/Q2785>

## Cassiodore (0490?-0580?) (Q2785)

Aucune description fournie

Cassiodore (0490?-0580?) | Cassiodore 0490?-0580? | Cassiodore | Cassiodore (v. 485-v. 580) | Cassiodore (v. 485-v. 580)

Plus de langues

Configurer

Langue	Libellé	Description	Également connu comme
français	Cassiodore (0490?-0580?)	Aucune description fournie	Cassiodore (0490?-0580?) Cassiodore 0490?-0580? Cassiodore Cassiodore (v. 485-v. 580) Cassiodore (v. 485-v. 580)

Toutes les langues saisies

## Déclarations

nature de l'élément	être humain	modifier
	▼ 0 référence	+ ajouter une référence
		+ ajouter une valeur

date de naissance	490 Grégorien	modifier
	qualité de l'information	prisé
	▼ 0 référence	+ ajouter une référence
		+ ajouter une valeur

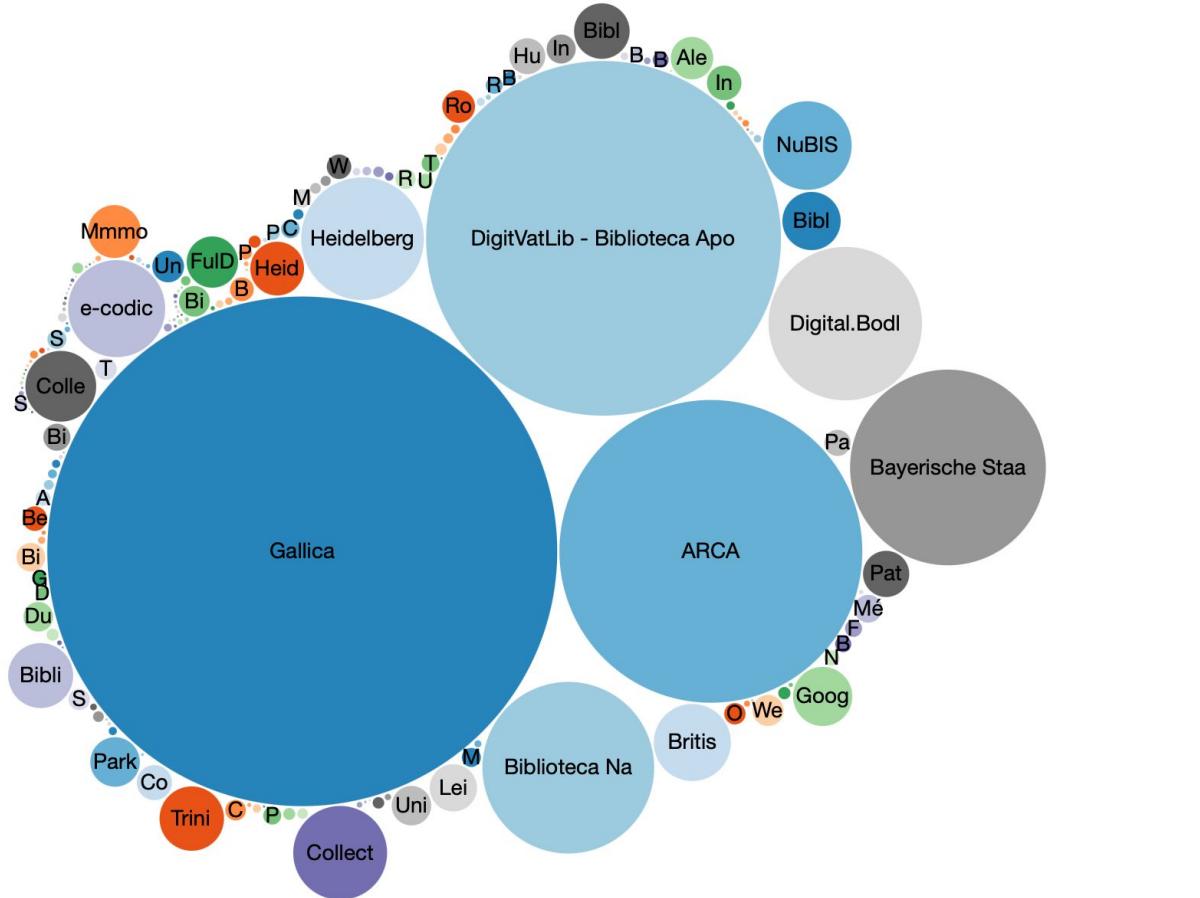
date de mort	580 Grégorien	modifier
	qualité de l'information	prisé
	▼ 0 référence	+ ajouter une référence
		+ ajouter une valeur

# The data.biblissima Hub

- Standalone wikibase instance (since March 2019)
- A data “Hub” to manage and share semantic authority data
  - Explicit Open Licence (Etalab Open Licence 2.0)
- Populated over time each time a new dataset is ingested in the Biblissima data cluster
- Open to contribution (contact us!)

# Main entity types

Individuals	<b>Alternative labels;</b> Dates of birth/death; Gender, Bibliography, <b>External identifiers</b>
Organisations	<b>Alternative labels</b> , Location, Bibliography, <b>External identifiers</b>
Places	<b>Alternative labels;</b> Country, Geographical coordinates, External identifiers
Manuscripts and early printed books shelfmarks	<b>Alternative labels</b> , digitisation url, IIIF manifest, Bibliography, <b>External identifiers</b>
Works	<b>Alternative labels;</b> Language, Author, Bibliography, <b>External identifiers</b>
Iconographic descriptors	<b>Alternative labels;</b> Broder concept, <b>External identifiers</b>



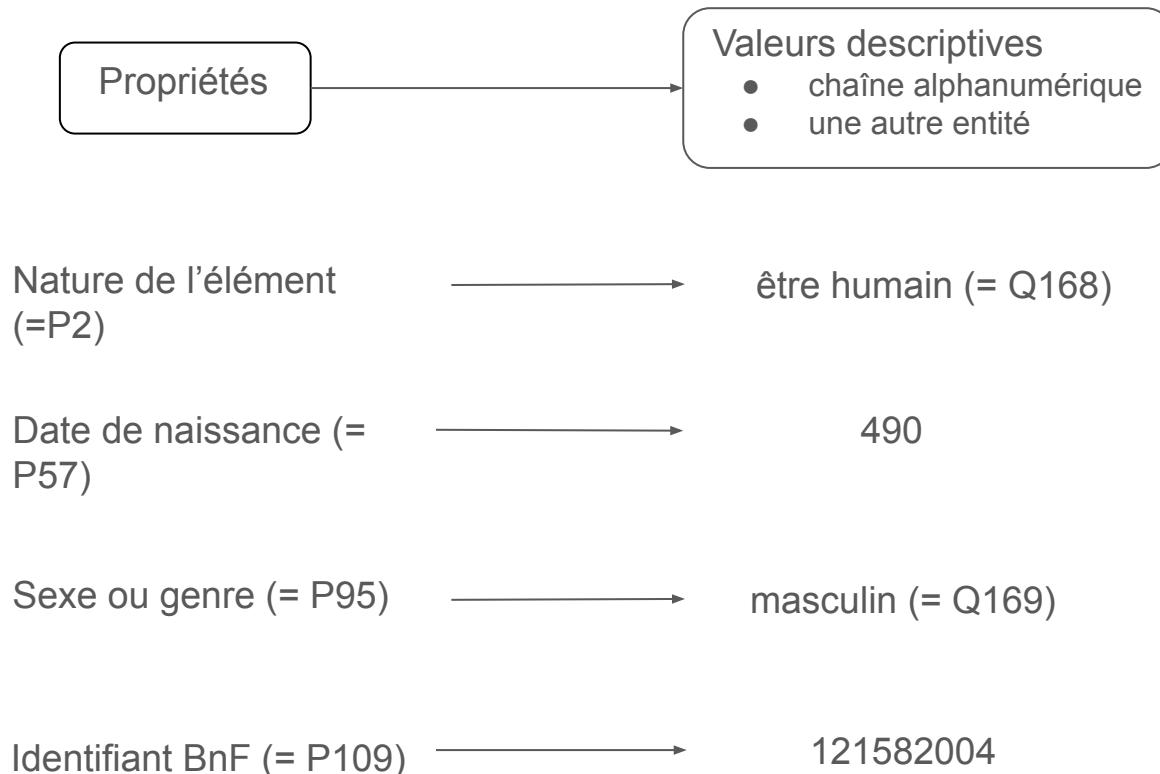


# Data model

- Based on the design principles of Wikibase:
  - Ad hoc, evolving inline with needs
- Based on library practices for authority files
- Aligned as much as possible with Wikidata or other ontologies
  - via properties like “equivalent property” or “equivalent class”
  - Other ontologies like Schema.org, SKOS, FOAF, DC
- 308 properties
  - ~182 for external identifiers ex. [CERL identifier](#)
- 88 item classes
  - Main classes: Human (Q168), Descriptor (Q304387), Oeuvre (Q282950) Manuscrit (Q32810), organisation (Q167)...

## Types d'entités

# Cassiodore



- Personnes physiques
- Familles
- Organisations
- Noms géographiques
- Cotes de manuscrits
- Cotes d'imprimés anciens
- Œuvres
- Descripteurs iconographiques



## Cassiodore (0490?-0580?)

(2785)

Aucune description fournie

Cassiodore (0490?-0580?) | Cassiodore 0490?-0580? | Cassiodore | Cassiodore (v. 485-v. 580)) | Cassiodore (v. 485-v. 580)

modifier

▼ Plus de langues

Configurer

Langue	Libellé	Description	Également connu comme
français	Cassiodore (0490?-0580?)	Aucune description fournie	Cassiodore (0490?-0580?) Cassiodore 0490?-0580? Cassiodore Cassiodore (v. 485-v. 580)) Cassiodore (v. 485-v. 580)

Toutes les langues saisies

### Déclarations

nature de l'élément	être humain	modifier
	▼ 0 référence	+ ajouter une référence + ajouter une valeur

date de naissance	490 Grégorien	modifier
	qualité de l'information	prisé
	▼ 0 référence	+ ajouter une référence + ajouter une valeur

date de mort	580 Grégorien	modifier
	qualité de l'information	prisé
	▼ 0 référence	+ ajouter une référence

## Chosen label

### Alternative labels:

Spelling forms found in the source data and other datasets (BnF, wikidata etc.)

sex or gender		male	
		- 0 references	

Biblissima note		Dizionario Biografico degli Italiani, 21, 1978: Cassiodoro (Flavius Magnus Aurelius Cassiodorus Senator; più spesso detto Cassiodorus Senator o semplicemente Senator, che è nome proprio e non designazione di dignità)	
		- 0 references	
		A. Molinier, Les Sources de l'histoire de France - Des origines aux guerres d'Italie (1494): 224. Magnus Aurelius Cassiodorus Senator	
		- 0 references	

Biblissima Resources			
Biblissima Portal ID		pdatad89fe276e17712275adbc02eb985600cabfa7390	
		- 0 references	

Search Result in IIIF-Collections		Q2785	
		- 0 references	

Identifiers			
BnF ID		121582004	
		- 0 references	

## Bibliographic note

Provide clues to help for the identification of the person

## Links to Biblissima web pages

- Biblissima portal
- IIIF Collections

## Identifiers

Identifiers		
BnF ID	<a href="#">121582004</a>	<a href="#">edit</a>
	▼ 0 references	<a href="#">+ add reference</a>
		<a href="#">+ add value</a>
Library of Congress authority ID	<a href="#">names/n79082349</a>	<a href="#">edit</a>
	▼ 0 references	<a href="#">+ add reference</a>
		<a href="#">+ add value</a>
GND ID	<a href="#">118519514</a>	<a href="#">edit</a>
	▼ 0 references	<a href="#">+ add reference</a>
		<a href="#">+ add value</a>
CERL ID	<a href="#">cnp01302134</a>	<a href="#">edit</a>
	▶ 2 references	<a href="#">+ add value</a>
Wikidata ID	<a href="#">Q192365</a>	<a href="#">edit</a>
	▼ 0 references	<a href="#">+ add reference</a>
		<a href="#">+ add value</a>
SUDOC authorities ID	<a href="#">030097703</a>	<a href="#">edit</a>
	▼ 0 references	<a href="#">+ add reference</a>
		<a href="#">+ add value</a>
BNE ID	<a href="#">XX878991</a>	<a href="#">edit</a>
	▼ 0 references	<a href="#">+ add reference</a>

## Identifiers section

- Alignments to the identifier in the source database
- External alignments  
(BnF, LoC, Geonames, GND, CERL, BNE, IdRef, Wikidata...)

# API Endpoints to data.biblissima

## API Mediawiki/wikibase

<https://data.biblissima.fr/w/api.php>

## RDF data about en entity

Content negotiation:

<https://data.biblissima.fr/entity/Q2785>

With indication of format:

<https://data.biblissima.fr/w/Special:EntityData/Q2785.json>

## SPARQL Endpoint (in-house use only)

Querying RDF data with advanced queries

## Reconciliation API

Endpoints for single queries  
or query batches (fr|en):

<https://data.biblissima.fr/reconcile/fr/api>

<https://data.biblissima.fr/reconcile/en/api>

Graphical clients...



OpenRefine

# What is data reconciliation in OpenRefine?

- **Reconciliation** = Comparing/aligning a local dataset with an external source to identify correspondances and create actionable links from them.
- A semi-automated process: suggestions from the algorithm that have to be validated (manually or programmatically) or rejected.
- Developed as a [specification of a protocol](#) for matching data on the web, now it's own [W3C Entity Reconciliation Community Group](#).
- OpenRefine acts as a graphical **client** that can connect to any webservice that follows that protocol, offering a user-friendly interface to perform reconciliation operations...

The image displays two side-by-side screenshots of the OpenRefine web application, illustrating the reconciliation process.

**Screenshot 1 (Left): Reconcile column Billeissima**

This screenshot shows the "Reconcile" dialog for a column named "Billeissima". The "Reconcile each cell to an entity of one of these types:" dropdown is set to "Any Human". Below it, a list of entities is shown, including "Olympe", "descripteur", "Q304387", "aureve", "Q3000550", "manuscrit", "Q38110", "organisation", "Q38111", "personnel militaire", "Q294837", "religieux", "Q294862", and "type iconographique", "Q294876". The "Auto-match candidates with high confidence" checkbox is checked. The "Maximum number of candidates to return" dropdown is set to 5. The "Reconcile against type" and "Reconcile against no particular type" radio buttons are also visible.

**Screenshot 2 (Right): Reconcile bhma\_names**

This screenshot shows the results of a reconciliation operation for the "name" column. It lists 20 rows of data, each with a person's name and their ID. The names include "Jean (duc de Berry, 1340-1416)", "Charles V d'Anjou", "King Charles V of France", "Charles IV (roi de France, 1338-1380)", "Jeanne de Bourbon (1338-1378)", "Limburg brothers Paul, John and Herman", "Charles IV of France", "Charles IV (roi de France, 1338-1380)", "Jeanne de Bourbon (1338-1378)", "Francis Cobham (153-1592)", "Marguerite of France (reine de France, 1469-1524)", "Marguerite of France (Duchesse de Savoie, 1523-1574)", "Francis Cobham (153-1592)", "Claude of France (reine de France, 1469-1524)", "Marguerite of France (Duchesse de Savoie, 1523-1574)", "Francis Cobham (153-1592)", "De Limbourg (fille)", "Barthélémy d'Eck (14.-1476)", "Jean Cobham (143-1493)", "The Duke of Savoy", and "Charles de Lorraine Aumatte (duc d', 1670-1736)". The interface includes a "name" facet on the left and various filtering and search options at the top.

# A very quick introduction to OpenRefine

# OpenRefine

- Formerly known as Google Refine, openSource via Github since 2012
- Desktop app, running in your browser
- Community driven (See [Forum](#))
- A powerful and user-friendly tool to clean and transform data
- Looking like a spreadsheet, but works more like a database
- A dedicated language expression (GREL for “General Refine Expression Language)

## Data cleaning

The process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

## Data transformation

The process of converting data from one format or structure into another.

source:

<https://www.tableau.com/learn/articles/what-is-data-cleaning>

# Demo 1

<http://127.0.0.1:3333>

OpenRefine interface and main operations



A power tool for working with messy data.

Create project

Open project

Import project

Language settings



Version 3.8.2 [TRUNK]

Preferences  
Help  
About

### Create a project by importing data. What kinds of data files can I import?

TSV, CSV, \*SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with OpenRefine extensions.

Get data from

Locate one or more files on your computer to upload:

aucun fichier sélectionné

**Next »**

**This Computer**

[Web Addresses \(URLs\)](#)

[Clipboard](#)

[Database](#)

[Google Data](#)

[Create project](#)[Open project](#)[Import project](#)[Language settings](#)

Version 3.8.2 [TRUNK]

[Preferences](#)[Help](#)[About](#)

All

	Last modified	Name	Tags	Creator	Subject	Description	Row Count
X	About Aug 2, 2024 at 2:43:23 PM	test dataset					19
X	About Aug 2, 2024 at 2:40:43 PM	coll hist berry					111
X	About Aug 1, 2024 at 4:25:47 PM	obma_names					20
X	About Aug 1, 2024 at 4:24:23 PM	dataset ex 1 2 csv					16
X	About Jul 31, 2024 at 3:32:53 PM	dexex1_v2					16
X	About Jul 31, 2024 at 10:20:34 AM	Clipboard					17
X	About Jul 31, 2024 at 10:19:33 AM	dhex1					16
X	PM	dh2024 biblissima exercice 1 csv					16
X	PM	dh2024-ex-1					89
X	PM	dh-2024-descripteurs-icono					54
X	PM	bbma dh2024					10
X	0 PM	bbma_ex_3					10
X	41 AM	biblissima 0 xml					20
X	About Jul 27, 2024 at 4:34:05 PM	Biblissima-names					24
X	About Jul 26, 2024 at 7:26:53 PM	Biblissima-names					18
X	About Jul 26, 2024 at 5:27:20 PM	Biblissima-names					12
X	About Jul 26, 2024 at 5:07:50 PM	Biblissima-names					19
X	About Jul 26, 2024 at 5:06:56 PM	Biblissima-names					11
X	About Jul 26, 2024 at 3:23:57 PM	cooccurrence Bibliothèque de Jean duc de Berry csv					266
X	About Jul 25, 2024 at 6:09:44 PM	Presse-papier					31
X	About Jul 25, 2024 at 5:57:01 PM	B372616101 Res 3537 manifest json					395
X	About Jul 25, 2024 at 5:52:09 PM	B372616101 Res 3537 manifest json					395
X	About Jul 25, 2024 at 5:25:15 PM	manifest roi sport json					4
X	About Jul 25, 2024 at 3:42:12 PM	DH-2024-ms-varia					36

[Browse workspace directory](#)

Facet / Filter

Undo / Redo 29 / 29

Using facets and filters 

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

19 rows

Extensions VIB-Bits ▾ RDF Transform ▾ RDF ▾ Wikibase ▾

Show as: **rows** records

Show: 5 10 25 50 100 500 1000 rows

« first &lt; previous 1 next » last

	All	nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance
1.	Augustinus	lat		Augustinus	Augustin (saint, 0354-0430) Choose new match	Q8018	Augustin d'Hippone Choose new match	Q1966		Q1966	0354
2.	Girolamo Balbi	it		Girolamo Balbi	Girolamo Balbi (1460-1535) Choose new match	Q673596	Girolamo Balbi Choose new match	Q5960		Q5960	1460
3.	Boccaccio	it		Boccaccio	Boccace (1313-1375)	Q1402	Boccace	Q2564		Q2564	1313
4.	Guillaume Ier Briçonnet	fr		Guillaume Briçonnet							1445
5.	Guillaume Ier Briçonnet	fr		Guillaume Briçonnet							1445
6.	Leonardo Bruni	it		Leonardo Bruni						Q10361	1370
7.	Charles VIII roi de France	fr		Charles VIII roi de France						Q2991	1470
8.	Charles de Croÿ	fr		Charles de Croÿ	Charles de Croÿ (duc de, 1488-1612) Choose new match	Q673597	Charles III de Croÿ Choose new match	Q2891		Q2891	1560
9.	Cosimo De Medici	fr		Cosimo De Medici	Cosme de Médicis (1389-1464)	Q48544	Cosme de Médicis Choose new match	Q233185		Q233185	1389
10.	Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?

## Row view vs Record view

In **Record** view, OpenRefine can link together **multiple rows** as part of the same Record.

Facet / Filter

Undo / Redo 29 / 29

19 rows

Extensions VIB-Bits ▾ RDF Transform ▾ RDF ▾ Wikibase ▾

**Using facets and filters**

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

**1**

Click on arrow to reveal a Dropdown menu operating on the column it belongs to

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

All	nom_entite	langue	to_recon	to_recon_2	Facet
1.	Augustinus				<ul style="list-style-type: none"><li>Text facet</li><li>Numeric facet</li><li>Timeline facet</li><li>Scatterplot facet...</li><li>Custom text facet...</li><li>Custom numeric facet...</li><li>Customized facets</li></ul>
2.	Girolamo Balbi				<ul style="list-style-type: none"><li>Sort...</li><li>View</li><li>Reconcile</li></ul>
3.	Boccaccio				
4.	Guillaume ler Brionnet				
5.	Guillaume ler Brionnet	fr	Guillaume ler Brionnet	Guillaume ler Brionnet	<ul style="list-style-type: none"><li><input checked="" type="checkbox"/> Create new item</li><li>Search for match</li></ul>
6.	Leonardo Bruni	it	Leonardo Bruni	Leonardo Bruni (1370-1444)	<ul style="list-style-type: none"><li><input checked="" type="checkbox"/> Create new item</li><li>Choose new match</li></ul>
7.	Charles VIII roi de France	fr	Charles VIII roi de France	Charles VIII (roi de France, 1470-1498)	<ul style="list-style-type: none"><li>Choose new match</li></ul>
8.	Charles de Croÿ	fr	Charles de Croÿ	Charles Croÿ (duc de, 1560-1612)	<ul style="list-style-type: none"><li>Choose new match</li></ul>
9.	Cosimo De Medici	fr	Cosimo De Medici	Cosme de Médicis (1389-1464)	<ul style="list-style-type: none"><li>Choose new match</li></ul>
10.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296)	<ul style="list-style-type: none"><li>Choose new match</li></ul>

**2**

Facet > Text facet  
groups together  
the entries with the  
same value

Facet / Filter

Undo / Redo 29 / 29

### 3 matching rows (19 total)

Extensions VIB-Bits ▾ RDF Transform ▾ RDF ▾ Wikibase ▾

Refresh

Reset all Remove all

langue change invert reset

3 choices Sort by: name count

Cluster

fr 45

it 3

lat 1

exclude

Facet by choice counts

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first < previous 1 next » last »

All	nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance	date_mort	http
2.	Girolamo Balbi	it		Girolamo Balbi	Girolamo Balbi (1460-1535) Choose new match	Q673596	Girolamo Balbi Choose new match	Q5960		Q5960 1460		1535 http
3.	Boccaccio	edit	it	Boccaccio	Boccace (1313-1375) Choose new match	Q1402	Boccace Choose new match	Q2564		Q2564 1313		1375 http
6.	Leonardo Bruni	it		Leonardo Bruni	Leonardo Bruni (1370-1444) Choose new match	Q314488	Leonardo Bruni Choose new match	Q10361		Q10361 1370		1444 http

3

Navigate in the dataset with facets

Facet / Filter

Undo / Redo 38 / 38

19 rows

Extensions

VIB-Bits

RDF Transform

RDF

Wikibase

Using facets and filters 

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

Show as: [rows](#) [records](#)Show: [5](#) [10](#) [25](#) [50](#) [100](#) [500](#) [1000](#) rows

« first &lt; previous 1 next &gt; last »

All	nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance
1. Augustinus	lat		Augustinus	Augustin (saint, 0354-0430) Choose new match	Q8018	Augustin d'Hippone Choose new match	Q1966		Q1966	0354
2. Girolamo Balbi	it		Girolamo Balbi	Girolamo Balbi Choose new match		Girolamo Balbi Choose new match	Q5960		Q5960	1460
3. Boccaccio	it		Boccaccio	Boccaccio Choose new match		Boccaccio Choose new match	Q2564		Q2564	1313
4. Guillaume Ier Brionnet	fr		Guillaume Ier Brionnet	Guillaume Ier Brionnet Choose new match		Guillaume Ier Brionnet Choose new match				1445
5. Guillaume Ier Brionnet	fr		Guillaume Ier Brionnet	Guillaume Ier Brionnet Choose new match		Guillaume Ier Brionnet Choose new match				1445
6. Leonardo Bruni	it		Leonardo Bruni	Leonardo Bruni Choose new match		Leonardo Bruni Choose new match	Q10361		Q10361	1370
7. Charles VIII roi de France	fr		Charles VIII roi de France	Charles VIII (roi de France, 1470-1498) Choose new match	Q134452	Charles VIII Choose new match	Q2991		Q2991	1470
8. Charles de Croÿ	fr		Charles de Croÿ	Charles Croÿ (duc de, 1560-1612) Choose new match	Q676807	Charles III de Croÿ Choose new match	Q2891		Q2891	1560
9. Cosimo De Medici	fr		Cosimo De Medici	Cosme de Médicis (1389-1464) Choose new match	Q48544	Cosme de Médicis Choose new match	Q233185		Q233185	1389
10. Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
11. Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
12. Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296)	Q2329414	Guillaume V Durand	Q6361		Q6361	1230?

1'

Star the rows you  
want to process  
later

Facet / Filter

Undo / Redo 38 / 38

19 rows

Extensions VIB-Bits ▾ RDF Transform ▾ RDF ▾ Wikibase ▾

Using facets and filters 

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started?  
[Watch these screencasts](#)

Show as: [rows](#) [records](#) Show: 5 10 25 50 100 500 1000 rows

« first &lt; previous 1 next » last »

All	nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance
Transform...		us	lat	Augustinus	Augustin (saint, 0354-0430) Choose new match	Q8018	Augustin d'Hippone Choose new match	Q1966	Q1966	0354
Edit all columns										
Facet	Facet by star									
Edit rows	Facet by flag									
Edit columns	Facet by blank (null or empty string)									
View	Blank values per column									
	4.	Guillaum Briçonne								1445
	5.	Guillaum Briçonne								1445
	6.	Leonardo Bruni	it	Leonardo Bruni	Leonardo Bruni Choose new match		Q10361		Q10361	1370
	7.	Charles VIII roi de France	fr	Charles VIII roi de France	Charles VIII (roi de France, 1470-1498) Choose new match	Q134452	Charles VIII Choose new match	Q2991	Q2991	1470
	8.	Charles de Croÿ	fr	Charles de Croÿ	Charles Croÿ (duc de, 1560-1612) Choose new match	Q676807	Charles III de Croÿ Choose new match	Q2891	Q2891	1560
	9.	Cosimo De Medici	fr	Cosimo De Medici	Cosme de Médicis (1389-1464) Choose new match	Q48544	Cosme de Médicis Choose new match	Q233185	Q233185	1389
	10.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361	Q6361	1230?
	11.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361	Q6361	1230?
	12.	Guillaume	fr	Guillaume	Guillaume Durand (1230?-1296)	Q2329414	Guillaume V	Q6361	Q6361	1230?

2'

Facet &gt; Facet by star

Facet / Filter Undo / Redo 38 / 38

Refresh Reset all Remove all

Starred Rows change  
2 choices Sort by: name count  
false 16 true 3

Facet by choice counts

19 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows « first < previous 1 next » last »

Extensions VIB-Bits RDF Transform RDF Wikibase

3'  
Use “true” et “false” links retrieve the starred rows or exclude them

					Current Wikidata	wd	Qid	latitude	test	date_naissance
1.					Augustin d'Hippone Choose new match	Q1966			Q1966	0354
2.					Girolamo Balbi Choose new match	Q5960			Q5960	1460
3.					Boccace Choose new match	Q2564			Q2564	1313
4.					Guillaume Ier de Nîmes Choose new match					1445
5.	Guillaume Ier Briçonnet	fr	Guillaume Ier Briçonnet	Guillaume Ier Briçonnet <input checked="" type="checkbox"/> Create new item Search for match						1445
6.	Leonardo Bruni	it	Leonardo Bruni	Leonardo Bruni (1370-1444) Choose new match	Q314488	Leonardo Bruni Choose new match	Q10361		Q10361	1370
7.	Charles VIII roi de France	fr	Charles VIII roi de France	Charles VIII (roi de France, 1470-1498) Choose new match	Q134452	Charles VIII Choose new match	Q2991		Q2991	1470
8.	Charles de Croÿ	fr	Charles de Croÿ	Charles Croÿ (duc de, 1560-1612) Choose new match	Q676807	Charles III de Croÿ Choose new match	Q2891		Q2891	1560
9.	Cosimo De Medici	fr	Cosimo De Medici	Cosme de Médicis (1389-1464) Choose new match	Q48544	Cosme de Médicis Choose new match	Q233185		Q233185	1389
10.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
11.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
12.	Guillaume	fr	Guillaume	Guillaume Durand (1230?-1296)	Q2329414	Guillaume V	Q6361		Q6361	1230?

Facet / Filter

Undo / Redo 29 / 38

Refresh

Reset all Remove all

**langue** change  
3 choices Sort by: name count Cluster

- fr 15
- it 3
- lat 1

Facet by choice counts

**edit include**

19 rows

Extensions VIB-Bits ▾ RDF Transform ▾ RDF ▾ Wikibase ▾

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first &lt; previous 1 next » last »

<input type="checkbox"/>	All	<input type="checkbox"/>	nom_entite	<input type="checkbox"/>	langue	<input type="checkbox"/>	to_recon	<input type="checkbox"/>	to_recon_2	<input type="checkbox"/>	identifiant Wikidata	<input type="checkbox"/>	wd	<input type="checkbox"/>	
1.	Augustinus	lat	Augustinus	Augustin (saint, 0354-0430) Choose new match	Q8018	Augustin d'Hippone Choose new match	Q1966			Q1966	0354				
2.	Girolamo Balbi	it	Girolamo Balbi	Girolamo Balbi (1460-1535) Choose new match	Q673596	Girolamo Balbi Choose new match	Q5960			Q5960	1460				
3.	Boccaccio	it	Boccaccio	Boccace (1313-1375) Choose new match	Q1402	Boccace Choose new match	Q2564			Q2564	1313				
4.	Guillaume Ier Briçonnet	fr	Guillaume Ier Briçonnet	Guillaume Ier Briçonnet <input checked="" type="checkbox"/> Create new item	Q19629233	Guillaume Ier de Nîmes Choose new match					1445				
5.	Guillaume II Briçonnet										1445				
6.	Leonardo Bruni										Leonardo Bruni Choose new match	Q10361		Q10361	1370
7.	Charles VIII de France										Charles VIII Choose new match	Q2991		Q2991	1470
8.	Charles III de Croÿ										Charles III de Croÿ Choose new match	Q2891		Q2891	1560
9.	Cosimo de' Medici										Cosme de Médicis Choose new match	Q233185		Q233185	1389
10.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361			Q6361	1230?				
11.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361			Q6361	1230?				
12.	Guillaume	fr	Guillaume	Guillaume Durand (1230?-1296)	Q2329414	Guillaume V	Q6361			Q6361	1230?				

4'

Hover on the line of a facet to reveal buttons and click on edit

Facet / Filter

Undo / Redo 29 / 38

Refresh

Reset all Remove all

 langue change

3 choices Sort by: name count

Cluster

fr 15

it 3

lat 1

Facet by choice counts

19 rows

Extensions

VIB-Bits

RDF Transform

RDF

Wikibase

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

« first &lt; previous 1 next &gt; last »

All	nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance
1.	Augustinus	lat	Augustinus	Augustin (saint, 0354-0430) Choose new match	Q8018	Augustin d'Hippone Choose new match	Q1966		Q1966	0354
2.	Guillaume Ier de Nîmes	fr	Guillaume Ier de Nîmes	Guillaume Ier de Nîmes Choose new match	Q673596	Girolamo Balbi Choose new match	Q5960		Q5960	1460
3.	Guillaume Ier Brionnet	fr	Guillaume Ier Brionnet	Guillaume Ier Brionnet <input checked="" type="checkbox"/> Create new item Search for match	Q19629233	Guillaume Ier Brionnet Choose new match	Q2564		Q2564	1313
4.	Guillaume Ier Brionnet	fr	Guillaume Ier Brionnet	Guillaume Ier Brionnet <input checked="" type="checkbox"/> Create new item Search for match	Q19629233	Guillaume Ier Brionnet Choose new match				1445
5.	Guillaume Ier Brionnet	fr	Guillaume Ier Brionnet	Guillaume Ier Brionnet <input checked="" type="checkbox"/> Create new item Search for match						1445
6.	Leonardo Bruni	it	Leonardo Bruni	Leonardo Bruni (1370-1444)	Q314488	Leonardo Bruni Choose new match	Q10361		Q10361	1370
7.	Charles VIII	de	Charles VIII	Charles VIII Choose new match		Charles VIII Choose new match	Q2991		Q2991	1470
8.	Charles III de Croÿ	de	Charles III de Croÿ	Charles III de Croÿ Choose new match		Charles III de Croÿ Choose new match	Q2891		Q2891	1560
9.	Cosme de Médicis	Médicis	Cosme de Médicis	Cosme de Médicis Choose new match		Cosme de Médicis Choose new match	Q233185		Q233185	1389
10.	Guillaume Durand	Dur	Guillaume Durand	Guillaume Durand Choose new match		Guillaume V Durand Choose new match	Q6361		Q6361	1230?
11.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
12.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?

5'

You can modify the label  
e.g. "italian" instead of "it"

Facet / Filter

Undo / Redo 38 / 38

Refresh

Reset all Remove all

## Starred Rows

change

2 choices Sort by: name count

false 16

true 3

Facet by choice counts

19 rows

Extensions

VIB-Bits

RDF Transform

RDF

Wikibase

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

&lt; first &lt; previous 1 next last

					Current Wikidata	wd	Qid	latitude	test	date_naissance
1.					Augustin d'Hippone Choose new match	Q1966			Q1966	0354
2.					Girolamo Balbi Choose new match	Q5960			Q5960	1460
3.					Boccace Choose new match	Q2564			Q2564	1313
4.					Guillaume Ier de Nîmes Choose new match					1445
5.	Guillaume Ier Brionnet	fr	Guillaume Ier Brionnet	Guillaume Ier Brionnet <input checked="" type="checkbox"/> Create new item Search for match						1445
6.	Leonardo Bruni	it	Leonardo Bruni	Leonardo Bruni (1370-1444) Choose new match	Q314488	Leonardo Bruni Choose new match	Q10361		Q10361	1370
7.	Charles VIII roi de France	fr	Charles VIII roi de France	Charles VIII (roi de France, 1470-1498) Choose new match	Q134452	Charles VIII Choose new match	Q2991		Q2991	1470
8.	Charles de Croÿ	fr	Charles de Croÿ	Charles Croÿ (duc de, 1560-1612) Choose new match	Q676807	Charles III de Croÿ Choose new match	Q2891		Q2891	1560
9.	Cosimo De Medici	fr	Cosimo De Medici	Cosme de Médicis (1389-1464) Choose new match	Q48544	Cosme de Médicis Choose new match	Q233185		Q233185	1389
10.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
11.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
12.	Guillaume	fr	Guillaume	Guillaume Durand (1230?-1296)	Q2329414	Guillaume V	Q6361		Q6361	1230?

1"

click on the undo/redo tab to access your operation history

Extract... Apply...

Filter

23. Edit single cell on row 19, column langue

24. Reconcile cells in column to\_recon\_2 to type Q168

25. Create new column test based on column to\_recon\_2 by filling 12 rows with cell.recon.match.id

26. Extend data at index 4 based on column to\_recon\_2

27. Edit single cell on row 4, column identifiant Wikidata

28. Create new column wd based on column identifiant Wikidata by filling 13 rows with gref:value

29. Reconcile cells in column wd to type Q5

30. Star row 4

31. Star row 6

32. Star row 2

33. Unstar row 2

34. Star row 1

35. Unstar row 4

36. Unstar row 6

37. Star row 7

38. Star row 3

Show as: rows records

Show: 5 10 25 50 100 500 1000 rows

&lt; first &lt; previous 1 next &gt; last &gt;

All	nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance
1. Augustinus	lat	Augustinus	Augustin (saint, 0354-0430)	Choose new match	Q8018	Augustin d'Hippone	Q1966		Q1966	0354
2. Girolamo Balbi	it	Girolamo Balbi	Girolamo Balbi (1460-1535)	Choose new match	Q673596	Girolamo Balbi	Q5960		Q5960	1460
3. Boccaccio	it	Boccaccio	Boccace (1313-1375)	Choose new match	Q1402	Boccace	Q2564		Q2564	1313
4. Guillaume Ier Briçonnet	fr	Guillaume Ier Briçonnet	Guillaume Ier Briçonnet	<input checked="" type="checkbox"/> Create new item <input checked="" type="checkbox"/> Search for match	Q19629233	Guillaume Ier de Nîmes	Choose new match			1445
5. Guillaume Ier Briçonnet	fr	Guillaume Ier Briçonnet	Guillaume Ier Briçonnet	<input checked="" type="checkbox"/> Create new item <input checked="" type="checkbox"/> Search for match						1445
6. Leonardo Bruni	it	Leonardo Bruni	Leonardo Bruni (1370-1444)	Choose new match	Q314488	Leonardo Bruni	Q10361		Q10361	1370
7. Charles VIII roi de France	fr	Charles VIII roi de France	Charles VIII (roi de France, 1470-1498)		Q134452	Charles VIII	Q2991		Q2991	1470
8. Charles	fr	Charles	Charles		Q676807	Charles III de Croÿ	Q2891		Q2891	1560
9. Cosimo Medici	fr	Cosimo Medici	Cosimo Medici		Q48544	Cosme de Médicis	Q233185		Q233185	1389
10. Guillaume Durand	fr	Guillaume Durand	Guillaume Durand		Q2329414	Guillaume V Durand	Q6361		Q6361	1230?
11. Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296)		Q2329414	Guillaume V Durand	Q6361		Q6361	1230?
12. Guillaume Durand	fr	Guillaume Durand	Guillaume Durand		Q2329414	Guillaume V Durand	Q6361		Q6361	1230?

2"  
Select the operations you want to undo

Facet / Filter

Undo / Redo 29 / 38

Refresh

Reset all Remove all

**langue** change  
 3 choices Sort by: name count Cluster

fr 15

it 3

lat 1

Facet by choice counts

19 rows

Extensions

VIB-Bits

RDF Transform

RDF

Wikibase

Show as: [rows](#) [records](#)Show: [5](#) [10](#) **25** [50](#) [100](#) [500](#) [1000](#) rows

« first

&lt; previous

1

next &gt;

last »

	All	nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance
1.	Facet			Augustinus	Augustin (saint, 0354-0430) Choose new match	Q8018	Augustin d'Hippone Choose new match	Q1966		Q1966	0354
2.	Text filter										
3.	Edit cells			Transform...		ihi (1460-1535)	Q673596	Girolamo Salvi Choose new match	Q5960		Q5960 1460
4.	Edit column			Common transforms		Trim leading and trailing whitespace					
	Transpose			Fill down		Collapse consecutive whitespace					
	Sort...			Blank down							
	View			Unescape HTML entities							
	Reconcile			Replace smart quotes with ASCII							
				To titlecase							
5.	Guillaume Ier Briçonnet	fr		Cluster and edit...		To uppercase					1445
6.	Leonardo Bruni	it		Replace...		To lowercase					
					Bruni	Choose new match					
7.	Charles VIII roi de France	fr		Charles VIII roi de France	Charles VIII (1498) Choose new match		Leonardo Bruni Choose new match	Q10361		Q10361	1370
8.	Charles de Croÿ	fr		Charles de Croÿ	Charles Croÿ (1612) Choose new match		Charles VIII Choose new match	Q2991		Q2991	1470
9.	Cosimo De Medici	fr		Cosimo De Medici	Cosme de Médicis (1389-1464) Choose new match	Q48544	Charles III de Croÿ Choose new match	Q2891		Q2891	1560
10.	Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
11.	Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?
12.	Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296)	Q2329414	Guillaume V Durand Choose new match	Q6361		Q6361	1230?

## Facet / Filter

Undo / Redo 29 / 38



19 rows

## Extensions

VIB-Bits ▾

## RDF Transform

RDF ▾

Wikibase ▾

Refresh

[Reset all](#) [Remove all](#)

  langue change

change

### 3 choices Sort by: Name

11

It 3

lat

## Facet by choice counts

Show as:		rows	records	Show: 5 10 25 50 100 500 1000 rows		< first		< previous		1	next >	last »
All		nom_entite	langue	to_recon	to_recon_2	identifiant Wikidata	wd	Qid	latitude	test	date_naissance	
1.	Facet			Augustinus	Augustin (saint, 0354-0430) Choose new match	Q8018	Augustin d'Hippone Choose new match	Q1966		Q1966	0354	
2.	Edit cells			Transform...		Illi (1460-1535)	O673596	Girolamo Balbi choose new match	Q5960		Q5960 1460	
	Edit column			Common transforms		Trim leading and trailing whitespace		Boccace choose new match	Q2564		Q2564 1313	
	Transpose			Fill down		Collapse consecutive whitespace						
3.	Sort...			Blank down		Unescape HTML entities						
4.	View			Split multi-valued cells...		Replace smart quotes with ASCII		Guillaume le de Nîmes choose new match			1445	
	Reconcile			Join multi-valued cells...		To titlecase						
5.	Guillaume le Briçonnet	fr		Cluster and edit...		To uppercase					1445	
6.	Leonardo Bruni	it		Replace...		To lowercase		Leonardo Bruni choose new match	Q10361		Q10361 1370	
				Bruni	Choose new match	To number						
						To date						
7.	Charles VIII roi de France	fr		Charles VIII roi de France	Charles VIII 1498 Choose new match	To text		Charles VIII choose new match	Q2991		Q2991 1470	
8.	Charles de Croÿ	fr		Charles de Croÿ	Charles Croÿ 1612 Choose new match	To null		Charles III de Croÿ choose new match	Q2891		Q2891 1560	
						To empty string						
9.	Cosimo De Medici	fr		Cosimo De Medici	Cosme de Médicis (1389-1464) Choose new match	Q48544	Cosme de Médicis choose new match	Q233185		Q233185	1389	
10.	Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand choose new match	Q6361		Q6361	1230?	
11.	Guillaume Durand	fr		Guillaume Durand	Guillaume Durand (1230?-1296) Choose new match	Q2329414	Guillaume V Durand choose new match	Q6361		Q6361	1230?	
12.	Guillaume	fr		Guillaume	Guillaume Durand (1230?-1296)	Q2329414	Guillaume V	Q6361		Q6361	1230?	

Facet / Filter

Undo / Redo 29 / 38

19 rows

Extensions VIB-Bits ▾ RDF Transform ▾ RDF ▾ Wikibase ▾

Refresh

Reset all Remove all

langue

change

3 choices Sort by: name count

Cluster

fr 15

it 3

lat 1

Facet by choice counts

## Add column based on column date\_naissance

New column name cleaned\_date\_naissance

On error  set to blank  store error  copy value from original column

Expression

Language General Refine Expression Language (GREL)

value.replace(/^0|\?/, "")

No syntax

Preview

History

Starred

Help

row	value	value.replace(/^0 \?/, "")
1.	0354	354
2.	1460	1460
3.	1313	1313
4.	1445	1445
5.	1445	1445
6.	1370	1370
7.	1370	1370

OK

Cancel

10.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296)	Q2329414	Choose new match	Guillaume V Durand	Q6361	Q6361	1230?
11.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296)	Q2329414	Choose new match	Guillaume V Durand	Q6361	Q6361	1230?
12.	Guillaume Durand	fr	Guillaume Durand	Guillaume Durand (1230?-1296)	Q2329414	Choose new match	Guillaume V Durand	Q6361	Q6361	1230?

# Demo 2

<http://127.0.0.1:3333>

Reconciliation with Wikidata

# **Exercises link**

<https://github.com/emmamorlock/workshop/>

Go to ‘exercises’ folder

# **Exercise 1**

Reconcile a dataset  
against datablissima

~ 30 min

## **Exercise 2**

Improve the results

# **Exercise 3**

## Export an enriched dataset in XML/TEI

# Conclusion & discussion

# **Thank you!**

Visit <https://projet.biblissima.fr/>

# Credits

<https://portail.biblissima.fr/>

<https://data.biblissima.fr/w/Accueil>

BOIS K., FRUNZEANU E., ROBINEAU R., *Gérer, partager et interroger grâce à Wikibase des référentiels d'autorité pour la recherche : l'exemple de data.biblissima*, « Atelier Wikidata », Atelier Données de la Mission pour les Initiatives Transverses et Interdisciplinaires du CNRS, 21 avril 2022. Disponible sur <https://doi.org/10.5281/zenodo.6581780>

BOIS K., FRUNZEANU E., ROBINEAU R., *The Biblissima Authority File of Geographical Names*, Atelier Campus Condorcet « Référentiels géo-historiques sémantisés pour les humanités », Ecole nationale des chartes, 14 mai 2019. Disponible sur <https://projet.biblissima.fr/fr/biblissima-authority-file-geographical-names>

BOIS K., FRUNZEANU E., ROBINEAU R., *Les référentiels Biblissima : épine dorsale du portail Biblissima et de IIIF-Collections*, Séminaire sur l'étude des provenances dans les bibliothèques territoriales françaises (CERL, MCC, Bibliothèque municipale de Lyon), Lyon, 3 avril 2019. Disponible sur <https://projet.biblissima.fr/fr/referentiels-biblissima-epine-dorsale-portail-biblissima-iiif-collections>

FRUNZEANU E., ROBINEAU R., *What's in a Name? Text and Image for Indexing Prosopographical Data*, Ecole d'été « Bibliotheca Digitalis. Reconstitution de réseaux culturels au début des Temps modernes. Des sources primaires aux données », Le Mans, 5 juillet 2017. Disponible sur <https://projet.biblissima.fr/fr/what-is-in-a-name-text-image-indexing-prosopographical-data>