

# Gérer, partager et interroger grâce à Wikibase des référentiels d'autorité pour la recherche

*Webinaire Wikidata - GT “Atelier Données” MITI CNRS*

Régis ROBINEAU  
Coordinateur de l'équipe Portail Biblissima+

21 avril 2022

Biblissima 



# Biblissima+ (2021-2029)

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



# Biblissima+ (2021-2029)

---

- 17 partenaires, ~12 millions €
- Infrastructure numérique multipolaire de recherche et de service consacrée à **l'histoire de la transmission des textes anciens**
- Portail national de ressources numériques interopérables



ARCHIVES  
NATIONALES



centre  
Michel  
de  
Bouard  
CRAHOM



crh

CAMPUS  
CONDORCET  
GRAND EQUIPEMENT  
DOCUMENTAIRE

HISOMA  
Histoire et sources des mondes antiques

IRHT  
Institut de recherche  
et d'histoire des textes

P D N  
—  
M R S H

persée



FranceArchives  
PORTAIL NATIONAL DES ARCHIVES

T E K L I A



# Biblissima+

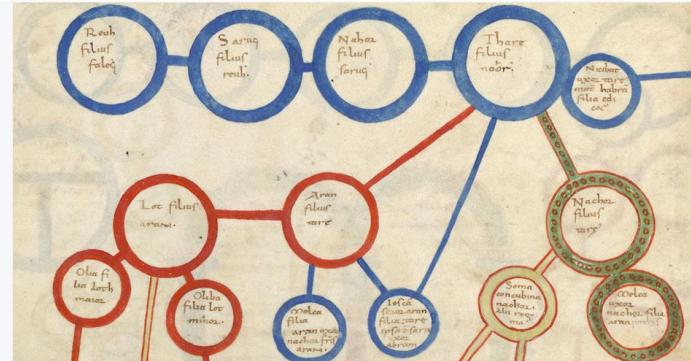
Observatoire des cultures écrites anciennes, de l'argile à l'imprimé

Biblissima+ est une infrastructure numérique multipolaire de recherche fondamentale et de service consacrée à l'histoire de la transmission des textes anciens, de l'Antiquité à la Renaissance en Orient comme en Occident.

## De Biblissima à Biblissima+

L'infrastructure Biblissima+ (2021-2029) s'articule autour de **7 domaines d'expertise ou « clusters »** regroupant chercheurs, conservateurs et ingénieurs. Ils forment autant de pôles de compétences à l'échelle nationale autour de l'étude des textes anciens et de leur transmission à travers l'Histoire...

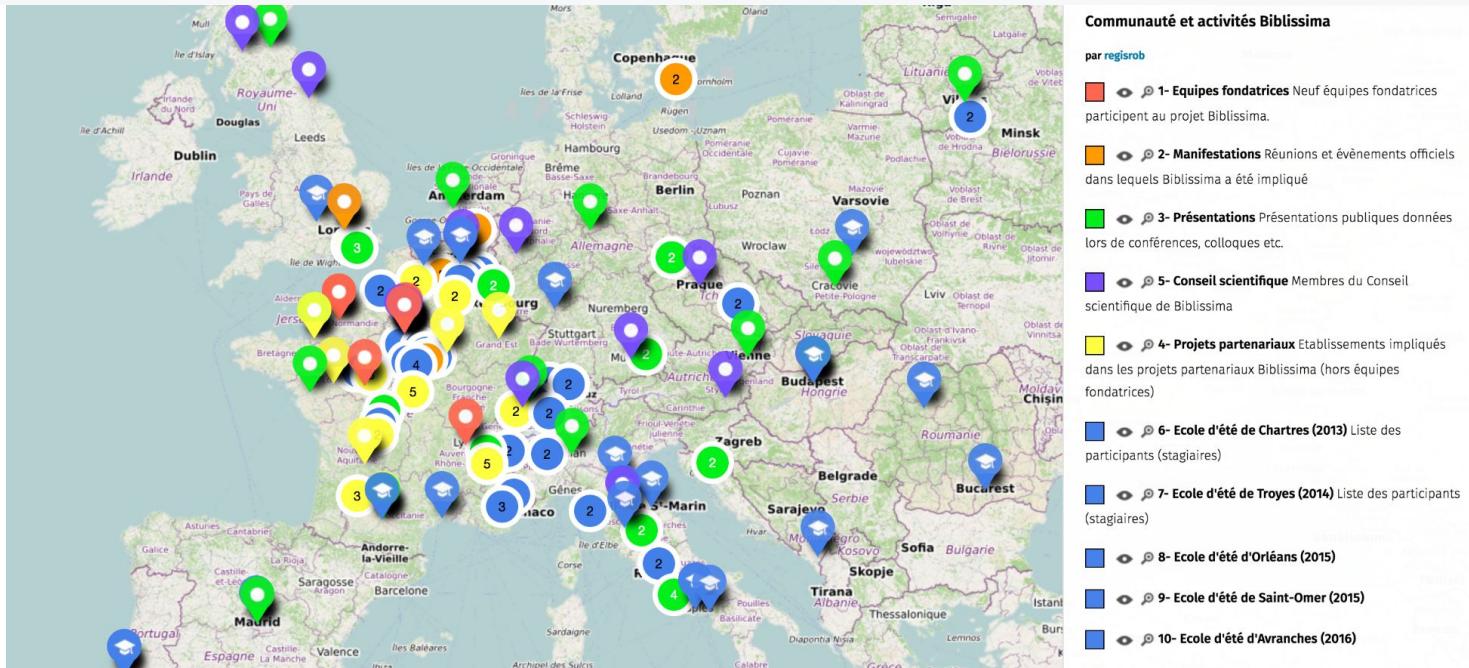
[En savoir plus](#)



Site d'information de Biblissima+ : [projet.biblissima.fr](http://projet.biblissima.fr)

# Socle de Biblissima+

→ Biblissima (2012-2021) et son réseau : 67 projets / 70 partenaires



# Socle de Biblissima+

---

- Infrastructure de mise en interopérabilité des données :
  - Portail Biblissima (2017)
    - propulsé par *Cubicweb*
  - IIIF-Collections (2018)
    - application PHP basée sur *ElasticSearch*
  - Référentiels d'autorité (2019)
    - instance *Wikibase*



Tout Biblissima

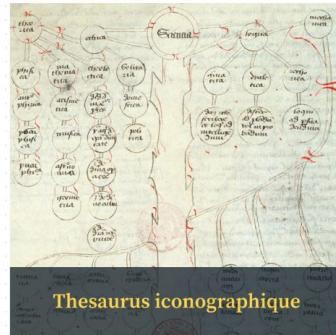
Rechercher une œuvre, une personne, un lieu, une cote, une enluminure...



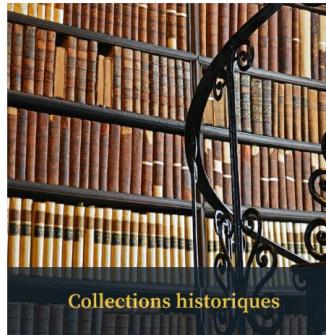
Bibliothèque virtuelle des bibliothèques, ce portail vous invite à découvrir l'histoire d'une partie des textes et livres qui ont été écrits, traduits, enluminés, collectionnés ou inventoriés depuis l'Antiquité jusqu'au XVIII<sup>e</sup> siècle.

[A propos de ce portail](#)

## Rechercher par type de données



Thesaurus iconographique



Collections historiques



Manuscrits et imprimés anciens

*Portail Biblissima et ses ~650 000 pages : [portail.biblissima.fr](http://portail.biblissima.fr)*



# Search for interoperable digitized manuscripts and rare books...

This prototype application allows you to search across IIIF-compliant manuscripts and rare books dated before 1800 coming from many digital libraries in the world. It is a work in progress, the platform is updated and enriched on a regular basis. [Read more...](#)

## Browse collections



Gallica (Bibliothèque nationale de France)

50149 manifests

Digitized manuscripts from Gallica, the digital library of the National Library of France (BnF)

[Browse collection](#)

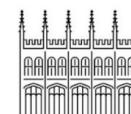


BVMM (IRHT-CNRS)

8095 manifests

Digital library of medieval manuscripts (IRHT-CNRS)

[Browse collection](#)



BODLEIAN LIBRARIES

Digital Bodleian (Oxford University)

5882 manifests

Bodleian Libraries' digital collections (Oxford)

[Browse collection](#)

Moteur IIIF-Collections et ses 24 bibliothèques numériques : [iiif.biblissima.fr/collections](http://iiif.biblissima.fr/collections)



## Accueil

Cette page contient des modifications qui ne sont pas marquées à traduire.

Autres langues : [English](#) · [français](#)

### Sommaire [masquer]

- 1 Référentiels d'autorité Biblissima
  - 1.1 Référentiels disponibles
  - 1.2 Référentiels à venir
  - 1.3 Accès aux données et API
  - 1.4 Licence des données

## Référentiels d'autorité Biblissima

Nombre total d'items présents dans les référentiels : **294 446** (contenant au total plus de 12 millions de données)

Cette plateforme met à disposition sous forme de données structurées les **référentiels d'autorités de Biblissima**. Sa réalisation s'inscrit dans le cadre du programme Biblissima, l'observatoire du patrimoine écrit du Moyen Âge et de la Renaissance du 8e au 18e siècle.

Ces référentiels d'autorité sont constitués à partir des données collectées et traitées dans le cadre de Biblissima.

Ils constituent l'épine dorsale du [portail Biblissima](#) et du prototype [IIIF Collections of Manuscripts and Rare Books](#).

Leur création est le fruit d'un travail d'identification, de désambiguisation, d'harmonisation et d'alignement d'entités nommées issues d'un large éventail de sources de données (bibliothèques numériques, catalogues et bases de données spécialisées portant sur les manuscrits et imprimés anciens). Cet effort est mené depuis les débuts de Biblissima en 2013.

Pour toute question, ou si vous souhaitez contribuer, vous pouvez nous contacter à l'adresse [team@biblissima-condorcet.fr](mailto:team@biblissima-condorcet.fr)

**Avertissement :** Dans sa première phase, cette plateforme est conçue comme un outil de gestion, de publication et d'exposition des référentiels d'autorité de Biblissima. Son but premier est de constituer un "hub" d'identifiants et de données structurées, accessibles, interopérables et réutilisables.

Ce hub joue un rôle central pour l'interopérabilité des ressources traitées dans le cadre de Biblissima, tout en permettant à d'autres projets de bénéficier du travail de réconciliation et d'harmonisation effectué en ce sens. Les données sont en effet accessibles de façon automatisée par divers mécanismes destinés

*Référentiels d'autorité data.biblissima et ses ~300 000 entités : [data.biblissima.fr](http://data.biblissima.fr)*

# Types de données traités par Biblissima

---

- Cotes de manuscrits et imprimés anciens
- Personnes physiques
- Organisations (dont établissements de conservation)
- Noms géographiques
- Oeuvres textuelles
- Descripteurs iconographiques



# Périmètre élargi avec Biblissima+

---

- Inscriptions épigraphiques
- Sceaux
- Monnaies
- Reliures



# Traitement des données d'autorité

---

- **Réconciliation** : identifier, désambiguïser et “clusteriser” des entités nommées
- **Alignement** vers des fichiers d'autorité de bibliothèques et d'autres jeux de données du *Linked Open Data*
- Forger des **identifiants** uniques et stables

# La plateforme data.biblissima

---

- Mise en ligne en mars 2019 : instance Wikibase gérée en interne
- Conçue comme un hub pour gérer et partager des données d'autorité
  - données sous Licence Ouverte / Open License (Etalab)
- Alimentée graduellement au fil du traitement des jeux de données
- Ouverte à des contributeurs extérieurs (sur demande)

# Pourquoi Wikibase ?

---

- Technologie wiki : nativement collaborative, versionnage
- Gestion d'identifiants URIs (y compris fusion/redirection)
- Exports RDF et négociation de contenu (compatibilité “*Linked Data*”)
- Formulaires faciles à prendre en main
- Points d'accès pour des machines : API Web + SPARQL
- GUI SPARQL (Wikidata Query Service - WDQS)
- Communauté d'utilisateurs :
  - Wikimedia (Wikidata, Commons), [Wikibase Stakeholder Group](#), FactGrid, NFDI4Culture, EU, ABES, BnF...

# Pourquoi pas Wikidata ?

---

- Besoin d'une plateforme autonome, tant au niveau technique qu'organisationnel
  - question de visibilité du projet, de son identité
- Besoin de souplesse au niveau du modèle de données
- Risque de bruit (cf. question du périmètre)

# Publication des référentiels Biblissima

---

- ✓ Personnes physiques
- ✓ Organisations (dont établissements de conservation)
- ✓ Noms géographiques
- ✓ Cotes de manuscrits et imprimés anciens
- ⚠ Oeuvres
- ⚠ Descripteurs iconographiques

~300 000 entités publiées sur [data.biblissima.fr](https://data.biblissima.fr) (avril 2022)

# Modèle de données

---

## → Modèle *ad hoc* (principe de Wikibase)

- basé sur les pratiques des fichiers d'autorité des bibliothèques
- une partie des propriétés reprises de Wikidata et importées initialement via *WikibaseImport*
- alignement systématique des classes et propriétés “locales” avec Wikidata (si équivalence) ou d'autres ontologies :
  - via les propriétés “*equivalent property*” ou “*equivalent class*” (vers Wikidata, Schema.org, SKOS, Foaf, DC etc.)
- évolution du modèle au fil de l'eau en fonction des besoins

# Modèle de données

---

- 283 propriétés
  - dont 182 propriétés d'identifiant (ex. identifiant CERL)
- 88 classes (types d'item)
  - dont **12 principales** (être humain, organisation, manuscrit, imprimé, oeuvre, lieu géographique etc.)



# Volumes de données (avril 2022)

---

294 447 entités, dont :

- **36 356 personnes** (798 familles). Ex. [Item:Q9866](#)
- **5780 organisations**. Ex. [Item:Q81208](#)
- **6550 noms géographiques**. Ex. [Item:Q27525](#)
- **211 695 cotes**. Ex. [Item:Q58639](#)
- **4603 œuvres**. Ex. [Item:Q282970](#) (*en cours d'intégration*)
- **~17 000 descripteurs**. Ex. [Item:Q294672](#) (*en cours d'intégration*)



# Cassiodore (0490?-0580?)

(Q2785)

Aucune description fournie

Cassiodore (0490?-0580?) | Cassiodore 0490?-0580? | Cassiodore | 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))
English	Cassiodorus (0490?-0580?)	Aucune description fournie	Cassiodorus, Senator, approximately 487–approximately 580 Cassiodorus Cassiodorus, Senator

Toutes les langues saisies

## Déclarations

nature de l'élément	être humain
▾ 0 référence	

date de naissance	490 Grégorien
	qualité de l'information      présumé
▾ 0 référence	

date de mort	580 Grégorien
	qualité de l'information      présumé

## Forme retenue du libellé :

- soit récupérée du fichier d'autorité BnF ou Library of Congress
- soit basés sur les choix faits par d'autres référentiels et dictionnaires

## Formes alternatives :

- graphies du nom figurant dans les bases et jeux de données sources



sex ou genre

masculin

▼ 0 référence

note Biblissima

⌚ 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 référence

⌚ A. Molinier, Les Sources de l'histoire de France - Des origines aux guerres d'Italie (1494): 224. Magnus Aurelius Cassiodorus Senator

▼ 0 référence

### Note bibliographique :

- ajoutées pour donner des indices sur l'identité de certaines personnes

Ressources Biblissima

identifiant Biblissima

⌚ pdatad89fe276e17712275adbc02eb985600cabfa7390

▼ 0 référence

Occurrences de l'entité dans

IIIF-Collections

⌚ Q2785

▼ 0 référence

### Lien vers les pages Biblissima :

- sur le Portail et IIIF-Collections

Identifiants

identifiant Bibliothèque  
nationale de France

⌚ 121582004

▼ 0 référence

## Identifiants



identifiant Biblissima	↳ d89fe276e17712275adbc02eb985600cabfa7390 0 référence	identifiant Bibliothèque royale des Pays-Bas	↳ 070138427 0 référence
identifiant Bibliothèque nationale de France	↳ 121582004 0 référence	BEACON ID	↳ 118519514 0 référence
identifiant Bibliothèque du Congrès	↳ names/n79082349 0 référence	identifiant Collections numériques de la Sorbonne	↳ 0000-0000_1901_num_1_1_855_t1_0087_0000_5 0 référence
identifiant Gemeinsame Normdatei	↳ 118519514 0 référence	Medieval Manuscripts in Oxford Libraries ID English	↳ person_95230707 0 référence
identifiant CERL	↳ cnp01302134 2 références	identifiant Dizionario Biografico degli Italiani	↳ cassiodoro 0 référence
identifiant Wikidata	↳ Q192365 0 référence	identifiant Pinakes (Textes et manuscrits grecs) (IRHT)	↳ auteur/624 0 référence
identifiant IdRef	↳ 030097703 0 référence	identifiant Bibliothèques françoises (CESR, Université de Tours)	↳ bvhbibfr_01404 0 référence
identifiant Bibliothèque	↳ XX878991		

### Section Identifiants :

- liens vers les bases sources (IRHT, CESR etc.)
- alignements externes (BnF, LoC, GND, CERL, BNE, IdRef, Wikidata...)

# Référentiel des lieux géographiques

---

## → Types de lieux concernés :

- Lieux des établissements de conservation
- Lieux des collectivités renseignées en tant qu'agent, en lien avec un document (ancien possesseur, producteur d'archives etc.)
- Lieux de production des manuscrits ou de publication des imprimés
- Lieux en tant que descripteur iconographique (Mandragore et Initiale)



# Saint-Benoît-sur-Loire (Loiret, France)

(Q31809)

Aucune description fournie

Saint-Benoît-sur-Loire | Saint-Benoît-sur-Loire, Centre, France | saint-benoît-sur-loire (france) | SAINT-BENOIT-SUR-LOIRE

▼ Plus de langues

Configurer

Langue	Libellé	Description	Également connu comme
français	Saint-Benoît-sur-Loire (Loiret, France)	Aucune description fournie	Saint-Benoît-sur-Loire Saint-Benoît-sur-Loire, Centre, France saint-benoît-sur-loire (france) SAINT-BENOIT-SUR-LOIRE

## Déclarations

nature de l'élément	<span>lieu géographique</span> <span>▼ 0 référence</span> <div style="border: 1px solid red; padding: 2px; margin-top: 10px;">lieu habité</div> <span>▼ 0 référence</span>
---------------------	---

partie de	<span>géographie: france et monaco</span> <span>▼ 0 référence</span>
-----------	---

pays	<span>France</span> <span>▼ 0 référence</span>
------	---

## Forme retenue du libellé :

- formes en français récupérées de la BnF
- formes en anglais : Geonames, Wikidata ou autres référentiels (ex. Pleiades)

## Formes alternatives :

- graphies du nom figurant dans les bases et jeux de données sources

## Type de lieu (Geonames) :

- code d'élément GeoNames (*feature code* ; ex. P.PPL pour “lieu habité”)

## Classement thématique repris de la base Mandragore

- 36 catégories inspirées de la classification Dewey



latitude

47.80561

► 2 références

longitude

2.31274

► 2 références

coordonnées géographiques

47°48'20.20"N, 2°18'45.86"E

▼ 0 référence

### Coordonnées géographiques :

- extraites de GeoNames ou de la BnF (Cartes et Plans)

### Ressources Biblissima

identifiant Biblissima

ldataf154bc4ac59d01d9d123a028ed677c51f873ba47

▼ 0 référence

### Lien vers la page du lieu sur le Portail

### Identifiants

identifiant Bibliothèque nationale de France

15260662n

▼ 0 référence

11962129k

répertoire d'autorité matière

<http://rameau.bnf.fr>

▼ 0 référence

### Identifiants externes :

- GeoNames et BnF (ID RAMEAU + ID Cartes et Plans)

identifiant GeoNames

2981393

[data.biblissima.fr/entity/O31809](http://data.biblissima.fr/entity/O31809)



# Points d'accès API de data.biblissima

---

- API Mediawiki / Wikibase :  
<https://data.biblissima.fr/w/api.php>
- Données RDF d'une entité (négociation de contenu) :  
<https://data.biblissima.fr/entity/Q{id}>
- Webservice de réconciliation :  
<https://data.biblissima.fr/reconcile/fr/api>
  - Manifest de l'instance Wikibase (pour OpenRefine) :  
<https://github.com/OpenRefine/wikibase-manifests/blob/master/biblissima-data-manifest.json>
- Interface SPARQL (en test)

# Étapes principales du workflow

---

1. Extraction des entités (agents, lieux, cotes etc.) à partir des données source (MySQL, XML, JSON...)
2. Traitement des entités :
  - dédoublonnage
  - réconciliation via le module dédié dans *OpenRefine* (webservice [openrefine-wikibase](#) d'A. Delpeuch ; seuil de matching auto. fixé à 80)
  - alignement externe vers d'autres référentiels (optionnel)
    - récupération d'informations tierces ( coordonnées géo, libellés alternatifs, dates, références biblio etc.)

# Étapes principales du workflow

---

## 3. Import des données dans Wikibase :

- Fichiers CSV préparés en amont dans *OpenRefine* (suivant une convention de nommage des colonnes)
- Scripts Pywikibot *ad hoc*, développés en interne
  - création ou mise à jour d'entités
- Mise à jour des données du triplestore (*Blazegraph*)

# SPARQL endpoint

---

→ Toujours en phase de test :

- problème de performance et ressources machine
- automatisation de la mise à jour Wikibase ⇒ triplestore (*Blazegraph*)
- fork de *wikidata-query-rdf* (back-end) et ajustements manuels de *wikidata-query-gui* (état de 2020), à cause de parties de code spécifiques à Wikidata

# SPARQL endpoint

---

→ À usage interne pour l'instant :

- pour récupérer les QID lors de l'import de données dans le Portail  
(requête SPARQL intégrée au module d'import de *Cubicweb*)
- pour extraire les données sur les items numérisés (URL et type) à partir du référentiel des cotes de manuscrits et imprimés anciens
- pour faire rapidement des extractions en masse, ou faire des requêtes plus poussées sur les données



## Cassiodore (0490?-0580?)



Autres graphies : Aurelius Cassiodorus (0490?-0580?)  
Flavio Magno Aurelio Cassiodoro (0490?-0580?)  
Flavius Magnus Aurelius Cassiodorus (0490?-0580?)  
Magnus Aurelius Cassiodorus (0490?-0580?)  
CASSIODORIUS

Naissance : 490  
Mort : 580

Notes : Préfet du prétoire et historiographe de Théodoric. - Fondateur du monastère de Vivarium (Source : *BnF*)  
Date : (Source : *Initiale*)

Référentiel d'autorité Biblissima : <https://data.biblissima.fr/entity/Q2785>

URI de l'entité personne sur data.biblissima

## Interventions

Auteur de l'oeuvre :

- > [Cassiodore \(0490?-0580?\) : De l'âme](#)
- > [Cassiodore \(0490?-0580?\) : Excerpta](#)
- > [Cassiodore \(0490?-0580?\) : Excerpta Quintiliani de arte rhetorica Cassiodori Institutionibus inserta](#)
- > [Cassiodore \(0490?-0580?\) : Expositio psalmorum](#)
- > [Cassiodore \(0490?-0580?\) : Historia ecclesiastica tripartita](#)
- > [Cassiodore \(0490?-0580?\) : Institutiones](#)

*Page de Cassiodore sur le Portail Biblissima*

Référentiel d'autorité Biblissima : <https://data.biblissima.fr/entity/Q120888>Ma sélection 

Nom d'usage : Evangiles de Saint-Riquier  
 Conservé à : Abbeville, Bibliothèque municipale  
 Date de fabrication : 750-799 (08e s. (seconde moitié)) (Source : Initiale)  
     fin du 8e siècle (Source : Europeana Regia)  
 Lieu de fabrication (Note) : France ? (Source : Initiale)  
 Support : parchemin  
 Nombre de feuillets : 198 ff.  
 Format : 350 × 250 mm.

URI de l'entité cote sur data.biblissima


<https://gallica.bnf.fr/ark:/12148/btv1b55005654j/manifest.json>

Numérisation (décor)

[https://bvmm.ihrt.cnrs.fr/mirador/index.php?  
manifest=https://bvmm.ihrt.cnrs.fr/iiif/15028/manifest](https://bvmm.ihrt.cnrs.fr/mirador/index.php?manifest=https://bvmm.ihrt.cnrs.fr/iiif/15028/manifest)
Numérisation : <https://gallica.bnf.fr/ark:/12148/btv1b55005654j>

Infos liées à la numérisation  
 (URL BVMM, URL Gallica,  
 URL du Manifest IIIF, type de  
 numérisation)

*Page du manuscrit Abbeville Ms. 4 sur le Portail Biblissima*

# Project:SPARQL query service/examples

Sommaire [masquer]

1 Exemples de requêtes SPARQL

- 1.1 Requêtes générales
  - 1.1.1 Tous les triplets d'une entité donnée
  - 1.1.2 Vue globale des types d'entités
  - 1.1.3 Lieux (liste complète)
  - 1.1.4 Lieux géolocalisés (carte)
  - 1.1.5 Personnes physiques (liste complète)
  - 1.1.6 Organisations (liste complète)
  - 1.1.7 Manuscrits (liste complète)
  - 1.1.8 Manuscrits (répartition par bibliothèque)
  - 1.1.9 Manuscrits numérisés (répartition par bibnum)
  - 1.1.10 Bibliothèques numériques IIIF
- 1.2 Autres requêtes
  - 1.2.1 Personnes avec 2 ID bbma
  - 1.2.2 Organisations avec RCR et disposant de moins de 20 cotes
  - 1.2.3 Manuscrits détruits
  - 1.2.4 Entités alignées vers le catalogue BnF
  - 1.2.5 Entités alignées avec la DNB
  - 1.2.6 Entités alignées avec Persée ou IdRef
  - 1.2.7 Entités personnes issues des bases de l'IRHT
  - 1.2.8 Données sur les items numérisés
  - 1.2.9 Entités fusionnées/redirigées

## Exemples de requêtes SPARQL

Biblissima Query Service

Exemples

Assistant de requête ⓘ

+ Filtrer

+ Afficher

Limiter à 100

Type to filter

Requêtes générales

- Tous les triplets d'une entité donnée ⓘ
- Vue globale des types d'entités ⓘ
- Lieux (liste complète) ⓘ
- Lieux géolocalisés (carte) ⓘ
- Personnes physiques (liste complète) ⓘ
- Organisations (liste complète) ⓘ
- Organisations (liste complète) ⓘ
- Manuscrits (liste complète) ⓘ
- Manuscrits (répartition par bibliothèque) ⓘ
- Manuscrits numérisés (répartition par bibnum) ⓘ
- Bibliothèques numériques IIIF ⓘ

Autres requêtes

20

Requêtes SPARQL d'exemple (page wiki ⇒ interface SPARQL)

```

1 #Données sur les items numérisés
2 SELECT ?entity ?idbbma
3 (GROUP_CONCAT(DISTINCT ?surrogates; SEPARATOR = "###") AS ?surrogate) (GROUP_CONCAT(DISTINCT ?manifests; SEPARATOR = "###") AS ?manifest) (GROUP_CONCAT(DISTINCT ?digitTypeLabel; SEPARATOR = "###") AS ?dsTypeLabel)
4 WHERE {
5   VALUES ?filter { wd:Q32810 wd:Q113619 }
6   ?entity wdt:P129 ?uncleaned_idbbma;
7   wdt:P2 ?filter; wdt:P197 ?surrogates;
8   OPTIONAL {
9     ?entity p:P197 ?ds_stmt.
10    ?ds_stmt pq:P279 ?dsType.
11    ?dsType rdfs:label ?_dsTypeLabel.
12    FILTER((LANG(?_dsTypeLabel)) = "fr")
13  }
14  OPTIONAL {
15    ?entity wdt:P196 ?manifests ;
16    OPTIONAL {
17      ?entity p:P196 ?manifest_stmt.
18      ?manifest_stmt pq:P279 ?digitType.
19      ?digitType rdfs:label ?_digitTypeLabel.
20      FILTER((LANG(?_digitTypeLabel)) = "fr")
21    }
22  }
23  BIND(REPLACE(?uncleaned_idbbma, '^{\imj}data', '') AS ?idbbma)
24  BIND(COALESCE(?_dsTypeLabel, "<no label>") AS ?dsTypeLabel)
25  BIND(COALESCE(?_digitTypeLabel, "<no label>") AS ?digitTypeLabel).
26 }
27 GROUP BY ?entity ?idbbma
28 ORDER BY ?entity
29 LIMIT 20000

```

   20000 results in 19824 ms   

entity	idbbma	surrogate	manifest	type_label	dsTypeLabel	digitTypeLabel	dsType
<a href="https://data.biblissima.fr/entity/Q100014">Q100014</a>	937b6b9095808c1e92d36bd745deb5682ce396a1	<a href="https://digi.vatlib.it/view/MSS_Vat.lat.5858">https://digi.vatlib.it/view/MSS_Vat.lat.5858</a>	<a href="https://digi.vatlib.it/iiif/MSS_Vat.lat.5858/manifest.json">https://digi.vatlib.it/iiif/MSS_Vat.lat.5858/manifest.json</a>	<no label>	<no label>		
<a href="https://data.biblissima.fr/entity/Q100014">Q100014</a>							

Exemple de requête SPARQL : données associés à un item numérisé (URL numérisation, URL Manifest IIIF, type de numérisation)



## Orthelius

Type : personne

Identifiant BVH : bvhbibfr\_03471

Lien BVH : [bibfr.bvh.univ-tours.fr/personne/bvhbibfr\\_03471](http://bibfr.bvh.univ-tours.fr/personne/bvhbibfr_03471)

Export XML-TEI : [bibfr.bvh.univ-tours.fr/personne/bvhbibfr\\_03471/xml](http://bibfr.bvh.univ-tours.fr/personne/bvhbibfr_03471/xml)

### Référentiels

<https://data.biblissima.fr/entity/Q262>

## Résultats

La Croix du Maine

Aucune notice principale (nom en vedette)

1 notice secondaire (nom cité) [ développer ]



Biblissima  
Patrimoine écrit du Moyen Âge  
et de la Renaissance



consortium  
cahier



Huma-Num



Laboratoire  
Arts, civilisation et histoire de l'Europe | ARCHE | UR 3400  
Université de Strasbourg

*Exemple de récupération et insertion des URI data.biblissima dans une base de données  
d'un partenaire de Biblissima (CESR)*

# Alignons-nous !

The image shows two screenshots of the OpenRefine interface. The left screenshot displays the 'Reconcilier la colonne "agents\_orig\_labels"' (Reconcile column "agents\_orig\_labels") dialog. It lists three entity types: organisation (Q167), être humain (Q168), and manuscrit (Q32810). Below this, there's a table for selecting columns to include and their properties. The right screenshot shows the main OpenRefine workspace with a table of 15 matching lines from the 'bremen\_books.csv' file. The table includes columns for label, canonical, agents\_pref\_label, agents\_alt\_labels, agents\_orig\_labels, and agents\_role. A sidebar on the right provides options for creating new elements and searching for correspondences.

Réconciliation de données dans OpenRefine grâce à l'API dédiée, utilisable par tous les projets souhaitant s'aligner avec les référentiels Biblissima (voir le "[Manifest Wikibase](#)" de l'instance)



# Obstacles techniques

---

- Installation initiale et configuration de Wikibase peu évidente
  - manque de documentation en 2019
  - gestion des extensions peu pratique
- Prise en main de la bibliothèque *Pywikibot*
- *Wikidata Query Service (WDQS)*
  - code spécifique à Wikidata (côté backend et frontend)
  - *Blazegraph* gourmand en ressources (Java)

# Limites de Wikibase

---

→ Moteur de recherche limité :

- problème de prise en compte des caractères spéciaux
- pas de support natif des troncatures
- absence de filtre par type d'entité

→ Prise en main difficile de la logique de saisie et du modèle de données :

- connaissance préalable du modèle de données
- manque de guidage lors de la création ou édition de page

# Suite du travail...

---

- Finaliser la publication des référentiels existants :
  - œuvres + descripteurs iconographiques
- Développer la synchronisation Wikibase ⇒ Portail (via SPARQL)
- Documenter le modèle de données
- Alimenter et étendre les référentiels avec les jeux de données produits dans le cadre de Biblissima+ (→ 2029)

# Merci de votre attention !

Équipe Biblissima+ :

*Kévin BOIS*

*Eduard FRUNZEANU*

*Régis ROBINEAU*

Biblissima 

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

[biblissima.fr](http://biblissima.fr)

 [@biblissima](https://twitter.com/biblissima)