



**HAL**  
open science

## Rédiger un PGD... et après ? Retour d'expérience sur le PGD de Biblissima+

Emmanuelle Morlock

### ► To cite this version:

Emmanuelle Morlock. Rédiger un PGD... et après ? Retour d'expérience sur le PGD de Biblissima+. Printemps de la Donnée 2024, INRAE; Université Haute-Alsace; Université de Strasbourg; INSA; PNDB; AgroParisTech; Université de Lille; Sorbonne Université; Data Terra, Jun 2024, Paris, France. 10.5281/zenodo.12165936 . hal-04683432

**HAL Id: hal-04683432**

**<https://hal.inrae.fr/hal-04683432v1>**

Submitted on 2 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Rédiger un PGD... et après ?

## Retour d'expérience sur le PGD de Biblissima+

**Emmanuelle Morlock**  
IR CNRS, Laboratoire HiSoMA (Lyon)



# Résumé

Biblissima+ est une infrastructure numérique multipolaire de recherche fondamentale et de service consacrée au patrimoine écrit ancien, de l'antiquité à la renaissance. Elle réalise la mise en interopérabilité de ressources numériques hétérogènes au sein d'un portail national et s'articule autour de 7 domaines d'expertise ou « clusters » organisés selon le cycle de vie des données. Son PGD, dont la première version a été rédigée en avril 2022, s'appuie sur un plan original combinant une double dimension de PGD d'entité et de projet. Un site web sur gitlab a également été mis en place pour faciliter son utilisation pratique. En effet, de nombreuses questions se posent encore concernant le suivi de son implémentation concrète et la remontée d'information à mettre en place pour sa mise à jour.

<https://gtsodataweb9.sciencesconf.org/>



Le contexte :

L'infrastructure Biblissima+

# Bibliissima+

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

Bibliissima+ 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. [Voir les actualités](#)

**Bibliissima** : un projet conçu il y a 13 ans, commencé il y a 11 ans

De Bibliissima à Bibliissima+

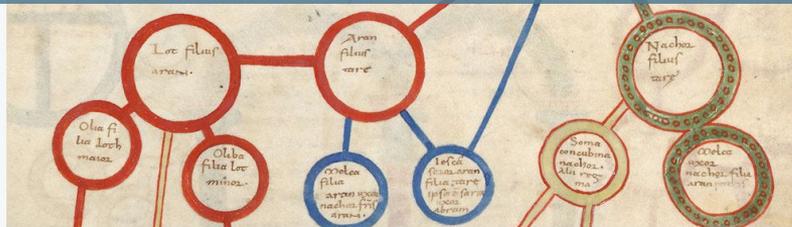
**Bibliissima+** : l'infrastructure s'agrandit et se développe de fin 2021 à 2029

L'infrastructure Bibliissima+ (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](#)

<https://bibliissima.fr/>





École nationale des chartes

Catalogues de vente de bibliothèques  
Manuscrits mis en vente

éditions en ligne de l'école des chartes

Miroir des classiques  
Frédéric Duval et Françoise Viell

Présentation | Description des notices | Rechercher Catalogue | Rechercher par mot-clé | Exemple

3029 résultats ont été trouvés pour votre recherche

Titre du manuscrit

Sélectionner

Auteur

Sélectionner

Traducteur

Nom du

La Vallière, Louis-César  
Baume-le-Blanc, duc de  
La Vallière, Louis-César de La  
Baume-le-Blanc, duc de  
Guyon, Jean-Baptiste-Denis

La forteresse de la foy  
Les livres ystorialuz d  
translates de latin en f  
escolastres



Mandragore, base des manuscrits enluminés de la BnF

# Biblissima : objectif initial

## Créer un accès unique à la documentation savante sur l'histoire de la transmission des textes (8e-18e siècles)



About the project

Partners

About the project

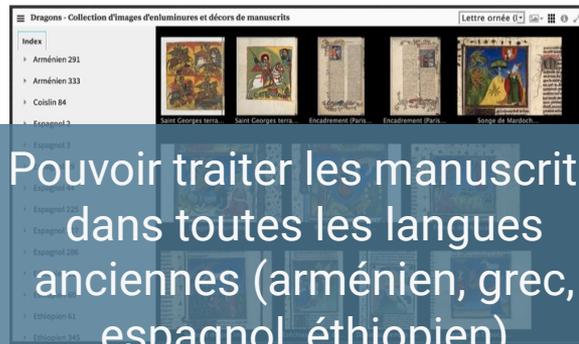
"Books within books: Hebrew Fragments in European Libraries" working on fragments of medieval Hebrew books and documents rec files in various libraries and archives in Europe, Israel and USA. Heb unique witnesses of Jewish presence and intellectual activities in me of the books and writings produced in the past have been preserved bindings has considerably enriched our knowledge of medieval Hebr fragments have been identified in various libraries and collections in Hungary, Italy, Poland, Slovakia, Spain, Switzerland and Czech Republic. The richness at

RELIURES | ATELIERS ET RELIEURS | POSSESSEURS





S'ouvrir à tous les supports porteurs d'écrits



Pouvoir traiter les manuscrits dans toutes les langues anciennes (arménien, grec, espagnol, éthiopien)

**Biblissima+ :**  
élargissement du périmètre avec de multiples défis scientifiques



Question des écritures sans polices unicode



Donner accès à la bibliographie secondaire en texte intégral : Persée, Hal, ISTEK...



# 2 périodes de financement

Direction scientifique et technique : **Anne-Marie Turcan Verkerk** (EPHE)

Etablissement porteur : **Campus Condorcet**

## Biblissima (2012-2021)

- ~7,1 M € (PIA 1)
- 9 partenaires
- Circulation des textes du **Moyen-Âge à la Renaissance**
  
- Mise en place de l'infrastructure technique : portail biblissima + hub de référentiels d'autorité data.biblissima.fr

## Biblissima+ (2021-2029)

- ~11,9 M € (PIA 3)
- 17 partenaires
- Elargissement à l'**antiquité / tous supports / toutes langues**
  
- Développement de l'infrastructure autour de 7 domaines d'innovation
  
- 25 contrats de travail en 2022, 42 en 2023 (hors prestations)

# 17 équipes de recherche



ARCHIVES  
NATIONALES



centre  
Michel  
de  
Bouard  
CRAHAM



crh

CAMPUS  
CONDORCET  
GRAND EQUIPEMENT  
DOCUMENTAIRE

HISOMα  
Histoire et sources des mondes antiques

IRHT  
Institut de recherche  
et d'histoire des textes

P D N  
M R S H

perseé



FranceArchives  
PORTAIL NATIONAL DES ARCHIVES

T E K L I A

# 16 Établissements et 1 entreprise



**Campus Condorcet**  
PARIS - AUBERVILLIERS

**Campus Condorcet (EPCC)**  
Coordinateur de Biblissima  
Établissement public (EP)  
Adresse : 8, cours des Humanités 93322 AUBERVILLIERS Cedex  
Président : Jean-François BALAUDE  
Responsable de Biblissima : Anne Marie TURCAN-VERKERK



**Centre national de la recherche scientifique (CNRS)**  
Établissement Public à Caractère Scientifique et Technologique (EPICST)  
Adresse du siège : 3 rue Michel Ange - 75794 PARIS Cedex 16  
Président : Antoine PETIT



**Université Paris Sciences & Lettres (PSL)**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 60 rue Mazarine 75006 PARIS  
Président : Alain FUCHS



**École des hautes études en sciences sociales (EHESS)**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 54 boulevard Raspail, 75006 PARIS  
Président : Christophe PROCHASSION



**Ministère de la Culture**  
Adresse : 181 rue Saint-Honoré 75001 PARIS  
Ministre : Roselyne BACHELOT-NARQUIN



**Muséum national d'Histoire naturelle (MNHN)**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 57, rue Cuvier, 75005 PARIS  
Président : Bruno DAVID



**Teklia**  
Société par actions simplifiée unipersonnelle (SASU)  
Adresse : 30 rue Raymond Losserand, 75014 PARIS  
Président : Christopher KERMORVANT



**Avignon Université**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 74 rue Louis Pasteur, 84029 AVIGNON Cedex 1  
Président : Philippe ELLERKAMP



**École nationale des chartes (ENC-PSL)**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 65, rue de Richelieu 75002 PARIS  
Directrice : Michelle BUBENICEK



**École normale supérieure de Lyon (ENS Lyon)**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 15, parvis René Descartes, BP 7000, 69342 LYON Cedex 7  
Président Directeur général : Jean-François PINTON



**École normale supérieure de Paris (ENS-PSL)**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 45 rue d'Ulm, F-75230 PARIS cedex 05  
Directeur : Marc MEZARD



**École pratique des hautes études (EPHE-PSL)**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : Les Patios Saint-Jacques, 4-14 rue Fernus, 75014 PARIS  
Président : Jean-Michel VERGIER



**Université de Caen Normandie**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : Esplanade de la Paix, 14032 CAEN Cedex  
Président : Lamri ADOUI



**Université Lumière Lyon 2**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 18 quai Claude Bernard 69365 LYON Cedex 07  
Président : Nathalie DOMPIER



**Université Jean Moulin Lyon 3**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 1C avenue Des Frères Lumière CS 78242 69372 LYON Cedex 08  
Président : Eric CARPANO



**Université de Poitiers**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
15 rue de l'Hôtel Dieu TSA 71117 86073 POITIERS Cedex 9  
Présidente : Virginie LAVAL



**Université de Tours**  
Établissement public à caractère scientifique, culturel et professionnel (EPSCP)  
Adresse : 60 rue du Parc d'Evain - BP 12930, 37020 TOURS Cedex 1  
Président : Arnaud GIACOMETTI

# Programme de travail de Biblissima+

## Volet A

### Nouvelles données

- A1 - Mécanismes et grandes campagnes de **collecte de données ouvertes**  
+ bibliographie secondaire / liens avec Persée, OpenEditions, HAL...
- A2 - **Appel à projets** annuel

## Volet B

### Ecosystème

- Consolider le portail Biblissima
- Créer les briques = référentiels / éditions / bases de données / outils / formations...
- Structurer les communautés en finançant des “clusters” thématiques

# Les 7 Clusters Biblissima+

**1**  
**Images**  
Numérisation et  
protocole IIF

**3**  
**IA**  
Reconnaissance  
de de formes et  
d'écritures

**4**  
**Traitements**  
Systèmes  
graphiques -  
analyse de  
documents

**5**  
**TEI**  
Edition de  
sources

**6**  
**Patrimoine  
musical**

**2**  
**Données  
d'analyse  
de matériaux**

Tout Biblissima

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



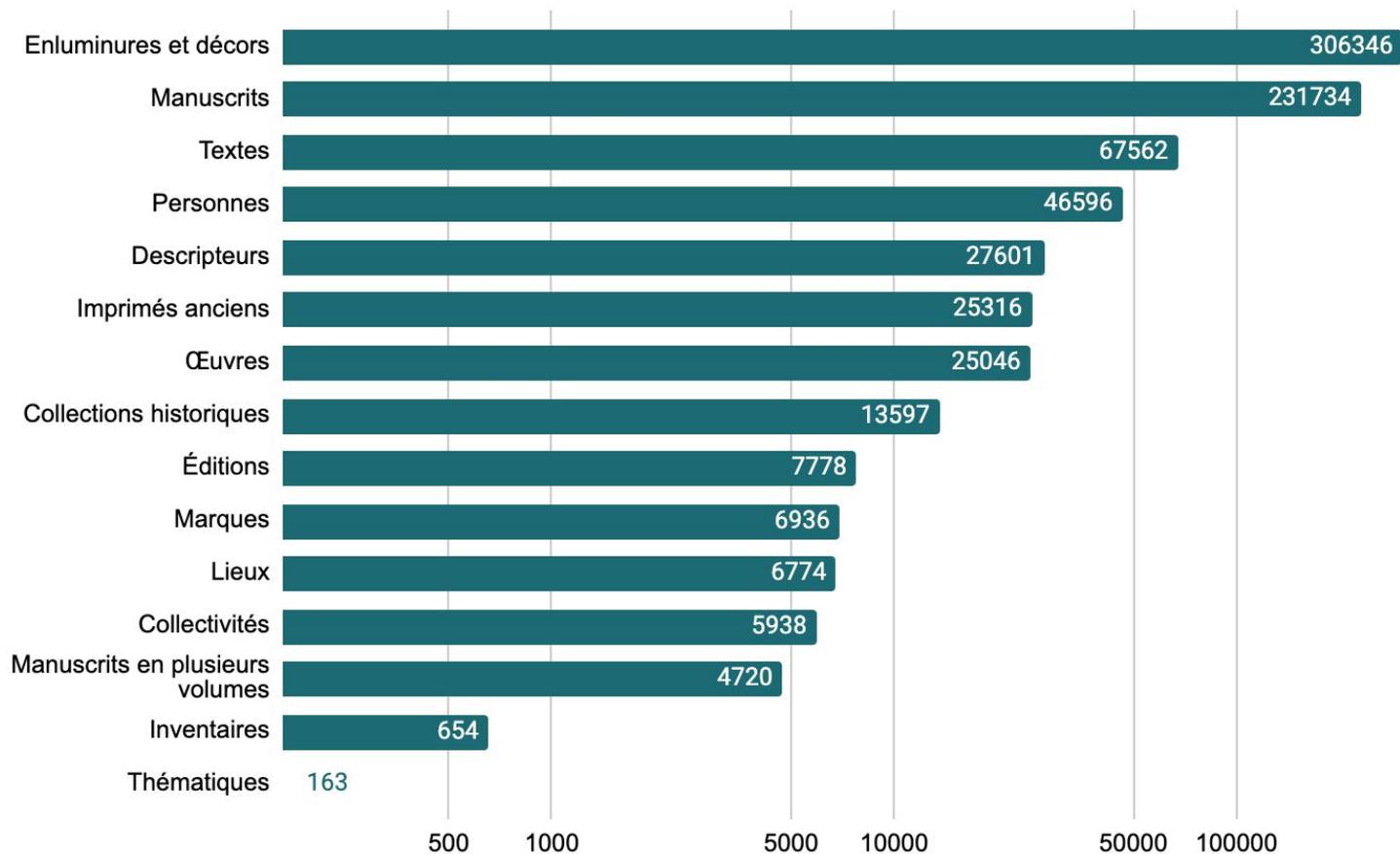
Bibliothèque virtuelle des bibliothèques, ce portail vous invite à découvrir l'histoire d'une partie des textes et livres qui ont été écrits, traduits, enluminés, collectionnés ou inventoriés depuis l'Antiquité jusqu'au XVIIIe siècle.

[A propos de ce portail](#)

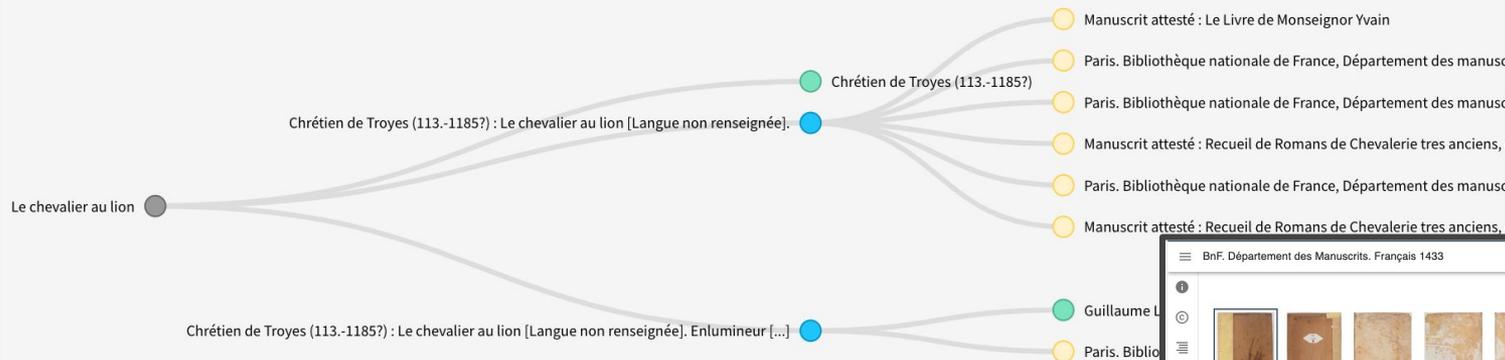
## Explorer



## 776 761 pages d'entités dans le portail **Biblissima** au 16/06/2024



# Vie de l'œuvre



International Image Interoperability Framework

BnF. Département des Manuscrits. Français 1433

IIIF manifest  
<https://gallica.bnf.fr/iiif/ark:/12148/btv1b105096493/manifest.json>

Digitisation:  
<https://gallica.bnf.fr/ark:/12148/btv1b105096493>

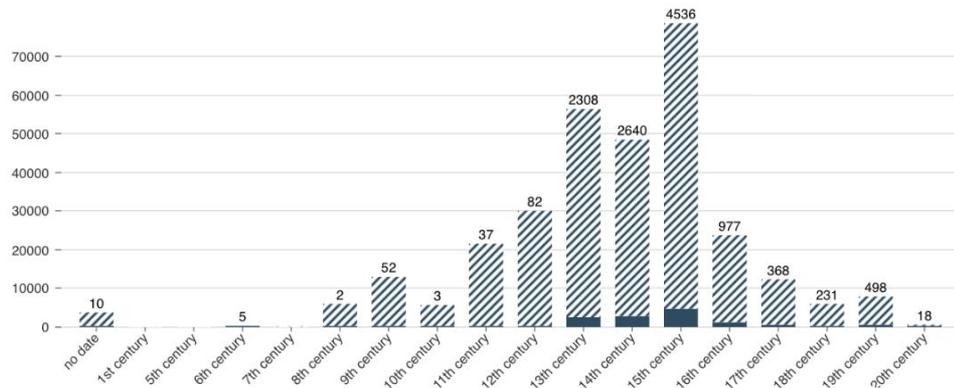
Remove filters

Thématique: Sport

Please select more filters

11767 illuminations and decoration  
(including 9831 with image)

### CENTURIES



### TOPICS

ARTS

Architecture  
Arts du spectacle  
Jeux et jouets  
**Sport**

### DESCRIPTORS

anneau (54)  
carquois (803)  
cavalier (10665)  
ceste (2)  
champ clos (59)  
juge diseur (36)  
lutte (148)  
manège (équitation) (4)  
marche (10)  
mât (12)

Search ...

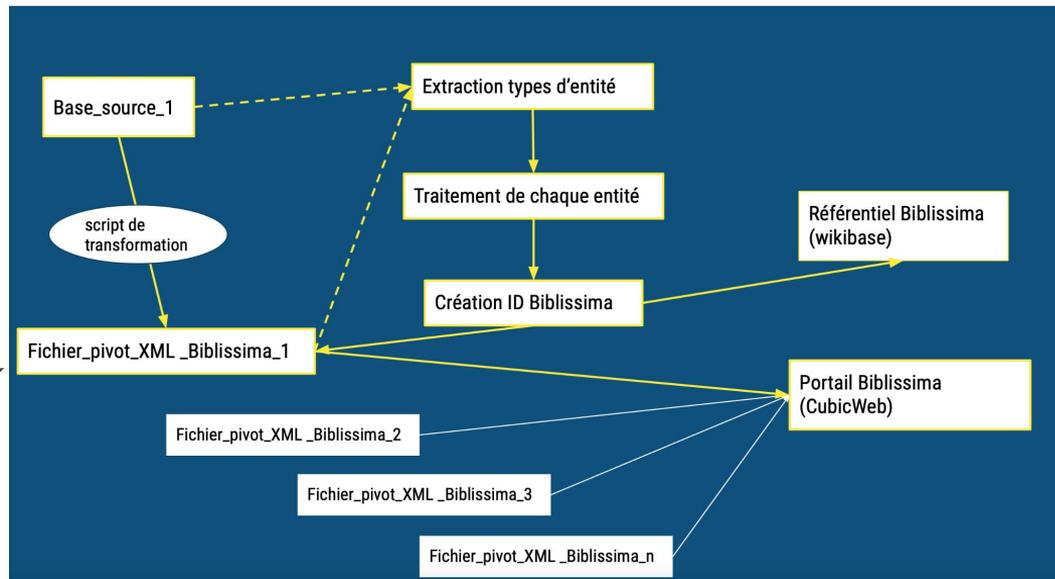
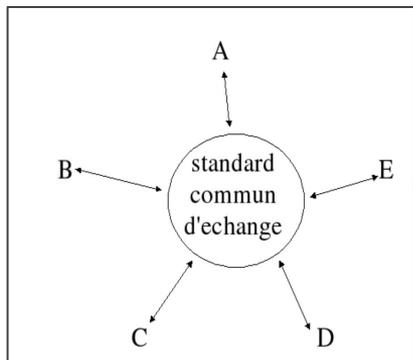
Lancelot-Graal (67)  
Les passages d'Outremer (66)  
Traité de la Toison d'or (65)  
**Chroniques d'Alexandre (57)**  
Caballero Cifar (55)  
Anciennes chroniques d'Angleterre (54)  
Voeux du paon (54)  
Fleur des histoires d'Orient (53)

FILTER

Descriptors  
Works  
Authors  
Illuminators  
Places of origin



# Mise en intéropérabilité



# Services, outils, documentation, données, codes



**Le Portail Biblissima**

Point d'entrée unique sur le patrimoine des cultures écrites anciennes de l'Antiquité à la Renaissance.



**IIF Collections**

Moteur de recherche de manuscrits et imprimés anciens numérisés et interopérables via les protocoles IIF.



**Référentiels Biblissima**

Plateforme collaborative de gestion et de publication des référentiels d'autorité Biblissima.



**Outils Biblissima**

Principales: Aide à la lecture et à l'apprentissage des langues anciennes, outils et environnements de travail en XML.

Archetype, Arkindex, Collatinus, e-scriptorium, Extractor, Eulexis, Filigranes pour tous, Kraken, Max, Multipal, Schema TEI, Thecae, Titulus, etc.

Accès par API et web services



## Ressources

Ressources Biblissima

Données et publications

Plan de gestion des données

Formations et événements

Documentation

IIF@Biblissima



### 3 ingénieurs en CDI depuis janvier 2018 :

- **Kévin Bois** (IE, développeur)
- **Eduard Frunzeanu** (IE, responsable des référentiels et du traitement des données)
- **Régis Robineau** (IE, coordinateur de l'équipe, référent IIIF)

### + 3 recrutements entre fin 2022 et avril 2024 :

- **Cécile Sajdak** (IE, CDD 4 ans) : développement de IIIF dans les archives de France
- **Laurie-Anne Tuairé** (IE, CDD) : gestionnaire d'opérations scientifiques
- **Florian Langelé** (IE, CDD) : traitement de données bibliographiques

# Contexte du point de vue de la gestion des données

## FORCES

Positionnement Science ouverte explicite

FAIR by design (web sémantique, IIF, TEI)

Même équipe technique (expertise et bonnes pratiques en place) - Serveurs : Campus Condorcet / Huma-Num

## OPPORTUNITES

Vrais enjeux de partage et réutilisation (HTR, référentiels)

Nombreux jeunes chercheurs et ingénieurs recrutés

Possibilité d'expérimenter, rôle potentiel des clusters

## FAIBLESSES

Pas de personnel dédié dans l'équipe tech.

Travail sur les briques : personnels en CDD

Socle technique en évolution

Hétérogénéité et volumes (124 livrables x n datasets)

## MENACES

Résistances ou non intérêt des acteurs

Particularités du FAIR en SHS

Multiplés tutelles et politiques de Science ouverte

Hétérogénéité des compétences numériques

## Les défis de la gestion et du partage des données en Lettres Arts et Sciences humaines

- **Hétérogénéité et variété** des données
- Données **difficilement décontextualisables** (Importance des institutions d'origine)
- **Complexités et incertitudes juridiques** impactant les pratiques de partage et la gestion des données
- Nécessité des métamoteurs et agrégateurs mais **risque de perte de la complexité des données** et biais dans la présentation (comment représenter les manques)
- Paradigme de la “data driven science” non mûr...
- **Délimitations difficiles à réaliser** entre matériaux de départ, analyses et interprétations, agencements...



The risk of losing thick description: Data management challenges Arts and Humanities face in the evolving FAIR data ecosystem

Erzsébet Tóth-Czifra

<https://shs.hal.science/halshs-02115505>

Quel PGD ?

Modèle, méthode, outils... ?

# Cadre contractuel ANR

## **Article 8 de la convention attributive d'aide**

- V1 à 6 mois (soit le 26/04/2022)
- 1 version tous les 2 ans
- 1 version finale

# 2 versions + 1 site en ligne — DOI : [zenodo.org/records/10131101](https://zenodo.org/records/10131101)

**Biblissima!**  
Observatoire des cultures écrites  
de l'argile à l'imprimé

---

## Plan de gestion des données

*Version Initiale (V1.07)*

---

Tableau de suivi du document	
Titre	Plan de gestion des données du projet Biblissima+ V1 (version initiale à 6 mois)
Auteurs	Emmanuel Mörtek, Régis Robineau, Edouard Frouzeau, Kevin Bois
Contributeurs	Anne-Marie Turcan-Verkerk, Marie-Agnès Avenel, François Bougard
Rélecteurs	Responsables scientifiques et techniques de livrables
Validé par	Anne-Marie Turcan-Verkerk (responsable scientifique et technique de Biblissima+)
Date de création	24 mars 2022
Type	Texte
Langage	fr-FR
Confidentialité	Public
Statut	<input checked="" type="checkbox"/> En cours de rédaction <input checked="" type="checkbox"/> Pour réécriture <input checked="" type="checkbox"/> Pour validation <input checked="" type="checkbox"/> Validé

Biblissima+ Plan de Gestion des données - version initiale - avril 2022 1

**Biblissima!**  
Observatoire des cultures écrites  
de l'argile à l'imprimé

---

Plan de gestion de données de l'infrastructure  
Biblissima+, observatoire des cultures écrites  
anciennes de l'argile à l'imprimé (ANR-21-  
ESRE-0005)

---

**Version 2.0**  
Octobre 2023

---

**PGD de Biblissima+**

Gitlab

- Introduction** >
- Le PGD de Biblissima+ (cadre de référence) >
- Vue d'ensemble des jeux de données >
- PGD détaillé (Périmètre 1) >
- Annexes >

estion des données (PGD) de l'observatoire des cultures  
édérant 16 établissements et une entreprise privée, réunis  
rée une infrastructure numérique multipolaire de  
service consacrée à l'histoire de la transmission des textes  
d'argile mésopotamiennes aux premiers livres imprimés,  
tes les langues. Biblissima+ concerne donc l'ensemble des  
nettant des textes anciens, y compris les sources  
onnaies, mais aussi les archives d'érudits modernes et de  
d elles apportent des informations originales sur les textes

[dmp.biblissima.fr](https://dmp.biblissima.fr)



# Calendrier de la V1

## Enquête (février-mars 2022)

- 125 jeux de données identifiés
- 43 réponses

## Choix clés

- Limiter le “PGD projet” au périmètre géré par l’équipe technique
- Mais inventorier toutes les contributions
- Responsabiliser les acteurs
- Faciliter les dépôts / sensibiliser

## Rédaction avec l’équipe technique (mars-avril)

- Lignes directrices
- PGD détaillé de P1
- Tableaux synthétiques (P2-P3)
- Annexe : synthèse de l’enquête

## Relecture (15 avril)

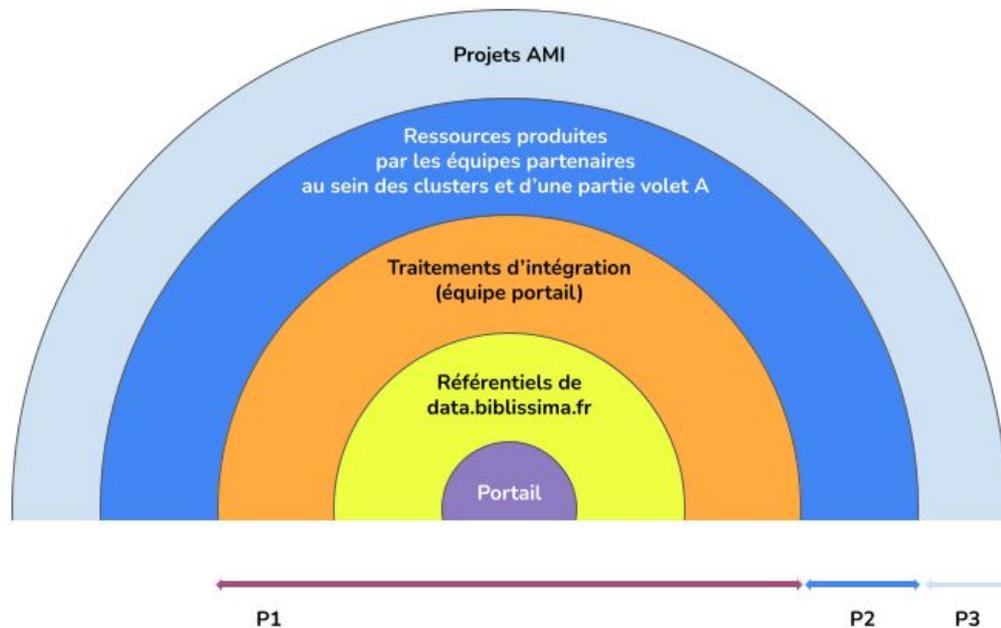
## Envoi à l’ANR (26 avril)

# Le “PGD simplifié” utilisé pour l'enquête

Formulaire renvoyer par mail à emmanuelle.morlock@cnsr.fr le 01/03/2022 au plus tard	
	Date :
	Nom du répondant :
<b>Identification du jeu de données</b>	
Si possible, référence du livrable auquel le jeu de données est rattaché (cf. feuille de référence jointe). À défaut, équipe et cluster concerné.	
<b>Titre du livrable</b>	Intégration de plusieurs volumes de la série « Documents, études et répertoires de l'Institut de Recherche et d'Histoire des Textes » (DER) disponibles sur Persée
<b>Description du jeu de données</b>	Catalogues de manuscrits de plusieurs bibliothèques édités par l'IRHT dans la série DER. Ces volumes imprimés ont été ocrés et mis en ligne par Persée. La constitution du jeu de données implique une extraction des notices descriptives selon une structuration compatible avec le modèle du portail Biblissima, en fonction de la nature des catalogues. Le processus implique aussi une relecture et correction de FOCR ainsi qu'une restructuration du format TEI source.
<b>Période de début des travaux prévue</b> (en indiquant une tranche annuelle ou semestrielle par rapport à la durée totale du financement ANR, par ex. T18, T60...)	
<b>Période de livraison prévue</b> (idem)	01/04/2022 (T2)
<b>Modalités d'utilisation des référentiels Biblissima</b>	31/03/2023 (T6)
<b>Caractérisation des données collectées (produites ou réutilisées)</b>	
<b>Type de provenance</b> (création de nouvelles données / réutilisation et transformation de données existantes).	réutilisation et transformation de données existantes
<b>Type de données</b> (textuelles / numériques / images / vidéo / médias divers / de simulation / code informatiques).	textuelles
<b>Format(s) des données et Standard(s) utilisés</b> (exprimés à l'aide de l'extension du nom de fichier .txt, .pdf, .csv) - préciser s'il y a un encodage standard (XML/TEI, EAD)...	.xml (XML « maison » et XML-TEI), .csv
<b>Informations sur la volumétrie</b> (exprimés en espace de stockage requis (octets), et/ou en quantités d'objets, de fichiers, de lignes, et colonnes).	
<b>Métadonnées et documentation</b>	

Merci d'inscrire vos réponses dans cette colonne dans la mesure du possible. Indiquez "Sans objet" quand la question n'est pas pertinente pour le jeu de données traité.

# Périmètres de responsabilités



# Un résultat “hybride”

- Une introduction “pédagogique” sur le rôle intégrateur d’un PGD au sein d’une équipe projet
- Une partie “cadre de référence” s’apparentant à un PGD “entité” ou “structure”
- Des inventaires de jeux de données centrés sur la stratégie de partage / conservation
- 1 PGD de projet traditionnel (pour P1) suivant un modèle ANR
- Des fiches pratiques pour faciliter les dépôts (en annexes)

# Plan

<b>Introductions</b>	<ul style="list-style-type: none"><li>• Définition et objectifs d'un plan de gestion des données</li><li>• Champ d'application et politiques d'établissements applicables</li><li>• Spécificités du projet Biblissima+</li></ul>
<b>Cadre de référence</b>	<ul style="list-style-type: none"><li>• Lignes directrices (12)</li><li>• Exigences minimales (7)</li><li>• Pratiques souhaitées (8)</li><li>• Enjeux des dépôts et identifiants pérennes pour la citation</li><li>• Responsabilités et ressources</li></ul> <p><b>PGD structure</b></p>
<b>PGD périmètre 1</b>	Selon le plan du modèle ANR <b>PGD projet</b>
<b>Annexes</b>	Listes des livrables, méthode, fiche pratique Zenodo...



*Un PGD est enfin le fruit d'une coopération inter-métiers entre scientifiques, informaticiens, documentalistes, archivistes et éditeurs qui peuvent participer à des degrés divers à son élaboration et à sa mise à jour. Quand elle est correctement mise en œuvre, cette dimension collaborative du PGD peut jouer un rôle particulièrement utile pour contribuer à forger une vision commune au sein d'une équipe. Le PGD offre en effet, sous l'angle spécifique des données et des procédures numériques, une vision d'ensemble du projet à la fois synthétique et concrète. Il peut ainsi faciliter grandement l'intégration des nouveaux collaborateurs ou collaboratrices. Il peut aussi servir de document de référence pour transmettre de manière efficace des informations clé pour comprendre le socle technique du projet.*

# Lignes directrices - exemples

- **Recommandations générales**

- Principes FAIR
- Principe “aussi ouvert que possible, pas plus fermé que nécessaire”
- Un PGD “particulier” (autonome) pour chaque livrable devant tenir compte des politiques de l’établissement d’appartenance des porteurs
- ...

- **Exigences minimales**

- Convention de nommage des fichiers
- Publication des PGDs dans les communautés Zenodo des Clusters
- Préparation des dépôts dans des clouds Sharedocs mis à disposition par Huma-Num
- Justifier les restrictions à l’ouverture et au partage
- ...

- **Pratiques individuelles souhaitées**

- Rappel des exigences de dépôt dans HAL pour les publications (articles, communications)
- Porter le PGD à la connaissance de tout nouvel arrivant dans le projet
- Bien anticiper le recours à l’équipe portail
- ...

# Les inventaires (tableaux)

- **Périmètre**

- Biblissima et Biblissima+

- **Classement**

- Infrastructure (P1)
- Outils de traitement pour l'intégration des données (P1)
- Ressources intégrées (P2-P3)
- Chaînes d'outils (P2)
- Autres ressources (P2-P3)

- **Focus**

- Partage (au sens de "dépôt dans un entrepôt de confiance")
- Archivage
- Actions à mener



## Lignes directrices 2

- **Dépôts de jeu de données**

- DOI
- Métadonnées du dépôt
- À **chaque version majeure** de la ressource
- sur Zenodo ou Nakala

- **3 fichiers demandés**

- README
- LICENSES
- AUTHORS

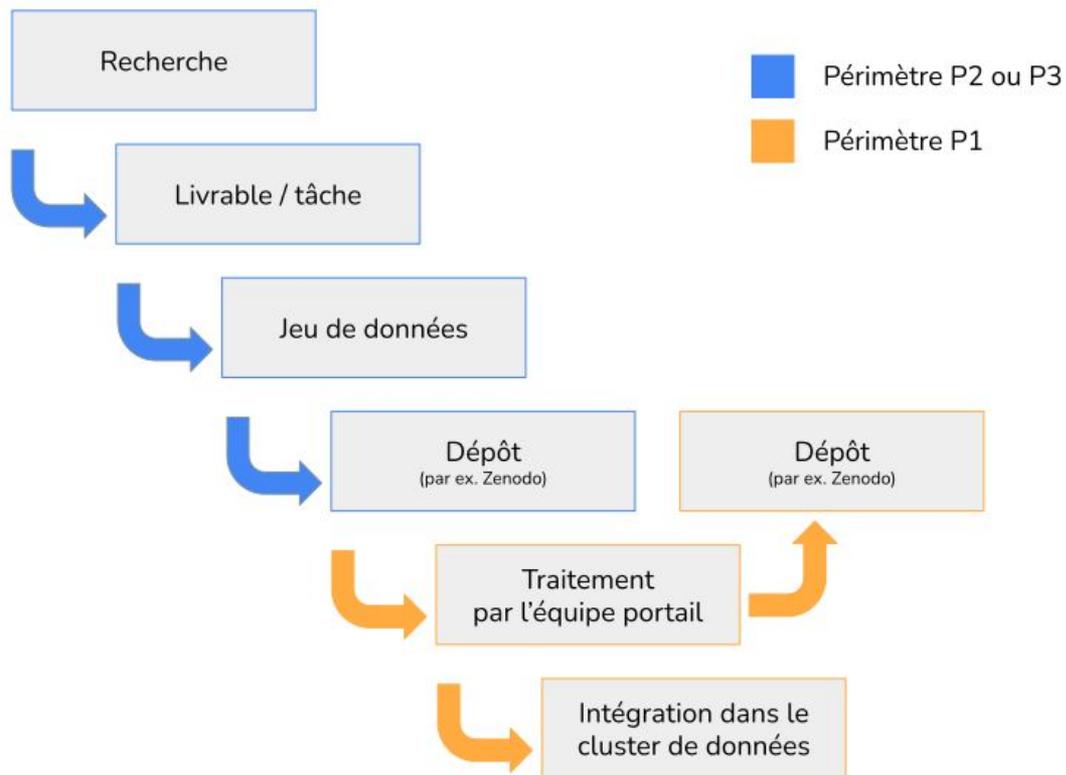


Métadonnées : Datacite



Métadonnées : Dublin Core

# Workflow des dépôts



## PGD\_V1 - Tableau 1 - Infrastructure numérique (P1)

Produit de recherche	Description	Nature des données	Formats / standards	Volumétrie	Politique de partage	Politique de conservation à long terme	Actions de Fairisation à mener
Cluster de données (Portail)	Données importées dans le Portail (base de données PostgreSQL), publiées via une application web (Portail) / Biblissima	Données textuelles formalisées	MARC, XML, JSON, RDF	645 839 entités (29/03/2022), pour 1,1 Go de données XML	Publication des données via le Portail, exposition de données dans le Web sémantique Partage via les dépôts des jeux de données au format pivot et enrichis.	Archivage en fin de projet d'un export des données au format RDF sur Zenodo. Les ressources issues des partenaires sont déposées de manière autonome par leurs producteurs à chaque version majeure (cf. processus d'intégration de sources de données dans le portail).	Préparer l'archivage final à la fin du projet en même temps que les spécifications des développements technologiques de l'infrastructure héritée de Biblissima 1. Sur les notices de ressources afficher les DOI des dépôts des jeux de données.
Cluster de données (IIIF-Collections)	Données importées dans IIIF-Collections (ElasticSearch), publiées via une application PHP	Données textuelles formalisées	CSV, JSON	82 065 items (29/03/2022)	Publication des données via le site IIIF-Collections	Les données vont être transformées en XML et versées dans le Portail (cf. ligne ci-dessus)	N/A
Interfaces et applications web	Moteurs d'indexation et interfaces de recherche et de visualisation des données (Portail et IIIF-Collections), développées en interne et en lien avec des prestataires Intégration de la recherche sur les matériaux dans l'interface du portail (avec CRC)	Codes informatiques	PHP, Python, JSON, Javascript	Portail : ~500Mo de code + ~5.5Go de caches et tests d'import IIIF Collections (app web) : ~1Go	Via Github / Gitlab et Zenodo pour les versions majeures	Moissonnage par l'archive pérenne de logiciels Software Heritage. Les logiciels ou modules dotés d'un potentiel de réutilisation dans la communauté feront l'objet d'un dépôt avec métadonnées modéré via la voie couplée HAL + Software Heritage	N/A
Visualiseur d'images Mirador	Version packagée du visualiseur (avec des plugins) pour le Portail Biblissima	Codes informatiques	Javascript, IIIF	~13 Mo	Github / Gitlab	N/A	N/A

Vue synthétique des 2 objectifs clés : Partage / Conservation

## 6 – Ressources du périmètre P3

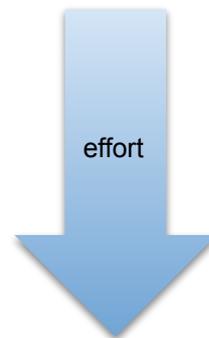


Ce tableau recense les jeux de données produits par les projets lauréats de l'appel à manifestation d'intérêt annuel (2022-2027).

Type de données brutes fournies	Nom	Équipe fondatrice associée	Catégorie de soutien	Type de ressource du projet global selon catégories du Portail	Utilisation native des référentiels Biblissima	Suivi intégration	Licence de diffusion du partenaire pour la base source	Politique de conservation à long terme du partenaire pour la base source	Lien
XML-ALTO (données d'entraînement)	Fabliaux	CIHAM	Projet partenarial	Corpus ou édition de source	oui	N / A	CC-BY 4.0	non	<a href="https://">https://</a>
images .jpg	Fabliaux	CIHAM	Projet partenarial	Corpus ou édition de source	oui	N / A	CC-BY 4.0	non	<a href="https://">https://</a>
XML-TEI	Fabliaux	CIHAM	Projet partenarial	Corpus ou édition de source	oui, pour les métadonnées des manuscrits	B+, non intégré	Etalab + CC-BY 4.0	Dépôt sur Nakala prévu fin 2023	<a href="https://num.fdiffus">https://num.fdiffus</a>
TXT	Fabliaux	CIHAM	Projet partenarial	Corpus ou édition de source	non	N / A	Etalab	non	<a href="https://num.fdiffus">https://num.fdiffus</a>
MLmodel (model HTR Kraken)	Fabliaux	CIHAM	Projet partenarial	Corpus ou édition de source	non	N / A	CC-BY 4.0	non	<a href="https://num.f">https://num.f</a>
XML-Alto (prédiction HTR)	Fabliaux	CIHAM	Projet partenarial	Corpus ou édition de source	non	N / A	CC-BY 4.0	non	<a href="https://num.f">https://num.f</a>
non connu	Bibliotheca Carnotensis Nova	IRHT	Projet partenarial	Catalogue ou répertoire	non connu	non connu	non connu	non connu	-

# Cas particulier des codes sources

- Principes à mettre en place pour l'infra numérique (P1) mais utilisables pour les autres périmètres si le code est géré dans Gitlab / Github
- Réponse graduée :
  - **Archivage « instantané »**
    - sans ajout de métadonnées, identifiant citable Software Heritage
  - **Dépôt via les tags de version de Github**
    - métadonnées + DOI Zenodo
  - **Dépôt vers Software Heritage via HAL**
    - métadonnées + identifiant citable Software Heritage, dépôt modéré



## V2 - octobre 2023

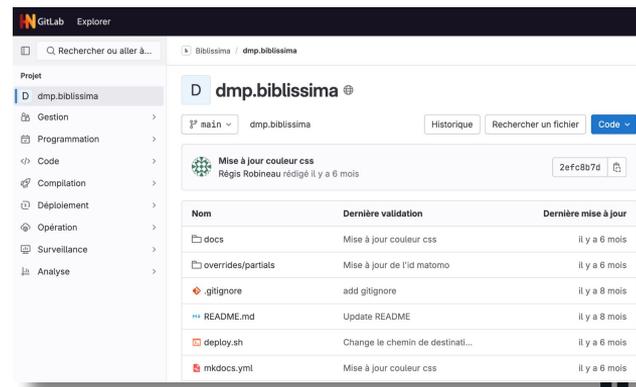
- Questionnaire envoyé le 3/10/2024 aux responsables de livrables
- 12 réponses seulement

question	oui	non
Souhaitez-vous demander des corrections ou ajouts ?	5	6
Si vous avez répondu "Oui" à la question précédente, souhaitez-vous le porter ce document à la connaissance du bureau exécutif et de l'équipe technique de Biblissima+ ?	1	3
Si vous avez répondu "Oui" à la question précédente, le lien ou le fichier peuvent-ils être diffusés ?	0	1
Avez-vous rédigé un PGD pour votre projet financé ou co-financé par Biblissima+	4	7



# Outils de rédaction et de publication

- V1 : Choix du document bureautique partagé (Googledoc...) étant donné le calendrier et génération d'un fichier pdf pour l'ANR
  - Mais recommandation d'utiliser le modèle structuré de DMP Opidor ou le modèle Argos d'OpenAire pour les "PGD particuliers"
- V2 : création d'un site en ligne le générateur de sites statiques de documentation MKdocs
  - Pour permettre la citation via des liens et le traçage des modifications sur un gestionnaire de code gitlab
  - La version pdf pour l'ANR a été générée à partir de cette version numérique.



Bilan

# Bilan

- **Réception**

- Zenodo : 1000 vues / 660 téléchargements
- Document apprécié (pour ceux qui l'ont lu !)
- Peu lu au sein du projet (P2-P3) ?
- Principes non suivis (PGDs particuliers pour P2-P3)

- **Dépôts**

- Oui quand il y a un enjeu perçu de réutilisation (thésaurus icono, HTR)
- Principe du dépôt avant traitement par l'équipe technique non pertinent
- Automatisation Github-Zenodo non connu

- **Accompagnement**

- A bien fonctionné pour l'appel à projet
- Clairement insuffisant (règles, recommandations et aides ne sont pas connus ou compris)

Published May 2, 2024 | Version v2

Dataset

🔓 Open



Edit



New version



Share

# Thesaurus iconographique Biblissima

Frunzeanu, Eduard<sup>1</sup>; Bois, Kévin<sup>1</sup>; Robineau, Régis<sup>1</sup>

Show affiliations

Ce dépôt met à disposition le thésaurus iconographique du [portail Biblissima](#) conformément au modèle [SKOS](#) (formats RDF/XML et Turtle). Ce thésaurus est constitué à partir des données des bases partenaires du projet [Biblissima](#).

Il est actuellement utilisé dans le portail Biblissima pour permettre la recherche et la navigation dans un corpus de plus de 306 000 notices d'enluminures et décors de manuscrits et imprimés anciens issues de deux bases iconographiques : Mandragore (BnF) et Initiale (IRHT-CNRS).

Voir sur le portail Biblissima :

- la page [Thésaurus iconographique](#)
- l'[interface d'exploration et de visualisation de l'iconographie](#)

Les concepts de ce thésaurus sont stockés et gérés dans la plateforme des [référentiels d'autorité de Biblissima](#) s'appuyant sur le logiciel Wikibase.

Pour plus d'informations sur la constitution du référentiel des descripteurs iconographiques, voir [cette page de présentation](#).

## Notes

Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme d'Investissements d'avenir intégré à France 2030, portant la référence ANR-21- ESRE-0005 (Biblissima+).

183

VIEWS

55

DOWNLOADS

▶ Show more details

## Versions

Version v2

May 2, 2024

10.5281/zenodo.11102986

Version v1

May 30, 2023

10.5281/zenodo.7988061

[View all 2 versions](#)

**Cite all versions?** You can cite all versions by using the DOI [10.5281/zenodo.7988060](#). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)



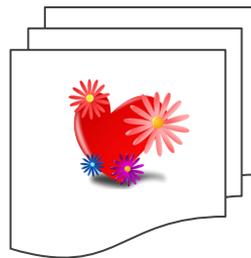
Perspectives

Un PGD plus...

*clair, simple, FAIR et sobre ?*

# Accompagner

- Mieux communiquer (fiches pratiques plus visuelles)
- Solliciter et mettre la main à la pâte pour le dépôt des codes sources
- Organiser des webinaires sur le PGD en impliquant l'équipe technique

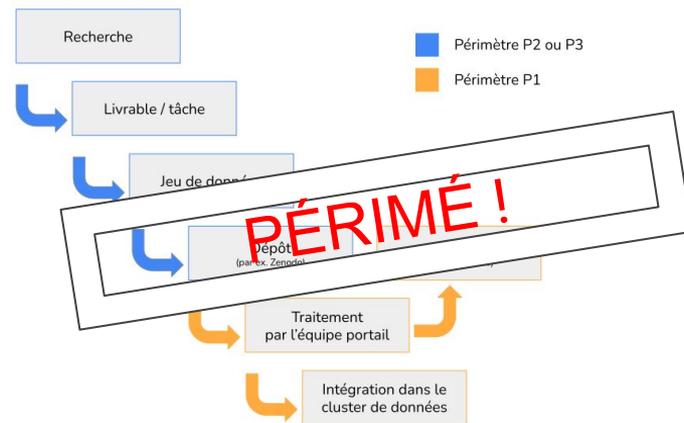


Software Heritage



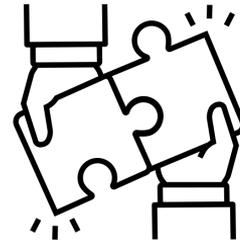
# Simplifier

- Supprimer le lien entre dépôt et traitement par l'équipe technique
- PGDs minimalistes à imaginer
  - via le formulaire de reporting annuel pour l'ANR ?
  - via DMP Opidor
  - via Argos ?



# Recentrer sur les nouveaux enjeux

- Organiser un atelier FIP (Fair Implementation Profile) avec le CS
- Préparer la candidature à la feuille de route des infrastructures nationales
- Le PGD, un bon outil pour consigner des objectifs de sobriété numérique ?



# Conclusion

# Conclusion

- Un document hybride Structure / Projet
  - Structure : plus orienté “incitations” que “règles”
  - Projet : vu comme une forme de documentation utile
- Un document évolutif : contenu / forme / objectifs
- Finalement, le PGD reste-t-il un “document administratif”... ou peut-il jouer un rôle plus au coeur de la production des données et de la gouvernance du projet ?

Questions ?

