

---

# DMP du projet "Hoppe Droit - Hypothèses d'Observation des Productions Pédagogiques Editées en Droit"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - Modèle de PGD (français)" fourni par Agence nationale de la recherche (ANR).

## Renseignements sur le plan

<b>Titre du plan</b>	DMP du projet "Hoppe Droit - Hypothèses d'Observation des Productions Pédagogiques Editées en Droit"
<b>Version</b>	Version finale
<b>Domaines de recherche (selon classification de l'OCDE)</b>	
<b>Langue</b>	fra
<b>Date de création</b>	2019-10-14
<b>Date de dernière modification</b>	2021-09-27
<b>Identifiant</b>	3827
<b>Type d'identifiant</b>	identifiant local
<b>Licence</b>	Creative Commons Attribution 4.0 International
<b>Documents (publications, rapports, brevets, plan expérimental....), sites web associés</b>	<ul style="list-style-type: none"><li>• Site web : <a href="https://bcujas-hoppe.univ-paris1.fr/">https://bcujas-hoppe.univ-paris1.fr/</a></li><li>• Rapport scientifique : <a href="https://www.collexpersee.eu/wp-content/uploads/2021/07/Rapport_scientifique-HOPPE.pdf">https://www.collexpersee.eu/wp-content/uploads/2021/07/Rapport_scientifique-HOPPE.pdf</a></li></ul>

## Renseignements sur le projet

**Titre du projet** Hoppe Droit - Hypothèses d'Observation des Productions Pédagogiques Editées en Droit

**Résumé** L'objectif du projet Hoppe Droit est l'élaboration d'un outil d'observation des productions pédagogiques en droit permettant d'étudier les évolutions du droit français et de son enseignement entre 1804 et 1954. Le travail repose sur un partenariat fort entre chercheurs et bibliothécaires à chaque étape : définition des objets, modélisation des données, délimitation de la période étudiée, création et enrichissement du corpus, élaboration d'un outil de visualisation. Il s'appuie sur les collections de la BIU Cujas.  
Ce partenariat a permis d'élaborer une méthodologie documentée, et reproductible pour d'autres disciplines

**Sources de financement**

- CollEx :

**Date de début** 2019-04-03

**Date de fin** 2021-04-03

**Produits de recherche :**

1. Corpus extrait du catalogue Cujas dans le cadre du projet Collex Hoppe Droit (Jeu de données)
2. Base de données Hoppe Droit (Heurist-HumaNum) (Jeu de données)
3. Interface de visualisation de Hoppe Droit (Resource interactive)

**Contributeurs**

Nom	Affiliation	Rôles
Catherine Renard	Université Paris 1 - BIU Cujas	<ul style="list-style-type: none"><li>• Responsable du plan de gestion de données</li></ul>
Gottely Alexandra	Université Paris 1 - BIU Cujas	<ul style="list-style-type: none"><li>• Coordinateur du projet</li><li>• Personne contact pour les données (Algo_Hoppe, Interface Hoppe, CorpusHoppeDroit)</li></ul>
Chambost Anne-Sophie	Sciences Po Lyon	

**Droits d'auteur :**

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

# DMP du projet "Hoppe Droit - Hypothèses d'Observation des Productions Pédagogiques Editées en Droit"

## 1. Description des données et collecte ou réutilisation de données existantes

### Corpus extrait du catalogue Cujas dans le cadre du projet Collex Hoppe Droit

#### *Données collectées*

Postulat de départ :

Le corpus est issu des collections de la bibliothèque Cujas, héritière de la bibliothèque de la faculté de droit de Paris. Bien que non exhaustives, les chercheurs associés au projet ont estimé qu'elles offraient une représentation significative d'une production destinée à étudier et enseigner le droit. Ainsi, l'étude porte « sur les productions pédagogiques éditées en droit francophone, sur une période courant de la réouverture des facultés de droit en 1804 jusqu'à la réforme de la licence en 1954 ».

#### *Origine des données collectées*

- **Extraction initiale à partir du catalogue de la bibliothèque Cujas géré par le logiciel Aleph (Ex-Libris).**

13376 notices bibliographiques au format Unimarc ont été extraites du catalogue Cujas au format xlsx à partir des termes des requêtes choisis par les chercheurs associés au projet et présentés dans le tableau 1.

Pour cette première extraction, toutes les zones Unimarc n'ont pas été exportées ; Après validation du corpus, de nouvelles extractions ont été effectuées avec l'ensemble des zones Unimarc utiles pour l'importation vers la base Heurist (voir tableau 2).

**Tableau 1 - Termes de requête**

<b>TYPE</b> (Verbe, NOM, NOM propre, ADVERBE, ADJECTIF)	<b>MOT</b>	<b>FORME DU PLURIEL</b>
NOM	guide	
NOM	introduction	
NOM	élément	
NOM	leçon	
NOM	institute	
NOM	répétition	
NOM	principe	
NOM	mémento	
NOM	précis	
NOM	traité	traités
NOM	manuel	
NOM	question	
NOM	cours	
NOM	méthode	
EXP	tableau synoptique	tableaux synoptiques

**Tableau 2 - Métadonnées extraites des notices Unimarc**

## Données descriptives de l'objet production pédagogique en droit

Données retenues comme pertinente	Correspondance(s) UNIMARC (zones et champs)
Titre	200 \$a
Complément de titre	200 \$e
Date d'édition	<a href="#">100 \$a / 219 \$d ou \$i (\$r avant 1830) (210 pour les périods)</a>
Mention d'édition	205 \$a
Collection	225 \$a / 410 \$0
Titre d'ensemble	200 \$a
Titre de partie	200 \$h numéro de partie \$i titre de partie
Auteur(s) principal(aux)	200 \$f / 700 ou 701 \$3lien vers la notice d'autorité
Auteur(s) secondaire(s)	200 \$g / 702 \$3lien vers la notice d'autorité
Fonction de l'auteur	<a href="#">700 ou 701 ou 702\$4code de fonction</a>
Sujet(s)	<a href="#">600. 601. 606. 607...</a>
Format	215 \$d
Nombre de pages	215 \$a
Illustration	215 \$c
Editeur(s)	219 \$c
ISBN	010 \$AISBN13 ou \$aISBN10
ISSN	011 \$a
PPN	003 <a href="http://www.sudoc.fr/PPN">http://www.sudoc.fr/PPN</a>
Langue(s)	101 \$a (+ éventuelle note en 302 si multilingue)
Cote	930 \$a
ID N°Bibliographique	L035 ##\$aALPnuméro aleph
Date du cours	
lieu du cours	
Niveau du cours	
Discipline/Classification	

- **Traitement des données collectées**

- - Nettoyage et validation des titres du corpus

Le corpus constitué lors de cette première extraction a fait l'objet de tris successifs : documents hors sujet puis tri en fonction des éditeurs, des collections et des auteurs ; à l'issue de ces opérations de sélection, 5600 notices ont été conservées.

Ce nouveau corpus a fait l'objet d'une vérification titre à titre et 400 titres suscitant des doutes ont été ensuite validés « livre en main » par la chercheuse associée au projet. Le corpus final validé comporte 3390 items et concerne « les cours professés dans les facultés de droit donnant lieu à examen et/ou diplôme ; les cours de droit professés en dehors des facultés de droit pour des étudiants et validés par un examen et/ou un diplôme ; toute publication en droit dont l'objectif affirmé était l'aide à l'étude, la révision ou le passage d'examen des étudiants. Tout ce qui relevait des enseignements pratiques a été exclu. Enfin, seuls les ouvrages faisant l'objet d'une publication commerciale ont été intégrés » (*Extrait du rapport scientifique*).

Le corpus nettoyé et enrichi a été validé par Anne-Sophie Chambost avant son importation vers Heurist.

### **Formats des données**

Les 13376 notices bibliographiques extraites du catalogue Cujas et validées sont originellement au **format Unimarc**.

Pendant la première phase du projet, les données sont extraites et stockées sous la forme d'un **tableau Excel au format xls**. Ce tableau sera stocké et archivé au format csv en fin de projet.

### **Volume des données**

Volume de l'extraction du catalogue Aleph = 13376 notices

Volume au 1er tri = 5600 notices

Volume du corpus final validé = 3390 notices (productions pédagogiques)

Volume du corpus (format xml) importé dans Heurist (octobre 2020) est de **1,11 Mo**

## **Base de données Hoppe Droit (Heurist-HumaNum)**

Le corpus importé dans la base Heurist Algo\_Hoppe est constitué des 3390 notices validées lors la 1ère phase du projet.

### **Enrichissement des données dans la base Heurist Algo\_Hoppe**

L'enrichissement s'est effectué sur le corpus préalablement validé par la chercheuse associée au projet.

- complétude des notices et normalisation des données Auteur (notices d'autorité) et Editeur (s'appuie sur la Bibliographie de la France)
- enrichissement des notices : ajout d'un indice du plan de classement Cujas si manquant, ajout d'identifiants et de liens vers les notices d'autorité.

Les données sont réparties au départ du travail dans 4 entités :

- 1) Production pédagogique qui correspond à la notice bibliographique du document
- 2) Auteur qui correspond à l'auteur de la production pédagogique
- 3) Editeur qui correspond à l'éditeur commercial de la production pédagogique
- 4) Collection qui correspond à la collection dans laquelle la production pédagogique est éditée (non encore traitée)

### **Modèle de données pour la base Heurist**

<b>Donnees descriptives de l'entité Production pédagogique</b>	<b>Donnees descriptives de l'entité Auteur (personne physique)</b>	<b>Donnees descriptives de l'entité Editeur (personne morale)</b>	<b>Données descriptives de l'entité Collection</b>
<i>Données retenues comme pertinentes</i>	<i>Données retenues comme pertinentes</i>	<i>Données retenues comme pertinentes</i>	<i>Données retenues comme pertinentes</i>
	Nom		
Titre	Prénom	Nom entité morale	Titre de collection
Complément de titre	Date de naissance - date de mort	Lieu	Sous-collection(s)
Date d'édition	Nationalité	Propriétaire	ISSN
Mention d'édition	Lieu de naissance	Relation avec structure(s) précédente(s)	Directeur de collection
Collection	Grade universitaire	Relation avec structure(s) suivante(s)	Suite de
Titre d'ensemble	Université et/ou école de rattachement	Relation avec groupe éditorial	Devient
Titre de partie	Niveau(x) enseigné(s)	Date de création	Date de création
Auteur(s) principal(aux)	Matière(s) enseignée(s)	date de disparition	Date de disparition
Auteur(s) secondaire(s)	Intitulé(s) cours	Nature de la disparition	Identifiant (PPN)
Fonction de l'auteur	Langue d'enseignement	Lien Collection(s)	Editeur(s)
Sujet(s)	fonction éditoriale	Lien Editeur(s)	Lien Editeur(s) pour coédition ou pour changement d'éditeur
Format	responsabilité(s) administrative(s)	Domaine(s)	Nombre d'items dans la collection
Nombre de pages	Activité(s) professionnelles, associatives et/ou éditoriales	Nature de l'activité (Editeur/libraire)	
Illustration	lien documentation extérieure (bibliographie...)	Mention ds les publications de l'université	
Editeur(s)	lien illustration et/ou photo		
ISBN	Nom d'épouse		
ISSN	Nom d'auteur		
PPN			
Langue(s)			
Cote			
ID N°Bibliographique			
Date du cours			
lieu du cours			
Niveau du cours			
Discipline/Classification			
	Référentiels	REFERENTIELS	
	SIPROJURIS : <a href="http://siprojuris.symogih.org/">http://siprojuris.symogih.org/</a>	BnF : <a href="http://catalogue.bnf.fr/recherche- autorite.do?pageRech=rat">http://catalogue.bnf.fr/recherche- autorite.do?pageRech=rat</a>	Bibliographie de la France
	BnF : <a href="http://catalogue.bnf.fr/recherche- autorite.do?pageRech=rat">http://catalogue.bnf.fr/recherche- autorite.do?pageRech=rat</a>	IdRef : <a href="https://www.idref.fr">https://www.idref.fr</a>	
	IdRef : <a href="https://www.idref.fr">https://www.idref.fr</a>		

Un mappage des données en **xml** a permis l'import des données sous ce format dans la base Heurist.

La base de données Hoppe Droit sur Heurist peut être exportée sous les formats suivants : csv, xml, json, Gephi. Les deux premiers formats sont utilisés pour le stockage et le seront pour l'archivage.

#### **Volume des données**

Volume du 1er import du corpus (format xml) dans Heurist (octobre 2020) est de **1,11 Mo**

Volume de l'export de la base de données Heurist (janvier 2022) est de **13,686 Mo**

## **Interface de visualisation de Hoppe Droit**

Données moissonnées telles quelles de la base Heurist Algo\_Hoppe : aucun autre traitement

Dans l'outil de visualisation, les données issues d'une recherche sont exportables au format csv.

## **2. Documentation et qualité des données**

### **La documentation du projet**

La documentation concernant les opérations de tri et d'enrichissement du corpus est disponible dans l'espace partagé par l'équipe projet. Par ailleurs, la synthèse méthodologique est également disponible en accès libre dans le rapport scientifique disponible sur le site CollEx Persée [https://www.collexpersee.eu/wp-content/uploads/2021/07/Rapport\\_scientifique-HOPPE.pdf](https://www.collexpersee.eu/wp-content/uploads/2021/07/Rapport_scientifique-HOPPE.pdf)

### **Structuration des données et les métadonnées**

Les données extraites ont été retravaillées pour correspondre aux attributs des entités identifiées et sur lesquelles repose la structure de la base Heurist. En effet, si le mappage a permis de faire correspondre au maximum les zones Unimarc des notices d'origine avec les attributs définis pour la base de données, certains attributs assignés aux entités correspondaient quelquefois qu'à une partie de zone Unimarc, et d'autres ne correspondaient à aucune zone comme par exemple le niveau du cours, le lieu du cours, etc.

Les 4 entités identifiées :

- 1) Production pédagogique qui correspond à la notice bibliographique du document
- 2) Auteur qui correspond à l'auteur de la production pédagogique
- 3) Editeur qui correspond à l'éditeur commercial de la production pédagogique
- 4) Collection qui correspond à la collection dans laquelle la production pédagogique est éditée

Les référentiels utilisés pour le contrôle des métadonnées Auteur et Editeur : données d'autorité BnF, données d'autorité IdRef.

Les données concernant les éditeurs ont été consolidées quand nécessaire par la consultation de la *Bibliographie de la France ou Journal général de l'Imprimerie*

- Validation du corpus par le chercheur partenaire du projet.
- Contrôle des données dans la base de données Heurist

Pour enrichir le corpus, plusieurs listes de contrôle (ou vocabulaires) dans la grille de saisie sont mises en place sur la base de données Heurist afin de garantir la qualité des données dans le futur : Eléments bibliographiques, Domaine d'activité, Eléments de carrière, Grade ou fonction, Niveau de cours, Organisation.

- Identifiants et référentiels

Les notices d'autorité IdRef et/ou BnF ainsi que le renvoi à la notice du catalogue Cujas sont ajoutées systématiquement dans une zone de lien pour les productions pédagogiques, pour les auteurs et pour les organisations. De plus, les notices de production pédagogique sont accompagnées de numéros PPN (numéro d'identification unique de la notice SUDOC), de l'ID numéro bibliographique (correspondance avec la notice bibliographique du catalogue Cujas), de la cote Cujas du document.

- Outil(s) mis en place

A l'issue du projet, un manuel de saisie dans la base Heurist sera rédigé pour garantir une saisie ou des modifications homogènes. Certaines procédures de contrôle qualité seront mises en place.

Un fichier Readme accompagnera le dépôt des données dans Zenodo

### 3. Stockage et sauvegarde pendant le processus de recherche

#### Corpus extrait du catalogue Cujas dans le cadre du projet Collex Hoppe Droit

##### **Stockage des données**

Le corpus de travail sous forme d'un classeur Excel est stocké sur l'espace partagé du service du patrimoine sur l'espace commun de stockage de la BIU Cujas (sauvegarde quotidienne) ainsi que sur les postes personnels des responsables Cujas du projet.

##### **Nommage des fichiers**

**Forme générique :** Nom du projet\_Etat des données\_JJMMAAAA

**Par exemple :**

- Fichier correspondant à l'extraction effectuée à partir du catalogue (ALEPH) :HOPPE\_Extraction\_28092020
- Fichier correspondant à l'étape de tri dans les titres issus de l'extraction :HOPPE\_Corpus\_Tri\_28092020
- Fichier correspondant à l'étape de validation du tri : HOPPE\_Corpus\_Valide\_28092020
- Fichier correspondant au modèle de données (Heurist) :HOPPE\_Corpus\_par\_entites\_En\_cours\_28092020, HOPPE\_Corpus\_par\_entites\_Valide\_28092020
- Fichier correspondant au corpus préparé pour l'import dans Heurist : HOPPE\_Corpus\_pour\_import\_En\_cours\_28092020,HOPPE\_Corpus\_pour\_import\_Valide\_28092020

Sauvegarde quotidienne

Aucune donnée sensible dans ce projet

#### Base de données Hoppe Droit (Heurist-HumaNum)

Stockées et sauvegardées par HumaNum (hébergeur de la base Heurist)

Sauvegarde assurée par HumaNum

pas de données sensibles dans ce projet

#### Interface de visualisation de Hoppe Droit

Sur le serveur du prestataire Ouestware en phase de conception et de test.

Sur un serveur de la BIU Cujas lors de la mise en production : pas de sauvegarde car le code de l'application est disponible sur gitlab et les données sont moissonnées quotidiennement à partir de la base de données Heurist (HumaNum).

Idem

### 4. Exigences légales et éthiques, codes de conduite

Aucune donnée à caractère personnel



---

Toutes les données traitées sont libres de droit.

Les applications créées ou utilisées pour le projet sont libres et ouvertes :

- Base [Heurist](#) (algo\_hoppe) hébergée sur l'infrastructure HumaNum.
- Code de l'interface Hoppe conçue par Ouestware déposé sur [gitlab](#)

---

----

---

## 5. Partage des données et conservation à long terme

### Corpus extrait du catalogue Cujas dans le cadre du projet Collex Hoppe Droit

Le corpus est déposé sur la plateforme zenodo sous une licence CC BY associé à un fichier Readme :

<https://doi.org/10.5281/zenodo.5931542>

Il est prévu de faire un versioning du dépôt du corpus au fil des enrichissements apportés et/ou un dépôt sur la plateforme Recherche Data Gouv après son ouverture.

Aucune restriction ou embargo

---

Les données devraient être archivées de façon pérenne au Cines dans le cadre de la convention d'archivage pérenne entre la BIU Cujas et le Cines. Le projet est actuellement en cours de réflexion. A priori, seront archivés de façon pérenne : le corpus issu de la base de données Heurist à l'issue du projet au format xml et/ou CSV, le PGD, le dossier de candidature à AAP CollEx, la documentation concernant le traitement des données (rapport scientifique et autre documentation et/ou publication). Possibilité de versioning pour l'archivage. L'archivage pérenne devrait intervenir en 2022.

---

DOI attribué lors du dépôt dans Zenodo : 10.5281/zenodo.5931542

URL : <https://doi.org/10.5281/zenodo.5931542>

### Base de données Hoppe Droit (Heurist-HumaNum)

La base Heurist algo\_Hoppe est hébergée sur HumaNum. L'accès est réservé aux partenaires du projet.

---

Dans le cadre du projet d'archivage au Cines, seront archivés de façon pérenne : la documentation concernant la structure de la base de données Heurist Algo\_hoppe.

---

-----

## Interface de visualisation de Hoppe Droit

Code déposé sur gitlab disponible sans restriction ou embargo à l'adresse suivante : <https://gitlab.com/ouestware/cujas-hoppe/>

Archivage sur Software Heritage effectif au premier semestre 2022 :

<https://www.softwareheritage.org/?lang=fr>

Identifiant : swh:1:dir:90f0dc505022ec3e656a33f05fe2c3e6f456ca23;origin=https://gitlab.com/ouestware/cujas-

hoppe.git;visit=swh:1:snp:869616963e66fea3e4b030635731f3799dbd2725;anchor=swh:1:rev:9f41c01edbb431986949e587ed64306611e57087

Lien pérenne :

<https://archive.softwareheritage.org/swh:1:dir:90f0dc505022ec3e656a33f05fe2c3e6f456ca23;origin=https://gitlab.com/ouestware/cujas-hoppe.git;visit=swh:1:snp:869616963e66fea3e4b030635731f3799dbd2725;anchor=swh:1:rev:9f41c01edbb431986949e587ed64306611e57087>

---

---

## 6. Responsabilités et ressources en matière de gestion des données

### contacts

- Co-coordinatrice du projet, gestion et qualité des données : alexandra.gottely@univ-paris1.fr
- Archivage pérenne, PGD et dépôt des données : renard@univ-paris1.fr

l'équipe projet de la bibliothèque a la charge d'assurer la gestion des données (contact : HoppeDroit@univ-paris1.fr)

L'archivage pérenne des données du projet rentre dans le cadre de la convention Cines/Cujas (prix indicatif pour 2021 :1025 Euros TTC par To par an jusqu'à 11 To )