

---

# PGD HyperOtlet

*Plan de gestion de données créé à l'aide de DMP OPIDoR*

**Créateur du PGD :** Henri Sergent

**Affiliation du créateur principal :** Ecole Normale Supérieure

**Modèle du PGD :** ANR - Modèle de PGD (français)

**Dernière modification du PGD :** 22/11/2019

**Financier :** Agence nationale de la recherche (ANR)

**Numéro de subvention :** ANR-17-CE38-0011

## Résumé du projet :

**HyperOtlet** est un projet pluri- et transdisciplinaire qui porte sur une œuvre majeure de l'histoire de la documentation: le Traité de documentation (TD, 1934), par le juriste belge Paul Otlet († 1944).

Pacifiste, convaincu de l'importance de la connaissance comme instrument de paix, internationaliste, engagé dans les associations internationales, auteur d'un essai sur la Société des nations (1917), passionné de bibliographie et de classification, il est aussi l'ardent militant d'un domaine nouveau : la documentation.

Le mot est récent (environ 1870) comme l'activité qu'il désigne, articulée sur les nouveaux impératifs « documentaires » des sociétés industrielles en voie de « démocratisation ».

Sensible à l'accélération de la demande/production documentaire, à la prolifération des supports (photographie, phonographe, téléphone, cinéma), à la diversification des formes (livre, revues, journaux, etc.), P. Otlet crée de nombreux instruments d'organisation des connaissances et « invente » une tradition de savoirs qui prépare le développement d'un « nouveau régime documentaire » (B. Müller).

Objet de nombreuses « relectures » et appropriations, revendiqué comme un « précurseur » (Internet), un « visionnaire » (société de l'information), ou un « utopiste » (cité mondiale de la connaissance, Mundaneum), Otlet est d'abord un acteur de son temps. Le projet HyperOtlet se propose précisément de revenir sur son œuvre séminale afin de la replacer dans sa complexité contextuelle.

Théorie, méthodologie et pédagogie de la documentation, le TD est un objet paradoxal dès lors que le projet documentaire de son auteur est de dépasser le livre, de le dé-livrer de sa clôture, de désintégrer son contenu sur des fiches re-déployées dans une encyclopédie-documentaire. Objet hybride, livre par sa reliure et sa clôture, il est aussi une accumulation de pages/fiches, une agglomération de dossiers, une collection de classeurs, un répertoire et une encyclopédie.

Nous envisageons de le traiter comme un « hyperdocument », comme l'espace d'un dispositif de régulation d'une nouvelle technologie intellectuelle (Pascal Robert) et d'une nouvelle architecture de la connaissance : la documentation.

Comme « hyperdocument », le TD contient en lui-même les conditions et les possibilités de sa mise en « documents », de sa « documentarisation », comme livre il contient la matière pour fixer une nouvelle tradition intellectuelle. C'est donc dans

ces tensions entre technologie, matérialité, organisation des savoirs et invention d'une tradition que nous plaçons notre projet.

HyperOtlet articulera des recherches historiques et documentaires à un dispositif numérique collaboratif en élaborant un écosystème numérique autour de quatre axes :

- la (re)contextualisation (Centre Maurice Halbwachs, CMH) de la culture numérique ;
- une réflexion sur les problématiques actuelles de la documentation dans les humanités numériques (Laboratoire MICA (Médiations, Informations, Communication, Arts), de la « documentarité » et de la « documentalité » ;
- l'élaboration de nouveaux instruments de lecture, de consultation, de circulation dans les œuvres (Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques, Enssib), Maison des sciences de l'homme Paris Nord, Msh-PN).
- l'animation d'une communauté épistémique qui pourra porter une tradition documentaire européenne (MICA, Enssib, Mundaneum);

Inscrit dans les sciences sociales et le développement des Humanités numériques, HyperOtlet se propose de dépasser le modèle désormais classique de l'édition augmentée sur le WEB (texte, hyperliens, multimédias) et ambitionne de développer un nouvel instrument autour d'un nouvel objet numérique, d'un nouveau dispositif : l'hyperdocument (base de données, documents, données). Non seulement un texte augmenté, mais un texte décomposé en documents, susceptibles d'engendrer de nouveaux enrichissements, combinaisons, organisations, régularités, modélisations, simulations.

COORDINATEUR DU PROJET

**Monsieur Bertrand Müller** (CENTRE MAURICE HALBWACHS)

PARTENAIRE

- MUNDA Mundaneum
- Enssib Ecole Nationale Supérieure des Sciences
- de l'Information et des Bibliothèques
- MICA MEDIATIONS, INFORMATIONS, COMMUNICATION, ARTS
- MSHPN Maison des sciences de l'Homme Paris Nord
- CMH CENTRE MAURICE HALBWACHS

**Aide de l'ANR 469 296 euros**

**Début et durée du projet scientifique : décembre 2017 - 36 Mois**

**Chercheur Principal :** Bertrand Müller

**Identifiant ORCID :** <https://orcid.org/0000-0002-0705-2837>

**Contact pour les Données :** Henri Sergent

**Produits de recherche :**

1. CMS <http://hyperotlet-collections.huma-num.fr/> ( Interactive Resource )
2. [http://hyperotlet-collections.huma-num.fr/api/items?item\\_set\\_id=1](http://hyperotlet-collections.huma-num.fr/api/items?item_set_id=1) ( Dataset )
3. Sharedocs du projet HyperOtlet ( Service )

## 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.

# PGD HyperOtlet

---

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

La production des données textuelles est réalisée par les participants au projet. Elle pourra comprendre la mise en place de datavisualisations, ainsi que de contributions diverses recueillies et centralisées sur notre CMS (Omeka S) . Le programme de recherche s'appuie sur des corpus d'archives dispersées (en ligne, au Mundaneum, présentes chez les partenaires, issues de numérisations, et/ou traitements automatisés, extractions manuelles de contenus à partir de textes OCRisés) afin de produire, à l'aide de méthodologies issues des humanités numériques (fouille de texte en python, web sémantique, indexation des entités nommées) des synthèses et des publications scientifiques.

Parmi les outils utilisés figurent :

Omeka S <https://omeka.org/s/>

Apache Solr et/ou Tika <https://lucene.apache.org/solr/>

Open semantic Search <https://github.com/opensemanticsearch>

Python 3 et plus précisément la bibliothèque nltk <https://www.nltk.org/>

Gephi <https://gephi.org/>

D3.js <https://d3js.org/>

La provenance des documents sera inscrite dans des métadonnées Dublin Core.

Parmi les contenus moissonnés et produits figurent :

- pages web (html)
- *notebooks* jupyter (*pynb, py*)
- contenus textuels non structurés issus d'OCR (*txt, pdf*), corpus nettoyés (csv, json) , corpus structurés en xml (epub, TEI-XML).
- images (numérisées puis versées dans une visionneuse IIF (jpg)
- visualisations de données (D3.js)

Le volume est estimé 250Go > 1To d'ici la fin du projet.

Dans la mesure du possible des formats libres et ouverts seront privilégiés à des fins d'interopérabilité et de réutilisation.

## 2. Documentation et qualité des données

### CMS <http://hyperotlet-collections.huma-num.fr/>

Les données sont organisées selon un modèle de données s'appuyant sur les vocabulaires contrôlés suivants afin de structurer nos ressources dans OmekaS (CMS):

- The RDF Concepts Vocabulary (RDF)
- Dublin Core
- Dublin Core Type
- Friend of a Friend
- ORE
- SKOS
- The RDF Concepts Vocabulary (RDF)
- Bibliographic Ontology(bibo)

Les scripts et la documentation du projet sont centralisés sur <https://gitlab.in2p3.fr/henri-louis.sergent/HyperOtlet>

Relecture par les ingénieurs d'études et les chercheurs associés au projet. Validation des formats par

<https://facile.cines.fr/> a des fins d'archivage pérenne.

**[http://hyperotlet-collections.huma-num.fr/api/items?item\\_set\\_id=1](http://hyperotlet-collections.huma-num.fr/api/items?item_set_id=1)**

Question sans réponse.

Question sans réponse.

**Sharedocs du projet HyperOtlet**

Question sans réponse.

Question sans réponse.

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

**CMS <http://hyperotlet-collections.huma-num.fr/>**

Durant le projet ANR : les données sont hébergés sur : cchumflt01.in2p3.fr  
+ sauvegardes locales sur les postes des membres du projet HyperOtlet.

la documentation du projet ainsi que les notebooks python sont disponibles sur : <https://gitlab.in2p3.fr/henri-louis.sergent/HyperOtlet>

La bibliographie partagée (Zotero) du projet est située sur :  
<https://www.zotero.org/groups/1798805/hyperotlet/items?>

En cas d'incident, nous nous repons sur l'expertise de la TGIR Humanum. Nous disposons cependant de sauvegardes locales de nos contenus.

L'accès à nos environnement de travail passe par une identification par eduGain (RENATER),  
<https://services.renater.fr/federation/documentation/generale/presentation-edugain>

**[http://hyperotlet-collections.huma-num.fr/api/items?item\\_set\\_id=1](http://hyperotlet-collections.huma-num.fr/api/items?item_set_id=1)**

Question sans réponse.

Question sans réponse.

## Sharedocs du projet HyperOtlet

Question sans réponse.

Question sans réponse.

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

### **CMS <http://hyperotlet-collections.huma-num.fr/>**

Pas de collecte de données personnelles des utilisateurs (autres que identifiants de connexion à nos outils en ligne n'est envisagée à ce stade du projet [Oct 2019].

Code et données produites sous CC-BY-NC sauf mention explicitement contraire.

L'accord de consortium précise en son article 7 les modalités liées à la propriété intellectuelle des résultats. L'article 8 précise les modalités d'utilisation et d'exploitation des résultats. [Accord de Consortium](#)

Pas de questions éthiques soulevés pour l'instant. Le cas échéant nous nous reporterons sur le Guide "Pratiquer une recherche intègre et responsable" mis en place par le COMETS du CNRS qui décline les principes qui figurent dans la Charte nationale de déontologie des métiers de la recherche, signée en 2015 par la majorité des organismes de recherche français et par la Conférence des Présidents d'Université (CPU).

<http://www.cnrs.fr/comets/IMG/pdf/guide2017-fr-2.pdf>

### **[http://hyperotlet-collections.huma-num.fr/api/items?item\\_set\\_id=1](http://hyperotlet-collections.huma-num.fr/api/items?item_set_id=1)**

Question sans réponse.

Question sans réponse.

Question sans réponse.

## Sharedocs du projet HyperOtlet

Question sans réponse.

Question sans réponse.

Question sans réponse.

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

### CMS <http://hyperotlet-collections.huma-num.fr/>

Partage des données par une API rest :

[http://hyperotlet-collections.huma-num.fr/api/items?item\\_set\\_id=1](http://hyperotlet-collections.huma-num.fr/api/items?item_set_id=1)

Et/ou un entrepot OAI-PMH :

<http://hyperotlet-collections.huma-num.fr/oai?>

Pas d'embargo prévu. Publication, mise à disposition au cours et avant la fin du programme ANR en décembre 2020.

A l'issue du projet le CINES ou le Mundaneum sont des destinations envisagées pour l'archivage des données (présentées sous la forme d'un conteneur Docker ou lxd)

Archivage au CINES après validation par les chercheurs associés au projet, et/ou Versement sur les serveurs du Mundaneum si possible. Au delà des 3ans de l'ANR une mise en containers (Docker lxd) aisément transposables et contenant la configuration logicielle (OS et version des bibliothèques de code en particulier) requise est à envisager.

Navigateurs web à jour pour les données mises en ligne. editeur de type Notepad++, Sublime, Atom. Python 3/anaconda navigator pour les notebooks.

Moissonnage des données dans Isidore, Mise en ligne de tout ou partie des corpus dans OpenEditions.

**[http://hyperotlet-collections.huma-num.fr/api/items?item\\_set\\_id=1](http://hyperotlet-collections.huma-num.fr/api/items?item_set_id=1)**

Question sans réponse.

Question sans réponse.

Question sans réponse.

Question sans réponse.

### **Sharedocs du projet HyperOtlet**

Question sans réponse.

Question sans réponse.

Question sans réponse.

Question sans réponse.

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

### **CMS <http://hyperotlet-collections.huma-num.fr/>**

Henri Sergent (IE) est le gestionnaire de données du projet pour la durée de l'ANR. Charge à lui de gérer le PGD, l'architecture de l'information du projet HyperOtlet.



Chiffrage estimé à 10% du temps de travail du gestionnaire de données

**[http://hyperotlet-collections.huma-num.fr/api/items?item\\_set\\_id=1](http://hyperotlet-collections.huma-num.fr/api/items?item_set_id=1)**

Question sans réponse.

Question sans réponse.

**Sharedocs du projet HyperOtlet**

Question sans réponse.

Question sans réponse.