Plan de gestion de données : Les dossiers de Bouvard et Pécuchet

Stéphanie Dord-Crouslé, Laurene L'Hermite

Plan de Gestion de Données : Les dossiers de *Bouvard et Pécuchet*

Table des matières

---

1) Plan de gestion de données (PGD) du projet Les dossiers de Bouvard et Pécuchet

☐ ***Présentation de la section***

Cette section décrit le PGD: elle présente l'auteur du PGD, les relecteurs du PGD, les autres intervenants assurant la gestion du PGD et, le cas échéant,ses mises à jour.

☐ ***Recommandations :***

Il est utile de désigner un responsable du PGD qui sera la personne à contacter. Il n'est pas nécessairement le responsable scientifique du projet. Il est recommandé d'associer ce responsable à son identifiant ORCID, IdRef, ISNI, IdHal et de nommer l'ensemble des personnes ayant contribué à la rédaction et à la relecture du PGD.
Le PGD évolue au fur et à mesure de l'avancée du projet de recherche et de l'enrichissement des données. Afin de faciliter sa rédaction, il est conseillé d'en produire une première version au début du projet, qui sera modifiée éventuellement en cours de projet, ainsi qu'à la fin du projet et d'indiquer les versions du PGD dans leur ordre antéchronologique en commençant par l'actuelle.

Auteur du plan de gestion des données :

Dord-Crouslé, Stéphanie, IdHAL : stephanie-dord-crousle ; ORCID : 0000-0002-6683-9509, CNRS (UMR 5317 IHRIM), France
Rôle dans le projet : responsable scientifique

L'HERMITE, Laurène, IdRef : https://www.idref.fr/236176927 ; Université de La Rochelle, Centre de recherches en histoire internationale et atlantique (EA1163), La Rochelle, France
Rôle dans le projet : co-auteure du PGD

Version du plan de gestion des données :

PGD V1: 30/10/2021
1 version de ce PGD est actuellement prévue

2) Présentation du projet et responsabilités

☐ ***Présentation de la section***

Cette section décrit le projet ou le corpus sur lequel porte le PGD. Elle décrit le projet, ses objectifs, participants, etc. Ici, nous décrivons le Consortium CAHIER mais vous trouverez en annexe des exemples plus précis basés sur des projets membres de CAHIER

☐ ***Recommandations :***

Si le nom du projet est un acronyme, indiquez également la version développée.
Exemple : Antonomaz ("ANalyse auTOmatique et NumérisatiOn des MAZarinades")
Identifier la/le responsable scientifique du projet : Nom, Prénom, Institution, Laboratoire, Unité de rattachement, Ville, Pays. Mettre en lien son identifiant ORCID (ou ISNI, IdRef, IdHal, ...). Si possible indiquez des données de contact (courriel, téléphone professionnel)
Exemple : Karine ABIVEN (https://orcid.org/0000-0001-9518-1040),
Sens-Texte-Informatique-Histoire (STIH), EA 4509, Université Paris - Sorbonne, Paris IV, France
Précisez également si le projet s'inscrit dans une programmation scientifique financée et les axes scientifiques liés à cette programmation :

- *Axes scientifique d'un Labex*
- *Programme de financement d'un projet ANR, H2020*
- *Axe ou programme scientifique d'une structure de recherche liée au porteur ou à l'équipe projet...*

Nom du projet

Les dossiers de *Bouvard et Pécuchet*

Responsable du projet (principal researcher) et unité de rattachement

Dord-Crouslé, Stéphanie, IdHAL : stephanie-dord-crousle ; ORCID : 0000-0002-6683-9509, CNRS (UMR 5317 IHRIM), France

Rôle dans le projet : responsable scientifique

Financeur(s) du projet et type de financement

Le projet a bénéficié d'un soutien financier spécifique

- du CNRS (appel d'offres « ATIP Jeunes chercheurs » 2006 du Département Sciences humaines et sociales)
- de l'ANR (appel à projets « Corpus et outils de la recherche en Sciences humaines et sociales » du programme Sciences humaines et sociales 2007 ; https://anr.fr/Projet-ANR-07-CORP-0009).
- de la Région Rhône-Alpes (allocation doctorale allouée au projet dans le cadre du Cluster de recherche n° 13 « Culture, patrimoine, création » 2007-2010)
- du Ministère des Affaires étrangères et européennes (Partenariat Hubert Curien Galilée 2009)

- de la TGIR Huma-Num par l'intermédiaire du consortium CAHIER (2012, 2013 et 2018).

Voir http://www.dossiers-flaubert.fr/projet-partenaires-soutiens

Institution / organisme / unité porteuses du projet

l'UMR 5317 IHRIM qui a pris la suite de l'UMR 5611 LIRE le 1er janvier 2016.

Partenaires (identifier les organismes partenaires, ressources et co-financeurs du projet)

Le projet – dans sa version initiale – a été réalisé entre 2006 et 2012

- dans le cadre de l'UMR 5611 LIRE (« Littérature, Idéologies, REprésentations, XVIIIe et XIXe siècles »), unité mixte de recherche qui associait le CNRS, l'Université Lumière - Lyon 2, l'Université Jean Monnet de Saint-Étienne, l'Université Stendhal - Grenoble 3 et l'ENS de Lyon
- avec le soutien technique de l'USR 3385 ISH (unité mixte de services de l'Institut des Sciences de l'Homme)
- de l'ENS-LSH (École normale supérieure - Lettres et sciences humaines) devenue l'ENS de Lyon (École normale supérieure de Lyon)
- et du TGE Adonis.

Il se poursuit aujourd'hui :

- dans le cadre de l'UMR 5317 IHRIM qui a pris la suite de l'UMR 5611 LIRE le 1er janvier 2016
- avec le soutien technique de la TGIR Huma-Num
- au sein du consortium CAHIER.

Voir http://www.dossiers-flaubert.fr/projet-partenaires-soutiens

Descriptif et objectif(s) du projet

Conservés à la bibliothèque municipale de Rouen, les dossiers de *Bouvard et Pécuchet*, le dernier roman – posthume et inachevé – de Gustave Flaubert (1821-1880), constituent un ensemble patrimonial imposant (2 400 feuillets), cohérent, d'importance scientifique et culturelle reconnue. Ils sont porteurs d'une dimension épistémologique singulière : composés pour rédiger une « encyclopédie critique en farce », ils proposent une configuration critique des savoirs au XIXe siècle, originale et révélatrice. Ils forment le socle de la présente édition. Mais d'autres dossiers existent ailleurs qui ont vocation à enrichir le site en rejoignant progressivement et virtuellement leurs semblables. Car c'est l'ensemble de ce chantier documentaire qui a servi à rédiger le premier volume de l'œuvre et aurait dû être réutilisé pour la composition d'un second volume, jamais écrit en raison de la mort soudaine du romancier.

Or, en raison de leur volume, de leur organisation complexe et indéfiniment mouvante, ainsi que de leurs contenus scientifiques extrêmement variés, les dossiers ne peuvent pas être

édités de manière satisfaisante sous une forme imprimée. C'est particulièrement vrai pour les pages préparées en vue du second volume du roman : les annotations que l'écrivain y a portées, indiquant le lieu probable du classement, sont souvent plurielles et obligent à conserver aux fragments textuels une mobilité qui est nécessairement défaite par la fixité d'une édition imprimée.

Dépassant cette limite en recourant au support électronique et à l'encodage XML-TEI intégral du corpus, la présente édition offre l'accès :

- aux images, à la transcription (formats diplomatique et textuel) et aux métadonnées des pages du corpus,
- à un moteur de recherche plein texte,
- à trois bibliothèques permettant d'identifier les références utilisées par Flaubert et de circuler dans le corpus
- et à un outil de production de « seconds volumes » possibles : l'agenceur.

Date et durée

Date de début des travaux : 2006
Date de fin des travaux : non prévue (tant qu'il y aura des dossiers à intégrer)

Mots clés du projet

- Roman -- Dossiers documentaires ;
- Text Encoding Initiative (langage de balisage) ;
- Bibliothèques numériques ;
- Flaubert, Gustave (1821-1880) Bouvard et Pécuchet ;
- Oeuvre inachevée ;
- Reconstitution conjecturale ;
- Edition posthume ;
- Manuscrits inédits ;

Publications (articles, pré-proposition, site web, …)

Site web du projet : http://www.dossiers-flaubert.fr
Le site Les dossiers de Bouvard et Pécuchet s'est vu attribuer un ISSN (International Standard Serial Number) par la Bibliothèque nationale de France : ISSN 2495-9979.
Sont ainsi soulignés et valorisés les enrichissements progressifs du site qui le constituent en ressource intégratrice.
Ressortent en particulier d'une publication en série :

- les reconstitutions conjecturales du « second volume » existant déjà sous forme papier ainsi que les agencements créés par des internautes (après validation par le comité scientifique du projet) qui seront progressivement mis en ligne sur le site
- et l'ajout à venir, sur la plateforme éditoriale, de nouveaux dossiers de notes non conservés à la bibliothèque municipale de Rouen.

Carnet de recherche du projet : https://flaubert.hypotheses.org/

Listes des articles publiés par le projet :
https://halshs.archives-ouvertes.fr/ANR-07-CORP-009

Autres livrables (guides, recommandations, etc.) :
Compte rendu de fin de projet ANR :

- **S. Dord-Crouslé**. Compte-rendu de fin de projet -Projet ANR-07-CORP-009 BOUVARD - *Les Dossiers de Bouvard et Pécuchet de Flaubert. Enrichissement, valorisation, documentation d'un corpus multi supports : Programme " Corpus et outils de la Recherche en Sciences Humaines et Sociales "* 2007. [Rapport de recherche] ANR (Agence Nationale de la Recherche - France). 2012. ⟨halshs-00760914⟩

3) Présentation et description du corpus

☐ ***Présentation de la section***

Cette section décrit le corpus et ses données. Elle décrit de façon plus précise les données du projet, les méthodes appliquées pour les collecter, etc. Ici, nous décrivons les données du Consortium CAHIER dans leur ensemble mais vous trouverez en annexe des exemples plus précis basés sur des projets membres de CAHIER

☐ ***Recommandations :***

Il s'agira de préciser le mode de collecte et l'origine des données, les centres d'archives, bibliothèques ou centres d'études hébergeant les données y compris si les données procèdent d'un moissonnage de ressources en ligne. L'organisation du corpus, l'arborescence des fichiers, le système de nommage et de gestion des répertoires et des fichiers doit être décrite. De même que la nature des données, leurs formats, leur volumétrie (en poids et nombre de fichiers), leur état, etc. Pour que les données soient réutilisables sur le long terme, les formats doivent être ouverts et non propriétaires et les données stockées dans des entrepôts accessibles.

Présentez et décrivez le corpus

Voir http://www.dossiers-flaubert.fr/edition-corpus
Au corpus originel intégralement conservé à la bibliothèque municipale de Rouen s'ajoutent maintenant, peu à peu, les dossiers conservés ailleurs. En dépassant les strictes limites qu'il s'était d'abord fixées (l'édition d'un ensemble patrimonial cohérent conservé à la bibliothèque municipale de Rouen), le site commence à réaliser pleinement le projet scientifique d'ampleur qui est le sien : donner à lire l'ensemble de la documentation réunie par Flaubert pour son entreprise encyclopédique « en farce » et permettre à l'agenceur d'y puiser des matériaux – pour certains inconnus – en vue de la création ou de l'enrichissement de « seconds volumes » possibles.
Ont été ajoutés les dossiers Rousseau, Hegel et Mirabeau.

Période couverte par le corpus, auteur(s) concerné(s)

Auteur : Flaubert, Gustave (1821-1880)
ISNI : 0000000122762442
ARK BnF : http://catalogue.bnf.fr/ark:/12148/cb11902894q
IDRef : 026866552
Période du corpus : Les documents produits par Flaubert l'ont été entre 1860 et 1880.
Mais les ouvrages pris en note ont un empan chronologique bien plus large: antiquité - 1880.

Organisation du corpus

Unités constitutives du corpus
(voir http://www.dossiers-flaubert.fr/edition-unites-constitutives) :

- **La page** (unité matérielle) : L'unité constitutive matérielle du corpus des dossiers documentaires de Bouvard et Pécuchet est *la page*. Un feuillet est formé de deux pages ou folios, un recto et un verso – dont l'un peut être vierge.
- **Les transcriptions** (associées aux pages). Elles sont de 4 types :

    - La *transcription ultra-diplomatique* se présente sous la forme d'un fichier PDF généré à partir d'un logiciel de traitement de texte. Elle reprend toutes les particularités de la graphie du scripteur.
    - La *transcription diplomatique* (format HTML et générée à partir des fichiers XML/TEI) conserve tous les traitements textuels décrits pour la version ultra-diplomatique. En revanche, elle homogénéise et rationalise une partie des dimensions topographiques et graphiques.
    - La *transcription normalisée* (format HTML et générée à partir des fichiers XML/TEI) achève d'homogénéiser le rendu topographique des pages en déterminant et en ne conservant que quelques espaces signifiants (essentiellement deux : la marge et le corps du texte). Mais surtout, elle propose un texte intelligible par tous les lecteurs, débarrassé des particularités et des graphies déviantes propres à chaque scripteur.
    - La *transcription enrichie* (format HTML et générée à partir des fichiers XML/TEI) permet de faire le lien entre les versions diplomatique et normalisée.
- **Les textes** (unités logiques à fondement matériel) : les pages regroupées selon un ordre validé scientifiquement forment des *textes* qui appartiennent à des catégories typologiques homogènes. Il s'agit d'un autre point d'entrée vers la lecture et l'exploitation des Dossiers. Techniquement, chaque texte, que ce soit en version diplomatique ou en version normalisée, présente l'agrégation – au sein d'une page HTML – du contenu balisé en XML/TEI de l'ensemble des pages concernées ; il est doté d'une URL spécifique et est accessible sur le site à partir d'une page de sommaire permettant de lister, type par type, la totalité des textes du corpus selon différents ordres (classement patrimonial, ordre alphabétique des titres, etc.)
- **Les fragments** : les pages sont composées de *fragments textuels*. Ce sont les unités logiques fondamentales de l'édition électronique du corpus : à leur niveau va être vérifiée et promue la mobilité des éléments constitutifs des dossiers documentaires de Bouvard et Pécuchet. La possibilité de créer des reconstitutions conjecturales du second volume du roman repose sur le découpage de l'intégralité du corpus en fragments textuels, opération qui le rend manipulable et infiniment réagençable. Chaque fragment textuel est accessible par l'intermédiaire d'une métadonnée (« Référence bibliographique de fragment ») attachée à la page où il apparaît, et élucidant la référence bibliographique exacte du fragment copié par Flaubert ou l'un de ses collaborateurs.
- **La citation** est le regroupement de tous les fragments présentant la réalisation textuelle de la même référence bibliographique. Chaque citation possède une page dédiée, pourvue d'une URL et présentant toutes les informations nécessaires à son identification.

Plan de nommage des fichiers :

Collection (nom)	**Collection (description)**	**Volume (nom)**	**Volume (description)**	
BnF	NAF	28825	"Littérature - esthétique"	
Montmorency	Musée JJ	495	"Notes sur	Montmorency_4

	Rousseau		rousseau”	95_f_001_r.jpg
Rouen	BM	g 225-3	Feuillets épars	
[Rouen]		Volume 1 cote g 226-1 => à corriger en :	g 226-1	
	Information requise uniquement dans le TEIHeader			
Antibes	Vente Caroline Franklin Groult, 1931	066	“Esthétique de Hegel”	Antibes_066_f_001_r

Mode de collecte et origine des données

Origine des images, manuscrits et autres pièces des dossiers :

- Bibliothèque municipale de Rouen

La numérisation du microfilm de sauvegarde des documents visés par le projet (microfilm acquis au prix public) nous a permis de constituer une base de 3500 images en noir et blanc de qualité médiocre .
Parallèlement, achat de près de 300 images HD couleur (sur 3500) au prix public grâce à une partie du financement reçu de l’ANR.
Manuscrit “définitif” Premier volume, notes, brouillons, plans, scénarios, notes de lecture. Pages préparatoires Second volume
Puis mise à disposition à titre gracieux par la bibliothèque de la numérisation des deux dossiers concernant le *Dictionnaire des idées reçues* (soit 130 images).

- Musée Jean-Jacques Rousseau et Bibliothèque d’études rousseauistes, Montmorency : mise à disposition des images à titre gracieux par convention
- Antibes, vente Caroline Franklin Groult, 1931: images uniquement, issues de collections privées, non référencées.

Etat du corpus numérique, types de données et volumétrie

Le corpus est ouvert et en cours d’enrichissement. De nouveaux dossiers sont régulièrement ajoutés. L’encodage est toujours en cours et en phase d’amélioration.

Il contient :

- Base de données SQL : Données de travail du projet, références bibliographiques (actuellement plus de 20000), etc. Voir par exemple http://www.dossiers-flaubert.fr/index.php?node=bibliotheques – 100 Mo
- Base de données XML : Transcriptions TEI – 3500 transcriptions à terme ; 2000 disponibles actuellement
- Images : Fac-similés de manuscrits – 3500 images (toutes disponibles en ligne), 5 Go
- Images : Fragments d’images découpés (partiellement disponibles – pour 300 images) – 21000 images (prévision), 3Go

Modifications effectuées sur les données, versions, …

Transcription issue d'un traitement de texte puis balisage en TEI.

Autres données créées ou collectées pour documenter et/ou enrichir les corpus constitués.

Trois bibliothèques de références bibliographiques destinées à enrichir et lier les données.
Voir : http://www.dossiers-flaubert.fr/index.php?node=bibliotheques

L'agenceur :
Il s'agit d'un outil informatique de production de "seconds volumes" possibles. Pour utiliser cet outil, il faut préalablement s'identifier (se connecter ou, lors de sa première visite, créer un compte).
Cette démarche permet à chaque demandeur de disposer d'un espace de travail personnel et privé.
Des informations utiles à la prise en main de l'agenceur sont disponibles :

- sur la page "Agencements" de présentation de l'édition
- en cliquant sur le point d'interrogation ("?") qui se trouve en haut et à droite sur certaines pages de l'espace de travail
- ou bien en consultant les pages dédiées du carnet de recherche du projet qui comportent plusieurs tutoriels (par exemple, ici).

Vous pouvez aussi regarder la vidéo expliquant le fonctionnement du site et plus particulièrement celui de l'agenceur.

Métadonnées, créées et standards et formats utilisés

Les métadonnées sont entièrement accessibles sur le site des "Dossiers…".
Exemple : http://www.dossiers-flaubert.fr/cote-Antibes_066_f_002__r-meta
Les métadonnées ne sont pas standardisées et les champs d'indexation sont libres.
Des transcriptions XML/TEI et des descriptions sont toujours en cours de réalisation.

Les métadonnées descriptives, administratives et techniques

Cote, Scripteur du manuscrit, ensemble textuel d'où provient l'extrait, nom du transcripteur.

Les métadonnées structurelles et l'annotation sémantique

Chronologie du document, provenance, classement typologique* et caractéristiques matérielles.
**Les métadonnées de classement* : pour chaque page sont proposés un classement typologique (en fonction des différents types de pages qui existent dans le corpus : notes de lecture, pages préparées pour le second volume, documentation brute imprimée, etc.) ; un classement chronologique (selon la datation plus ou moins précise qui peut être affectée à chaque page en fonction d'informations internes, comme les filiations génétiques, ou externes, la date d'emprunt d'un ouvrage consignée dans le registre d'une bibliothèque ou la mention, dans une lettre, de la période à laquelle une lecture a été faite par le romancier) ; et un classement par scripteur (Flaubert est évidemment le plus largement représenté, mais bien d'autres personnes lui ont apporté leur aide et ont laissé des traces manuscrites dans les dossiers de Rouen, au premier rang desquelles son ami Edmond Laporte, mais aussi

son « disciple » Guy de Maupassant). Ces classements permettent de proposer trois points d'accès au corpus qui s'ajoutent à celui que fournit, par défaut, le classement patrimonial, accessible par les sections du descriptif établi par l'institution de conservation ou par cotes.

Annotations ou métadonnées d'enrichissement

Annotations critiques.
Transcriptions TEI destinées à enrichir les manuscrits : transcription ultra diplomatique, diplomatique, transcription normalisée, transcription enrichie.
Références bibliographiques

4) Modalités de partage, de sauvegarde et de protection des données. Volumétrie des données stockées et espaces choisis.

☐ ***Présentation de la section***

Cette section décrit la documentation produite au cours projet. Il s'agit d'une documentation autre que numérique (sur support papier par exemple). Si elle existe, il est important de la décrire. Cette section décrit également les lieux et infrastructures de stockage des données pendant le projet.

☐ ***Recommandations :***

Il s'agira de préciser ici le matériel physique et les lieux de stockage des données. Idéalement, il faudrait stocker les données dans au moins 2 endroits, éviter le stockage externe et privilégier les outils mis à disposition par l'institution. Pour cela, il peut être nécessaire de savoir quel est le volume approximatif des données à sauvegarder, l'espace de stockage nécessaire, la périodicité des sauvegardes, le nom et la nature du service fourni par l'institution, etc. On peut également indiquer les procédures de sauvegarde mises en place (fréquence des sauvegardes, automatisée ou non ?), les personnes en charge de la protection de ces données et du contrôle de l'accès, le mode de récupération des données en cas d'incident…

Stockage

Actuellement les données sont stockées via les services dédiés d'Huma-Num (Huma-Num Box ?)
Le transfert des données sur l'entrepôt Nakala (service Huma-Num) est prévu sous peu.

Volume des données stockées (qui sera également celui des données à sauvegarder) :

- PHP/JavaScript/CSS...: 892 Mo
- MySQL: 125 Mo
- SOLR: 38 Mo
- XML/TEI: 83 Mo
- Images (des manuscrits sous différents formats): 4,9 Go

Accès, partage et limites (d'accessibilité) des données

Une collection de 900 pages est moissonnée par Isidore

Concernant certaines images et manuscrits issus de collections de bibliothèques ou de fonds privés, il est à prévoir des règles de partage et de réutilisation :
Pour les cotes Rouen g226, g227 et g228 :

- Images consistant en des reproductions de microfilms : « Collections Bibliothèque municipale de Rouen ».
- Images consistant en des reproductions des manuscrits du Dictionnaire des idées reçues : « Collections Bibliothèque municipale de Rouen - photographie société Arkhénûm ».

- Autres images consistant en des reproductions des manuscrits de Bouvard et Pécuchet : « Collections Bibliothèque municipale de Rouen – photographie Thierry Ascencio-Parvy ».

Toute utilisation publique ou commerciale des images doit faire l'objet d'une autorisation préalable. Les demandes sont à adresser à la bibliothèque municipale de Rouen :
par courrier : Bibliothèque de Rouen, 3 rue Jacques Villon, F-76043 ROUEN CEDEX
ou par courriel : bibliotheque@rouen.fr

Pour la cote Montmorency : « Collection musée Jean-Jacques Rousseau - Ville de Montmorency - photographe Laure Querouil »
Toute utilisation publique ou commerciale des images doit faire l'objet d'une autorisation préalable. Les demandes sont à adresser au Musée Jean-Jacques Rousseau et Bibliothèque d'études rousseauistes par courrier :
Musée Jean-Jacques Rousseau et Bibliothèque d'études rousseauistes
4 rue du Mont-Louis
95160 Montmorency
ou par courriel : Rousseau-museum@ville-montmorency.fr

Pour la cote Antibes : « Collections privées »

Concernant l'utilisation des transcriptions :
L'utilisation des transcriptions à des fins privées, à des fins d'enseignement ou de recherche scientifique est autorisée, sous réserve de mentionner ainsi leur origine :

- « Transcription(s) réalisée(s) par [nom du transcripteur] pour l'édition des *Dossiers documentaires de Bouvard et Pécuchet*, sous la dir. de S. Dord-Crouslé, 2012-..., http://www.dossiers-flaubert.fr, ISSN 2495-9979. »

Pour toute publication, demander préalablement l'autorisation à la responsable de l'édition : Stéphanie Dord-Crouslé.
Le corpus complet est à citer comme suit :

- **Gustave Flaubert,** *Les dossiers documentaires de Bouvard et Pécuchet*. Édition intégrale balisée en XML-TEI accompagnée d'un outil de production de « seconds volumes » possibles, sous la dir. de Stéphanie Dord-Crouslé, 2012-..., http://www.dossiers-flaubert.fr, ISSN 2495-9979.

5) Responsabilités et ressources pour la gestion des données

☐ ***Présentation de la section***

Cette section décrit, identifie, présente et nomme les responsables de la gestion des données.

☐ ***Recommandations :***

Afin de respecter les principes FAIR, CAHIER recommande le dépôt de celles-ci dans l'entrepôt Nakala (https://www.nakala.fr/). Ce service de dépôt et de stockage des données est proposé par la TGIR HumaNum pour les SHS. Il assure la gestion pérenne et sûre des données. Utiliser Nakala n'empêche pas de recourir à un second dépôt sur un autre entrepôt ou sur une plateforme institutionnelle.

Responsable de la gestion des données :
Dord-Crouslé, Stéphanie, IdHAL : stephanie-dord-crousle ; ORCID : 0000-0002-6683-9509, CNRS (UMR 5317 IHRIM), France

Évaluation des coûts (budgets, personnels et temps) dédiés à rendre les données FAIR (temps et budgets pour la collecte et la diffusion des données, pour le stockage et l'archivage).

Dans le cadre du Consortium CAHIER, les moyens assumés par l'infrastructure Huma-Num ont concerné les tâches suivantes:

- mise à disposition de moyens matériels tels que des serveurs, machines virtuelles, logiciels dédiés et licences supplémentaires dont les coûts et abonnements ne sont pas supportés par les projets, soit une économie estimée à ~5000€ / an pour chaque projet membre
- mise à disposition de moyens humains (ETP) pour des tâches spécifiques relevant à la fois de la gestion des moyens matériels (serveurs, machines, etc.), du stockage des données et des actions de formation, soit une économie estimée à plus de ~50000€ / an pour chaque projet membre

Equipe engagée dans la gestion des données à différentes étapes du projet (voir http://www.dossiers-flaubert.fr/projet-equipe-technique) :
Développements informatiques

- [2016-...] Pierre-Yves Jallud (CNRS-IHRIM)
- [2014-2015] Jean-Eudes Trouslard (jet@zoulous.com) pour le module "seconds volumes à la demande" :
  Parties plans et agencements de l'espace de travail.
  - module d'extraction des données de la base TEI vers la base mySql
  - conception et développement de l'affichage des agencements et plans d'une reconstitution conjecturale
  - outils de modifications de l'arborescence des agencements et plans
  - génération en pdf du texte de la reconstitution conjecturale
- [2011-2012] Hugo Schuler (CDD)
- [2009] Stéphane Wustner (Stagiaire)
- [2008-2009] Jérémie Lagravière (CDD)

- [2007-2008] Martial Tola (CNRS-ISH)
- [2007-2012, responsable] Raphaël Tournoy (CNRS-ISH)

TEI et ingénierie documentaire
- [2016-…] Maud Ingarao (ENS Lyon-IHRIM)
- [2016-…] Paul Gaillardon (CDD puis CNRS-IHRIM)
- [2016] Christelle Cluze (Stagiaire)
- [2012-...] Nathalie Arlin (Vacataire)
- [2011] Marjorie Burghart (EHESS, CIHAM)
- [2010-2015] Laetitia Faure (CDD puis CNRS-LIRE)
- [2008-2012, responsable] Emmanuelle Morlock-Gerstenkorn (CNRS-ISH)
- [2008] Vanessa Le Rolle (Stagiaire)
- [2007-2011] Christine Berthaud (CNRS-ISH)

Aide à la transcription
- [2011] Cécile Cordier (CDD)
- [2007-2008] Claire Giguet (CDD)

Traitement des images
- [2016-…] Florence Poncet (CDD IHRIM)
- [2011-2013] Françoise Notter-Truxa (CNRS-LIRE)
- [2007-2011] Véronique Églin (INSA, LIRIS)
- [2007-2011] Vincent Malleron (Doctorant, Université Lyon 2, LIRE et LIRIS)
- [2007-2010] Christophe Lemius (CNRS-LIRE)

6) Archivage des données

☐ ***Présentation de la section***

Cette section décrit les données à conserver à court, moyen et long terme, les éventuelles données à détruire ou à laisser sous embargo et indique la durée de cette restriction.

☐ ***Recommandations :***

A l'issue du projet, des jeux de données se prêteront à une conservation à long terme pour une utilisation future, tandis que d'autres données ne nécessitent qu'une préservation à moyen terme car jugées moins essentielles et au potentiel de réutilisation limité, voire, elles pourront être destructibles pour des raisons de légalité ou de confidentialité.

L'archivage n'est pas encore mis en place mais souhaité, et ce pour la globalité des données du projet. Les informations qui suivent sont donc de l'ordre du prospectif.

Plateforme pour l'archivage pérenne des données

CINES

Durée de conservation des données

Illimitée

Volume des données à conserver

La totalité

Outils, méthodes, procédures nécessaires pour accéder à ces données archivées et les réutiliser

voir conditions CINES

7) Partage des données à l'issue / au fil du projet

☐ ***Présentation de la section***

Cette section décrit la politique de dissémination des données. Elle indique s'il existe des limites à la diffusion des données, comment les données pourront être trouvées et réutilisées par les pairs, voire par le grand public.

☐ ***Recommandations :***

Une bonne dissémination des données requiert, dans la mesure du possible, le respect des principes FAIR: les données doivent être trouvables (findables), accessibles, interopérables et réutilisables. Pour être réutilisables, les données doivent être faciles d'accès, identifiables et citables grâce à des identifiants uniques (DOI) et leur usage facilité par l'accompagnement d'une description et de documentations, par des formats ouverts et non propriétaires et par une disponibilité facilitée par un lieu de stockage (entrepôt) ouvert, gratuit et référencé par les moteurs de recherche.

Les données primaires sont accessibles (images + transcriptions, certaines encore en cours). Les métadonnées du corpus sont partiellement moissonnables via OAI-PMH (voir https://www.rechercheisidore.fr/search/?source=10670/2.q6yiyl)

Eléments d'accompagnement qui permettent la réutilisation des données.

Des informations et tutoriels sont présents sur le site des *Dossiers*. Notamment sur la page "Espace de travail" qui renvoie à l'utilisation de l'Agenceur.

Publications sur les données destinées à en améliorer l'exposition

Le carnet de recherche dédié au projet de l'édition numérique des *Dossiers de Bouvard et Pécuchet* : https://flaubert.hypotheses.org/
Ce carnet diffuse les informations et actualités liées au projet et à son évolution, de même qu'il est un lieu d'échanges et de valorisation pour les chercheurs qui souhaitent réutiliser les sources des *Dossiers*.

L'inventaire des pièces du dossier de genèse de *Bouvard et Pécuchet* :
https://flaubert.univ-rouen.fr/ressources/bp_sphere_inventaire.php

Bibliographie non exhaustive :

Stéphanie Dord-Crouslé. Vers une édition électronique des dossiers de Bouvard et Pécuchet. Stéphanie Dord-Crouslé, Stella Mangiapane et Rosa Maria Palermo Di Stefano. *Éditer le chantier documentaire de Bouvard et Pécuchet. Explorations critiques et premières réalisations numériques*, Andrea Lippolis Editore, pp.15-20, 2010. ⟨halshs-00549160⟩

Alexei Lavrentiev, Serge Heiden. Exploration textométrique du corpus des dossiers de Bouvard et Pécuchet. *Revue Flaubert,* Centre Flaubert, 2014, pp.1-12. ⟨halshs-00678874⟩

Stéphanie Dord-Crouslé. Le site et l'état d'avancement du projet Bouvard. Édition des dossiers documentaires de Bouvard et Pécuchet. *Journées d'études internationales des 11*

et 12 décembre 2008, Lyon, École Normale Supérieure - Lettres et Sciences humaines, Dec 2008, Lyon, France. ⟨halshs-00368846⟩

Pierre-Edouard Portier. Manipulations multimodales pour la construction de documents multistructurés. *Colloque: "Bouvard et Pécuchet : les " seconds volumes " possibles - Documentation, circulations, édition"*, ENS de Lyon, dir. Stéphanie Dord-Crouslé, Mar 2012, Lyon, France. ⟨halshs-00678876⟩

Caroline Angé. Édition de fragments : les enjeux de la mise en forme numérique*. colloque: "Bouvard et Pécuchet : les " seconds volumes " possibles - Documentation, circulations, édition"*, ENS de Lyon, dir. Stéphanie Dord-Crouslé, Mar 2012, Lyon, France. ⟨halshs-00678861⟩

Emmanuelle Morlock-Gerstenkorn. La pratique de l'encodage dans le projet d'édition électronique des Dossiers de Bouvard et Pécuchet : quelques exemples. Textes numériques : l'encodage, pratique savante ?, *Séminaire "Édition savante et humanités numériques"* (EHESS), Dec 2011, Paris, France. ⟨halshs-01141447⟩

Emmanuelle Morlock-Gerstenkorn. Les dossiers de Bouvard et Pécuchet de Flaubert - Fragments visuels et fragments logiques au sein du projet d'édition électronique. *Séminaire publication électronique* - IRHT Orléans, Dec 2009, Orléans, France. ⟨halshs-00438078⟩

Vincent Malleron. Outils d'analyse d'image pour les dossiers de Bouvard et Pécuchet : un panorama. *Édition des dossiers documentaires de Bouvard et Pécuchet. Journées d'études internationales des 11 et 12 décembre 2008, Lyon, École Normale Supérieure - Lettres et Sciences humaines*, Dec 2008, Lyon, France. ⟨halshs-00377381⟩

Stéphanie Dord-Crouslé. Fragments textuels et catégories de classement. Un cas d'utilisation de XML-TEI dans le dispositif éditorial du corpus BOUVARD. *Éditions critiques et génétiques en Rhône-Alpes*, Jun 2013, Grenoble, France. ⟨halshs-00838143⟩

Vincent Malleron. Le numérique et l'interdisciplinarité au service des dossiers de Bouvard et Pécuchet : Vers une mobilité retrouvée. *Séminaire de bilan et de prospective du Cluster 13 « Culture, patrimoine, création » mis en place et soutenu par la Région Rhône-Alpes*, vendredi 23 octobre 2009, Château de Montchat, Oct 2009, Lyon, France. ⟨halshs-00426391⟩

Stéphanie Dord-Crouslé, Emmanuelle Morlock-Gerstenkorn. Le "modèle abstrait" du corpus Bouvard : première approche. J*ournée d'étude " Constitution et exploitation de corpus issus de manuscrits - Lectures, écritures et nouvelles approches en recherche documentaire " organisée par Cécile Meynard et Thomas Lebarbé*, Mar 2009, Grenoble, France. ⟨halshs-00368044⟩

Vincent Malleron, Véronique Eglin, Hubert Emptoz, Stéphanie Dord-Crouslé, Philippe Régnier. *Hierarchical decomposition of handwritten manuscripts layouts. Computer Analysis of Images and Patterns*, Sep 2009, Muenster, Germany. pp.221-228, ⟨10.1007/978-3-642-03767-2⟩. ⟨halshs-00420059⟩

Stéphanie Dord-Crouslé, Emmanuelle Morlock, Raphaël Tournoy. **Nouveaux objets éditoriaux. Le site d'édition des dossiers documentaires de Bouvard et Pécuchet (Flaubert)**
Les Cahiers du numérique, Lavoisier, 2012, 7 (3-4/2011 " Empreintes de l'hypertexte. Rétrospective et évolution ", sous la dir. de Caroline Angé), pp.123-145. ⟨10.3166/LCN.7.3-4.123-145⟩

Vincent Malleron, Stéphanie Dord-Crouslé, Véronique Eglin, Hubert Emptoz, Philippe Régnier. Extraction automatisée de lignes et de fragments textuels dans les images de manuscrits d'auteur du XIXe siècle. *MAnifestation des JEunes Chercheurs en Sciences et Technologies de l'Information et de la Communication*, Nov 2009, Avignon, France. ⟨halshs-00443548⟩

Conditions de réutilisation : licences et contrats pour l'ensemble du projet

Conditions indiquées plus haut pour les images et manuscrits (voir : Accès, partage et limites d'accessibilité des données).
Pour les transcriptions : obligation de citation.