



HAL
open science

IThAC Plan de gestion de données à 6 mois

Malika Bastin-Hammou, Pascale Paré-Rey, Emmanuelle Morlock, Anne Garcia-Fernandez

► **To cite this version:**

Malika Bastin-Hammou, Pascale Paré-Rey, Emmanuelle Morlock, Anne Garcia-Fernandez. IThAC Plan de gestion de données à 6 mois. ANR (Agence Nationale de la Recherche - France). 2024. halshs-04624988

HAL Id: halshs-04624988

<https://shs.hal.science/halshs-04624988v1>

Submitted on 25 Jun 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.



Open licence - etalab

IThAC

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

Créateurs du PGD : Malika Bastin, Pascale Paré, Anne Garcia-Fernandez

Affiliation du créateur principal : CNRS

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

Dernière modification du PGD : 28/07/2020

Financier : Agence nationale de la recherche (ANR)

Résumé du projet :

L'ANR IThAC a pour objectif l'étude de la réception du théâtre antique dans l'Europe du XVI^e s. à travers l'analyse du corpus des paratextes savants imprimés qui lui sont alors consacrés, et la mise à disposition de la communauté scientifique de la traduction de ce corpus en français, grâce à la construction d'une interface numérique évolutive. On fait l'hypothèse que la collecte, la traduction et l'analyse de ce corpus, longtemps négligé parce que difficilement accessible matériellement et parce que très largement rédigé en latin, voire en grec, permettront de saisir à la fois comment le théâtre antique a été reçu et compris par ses « inventeurs » dans l'Europe du XVI^e s., mais aussi comment les idées et les méthodes qu'ils véhiculent, à l'heure où s'inventaient aussi bien le théâtre moderne que la philologie, ont circulé et se sont développées grâce notamment à leur large diffusion rendue possible par l'imprimé.

Piloté par Malika Bastin-Hammou et mené en collaboration avec l'UMR HiSoMA autour d'une équipe de 15 chercheurs membres de Litt&Arts et d'HiSoMA, le projet a obtenu un financement de 48 mois par le plan d'action de l'ANR 2019.

Chercheur Principal : Malika Bastin-Hammou

Identifiant ORCID : 0000-0003-3478-3368

Contact pour les Données : Malika Bastin-Hammou

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

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données préexistantes réutilisées sont des facsimilés soit accessibles en ligne, soit commandés à des bibliothèques. La provenance des données sera indiquée dans la présentation de chaque fichier sous forme de lien vers la ressource en ligne, quand elle est libre de droits

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

1. Des données textuelles seront collectées au format pdf (données sources).
2. Des données textuelles seront produites aux formats doc (données intermédiaires). Le choix du format doc est guidé par un souci de maniabilité pour la majorité des membres du consortium.
3. Des données textuelles seront produites au format xml-tei (données finales). Le choix du format tei correspond à l'usage le plus répandu dans la communauté de recherche en édition numérique de textes.

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

1. Les données intermédiaires sont assorties de métadonnées détaillées qui ont fait l'objet d'une réflexion collective et qui sont récapitulées dans un fichier type. Elles sont identifiées par un titre standardisé (convention de nommage).
 2. Les données finales en TEI sont assorties de métadonnées récapitulées dans un fichier type (schéma d'encodage). Elles sont identifiées par un titre standardisé (convention de nommage) identique aux fichiers de données intermédiaires correspondants (Ar1586_Frischlinus_p1). Ces métadonnées utilisent les standards de la TEI.
- Ces deux catégories de données sont stockées sur le Cloud de l'UGA dans trois dossiers distincts, identifiables par leur format (pdf, doc, tei).

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

1. Données sources : un responsable par auteur antique assure la collecte du corpus dont il a la charge et en contrôle la qualité, en comparant quand c'est possible diverses versions du texte.
2. Données intermédiaires : un ou plusieurs responsable(s) par auteur antique assure(nt) la transcription et la traduction des données sources; il soumet son travail au collectif des transcrip-teurs-traducteurs qui contrôlent la qualité des données produites.
3. Données finales : un responsable par auteur antique supervise l'encodage du corpus dont il a la charge. Une responsable de la TEI contrôle le respect du schéma d'encodage. Le logiciel Oxygen valide automatiquement les fichiers produits.

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données et métadonnées sont stockées et sauvegardées sur le Cloud de l'UGA en première phase du projet. En phase suivante, un laboratoire numérique permettra à chaque membre autorisé d'interagir via un fil de discussion sur les documents en cours de travail.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

L'accès au Cloud de l'UGA est contrôlé par la coordinatrice du projet. Elle gère les autorisations permanentes et temporaires. Les membres du consortium ont tous une autorisation d'accès permanente. Les collaborateurs temporaires (stagiaires, vacataires) ont une autorisation d'accès temporaire.

Le risque principal concerne le droit d'auteur pour les traducteurs-trices. Chaque fichier TEI mentionne dans les métadonnées le nom du ou des traducteurs-trices.

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Dans l'objectif de gestion de comptes utilisateurs sur l'interface web du projet (email, nom et prénom), des données personnelles sont traitées.

Ces données sont complétées par l'utilisateur lors de son inscription : courriel et nom d'utilisateur. Elles sont accessibles de façon sécurisée par les administrateurs du site web. Elles sont stockées sur le serveur du projet (hébergement institutionnel et sécurisé proposé par la TGR Huma-Num). Elles ne feront l'objet d'aucun transfert, seront accessibles et modifiables selon la loi en vigueur et seront conservées pour une durée limitée (à fixer) après la date de dernière utilisation du service.

De plus, conformément à l'Art. 6-II de la loi n°2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique ; décret n°2011-219 du 25 février 2011 relatif à la conservation et à la communication des données des journaux, nous conservons les journaux de connexion pendant une durée minimale d'un an.

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Les données sources et intermédiaires du corpus seront accessibles aux seuls membres du projet. Les données finales du corpus seront librement accessibles au terme du projet.

Les droits de propriété intellectuelle des traductions du corpus sont détenus par les traducteurs qui les diffusent avec une licence creative commons.

Les droits de propriété intellectuelle des annotations et commentaires sont détenus par les commentateurs qui les diffusent avec une licence creative commons.

Pas d'usage commercial.

En cas de litige, le service juridique des tutelles interviendra.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Sans objet

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

Les données intermédiaires ne seront pas partagées; les données finales le seront librement au terme du projet (2024)

Une sélection des données sera partagée via une publication papier.

Plan de préservation des données et archivage pérenne : service Nakala de la TGIR HumaNum qui propose une procédure simplifiée d'archivage sur le CINES

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Toutes les données finales seront conservées sur le long terme.

Les publications réalisées à partir des données le seront également.

Elles seront préservées par le CINES.

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les données textuelles en xml-tei peuvent être lues par n'importe quel éditeur textuel.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

Nakala fournit des identifiants de type Handle

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données

(c'est-à-dire le gestionnaire des données) ?

Malika Bastin-Hammou, collaboratrice principale, UGA, est responsable de la mise en œuvre du PGD, de sa révision, et de ses mises à jour régulières. Elle est également responsable de la saisie des données, production des métadonnées, qualité des données, stockage et sauvegarde, archivage et partage des données en collaboration avec :

Pascale Paré (responsable du WP2 collecte, transcription et traduction des paratextes) est responsable de la saisie des données

Anne Garcia Fernandez (responsable du WP3 Encodage, construction des outils d'analyse et diffusion électronique) est responsable du stockage et de la sauvegarde des données

La coordination des responsabilités de gestion des données entre partenaires se fait via le Cloud de l'UGA et des réunions régulières.

Le PGD sera mis à jour annuellement.

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Les données seront traitées par les 14 membres du consortium ainsi que par une équipe de stagiaires et de vacataires.

Les frais de stockage pris en charge par la TGIR HumaNum et nos tutelles (Cloud de l'UGA).

Le budget de l'ANR comprend le coût de matériel (4 ordinateurs et clés USB) et de prestations (stagiaires et vacations) pour le traitement des données.