



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



I-BioLex

Fragmentation and defragmentation of the law on biomedical innovations



Deliverable D9-1: Data Management Plan

Version	Modifications	Date	Authors
0.0	Initial Document Creation	09/03/2021	A. Mahalatchimy
1.0	First draft	14/06/2021	A. Mahalatchimy, E. Gennet, M. Mbengue, M. Perron; C.-S. Donati



I- BioLex Deliverable D9-1: Data Management Plan





I. SOMMAIRE

TABLE DES MATIERES

- I. Sommaire 2**
- II. Concernant les auteurs..... 4**
- III. Introduction..... 5**
 - 1. Le projet ANR I-Biolex..... 5**
 - 2. Objectifs..... 6**
 - 3. Méthodologie 6**
 - 4. Métadonnées du plan de gestion des données 7**
- IV. Plan de gestion des données 8**
 - 1. Description des données et collecte ou réutilisation de données existantes 8**
 - 1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ? 8
 - 1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ? 9
 - 2. Documentation et qualité des données 11**
 - 2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ? 11
 - 2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ? 12
 - 3. Stockage et sauvegarde pendant le processus de recherche 13**
 - 3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ? 13
 - 3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ? 13
 - 4. Exigences légales et éthiques, codes de conduite..... 14**
 - 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 ? 14
 - 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 ? 15
 - 4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ? 16
 - 5. Partage des données et conservation à long terme 16**
 - 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 ? 16
 - 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) ? 17
 - 5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ? . 18
 - 5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ? 18





I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



6. Responsabilités et ressources en matière de gestion des données 19
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) ? 19
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) ? 19

V. Définitions..... 20



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



II. CONCERNANT LES AUTEURS

Aurélié MAHALATCHIMY est chargé de recherche en droit au CNRS, au sein de l'UMR 7318 DICE CERIC, CNRS-Aix-Marseille Université-Université de Pau et des Pays de l'Adour-Université de Toulon et du Var. Elle est la coordinatrice scientifique du projet ANR I-BioLex. Elle est aussi co-responsable avec M. Perron de la tâche 9 'Coordination, management, communication and spreading' dont relève ce rapport.

Eloïse GENNET est post-doctorante en droit, au sein de l'UMR 7318 DICE CERIC, CNRS-Aix-Marseille Université-Université de Pau et des Pays de l'Adour-Université de Toulon et du Var, dans le cadre du projet ANR I-BioLex.

Mbathie MBENGUE est étudiant en M2 Information Communication à l'Ecole de Journalisme et de Communication d'Aix-Marseille. Il effectue un stage en communication au sein de l'UMR 7318 DICE CERIC, du 15 février 2021 au 30 juillet 2021, dans le cadre du projet ANR I-BioLex.

Martine PERRON est ingénieure d'étude en communication et valorisation au CNRS, au sein de l'UMR 7318 DICE CERIC, CNRS-Aix-Marseille Université-Université de Pau et des Pays de l'Adour-Université de Toulon et du Var. Elle est co-responsable avec A. Mahalatchimy de la tâche 9 'Coordination, management, communication and spreading' dont relève ce rapport.

Caroline Sophie DONATI est Chargée de mission données de la recherche, Département de l'ingénierie documentaire - Cellule Science Ouverte - Service commun de la documentation, à l'Université d'Aix-Marseille.

III. INTRODUCTION

1. LE PROJET ANR I-BIOLEX

Le projet ANR JCJC AAPG 2020 I-BioLex 'Fragmentation et défragmentation du droit des innovations biomédicales' est porté par Aurélie Mahalatchimy, chargé de recherche en droit au CNRS, au sein de l'UMR 7318 DICE CERIC, CNRS-Aix-Marseille Université-Université de Pau et des Pays de l'Adour-Université de Toulon et du Var. Ce projet de 4 ans (2021-2024) visant à étudier les processus de fragmentation et de défragmentation du droit européen applicable aux innovations biomédicales (principalement thérapie génique, médecine régénératrice et nanomédecine).

Les objectifs du projet sont d'explorer et d'expliquer ces processus juridiques et d'en décrypter les développements dans le temps dans le domaine du droit européen des innovations biomédicales. Tout en s'appuyant sur les travaux existants consacrés au cadre juridique des innovations biomédicales, à la temporalité du droit et au phénomène de fragmentation en droit, le projet I-BioLex utilisera des approches comparatives/comparatistes et interdisciplinaires, et combinera une portée théorique et conceptuelle à une dimension pratique de modélisation des processus décisionnels afin de déterminer comment le droit des innovations biomédicales peut remplir différents objectifs sociétaux.

Ce projet est principalement un projet de recherche en sciences humaines et sociales, mais il est aussi interdisciplinaire en ce qu'il implique des chercheurs en sciences biomédicales et en sciences pharmaceutiques.

L'équipe du projet I-BioLex mobilise :

- 5 équipes françaises rattachées à des laboratoires de l'Université d'Aix-Marseille :
 - o Au sein du Centre d'Etudes et de Recherches Internationales et Communautaires (CERIC, UMR 7318 Droits international, comparé et européen, Faculté de droit et de science politique) : Estelle Brosset (Professeure de droit public) ; Aurélie Mahalatchimy (Chargée de recherche au CNRS, Droit) ; Sandrine Maljean-Dubois (Directrice de recherche au CNRS, Droit) ; Martine Perron (Ingénieure d'études au CNRS, Communication et valorisation), Nathalie Rubio (Professeure de droit public), Pascal Gauttier (Gestionnaire de projets de recherche) ; Eloïse Gennet (Post-doctorante en droit) ; Mbathe Mbenge (Stagiaire en communication).
 - o Au sein de l'Institut Louis Favoreu- Groupe d'Etudes et de Recherches sur la Justice Constitutionnelle (ILF-GERJC, UMR 7318 Droits international, comparé et européen, Faculté de droit et de science politique) : Laurence Gay (Chargée de recherche au CNRS, Droit) ; Xavier Magnon (Professeur de droit public).
 - o Au sein du Laboratoire méditerranéen de sociologie (LAMES UMR 7305, Maison méditerranéenne des sciences de l'homme) : Marc Bernardot (Professeur de sociologie).
 - o Au sein du Centre de recherche en cancérologie de Marseille (CRCM) : Christian Chabannon (Professeur de biologie cellulaire), Hashim Wamutitu (Stagiaire en sciences biomédicales).
 - o Au sein du Centre de recherche en Cardiovasculaire et Nutrition (C2VN, Faculté de Pharmacie) : Florence Sabatier (Professeure d'hématologie et de biothérapies).
- 2 laboratoires britanniques de l'Université d'Oxford (Michael Morrisson, Research Fellow in Science and technology Studies, HeLEX) et de la Queen's University de Belfast (Mark Flear, Reader in law)
- 1 laboratoire de l'Université de Monash en Australie (John Gardner, Senior Lecturer in Sociology).

Le Comité scientifique consultatif regroupe quatre personnalités connues pour leur expertise dans les domaines du projet : Alex Faulkner (Sociology/STS, University of Sussex), Emmanuelle Rial-Sebbag (Law, Université Toulouse III-InsERM), Virginie Tournay (Sciences Po CEVIPOF) et Andrew Webster (Sociology/STS, University of York).



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



Dans I-BioLex, la tâche 9 est dédiée à la coordination, la gestion, la communication et la diffusion de l'information. Elle inclut l'élaboration d'un Plan de Gestion des Données (PGD) au mois 6, ainsi que ses mises à jour. Ce rapport contient la première version du PGD d'I-BioLex et constitue donc le 'Deliverable D9-1 : Data Management Plan'.

2. OBJECTIFS

Ce rapport a pour objectifs de montrer comment, tout au long du projet I-BioLex et après, les données issues de ce projet sont gérées :

- Selon des bonnes pratiques concernant notamment la collecte, le stockage, la documentation, etc...
- Dans le respect des exigences éthiques et juridiques relatives à la protection et la diffusion des données, aux conditions de partage et aux modalités de réutilisation, etc...
- Dans le respect des principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable)
- Selon une approche visant à faciliter les échanges entre les collaborateurs du projet
- Selon une approche d'anticipation des besoins afin de gagner du temps pour les publications et la valorisation des résultats.

3. METHODOLOGIE

L'élaboration de ce plan de gestion des données résulte de l'alliance des compétences multiples et d'un travail collaboratif approfondi de ses auteurs, basé sur la méthodologie suivante :

- Recherches générales sur la définition, les objectifs et le contenu d'un plan de gestion des données d'un projet de recherche : consultation du site de l'ANR sur le PGD, du site DMP OPIDOR, de cours en ligne de l'INIST, des modèles existants de PGD et de PGD publics.
- Formation « Plan de gestion des données » en ligne de 3 heures sur le PGD, organisée par l'Université d'Aix-Marseille (formation assurée par Mme Caroline Sophie Donati - Chargée de mission données de la recherche - Département de l'ingénierie documentaire - Cellule Science ouverte - Service commun de la documentation) à laquelle tous les auteurs ont participé au début du projet.
- Choix et prise en main de l'outil DMP OPIDOR et du modèle de PGD de l'ANR.
- Identification des jeux de données d'I-BioLex : Jeu 1- Textes ; Jeu 2- Questionnaires ; Jeu 3- Images
- Recherches sur les outils nécessaires à une gestion FAIR des données d'I-BioLex ; choix et prise en main de ses outils :
 - o Pour les questionnaires, Limesurvey a été privilégié pour éviter de recourir à Google Form ou MS Form, dont l'utilisation n'est pas recommandée par l'ANR.
 - o Pour le partage et le stockage des données, AMUbox et MyCore, outils issus d'établissements publics (respectivement Aix-Marseille Université et le CNRS), ont été privilégiés à Google Drive et DropBox.



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



- Pour l'archivage des données, HAL-SHS déjà utilisée au sein de l'UMR 7318 DICE CERIC continuera à l'être dans le cadre d'I-BioLex. L'entrepôt Zenodo a été choisi pour sa dimension interdisciplinaire.
- Réunions et discussions régulières entre les membres d'I-BioLex pour décider comment assurer une gestion FAIR des données du projet.

4. METADONNEES DU PLAN DE GESTION DES DONNEES

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

Créateur du PGD : Aurélie Mahalatchimy

Affiliation du créateur principal : Université d'Aix-Marseille

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

Dernière modification du PGD : 14/06/2021

Financier : Agence nationale de la recherche (ANR)

Numéro de subvention : ANR-20-CE26-0007-01

Résumé du projet :

Le droit européen des Innovations Biomédicales (IB), dans les domaines tels la thérapie génique, la médecine régénératrice ou la nanomédecine, implique deux processus distincts : la fragmentation (division, segmentation) et la défragmentation (rapprochement, regroupement). Au lieu de les considérer comme des forces opposées, l'hypothèse de la recherche est qu'ils participent conjointement à l'adaptabilité et à la cohérence du droit européen des IB. Les objectifs sont d'explorer et d'expliquer ces processus juridiques et de produire des généalogies –au sens d'histoires de développements successifs- de ces deux processus. S'appuyant sur les travaux liés au cadre juridique des IB, à la temporalité du droit et au phénomène de fragmentation en droit, I-BioLex recourt aux approches comparative et interdisciplinaire et combine éléments théoriques et empiriques pour tester cette hypothèse et savoir comment le droit peut servir divers objectifs sociétaux.

Chercheur Principal : Aurélie Mahalatchimy

Identifiant ORCID : <https://orcid.org/0000-0002-4365-2143>

Contact pour les Données : Aurélie Mahalatchimy

Produits de recherche :

1. Jeu 1- Textes : Jeu 1- Analyse de textes (Texte)
2. Jeu 2-Questionnaires : Jeu 2-Questionnaires (Texte)
3. Jeu 3- Images : Jeu 3- Images (Image)

Nom du DMP sur OPIDoR : I-BioLex- FR Version 1 (14062021)

IV. PLAN DE GESTION DES DONNEES

1. DESCRIPTION DES DONNEES ET COLLECTE OU REUTILISATION DE DONNEES EXISTANTES

1A. COMMENT DE NOUVELLES DONNEES SERONT-ELLES RECUEILLIES OU PRODUITES ET/OU COMMENT DES DONNEES PREEXISTANTES SERONT-ELLES REUTILISEES ?

JEU 1- ANALYSE DE TEXTES

Collecte de données existantes

- Recherche documentaire : collecte de textes à caractère normatif
 - Collecte de texte en droit européen ou en droit français et britannique, en libre accès sur les sites internet concernés tels que eurlex.europa.eu ; coe.int ; legifrance.fr ou encore legislation.gov.uk.
 - Collecte de textes à caractère normatif ou lignes directrices d'acteurs concernés (institutionnels ou de la société civile), en libre accès sur les sites internet concernés, pour ne prendre que quelques exemples, le Comité Consultatif National d'Éthique (ccne-ethique.fr) ou encore les entreprises du médicament (leem.org)

Ces données publiques ne font pas l'objet de restrictions pour leur réutilisation. Lorsqu'elles seront utilisées, elles feront l'objet d'une citation complète conformément aux exigences de déontologie de la recherche scientifique.

- Recherche doctrinale : recherche bibliographique
 - En libre accès sur Internet,
 - À partir des bases de données spécialisées (accès payant grâce aux abonnements de l'Université Aix-Marseille ou du CNRS) : par exemple Dalloz, Cairn, Westlaw, HeinOnline, PubMed, BNDS et de nombreuses autres bases de données sélectionnées pour leur pertinence vis-à-vis des thématiques traitées dans I-BioLex.
 - À partir des ouvrages et revues en format papier disponibles à Aix-Marseille Université
 - À partir d'ouvrages achetés par les membres de I-BioLex

Ces données, soit en accès libre, soit en accès payant, feront l'objet d'une citation complète conformément aux exigences de déontologie de la recherche scientifique.

Méthodologie de sélection

Les critères de pertinence, de sélection et donc d'exclusion des articles de doctrine et des textes à caractère normatif sont multiples. Ils reposent et dépendent des analyses thématiques menées.

En particulier en ce qui concerne les articles de doctrine, chaque membre de l'équipe effectuant des recherches rédigera des fiches descriptives (standardisées pour le projet) de la méthode utilisée dans les bases de données pour la recherche doctrinale (décrivant la date de la recherche, les mots-clés et recherches avancées exactes, les bases de données utilisées, le nombre de résultat, les critères d'exclusion etc.). Ces fiches descriptives seront à usage interne pour une meilleure collaboration et répartition des recherches entre collaborateurs.

Production de nouvelles données



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



Production de données grâce à l'analyse des données précédemment recueillies :

- Recherche documentaire/normative : analyse qualitative via le logiciel NVivo et sur documents .docx, .xlsx, .pptx
- Recherche bibliographique/doctrinale : analyse textuelle des textes dans Acrobat Reader via le logiciel Zotero, et sur documents .docx, .xlsx, .pptx

JEU 2-QUESTIONNAIRES

De nouvelles données seront produites par la diffusion et la collecte de questionnaires :

- En ligne par email grâce à l'outil Limesurvey d'Aix-Marseille Université dont la sécurité est assurée par la Direction Opérationnelle des Systèmes d'Information (DOSI) de l'Université. Une première liste de destinataires cibles sera élaborée grâce à l'identification des organisations pertinentes et diffusion à leurs membres par un lien URL ouvert. Les destinataires seront également invités à partager ce lien à toute personne de leur réseau qu'ils estimeraient concernée (snowball sampling)
- Distribution de ces mêmes questionnaires sous format papier dans des conférences ciblées. Les réponses collectées sous format papier seront ensuite rentrées dans Limesurvey afin de centraliser l'analyse par la suite.

À partir de ces nouvelles données, des données d'analyse seront produites et mises par écrit (cf. jeu 1 Textes) grâce aux fonctionnalités d'analyse qualitative et quantitative de Limesurvey.

JEU 3- IMAGES

Utilisation de données existantes

- Recherche d'images vectorielles de stocks libres de droit dans les collections *Shutterstock* ou *Adobe Stock* (ex Fotolia)

Si l'accès à ces deux banques est payant, il est disponible pour le projet I-BioLex grâce aux abonnements contractés par Aix-Marseille Université.

- Utilisation fortuite de produits graphiques sous licence

Ces produits pourraient être utilisés dans le projet en citant la source après avoir obtenu l'accord de l'auteur pour son utilisation et sa gratuité.

Production de nouvelles données

- Création d'images à partir de logiciels tels Adobe Indesign, Photoshop ou Illustrator pour produire principalement du jpeg, gif, tiff, png. L'abonnement payant à ses logiciels de la chargée de communication d'I-BioLex implique la propriété des images créées par les tutelles institutionnelles de la coordinatrice du projet. Ces images seront rendues disponibles et pourront être réutilisées avec citation de l'auteur et de la source, conformément aux droits d'auteurs (Cf. 5a).

1B. QUELLES DONNEES (TYPES, FORMATS ET VOLUMES PAR EX.) SERONT COLLECTEES OU PRODUITES ?



JEU 1- ANALYSE DE TEXTES

Collecte de données existantes

Textes au format docx, pdf, png, jpeg. Les textes individuels ne dépassent que rarement 1Mo.

Production de données

- Données produites à partir du logiciel NVivo

Les « projets » générés à partir du logiciel NVivo (recueil et codage des textes à caractère normatif) sont aux formats nvpj et qdpj et à terme, pourront dépasser 1Go. Plusieurs projets seront créés (par exemple, un projet droit de l'Union, un projet droit français, un projet droit britannique etc).

- Données produites à partir du logiciel Zotero

Les « bibliothèques » générées à partir du logiciel Zotero (recueil de la doctrine) sont au format rdf, et peuvent être exportées sous plusieurs formats. Au besoin, Zotero RDF sera privilégié.

Des "bibliographies" seront créées à partir de Zotero au format rtf ou html, en vue de leur intégration dans des documents au format docx ou pdf.

Des "Rapports Zotero" ou des "chronologies Zotero" pourront être créés au format html ou pdf.

Aurélié Mahalatchimy centralise toutes les bibliothèques Zotero sur son espace Zotero de stockage illimité en ligne.

Le volume de ces données n'est pas connu à ce jour.

- Autres données produites

Les analyses seront rédigées aux formats docx, pptx, xlsx et éventuellement convertis au format pdf car plus sécurisé pour la diffusion. Le format docx, pptx et xlsx seront systématiquement préférés puisque compatibles avec l'espace de travail collaboratif AMUbox (cf ci-après), contrairement aux formats anciens doc, ppt et xls.

Le volume de ces données n'est pas connu à ce jour.

JEU 2-QUESTIONNAIRES

Questionnaires numériques

- La structure du questionnaire peut être exportée à divers formats depuis Limesurvey, notamment lss, lsa, xml, html, txt, mais également pdf grâce à l'export en format imprimable. Le format pdf sera privilégié.
- Les réponses collectées sur Limesurvey peuvent être exportées directement dans des logiciels d'analyse statistique tels SPSS. Mais dans I-BioLex, ils le seront plus probablement aux formats pdf, docx, xlsx, html ou csv, formats proposés pour l'exportation dans Limesurvey.

Questionnaires papiers

- À partir des fichiers PDF générés sur Limesurvey, des questionnaires anonymes (associés à une note d'information et à un formulaire de non opposition) seront imprimés pour distribution en main propre à l'entrée des salles des sessions ciblées ou sur les sièges, si possible.
- Les réponses collectées au format papier (en fin de session) seront ensuite intégrées directement dans Limesurvey, afin d'être transformées en réponses numériques.



JEU 3- IMAGES

Supports graphiques virtuels

- Site internet (ELSIBI) créé avec le logiciel CMS Wordpress, logiciel libre utilisé pour créer des sites. Ce choix découle de l'expertise particulière de la chargée de communication Martine Perron et du stagiaire de M2 en Information et Communication Mbathie Mbengue.
- Formats : BMP, TIFF, JPEG ou JPG, GIF et PNG

Supports graphiques matériels

- Affiches, kakemonos, posters, flyers

Les formats privilégiés sont des modèles AO, A4, A3, A5, guidés par l'expertise du personnel de l'organisme, la volonté de réaliser des formats ouverts et standards couramment utilisés dans le domaine du droit lors de manifestations.

Les formats de fichiers utilisés ont des extensions standards : textes (pdf, xls, doc, txt, ou rdf); images (BMP, TIFF, JPEG ou JPG, GIF et PNG).

2. DOCUMENTATION ET QUALITE DES DONNEES

2A. QUELLES METADONNEES ET QUELLE DOCUMENTATION (PAR EXEMPLE METHODOLOGIE DE COLLECTE ET MODE D'ORGANISATION DES DONNEES) ACCOMPAGNERONT LES DONNEES ?

JEU 1- ANALYSE DE TEXTES

Nommage

- Textes à caractère normatif : Liste des références dans un document Word sur AMUbox, et classement de ces textes, en fonction des besoins, sur AMUbox en version PDF nommés de la manière suivante : Institutions ou abréviation du pays- Type de texte (eg. Loi) - Numéro- mots clés du titre- année.
- Références doctrinales/bibliographiques : Nommage automatique sur Zotero avec collection automatique des métadonnées des documents. Sinon : - Nom de l'Auteur en majuscule, 1ère lettre du prénom- mots clés du titre- Abréviation de la revue le cas échéant- année.

Classement

- Les textes à caractère normatif seront classés dans NVivo par système juridique (Union européenne, Conseil de l'Europe, France, Grande-Bretagne), puis par thématique (médicament, dispositif médical etc.), puis encore par étape de développement (recherche fondamentale, essais cliniques, mise sur le marché, fabrication, pharmacovigilance etc.) et enfin par degré de normativité (droit abrogé, droit contraignant, droit non-contraignant).

Les références pertinentes pour plusieurs des thématiques identifiées seront ajoutées dans tous les dossiers et sous-dossiers thématiques concernés.

- Le même système de classement sera adopté pour les références bibliographiques/doctrinales (à l'exception du "degré de normativité"), en utilisant la fonction "tag" de Zotero pour établir des listes de bibliographies thématiques.



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



Les références pertinentes pour plusieurs des thématiques identifiées seront ajoutées dans toutes les bibliographies thématiques concernées.

Citation

Capture standardisée grâce à l'outil Zotero (de préférence par captation automatique grâce au connecteur Zotero sur le navigateur, référencement manuel en cas de besoin ou après vérification pour rectification d'éventuelles anomalies) Pour les textes à caractère normatif comme pour les références bibliographiques, le mode de citation dépendra des styles requis par les éditeurs de revues scientifiques dans lesquelles les résultats du projet seront publiés. Il pourra s'agir soit de styles existants et standardisés prévus par Zotero, soit de styles propres à la discipline juridique, non harmonisés à ce jour.

Métadonnées

Les données seront déposées et publiées sur Zenodo. Elles seront donc exprimées au moment du dépôt dans le standard de métadonnées correspondant et seront dans tous les cas converties/convertibles en Dublin Core.

JEU 2-QUESTIONNAIRES

Un registre descriptif sur Word ou Excel sera élaboré et mis à jour afin de rappeler les organisations et types de personnes identifiées à qui seront envoyés les questionnaires, la date d'envoi et le lieu/format de diffusion des questionnaires (numériques ou papier dans des conférences) .

Ce registre permettra également, notamment pour les questionnaires papiers, de recenser le nombre de questionnaires distribués et le nombre de réponses reçues.

Ce registre et les réponses aux questionnaires seront d'abord stockés sur AMUbox pour analyse. Puis, ces données seront déposées et publiées sur Zenodo. Elles seront donc exprimées au moment du dépôt dans le standard de métadonnées correspondant.

JEU 3- IMAGES

Sur AMUbox, des dossiers seront créés par "projet" ou "événement". Ils comprendront les images collectées et les images modifiées ou produites dans le cadre d'I-BioLex.

Chaque dossier comprendra un fichier avec les métadonnées des images :

- Pour les images collectées : nom de la banque de données- source, date de téléchargement (AAAAMMJJ) et auteur de l'image (Nom et première lettre du prénom. Par exemple : PerronM)
- Pour les images modifiées : nature des modifications, date des modifications (AAAAMMJJ), auteur des modifications (Nom et première lettre du prénom. Par exemple : PerronM)
- Pour les images nouvellement créées : date de création (AAAAMMJJ), auteur (Nom et première lettre du prénom. Par exemple : PerronM)

2B. QUELLES MESURES DE CONTROLE DE LA QUALITE DES DONNEES SERONT MISES EN ŒUVRE ?

JEU 1- ANALYSE DE TEXTES

Vérification de la méthodologie de collecte par plusieurs personnes de l'équipe (revue par les pairs), et reproductibilité de la méthode de recherche de textes grâce aux fiches descriptives. En effet, en particulier pour les articles de doctrine, chaque membre de l'équipe effectuant des recherches rédigera des fiches descriptives (standardisées pour le projet)



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



de la méthode utilisée dans les bases de données pour la recherche doctrinale (décrivant la date de la recherche, les mots-clés et recherches avancées exactes, les bases de données utilisées, le nombre de résultat, les critères d'exclusion etc.). Ces fiches descriptives seront à usage interne pour une meilleure collaboration et répartition des recherches entre collaborateurs.

Pour le nommage des textes, utilisation par plusieurs personnes de ces textes, donc vérification (par les pairs) du nommage correct.

JEU 2-QUESTIONNAIRES

Validation du questionnaire par les pairs (membres du projet), diffusion des questionnaires dans plusieurs conférences ou vers plusieurs communautés (répétition des échantillons), saisie automatique des données sur Limesurvey par les personnes répondant directement aux questionnaires en ligne ou par les membres de l'équipe pour l'intégration numérique des réponses aux questionnaires- papier. La saisie des données sera vérifiée par les pairs dans l'équipe du projet.

JEU 3- IMAGES

La présence des métadonnées associées aux images collectées/modifiées/produites sera vérifiée par les membres de l'équipe (vérification par les pairs au sein de l'équipe) à chaque utilisation des images. Ces métadonnées seront associées aux images lors de leur dépôt sur Zenodo.

3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

3A. COMMENT LES DONNEES ET LES METADONNEES SERONT-ELLES STOCKEES ET SAUVEGARDEES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

L'ensemble de la documentation et du travail en cours est stocké sur AMUbox, y compris les versions successives des projets NVivo, régulièrement exportées et sauvegardées sur AMUbox (nouvelle sauvegarde après chaque activité de codage sur NVivo).

Les services informatiques de l'Université Aix-Marseille font des sauvegardes régulières :

- 1 snapshot par jour sur 90 jour est gardé, permettant de retrouver des versions antérieures des fichiers en cas de problèmes
- 1 sauvegarde par jour sur un robot de sauvegarde est faite également tous les jours, et 1 an de sauvegardes sont conservées

En outre, Aurélie Mahalatchimy effectuera une sauvegarde mensuelle supplémentaire de toutes les données du projet sur le cloud du CNRS MyCoRe.

3B. COMMENT LA SECURITE DES DONNEES ET LA PROTECTION DES DONNEES SENSIBLES SERONT-ELLES ASSUREES TOUT AU LONG DU PROCESSUS DE RECHERCHE ?

Sécurité des données

- Les données relatives au projet sont stockées et sauvegardées sur AMUbox. Sur cette plateforme, deux types de partage existent, mais le plus protecteur sera privilégié : le partage par invitation personnelle à une adresse mail (et a contrario, le partage par lien, qui donne accès aux documents à toute personne possédant le lien créé, sera écarté).



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



- L'ensemble des données du projet sera régulièrement sauvegardé sur MyCore en accès restreint.
- En cas d'incident, les données seront récupérées soit sur AMUbox soit sur MyCore, en fonction de la sauvegarde la plus récente.
- L'accès aux données sera contrôlé par Aurélie Mahalatchimy (coordinatrice du projet I-BioLex) et/ou par Eloïse Gennet (post-doctorante en droit du projet I-BioLex). Les données sont classées dans des fichiers correspondant à chaque tâche du projet. Chaque "fichier-tâche" (I-BioLex-T1; I-BioLex-T2....) est entièrement accessible aux responsables de la tâche, et il contient un sous-dossier partagé avec l'ensemble des collaborateurs impliqués dans des travaux communs concernant cette tâche. Des sous-dossiers peuvent également être accessibles en accès plus restreint ou plus large (implication de collaborateurs externes au projet) si besoin. A côté des "fichiers-tâches", un fichier "I-BioLex- SAG-Shared" est partagé avec les membres du Scientific Advisory Group et un fichier "I-BioLex Consortium- SHARED" est partagé avec l'ensemble des membres du projets.

Protection des données sensibles

- Les recherches documentaires et doctrinales ne comporteront pas de donnée sensible.
- Les questionnaires seront sous forme anonyme, par conséquent, le projet ne traitera aucune donnée personnelle/sensible au cours du processus de recherche
- Les images collectées seront libres de droit ou lorsqu'il s'agit de produits graphiques sous licence, en citant la source après avoir obtenu l'accord de l'auteur pour son utilisation et sa gratuité.

4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE

4A. SI DES DONNEES A CARACTERE PERSONNEL SONT TRAITÉES, COMMENT LE RESPECT DES DISPOSITIONS DE LA LEGISLATION SUR LES DONNEES A CARACTERE PERSONNEL ET SUR LA SECURITE DES DONNEES SERA-T-IL ASSURE ?

JEU 1- ANALYSE DE TEXTES

Aucune donnée personnelle qui n'ait pas déjà été rendue publique auparavant (par exemple le nom, l'affiliation ou l'adresse email des auteurs des articles) ne sera traitée.

JEU 2-QUESTIONNAIRES

Les questionnaires ne contiendront pas de données à caractère personnel.

Ils comprendront d'abord une question sur la catégorie socio-professionnelle des répondants (par exemple, au choix entre industrie, agence réglementaire, chercheur hospitalo-universitaire, autres), puis des questions uniquement d'ordre général ne permettant pas d'identifier les répondants ou de donner des indices sur leur identité en fonction de leurs réponses (par exemple : "Nous visons l'exhaustivité dans le domaine de.... Pensez-vous que nous avons oublié un texte réglementaire parmi la liste des textes figurant ci-dessous ? Et si oui, lequel ou lesquels ?"). Dans tous les cas, les chercheurs du projet I-BioLex n'ont aucun intérêt à connaître l'identité des répondants et ne chercheront pas à le faire.

La collecte des réponses aux questionnaires se fera aussi de manière anonyme : grâce à l'outil Limesurvey proposant cette option pour les questionnaires en ligne; et en fin de sessions de conférences ciblées pour les questionnaires-papier.



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



Enfin, les questionnaires et leurs processus de diffusion seront soumis à l'avis du comité d'éthique d'Aix-Marseille Université avant le début des enquêtes de terrain.

JEU 3- IMAGES

Aucune image comportant des données à caractère personnel ne sera traitée, aucune image de personne identifiable ne sera utilisée.

4B. COMMENT LES AUTRES QUESTIONS JURIDIQUES, COMME LA TITULARITE OU LES DROITS DE PROPRIETE INTELLECTUELLE SUR LES DONNEES, SERONT-ELLES ABORDEES ? QUELLE EST LA LEGISLATION APPLICABLE EN LA MATIERE ?

JEU 1- ANALYSE DE TEXTES

- La collecte de textes à caractère normatif seront issus, sauf exception, de bases de données en libre accès, qu'elles soient juridiques, institutionnelles ou scientifiques.
- La doctrine analysée sera en revanche en majorité à accès payant, c'est pourquoi seules les métadonnées des références bibliographiques (et non les documents PDF eux-mêmes) seront partagées au public, principalement via le site ELSIBI.
- Le CNRS (employeur de la coordinatrice du projet I-BioLex) et l'université d'Aix-Marseille (tutelle de l'Unité de recherche de la coordinatrice du projet I-BioLex) sont copropriétaires des données.
- Aurélie Mahalatchimy et Éloïse Gennet contrôleront l'accès aux données.
 - Les données chaudes (en cours d'analyse) seront accessibles aux membres du projet uniquement via le dossier partagé sur AMUbox
 - Les données froides (publications) seront disponibles en libre accès (licence CC-BY) à la fin de l'embargo de publication prévu par les revues le cas échéant, sur HAL-SHS. Les données froides seront stockées et mises à disposition par l'intermédiaire de l'entrepôt Zenodo (<https://zenodo.org/>).
- Ce projet est soumis à la réglementation classique des droits d'auteurs.

JEU 2-QUESTIONNAIRES

Cf. réponse jeu 1.

JEU 3- IMAGES

Respect des licences et des règles relatives aux droits d'auteur pour les images collectées

- Pour chaque image collectée auprès des banques d'images, l'auteur sera cité conformément aux consignes de la banque en question. D'une part, il sera toujours vérifié que la licence autorise l'usage de l'image (Images libres de droit ou appartenant au domaine public, ou extraites des banques pédagogiques autorisées comme Eduthèque, ou bien licences Creative Commons à condition de respecter les termes de la licence). D'autre part, l'auteur, la source, et la licence seront toujours cités.



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



- Les éventuelles modifications et création d'images seront décrites lors du dépôt dans Zenodo.

Propriété et droits d'auteurs des images nouvellement créées

- Le CNRS (employeur de la coordinatrice du projet I-BioLex) et l'université d'Aix-Marseille (tutelle de l'Unité de recherche de la coordinatrice du projet I-BioLex) sont copropriétaires des données.
- Aurélie Mahalatchimy et Martine Perron contrôleront l'accès aux données.
 - Les données chaudes (en cours de création) seront accessibles aux membres du projet uniquement, voire à certains membres seulement, via le dossier partagé sur AMUbox
 - Les données froides (images finales créées) seront disponibles en libre accès (licence CC-BY) sur les différents supports du projet. Les données froides seront stockées et mises à disposition par l'intermédiaire de l'entrepôt Zenodo (<https://zenodo.org/>).
- Ce projet est soumis à la réglementation classique des droits d'auteurs.

4C. COMMENT LES EVENTUELLES QUESTIONS ETHIQUES SERONT-ELLES PRISES EN COMPTE, LES CODES DEONTOLOGIQUES RESPECTES ?

JEU 1- ANALYSE DE TEXTES

Respect des chartes de citation et de référencement et des bonnes pratiques d'intégrité scientifique.

JEU 2-QUESTIONNAIRES

L'avis du comité d'éthique d'Aix-Marseille Université sera demandé avant la diffusion des questionnaires. Les questionnaires seront anonymes. Leur diffusion et leur collecte le seront également. Dans la mesure où I-BioLex est un projet en Sciences Humaines et Sociales, il est généralement admis en France, et en particulier à l'Inserm, que le consentement n'est pas requis et que seule la non-opposition des participants à la recherche est nécessaire. C'est pourquoi les questionnaires seront accompagnés d'un formulaire de non-opposition et d'une fiche d'information sur le projet.

JEU 3- IMAGES

La collecte et la création des images au sein d'I-BioLex ne soulèvent pas de question éthique particulière. Comme l'ensemble du projet, ces activités seront menées dans le respect de la Charte française de déontologie des métiers de la recherche, signée notamment par l'ANR et le CNRS.

5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

5A. COMMENT ET QUAND LES DONNEES SERONT-ELLES PARTAGEES ? Y-A-T-IL DES RESTRICTIONS AU PARTAGE DES DONNEES OU DES RAISONS DE DEFINIR UN EMBARGO ?

JEU 1- ANALYSE DE TEXTES

Données collectées



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



- Les liens hypertextes vers les documents en accès libre (notamment les instruments de droit) pourront être fournis dans les bibliographies et collections de textes, de même que les liens vers les pages des bases de données de la doctrine (sans que ce lien n'ouvre droit aux PDF si l'accès est payant)

Données produites

- Partage des données dans Zenodo.
- Partage avec Hypothèses.org via le site internet ELSIBI mis en place par la chargée de communication Martine Perron et le stagiaire en communication Mbathie Mbengue.
- L'adaptation et la faisabilité de l'outil Omeka sont à ce stade encore à l'étude.
- Publication d'un policy briefing vers les institutions règlementaires
- Autres moyens de publier vers le public : revues.org, open edition books, Calenda, sites internet des partenaires, site ANR, futur site du projet européen EuroGCT, sites des réseaux de recherche, publications scientifiques (ouvrages, articles), supports de communications graphiques, réseaux sociaux, presse et médias
- Les données produites qui ne seront pas publiées en 'open access' seront systématiquement ajoutées sur HAL-SHS avec respect, le cas échéant, de l'embargo
- Toutes les données rendues disponibles pourront être réutilisées avec citation de l'auteur et de la source, conformément aux droits d'auteurs.

JEU 2-QUESTIONNAIRES

Pour les données d'analyses produites à partir des questionnaires, cf. Jeu 1.
Une fois les données d'analyse publiées, les données brutes seront archivées sous accès contrôlé sur Zenodo, avec validation des demandes d'accès auprès d'Aurélié Mahalatchimy : mais cela pourrait évoluer vers un accès libre.

JEU 3- IMAGES

Les images produites pourront être partagées

- Sur le site Internet ELSIBI ;
- Sur divers supports numériques ou papiers liés au projet
- Sur les réseaux sociaux du blog de recherche ELSIBI (en cours de création)

Toutes les données rendues disponibles pourront être réutilisées avec citation de l'auteur et de la source, conformément aux droits d'auteurs.

5B. COMMENT LES DONNEES A CONSERVER SERONT-ELLES SELECTIONNEES ET OU SERONT-ELLES PRESERVEES SUR LE LONG TERME (PAR EX. UN ENTREPOT DE DONNEES OU UNE ARCHIVE) ?

JEU 1- ANALYSE DE TEXTES



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



Zenodo et HAL-SHS.

JEU 2-QUESTIONNAIRES

Dépôt dans Zenodo.

JEU 3- IMAGES

Dépôt dans Zenodo

5C. QUELLES METHODES OU QUELS OUTILS LOGICIELS SERONT NECESSAIRES POUR ACCEDER ET UTILISER LES DONNEES ?

JEU 1- ANALYSE DE TEXTES

Les seules restrictions à l'utilisation des données peuvent résider dans le manque d'accès aux bases de données doctrinales à accès payant (les références bibliographiques pourront être utilisées sans outil logiciel spécifique, mais les articles auxquels ces références feront allusion seront pour la plupart à accès payant).

La majorité des données seront en accès libre via HAL-SHS, le site Internet ELSIBI et/ou l'entrepôt Zenodo. En cas d'accès restreint sur Zenodo, les demandes d'accès seront traitées en direct.

JEU 2-QUESTIONNAIRES

Cette question est à l'étude. Il s'agit de savoir sous quel format la structure du questionnaire ainsi que les réponses collectées pourront être sauvegardées et réutilisées.

JEU 3- IMAGES

Aucun outil spécifique ne sera nécessaire pour accéder aux images produites, qui seront entreposées en accès libre sur Zenodo.

5D. COMMENT L'ATTRIBUTION D'UN IDENTIFIANT UNIQUE ET PERENNE (COMME LE DOI) SERA-T-ELLE ASSUREE POUR CHAQUE JEU DE DONNEES ?

JEU 1- ANALYSE DE TEXTES

Grâce au partage sur la plateforme Zenodo qui permet d'attribuer un identifiant DOI unique et pérenne.

JEU 2-QUESTIONNAIRES

Le dépôt dans Zenodo procure un DOI.

JEU 3- IMAGES

Le dépôt dans Zenodo procurera un DOI.



I-BioLex is funded by ANR AAPG 2020. GA n°ANR-20-CE26-0007-01



6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

6A. QUI (PAR EXEMPLE ROLE, POSITION ET INSTITUTION DE RATTACHEMENT) SERA RESPONSABLE DE LA GESTION DES DONNEES (C'EST-A-DIRE LE GESTIONNAIRE DES DONNEES) ?

Les principales responsables de la gestion des données et de la bonne mise en œuvre du PGD seront Aurélie Mahalatchimy (Chargée de recherche CNRS et PI de I-BioLex) et Éloïse Gennet (post-doctorante pour I-BioLex). Une première version de ce PGD est élaborée 6 mois après le lancement du projet (M6). Une première mise à jour sera également faite 6 mois plus tard (M12), puis une fois par an (M24 ; M36 et M48).

6B. QUELLES SERONT LES RESSOURCES (BUDGET ET TEMPS ALLOUES) DEDIEES A LA GESTION DES DONNEES PERMETTANT DE S'ASSURER QUE LES DONNEES SERONT FAIR (FACILE A TROUVER, ACCESSIBLE, INTEROPERABLE, REUTILISABLE) ?

Plusieurs membres de l'équipe sont impliqués dans la gestion des données, de la rédaction du plan de gestion des données et de son suivi afin de s'assurer que les données seront conformes aux principes FAIR : Aurélie Mahalatchimy, Éloïse Gennet, Martine Perron, Mbathie Mbengue et Pascal Gauttier.

Le temps cumulé alloué à la première version du PGD est d'approximativement 3 mois. Le temps alloué à chacune des révisions du PGD est évalué à environ 10 jours.

Le temps alloué à la gestion des données permettant de s'assurer que les données seront FAIR tout au long du projet est évalué à environ 1 mois en début du projet (découverte, choix, et prise en main d'AMUbox, MyCore et Zenodo, puis au moins 2 jours par mois tout au long du projet.



V. DEFINITIONS

AMUbox : Service de sauvegarde et de partage de documents en ligne. Les documents sont stockés sur des serveurs d'Aix-Marseille Université (AMU) et accessibles depuis l'internet. La collaboration sur un même document est aussi possible directement depuis AMUbox.

Digital Object Identifler (DOI) : Méthode standardisée pour l'identification pérenne et unique d'un document publié, développée par l'*International DOI Foundation* (www.doi.org).

Données chaudes : Données en cours d'analyse en cours d'analyse qui comprennent à la fois les données brutes (données collectées, non organisées, non mises en forme) et les données traitées (organisées, structurées et prêtes à être analysées).

Données froides : Données achevées ou analysées, telles qu'elles apparaissent dans les publications.

ELSIBI : Nom du site internet créé dans le cadre du projet I-BioLex grâce au logiciel CMS Wordpress et hébergé sous forme de Carnet Hypothèse. Il signifie : *Ethical, Legal and Social Implications of Biomedical Innovations* (Implications Éthiques, Juridiques et Sociales des Innovations Biomédicales) (<https://elsibi.hypotheses.org/>).

HAL-SHS : Archive ouverte pour les Sciences de l'Homme et de la Société destinée au dépôt et à la diffusion d'articles scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, dans toutes les disciplines des sciences humaines et de la société (<https://halshs.archives-ouvertes.fr/>).

Limesurvey : Logiciel gratuit et multilingue d'enquête en ligne permettent de construire questionnaires ou sondages simples comme des formulaires complexes, d'en collecter et d'en analyser les réponses (<https://www.limesurvey.org/fr/>).

Plan de gestion des données (PGD) : Le plan de gestion des données vise à décrire la façon dont les données de recherche seront produites, collectées, réutilisées, stockées, protégées, disséminées ou encore conservées à long terme.

Métadonnées : informations relatives aux données de la recherche permettant de les documenter, d'en assurer la qualité afin qu'elles puissent être facilement retrouvées et réutilisées. En sciences humaines et sociales, le standard usuel des métadonnées est Dublin Core.

MyCore : Service de stockage, sauvegarde et partage sécurisé de documents en ligne, sur les serveurs du Centre National de la Recherche Scientifique (CNRS) et accessibles depuis l'internet.

NVivo : Logiciel d'analyse qualitative de données qui permet de créer un espace de travail, d'organiser et de visualiser facilement ses données. NVivo peut recevoir de multiples sources et formats de données (résultats d'enquête, fichiers vidéos, audios, pages web, articles etc). Nvivo permet aussi de coder les données, d'en identifier et visualiser les liens entre les différentes sources.

Zenodo : Plateforme de dépôt d'archives ouvertes en libre accès développée et hébergée par le service informatique du Centre Européen de Recherche Nucléaire (CERN) (<https://giving.web.cern.ch/content/zenodo-0>).

Zotero : Logiciel gratuit de gestion de références bibliographiques qui permet d'importer, de stocker, d'organiser, de synchroniser et de partager les références, d'en générer des bibliographies ainsi que différentes formes de citation.