

Politique de gouvernance des données de l'Institut du Cerveau

Garantir l'intégrité de la recherche, respecter les valeurs de la science ouverte et valoriser les découvertes issues de la recherche



Table des matières

Table des matières	1
Préambule	3
Objet de la politique	4
Contexte juridique	4
Domaine d'application	5
Résumé	5
Répartition des responsabilités	6
Rôle, droits et responsabilités de l'Institut	7
Rôle, droits et responsabilités du chercheur	8
Création, archivage et mise à jour du plan de gestion données	8
Préparation de la collecte des données	8
Préparation de la collecte des données personnelles	9
Collecte des données	10
Collecte des données personnelles	10
Réutilisation des données	10
Réutilisation des données personnelles	10
Traitement des données	11
Publication des données	11
Archivage des données	12
Archivage des données personnelles	12
Développement de logiciels, de conceptions d'équipement et de scripts d'analyse	13
Annexes	15
Glossaire	16
Documents de soutien et références	18
Crédits	19

Préambule

Ce document présente la politique de l'Institut du Cerveau couvrant l'ensemble des données de recherche. Elle vise à garantir l'intégrité de la recherche et un accès aux résultats aussi ouvert que possible, maximisant leur valeur pour le progrès scientifique.

Le présent document a été rédigé par le Comité de gouvernance des données de l'Institut du Cerveau. Celui-ci se compose de membres du Comité de pilotage scientifique et médical et de représentants de toutes les parties impliquées dans la gestion des données.

Ce document fixe également les lignes directrices de l'Institut quant à la gestion et au partage des données, des logiciels et du code de recherche. Il synthétise les meilleures pratiques imposées ou recommandées aux chercheurs par l'Institut du Cerveau sur l'ensemble du processus de recherche. Il établit un cadre sur lequel les plateformes et les équipes de recherche peuvent s'appuyer (autant que possible) pour élaborer leur propre protocole de gouvernance des données.

L'Institut du Cerveau accompagne les chercheurs dans l'application de cette politique en mettant à disposition les ressources opérationnelles dont ils ont besoin pour mettre en œuvre les meilleures pratiques possibles.

La politique de gouvernance des données s'applique à l'ensemble du personnel travaillant à l'Institut du Cerveau, quels que soient leur employeur et le lieu d'exercice de leur activité. Elle s'applique au personnel scientifique, y compris aux chercheurs, ingénieurs, chefs de projet, doctorants et postdoctorants ainsi qu'à l'ensemble des équipes de soutien techniques, réglementaire et administratif impliquées dans la conception et la mise en œuvre de projets scientifiques. Les équipes d'organismes tiers en collaboration sont également tenus de respecter et d'appliquer cette politique dans leurs activités respectives.

La mise en œuvre de cette politique s'exerce sous la responsabilité globale de la Direction Générale de l'Institut du Cerveau qui soutient avec conviction et engagement cette politique.

Chaque directeur, chaque responsable de plateforme ou de biobanque/banque de données et chaque chef d'équipe est responsable, dans son domaine, de la mise en œuvre de la politique et est tenu d'en garantir l'application avec l'assistance des différentes fonctions support et des réseaux de référents et organes compétents.

La mise en œuvre de cette politique implique le soutien et la responsabilité de chacun. Le personnel pourra bénéficier du déploiement progressif d'équipements et outils essentiels ainsi que d'une organisation visant à faciliter l'appropriation par chacun d'une véritable culture de la gestion des données. Il pourra également s'appuyer sur des règles de bonnes pratiques favorisant la bonne mise en œuvre de cette politique.

L'Institut du Cerveau considère que les jeux de données et logiciels de recherche sont des produits de recherche légitimes et éligibles à une évaluation. De plus, il reconnaît que les activités de gestion des données et de développement logiciel font partie intégrante du processus de recherche et contribuent à la qualité de la recherche. En outre, l'Institut du Cerveau considère les logiciels comme produits de recherche valorisables pouvant donner lieu à rémunération selon l'accord en vigueur avec l'Institut. Les évaluations de l'Institut couvriront notamment les plans de gestion des données et les autres initiatives ayant trait à la gestion et à la diffusion de données et de code logiciel.

Objet de la présente politique

Les données de recherche représentent, pour la recherche académique, une véritable force motrice. La gestion des données de recherche implique l'organisation et la gestion rigoureuses des données de recherche sur l'ensemble du cycle du projet de recherche. Une manipulation rigoureuse des données de recherche est essentielle, non seulement pour assurer la fiabilité, la qualité et la reproductibilité de la recherche, mais également pour favoriser la réutilisation (et la réutilisabilité) de ces données.

À l'Institut du Cerveau, l'écosystème de la recherche (équipes et plateformes) utilise, manipule et génère de grandes quantités de données numériques et non numériques et s'appuie sur le développement de logiciels plus ou moins complexes, allant de scripts courts à des applications plus élaborées. La collecte et la production de données doivent respecter le cadre juridique établi pour protéger les individus vis-à-vis du traitement de leurs données personnelles, préserver la qualité et l'intégrité des données et satisfaire aux droits de propriété intellectuelle de tous les tiers. La communauté de recherche internationale doit également se mobiliser pour assurer la préservation, le partage et la réutilisation des produits de la recherche scientifique. Ce mouvement mondial, impliquant des chercheurs, des décideurs politiques et des financeurs, vise à améliorer la qualité, l'intégrité et la reproductibilité de la recherche.

L'Institut du Cerveau collabore étroitement (et a établi des accords de partenariat) avec des institutions publiques impliquées dans la recherche, l'enseignement supérieur et la médecine (CNRS, INSERM, AP-HP et Sorbonne Université, par exemple) avec d'autres IHU (Instituts Hospitalo-Universitaires) et UMR (unités mixtes de recherche) ainsi qu'avec des partenaires industriels. De nombreuses collaborations avec des institutions internationales sont également en cours, par exemple, via le financement de la Commission européenne ou, aux États-Unis, du NIH (National Institutes of Health). Dans ce contexte, il convient de définir des principes communs, fondements d'une véritable confiance à l'égard de la qualité des données, ainsi que des règles visant à sécuriser l'accès aux données et à en assurer un usage approprié. Cette démarche nécessite un consensus sur des lignes directrices, l'engagement de chacun à les respecter et une organisation s'appuyant sur une expertise spécialisée, aidant les chercheurs à honorer cet engagement. La présente politique représente un jalon majeur de l'initiative de renforcement de la stratégie d'intégrité scientifique entreprise par l'Institut.

Contexte juridique

Toutes les phases de la recherche doivent respecter les dispositions juridiques et réglementaires applicables, notamment le RGPD (règlement général sur la protection des données¹), la loi du 6 janvier 1978 dite « loi Informatique et Libertés »² et le règlement européen et national sur les essais cliniques³. Durant toutes les phases de la recherche, le travail doit être mené dans le respect des codes de conduite applicables, y compris le Code de conduite pour l'intégrité en recherche (Code of Conduct for Research Integrity⁴) et la Charte éthique et déontologique de l'Institut du Cerveau. D'autres codes de conduite relatifs à certaines disciplines telles que la génomique ou l'intelligence artificielle ainsi que ceux développés à l'avenir doivent également être respectés. Les lignes directrices de l'Institut du Cerveau relatives à la gestion et au partage de données et de code logiciel de recherche s'alignent sur les exigences des organismes qui financent la recherche et d'éditeurs de revues. Elles sont conformes aux principes **FAIR**⁵ (Faciles à trouver, Accessibles, Interopérables et Réutilisables) promus par l'association EOSC (European Open Science Cloud⁶).

Domaine d'application

La présente politique s'applique à l'ensemble du personnel scientifique public et privé (chercheurs, ingénieurs, chefs de projet, doctorants et postdoctorants, etc.) travaillant à l'Institut du Cerveau ainsi qu'aux équipes de soutien technique, réglementaire et administratif impliquées dans la conception et la mise en œuvre de projets scientifiques.

Cette politique couvre l'ensemble des données de recherche numériques et non numériques (par exemple les enregistrements factuels (fichiers numériques, données textuelles, images et contenus vidéo ou audio) utilisés comme sources principales pour la recherche scientifique et largement acceptés dans la communauté scientifique, nécessaires pour valider les résultats de recherche).

Elle s'applique à l'ensemble des projets de recherche menés à l'Institut du Cerveau, faisant l'objet de contrats ou d'accords existants. Elle s'applique également à l'ensemble des composants logiciels (scripts, programmes ou workflows) et aux logiciels plus élaborés développés à l'Institut du Cerveau.

Résumé

L'Institut du Cerveau a pour objectif de garantir l'accès le plus large possible à ses recherches de dimension mondiale, conformément au principe « ouvert autant que possible, fermé autant que nécessaire ». La politique de gouvernance des données vise à sensibiliser les chercheurs aux meilleures pratiques et à mettre à leur disposition les ressources opérationnelles nécessaires à leur mise en œuvre. La présente politique de gouvernance des données a pour but de préciser les rôles, droits et responsabilités de chaque partie prenante dans le cadre de la politique de gestion des données et logiciels de recherche de l'Institut du Cerveau. La stratégie de l'Institut du Cerveau consiste à fournir le soutien et les ressources nécessaires à la mise en œuvre de la politique et à solliciter l'avis des responsables, investigateurs principaux et chercheurs en vue d'une révision régulière de la politique.

1 <https://gdpr-info.eu/>

2 <https://fra.europa.eu/en/law-reference/act-ndeg78-17-6-january-1978-data-processing-data-files-and-individual-liberties>

3 <https://www.ema.europa.eu/en/human-regulatory/research-development/clinical-trials/clinical-trials-regulation>

4 https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/guideline-for-promoting-research-integrity-in-research-performing-organisations_horizon_en.pdf

5 <https://www.go-fair.org/fair-principles/>

6 <https://eosc-portal.eu/>

Répartition des responsabilités

Le Conseil d'administration établit le cadre stratégique dans lequel s'inscrit la gestion des données et l'évalue régulièrement (des processus d'évaluation externe seront mis en place). Il met à disposition les infrastructures et ressources adéquates au niveau central pour favoriser une gestion responsable des données. Il supervise la conformité à la politique de données de l'Institut du Cerveau.

La Direction générale de l'Institut du Cerveau a notamment pour responsabilités d'approuver la politique de gouvernance des données de l'Institut et de garantir l'évaluation régulière du protocole de gouvernance des données de l'Institut. Elle endosse la responsabilité, envers le Conseil d'administration, de l'élaboration et de la mise en œuvre de la politique de gouvernance des données au sein de l'Institut et de tout écart vis-à-vis de la politique de données.

Le Comité délégué aux données a pour responsabilité d'établir et de mettre à jour la politique de gouvernance des données avant révision par le Conseil d'administration.

Les responsables en charge de l'encadrement des chercheurs, des collaborateurs ou des étudiants sont tenus d'informer ces derniers des termes de la politique de gouvernance des données, de veiller à ce qu'ils s'y conforment et, lorsque nécessaire, de les sensibiliser à leur responsabilité dans sa mise en œuvre.

L'investigateur principal ou PI (responsable de recherche, chef de projet ou équivalent) porte la responsabilité quotidienne d'assurer une gestion efficace des données générées par ou collectées dans le cadre de ses recherches, de la collecte des données à leur publication, en passant par leur traitement. Il est responsable de la conformité à la politique de gouvernance des données de l'Institut et aux exigences juridiques, éthiques et contractuelles ayant trait aux données de recherche, et ce, tout au long du projet de recherche.

L'investigateur principal rend compte de ces aspects à son superviseur. Il a pour responsabilité d'élaborer le DMP, plan de gestion des données, ou d'en déléguer l'élaboration et d'assurer sa mise à jour et son archivage.

L'investigateur principal a également pour responsabilité d'établir avec les collaborateurs et étudiants impliqués dans le projet de recherche les accords nécessaires ayant trait à la gestion des données de recherche.

Les chercheurs et ingénieurs sont tenus de travailler dans le respect de la politique de données de l'Institut du Cerveau et du protocole de gouvernance des données de l'Institut et, dans le cadre du projet de recherche, d'assurer la gestion des données conformément au plan de gestion des données ainsi qu'aux lignes directrices et codes de conduite applicables. À cet égard, ils doivent notamment s'efforcer de gérer un dossier complet et précis sur leur recherche, documentant des procédures claires sur la collecte des données de recherche, leur utilisation et leur réutilisation, l'accès à ces données et leur conservation ou suppression.

Rôle, droits et responsabilités de l'Institut

L'Institut du Cerveau s'engage à fournir :

- L'assistance, le conseil et les informations nécessaires sur tous les aspects de la gestion des données de recherche.
- Une infrastructure robuste et sûre pour favoriser la gestion appropriée et le stockage sécurisé des données et logiciels de recherche.
- Les outils nécessaires à l'infrastructure de stockage et à la gestion de la capacité de stockage et des accès ainsi que les outils nécessaires à la mise en œuvre de la politique de conservation et de sauvegarde des données.
- Les outils permettant une gestion robuste et conforme des données de recherche.
- Des check-lists et une documentation favorisant l'adoption des bonnes pratiques et la navigation dans les ressources disponibles.
- Formation et soutien individuel pour aider tous ses collaborateurs à adopter les bonnes pratiques et à utiliser les ressources disponibles.

Rôle, droits et responsabilités du chercheur

Création, archivage et mise à jour du plan de gestion des données

Un DMP (**plan de gestion des données**) doit être mis en place dès que possible, au plus tard au moment de la soumission du projet. Une première version du DMP doit être créée sur la plateforme OPIDOR⁷ à l'aide du modèle de l'Institut et mise à jour régulièrement jusqu'à la fin du projet. Le modèle est adapté à la politique de gouvernance des données de l'Institut du Cerveau. Il est accompagné d'explications et d'exemples correspondant aux activités de recherche de l'Institut et est conforme aux exigences de l'ERC (Conseil européen de la recherche) et de l'ANR (Agence nationale de la recherche). OPIDOR est une plateforme développée par l'Inist-CNRS, une unité de service du CNRS. Un groupe de travail associé, à l'Institut du Cerveau, gère les modèles en s'appuyant sur les retours d'information des chercheurs et des plateformes.

L'Institut du Cerveau offre au chercheur un service support via l'adresse dmp@icm-institute.org pour la création de DMP. La plateforme DAC (Data Analysis Core, ou centre d'analyse des données) prendra contact avec le demandeur pour planifier une session d'assistance. Des sessions de formations sont également dispensées régulièrement⁸.

Préparation de la collecte des données

Avant d'entreprendre une recherche impliquant des tiers, il convient d'établir des accords clairs sur la manière dont les données de recherche seront collectées, traitées, utilisées et stockées (et sur la manière d'y accéder) ainsi que sur les droits d'auteur et droits d'exploitation. Ils prennent la forme d'un accord de partenariat, d'un accord de transfert de données (et, le cas échéant, de matériel), d'un accord de consortium ou d'accords similaires et sont rédigés ou revus par le pôle juridique de l'Institut. Une check-list simplifiant la rédaction de l'accord est mise à disposition⁹. Il est fortement recommandé de finaliser préalablement le DMP de sorte qu'il puisse être annexé à l'accord et qu'il ne soit donc pas nécessaire de le redéfinir dans l'accord.

L'Institut du Cerveau encourage les chercheurs à réutiliser les codes et les logiciels (open source) existants avant d'en développer de nouveaux (catalogue de solutions informatiques disponible sur l'intranet). En cas d'utilisation de codes ou de logiciels existants, les conséquences potentielles liées aux licences utilisateur qui leur sont associées doivent être examinées. La DSI (Direction des Systèmes d'Information) offre son assistance en mettant à disposition les espaces de stockage existants au sein de l'infrastructure informatique et héberge un serveur GitLab pour le développement collaboratif et le partage de code.

Les chercheurs doivent vérifier si leur projet de recherche est soumis à des obligations juridiques avant toute création ou collecte de données. Une check-list des obligations juridiques est mise à disposition¹⁰. Les chercheurs peuvent consulter la CART (Cellule d'Appui Règlementaire et Technique) pour obtenir des éclaircissements sur les exigences juridiques, réglementaires et techniques et sur la façon de s'y conformer. La CART étudie les projets qui lui sont soumis chaque mois. Il se compose du pôle juridique, de la DPO (Délégation à la Protection des Données), de la cellule RIPH (recherche impliquant la personne humaine), de la DSI, de la Référente à l'Intégrité Scientifique (RIS), du pôle ROQ (Risques, Organisation et Qualité) et de la plateforme DAC.

Les exigences de stockage, besoins informatiques et autorisations d'accès aux données doivent être définis conjointement avec la DSI (Direction des Systèmes d'Information) avant toute création ou collecte de données et mis à jour tout au long du projet de recherche. Chaque chercheur dispose de son propre espace de stockage sur les serveurs de la DSI (pour le stockage des données de recherche) et peut demander la création d'un espace supplémentaire pour chaque projet de recherche, avec un accès réservé aux personnes impliquées dans le projet. Il est recommandé d'organiser ces espaces de stockage selon un **plan de classement**. Une convention d'affectation de nom commune, précise et cohérente, simplifiant la recherche et l'identification des données selon les **principes FAIR**, doit être adoptée¹¹.

Préparation de la collecte des données personnelles

L'autorisation de la CNIL¹² est requise préalablement à la mise en œuvre du projet à moins que celui-ci ne soit conforme à une méthodologie de référence (MR), ce qui doit être documenté en interne. Les méthodologies de référence sont disponibles ici¹³. Une check-list de conformité aux MR et des modèles de soumission à la CNIL sont disponibles¹⁴. L'évaluation de conformité de premier niveau sera menée par le chercheur à l'aide de ces outils. Il bénéficiera non seulement de l'assistance de la CART, mais également de celle de la DPO pour l'établissement des demandes d'autorisation liées à des recherches stratégiques, complexes ou innovantes, qui doivent être soumises à la CNIL.

L'Institut demande aux chercheurs d'anticiper toutes les solutions concernant l'utilisation future de données personnelles. À cet égard, la meilleure pratique consiste à informer les participants, lors de l'obtention de leur consentement éclairé, de l'utilisation potentielle des données et échantillons biologiques à d'autres fins liées à la recherche médicale et à les orienter vers une ressource d'information spécifique (site Web, par exemple), fournissant des explications ou des renseignements actualisés sur l'utilisation des données. Notre portail de **transparence des données** est disponible à cette fin¹⁵.

Une analyse d'impact sur la protection des données personnelles doit être menée lorsqu'un traitement de données est susceptible de faire peser un risque élevé sur les droits et libertés des personnes physiques, en fonction de la sensibilité des données et du niveau de risque. À titre d'exemple, les données sensibles associées à un risque élevé incluent les données révélant l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale. Il est interdit de traiter des données génétiques et biométriques dans le but d'identifier de manière unique une personne physique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle des personnes. Plus généralement, une analyse d'impact sera systématiquement requise et doit être impérativement fournie à l'appui de la demande d'autorisation en cas de recherche médicale portant sur des patients et/ou sur des mineurs et incluant le traitement de leurs données génétiques ou la constitution d'un registre ou d'une base de données (« entrepôt ») ayant vocation à être ouvert à la communauté de recherche. Une check-list permettant de déterminer la sensibilité des données et le niveau de risque sera mise à disposition. L'analyse est menée sous la responsabilité opérationnelle du responsable scientifique du projet, par les équipes chargées de la mise en œuvre de l'étude, à l'aide des services d'appui réglementaire et technique, et validée par le responsable du traitement après avis de la DPO et du RSSI.

En cas d'échange de **données personnelles** avec des tiers, un accord sur le transfert et l'usage de ces données doit être établi conformément à la législation et vigueur et supervisé par le pôle juridique de l'Institut. Les transferts de données personnelles d'un pays de l'Union européenne vers un pays hors Union européenne n'appliquant pas rigoureusement les dispositions européennes relatives aux données doivent être anticipés et encadrés. Une check-list et des modèles sur les accords de partage de données sont disponibles¹⁶.

⁷ <https://opidor.fr/planifier/>

⁸ <https://dac.institutducerveau-icm.org/service/training/>

⁹ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#checklist-and-templates-data-sharing>

¹⁰ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#checklist-and-templates-data-collection>

¹¹ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#file-naming>

¹² <https://www.cnil.fr/fr>

¹³ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#cnil-reference-methodologies>

¹⁴ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#cnil-checklist-and-templates>

¹⁵ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#data-transparency-portal>

¹⁶ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#checklist-and-templates-data-sharing>

Collecte des données

Pour simplifier l'utilisation et la réutilisation des données, les investigateurs doivent utiliser des formats ouverts, standard et interopérables et créer des **métadonnées** dès la génération des premières données.

Dans la mesure du possible, les **métadonnées** associées à des données non numériques doivent être numérisées conformément aux principes **FAIR**.

L'Institut encourage les chercheurs à contacter la Direction des Systèmes d'Information avant la collecte de données pour se renseigner sur les logiciels et ressources de collecte, de stockage et d'analyse recommandés.

Collecte des données personnelles

Pour les projets impliquant le traitement de données personnelles, les données doivent être **pseudonymisées** avant leur utilisation à des fins de recherche à l'aide de l'outil interne Identity Manager¹⁷. Celui-ci offre le niveau de sécurité approprié, à savoir celui des cahiers d'observation électroniques eCRF (Electronic Case Report Forms), conçus pour assurer la conformité aux obligations juridiques. Une formation à l'utilisation de l'outil Identity Manager est dispensée régulièrement¹⁸.

Réutilisation des données

L'Institut du Cerveau encourage les chercheurs à réutiliser les données dans la mesure du possible. Les **jeux de données** réutilisés doivent être cités dans les publications. L'Institut du Cerveau considère que les **jeux de données** sont des produits de recherche légitimes et citables, en accord avec les principes décrits dans la Joint Declaration of Data Citation Principles¹⁹.

La réutilisation des données doit respecter les dispositions juridiques et exigences contractuelles. Les chercheurs peuvent consulter la CART (Celle d'Appui Règlementaire et Technique) pour obtenir des éclaircissements sur les exigences juridiques et sur la façon de s'y conformer et pour s'assurer que le promoteur de l'étude accepte la réutilisation des données ou échantillons. La CART étudie les projets qui lui sont soumis chaque mois. Il se compose du pôle juridique, de la DPO, de la cellule RIPH (recherche impliquant la personne humaine), de la DSI, du RIS, du pôle ROQ, du RSSI et de la plateforme DAC.

Les jeux de données réutilisés doivent être cités dans les publications.

Réutilisation des données personnelles

Les données doivent être réutilisées de manière transparente pour les titulaires des données (ou personnes concernées). La réutilisation de données personnelles est considérée comme traitement de données et les personnes participant à la recherche doivent donc en être informées. Il est fortement recommandé d'anticiper cette exigence en utilisant un **portail de transparence** des données dès le début de l'étude. Ceci allégera la charge administrative pour les études suivantes. Seule la réutilisation de données personnelles anonymes (et non pseudonymisées) ne fait pas l'objet d'une obligation d'information.

Traitement des données

Les données de recherche doivent être manipulées d'une manière sécurisée, garantissant l'intégrité, la disponibilité et (si nécessaire) la confidentialité des données. Toutes les mesures nécessaires doivent être prises pour prévenir la perte, la dégradation ou toute utilisation ou traitement non autorisé, conformément aux politiques appropriées sur la sécurité et la confidentialité des informations. L'accès aux données et les actions effectuées sur les données doivent être suivis et documentés. Par ailleurs, des contrôles de qualité et de cohérence des données appropriés doivent être mis en place, conformément à ceux définis au début de l'étude.

Les données de recherche doivent être sauvegardées régulièrement sur des espaces de stockage maîtrisés et contrôlés. L'Institut du Cerveau recommande fortement l'utilisation de l'infrastructure fournie par la DSI (Direction des Systèmes d'Informations) pour le traitement et le stockage des données de recherche et des logiciels scientifiques, en fonction de la sensibilité des données.

Des **métadonnées** décrivant les processus appliqués aux données doivent être créées et liées aux données dès leur création et pour toute la durée de leur traitement. Les **métadonnées** doivent décrire la catégorie de données (données de recherche, données personnelles, données sensibles, etc.) afin de mettre en œuvre des solutions de traitement et de stockage des données appropriées. Il est fortement recommandé d'utiliser des cahiers d'observation électroniques eCRF (Electronic Case Report Forms) ainsi que les outils numériques disponibles, favorisant la conduite des processus de gestion des données (tels que REDCap, XNAT et OMERO). La DSI met à disposition une liste des solutions logicielles disponibles²⁰ et peut répondre à toute autre demande.

Publication des données

Avant toute publication de **jeux de données**, les données doivent être stockées et partagées à l'aide d'outils adaptés au niveau de sensibilité de chaque type de données et doivent être accessibles aux personnes autorisées uniquement.

L'Institut du Cerveau encourage la diffusion, la publication et le partage des données et du code logiciel de recherche conformément au principe « aussi ouvert que possible, aussi fermé que nécessaire ». En d'autres termes, les données et le code doivent être partagés dans le respect des obligations juridiques et contractuelles : dispositions ayant trait à la confidentialité (RGPD), obligations contractuelles à l'égard du promoteur ou du responsable de traitement, etc. Une check-list aidant à déterminer si les données peuvent être partagées (et de quelle manière) sera mise à disposition²¹.

L'Institut du Cerveau demande aux chercheurs de publier les données qui sous-tendent l'étude au moment de sa publication. Il les encourage tout particulièrement à publier les résultats négatifs (y compris les données qui ne donneront pas lieu à un article évalué par les pairs). Une liste des entrepôts de données en ligne sera mise à disposition²².

L'Institut du Cerveau encourage la publication de **data papers**, c'est-à-dire des publications évaluées par les pairs, présentant et décrivant un **jeu de données**, citables au même titre que toute autre publication.

Si l'accès aux données après publication doit demeurer restreint, l'Institut du Cerveau demande à ce que les **métadonnées** soient accessibles, de sorte que l'existence des données soit connue de même que leurs conditions d'accès.

Lorsque des jeux de données sont publiés, ils doivent être correctement décrits, documentés et contextualisés. Pour satisfaire aux exigences des organismes financeurs, la publication des données doit respecter les **principes FAIR** et doit toujours être associée à une **licence de diffusion et d'exploitation**.

¹⁷ <https://idmanager.icm-institute.org/login>

¹⁸ <https://dac.institutducerveau-icm.org/service/training/>

¹⁹ <https://force11.org/info/joint-declaration-of-data-citation-principles-final/>

²⁰ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#checklist-and-templates-data-sharing>

²¹ <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#data-sharing-and-archiving-repositories>

Archivage des données

À la fin de chaque projet de recherche et avant sa clôture, les chercheurs sont tenus de déposer les données produites dans un espace de stockage spécifique pour en assurer la pérennité. Pour garantir que les données déposées sont compréhensibles et réutilisables, ils doivent les documenter correctement et les enregistrer dans un format durable, dans la mesure du possible. Une liste de formats durables et d'outils de conversion est disponible [ici](#)²².

Les chercheurs doivent vérifier les règlements et obligations juridiques ou contractuelles de conservation applicables à leurs données. Dans le cas de données personnelles, la loi exige qu'elles soient collectées et conservées uniquement si elles sont « adéquates, pertinentes et limitées » à ce qui est nécessaire au regard des finalités pour lesquelles elles sont collectées ou traitées. D'une manière générale, seules les données dont la conservation est obligatoire ou présente un intérêt à long terme doivent être stockées, notamment :

- o Les données soumises à des obligations juridiques ou contractuelles de conservation.
- o Les données ayant valeur de preuve (preuve d'antériorité, par exemple).
- o Les données pouvant servir à la reproductibilité de travaux scientifiques.
- o Les données uniques, non reproductibles ou difficilement reproductibles.

Archivage des données personnelles

Les données personnelles doivent être conservées, détruites ou anonymisées pendant un projet de recherche ou après sa réalisation conformément au RGPD. Si des données personnelles doivent être mises à disposition d'autres équipes de recherche par la suite, il est important de déterminer :

- o la durée de conservation des données aux fins du traitement de données personnelles, nécessaire à l'étude faisant l'objet de la demande d'autorisation ;
- o la durée de conservation des données aux fins du traitement nécessaire à la constitution de la collection d'échantillons biologiques/de l'**entrepôt de données** destiné à la réutilisation des données à des fins de recherche.

La durée minimale de conservation des données de recherche (à l'exclusion des **données personnelles**) est définie par la législation, des codes de conduite, des conditions contractuelles ou des accords avérés spécifiques, relatifs à la discipline concernée. Il est donc important de veiller à ce que la période de conservation des **données personnelles** soit adaptée aux fins pour lesquelles elles sont traitées et soit justifiée par ces fins. Il convient de faire la distinction entre :

- o l'utilisation en cours (conservation des données en « base active ») lorsqu'elles sont nécessaires, dans l'environnement de travail immédiat, pour les équipes en charge de la mise en œuvre de l'étude (durée des inclusions et des contrôles de suivi et post-publication) ;
- o et l'archivage des données conformément à la réglementation en vigueur, lorsque les données ne sont plus utilisées mais continuent de présenter un intérêt administratif, historique, scientifique ou statistique pour l'organisation.

Développement de logiciels, de conceptions d'équipement et de scripts d'analyse

L'Institut du Cerveau a pour objectif de garantir l'accès le plus large possible à ses recherches de dimension mondiale, conformément au principe « aussi ouvert que possible, aussi fermé que nécessaire. » Si des logiciels, des conceptions matérielles, du code ou des scripts d'analyse peuvent être diffusés, l'Institut du Cerveau recommande de rendre le code source public, de fournir des packages et d'enregistrer les logiciels dans un catalogue de logiciels.

L'Institut du Cerveau encourage tous les chercheurs qui développent des composants logiciels (tels que des scripts, des programmes, des conceptions d'équipement ou des workflows) à suivre les meilleures pratiques de développement, notamment l'utilisation d'outils de gestion de versions, de documentation et de test et d'espaces d'hébergement de code source (par exemple, GitLab). La DSI peut les conseiller sur les directives et normes de programmation. Les procédures d'intégration, de développement et d'hébergement de nouvelles applications sont disponibles sur l'intranet²².

Les chercheurs sont encouragés à inclure les versions de logiciels, de code, de conceptions matérielles, de bibliothèques et de systèmes d'exploitation permettant la régénération de l'environnement d'exécution des logiciels.

Avant la diffusion d'un logiciel, les chercheurs doivent vérifier s'il peut être utilisé à des fins commerciales/industrielles conformément aux règles imposées par les organismes financeurs. Si tel est le cas, ils doivent contacter la Direction de l'innovation sans délai pour établir la déclaration d'invention. De même, lorsqu'une société commerciale contacte des chercheurs au sujet d'un logiciel qu'ils ont publié, ces derniers doivent prendre contact avec la Direction de l'innovation.

Les logiciels publiés doivent toujours être accompagnés d'une **licence de diffusion et d'exploitation**. Le choix de la licence peut avoir des répercussions sur l'exploitation future du logiciel (commerciale, académique, gratuite, etc.).

Pour permettre à des tiers d'identifier et de citer la version correcte du code logiciel (par exemple, la version utilisée et citée dans une publication), l'Institut du Cerveau recommande de soumettre la version publiée dans Software Heritage²³, de sorte qu'elle demeure disponible et citable de façon permanente.

Lorsque des chercheurs réutilisent des logiciels ou composants logiciels de tiers, ils doivent citer chaque composant réutilisé dans leurs publications. L'Institut du Cerveau considère que les logiciels sont des produits de recherche légitimes et citables, en accord avec les Software Citation Principles²⁴.

²² <https://gitlab.com/icm-institute/dac/data-management/-/wikis/home#sustainable-formats-conversion-tools-and-code>

²³ <https://www.softwareheritage.org/>

²⁴ <https://force11.org/info/software-citation-principles-published-2016/>

ANNEXES

Glossaire

Plan de classement

Arborescence de répertoires et de fichiers, nommés de manière intelligible par tous (avec l'acronyme du projet, par exemple), commune à un projet ou à une entité. Le plan de classement simplifie pour chacun la recherche de données et la transmission des informations dans un contexte de rotation régulière du personnel (doctorants, postdoctorants, collaborateurs en CDD, etc.)

Data paper (ou data article)

Publication scientifique évaluée par les pairs dont le but principal est de décrire un ou plusieurs jeux de données. Les données décrites doivent être accessibles, soit sous forme de fichiers annexés, soit par un lien permanent vers l'entrepôt où elles sont stockées. Le data paper peut être publié dans un data journal (une revue qui publie exclusivement des data papers) ou dans une revue scientifique classique.

Données qui sous-tendent les publications

Données nécessaires à la validation des résultats présentés dans les publications scientifiques.

DPM : plan de gestion de données ou Data Management Plan

Document synthétique et évolutif qui aide à organiser et anticiper toutes les étapes du cycle de vie de la donnée. Il est rédigé au début d'un projet de recherche, puis régulièrement mis à jour et explique la façon dont les données de recherche collectées ou générées seront gérées durant le projet et après son terme.

Entrepôt de données

Référentiel (ou dépôt) permettant non seulement de stocker des données de recherche, mais également de les retrouver et de les réutiliser grâce à une description par des métadonnées. Il existe de nombreux entrepôts de données de différents types : disciplinaires, pluridisciplinaires, propres à un éditeur, institutionnels, etc.

Jeu de données (ou dataset)

Groupement de données similaires ou connexes, rassemblées pour former un ensemble cohérent. Un jeu de données doit toujours être accompagné de métadonnées, simplifiant sa compréhension et sa réutilisation.

Licence de diffusion et d'exploitation

Instrument juridique complémentaire au droit d'auteur. Elle permet au titulaire des droits sur une œuvre d'accorder à l'avance aux utilisateurs certains droits d'exploitation de cette œuvre.

Métadonnées

Informations structurées, décrivant une donnée ou une ressource. Elles accompagnent les données pour permettre à la communauté scientifique d'y accéder, de les comprendre et de les réutiliser.

Principes FAIR

Les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable), publiés en 2016, correspondent à des lignes directrices et des bonnes pratiques dont l'objectif est d'améliorer la réutilisation des données de recherche. Ces principes peuvent également s'appliquer aux logiciels de recherche moyennant une légère adaptation.

Standards de métadonnées

Modèles qui spécifient toutes les métadonnées nécessaires pour décrire une ressource. Par exemple, le standard MIGen (Minimum Information about a Genotyping Experiment) décrit toutes les métadonnées nécessaires pour pouvoir comprendre et réutiliser des données issues d'une expérience de génotypage.

Règles de nommage

Il est primordial d'adopter des règles de nommage précises et communes pour localiser et identifier les données, éviter les problèmes lors des transferts et permettre la conservation des données à moyen et long terme. Il convient, par exemple, de donner un nom bref et explicite (incluant l'acronyme du projet), d'éviter les espaces et caractères spéciaux, d'indiquer les versions, etc.

Pseudonymisation

Procédure de désidentification des données : les données directement identifiantes (ex : nom, prénom) dans un enregistrement de données sont remplacés par un ou plusieurs codes de substitution ou pseudonymes. La pseudonymisation permet de rendre un jeu (de données) moins identifiable, tout en restant adapté à l'analyse et au traitement des données.

La pseudonymisation est l'une des méthodes permettant de respecter les exigences du RGPD (règlement général sur la protection des données) relatives au stockage sécurisé des données personnelles. Les données pseudonymisées peuvent être restaurées (récupération de l'état initial) moyennant l'ajout d'informations permettant de réidentifier les individus. L'anonymisation, quant à elle, a pour but d'empêcher la réidentification des individus dans le jeu de données.

Données personnelles

Toute information se rapportant à une personne physique identifiée ou identifiable. La notion de données personnelles est à comprendre de façon très large et fait référence aux données permettant d'identifier une personne directement ou indirectement. L'identification directe d'une personne physique peut être réalisée à partir d'une seule donnée (par exemple, prénom, nom et adresse mail nominative). L'identification indirecte d'une personne physique peut être réalisée à partir d'un identifiant (code de patient, par exemple) ou du croisement d'un ensemble de données correspondant à des facteurs spécifiques propres à l'identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale d'une personne.

Dans la pratique, pour réduire la sensibilité des données, les chercheurs utilisent la pseudonymisation en remplaçant les données d'identification directe (prénom et nom) dans un jeu de données par des données d'identification indirecte (code de patient, par exemple)²⁸.

Plan de gestion de logiciel (ou SMP pour Software Management Plan)

Document évolutif décrivant un logiciel de recherche, la façon dont il est conçu et développé, ses objectifs, à qui il s'adresse, les résultats attendus et obtenus, son éventuelle diffusion, les informations de propriété intellectuelle, etc. tout au long de son cycle de vie.

Recherche

Travail original entrepris sur une base systématique visant à augmenter le stock de connaissances, incluant sciences de l'homme, des civilisations et des groupes humains, et utilisation ce stock de connaissances pour concevoir de nouvelles applications.

Données de recherche

Informations enregistrées (quel que soit le type de support sur lequel elles sont stockées) nécessaires pour étayer ou valider les observations, découvertes ou résultats d'un projet de recherche.

Format durable

Format ouvert (dont les spécifications internes sont librement accessibles), largement utilisé et standardisé ou normalisé (dans la mesure du possible).

Format ouvert, standard et interopérable

Format dont les spécifications fonctionnelles et techniques sont publiques et disponibles gratuitement (ou à faible coût). Un format ouvert et standardisé ou normalisé (par exemple, le format PDF/A normalisé ISO 19005) est interopérable : des données enregistrées dans ce type de format sont indépendantes du logiciel utilisé pour les créer et elles pourront être lues et modifiées à l'aide de tout logiciel conçu pour traiter le type de fichier correspondant (image, texte, audio, etc.).

Identifiant unique et pérenne (ou PID pour Persistent Identifier)

Identifiant permettant l'identification unique d'une ressource numérique et sa citation. Il garantit un lien stable à la ressource en ligne en faisant correspondre en permanence l'identité de la ressource à sa localisation sur le Web.

²⁸ <https://gdpr.eu/article-4-definitions/>

Documents de soutien et références

- Rapport de l'office parlementaire des choix scientifiques et technologiques sur « Pour une science ouverte, réaliste, équilibrée et respectueuse de la liberté académique » - *mars 2022*
- Rapport de l'office parlementaire des choix scientifiques et technologiques sur «Promouvoir et protéger une culture partagée de l'intégrité scientifique » - *mars 2021*
- Feuille de route du CNRS pour la science ouverte - *novembre 2019*
- Plan national pour la science ouverte - *juillet 2018*
- Site de la Commission Européenne : www.ec.europa.eu
- Promouvoir et protéger une culture partagée de l'intégrité scientifique Pratiquer une recherche intègre et responsable - guide - *Comité d'éthique du CNRS - mars 2017*
- The European Code of Conduct for Research Integrity - *All European Academies - 2017*
- Charte de déontologie et d'intégrité scientifique de l'Agence nationale de la recherche - *2018*
- Le cadre juridique dans lequel s'inscrit l'utilisation de données personnelles dans le contexte des activités de recherche de l'Institut repose sur les principaux textes suivants :
 - o Le règlement européen n° 2016-679 du 27 avril 2016, dit RGPD pour « règlement général sur la protection des données » ; <https://www.cnil.fr/fr/reglement-europeen-protection-donnees>
 - o La loi n° 78-17 du 6 janvier 1978, dite « loi Informatique et Libertés », modifiée par la loi n° 2018-493 du 20 juin 2018 et son décret d'application n° 2019-536 du 29 mai 2019 ;
 - o La loi n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine, dite loi Jardé, modifiée par l'ordonnance n° 2016-800 du 16 juin 2016 ;
 - o Les lois n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé et n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé, pour ce qui concerne le Système national des données de santé.
 - o La section « Délégué à la protection des données » de l'intranet de l'Institut du Cerveau inclut une présentation de ces différents textes et des recommandations spécifiques à l'Institut du Cerveau
- Engagement de la Direction et Déclaration de la politique de protection des données personnelles - *28 janvier 2022*
- Management Commitment and Privacy Policy Statement Paris, *28 January 2022*
- Ethics and professional conduct charter of Paris Brain Institute
- Charte d'éthique et de déontologie de l'Institut du Cerveau
- Guide de la recherche de l'Institut du Cerveau
- Guide de la DSI de l'Institut du Cerveau

Crédits

Comité de pilotage scientifique : Jean Christophe Corvol, Brian Lau, Liane Schmidt et Jacobo Sitt

Membre du Comité d'éthique et experte en gestion des données : Ségolène Aymé
DAC (Data Analysis Core) : Mathias Antunes, Stephen Whitmarsh, Violetta Zujovic
DSI (Direction des Systèmes d'Information) : Stéphane Chaillou
DPO (Déléguée à la protection des données) : Frédérique Lesaulnier
Responsable du pôle juridique : Mathilde Gibert
RIS (Déléguée à l'intégrité de la recherche) : Claire Levy-Marchal



Hôpital Pitié-Salpêtrière — 47, bd de l'Hôpital — 75013 Paris

www.institutducerveau-icm.org

ASSISTANCE
PUBLIQUE  HÔPITAUX
DE PARIS



 **Inserm**
La science pour la santé
From science to health

 **SORBONNE
UNIVERSITÉ**