

Règlement sur la gestion des données de la recherche à l'UPHF et à l'INSA HDF.

Résumé

Le règlement des données de la recherche de l'Université Polytechnique Hauts-de-France et de l'INSA Hauts-de-France fixe les **lignes directrices pour la gestion et le partage des données de la recherche**. Ce règlement rappelle les bonnes pratiques que l'UPHF et l'INSA HDF recommandent pour l'ensemble du processus de recherche. Plusieurs services de l'UPHF et de l'INSA HDF accompagnent les personnels pour mettre en œuvre et appliquer la présente politique. Des formations et des sensibilisations aux bonnes pratiques en matière de gestion des données de la recherche sont mises en place et assurées par les acteurs de la donnée au sein des différents services supports de l'établissement.

Contexte

Aux niveaux international et national la science ouverte et l'ouverture des données de la recherche visent à assurer le partage, la pérennisation et la réutilisation des productions et des données de recherche produites par les communautés scientifiques.

L'UPHF a mené une grande enquête suite à l'obtention d'un financement ministériel, et l'analyse de ses résultats a démontré qu'il existait des manquements et des méconnaissances dans le traitement et la gestion des données de la recherche au sein de l'Université Polytechnique Hauts-de-France et de l'INSA Hauts-de-France. Le Comité de Pilotage de la donnée de l'UPHF et de l'INSA HDF s'est emparé de la problématique et a travaillé à la mise en place et à la diffusion de bonnes pratiques en matière de gestion des données de la recherche.

Le déploiement d'un règlement permet aux chercheurs de mieux connaître les obligations et pratiques souhaitées par l'établissement en matière de gestion des données de la recherche.

Domaine d'application

Le présent document est destiné aux membres du personnel scientifique de l'UPHF et de l'INSA HDF :

- Chercheurs,
- Enseignants chercheurs
- Ingénieurs de recherche,
- Ingénieurs d'étude,
- Techniciens,
- Contractuels de la recherche (doctorants, post-doctorants, ATER)

Toutes les données de la recherche produites dans les laboratoires sont concernées par le présent document et ce quels que soient leurs supports et leurs formats :

- données d'observations ;
- données expérimentales ;
- données de simulation ;
- données de référence ;

- données dérivées ou compilées.

Les données de la recherche sont définies comme des « enregistrements factuels (chiffres, textes, images, sons, etc.) qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche ».¹

Le règlement de l'UPHF et de l'INSA HDF a vocation à aider et à offrir aux chercheurs un cadre définissant des principes généraux d'utilisation, de stockage, de diffusion et de réutilisation des données.

Ce règlement permet de définir les droits, responsabilités et rôles :

- **des scientifiques** dans le traitement et la gestion des données de la recherche ;
- **de l'université** dans l'adoption et la mise en place de la politique.

Ce règlement a pour objectif de sensibiliser et de guider les personnels producteurs et utilisateurs de données afin de les aider à adopter de bonnes pratiques tout au long du cycle de vie de la donnée, de la collecte à la réutilisation.

Le Plan de Gestion des Données (PGD) institutionnel est un document complémentaire du présent règlement permettant de rappeler les moyens et services existants au sein de l'UPHF et de l'INSA HDF pour répondre aux différents besoins que peuvent rencontrer les personnels scientifiques dans la gestion de leurs données de recherche dans le cadre d'un projet de recherche.

Cycle de vie de la donnée

Planification

Le décret n°2021-1572 du 3 décembre 2021² relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et fondations reconnues d'utilité publique ayant pour activité principale la recherche publique, **rend obligatoire la rédaction d'un Plan de Gestion des Données** pour chaque projet de recherche. Élaborer un Plan de Gestion des Données est désormais une obligation réglementaire (et pas seulement contractuelle vis-à-vis des financeurs publics). L'UPHF et l'INSA HDF ont pour responsabilité de veiller à la mise en œuvre de ce décret et proposent un accompagnement sur la rédaction et la relecture des Plans de Gestion des Données.

¹ Définition de l'Organisation de Coopération et de Développement Économique (OCDE) *Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics*, Paris, 2007, 28 p., <http://www.oecd.org/fr/science/sci-tech/38500823.pdf> [consulté le 28/08/2023]

² « Décret n° 2021-1572 du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche et les fondations reconnues d'utilité publique ayant pour activité principale la recherche publique - Légifrance », <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000044411360>.

L'Université Polytechnique Hauts-de-France et l'INSA Hauts-de-France demandent à ses personnels scientifiques de mettre en place un Plan de Gestion des Données au début de chaque projet de recherche, qu'il soit financé par un financeur public (collectivité, ANR, Europe, etc.) ou privé (opérateur économique privé, etc.).

La mise en place d'un Plan de Gestion des Données dès le début d'un projet permet de faciliter **l'accès, la préservation et la valorisation** des données de recherche. L'établissement dispose d'un espace et d'un Plan de Gestion des données institutionnels sur la plateforme DMP Opidor.

Collecte / Création

L'Université Polytechnique Hauts-de-France et l'INSA Hauts-de-France apportent aux chercheurs une aide pour les accompagner dans la **réutilisation de données déjà existantes** afin d'éviter la duplication des jeux de données. Les entrepôts de données de la recherche sont des outils qui permettent d'accéder facilement à une masse de données réutilisables.

Les personnels scientifiques qui réutilisent des données **s'engagent à le faire dans le respect du cadre juridique** en citant les auteurs des jeux de données dans les publications.

La réutilisation d'un jeu de données public implique de mentionner la source et la date de sa dernière mise à jour ainsi que de respecter l'obligation de ne pas les altérer, ni d'en dénaturer le sens.

Tout travail de collecte ou de création de nouveaux jeux de données dans le cadre d'un projet implique que les chercheurs s'interrogent sur la nature des données qui seront créées. Les données à caractère personnel et les recherches **impliquant la personne humaine** imposent d'informer et de **prendre contact** avec le **Délégué à la Protection des Données (DPO)** de l'université afin d'adapter les pratiques et le traitement pour veiller au respect du Règlement Général de la Protection des Données (RGPD).

Les données de la recherche, même celles qui ne sont pas publiques, doivent être FAIR « [Il convient de] soutenir la diffusion de données de la recherche qui soient traçables, accessibles, interopérables et réutilisables (principe FAIR) », *Directive du 20 juin 2019 sur les données ouvertes*³. La communauté scientifique s'engage à documenter ses données à l'aide de métadonnées et à rendre ces dernières consultables, analysables et traitables (quand le statut juridique le permet). Les formats ouverts et les standards existants doivent être privilégiés quand cela est possible afin de faciliter la réutilisation et l'interopérabilité des données.

³ Directive (UE) 2019/1024 du Parlement européen et du Conseil du 20 juin 2019 concernant les données ouvertes et la réutilisation des informations du secteur public.

<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000038901825>

Traitement, Analyse et Stockage

Pour traiter, analyser et stocker les données produites, collectées ou réutilisées, l'UPHF et l'INSA HDF encouragent l'utilisation des **logiciels fournis par l'établissement**.

Les personnels scientifiques ayant identifié la nature des données qu'ils traitent dès le début du projet seront en mesure de connaître le degré de sensibilité des données et les règles de manipulation applicables en la matière. Les droits d'accès aux données doivent être formalisés et intégrés dans le Plan de Gestion des Données.

Les données nécessitant un traitement particulier doivent être stockées et partagées dans le respect des exigences légales.

La communauté scientifique s'engage à effectuer des sauvegardes régulières de ses jeux de données. L'établissement demande de stocker les données sur des espaces de stockage maîtrisés et contrôlés.

L'Université Polytechnique Hauts-de-France et l'INSA Hauts-de-France préconisent de réaliser des plans de classement pour organiser les espaces de stockage.

Les acteurs et partenaires du projet doivent se concerter sur une convention de nommage des fichiers qui définira des règles précises qui permettront de garantir la pérennité des fichiers. Les outils privés de stockage en ligne sont des espaces non-sécurisés et non-conformes (*GoogleDrive, Dropbox, Onedrive...*) et il est interdit de les utiliser.⁴

Pour garantir que les données soient FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) les chercheurs s'engagent à décrire leurs données dès le début du projet en utilisant des standards de métadonnées adaptés à leur discipline scientifique. La communauté scientifique doit documenter et tracer toutes actions effectuées sur les données tout au long du projet.

Diffusion, Partage et valorisation.

La loi « **Pour une République Numérique** » du **7 octobre 2016**⁵ vient rendre obligatoire la diffusion des données de la recherche dès lors que la nature des données ne met pas en jeu le droit d'un tiers et qu'elles ont été publiées par l'établissement, l'organisme de recherche ou le chercheur.

Le droit ne protège aucune donnée, exceptées celles contenant des informations sur lesquelles un tiers détient des droits :

⁴Circulaire n° 6282-SG du 5 juillet 2021 relative à la doctrine d'utilisation de l'informatique en nuage par l'État
[HTTPS://WWW.LEGIFRANCE.GOUV.FR/CIRCULAIRE/ID/45205](https://www.legifrance.gouv.fr/circulaire/id/45205)

⁵ LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique, 2016.
<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/>

- données personnelles ;
- données secrets industriels et commerciaux ;
- données touchant à la défense nationale ou sécurité publique ;
- données figurant dans une base de données protégée par le droit *sui generis* ;
- données protégées par un droit de propriété intellectuelle : droit de brevet sur une invention, droit d'auteur, droits voisins d'interprètes ou de producteurs sonores ou audiovisuels.

Les personnels scientifiques s'engagent à **vérifier les critères mentionnés ci-dessus avant de partager, d'ouvrir et de déposer des données.**

La communauté scientifique veille à respecter les obligations d'ouverture des données de la recherche imposées par les financeurs comme l'ANR et l'Europe, selon le principe de la Commission Européenne « aussi ouvert que possible, aussi fermé que nécessaire ». Il existe des alternatives au partage public de données sensibles ou personnelles :

- le dépôt des données de recherche dans des entrepôts dont l'accès est contrôlé ;
- l'anonymisation des données avant leur partage public ;
- le seul partage des métadonnées relatives aux données de recherche ;
- la mention des procédures d'accès aux données de recherche dans l'article qui leur est associé et la gestion des demandes d'accès aux données émises par les autres chercheurs.

Cependant les métadonnées doivent être accessibles pour permettre aux chercheurs d'avoir connaissance des données produites et des conditions d'accès.

L'UPHF et l'INSA HDF alertent sur le risque de captation des droits par l'éditeur lors de la publication de données comme matériels supplémentaires.

L'Université recommande à **ce qu'au minimum les données liées aux publications** soient partagées avec la communauté scientifique dans un souci de transparence et de preuve en appui aux résultats présentés et dans le respect des principes de l'intégrité scientifique. Les **jeux de données publiés** doivent être de **qualité** et **respecter les principes FAIR**. Pour diffuser ces données l'UPHF et l'INSA HDF demandent aux chercheurs de privilégier **des canaux de partage adaptés** tels que les entrepôts de données, bases de données et data papers.

Il existe de nombreux entrepôts de données de la recherche permettant de publier des jeux de données de manière sécurisée et pérenne pour les différentes communautés scientifiques. Les entrepôts de données offrent également la possibilité de partager en accès restreint un jeu de données (nécessite un accord pour accéder aux données, possibilité de définir un embargo).

L'Université incite la communauté scientifique à déposer dans un entrepôt de données de confiance disciplinaire ou thématique en privilégiant ceux ayant la certification CoreTrustSeal.

Les jeux de données partagés doivent se voir attribuer un identifiant pérenne et comporter au minimum les informations suivantes : auteurs, date et titre. Concernant les auteurs il est conseillé d'identifier les rôles possibles et les contributions, car les auteurs d'un jeu de données peuvent être différents de ceux d'un article associé.

L'UPHF s'est dotée d'un espace institutionnel au sein de l'entrepôt national Recherche Data Gouv pour offrir à sa communauté scientifique un lieu où ils peuvent déposer et valoriser leurs jeux de données dans le cas où un entrepôt disciplinaire n'existe pas.

Les personnels scientifiques de l'UPHF et de l'INSA HDF qui ouvrent leurs jeux de données doivent recourir à une licence pour indiquer que les données sont gratuites et les informations publiques. Il est important d'employer des licences reconnues juridiquement, interprétable par des machines et qui répondent aux principes FAIR.

Selon le décret n° 2017-638 du 27 avril 2017 ⁶relatif aux licences de réutilisation à titre gratuit des informations publiques et aux modalités de leur homologation, deux choix de licences existent :

- la licence ouverte/Open licence élaborée par Etalab et l'Open Database (Odbl) ;
- les licences Creative Commons peuvent être employées pour les projets internationaux.

L'Université encourage les chercheurs à valoriser leurs jeux de données à travers la publication de data papers : cet article publié dans une revue par les pairs est citable au même titre qu'une autre publication.

Conservation et archivage

L'archivage pérenne a pour fonction d'**assurer la conservation** à long terme **des données**, leur accessibilité tout en préservant leur intelligibilité.

À l'issue de chaque projet de recherche, les chercheurs doivent **vérifier la réglementation et les contraintes légales ou contractuelles** de conservation qui sont associées aux données du projet.

Les données qui ne doivent pas être conservées comme **les données à caractère personnel doivent faire l'objet d'un traitement particulier de la part du DPO** afin de conserver celles qui sont « adéquates, pertinentes, et non excessives » au regard des finalités pour lesquelles elles ont été collectées ou traitées.

Les données doivent également être conservées dans un espace de conservation dédié. Des données bien documentées et décrites permettent aux autres membres de la communauté scientifique d'avoir accès à des informations qui permettent la réutilisation des données.

Toutes les données n'ont pas vocation à être conservées, cela implique que la communauté scientifique s'interroge sur :

- le caractère obligatoire de conservation des données (légales ou contractuelles) ;
- les données ayant valeur de preuve ;
- les données pouvant servir à la reproductibilité des travaux scientifiques ;
- le caractère unique des données non reproductibles ou difficilement reproductibles.

⁶ Décret n° 2017-638 du 27 avril 2017 relatif aux licences de réutilisation à titre gratuit des informations publiques et aux modalités de leur homologation <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000034502557>

Droits et responsabilités de l'Université

Former et Sensibiliser

L'Université Polytechnique Hauts-de-France et l'INSA Hauts-de-France accompagnent la communauté scientifique en proposant des formations sur les bonnes pratiques en matière de gestion des données de la recherche (rédaction d'un PGD, Initiation à la gestion des données de la recherche, Open Data et Science ouverte, etc.). L'Université s'engage à offrir des services de soutien liés à la gestion des données de recherche. Elle accompagne les porteurs de projet afin de les aider à répondre aux exigences des financeurs en matière de données de la recherche et d'Open Data.

Proposer des infrastructures

L'Université Polytechnique Hauts-de-France et l'INSA Hauts-de-France mettent à disposition de la communauté scientifique des infrastructures adéquates, robustes et sécurisées pour la gestion des données de la recherche. Elle veille à en assurer la mise à jour et le bon fonctionnement.

Reconnaître et valoriser les bonnes pratiques

Les bonnes pratiques en matière de gestion des données de la recherche pourront éventuellement être utilisées dans le cadre de l'évaluation des activités scientifiques des personnels.

Références Législatives

Les contrats de financement ERC (Horizon 2021-2027) exigent que les métadonnées relatives aux données de la recherche ainsi financées soient rendues FAIR, notamment en recourant à la mention CC0 (General model grant agreement, 1.0, 1st June 2021, Annotated Model grant agreement, pre-draft 1.0, 23 July 2021) : « Ensure open access to the deposited data as soon as possible and within the deadlines and conditions set out in the DMP, via the repository. Access has to be provided under a Creative Commons Attribution International Licence (CC BY) or a Creative Commons Public Domain Dedication (CC 0) or a licence/tool with equivalent rights, unless justified otherwise in your DMP. »

« Les [établissements publics contribuant au service public de la recherche] définissent une politique de conservation, de communication et de réutilisation des résultats bruts des travaux scientifiques menés en son sein. À cet effet, ils veillent à la mise en œuvre par leur personnel de plans de gestion de données et contribue[nt] aux infrastructures qui permettent la conservation, la communication et la réutilisation des données et des codes sources. », décret n°2021-1572 du 3 décembre 2021 relatif au respect des exigences de l'intégrité scientifique par les établissements publics.

« Les établissements publics d'enseignement et de recherche, qui pouvaient auparavant facturer une redevance, sont désormais soumis au régime de droit commun. Ils doivent permettre gratuitement la réutilisation de leurs informations publiques (sauf décret en Conseil d'Etat qui les autoriserait à percevoir des redevances). Les données de la recherche, lorsqu'elles sont publiques, sont gratuitement réutilisables. NB. Les administrations employant moins de 50 agents à temps plein et les collectivités territoriales de moins de 3 500 habitants ne sont pas tenues de publier en ligne les données publiques. » [art. L312-1-1 CRPA.](#)

"La recherche publique a pour objectifs :

- a) *Le développement et le progrès de la recherche dans tous les domaines de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie.*
- b) *La valorisation des résultats de la recherche au service de la société, qui s'appuie sur l'innovation et le transfert de technologie ;*
- c) *Le partage et la diffusion des connaissances scientifiques en donnant priorité aux formats libres d'accès ;*
 - c) bis) *Le développement d'une capacité d'expertise et d'appui aux associations et fondations, reconnues d'utilité publique et aux politiques publiques menées pour répondre aux défis sociétaux, aux besoins sociaux, économiques et du développement durable ;*
- d) *La formation à la recherche et par la recherche ;*
- e) *L'organisation de l'accès libre aux données scientifiques."*

Article L112-1 du Code de la recherche.

Glossaire

- **Data paper**

Les data papers sont des publications à part entière qui suivent le même processus éditorial que les articles scientifiques. À la différence d'un article scientifique classique qui exploite, analyse et interprète les données scientifiques, un data paper décrit finement un/des jeu(x) de données de façon à en faciliter la compréhension et à rendre les données accessibles, interprétables et réutilisables.⁷

- **Data journal**

Journal qui est dédié à la publication des articles de données (data paper). Il fournit habituellement des modèles de description des données et guide les chercheurs sur les lieux de dépôt et sur la façon de décrire et de présenter leurs données.⁸

- **Délégué à la Protection des données**

Le Délégué à la Protection des données, communément appelé DPO est selon la CNIL « chargé de mettre en œuvre la conformité au règlement européen sur la protection des données au sein de l'organisme qui l'a désigné s'agissant de l'ensemble des traitements mis en œuvre par cet organisme ».

- **DMP Opidor**

Plateforme développée par l'INIST-CNRS qui permet de créer facilement un Plan de Gestion des Données en suivant un modèle structuré. Cette plateforme est libre, gratuite et collaborative. DMP Opidor est recommandé par l'Agence Nationale de la Recherche. L'Université Polytechnique Hauts-de-France dispose d'un espace personnalisé.

- **Donnée à caractère personnel**

« Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres »

Loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés.

- **Entrepôt de données de la recherche**

Espace dédié permettant de déposer, rechercher et réutiliser des données grâce à une description par des métadonnées. Un identifiant pérenne est attribué à chaque jeu de données. Il existe des entrepôts généralistes (Zenodo, Recherche Data Gouv) et disciplinaires.

⁷ Ministère de l'Enseignement Supérieur, de la Recherche et de l'Innovation. Deuxième Plan national pour la science ouverte. Juillet 2021. <https://www.ouvrirlascience.fr/deuxieme-plan-national-pour-la-science-ouverte/>

⁸ L'Hostis, D., Hamelin, M., Lelievre, V., Aventurier, P. (2016). Publier un Data Paper pour valoriser ses données (Cours). <https://hal.inrae.fr/hal-02801638/>

- **Jeu de données**

Ensemble de données brutes ou dérivées sous une forme lisible présentant une certaine unité, rassemblées pour former un ensemble cohérent.

- **Licences Creative Commons**

Creative Commons correspond à un contrat-type pour lequel les chercheurs déterminent les droits attachés à l'article ou aux données qu'ils ont produits dans le cadre de leurs recherches. Six licences sont possibles combinant quatre éléments de base qui sont : la paternité, la modification, l'utilisation commerciale et la réutilisation dans les mêmes conditions du contrat d'origine.

- **Métadonnées (metadata)**

Ensemble d'informations structurées qui décrit, explicite, localise une ressource informationnelle, dans le but d'en faciliter la recherche, l'usage et la gestion.⁹

- **Plan de Gestion des Données PGD/ Data Management Plan (DMP)**

Le Plan de gestion des données, ou Data Management Plan est un document conçu dès le début du projet permettant de formaliser la manière dont les données de recherche seront gérées durant et après le projet.

- **Principes FAIR**

FAIR, facile à trouver, accessible, interopérable et réutilisable. Les principes FAIR permettent de garantir que les données respecteront des standards de partage pour permettre à l'humain et aux systèmes informatiques qu'ils pourront facilement les trouver, les interpréter et les utiliser.

- **Pseudonymisation**

Processus par lequel des données à caractère personnel ne peuvent plus être attribuées à une personne précise sans avoir recours à des informations supplémentaires.¹⁰

- **Règlement général sur la protection des données personnelles RGPD**

Le Règlement Général sur la protection des données est un texte réglementaire européen qui encadre le traitement des données à caractère personnel afin de renforcer les droits des personnes, responsabiliser les acteurs et crédibiliser la régulation. Il est entré en application le 25 mai 2018.

⁹ NISO. [Understanding Metadata](#). 2004

¹⁰Article 4, Règlement Général sur la Protection des Données.

- **Standard de métadonnées**

A pour objectif de fournir un ensemble d'éléments caractéristiques qui permettent de décrire les productions scientifiques. Le standard est choisi en fonction de la destination des données : dépôt, publication, archivage, etc. Il existe de nombreux standards de métadonnées tels que le Dublin Core, EXIF, etc.

- **Science Ouverte**

Diffusion sans entrave des publications et des données de la recherche.¹¹

¹¹Ministère de l'enseignement supérieur et de la recherche. [Plan national pour la Science ouverte](#). 4 juillet 2018