

Politique d'établissement pour la Gestion et l'Ouverture des Données de la Recherche

1^{er} janvier 2021

Table des matières

1. Préambule	2
2. Objectifs	3
3. Responsabilités juridiques de l'établissement	4
4. Définition des données de la recherche	4
5. Principes et actions	7
5.1 Principes de gestion des données de la recherche d'AgroParisTech	8
5.2 Ouverture des données de la recherche d'AgroParisTech	11
5.3 Les données de la recherche dans les activités de formation d'AgroParisTech	13
5.4 Valorisation et réutilisation des données	14
6. Moyens	15
6.1 Infrastructures	15
6.2 Accompagnement et formation	16
6.3 Financement et évaluation de la recherche	17
6.4 Indicateurs et traçabilité	18
7. Responsabilités : Synthèse	18
7.1 De l'établissement	18
7.2 Des agents et doctorants	18
7.3 Des responsables de projet de recherche	19
7.4 Des Directeurs d'unités	19
8. Suivi et mise en œuvre	19

1. Préambule

AgroParisTech, Institut des sciences et industries du vivant et de l'environnement, exerce une double mission de formation et de recherche, dont l'objectif est de développer des compétences et des savoirs afin de répondre aux enjeux de la production agricole et forestière, des transformations alimentaires et non-alimentaires, de la gestion durable des ressources naturelles et de l'environnement, ou encore de la santé humaine.

Ces recherches, menées en collaboration avec des partenaires publics ou privés, génèrent la collecte, le traitement et la gestion d'une large variété de données, dont la nature et la forme reflètent la diversité des disciplines et des objets d'études représentés dans les travaux de l'établissement. Par ailleurs, la spécialisation d'AgroParisTech dans les sciences du vivant induit la production de données dans des contextes expérimentaux complexes impliquant humains, organismes, etc.

Ces familles de données font aujourd'hui l'objet d'une attention soutenue par les acteurs qui les génèrent et les exploitent, sous la quadruple impulsion de :

- (i) **l'essor des technologies numériques**, générant une accélération sans précédent du volume et de la variété des données produites ;
- (ii) **l'essor du mouvement de la science ouverte**, aspirant à développer la transparence et la diffusion sans barrière des résultats et méthodes de la recherche financée par des fonds publics ;
- (iii) **l'essor du mouvement d'ouverture des données publiques (*Open Data*)**, adossée à la définition d'un cadre légal de plus en plus précis, visant à élargir l'accessibilité et la réutilisabilité des données produites par les administrations ;
- (iv) **l'essor d'une volonté partagée de mieux définir et maîtriser** les conditions de circulation et d'utilisation des données de la recherche dans une approche plus globale de gouvernance des données produites et traitées par les établissements.

Plusieurs prises de conscience au sein de la communauté scientifique s'inscrivent dans ce contexte :

- **Les fragilités nées de l'évolution rapide des technologies associée au manque récurrent (sauf exceptions dans certaines disciplines) de pratiques formalisées** et partagées dans la gestion des données. Les exemples de pertes ou d'inexploitabilité de données¹ dues à l'obsolescence des supports, à l'absence de procédures de sauvegardes, de métadonnées suffisamment explicites ou tout simplement à un manque d'informations limitant la capacité d'utilisation des données à la personne les ayant produites, sont largement répandus.

¹ Inist, "Risque de perte des données de la recherche", *Introduction à la gestion des données de la recherche*, 2014 https://www.inist.fr/wp-content/uploads/donnees/co/module_Donnees_recherche_21.html ; Timothy H. Vines, et al., "The Availability of Research Data Declines Rapidly with Article Age", *Current Biology*, vol. 24, issue 1, pp. 94-97, <https://doi.org/10.1016/j.cub.2013.11.014>

- **La valeur de nombreuses données générées dans le cadre de projets de recherche**, non seulement lorsqu'elles sont coûteuses à produire (temps, équipement, etc.) et/ou lorsqu'elles ne sont pas reproductibles (par exemple : données issues d'un synchrotron, données issues de suivis naturalistes, etc.), mais également lorsqu'elles peuvent contribuer à la production de nouvelles connaissances au-delà de leur contexte premier de production. Cette logique s'inscrit aussi bien dans un souci de rationalisation des dispositifs expérimentaux financés par des fonds publics lorsque cela se justifie, que dans une démarche de renforcement de nouvelles collaborations. De plus, l'affichage et l'ouverture des productions et résultats de recherche contribuent à l'attractivité des établissements publics.
- **Les évolutions rapides du rapport à la validation des résultats scientifiques.** Les tensions qui se cristallisent actuellement d'une part au sein des communautés scientifiques (autour des questions liées à l'intégrité scientifique, ou encore à l'évaluation par les pairs notamment), et d'autre part dans les rapports de confiance entre la société et la recherche publique (défaillances et dérives dans le relai de résultats scientifiques, aspiration à davantage de transparence sur les outils, méthodes et résultats de la recherche, etc.) entraînent un besoin accru pour une meilleure disponibilité des données sous-tendant la constitution des savoirs scientifiques, au sein et au-delà des cercles académiques.

Dans ce contexte², les organismes et les communautés exerçant une mission de recherche publique formalisent procédures et recommandations visant à assurer une gestion de ces données organisée et durable pour en assurer la pérennité et l'exploitabilité, tout en veillant au respect du cadre juridique et réglementaire national et international. Celui-ci intègre de plus en plus la notion des données issues de la recherche publique dans son périmètre, afin d'en définir les conditions de communicabilité, de réutilisabilité, mais aussi de sécurité et de confidentialité dès lors que la nature de ces données le requiert.

En tant qu'organisme exerçant une activité de recherche dans le cadre d'une mission de service public, AgroParisTech est pleinement concerné par ces évolutions et ce contexte. L'établissement formalise ainsi dans ce document sa politique en matière de gestion et d'ouverture des données de la recherche produite par ses agents, ses doctorants et ses étudiants.

2. Objectifs

La présente politique vise à inscrire l'établissement dans la dynamique présentée ci-dessus et à structurer ses orientations en s'appuyant sur les démarches déjà initiées, en articulation avec les politiques de ses partenaires institutionnels et en accord avec les obligations juridiques s'exerçant sur l'établissement. Les principaux objectifs sont les suivants:

² Plus d'informations sur l'intranet de l'établissement : <https://intra.agroparistech.fr/>

- assurer une gestion maîtrisée des données de recherche collectées ou produites par les agents de l'établissement pour la bonne conduite des travaux de recherche et la contribution aux autres missions de l'établissement, notamment l'enseignement ;
- promouvoir l'ouverture des données de la recherche, en articulation avec d'une part les autres principes, et d'autre part les obligations légales des cadres scientifiques et de l'établissement ;
- sécuriser le patrimoine scientifique de l'établissement.

3. Responsabilités juridiques de l'établissement

En tant qu'établissement public exerçant une activité de recherche, AgroParisTech est responsable du respect du cadre juridique s'appliquant aux données de la recherche entendues comme des documents administratifs produits par ses agents³.

Par ailleurs, le Règlement Général sur la Protection des Données (RGPD)⁴ formalise la responsabilité de l'établissement sur l'ensemble des données personnelles collectées par ses agents, y compris dans le cadre d'une activité de recherche.

Dans le cas de partenariats avec des acteurs privés, les questions liées à la propriété, la confidentialité et l'ouverture des données doivent apparaître dans les contrats en amont des projets, et tenir compte des conditions de production des données (part d'investissement public, contexte de mission de service public ou pas, etc.)

De manière plus globale, et en accord avec les textes réglementaires, AgroParisTech s'inscrit dans une démarche de transparence vis-à-vis de ses différents publics, et à ce titre, souhaite organiser la gestion des données de la recherche produites par ses agents afin de répondre à cet enjeu.

4. Définition des données de la recherche

Les données de la recherche n'ont pas de définition légale stricte mais des éléments juridiques, scientifiques et pragmatiques permettent de proposer une définition des données de la recherche pour AgroParisTech. Dans la suite de ce document, les données de la recherche sont donc entendues comme correspondant aux critères présentés ci-après. Les nuances nécessaires seront précisées dans chaque sous-partie de cette politique.

³ Si un droit d'auteur s'applique ([Code de la Propriété Intellectuelle](#)), les personnels statutairement non soumis à une autorité hiérarchique conservent un droit de décision sur la gestion et l'ouverture du contenu : Professeurs et Maîtres de Conférences notamment. L'attribution de droits d'auteur sur des contenus considérés comme données de la recherche n'est cependant pas la norme, cf. document d'analyse juridique sur l'intranet : <https://intra.agroparistech.fr/>

⁴ Application du RGPD en France : <https://www.cnil.fr/fr/reglement-europeen-protection-donnees> ; application du RGPD à AgroParisTech : <https://intra.agroparistech.fr/spip.php?rubrique1440>

4.1 Définitions juridiques

4.1.1 Périmètres

La [directive européenne Public Sector Information](#) (directive européenne 2019/1024, initialement adoptée en 2003, actualisée en 2019) définit les données de la recherche ainsi : « des documents se présentant sous forme numérique, autres que des publications scientifiques, qui sont recueillis ou produits au cours d'activités de recherche scientifique et utilisés comme éléments probants dans le processus de recherche, ou dont la communauté scientifique admet communément qu'ils sont nécessaires pour valider des conclusions et résultats de la recherche ». Il n'existe pas de définition équivalente actuellement dans le droit français.

Légalement, les établissements publics ont obligation de publier, sous réserve de l'absence de droits contraires, leurs "documents administratifs" donc tous les documents, quels que soient leur date, leur lieu de conservation, leur forme et leur support, qui sont produits ou reçus, dans le cadre de leur mission de service public, par l'État, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Il est admis que ces documents administratifs incluent les données de la recherche produites dans le cadre d'une mission de service public ([guide d'analyse du cadre juridique](#), p. 8-10).

La diffusion au public ne peut se faire, sauf exception, que pour des documents dits « achevés » donc des données finalisées. Tous les documents préparatoires (incluant les cahiers de laboratoires et les données intermédiaires) sont par conséquent exclus de la définition légale. Il demeure néanmoins possible de partager ces documents intermédiaires s'il y a un intérêt scientifique à le faire et en l'absence de restrictions légales.

4.1.2 Données Sensibles

Certaines données de la recherche peuvent relever également des données sensibles. Ce sont les informations qui révèlent la prétendue origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale.

Ce sont également les données génétiques, les données biométriques aux fins d'identifier une personne physique de manière unique, les données concernant la santé, la vie sexuelle ou l'orientation sexuelle d'une personne physique.

Il est interdit de recueillir et d'utiliser ces données, sauf dans certains cas précis et notamment ([article 9 du RGPD](#)) :

- Si la personne concernée a donné son **consentement exprès** (écrit, clair et explicite) ;
- Si ces données sont nécessaires dans un **but médical** ou pour la recherche dans le domaine de la santé ;
- Si leur utilisation est justifiée par l'**intérêt public et autorisé par la CNIL** ;
- Si ces données ont été rendues publiques par la personne concernée ;

Dans le cadre des données de la recherche, les données sensibles sont aussi les données qui peuvent présenter un risque (pour l'objet biologique animal ou végétal ou les sujets relatifs aux données) et nécessiter des procédures de collecte, de gestion, de stockage et de diffusion spécifiques afin de tenir compte des possibles conséquences d'une diffusion ouverte (espèces ou aires protégées par exemple).

4.2 Définitions couramment admises

La définition la plus couramment utilisée à l'heure actuelle est [celle proposée par l'Organisation de coopération et de développement économiques \(OCDE\) \(2007\)](#) :

“Les données de la recherche sont définies comme des enregistrements factuels (chiffres, textes, images et sons), 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. Un ensemble de données de recherche constitue une représentation systématique et partielle du sujet faisant l'objet de la recherche.

Ce terme ne s'applique pas aux éléments suivants : carnets de laboratoire, analyses préliminaires et projets de documents scientifiques, programmes de travaux futurs, examens par les pairs, communications personnelles avec des collègues et objets matériels (par exemple, les échantillons de laboratoire, les souches bactériennes et les animaux de laboratoire tels que les souris). L'accès à tous ces produits ou résultats de la recherche est régi par d'autres considérations que celles abordées.”

Toutefois, il existe de nombreuses variantes, en fonction de l'angle privilégié par leurs auteurs : liste ouverte ou fermée selon une approche énumérative, typologique, ou relative au niveau de traitement des données⁵.

4.3 Définition des données de la recherche pour AgroParisTech

En tenant compte de ces précédents éléments, et en considérant la diversité des thématiques portées par l'établissement, AgroParisTech propose la définition suivante des données de la recherche dans le cadre de cette politique :

Les données de la recherche sont l'ensemble des enregistrements factuels (informations numériques, textuelles, visuelles ou sonores...) collectés, observés ou créés dans le cadre d'une activité de recherche. Ces données sont nécessaires à la construction de la recherche, à l'établissement et à la validation des résultats de recherche. Lorsque les données n'ont pas encore été traitées ou contextualisées, on parle de données brutes. Entre données brutes et résultats finaux, les données intermédiaires résultant des transformations successives des données brutes peuvent servir de base à d'autres travaux de recherche et peuvent donc, de ce fait, être également considérées comme données de la recherche.

Cette définition recouvre les types de données suivants :

⁵ Violaine Rebouillat, *Ouverture des données de la recherche : de la vision politique aux pratiques des chercheurs*. Thèse en Sciences de l'information et de la communication, Conservatoire national des arts et métiers - CNAM, 2019, Français, [\(NNT : 2019CNAM1254\)](#), [\(tel-02447653\)](#), pp. 25-46.

Données d'observation : issues d'observations ou de mesures, d'enquêtes, d'entretiens, d'expertises, collectées en temps réel ou générées *a posteriori* et qui sont généralement uniques et impossibles à reproduire

Données expérimentales : obtenues à partir d'équipements de laboratoire et/ou d'expérimentations de terrain, potentiellement reproductibles, parfois coûteuses

Données dérivées ou compilées : résultant de traitement ou de combinaison d'autres données

Données issues de modèles et de simulations : générées par des modèles informatiques ou de simulation, potentiellement reproductibles mais parfois longues à obtenir

Données de référence : collections de données, banques de données (génomiques, structures chimiques, herbiers, archives historiques...)

Données géographiques : faisant directement ou indirectement référence à un lieu ou une zone géographique spécifique

Données constitutives de programmes informatiques : codes sources

Remarque : les spécimens vivants, de quelque règne que ce soit, sont des objets de recherche dont la gestion, la manipulation et les modes d'échange entre spécialistes sont régis par des règles relevant du cadre légal, de l'éthique et des règles implicites ou explicites en vigueur dans chaque domaine scientifique. La présente politique concerne à part entière les données numériques ou non attenantes à ces spécimens.

5. Principes et actions

AgroParisTech définit ci-dessous les grandes orientations de sa politique et les moyens proposés pour sa mise en œuvre. Conscient que cette politique implique de nombreuses adaptations nécessaires pour ses agents et ses étudiants, **l'établissement précise que cette mise en œuvre s'organisera progressivement au cours des prochaines années, de manière réaliste, accompagnée et proportionnée au contexte interne de l'établissement et aux évolutions en cours⁶.**

Le périmètre des personnes concernées par l'ensemble des mesures de cette politique est le suivant :

Agents d'AgroParisTech : toute personne employée par AgroParisTech exerçant une activité de recherche scientifique : professeur.e, maître.sse de conférences, ingénieur.e, etc.

Les personnels des unités sous tutelle d'AgroParisTech mais non employés par l'établissement seront invités à prendre connaissance de la présente politique. Les dispositifs de formation et d'accompagnement leur seront ouverts.

Etudiants et doctorants d'AgroParisTech : toute personne inscrite dans un dispositif de formation et exerçant une activité de recherche ou d'ingénierie dans une unité sous tutelle d'AgroParisTech, ou bien sous la responsabilité d'un.e encadrant.e d'AgroParisTech. Seront considérés tous contextes de production de données (formation, expertise, etc.), dès lors que ces données sont ensuite destinées à une utilisation dans le cadre d'une activité de recherche.

⁶ Plus d'informations sur l'intranet : <https://intra.agroparistech.fr/>

Le périmètre d'application de la politique est le suivant :

Les éléments relatifs à la gestion des données de la recherche s'appliquent à l'ensemble des données produites dans le cadre d'une activité de recherche, et quel que soit leur niveau de traitement, par les agents, doctorants et étudiants d'AgroParisTech.

Les éléments relatifs à l'ouverture des données de la recherche s'appliquent à un périmètre plus restreint, en fonction du cadre juridique, réglementaire, éthique, etc. s'appliquant aux données. L'ouverture des données intervient après avoir identifié les obligations et les options possibles s'appliquant dans chaque cas spécifique.

5.1 Principes de gestion des données de la recherche d'AgroParisTech

5.1.1 Formalisation de la gestion des données

Enjeux

Une gestion des données formalisée permet d'anticiper les différentes dimensions de leur collecte, traitement, et utilisation dans le cadre d'un projet de recherche, ou dans le cadre de la gestion courante des activités d'une équipe ou d'une structure de recherche. Cette démarche implique plusieurs notions : l'organisation, la documentation, l'évolution, le stockage et la sécurisation des données. Cela concerne toutes les données quels que soient leur type, leur stade de maturité et de transformation.

Une gestion des données formalisée participe à l'intégrité scientifique en facilitant la preuve de la solidité des résultats (bonne pratique dans la collecte des données, fiabilité), leur originalité et leur reproductibilité.

Les acteurs de la recherche publique (financeurs, organismes de recherche, etc.) s'appuient depuis plusieurs années sur un modèle de document permettant de formaliser ces aspects : le Plan de Gestion de Données (PGD) ou Data Management Plan (DMP). Le PGD est désormais un livrable requis par des financeurs publics tels que la Commission européenne ou encore l'Agence nationale de la recherche. Le PGD est un document rédigé en amont d'un projet de recherche et périodiquement mis à jour, dont le but est de préciser la nature des données et les modalités de collecte, validation, stockage, partage, sécurisation, ouverture le cas échéant, traitement des démarches liées aux questions éthiques, et archivage. Ce document peut aussi être utilisé pour organiser la gestion de données d'une structure ou d'une équipe en dehors du contexte spécifique d'un projet.

Recommandations et actions

- AgroParisTech demande aux agents et doctorants de l'établissement porteurs d'un projet de recherche de définir en amont les modalités pratiques de collecte et de gestion des données qui seront traitées, pendant et après la fin du projet. La rédaction d'un PGD est recommandée si elle n'est pas imposée par le financeur du projet. La trame de PGD pourra

être choisie en fonction du contexte du projet : modèle institutionnel⁷ ou modèle propre au financeur.

- AgroParisTech invite les unités de recherche sous sa tutelle à formaliser les règles de gestion des données au sein de leur unité, notamment sur les aspects suivants : modalités courantes de stockage, documentation des données (métadonnées), modalités d'accès, modalités d'archivage, traitement des questions éthiques si l'objet des recherches le requiert. L'établissement s'engage à soutenir et accompagner selon des modalités adaptées à ses moyens cette démarche via le dispositif d'accompagnement décrit au paragraphe 6.2.

5.1.2 Mise en application des principes FAIR

Enjeux

La gestion des données, pour être efficace, améliorer et enrichir l'activité de recherche, doit suivre plusieurs principes généraux soutenus et partagés dans la communauté scientifique internationale. L'application de ces principes doit être initiée dès la création des données puis tout au long de la vie des données. Ces principes sont rassemblés sous l'acronyme FAIR⁸ :

- **Facilement trouvables** : les données doivent être référencées dans une base requêteable, structurée, bien décrites par des métadonnées riches et standardisées, et disposer d'un identifiant pérenne (DOI, etc.) ;
- **Accessibles** : le lieu de stockage et les conditions d'accès doivent être explicites, même si l'accès est soumis à conditions, les données doivent être téléchargeables ;
- **Interopérables** : les données doivent être enregistrées dans des formats standards pérennes appropriés selon les thématiques et les disciplines ;
- **Réutilisables** : les conditions de réutilisation doivent être clairement définies et les métadonnées les plus complètes et détaillées possibles.

Recommandations et actions

AgroParisTech recommande à ses agents, ses doctorants et ses étudiants :

- de prendre en compte les principes FAIR dès l'initiation des projets de recherche et tout au long de leur déroulement, ainsi qu'après leur clôture ;
- de les appliquer, dans la mesure du possible, au patrimoine de données précédemment acquis

5.1.3 Formalisation de procédures de stockage et d'archivage

Enjeux

Les infrastructures de stockage s'inscrivent dans une gestion de court terme des données. Au-delà du temps du projet, elles n'en garantissent la sauvegarde et l'accessibilité au mieux que sur une durée limitée. Les infrastructures d'archivage apportent des services

⁷ Modèle institutionnel de Plan de Gestion de Données (lien à insérer).

⁸ Les principes FAIR (*Findable, Accessible, Interoperable, Reusable*) ont été défini par le collectif Force 11 en 2016 : <https://www.force11.org/group/fairgroup/fairprinciples>.

complémentaires, permettant d'assurer la pérennité de la sauvegarde et de l'accès aux données sur des périodes plus longues, en s'adaptant à l'évolution du contexte technologique.

Dans le cadre de la prise en compte des enjeux environnementaux induits par le stockage numérique de données informatiques, une anticipation raisonnée du choix des données à conserver et la définition de durées de conservation adaptées à leur utilité est nécessaire afin de promouvoir un usage raisonné des capacités de stockage et d'archivage de longue durée.

Recommandations et actions

- AgroParisTech demande à ses agents, ses doctorants et ses étudiants exerçant une activité de recherche de définir, en amont de tout nouveau projet de recherche, une stratégie de sauvegarde incluant le stockage (données en cours d'utilisation) et l'archivage (données de projets achevés) des données qu'ils produisent et manipulent, sur laquelle s'accordent tous les partenaires du projet. Cette stratégie s'établira en cohérence avec l'organisation des données de l'unité le cas échéant, ou reposera sur les moyens mis à disposition par l'établissement (cf. « Moyens »), les partenaires co-tutelles ou des acteurs extérieurs, en considérant les éléments définis dans cette politique. Le coût humain et financier de ces démarches doit être anticipé dans le fonctionnement des unités et le déroulement des projets de recherche.
- AgroParisTech demande à ses agents, ses doctorants et ses étudiants de considérer attentivement, et pour chaque version d'un même jeu de données, l'utilité et la pertinence de conserver les données ou non. Une attention particulière doit être accordée aux données non-reproductibles, coûteuses, et/ou longues à produire.
- AgroParisTech met en place les dispositifs permettant à ses agents, ses doctorants et ses étudiants de définir une stratégie de stockage et d'archivage des données incluant les éléments suivants :

(1) Concernant la sécurité de l'accès aux données :

- l'identification claire des personnes bénéficiant d'un accès selon leur rôle dans le projet ou l'unité, et de leurs droits et devoirs sur les données concernées ;
- la modulation de leurs droits selon leurs rôles (lecture, écriture, modification, création/suppression) ;
- l'information claire des personnes sur le circuit de décision concernant les modalités de partage et de diffusion des données ;
- la prise en compte précise de la confidentialité des données ou de la présence de données sensibles dans les jeux de données (cf. 5.4.)

(2) Concernant la pérennité et la sécurité du stockage et de l'archivage des données à stocker et conserver :

- la sécurisation des mécanismes de sauvegarde des données : existence d'un système de maintenance, la duplication (*a minima*) des copies des données (avec synchronisation entre les différentes copies pour assurer l'homogénéité et l'intégrité des données) ;

- la documentation et la journalisation de toute opération sur les données afin que les modifications soient traçables ;
- la sécurisation du soutien financier et institutionnel de l'infrastructure : gouvernance, plan de financement, etc.

5.2 Ouverture des données de la recherche d'AgroParisTech

Enjeux

L'ouverture des données de la recherche s'inscrit dans un double contexte : le cadre juridique de l'*open data* s'appliquant aux données publiques,⁹ et le cadre politique et réglementaire relatif à la science ouverte¹⁰. En particulier, la politique Science Ouverte d'AgroParisTech intègre les éléments présentés ci-dessous.

L'ouverture des données de la recherche produites dans le cadre d'une mission de service public relève désormais d'un principe d'ouverture par défaut. Ce principe admet néanmoins une large variété de restrictions lorsque la nature des données ou leur contexte de production le requièrent. Cette logique est résumée par la maxime "Aussi ouvert que possible, aussi fermé que nécessaire."

L'instruction de la question de l'ouverture des données prend en compte notamment les éléments suivants :

- L'application ou non d'un droit contraire s'opposant à l'ouverture de ces données (secret, confidentialité, etc.) ;
- Le degré d'achèvement des données ;
- Le contexte de production des données : source de financement, production dans le cadre d'une mission de service public, etc.
- L'utilisation des données dans le cadre de publications scientifiques ;
- etc.

Recommandations et actions

a) Généralités

- En accord avec le cadre juridique, réglementaire, contractuel, politique ou éthique, AgroParisTech demande à ses agents, ses doctorants et ses étudiants d'ouvrir les données de la recherche et codes sources de logiciels selon un calendrier adapté aux objectifs d'exploitation et de communication envisagés. Une attention particulière sera apportée aux données permettant de valider les résultats présentés dans des publications scientifiques. **(Cf. Politique Science Ouverte d'AgroParisTech, articles 1.12 et 1.13)**

⁹ [Article L112-1 du Code de la recherche](#), [Art. L.312-1-1 du Code des relations entre le public et l'Administration](#), [Directive européenne Public Sector Information](#)

¹⁰ [Plan national pour la science ouverte](#), [Politique d'AgroParisTech pour la science ouverte \(citée dans le présent document sous le nom « Politique science ouverte »\)](#), [conditions contractuelles de financeurs publics](#) tels que la Commission Européenne ou l'Agence Nationale de la Recherche.

- AgroParisTech demande à ses agents, ses doctorants et ses étudiants de considérer les réflexions sur l'ouverture ou la confidentialité de leurs données comme un élément à part entière de leurs démarches de recherche, qui doivent impliquer tous les agents et collaborateurs concernés et être en accord avec la politique de l'établissement ainsi qu'avec celle des partenaires de recherche au sein des unités et des projets de recherche. Une anticipation de la démarche d'ouverture dès le montage du projet et une discussion entre partenaires sont encouragées afin d'éviter l'apparition de difficultés logistiques ou de désaccords au terme des projets (cf. 4.1.1 ci-dessus).

b) En pratique

- AgroParisTech demande à ses agents, ses doctorants et ses étudiants, lorsque le choix de l'ouverture est fait, de déposer leurs données dans des entrepôts thématiques identifiés comme étant de référence au sein de leurs communautés respectives, ou institutionnels si aucun entrepôt thématique ne répond à leurs attentes.
- AgroParisTech demande à ses agents, ses doctorants et ses étudiants de sélectionner les entrepôts où ils déposent leurs données en privilégiant un entrepôt certifié par des instances internationales reconnues (CoreTrustSeal, Data Seal of Approval, etc.), ou, à défaut, plusieurs critères :
 - métadonnées riches et standardisées ;
 - attribution obligatoire d'identifiants pérennes aux jeux de données déposés ;
 - définition obligatoire d'une licence de diffusion garantissant le respect des conditions de réutilisation des données choisies par leurs producteurs, et dans le respect du cadre légal ;
 - contrôle de l'intégrité des données avant dépôt ;
 - pérennité de l'entrepôt ;
 - gratuité de l'entrepôt, au moins pour le dépôt ;
 - sécurité de l'hébergement des données ;
 - gestion des droits d'accès aux données ;
 - gestion de l'entrepôt par des acteurs non commerciaux ;
 - adéquation avec les recommandations des partenaires et des éditeurs des publications sous réserve que les critères précédents soient respectés.
- AgroParisTech encourage ses agents, au-delà ou en parallèle du dépôt de leurs données dans des entrepôts appropriés, à rédiger dès que possible et pertinent, des Articles de Données (Data Papers). AgroParisTech demande à ses agents, parmi les choix propres à cette démarche, de privilégier :
 - une revue ayant un champ scientifique approprié et une bonne visibilité ;
 - une revue ayant obligatoirement un comité de lecture ;
 - une revue dans laquelle les articles de données sont les plus détaillés possible pour rendre le jeu de données compréhensible, réutilisable et visible.

La politique science ouverte d'AgroParisTech s'applique pour toute publication de Data Paper (cf. articles 1.6 à 1.23)

5.3 Les données de la recherche dans les activités de formation d'AgroParisTech

Enjeux

AgroParisTech exerce une double mission de formation et de recherche. Ces deux missions s'enrichissent mutuellement dans le cadre de la formation par et pour la recherche : intégration de résultats de recherche dans les formations proposées, stages de recherche, projets de master, formation doctorale, dispositifs de découverte des métiers de la recherche (par exemple, *La recherche et moi*). Les étudiants et doctorants d'AgroParisTech sont ainsi amenés à apprendre et à mettre en pratique des méthodes de collecte et de traitement des données de la recherche.

Recommandations et actions

- AgroParisTech encourage ses agents impliqués dans le développement et l'animation d'unités d'enseignement ou de toute autre intervention pédagogique, à privilégier, dans la mesure du possible, et en accord avec leur liberté pédagogique, des données ouvertes lors de la présentation de productions de recherche (**cf. politique science ouverte, article 7.4**).
- AgroParisTech s'engage à produire des supports d'information et de formation sur la gestion et l'ouverture des données de la recherche, permettant aux agents d'accompagner leurs actions pédagogiques intégrant le recours aux données de la recherche.
 - AgroParisTech demande à ses agents encadrant des étudiants en stage sans partenariat privé de les inviter à observer les recommandations de la politique institutionnelle pour la gestion des données produites durant leur stage. En l'absence de restrictions légales, il est notamment recommandé lorsque le contexte d'exercice du stage le permet de traiter comme un élément des conventions de ces stages le fait de restituer à ses maîtres de stage, en fin de stage, les données produites accompagnées de tous les éléments (métadonnées, protocoles, contexte) nécessaires à leur conservation et à leur réutilisation.
- AgroParisTech s'engage à développer la formation des doctorants de l'établissement, ainsi que de ceux inscrits dans les autres établissements co-accrédités, sur la gestion et l'ouverture des données de la recherche afin que ces principes soient appliqués durant leurs travaux de recherche.
- AgroParisTech demande à ses doctorants de rédiger et mettre à jour tout au long de leurs travaux de thèse un plan de gestion de leurs données produites ou agrégées et de les restituer à l'établissement en fin de thèse, s'il n'y a pas de restrictions légales ou contractuelles, accompagnées de tous les éléments (métadonnées, protocoles, contexte) nécessaires à leur conservation et à leur réutilisation. Le PGD pourra faire l'objet d'une annexe dans le manuscrit de thèse.

5.4 Valorisation et réutilisation des données

Enjeux

La politique de valorisation de la recherche et de l'innovation d'AgroParisTech implique un besoin de confidentialité et d'exclusivité pendant une période donnée sur des résultats pouvant donner lieu à un dépôt de brevet, ou encore à la création d'une entreprise.

L'ouverture structurée des données de la recherche offre de nouvelles opportunités dans la construction de savoirs, procédés ou produits, notamment pour :

- tester la robustesse et/ou la reproductibilité de résultats présentés (lorsque cela est cohérent avec les disciplines ou les méthodes employées) ;
- éviter de reproduire des protocoles expérimentaux coûteux en équipements, temps, ressources humaines, ou bien répondant à des critères éthiques stricts (expérimentation animale par exemple), si des données ouvertes répondent déjà aux besoins requis ;
- permettre à un même jeu de données de contribuer à plusieurs approches scientifiques (réutilisation, autre projet, méta-analyses, etc.) ;
- contribuer à la construction de procédés ou produits innovants (nouvelles modalités de collaboration, valorisation de jeux de données ouverts)¹¹.

Recommandations et actions

- AgroParisTech encourage ses agents, ses doctorants et ses étudiants à rechercher, parmi les données publiquement ouvertes, si des données nécessaires au projet ou à une partie du projet de recherche ou de valorisation qu'ils initient sont déjà disponibles. **(Cf. Politique science ouverte, article 10.7)**
- AgroParisTech encourage ses agents, ses doctorants et ses étudiants à adopter la plus grande prudence dans la mise en œuvre des procédures de stockage et de partage des données lorsqu'un potentiel de valorisation économique nécessitant des mesures de confidentialité spécifiques est identifié. Il est recommandé dans ce cas de figure de privilégier dans la mesure du possible les infrastructures de l'établissement sous réserve de l'absence d'obligations contraaires.
- AgroParisTech encourage ses agents et ses doctorants à anticiper les délais de confidentialité des recherches menées en fonction des éventuels projets de valorisation économique, afin de maintenir une période d'exclusivité adaptée et de définir les délais d'application des actions définies au 2. **(Cf. Politique science ouverte, article 10.5)**
- AgroParisTech propose des modèles de contrat faisant état des éléments de politique institutionnelle sur la gestion et l'ouverture des données de l'établissement.

¹¹ La Commission Européenne porte notamment le concept d'innovation ouverte (*open innovation*), en interaction étroite avec celui de science ouverte (*open science*). L'*Open Innovation* défend le principe d'une plus grande liberté d'échanges de connaissances entre acteurs impliqués dans les dispositifs d'innovation comme vecteur de création de produits et de services à valeur ajoutée. Pour plus d'informations : <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>

6. Moyens

6.1 Infrastructures

6.1.1 Infrastructures informatiques

Enjeux et contexte

Afin de mettre en application les bonnes pratiques de gestion des données de la recherche, notamment en termes de stockage, accès, manipulation, sécurisation et archivage, des infrastructures appropriées, performantes et dimensionnées selon les disciplines sont nécessaires.

Recommandations et actions

- AgroParisTech, via sa Direction des Systèmes d'Information¹², s'engage à accompagner ses agents, doctorants et étudiants dans les réflexions et les choix technologiques qu'implique une gestion maîtrisée des données de la recherche.
- AgroParisTech, via sa Direction des Systèmes d'Information, s'engage à développer à destination des agents, doctorants et étudiants de l'établissement une offre de services numériques adaptés aux besoins identifiés en matière de stockage, traitement, et archivage de données. Ces services pourront être mis à disposition par l'établissement, un partenaire ou des acteurs extérieurs (DataCenters, entrepôts). Une attention particulière sera portée aux besoins liés à la gestion de données confidentielles ou sensibles placées sous la responsabilité de l'établissement.
- AgroParisTech demande à ses agents d'identifier dans le montage de leurs projets de recherche les infrastructures nécessaires et d'anticiper, avec des délais suffisants, les ressources associées (financières, humaines, etc.) pour une gestion maîtrisée et une utilisation des données de la recherche conformes aux principes définis dans la présente politique (cf. 4.1).

6.1.2 Mise en conformité RGPD

Enjeux

En tant qu'établissement amené à collecter et traiter des données personnelles, AgroParisTech est soumis au Règlement Général pour la Protection des Données (RGPD), et, à ce titre, est responsable de la mise en application de la réglementation sur ces données afin d'en garantir la sécurité, la confidentialité, et un traitement conforme aux droits des personnes tels que définis par le RGPD.

¹² La DSI est en cours de préfiguration et évoluera à terme vers une direction des services numériques.

Recommandations et actions

- AgroParisTech rappelle à tous ses agents, doctorants et étudiants traitant, dans le cadre de leurs activités de recherche, des données personnelles, leur obligation de se conformer au RGPD.
- Conformément à la réglementation, AgroParisTech a nommé un Délégué à la Protection des Données (ou Data Protection Officer, DPO) (dpo@agroparistech.fr). L'établissement recommande à ses agents et ses doctorants de s'y référer en cas de questionnement sur le RGPD.
- AgroParisTech fournit, et s'engage à maintenir, un outil permettant d'effectuer les déclarations en ligne de traitement de données personnelles : <https://rgpd.agroparistech.fr/>.
- AgroParisTech s'engage, via sa Direction des Systèmes d'Information, à accompagner ses agents vers la mise en place de solutions appropriées de stockage et de sécurisation des données personnelles ou le choix de solutions externes répondant aux exigences du RGPD.

6.2 Accompagnement et formation

Enjeux

La mise en œuvre des principes et orientations pour la gestion et l'ouverture des données de la recherche couvre une large variété de compétences et savoirs, touchant à des objets souvent complexes (cadre juridique, infrastructures, etc.), au sujet desquels la formation et l'accompagnement des chercheurs constituent un enjeu prioritaire.

Recommandations et actions

- AgroParisTech s'engage à développer un dispositif d'accompagnement de ses agents et doctorants afin de répondre à ses besoins d'information, de formation, et d'appui sur les questions relatives à la gestion et l'ouverture des données de la recherche.
- Le.a(s) chargé.e(s) de mission sur les données de la recherche assure(nt) l'animation du dispositif d'accompagnement, et son articulation avec le réseau des correspondants en charge des questions liées aux données de la recherche dans les unités de recherche sous tutelle d'AgroParisTech.
- Une cellule d'appui interne assure, sur sollicitation, et selon des modalités définies en annexe¹³, l'information, la formation et l'accompagnement de la communauté de recherche d'AgroParisTech sur les aspects relevant de leurs périmètres respectifs. Cette cellule est composée *a minima* des personnes suivantes pour l'information, la formation et l'accompagnement :
 - Un.e ou des chargé.e(s) de mission : gestion et ouverture des données de la recherche

¹³ Disponible sur l'intranet : <https://intra.agroparistech.fr/>

- Un ou des représentants de la Direction des Systèmes d'Information : questions liées aux services numériques requis pour le stockage, l'archivage, l'accès et le traitement des données de la recherche
- Un ou des représentants de la Direction des Documentations, du Musée du Vivant et du CIRE : questions liées au traitement documentaire des données de la recherche

Le chargé de mission juridique et DPO d'AgroParisTech sera étroitement associé à la cellule pour toute question relevant de son champ d'expertise.

La cellule travaille en étroite collaboration avec les autres réseaux et dispositifs d'appui de l'établissement (réseau Eurêka, AgroParisTech Innovation etc.).

- AgroParisTech encourage la création d'un réseau de correspondants au sein des unités de recherche sous sa tutelle, en collaboration avec ses partenaires, afin de développer un accompagnement de proximité de ses agents en coordination avec la cellule d'appui.

Dans une perspective de plus long terme, compte tenu des besoins forts en matière de formation et d'accompagnement induits par l'objet de la présente politique auprès de sa communauté, le renforcement des compétences internes à AgroParisTech fera l'objet d'actions spécifiques.

6.3 Financement et évaluation de la recherche

Enjeux

L'adoption des modalités de gestion et d'ouverture des données de la recherche décrites dans ce document implique une prise en compte appropriée dans les dispositifs d'accompagnement au pilotage de la recherche, afin d'en soutenir l'appropriation par l'ensemble de la communauté d'AgroParisTech.

Recommandations et actions

a) Financement de la recherche

- Pour tout projet de recherche recevant un soutien financier d'AgroParisTech, et en l'absence de toute barrière juridique, réglementaire, contractuelle, politique ou éthique, AgroParisTech demande aux bénéficiaires :
 - (i) d'appliquer les modalités de gestion des données telles que définies au 4.1. de la présente politique ;
 - (ii) d'ouvrir les jeux de données et codes sources de logiciels produits dans le respect des réglementations en vigueur, selon les modalités telles que définies au 4.2. de la présente politique (**cf. politique science ouverte article 1.16**)
- Tout dossier déposé auprès de la Direction de la recherche et de la valorisation pour l'appel d'offre annuel interne (hors équipement scientifique) dont le porteur de projet serait un ancien bénéficiaire de l'appel d'offre qui n'aurait pas été en mesure de justifier l'absence d'ouverture des données et codes sources de logiciels issus de ce financement selon les critères prévus par les textes juridiques et réglementaires, sera jugé non recevable (**cf. politique science ouverte article 1.19**). Une période de transition d'une

durée minimale de trois ans (renouvelable) sera mise en place à compter de la date d'entrée en vigueur de la présente politique.

- AgroParisTech s'engage à développer des modalités de soutien financier à destination de projets de valorisation du patrimoine de données de l'établissement, non rattachées à un projet en cours ou à venir, à des fins de centralisation, de curation et/ou de publication en ligne.

b) Evaluation de la recherche

- AgroParisTech encourage la prise en compte d'actions en faveur de l'organisation de la gestion et de l'ouverture des données de la recherche dans le cadre des procédures de recrutement et auprès des instances en charge de l'évaluation de ses enseignants-chercheurs et de ses équipes (DGER, CNECA, HCERES), ou des instances de financement.

6.4 Indicateurs et traçabilité

- AgroParisTech demande à ses agents, ses doctorants et ses étudiants d'associer autant que possible l'établissement (utilisation des affiliations si possible, utilisation du nom de l'établissement correctement orthographié *a minima*) et leur identifiant chercheur (ORCID) aux dépôts de données réalisés dans des entrepôts publics.
- AgroParisTech organise le référencement des jeux de données publiés en entrepôts dans son application de suivi de ses productions scientifiques.
- Les dépôts de données en entrepôts et les data papers seront progressivement inclus dans les indicateurs officiels de l'établissement (rapport d'activité, etc.).

7. Responsabilités : Synthèse

7.1 De l'établissement

- Assurer la qualité et la traçabilité des données produites et traitées dans le cadre des travaux de recherche et de formation menés en son sein et avec ses partenaires publics ou privés et de son patrimoine de données
- Se mettre en conformité avec la loi s'appliquant aux données de la recherche et accompagner ses agents, doctorants et étudiants dans cette mise en conformité
- Mettre à disposition de sa communauté les moyens nécessaires pour la mise en œuvre de la présente politique : moyens matériels (infrastructures, etc.) ; moyens informationnels pour la bonne connaissance et la maîtrise des principes de bonne gestion des données de l'établissement ; moyens humains (accompagnement, etc.)
- Coordonner sa politique avec celle de ses partenaires (conventionnement sur les données de la recherche, etc.)
- Veiller à l'application des conditions définies dans le RGPD par ses agents, en articulation avec ses partenaires dans le cadre des unités mixtes de recherche (chaque tutelle étant co-responsable de traitement).

7.2 Des agents et doctorants

- S'informer et se former sur la gestion et l'ouverture des données de recherche ;

- Mettre en œuvre les recommandations de gestion des données de la recherche de l'établissement ;
- Utiliser les infrastructures recommandées ou mises à disposition.
- Consolider le patrimoine des données collectées au fil de sa carrière dans l'établissement.

7.3 Des responsables de projet de recherche

- Assurer la mise en place d'un plan de gestion de données en préparation du projet, veiller à son actualisation et sa mise en œuvre ;
- Instruire la question de l'ouverture ou de la confidentialité des données, prendre les mesures adéquates et faire remonter les problématiques rencontrées aux partenaires du projet ;
- Prendre les mesures nécessaires au respect des obligations légales, éthiques ou contractuelles s'appliquant aux données du projet.
- S'assurer que les modalités de gestion et d'ouverture des données au terme du projet figurent dans les contrats de collaboration et sont discutées dès que possible dans le projet.

Il est possible de déléguer tout ou partie de ces actions à une autre personne. Il est également possible de désigner un ou plusieurs responsables du traitement des données dans le PGD.

7.4 Des Directeurs d'unités

- Animer la réflexion de l'unité sur la gestion et l'ouverture de ses données en cohérence et en conciliation avec les politiques des différentes tutelles
- Faire figurer les modalités de gestion et d'ouverture des données de l'unité dans la convention d'unité lors de sa rédaction
- Désigner, en accord avec l'agent concerné, un relais local du réseau constitué pour informer et accompagner les agents dans la mise en place de la gestion et de l'ouverture des données de la recherche.
- Assurer la mise en conformité de leur unité avec le RGPD en lien avec les établissements tutelles.

8. Suivi et mise en œuvre

- Un.e ou plusieurs chargé.e(s) de mission, sous la responsabilité du Directeur de la Recherche et de la Valorisation, assure(nt) le suivi et la mise en œuvre de la présente politique en étroite collaboration avec les autres directions impliquées (Direction des Systèmes d'Information, Direction des Documentations, du Musée du Vivant et du CIRE, etc.)
- Les correspondants identifiés pour les missions d'accompagnement (cellule d'appui, etc.) assurent la mise en œuvre des actions définies en lien avec ces missions dans la présente politique.

- Les besoins, demandes, et retours liés à la présente politique seront instruits dans les instances de l'établissement (réunions avec les départements, Comité de Direction, Conseil Scientifique, etc. et CE et CEVE ponctuellement si besoin)