

**Note à l'attention de porteurs de PEPR, projets ciblés, WP
Données de la recherche**

**CNRS Sciences humaines & sociales
19 décembre 2023**

Cette note répond à un besoin qui a émergé des discussions entre les porteurs de PEPR, projets ciblés, WP relevant des sciences humaines et sociales et CNRS Sciences humaines & sociales, concernant la gestion et la diffusion des données de la recherche.

Synthèse des principaux points d'attention

- 1) Mettre en œuvre les principes FAIR pour la gestion des données issues des PEPR en rendant éligibles dans les appels à projets les coûts liés à la FAIRisation des données et en faisant des principes FAIR un élément d'évaluation des candidatures aux appels à projets ;
- 2) Pour le dépôts des données, donner la priorité à des entrepôts sûrs et reconnus dans leur champ (centres de références thématiques en priorité, puis infrastructures nationales de recherche, puis entrepôts institutionnels ou disciplinaires, puis recherche.data.gouv). En SHS, ces entrepôts sont Huma-Num et Progedo ;
- 3) Pour le soutien opérationnel à la gestion des données, s'appuyer en priorité sur des services de proximité qui s'organisent en ce moment dans le cadre des ateliers de la donnée labélisée par le MESR ; et/ou organiser au sein du PPR/PEPR un dispositif d'accompagnement des porteurs de projets financés par le programme pour assurer l'interface avec les entrepôts de données ; dans tous les cas, rendre systématique le plan de gestion de données pour les projets financés par le programme ;
- 4) Appliquer aux données issues des PEPR le principe d'ouverture par défaut (« Les données doivent être aussi ouvertes que possible et aussi fermés que nécessaire ») et créer au sein du PEPR un espace de discussion pour favoriser l'ouverture et le partage des données ;
- 5) Demander aux porteurs de projets financés dans le cadre des PEPR de s'engager sur des délais raisonnables de mise à disposition des données et veiller à ce que les données liées à des publications issues des travaux des PEPR soient rendues disponibles au moment de la publication.

Principes de gestion et de diffusion des données de recherche dans le cadre des PEPR

Cette note vise à expliciter les principes de gestion et de diffusion des données de recherche à appliquer dans le cadre de la conduite des PEPR (Programmes et Equipements Prioritaires de Recherche), d'accélération ou exploratoires.

Les PEPR couvrent un large spectre de disciplines, d'objets et de thématiques, auxquels sont associés des données pouvant être de natures très différentes. Néanmoins, ils constituent également un cadre privilégié pour favoriser l'adoption par les communautés de recherche de bonnes pratiques en matière de production, de gestion, de stockage et de diffusion des données.

L'[Union Européenne](#) et l'[ANR](#) ont déjà défini de grands principes pour la bonne gestion des données de recherche qu'elles répercutent dans le cadre de leurs appels à projets. Cette démarche doit être mise en œuvre pour la conduite des PEPR afin d'harmoniser les pratiques et tendre vers un haut niveau de qualité pour l'obtention de données plus robustes et mieux gérées tout au long de leur cycle de vie.

Le [Plan National pour la Science Ouverte du MESR](#) formule également des objectifs ambitieux en matière d'ouverture et de partage des données de recherche, qui peuvent trouver dans les PEPR un cadre intéressant de réalisation, en raison de l'échelle de ces programmes et leur capacité à fédérer des communautés organisées autour de thématiques délimitées.

Un enjeu important consiste également à ce que les données produites dans le cadre des PEPR s'appuient pour leur production, leur stockage et leur diffusion sur des infrastructures de recherche reconnues et sécurisées, garantissant que les communautés de recherche en conservent la maîtrise à long terme.

Les logiques d'ouverture et de partage des données doivent être conciliés avec celles de valorisation économique des résultats de la recherche, sachant que ces deux objectifs ne sont pas incompatibles. Le cadre légal applicable aux données de recherche fournit des ressources pour assurer la bonne articulation entre ces logiques (voir plus bas).

Mise en œuvre des principes FAIR

Définis d'abord au niveau européen, les [principes FAIR](#) se sont peu à peu imposés comme des standards partagés pour la bonne gestion et diffusion des données de recherche. Ils structurent la politique de données des agences de financement (ANR, Union européenne) ; ils sont au cœur du Plan National pour la Science Ouverte et la grande majorité des établissements et organismes de recherche s'appuient sur eux dans le cadre de leur politique interne de données (voir, par exemple, [Feuille de route Science Ouverte](#) et [Plan Données du CNRS](#)).

Pour rappel, les principes FAIR impliquent que les données soient :

- Faciles à trouver ;
- Accessibles ;
- Interopérables ;
- Réutilisables.

Plusieurs niveaux de « FAIRisation » sont possibles pour les données de recherche, le respect des principes pouvant être plus ou moins poussé selon la nature des données et les contraintes spécifiques qui s'y attachent.

Néanmoins, dans le cadre de la mise en œuvre des PEPR, un degré satisfaisant de compatibilité avec les principes FAIR devra être systématiquement recherché, afin de renforcer le caractère cumulatif des recherches et de constituer les données comme un patrimoine scientifique à transmettre à l'issue des programmes.

De plus, la mise en œuvre des principes FAIR a un coût, qui peut s'avérer élevé en fonction du niveau d'implémentation que l'on souhaite atteindre. Les dépenses liées à la « FAIRisation » des données doivent être rendues éligibles dans le cadre des appels à projets lancés par les PEPR, afin que les équipes de recherche financées puissent être en mesure d'y faire face. Cela inclut notamment le temps d'ingénierie nécessaire à la

production, au traitement et à la mise aux normes des données, que ce soit sous forme de prestations ou de temps de travail rémunéré sur contrat.

Chaque discipline a sa propre approche des principes FAIR et chaque objet de recherche appelle des discussions sur la meilleure manière d'implémenter ces directives. Les PEPR sont dans une position privilégiée pour favoriser cette discussion scientifique autour de la mise en œuvre des principes FAIR et les porteurs de programme devraient prévoir un dispositif pour traiter de ces enjeux ou recommander qu'ils soient abordés dans les projets.

Les Plans de Gestion des Données (PGD ou DMP pour *Data Management Plan*) constituent l'outil privilégié pour anticiper et planifier les opérations liées à la mise en œuvre des principes FAIR, tout en répartissant les rôles et les responsabilités.

L'ANR demande à présent systématiquement la [production d'un PGD](#) aux projets qui bénéficient de ses crédits. Il est donc logique que les projets financés dans le cadre des appels financés par les PEPR soient soumis à la même exigence. Il est également important que les PEPR intègrent dans les critères de leurs appels à projets des exigences relatives à la bonne gestion des données, basées sur les principes FAIR et en fassent un des éléments d'appréciation des candidatures.

Priorisation dans le choix des entrepôts

Il importe que pour le stockage et la conservation des données de recherche, les PEPR s'appuient au maximum sur des infrastructures reconnues et sécurisées, garantissant la maîtrise à long terme par les communautés de recherche. Le recours à ces infrastructures constitue également un enjeu pour la bonne dépense des crédits publics, en évitant la dispersion des moyens et la prolifération de solutions sans garantie de pérennité.

La question du choix des entrepôts de données se pose très différemment selon les disciplines, qui n'ont pas toutes le même niveau d'avancement et d'organisation concernant les infrastructures de données. Le MESR a par ailleurs engagé un vaste chantier au niveau national concernant les données de recherche, avec la mise en place de recherche.data.gouv.fr et des [ateliers de la donnée](#), dont le processus de labellisation au niveau national est bien avancé.

Une logique de subsidiarité doit être respectée pour le choix des entrepôts de données, sachant que recherche.data.gouv.fr a vocation à rester une solution de dernier recours, pour les communautés qui ne bénéficient pas déjà de solution structurée pour le dépôt des données.

Il conviendra donc pour les PEPR de privilégier d'abord les entrepôts de données considérés comme des centres de référence thématiques, puis les infrastructures nationales de recherche inscrites sur la feuille de route du MESR, puis les entrepôts disciplinaires ou d'établissement, et enfin seulement de se tourner vers recherche.data.gouv.fr, si aucune autre solution n'est à disposition.

Ainsi, dans un premier temps, pour les données de la recherche produites par les sciences humaines et sociales, Huma-Num et Progedo sont les interlocuteurs privilégiés.

Dans tous les cas, les projets financés dans le cadre des PEPR ne doivent pas recréer des solutions de stockage ou de diffusion de données avec les financements qu'ils reçoivent, afin d'éviter de contribuer à la dispersion des moyens. Les solutions publiques, offrant des garanties solides de maîtrise des données à long terme par les communautés scientifiques doivent être

systématiquement privilégiées par rapport à des solutions privées, notamment lorsque celles-ci induisent des restrictions, directes ou indirectes, à la réutilisation des données. Il importe en particulier de ne pas recourir aux solutions proposées par les éditeurs privés pour le dépôt des données associées aux publications.

Soutien opérationnel pour la gestion des données

Pour le soutien aux équipes de recherche concernant la gestion et la diffusion des données, le paysage national est en cours de structuration autour des « [ateliers de la donnée](#) » mis en place par le MESR à travers un processus de labellisation.

Ces ateliers s'organisent sur une base territoriale, et non disciplinaire, chaque site universitaire étant invité à présenter une offre de services autour des données de recherche. Quatorze ateliers sont ainsi déjà labélisés par le MESR et cinq sont sur une trajectoire de labellisation.

Les porteurs de projets financés par des PEPR devront en priorité rechercher à proximité le soutien des services labélisés dans le cadre des ateliers de données. Ces services peuvent jouer un rôle dans des opérations comme la réalisation d'un plan de gestion de données, le choix d'un entrepôt, l'application des principes FAIR, l'instruction de questions juridiques, l'accès à une infrastructure nationale, etc.

Si un centre de référence thématique ou une infrastructure nationale de recherche occupe un positionnement privilégié pour recevoir les données issues d'un PEPR, une formalisation particulière pourrait être opérée de manière à ce que l'infrastructure apporte un appui spécifique au projet.

Ouverture et partage des données

Les principes FAIR prévoient que les données de recherche doivent être accessibles et réutilisables. La question du degré d'ouverture à atteindre dans la diffusion de données de recherche est cadrée par l'adage « Les données doivent être aussi ouvertes que possibles et aussi fermées que nécessaires ».

Cette phrase signifie que les décisions liées au partage des données ne doivent pas découler de choix opérés simplement en opportunité, mais d'un diagnostic objectif quant à la possibilité d'ouvrir les données ou à la nécessité de les garder confidentielles. Cela veut dire que si l'ouverture des données est possible, elle doit alors être opérée et que ce n'est que dans l'hypothèse où des contraintes spécifiques se manifestent qu'on doit restreindre leur réutilisation. Ces principes se résument par l'idée que les données de recherche obéissent à un principe d'ouverture par défaut, assorti d'exceptions délimitées.

En France, ce principe d'ouverture par défaut découle, non seulement des principes FAIR, répercutés par des financeurs de la recherche comme l'ANR, mais aussi de la loi elle-même, qui établit un principe d'*Open Data* par défaut pour les données, lorsque celles-ci correspondent à des données publiques ([Loi République Numérique 2016](#) – la même règle valant pour les codes sources et logiciels de recherche). Cela signifie que les données de recherche doivent en principe être mises en ligne spontanément et rendues librement réutilisables, avec pour seule restriction le respect de leur intégrité et la mention de leur source.

Dans certains cas exceptionnels listés par la loi (droits de propriété intellectuelle appartenant à des tiers, secrets administratifs, données personnelles, protection du patrimoine scientifique et technique de la Nation, etc.), le principe d'ouverture par défaut s'écarte pour laisser place à une obligation de protection des données. Néanmoins, même pour des données couvertes par ces exceptions, il reste généralement possible d'organiser des formes de stockage et de partage sécurisés, permettant la circulation des données dans un cadre délimité et contrôlé.

Concernant l'articulation entre les logiques d'ouverture et de valorisation économique, il convient de noter que l'ouverture des données n'est pas antinomique avec le dépôt de brevets sur des inventions. Afin de préserver l'antériorité, il importe de phaser les processus en obtenant d'abord le brevet avant de divulguer les données associées. Certaines des exceptions citées ci-dessus (propriété intellectuelle de tiers, secret industriel et commercial, PPST) sont également invocables dans le cadre de partenariats public-privé qui justifieraient de réserver une exclusivité à un acteur économique. Les données de recherche ne sont en outre soumises au principe d'ouverture par défaut que si elles sont financées à au moins 50% par de l'argent public.

Les PEPR devront suivre cette logique du principe d'ouverture par défaut des données de recherche, en recherchant systématiquement le degré de partage des données le plus poussé, en fonction des contraintes liées à chaque type de données produites ou collectées. Ils pourraient s'appuyer sur les services des ateliers de la donnée à proximité pour effectuer les analyses juridiques pertinentes, déterminer les licences de réutilisation adaptées, effectuer les opérations nécessaires au partage des données (anonymisation, notamment, lorsqu'elle est possible) et évaluer le champ d'application des exceptions au principe d'ouverture. Lorsque des questions d'articulation entre ouverture et valorisation économique se posent, elles devraient être traitées en associant les compétences des ateliers de la donnée avec celles des services de partenariats et valorisation.

Lorsque des communautés de recherche sont encore peu familiarisées avec la problématique et/ou la pratique du partage des données, les PEPR devraient constituer en leur sein des espaces où ces questions peuvent être débattues de manière à définir des standards favorisant la diffusion des données, animés par la gouvernance des programmes. Le champ du PEPR et les équipes associées constituent déjà un premier cercle au sein duquel expérimenter concrètement le partage des données, avant de l'envisager à une échelle plus large.

Délais de mise à disposition des données

Même lorsque l'existence du principe d'ouverture par défaut des données de recherche est admise, la question du délai à partir duquel celui-ci doit être mis en œuvre peut rester difficile à trancher. La loi République Numérique indique que les données sont soumises au principe d'ouverture par défaut à partir du moment où elles sont « achevées ». Mais déterminer quand des données de recherche présente un caractère d'achèvement peut là aussi s'avérer complexe.

Le cadre des PEPR constitue néanmoins un champ favorable pour travailler cette question des délais raisonnables de mise à disposition des données de recherche dans un périmètre contrôlé. Il devrait ainsi être déterminé à l'avance que l'ensemble des données produites dans le cadre d'un PEPR ne doivent pas faire l'objet d'une réservation d'accès au-delà de la durée du PEPR lui-même, cette date constituant une limite maximale n'excluant par des diffusions plus rapides lorsque celles-ci sont possibles. De la même manière, les projets financés dans le cadre des

appels lancés par les PEPR doivent prendre des engagements de diffusion des données à l'issue de leur durée de financement, en anticipant cette question dans leur candidature.

Par ailleurs, si des résultats de recherche font l'objet de publications au cours de la durée de réalisation des PEPR, on doit considérer que les données sur lesquelles s'appuient ces écrits scientifiques présentent un degré d'achèvement suffisant pour déclencher l'application du principe d'ouverture par défaut. Cela signifie que les données associées à ces publications doivent être rendues immédiatement accessibles, notamment afin de favoriser la répliquabilité/reproductibilité des résultats. Ces données doivent être rendues librement réutilisables dès la publication des écrits, lorsque c'est juridiquement possible, ou faire l'objet de dépôt sur des entrepôts permettant le partage sécurisé, lorsque des restrictions à la réutilisation s'appliquent.

Si cette question du délai raisonnable de mise à disposition des données ne fait pas déjà l'objet d'un consensus disciplinaire, elle devrait faire l'objet d'une discussion organisée dans le cadre de la communauté rassemblée par le PEPR afin de déterminer un standard susceptible de garantir le partage réciproque des données parmi les équipes bénéficiant des financements.