

Université de Paris 8 - Vincennes-Saint-Denis: DMP - Université Paris 8 Vincennes-Saint-Denis (français)

INFORMATIONS ADMINISTRATIVES SUPPLEMENTAIRES

Historique des mises à jour du DMP

Exemple de réponse:

V2, 15/04/2021- Guattari, F., Deleuze G., Freedman J...

Rédacteurs et relecteurs du DMP

Exemple de réponse:

Deleuze, G. (université Paris 8), Nom, P. (laboratoire X), Nom, P. (SCD Paris 8) ...

Recommandations:

Indiquer l'ensemble des personnes ayant participé à la rédaction ou relecture du plan de gestion de données

1. PRESENTATION, OBJECTIFS ET ACTEURS DU PROJET

1.1. Indiquer l'identifiant interne du projet (facultatif)

1.2. Dans quel programme de financement s'inscrit le projet ?

Exemple de réponse:

Horizon Europe

1.3.1. Date de début de projet

1.3.2. Date de fin de projet

1.3.3. Durée du projet

1.4. Indiquer l'acronyme ou le titre du projet

Exemple de réponse:

DoRe_DATA_P8

Spatialité des violences conjugales & Covid-19

1.4. Indiquer les thèmes décrivant le projet

Recommandations:

Utiliser des mots clés.

Exemple de réponse:

Sciences de l'Homme et Société

Sciences de l'environnement

Sciences cognitives

1.5. Qui sont les partenaires du projet ?

Mentionner uniquement les partenaires ayant participé à la collecte, à l'analyse ou au traitement des données.

Exemple de réponse:

EA X - Université Paris 8 / Université Paris-Est Créteil / EPHE; UMR X - Université Paris 8 / CNRS; Assistance Publique des Hôpitaux de Paris

2. GESTION ET PRESENTATION DES DONNEES DU PROJET

2.1. Présentation des données collectées et produites

Quelles données (types, formats, mode) seront collectées et/ou produites au cours du projet ?

Recommandations:

Préciser entre autres le type, le périmètre, l'échelle, les méthodologies et les outils utilisés, la couverture temporelle, les formats, l'application éventuelle d'un processus qualité, etc.

Exemple de réponse:

Types de données : données d'observation, données expérimentales, données d'enquêtes, données textuelles, échantillons, base de données, images, données audiovisuelles, données issues de sites web, modèles...

Format de données : cela correspond à la manière selon laquelle les données sont codées pour le stockage, généralement reflétée par l'extension du nom de fichier (par exemple pdf, xls, odt, csv, doc, txt, rdf, mp4, png). Privilégiez les standards de formats ouverts pour les données (voir la liste indicative sur la plateforme de formation DoRANum : https://doranum.fr/wp-content/uploads/FS2_liste_indicative_formats_V1.pdf)

ou bien vérifiez si un fichier est valide via l'outil FACILE du Centre Informatique National de l'Enseignement Supérieur (CINES) : <https://facile.cines.fr/>

Mode de collecte : questionnaire, entretien semi-directifs, note d'observation, etc.

Outils de traitement des données : Excel, Stata, SPSS, Matlab v8 avec Signal Processing toolbox, Qgis, R, ...

2.2. Indiquer la volumétrie prévisionnelle et l'estimation de la taille du jeu de données au cours du projet

Recommandations:

Indiquer au début de projet, un volume prévisionnel de données produites.
En fin de projet, reporter le volume réel à la place du volume prévisionnel préalablement saisi.

Exemple de réponse:

Moins de 50 Mo, entre 50 et 100 Mo, entre 100 Mo et 1 Go, entre 1 et 10 Go, entre 10 et 50 Go, plus de 50 Go. D'autres unités de grandeur peuvent être utilisées comme le nombre de fichiers, la durée des enregistrements, etc.

Taille du jeu de données : 10 Mo, 100 Mo, 1 Go, ~ 200 Go, 1 To...

2.3. Articles déjà publiés en lien avec ces données

Recommandations:

Préciser la référence des articles, avec le lien, de préférence de type DOI.

2.4. Réutilisation des données existantes

Si les données utilisées dans le projet s'appuient sur la réutilisation de données existantes, citer leur origine.

Recommandations:

Préciser si les données existantes et le matériel utilisés sont protégées éventuellement par des droits spécifiques ou des restrictions (contraintes d'accès et de réutilisation...).

Indiquer si vous combinez vos nouvelles données avec des anciens jeux de données. Vérifiez qu'une licence de réutilisation permet ce cas de figure.

Exemple de réponse:

Exemple de réponse 1 :

Ce projet s'appuie sur des enquêtes qualitatives préexistantes qui ont été numérisées et mises à disposition de la communauté.

Exemple de réponse 2 :

Nous avons utilisé les données de 3 enquêtes réalisées par notre équipe depuis 2015.

2.5. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Recommandations:

Mentionner si la qualité et la conformité de la collecte des données sont contrôlées et documentées.

Préciser si toutes les données documentées dans le cahier de laboratoire ont été exploitées.

3. RESSOURCES ALLOUEES A LA GESTION DES DONNEES

3.1. Quelles sont les ressources (budget et temps alloué) dédiées à la gestion des données ?

Exemple de réponse:

Coûts matériels, installation de machines virtuelles

3.2. La gestion des données au cours du projet nécessite-t-elle des recrutements ou des actions de formations ciblées ?

Recommandations:

Pour des formations ou un accompagnement dans la gestion des données s'adresser au Service Commun de la Documentation, [Services aux chercheurs](#)

Exemple de réponse:

Un ETP de [n]% pour couvrir la gestion et la documentation des données.

Des formations spécifiques ont été prévues en matière d'accompagnement dans la gestion des données.

4. DESCRIPTION DES METADONNEES DES DONNEES

4.1. Quelles métadonnées décriront les données ?

Recommandations:

Indiquer les standards et les formats de métadonnées utilisés (Dublin Core, DDI, EAD, TEI, EML, MARC, CMDI...). Préciser s'il s'agit du standard recommandé pour la discipline. Si aucun standard disciplinaire n'existe, expliquez la solution choisie. Pour s'informer sur les métadonnées, standards et formats : <https://doranum.fr/metadonnees-standards-formats/fiche-synthetique/>

Contactez sur ces sujets le Service Commun de la Documentation, [Services aux chercheurs](#) .

Exemple de réponse:

Les métadonnées sont spécifiées par un chercheur/ingénieur et s'organisent selon les champs recommandés par le standard Dublin Core qualifié.

4.2. Comment les métadonnées ont-elles été produites (cahier de laboratoire, GPS, entrée manuelle etc.) ? Par qui ont-elles été produites ?

Recommandations:

Indiquer s'il existe une méthodologie spécifique pour collecter les données.

Préciser par qui les données ont été produites.

4.3. Gestion des fichiers : gestion et organisation des fichiers, conventions de nommage des fichiers, gestion des versions, etc...

Recommandations:

Il est important d'avoir une stratégie de nommage des fichiers de données et de la respecter pour récupérer et identifier plus facilement les fichiers de données recherchés, éviter les problèmes lors de transfert et de partage.

Quelques bonnes pratiques de nommage des fichiers :

- choisir un nom court (- de 100 caractères) et significatif,
- pas d'espace, pas d'accent, pas de caractères spéciaux,
- utiliser des abréviations standards (ex : CR pour compte rendu),
- indiquer le nom de l'auteur,
- renseigner la date et/ou l'heure (le format de date attendu est de type AAAA, AAAA-MM ou AAAA-MM-JJ),
- indiquer le numéro de version du fichier (si le document est destiné à évoluer), penser à la numérotation (saisir des 0 initiaux pour les tris).
- Documenter les règles de nommage.

Exemple de réponse:

Exemple de réponse 1 :

GD_20210512_tableaucomparatif_v2

Exemple de réponse 2 :

Soulie_entretien_20210413_transcription_VF

Exemple de réponse 3 :

Titre_projet_Jeudedonnées1(identifiant)_aaaammij

4.4. Documentation sur les données

Recommandations:

Mentionner s'il existe une documentation sur les données afin de faciliter la réutilisation par d'autres chercheurs ou dans un but commercial. Contacter sur ces sujets le Service Commun de la Documentation, [Services aux chercheurs](#).

5. RESPONSABILITES, EXIGENCES ETHIQUES ET DROITS DE PROPRIETE INTELLECTUELLE

5.1. Qui est responsable de la collecte des données ?

Exemple de réponse:

Nom prénom, institution, ville, adresse mail

Nom Prénom, UMR Structures Formelles du Langage (SFL), Paris, prenom.nom@univ-paris8.fr

5.2. Qui détiendra les droits de propriété intellectuelle pour les données créées lors du projet ? Quelle est la législation applicable en la matière ?

Exemple de réponse:

Nom Prénom, UMR Structures Formelles du Langage (SFL), Paris, prenom.nom@univ-paris8.fr

Recommandations:

Indiquer le propriétaire des données, c'est-à-dire qui aura le droit d'en contrôler l'accès.

Pour les questions de propriété intellectuelle et de droits sur les données de la recherche, contacter le [Délégué à l'intégrité scientifique](#) de Paris 8

Décrire le projet collaboratif (quelles institutions y prennent part, de quelles manières...). Si les données ont fait l'objet d'une convention entre les partenaires du projet, en préciser les termes. Pour les conventions, contacter la Direction de la recherche, [Service Valorisation de la recherche](#)

5.3. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Recommandations:

Indiquer si la réglementation sur la protection des données est appliquée (RGPD <https://www.cnil.fr/outil-pia-telechargez-et-installez-le-logiciel-de-la-cnil>).

Lorsqu'un projet comporte des données à caractère personnel, le Délégué à la Protection des données (DPO) doit être obligatoirement consulté avant sa mise en oeuvre ou sa modification. Contacter le [Délégué à la protection des données](#) de Paris 8.

Exemple de réponse:

Exemple de réponse 1 :

Données personnelles (collectées lors d'enquêtes) anonymisées ; utilisation [logiciel PIA de la CNIL](#) pour décrire la sensibilité des données...

Exemple de réponse 2 :

Recherche effectuée en partenariat avec des pays hors Europe : les recommandations du manuel Horizon Europe ont été respectées

5.4. Comment les questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Recommandations:

Indiquer si les questions d'éthique ont une incidence sur le stockage et la conservation des données.

Se renseigner auprès du [Délégué à l'intégrité scientifique](#) de Paris 8.

Exemple de réponse:

Les données produites s'appuient sur des processus de recherche et d'étude clinique. Des consultations avec le [comité d'éthique](#) de Paris 8 ont été instaurées notamment pour les données d'étude clinique.

6. STOCKAGE ET PARTAGE DES DONNEES AU COURS DU PROJET

6.1. Sur quels supports les données sont-elles produites ou échangées au cours du projet ?

Un même type de données peut être conservé sur plusieurs supports.

Exemple de réponse:

Exemple de réponse 1 :

Les questionnaires remplis par les enquêtés sont numérisés au fil de l'eau mais l'original est conservé sur support papier pendant la durée du projet.

Exemple de réponse 2 :

Les cahiers de laboratoire sont conservés sur support papier.

Exemple de réponse 3 :

Les données sont entièrement numériques.

6.2. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Recommandations:

Indiquer où sont stockées et sauvegardées les données. Mentionner si les données sont sécurisées. Le stockage des données sur des ordinateurs portables, des disques durs externes, des clés USB ou dans des nuages commerciaux (Dropbox, Google Drive, etc.) n'est pas recommandé.

Mentionner les procédures de sauvegarde ou de récupération des données

Exemple de réponse:

Les données seront stockées via le service [ShareDocs](#) de la TGIR Huma-Num sur un serveur hautement sécurisé, conformément au conventionnement établi.

Sur la sauvegarde des données : 3 copies sur 2 supports différents, dont 1 copie à distance.

6.3. Avez-vous besoin d'installation d'une machine virtuelle (VM) pour votre projet de recherche ?

Recommandations:

Indiquer si vous travaillez dans un environnement spécifique dans le cadre de votre projet que vous souhaitez conserver ?

Contactez les responsables du projet « Data Paris 8 » pour les demandes d'installation d'une machine virtuelle.

Exemple de réponse:

Dans le cadre de mon projet de recherche CAREXIL-FR je travaille dans un environnement TEI via la plateforme TEITOK et je souhaite conserver l'environnement Teitok. C'est pourquoi je demande à avoir accès à une machine virtuelle pour le déploiement des données de mon projet Carexil.

6.4. Modalités de partage et d'accès aux données pour les partenaires du projet ou pour d'autres personnes

Recommandations:

Indiquer les personnes ayant accès aux données et les niveaux d'accès pendant le processus de recherche.

Préciser comment le partage de données est organisé entre les différents partenaires pendant le processus de recherche.

Exemple de réponse:

Exemple de réponse 1 :

Les (jeux de) données seront déposé(e)s sur [ShareDocs](#) et partagé(e)s avec tous les partenaires du projet.

Exemple de réponse 2 :

Les données collectées seront partagées avec l'unité de recherche [...] de l'Université X... en charge de la thématique Y via la plateforme open source Open Science Framework : <https://osf.io/>

Exemple de réponse 3 :

Le projet XXX alliant à la fois des partenaires privés et publics, il sera nécessaire de détailler l'accès aux données selon les différents cas de figure.

7. DIFFUSION, PARTAGE ET REUTILISATION DES DONNEES A L'ISSUE DU PROJET

7.1. Quelles sont les modalités de diffusion de données à l'issue du projet de recherche ?

Recommandations:

Définir le type d'accès (ouvert, restreint, embargo). L'accès peut dépendre de la nature des données :

Les données produites dans le cadre du projet devront, autant que possible (« Aussi ouvert que possible, aussi fermé que nécessaire ») se conformer aux principes regroupés sous l'appellation « FAIR data » (Faciles à trouver, Accessibles, Interopérables et Réutilisables).

Dans le cas des programmes financés par certaines agences (ANR, ERC), la diffusion des données est obligatoire. Il peut cependant exister certaines restrictions liées à la nature des données. Vérifier à partir de [cet outil d'aide à la décision sur la diffusion des données](#) mis en place par le CIRAD.

Préciser s'il existe des restrictions au partage des données (embargo, délais de publication, pour certaines communautés...). La présence de données sensibles n'exclut pas nécessairement toute diffusion des données. L'embargo se présente justement comme un compromis permettant de signaler l'existence des données (enjeux d'impact) sans compromettre leur exclusivité d'exploitation. Il permet de garantir à moyen terme leur accessibilité et leur réutilisation sans risque de perte entre la fin du projet et le moment de la diffusion. Les données non diffusées ou non rendues librement accessibles sous licence open data restent soumises à des exigences d'archivage (voir infra « archivage et préservation à long terme des données »).

Enfin, il y a le cas de présence de données sensibles justifiant une dispense au principe de diffusion : données protégées, personnelles, stratégiques, issues de partenariats privés, etc. Indiquez les raisons pour lesquelles certains jeux de données (préciser lesquels) pourraient ne pas être diffusés ?

Exemple de réponse:

Exemple de réponse 1 :

Les données seront diffusées selon le principe "aussi ouvert que possible, aussi fermé que nécessaire"

Exemple de réponse 2 :

Les données seront diffusées après une période d'embargo de 6 mois.

Exemple de réponse 3 :

Les données cliniques sont destinées à être diffusées mais bénéficient d'une durée d'embargo d'un an pour priorité d'exploitation.

Exemple de réponse 4 :

Données d'enquête non anonymisées

7.2. Perspectives de réutilisation des données : quels potentiels de réutilisation verriez-vous pour ces données en dehors de votre projet ?

Recommandations:

Indiquer si la réutilisation, l'utilisation commerciale et la création de nouveaux ensembles de données ou produits dérivés sont autorisées.

Exemple de réponse:

Les données climatologiques peuvent servir de sources pour des travaux de recherche en histoire ou en géographie.

7.3. Où seront stockées les données en vue de leur diffusion ?

Recommandations:

Déposer les données issues de la recherche dans un entrepôt de données permet de les stocker de manière pérenne pour y avoir accès ensuite et pouvoir les réutiliser.

Exemple de réponse:

Les données de recherche issues de ce projet seront déposées dans l'entrepôt [Nakala](#) conformément au conventionnement établi avec la [TGIR Huma-Num](#), afin d'assurer leur accès à long terme par la communauté scientifique.

7.4. Standards et formats de métadonnées

Recommandations:

Documenter avec soin les données produites, tout au long de leur cycle de vie (métadonnées descriptives, métadonnées de gestion, métadonnées d'archivage etc.) est une garantie d'interopérabilité. Plus les métadonnées (champs et contenus) seront structurées plus il sera aisé de les exploiter à des fins de recherche ou d'indexation.

Il existe des formats très génériques de métadonnées pour qualifier des objets numériques (Dublin Core, DataCite Metadata Schema). [Nakala](#) utilise par défaut le format DublinCore qualifié. Les champs obligatoires sont : le type de dépôt, le titre, le(s) auteur(s), la date de création, la licence. Les autres champs du format Dublin Core sont facultatifs et peuvent, par exemple, être complétés plus tard.

La recherche de l'interopérabilité privilégiera des formats libres et ouverts à des formats propriétaires.

Pour des questions relatives aux métadonnées, s'adresser au Service Commun de la Documentation, [Services aux chercheurs](#).

Exemple de réponse:

Les métadonnées sont spécifiées par un chercheur/ingénieur et s'organisent selon les champs recommandés par le standard Dublin Core qualifié.

7.5. Quel est le volume des données à stocker ?

Il peut être différent de la volumétrie totale.

Exemple de réponse:

Exemple de réponse 1 : 100 Go

Exemple de réponse 2 : ~ 500 Go

Exemple de réponse 3 : 1 To

7.6. Identifiant pérenne attribué aux données par l'entrepôt

Recommandations:

Il existe différents types d'identifiant numérique (DOI, Handle, POI).

L'entrepôt Nakala fournit un DOI au format suivant : <https://doi.org/11280/23785026>.

Exemple de réponse:

DOI, Handle, Ark...

7.7. Type de licence utilisé pour la diffusion des données

Exemple de réponse:

Exemple de réponse 1 :

L'ensemble des données diffusées sera placé sous licence libre Etalab ou sous licence CC-BY.

Exemple de réponse 2 :

Les bases de données seront placées sous licence ODbL. Les outils et logiciels mis au point dans le cadre du traitement et de l'analyse des données seront placés sous licence GNU GPL.

Recommandations:

Afin de garantir l'accès, l'interopérabilité et la réutilisation la plus large possible des données, il est recommandé de privilégier des licences libres largement partagées et documentées type CC, CC-BY, CC-BY-NC, CC0, GNU, Open database License (ODbL), Licence ouverte.

Voir la fiche de DoRANum sur [les licences de réutilisation dans le cadre de l'Open Data](#).

7.8. Préciser s'il y a une stratégie mise en place de partage des données stockées via une bibliothèque numérique ou un portail de données (institutionnel...).

Recommandations:

Pour des questions relatives à la bibliothèque numérique de Paris 8 (Octaviana) s'adresser au Service Commun de la Documentation,

[Services aux chercheurs](#)

Exemple de réponse:

Les données déposées dans Nakala seront déversées dans la bibliothèque numérique [Octaviana](#) de l'Université Paris 8, dans la collection correspondant au projet.

7.9. Publications associées aux données et au projet

Recommandations:

Préciser s'il est prévu la publication d'un « data paper » ou de « supplementary data ».

Exemple de réponse:

Les données diffusées via l'entrepôt [Nakala](#) feront également l'objet d'un « data paper » dans la revue *Cybergeo*.

8. ARCHIVAGE ET PRESERVATION A LONG TERME DES DONNEES

8.1. Quel est l'intérêt à la préservation à long terme des données et quelles données ont vocation à être préservées sur le long terme (plus de 10 ans)

Recommandations:

La sélection peut être effectuée à partir de plusieurs critères : potentiel de réutilisation scientifique, valeur de preuve, valeur historique, collecte exceptionnelle etc. Préciser quelles sont les données à préserver. Mentionner si les données seront anonymisées.

Exemple de réponse:

Exemple de réponse 1 :

Les données X ont vocation à être préservées car elles ont une grande valeur historique

Exemple de réponse 2 :

Les données Y ne sont pas reproductibles et ont un intérêt à être préservées.

Exemple de réponse 3 :

Les données Z ont vocation à être préservées car leur coût de production est extrêmement élevé.

8.2. Où seront préservées les données ?

Recommandations:

Indiquer l'entrepôt choisi pour l'archivage pérenne.

Exemple de réponse:

Exemple de réponse 1 :

Les (jeux de) données sont déposé(e)s sur [Nakala](#) : <https://doi.org/10.34847/nkl.fde4s8ly>

Exemple de réponse 2 :

Les données seront archivées via les services de la TGIR HumNum (au Centre Informatique National de l'Enseignement Supérieur - [CINES](#)), conformément au conventionnement établi.

8.3. Destruction de données

Recommandations:

Précisez si certains types de données seront détruits (raisons légales, contractuelles, réglementaires, données à caractère personnel conformément aux recommandations de la CNIL).

Exemple de réponse:

Exemple de réponse 1 :

Les noms des personnes interviewés seront supprimés de l'enquête sur les pratiques culturelles.

Exemple de réponse 2 :

Suppression des fichiers intermédiaires de l'analyse de données.

8.4. Quelles méthodes ou quels outils seront nécessaires pour accéder et utiliser les données ?

Recommandations:

Indiquer éventuellement les outils et logiciels nécessaires pour avoir accès et pouvoir réutiliser les données.

Exemple de réponse:

Les données stockées au sein de l'infrastructure X seront accessibles via l'outil Y du laboratoire Z à tout utilisateur présent dans l'annuaire de la structure.

8.5. Indiquer si les données archivées seront partagées via une bibliothèque numérique ou un portail de données (institutionnel...).

Recommandations:

Pour des questions relatives à la bibliothèque numérique de Paris 8 (Octaviana) s'adresser au Service Commun de la Documentation, [Services aux chercheurs](#)

Exemple de réponse:

A l'issue du projet, les données déposées dans [Nakala](#) seront déversées dans la bibliothèque numérique [Octaviana](#) de l'Université Paris 8, dans la collection correspondant au projet.

8.6. Quelle est la durée de conservation des données ?

Recommandations:

Durée préconisée (tenir compte des exigences légales et/ou réglementaires existantes).

Exemple de réponse:

Exemple de réponse 1 :

Les données seront conservées 30 ans minimum, conformément au conventionnement établi.

Exemple de réponse 2 :

Les données seront conservées 20 ans minimum conformément au décret 2006-6 du 4 janvier 2006 fixant la conservation des données de patients.

Exemple de réponse 3 :

Seules les données sources ainsi que les données venant en appui des résultats publiés seront conservées pour une durée minimum de 15 ans. L'ensemble des données intermédiaires seront supprimées dans l'année suivant l'arrêt du projet.

8.7. Volume final des données archivées

Recommandations:

Une fois les données à archiver choisies, reporter leur volume final.

Exemple de réponse:

Exemple de réponse 1 : 80 Go

Exemple de réponse 2 : ~ 300 Go

Exemple de réponse 3 : 1 To

8.8. Responsable de l'archivage et de la conservation des données

Recommandations:

Nom et affiliation du responsable de l'archivage et de la conservation des données. Ou organisme producteur de l'entrepôt si celui-ci garantit l'archivage pérenne.

Exemple de réponse:

Exemple de réponse 1 :

Nom Prénom, UMR Laboratoire d'études de genre et de sexualité (LEGS), Aubervilliers, prenom.nom@univ-paris 8.fr

Exemple de réponse 2 :

TGIR Huma-Num