

## Université de Paris 7 - Denis Diderot: FAIR DMP Paris Descartes/Diderot - Contexte, acteurs, ressources

### Description du projet

Quel est l'organisme financeur du projet ?

Dans quel programme de recherche s'inscrit le projet ?

Indiquez l'acronyme ou le titre du projet

Rappelez les objectifs et les attendus du projet.

Indiquez les thèmes ou mots-clés décrivant le projet.

*Recommandations:*

Domaines ERC : [http://www.must-cofund.ch/pdf/ERC\\_keywords.pdf](http://www.must-cofund.ch/pdf/ERC_keywords.pdf)

Indiquez les références et la localisation de la documentation associée.

*Exemple de réponse:*

*Ex.1 : [Glossaire partagé](#) réalisé dans le cadre d'un projet interdisciplinaire d'imagerie du vivant (IDV)*

*Ex.2 : Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from <http://handbook.cochrane.org> : Handbook de la Cochrane Library pour la réalisation d'une revue systématique de la littérature dans le cadre d'un projet de recherche orienté Evidence Based Medicine (EBM)*

### Acteurs du projet

Qui sont les partenaires du projet ?

*Recommandations:*

**Partenaires du projet** - Ce champ comporte la liste des établissements et partenaires du projet, notamment ceux impliqués dans la production et la gestion de données. Dans le cas d'un projet associant des partenaires sur d'autres sites ou dans d'autres institutions, les responsabilités entre les partenaires doivent être précisées.

Qui sont le/les responsable(s) de la gestion des données et de la mise à jour du DMP au cours du projet ?

### Ressources nécessaires

La gestion des données au cours du projet (de la production à la diffusion) nécessite t-elle l'acquisition ou l'installation de matériel spécifique (serveurs, machines virtuelles, infrastructures, etc.) ? Si oui, précisez lesquels ?

*Recommandations:*

**Ressources nécessaires** - Les moyens mis en œuvre pour la gestion des données peuvent être humains ou matériels. Ils doivent garantir la réussite du projet à travers un nombre d'ETP suffisants, des compétences spécifiques (archivistiques, informatiques, documentaires, curation de données etc.), un soutien et un accompagnement institutionnels adaptés, la mise à disposition d'outils informatiques, etc.

Si des faiblesses persistent dans le dispositif, il peut être intéressant de les mentionner dans le DMP tant pour justifier d'éventuelles pratiques que pour permettre leur consolidation.

*Exemple de réponse:*

*Ex. Serveurs Petabox, machines virtuelles, logiciels dédiés, licences MatLab supplémentaires etc.*

La gestion des données au cours du projet (de la production à la diffusion) nécessite t-elle des recrutements ou des actions de formations ciblées ? Si oui, précisez lesquels ?

*Recommandations:*

**Besoins en formation** - Le fait de disposer de personnels formés ou à défaut sensibilisés à la gestion des données, à la rédaction ou aux enjeux d'un plan de gestion de données peut faciliter la collaboration des partenaires autour de ces sujets. Selon la nature du projet, des formations plus spécialisées à la conception de bases de données ou à la manipulation de données disciplinaires (*big data*, algorithmes de traitement, méthodes d'enrichissement, ontologies, etc.) peuvent également se révéler pertinentes à défaut de compétences identifiées parmi les partenaires ou de services supports dédiés au sein du projet.

**Personnes ressources** : services informatiques ; bibliothécaires ; archivistes ; documentalistes ; les URFIST

**Liens et outils :**

- [DoraNum](#): plateformes d'apprentissage numérique sur les données de recherche
- [Les URFIST](#) (Unité Régionale de Formation à l'Information Scientifique et Technique)
- [Webinar](#) (english) sur le Data management plan

*Exemple de réponse:*

*Ex. Un ETP de [n]% a été budgété pour couvrir la gestion, la documentation et la curation des données.*

-----  
*Des formations spécifiques en matière d'analyse et de curation de données ont été prévues en vue de la gestion du "big data".*

Sur la base des besoins identifiés ci-dessus, à combien estimez-vous le surcoût attaché à la gestion des données ? à leur diffusion ? à leur archivage ?

*Recommandations:*

#### **Estimation des coûts**

Pour rappel, dans le cadre des projets européens H2020, les coûts attachés à la gestion des données sont éligibles. Précisez les estimations avancées concernant :

- *Les coûts de stockage et de traitement* des données sont à prendre en compte dans cette estimation. Les prix sont généralement à la baisse pour des volumétries à la hausse mais il convient de ne pas uniquement raisonner en termes de stockage mais également en termes de services : entrées et sorties de données, maintenance, fréquence d'accès, sécurité, etc. Les services peuvent représenter des coûts masqués importants. Concernant les solutions de stockage externe, il n'existe pas de grilles tarifaires fixes. Elles varient selon le type de données, de services et de stockage (cloud, SATA, SAS etc.).
- *Les coûts de publication et de dissémination* des données : maintenance d'une base de données, création d'un site web, frais de publication (ou *articles processing charges* - APC) pour les *data papers*, frais de *supplementary data* d'un article etc. Les coûts peuvent fortement varier d'une discipline à l'autre (APC SHS < APC SVS ou STM).
- *Le coût de l'archivage* peut être évalué sur la base d'une volumétrie prévisionnelle en demandant des devis auprès des [plateformes d'archivage agréées](#).

*Exemple de réponse:*

*Ex. The [repository] has estimated their additional cost to archive the data is [insert euro amount]. This fee appears in the budget for this application as well.*

Comment seront couverts les besoins financiers ?

*Selon le financeur ou le programme de financement, les coûts liés à l'ouverture des données peuvent être éligibles, dans des conditions précisées par la convention de financement.*

# Université de Paris 7 - Denis Diderot: FAIR DMP Paris Descartes/Diderot - Stockage, partage, protection et diffusion

## Informations générales sur les données

Quels sont les méthodes et processus de production des données utilisés au cours du projet ?

*Exemple de réponse:*

*Ex. 1 : Les données ont été recueillies par questionnaire, puis intégrées dans une base de données Epidata, version 3.0.*

*Ex. 2 : Les données ont été collectées suivant la norme ISO/CEI 17025 sur les « exigences générales concernant la compétence des laboratoires d'étalonnages et d'essais ».*

Le projet s'appuie-t-il sur la réutilisation de données existantes ? Si oui, lesquelles et de quelle origine ?

*Exemple de réponse:*

*Ex. 1 : Ce projet s'appuie sur des enquêtes qualitatives préexistantes, qui ont été numérisées et mises à disposition de la communauté.*

*Ex. 2 : Nous avons utilisé les données de 4 enquêtes cas-témoins réalisées par notre équipe depuis 1998*

Comment se définit la propriété des données produites et collectées au cours du projet entre les différents partenaires ?

*Exemple de réponse:*

*Ex.1 : Résultats Propres : les Résultats seront la propriété du Partenaire qui a réalisé (ou fait réaliser par un sous-traitant, sans financement des autres Parties) les études/travaux qui sont à l'origine des Résultats. Le partenaire propriétaire des Résultats Propres sera seul juge de l'opportunité et du choix de leurs modes de protection et d'utilisation.*

*Ex.2 : Résultats Communs : Si les Résultats ont été générés par les études/travaux de plus d'un Partenaire (ou développés par un sous-traitant, financé conjointement par deux ou plusieurs Partenaires), chaque Partenaire aura une part égale de copropriété sur ces résultats.*

Précisez le type et la nature des données qui seront produites au cours du projet.

*Exemple de réponse:*

*Ex.1 Data and biological materials to be generated by the proposed project will mainly consist of images (e.g. confocal and transmission electron micrographs).*

*Ex. 2 Données d'enquêtes individuelles et données géographiques issues du recensement.*

Quels seront les principaux formats de production, de traitement ou de conversion des données utilisés au cours du projet ainsi que les outils et logiciels de lecture associés ?

*Exemple de réponse:*

*Ex. 1 : All data will be stored in digital form, either in the format in which it was originally generated (i.e. Metamorph files, for confocal images; Spectrum Mill files, for mass spectra with results of mass spectra analyses stored in Excel files; tiff files for gel images; Filemaker Pro files for genetics records), or will be converted into a digital form via scanning to create tiff or jpeg files (e.g. western blots or other types of results).*

*Ex. 2 : Quantitative survey data files generated will be processed and submitted to the [repository] as SPSS system files with DDI XML documentation. The data will be distributed in several widely used formats, including ASCII, tab-delimited (for use with Excel), SAS, SPSS, and Stata. Documentation will be provided as PDF. Data will be stored as ASCII along with setup files for the statistical software packages. Documentation will be preserved using XML and PDF/A.*

Avez-vous identifié des standards et/ou des formats de métadonnées permettant de décrire vos données ?

*Exemple de réponse:*

*Ex.1 Il n'existe pas à notre connaissance de standards de métadonnées pour décrire les données produites mais un modèle sera mis au point et documenté dans le cadre du projet. Les outils et logiciels de gestion développés dans ce contexte seront également documentés et autant que possibles rendu accessible.*

*Ex. 2: Les données issues de l'imagerie médicales seront décrites selon la norme et le format d'interopérabilité DICOM.*

## Stockage et partage des données au cours du projet

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

*Recommandations:*

**Support des données** - Un même type de données peut être conservé sur plusieurs supports (papier, serveur, disque dur, clé USB, etc.). Disposer d'une vision globale et consolidée des supports tout au long du cycle de vie est une garantie de sécurité et de pérennité. L'obsolescence des supports peut être tout aussi problématique que l'obsolescence des

formats (disparition des lecteurs de disquettes, fragilité des disques durs, disparition des magnétoscopes etc.) et peut nécessiter des transferts de données d'un support à un autre.

*Exemple de réponse:*

*Ex. 1 : Les questionnaires remplis par les patients sont numérisés au fil de l'eau mais l'original est conservé sur support papier pendant la durée du projet.*

*Ex. 2 : Les cahiers de laboratoire sont conservés sur support papier.*

Quels seront les lieux et les modalités de stockage et d'hébergement des données du projet ?

*Exemple de réponse:*

*Ex. Serveur local, serveur distant, disque dur externe, etc.*

De quelle manière seront gérés l'accès et le partage des données entre les partenaires au cours du projet ?

*Exemple de réponse:*

*Ex. The data will be processed and managed in a secure non-networked environment using virtual desktop technology.*

*Les fichiers de données sont diffusés aux équipes de recherche productrices par une plateforme d'accès à distance sécurisée. Un comité d'accès aux données sera mis en place.*

Indiquez le volume prévisionnel de données produites au cours du projet ?

*Exemple de réponse:*

*Ex. [n] Megaoctets ; [n] Teraoctets ...*

Indiquez les règles de nommage et de classement des informations et données produites sur les espaces de partage du projet.

*Recommandations:*

**Règles de nommage** - Dans les projets collaboratifs, il est souvent nécessaire de mettre en place des espaces de partage de données ou de fichiers. En l'absence d'application de gestion électronique de document (GED) ou de base de données structurée, l'adoption de règles de nommages communes peut faciliter le partage et la recherche de fichiers entre les partenaires.

Pour des raisons d'intégrité, le nommage doit respecter certaines règles formelles (pas d'espace, de ponctuation, de caractères spéciaux ou de lettres accentuées) et pour les fichiers non sériels porter certaines mentions (date, objet, version etc.).

Des métadonnées additionnelles peuvent également venir enrichir les propriétés des documents en complément des règles de nommage.

**Personnes ressources** : archivistes, bibliothécaires

**Liens et outils**

- un [outil de gestion des doublons](#)
- [des règles de nommage des fichiers](#) de l'Université de Lausanne ;
- [Les règles d'enregistrement](#) des domaines Internet
- des outils libres de renommage des fichiers : [ReNamer](#) et [Ant Renamer](#)

*Exemple de réponse:*

*Ex. Identifiant\_date; Objet\_annee-mois-jour\_initial ;*

*Brucker\_acqfile\_001\_20171108.dsc ;*

*Dupond\_entretien\_20120413\_transcription\_VF*

## Risques, sécurité et éthique des données

Quels sont les risques et menaces qui pèsent sur les données ?

*Recommandations:*

**Risques sur les données** - Différents facteurs sont susceptibles de menacer l'intégrité, la disponibilité et la confidentialité des données produites au cours du projet. Les risques peuvent être de différentes natures : des risques naturels pesant sur les infrastructures (zones sismiques, inondables etc.), des risques techniques (corruptions ou pertes de données, problèmes de serveurs etc.), des risques de confidentialité (accès non autorisés, fuites de données sensibles, etc.)

**Personnes ressources** : services informatiques ; CIL/Délégué à la protection des données

**Liens et outils**

- Quelques exemples et anecdotes de [risques et incidents avérés](#)
- [DRAMBORA](#) : outil interactif d'évaluation des risques

*Exemple de réponse:*

*Ex.1: Destruction accidentelle : Une fois le terrain fini, les données brutes sont sauvegardées sur un serveur BLAISE auquel*

trois ou quatre personnes ont accès. Ensuite, le chargé d'études les télécharge sur son poste pour la phase d'apurement, normalisation et enrichissement. Pendant cette étape et jusqu'au moment de leur diffusion, les données sont sauvegardées uniquement sur son poste de travail, ce qui, en cas de panne, pourrait signifier la perte de ces fichiers et du travail réalisé.

-----  
*Ex. 2 : De-synchronisation des fichiers diffusés sur les différentes plateformes de diffusion, vu la variété des procédures de publication et des formats des fichiers : xml sur Nesstar, zip avec pdf et SPSS sur le portail Quetelet.*

Quelles seront les mesures et les solutions mises en œuvre pour parer les risques identifiés ? Quels sont les précautions et les moyens prévus pour garantir la protection des données, notamment en matière de confidentialité ?

*Exemple de réponse:*

*Ex. 1: La sécurité est assurée par cryptage des fichiers système et par un accès en https. Par ailleurs, les journaux des opérations sont enregistrés quotidiennement, les liens de téléchargement sont à durée de vie courte, les dossiers à télécharger sont protégés par mot de passe et le nombre de téléchargements autorisés est limité à un par destinataire.*

-----  
*Ex. 2: La déclaration du dispositif ELIPSS a été inscrite au registre CIL du CNRS sous le numéro 2-12030 avec pour finalité du traitement la réalisation d'enquêtes à des fins de recherche. Cette déclaration décrit précisément le système d'information mis en œuvre, les moyens utilisés pour assurer la sécurité et la confidentialité des données, la procédure de gestion des enquêtes (assurée par le CDSP) et la gestion des contacts avec les panélistes (assurée à l'INED jusqu'en janvier 2015). Les données nominatives et les données d'enquêtes sont stockées dans deux systèmes d'information différents. Le traitement pour les données nominatives du panel a été enregistré au registre CIL de l'INED sous le numéro 2012-CIL-0012*

Le projet pose-t-il des questions éthiques ? Quelles sont les dispositions prises pour y répondre ?

*Exemple de réponse:*

*Ex. 1 Since the dataset consists of information extracted from articles, no ethical issues are expected.*

-----  
*Ex. 2 Les données produites s'appuient sur des processus de recherche et d'étude clinique. Des consultations régulières avec le Centre d'éthique Clinique ont été instaurées notamment pour les données d'étude portant sur relatives aux décisions de pose de sonde d'alimentation entérale chez les personnes âgées incompétentes.*

## **Dissémination et archivage des données**

Quel potentiel de réutilisation verriez-vous pour ces données en dehors de votre projet ?

*Exemple de réponse:*

*Ex. 1 : Les données climatologiques peuvent servir de sources pour des travaux de recherche en histoire ou en géographie.*

-----  
*Ex. 2: In addition to the research community, we expect these data will be used by practitioners and policymakers.*

Les données produites au cours du projet ont-elles vocation à être publiées ? Sont-elles soumises à un principe général de diffusion édicté par le financeur, un programme, un engagement etc. ?

*Exemple de réponse:*

*Ex. H2020 : Conformément à la révision de 2017 du programme de travail, les dispositions de l'Open Research Data Pilot (ORD pilote) lancé par la Commission Européenne dans le cadre de ses appels à projet H2020 est désormais étendu à l'ensemble des thématiques du programme (cf. article 29.3 du "Grant Agreement"). Les données produites dans le cadre du projet devront donc, autant que possible ("As open as possible as closed as necessary"), se conformer aux principes regroupés sous l'appellation "FAIR data" [Findable, Accessible; Interoperable, Reusable]. Cela suppose : 1) le dépôt dans un entrepôt ouvert des données et métadonnées permettant la validation des résultats publiés et de toutes autres données ou métadonnées jugées pertinentes dans le cadre du projet, 2) le placement des données sous licence ouverte autorisant à minima, l'accès, la fouille, l'exploitation, la reproduction et la dissémination des données (licence CC-By ou CCO), 3) le partage des informations relatives aux outils nécessaires à la validation ou à la lecture des données voire le partage des outils eux-mêmes (logiciels, algorithmes, codes etc.)*

Quelles sont les modalités de diffusion envisagées pour la dissémination des données et jeux de données ?

*Exemple de réponse:*

*Ex. 1 : The project team will create a dedicated Web site to manage and distribute the data because the audience for the data is small and has a tradition of interacting as a community. The site will be established using a content management system like Drupal or Joomla so that data users can participate in adding site content over time, making the site self-sustaining. The site will be available at a .org location.*

-----  
*Ex. 2 : Les données de recherche issues de ce projet seront déposées dans Zenodo afin d'assurer leur accès à long terme par la communauté scientifique.*

Quelles sont les modalités et garanties d'archivage envisagées pour vos données à l'issue du projet (plateformes, partenaires, durée de conservation etc.)

*Exemple de réponse:*

*Ex.1: Une convention d'archivage avec le CINES est à l'étude en vue de l'archivage à moyen et long terme des données*

*issues du projet.*

## Université de Paris 7 - Denis Diderot: FAIR DMP Paris Descartes/Diderot - Dissémination et archivage

### Identification des jeux de données

Combien de jeux de données destinés à être archivés et/ou diffusés comporte votre projet ?

*Exemple de réponse:*

*Ex. 1 : The dataset will be defined as all data generated by the proposed project.*

*Ex. 2 : Datasets will be defined by each workpackage there are from.*

Existe-t-il des liens ou des relations spécifiques entre certains jeux de données ? Si oui, précisez lesquels (parenté, complémentarité, etc.) ?

*Exemple de réponse:*

*Ex. 1 : Le jeu de données 2 correspond aux données dérivées issues des processus de traitement appliquées aux données brutes du jeu de données 1 .*

*Ex. 2 : Les jeux de données correspondent chacun aux données issues d'une enquête annuelle. Le dernier jeu de données correspond aux données de synthèse et d'analyse de l'ensemble des enquêtes annuelles.*

### Protection - Exception de diffusion

Indiquez les raisons pour lesquelles certains jeux de données (préciser lesquels) pourraient ne pas être diffusés ?

**Attention :** les jeux de données non diffusés ou non rendu librement accessibles sous licence open data restent soumis à des exigences d'archivage.

*Exemple de réponse:*

#### **Exception justification**

*Ex. 1 : This project will generate data designed to study the prevalence and correlates of DSM III-R psychiatric disorders and patterns and correlates of service utilization for these disorders in a nationally representative sample of over 8000 respondents. The sensitive nature of these data will require that the data be released through a restricted use contract.*

*Ex. 2 : Données à caractère personnel.*

*Ex. 3 : Demande de brevet en cours.*

### Description des jeux de données techniquement homogènes

#### **Champs communs (cohérence intellectuelle)**

- Intitulé du jeu de données
- Identifiant du jeu de données
- Auteurs et contributeurs
- Editeur (institution)
- Mots-clés, vocabulaires et ontologies (description des contenus)

#### **Champs à dupliquer pour chaque sous-jeu de données (cohérence technique)**

- Nature et type de données
- Format et outils de lecture associés
- Standards et formats de métadonnées
- Modalités de diffusion et localisation des données
- Durée de l'embargo (si nécessaire)
- Type de licence

*Exemple de réponse:*

#### **Identifiants**

*Ex. 1: doi:10.1594/PANGAEA.726855*

*Ex. 2: set\_1\_raw\_data*

#### **Nature ou type de données**

*Ex. Data and biological materials to be generated by the proposed project will mainly consist of images (e.g. confocal and transmission electron micrographs).*

#### **Vocabulaire et ontologies**

*Ex. La description des données et brutes dérivées s'appuiera sur les ontologies SNOMED et OntoPneum exprimé selon un formalisme OWL.*

### **Métadonnées**

*Ex. 1 : Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively.*

*Ex. 2 : Ecological Metadata Languages (EML) and Crystallographic Information Framework (CIF)*

### **Modalités diffusion**

*Ex.1 : Les données relatives aux sciences de la terre et environnementales ont été diffusées via l'entrepôt ouvert PANGAEA et feront également l'objet d'un "data paper" dans la revue X.*

*Ex. 2 : L'ensemble des données du projet ont été déposées pour diffusion et archivage sur la plateforme ZENODO et sont désormais visibles et accessibles sur le portail OpenAire*

### **Embargo**

*Ex. 1 : Les données seront diffusées après une période d'embargo de 6 mois.*

*Ex. 2 : Les données astronomiques sont destinées à être diffusées mais bénéficient d'une durée d'embargo d'un an pour priorité d'exploitation.*

### **Licence**

*Ex. 1 : L'ensemble des données diffusées sera placé sous licence libre Etalab ou sous licence CC-BY-NC. Les conditions de la licence seront rendues accessibles aux machines sous un format ONYX PL*

*Ex. 2 : Les bases de données seront placées sous licence ODbI. Les outils et logiciels mis au point dans le cadre du traitement et de l'analyse des données seront placés sous licence GNU GPL*

## **Description des jeux de données techniquement hétérogènes (cohérence intellectuelle)**

### **Champs communs (cohérence intellectuelle)**

- Intitulé du jeu de données
- Identifiant du jeu de données
- Auteurs et contributeurs
- Editeur (institution)
- Mots-clés, vocabulaires et ontologies (description des contenus)

### **Champs à dupliquer pour chaque sous-jeu de données (cohérence technique)**

- Nature et type de données
- Format et outils de lecture associés
- Standards et formats de métadonnées
- Modalités de diffusion et localisation des données
- Durée de l'embargo (si nécessaire)
- Type de licence

*Exemple de réponse:*

#### **Identifiants**

*Ex. 1: doi:10.1594/PANGAEA.726855*

*Ex. 2: set\_1\_raw\_data*

#### **Nature ou type de données**

*Ex. Data and biological materials to be generated by the proposed project will mainly consist of images (e.g. confocal and transmission electron micrographs).*

#### **Vocabulaire et ontologies**

*Ex. La description des données et brutes dérivées s'appuiera sur les ontologies SNOMED et OntoPneum exprimé selon un formalisme OWL.*

#### **Métadonnées**

*Ex. 1 : Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively.*

*Ex. 2 : Ecological Metadata Languages (EML) and Crystallographic Information Framework (CIF)*

#### **Modalités diffusion**

*Ex.1 : Les données relatives aux sciences de la terre et environnementales ont été diffusées via l'entrepôt ouvert PANGAEA et feront également l'objet d'un "data paper" dans la revue X.*

*Ex. 2 : L'ensemble des données du projet ont été déposées pour diffusion et archivage sur la plateforme ZENODO et sont désormais visibles et accessibles sur le portail OpenAire.*



## Embargo

*Ex. 1 : Les données seront diffusées après une période d'embargo de 6 mois.*  
-----

*Ex. 2 : Les données astronomiques sont destinées à être diffusées mais bénéficient d'une durée d'embargo d'un an pour priorité d'exploitation.*

## Licence

*Ex. 1 : L'ensemble des données diffusées sera placé sous licence libre Etalab ou sous licence CC-BY-NC. Les conditions de la licence seront rendues accessibles aux machines sous un format ONYX PL*  
-----

*Ex. 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*

## Tri et archivage des données

Quelles données ont vocation à être conservées sur le long terme (plus de 10 ans) ? La sélection peut être effectuée à partir de plusieurs critères : potentiel de réutilisation scientifique, valeur de preuve, valeur historique, etc.

*Exemple de réponse:*

*Ex. 1 : Les données climatiques ne sont pas reproductibles et ont un intérêt à être conservées.*  
-----

*Ex. 2 : Les données astronomiques ont vocation à être conservées car leur coût de production est extrêmement élevé.*

Combien de temps chaque jeu de données doit-il être conservé ?

*Exemple de réponse:*

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

*Ex. 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.*

Quel est le volume des données à archiver ?

Quelle est la plateforme d'archivage choisie ?

*Recommandations:*

**Choix de la plateforme d'archivage** - Les plateformes d'archivage électronique choisies doivent être agréées par le Ministère de la Culture. Certaines universités disposent de conventions avec des plateformes : contactez le service d'archives de votre institution. Des spécificités existent concernant les données de santé. L'hébergement de données de santé à caractère personnel sur support informatique nécessite un agrément, accordé par le ministre de la santé (articles L.1111-8 et R.1111-9 à R.1111-15-1 du Code de la santé publique).

### Liens et outils

- La liste des [plateformes d'archivage agréées](#) par le ministère de la Culture
- La liste des [hébergeurs agréés pour les données de santé](#)
- [Test FACILE](#) : le test d'éligibilité des formats auprès du CINES

*Exemple de réponse:*

*Ex. 1 Les données seront archivées via le centre informatique [X] sur la plateforme [Y]*  
-----

*Ex. 2 Les données seront archivées via les services de la TGIR HumaNum conformément au conventionnement établi.*