

MASA Consortium: ANR - Modèle de PGD structuré (français)

Description des données et collecte ou réutilisation de données existantes

Description générale du produit de recherche

Recommandations:

- Donnez des détails sur le type de données : par exemple numérique (bases de données, tableurs), textuel (documents), image, audio, vidéo, et/ou médias composites.
- Expliquez comment les données pourraient être réutilisées dans d'autres contextes. Les identifiants persistants (PID) devraient être appliqués de manière à ce que les données puissent être localisées et référencées de façon fiable et efficace. Ces PID aident aussi à comptabiliser les citations et les réutilisations.
- Indiquez s'il sera envisagé d'attribuer aux données un PID. Typiquement, un entrepôt pérenne digne de confiance attribuera des identifiants persistants.

Est-ce que des données existantes seront réutilisées ?

Recommandations:

- Énoncer les éventuelles contraintes à la réutilisation des données préexistantes.
- Indiquer brièvement le cas échéant, les raisons pour lesquelles l'utilisation de sources de données existantes a été envisagée mais écartée.

Comment seront produites/collectées les nouvelles données ?

Recommandations:

- Expliquer quelles méthodologies ou quels logiciels seront utilisés si de nouvelles données sont recueillies ou produites.
- Expliquer comment la provenance des données sera documentée.

Documentation et qualité des données

Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Recommandations:

- Indiquez quelles métadonnées seront fournies pour aider à la recherche et à l'identification des données.
- Indiquez quelles standards de métadonnées seront utilisés (par exemple DDI, TEI, EML, MARC, CMDI).
- Utiliser les standards de métadonnées de communauté lorsque ceux-ci existent.
- Indiquez comment les données seront organisées au cours du projet, en mentionnant par exemple conventions de nommage, contrôle de version et structures des dossiers. Des données bien classées et gérées de façon cohérente seront plus faciles à retrouver, à comprendre et à réutiliser.
- Pensez à la documentation qui serait nécessaire pour permettre une réutilisation des données. Il peut s'agir notamment de l'information sur la méthodologie utilisée pour collecter les données, sur les procédures et méthodes d'analyse utilisées, sur la définition des variables, des unités de mesure, etc.
- Tenez compte de la façon dont ces informations seront obtenues et enregistrées par exemple dans une base de données avec des liens vers chacun des fichiers, dans un fichier texte de type « lisez-moi », dans les en-têtes de fichiers, dans un livre de référence (« code book ») ou dans les cahiers de laboratoire.

Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?

Recommandations:

- Expliquer comment la qualité et la conformité de la collecte des données seront contrôlées et documentées. Il s'agit là de préciser les processus comme la calibration, la répétition des échantillons ou des mesures, la capture standardisée des données, la validation de saisie des données, la revue par les pairs, ou la représentation basée sur des vocabulaires contrôlés.

Exigences légales et éthiques, code de conduite

Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Recommandations:

- Lorsque vous manipulez des données à caractère personnel, veillez à ce que les lois sur la protection des données (par exemple, RGPD) soient appliquées, notamment :
 - o Obtenir un consentement éclairé pour la préservation et/ou le partage de données personnelles.
 - o Envisager l'anonymisation des données personnelles pour la préservation et/ou le partage (des données correctement anonymisées ne sont plus considérées comme des données personnelles).
 - o Envisager la pseudonymisation des données personnelles (la principale différence avec l'anonymisation est que la pseudonymisation est réversible).
 - o Envisagez le chiffrement des données, qui est considéré comme un cas particulier de pseudonymisation (la clé de cryptage doit alors être stockée séparément des données, par exemple par un tiers de confiance).
 - o Expliquez si une procédure d'accès spécifique a été mise en place pour les utilisateurs autorisés à accéder aux données personnelles.

Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Recommandations:

- Expliquez qui sera le propriétaire des données, c'est-à-dire qui aura le droit d'en contrôler l'accès :
 - o Expliquez quelles conditions d'accès s'appliqueront aux données ? Les données seront-elles librement accessibles, ou des restrictions seront-elles appliquées ? Si oui, lesquelles ? Envisager l'utilisation de licences concernant l'accès et la réutilisation des données.
 - o Assurez-vous de couvrir, dans l'accord de consortium, ces questions de droits de contrôle d'accès aux données pour les projets multipartenaires et en cas de propriété partagée des données.
- Indiquez si les droits de propriété intellectuelle (par exemple la directive bases de données, droits sui generis) sont affectés. Dans l'affirmative, veuillez expliquer lesquels et comment cela sera-t-il traité.
- Indiquez s'il y a des restrictions sur la réutilisation des données fournies par des tiers.

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

Recommandations:

- Déterminer si les questions d'éthique ont une incidence sur la façon dont les données sont stockées et transférées, qui peut les voir ou les utiliser, et quelle durée de conservation leur seront-elles appliquées. Démontrer que ces aspects sont bien pris en compte et planifiés.
- Adoptez les codes de conduite nationaux et internationaux et le code d'éthique institutionnel et vérifiez si une revue des pratiques (par exemple par un comité d'éthique) est requise pour ce qui concerne la collecte de données dans le cadre du projet de recherche.

Traitement et analyse des données

Comment et avec quels moyens seront traitées les données ?

Stockage et sauvegarde des données pendant le processus de recherche

Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

Recommandations:

- Décrire l'endroit où les données seront stockées et sauvegardées au cours du processus de recherche et la fréquence à laquelle la sauvegarde sera effectuée. Il est recommandé de stocker les données dans au moins deux lieux distincts.
- Privilégier l'utilisation de systèmes de stockage robustes, avec sauvegarde automatique, tels que ceux fournis par les services informatiques de l'institution d'origine. Le stockage des données sur des ordinateurs portables, des disques durs externes, ou des périphériques de stockage tels que des clés USB n'est pas recommandé.
- Expliquez comment les données seront récupérées en cas d'incident.
- Expliquer qui aura accès aux données au cours du processus de recherche et comment l'accès aux données est contrôlé, en particulier dans le cadre de recherches menées en collaboration.
- Tenez compte de la protection des données, en particulier si vos données sont sensibles (par exemple données à caractère personnel, politiquement sensibles des informations ou secrets commerciaux). Décrire les principaux risques et la façon dont ils seront gérés.
- Expliquer quelle politique institutionnelle de protection des données institutionnelles est mis en œuvre.

Partage des données et conservation à long terme

Comment les données seront-elles partagées ?

Recommandations:

- Détailler le format des données : la manière selon laquelle les données sont codées pour le stockage, généralement reflétée par l'extension du nom de fichier (par exemple PDF, XLS, DOC, TXT, ou RDF).
- Justifier l'utilisation de certains formats. Par exemple, les choix d'un format peuvent être guidés par l'expertise du personnel de l'organisme, ou par une préférence pour les formats ouverts, par les standards de format acceptés par les entrepôts de données, par l'usage largement répandu dans une communauté de recherche ou par le logiciel ou l'équipement qui sera utilisé.
- Privilégier les formats standards et ouverts car ils facilitent le partage et la réutilisation à long terme des données (plusieurs catalogues fournissent des listes de ces « formats préférés »).
- Donnez des détails sur les volumes (qui peuvent être exprimés en espace de stockage requis (octets), et/ou en quantités d'objets, de fichiers, de lignes, et colonnes).
- Expliquez comment les données pourront être retrouvées et partagées (par exemple, par dépôt dans un entrepôt de données de confiance, indexé dans un catalogue, par l'utilisation d'un service de données sécurisé, par traitement direct des demandes de données, ou l'utilisation de tout autre mécanisme).
- Expliquez à quel moment les données seront rendues disponibles. Indiquer les délais de publication prévus. Expliquez si une utilisation exclusive des données est revendiquée et, dans l'affirmative, pour quelle raison et pour combien de temps. Indiquez si le partage des données sera différé ou limité, par exemple pour des raisons de publication, pour protéger la propriété intellectuelle, ou dépôt de brevets.
- Indiquez qui pourra utiliser les données. S'il s'avère nécessaire de restreindre l'accès pour certaines communautés ou

d'imposer un accord pour le partage de données, expliquez comment et pourquoi. Expliquez les mesures qui seront prises pour dépasser ou minimiser ces restrictions.

- Décrire les utilisations (et/ou les utilisateurs) prévisibles des données dans un cadre de recherche.
- Indiquez où les données seront déposées. Si aucun entrepôt reconnu n'est proposé, démontrer dans le DMP que les données pourront être prises en charge efficacement au-delà de la durée de financement du projet. Il est recommandé de démontrer que les politiques des entrepôts et les procédures de dépôts (y compris les normes de métadonnées, et coûts mis en œuvre) ont été vérifiés.
- Indiquez si les utilisateurs potentiels ont besoin d'outils spécifiques pour l'accès et la (ré)utilisation des données. Tenez compte de la durée de vie des logiciels nécessaires pour accéder aux données.
- Indiquez si les données seront partagées via un entrepôt, si les demandes d'accès seront traitées en direct, ou si un autre mécanisme sera utilisé ?

Comment les données seront-elles conservées à long terme ?

Recommandations:

- Définissez le plan de préservation des données et fournir l'information sur la durée d'archivage pérenne des données.
- Indiquez quelles données ne doivent pas être divulguées ou doivent être détruites pour des raisons contractuelles, légales, ou réglementaires.
- Indiquez comment il sera décidé quelles données garder. Décrire les données qui seront à préserver à long terme.