

PACEA (UMR 5199 CNRS - Université de Bordeaux) : PACEA - Modèle de DMP

SECTION-1 : DESCRIPTION ET COLLECTE OU REUTILISATION D'UN JEU DE DONNEES

1. Réutilisation de données existantes (en dehors du projet)

Préciser le nom du jeu de données, l'origine (brève description) , l'identifiant pérenne, la licence. Y a-t-il des restrictions à la réutilisation des données fournies par des tiers ?

2. Description du jeu de données

Faire une brève description du jeu de données (contenu, contexte de production).

3. Format du jeu de données

- AVI
- WAVE
- JPG
- PDF
- CSV
- JSON
- XML
- DICON
- TIF
- Autre (précisez la liste des formats utilisés dans "Informations supplémentaires")

Recommandations:

Il est fortement recommandé d'employer des formats ouverts (CSV, TXT, XML, JSON, etc...).

4. Volume des données exprimé en espace de stockage requis (Ko, Mo, Go, To, Po, Eo)

5. Volume des données exprimé en quantités d'objets, de fichiers (nombre d'objets, nombre de lignes, nombre de colonnes)

6. Méthodes de collecte des données

- Tachéomètre
- Appareil photo
- Carnet de terrain papier
- GPS portatif, relevé manuel
- Scanner
- Autre (précisez laquelle dans "Informations supplémentaires")
- Saisie manuelle (tableur, portail web)
- Drone (lidar, photographie, photogrammétrie, etc...)

SECTION-2 : DOCUMENTATION ET QUALITE DU JEU DE DONNEES

1. Standard(s) de métadonnées utilisé(s) pour le jeu de données

Préciser lequel

- Dublin Core
- Autre (précisez la liste des métadonnées dans "Informations supplémentaires")
- Inspire
- DataCite

Recommandations:

Les données déposées dans un entrepôt certifié seront décrites en utilisant un standard de métadonnées. Celui-ci dépendra du type de données et de l'entrepôt choisi (ex. pour Zenodo, ce sera Datacite).

2. Convention de nommage du jeu de données

Indiquer la convention de nommage des fichiers/répertoires.

Exemple de réponse:

Si l'analyse a fait l'objet d'une demande spécifique via le portail Web "GATE", le nom des répertoires est généré automatiquement. Un code est créé à partir des données saisies et c'est le code qui sera utilisé pour nommer les répertoires.

Ex. de code : PL3_AN4_OB82_PR158_SI126_PE1_DA20191127100347

- PL3 => PLateforme n°3 (Microtomodensitométrie)
- AN4 => ANalyse n°4 (Scan3D)
- OB82 => OBjet n°82
- PR158 => PROjet n°158
- SI126 => Site de fouille n°126
- PE1 => PErsonne qui a fait la demande n°1
- DA201911... => Date de la demande AAAAMMJJHHMMSS

Recommandations:

Les données issues de certaines analyses (vidéo, image 3D, photo, radio, etc...), qui ne peuvent pas être stockées dans des bases de données, sont déposées dans des espaces de stockages dédiés.

Pensez à une convention de nommage. Elle permettra d'une part de savoir à quoi correspondent les données, mais aussi de les retrouver rapidement par quiconque. Une convention de nommage évitera de refaire des analyses qui existent déjà, réduisant ainsi les coûts et les pertes de temps inutiles.

3. Contrôle de version du jeu de données

Comment est-il envisagé de suivre les versions des jeux de données ?

Recommandations:

Si vous envisagez d'utiliser un identifiant unique et pérenne pour ce jeu de données, en cas de mise à jour des données, pensez à créer un nouvel identifiant.

4. Structuration et organisation des données dans une base de données

- Autre (précisez dans "Informations supplémentaires")
- BdD relationnelle libre (PostgreSQL/ MySQL)
- BdD relationnelle non libre (Filemaker/Access)

Recommandations:

Pour les données tabulaires, pouvant être stockées dans des bases de données, la convention de nommage suivra les bonnes pratiques de modélisation d'une base de données assurée par la personne en charge des bases de données de PACEA.

Une base de données sera toujours accompagnée d'un dictionnaire des données et d'un modèle physique des données. Celui-ci expliquera le lien entre chaque table et décrira les métadonnées des colonnes (type, format, etc...).

5. Vocabulaire contrôlé

Est-ce que les variables sont décrites à l'aide d'un vocabulaire contrôlé (si oui, préciser lequel) ?

- Non
- Oui (précisez lequel dans "Informations supplémentaires")

6. Contrôle et qualité du jeu de données

Préciser quelles méthodes d'assurance qualité ont été mises en place pour la qualité scientifique des données.

- Autre méthode (à préciser dans "Informations supplémentaires")
- Qualification des données
- Processus de calibration
- Validation des données
- Validation de saisie des données
- Une revue des données par des pairs
- Répétition des échantillons ou des mesures

Recommandations:

La validation des données : Les données sont comparées à des modèles statistiques par exemple (humainement ou numériquement)

La validation de la saisie des données : Le format de la donnée saisie est vérifié par le logiciel par rapport au format attendu (valeur bornée ? valeur négative possible ? entier ? etc ...)

La qualification des données : Un code qualité attribué à chaque valeur par un expert (1: Hors protocole, 2: Valeur bonne, 3: Mesure douteuse, etc ...)

Revue des données par des pairs : Equivaut à la qualification des données, mais sans attribuer de codes qualité

SECTION-3 : EXIGENCES LEGALES ET ETHIQUES, CODE DE CONDUITE

1. Données à caractère personnel contenues dans ce jeu de données.

Y a-t-il des données personnelles dans le jeu de données ?

Si oui, expliquer dans "Informations supplémentaires" les démarches / mesures qui ont été prises pour se conformer aux dispositions légales et assurer la sécurité des données. (Demande de consentement éclairé par écrit aux personnes concernées. Anonymisation, pseudonymisation des données. Accès aux données pour les utilisateurs autorisés)

- Oui (précisez lesquelles dans "Informations supplémentaires")
- Non

Recommandations:

Si le jeu de données comprend des données à caractère personnel (nom, prénom ou coordonnées d'un individu, ou des informations permettant d'identifier une personne de manière indirecte), vous devez respecter les consignes du RGPD (<https://www.cnil.fr/fr/rgpd-par-ou-commencer>). Pour toute question relative au RGPD à PACEA, écrivez à pacea-rgpd@services.cnrs.fr

La pseudonymisation est réversible, contrairement à l'anonymisation.

2. Données sensibles / confidentielles (politiquement sensibles, informations ou secrets commerciaux)

Y a-t-il des données sensibles dans le jeu de données ?

Si oui, dans "Informations supplémentaires", décrire les principaux risques associés à ces données et la façon dont ils seront gérés.

- Oui (précisez lesquelles dans "Informations supplémentaires")
- Non

3. Titularité et droits de propriété intellectuelle du jeu de données

Préciser qui sera le propriétaire des données et qui aura le contrôle de l'accès aux données. Comment ces droits sont-ils définis dans l'accord de consortium ?

Quelles seront les conditions d'accès aux données : libre ou restreinte ? Quelles seront les restrictions et pourquoi ?

Y a-t-il des droits de propriété intellectuelle qui s'appliquent ?

4. Ethique et code de déontologie sur le jeu de données

Y a-t-il une charte de déontologie ?

Si oui, préciser laquelle (organisme)

Si non, expliquer l'impact que cela peut avoir côté stockage, diffusion, durée de conservation. Préciser s'il est envisagé d'adopter un code de conduite ou d'éthique .

SECTION-4 : STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

1. Comment sera stocké le jeu de données pendant le projet ?

- Disque dur interne du poste de travail
- Serveur de base de données
- Serveur de fichiers de PACEA
- Disque dur externe
- Cloud (précisez lequel dans "Informations supplémentaires")
- Autre (précisez lequel dans "Informations supplémentaires")

Recommandations:

Privilégier les supports recommandés par les tutelles qui sont sécurisés et sauvegardés. Pour plus d'information, veuillez vous rapprocher de votre service informatique ou contacter le référent DMP de PACEA (pacea-dmp@services.cnrs.fr).

2. Sécurité des supports externes utilisés pour transporter les données ?

Si les supports ne sont ni chiffrés ni doublés, préciser dans "Informations supplémentaires" comment la sécurité des données sera assurée en cas de perte ou de vol des supports.

- Supports chiffrés
- Supports doublés

3. Sauvegarde des données

Si les données sont gérées par PACEA, enregistrer la réponse par défaut ci-dessous. Sinon expliquer les conditions de sauvegarde et de restauration en modifiant le texte par défaut.

4. Accès au jeu de données (en cas de collaboration)

Si les données sont accessibles comme expliqué ci-dessous, enregistrer la réponse par défaut. Dans le cas contraire, adapter le texte aux modalités d'accès.

SECTION-5 : PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

1. Restriction d'accès sur le jeu de données - Y-a-t-il un embargo ?

Si oui, préciser les raisons de l'embargo, sa durée, la date à laquelle les données seront disponibles.

- Non
- Oui (précisez la durée et les raisons dans "Informations supplémentaires")

2. Restriction d'accès sur le jeu de données - Y-a-t-il une exclusivité ?

Si oui, préciser la raison, pendant combien de temps. Durant cette exclusivité, le partage des données sera-t-il différé, limité ? Qui pourra utiliser les données pendant cette exclusivité ?

- Non
- Oui (précisez la durée et les raisons dans "Informations supplémentaires")

3. Conditions d'accès au jeu de données - L'accès aux données est-il ouvert, fermé ou restreint ?

Si un accord est requis pour partager les données, expliquer comment et pourquoi. L'accès aux données après l'embargo ou l'utilisation exclusive se fera-t-il par formulaire de demande, accès libre sans filtre, convention ?

- Ouvert
- Restreint (précisez pourquoi et comment dans "Informations complémentaires")
- Fermé (précisez la raison dans "Informations complémentaires")

Recommandations:

Un accès filtré ou restreint :

Cet accès peut se faire par exemple via un formulaire de demande de données à remplir en ligne pour savoir qui va utiliser ces données et en connaître l'usage.

Cela peut être aussi un accès libre limité à un certain niveau de données comme un accès libre uniquement à des données qualifiées "bonnes", pour les autres données il faut un accès autorisé

4. Partage du jeu de données

Quel que soit le moyen choisi, préciser l'URL du site web dans "Informations supplémentaires"

- Entrepôt de données de confiance (précisez lequel dans "Informations supplémentaires")
- Autre moyen de partage (précisez lequel dans "Informations supplémentaires")
- Portail web dédié (précisez l'URL dans "Informations supplémentaires")
- Portail d'accès aux données de PACEA

Recommandations:

Privilégiez des entrepôts de données de confiance qui distribuent et gèrent des identifiants pérennes (ArkID, DOI, autre) ou certifiés avec un certificat CoreTrustSeal par exemple (voir liste : <https://www.coretrustseal.org/why-certification/certified-repositories/>).

Les critères de sélection des entrepôts de données de confiance sont disponibles ici : <https://www.ouvrirlascience.fr/science-europe-guide-pratique-pour-une-harmonisation-internationale-de-la-gestion-des-donnees-de-recherche/>

Pour aller plus loin ou pour trouver un entrepôt dans votre discipline, voir le site re3data.org

5. Attribution d'un identifiant unique et pérenne

Est-il envisagé d'attribuer un identifiant unique et pérenne, lequel (DOI, ArkID, autre) ?

- Non
- Oui (précisez lequel dans "Informations supplémentaires")

6. Licence pour l'accès et la réutilisation des données

Quelle que soit la réponse, préciser avec exactitude la licence associée aux données (cf. recommandations)

- Open Database License (ODbL)
- Licences dites « permissives »
- Licences « avec obligation de réciprocité »
- Licence homologuée
- Licence ouverte de réutilisation d'informations publiques
- Autre

Recommandations:

Licence établie par le Gouvernement : Licence Ouverte

Licence avec obligation de partage à l'identique : Open Database License (ODbL)

Licences dites « permissives » : « Berkeley Software Distribution License », « Apache », « CeCILL-B » et « Massachusetts Institute of Technology License », « Creative Commons »

Licences « avec obligation de réciprocité » : « Mozilla Public License », « GNU General Public License » et « CeCILL »

Pour en savoir plus : <https://www.data.gouv.fr/fr/licences>

7. Méthodes et outils nécessaires pour l'accès et la réutilisation des données

Y-a-t-il besoin d'outils spécifiques pour accéder aux données et les réutiliser ? Pendant combien de temps ces outils seront-ils maintenus ?

8. Potentiel de réutilisation du jeu de données

Quels sont les utilisations et utilisateurs prévus de ces données ?

9. Sélection et conservation des données du jeu de données

Quelles données seront à conserver à long terme et pour quelle raison ?

Si l'entrepôt choisi précédemment ne le permet pas, identifier une archive (CINES, HAL Archive ouverte, autre).

Quelles données devront être détruites, quelles en sont les raisons ?

Recommandations:

Attention un site web que vous avez développé vous-même dans le cadre de votre projet pour partager les données n'est pas considéré comme un moyen de conservation sur le très long terme tel un entrepôt de données de confiance comme le CINES par exemple, qui garantit la réutilisation des données sur plusieurs dizaines de décennies.

SECTION-6 : RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

1. Rôles et responsabilités dans la gestion des données

2. Rôles et responsabilités dans la gestion des données - Collaboration

Expliquer comment s'effectue la coordination des responsabilités de gestion des données entre partenaires

3. FAIR Data, les ressources allouées

- Comment les ressources nécessaires (temps, budget, etc...) à la préparation des données pour le partage/préservation des données ont été chiffrées ? Ressources nécessaires pour le concept FAIR Data, temps et coût ?
- Comment tous ces coûts sont couverts ?

