CEA Commissariat à l'énergie atomique et aux énergies alternatives: Plan de gestion des données - Modèle CEA (FR)

Suivi des versions du PGD

Historique des versions du plan de gestion de données

Indice	Date	Nature de l'évolution

Recommandations:

OPIDoR ne gère pas les versions des plans de gestion de données. A vous de sauvegarder le PGD, en tant que de besoin, avant modification importante.

Vous pouvez créer une copie du plan et ajouter un indice de version au nouveau document.

Informations sur le projet

Nom du projet*

Acronyme du projet

Objectifs du projet

Identifiant de l'appel à projets*

Financeur (Europe, ANR, organisme, industriel,...)

Référence de la convention de financement

Coordinateur et partenaires du projet

Contact du projet

Recommandations:

Nom, prénom et email

Affiliation et unité du contact

Date début du projet

Durée du projet

Commentaires

Recommandations:

Vous souhaitez ajouter des informations sur le projet

Jeux de données

Règles de nommage des jeux de données

Recommandations:

Définir des règles à utiliser au cours du projet pour nommer de façon claire et cohérente les jeux de données, les dossiers et les fichiers.

- Eviter les caractères spéciaux, accentués et les espaces vides

Il est conseillé d’inclure les dates associées.

Liste de jeux de données

Recommandations:

Mettre un identifiant pour chaque ensemble de données

Arborescence de classement

Recommandations:

Prévoir une arborescence de classement des données dès que possible afin de faciliter l’accès et le stockage au cours du projet, puis à terme l’archivage des données

Description des données

Identifiant et nom du jeu de données*

Recommandations:

Indiquer l'ensemble des identifiants et noms des jeux de données concernés

But et relation avec les objectifs du projet*

Recommandations:

Exposez le but de la collecte des données et la relation avec les objectifs du projet.

A détailler si besoin en fonction des différents jeux de données.

Nature des données*

Recommandations:

Images, code, texte,...

Formats des fichiers*

Recommandations:

Privilégier des formats ouverts ou largement utilisés afin de faciliter le partage, mais également pérennes pour une meilleure conservation et un archivage facilité.

Décrire la correspondance entre le format originel et ceux résultant de conversion.

Préciser si besoin en fonction des différents jeux de données.

Réutilisation de données existantes*

Recommandations:

Préciser la source des données réutilisées si elles sont critiques pour le projet de recherche.

Indiquer le lien vers les données réutilisées **UNIQUEMENT si elles sont sur un entrepôt ouvert.**

Méthode de production des données*

Recommandations:

Comment les données ont-elles été générées ?

Expériences, instrument de mesure, observation, compilation, simulation, ...

Préciser les protocoles expérimentaux, appareils, conditions de mesures ; la source des données réutilisées ; les logiciels utilisés (version) ; formules ; algorithmes ; ...

Support des données

Recommandations:

Le support doit être choisi en fonction de l'utilisation souhaitée de la donnée

Hébergement des données

Recommandations:

Matériel et lieu physique de stockage.

Recours à un prestataire

Logiciel spécifique

Volume estimé des données*

Recommandations:

Volume de données estimé

A réévaluer en cours de projet

Politique appliquée aux données du projet

Recommandations:

Politique des financeurs, des institutions, ...
Ex : politique d'Open Data, ...

Propriété des données

Recommandations:

Il est possible que le contrat ou l'accord de consortium qui encadre le projet ayant généré les données contienne des dispositions relatives au régime de ces données. A qui appartiennent les données produites ? Peuvent-elles, et à quelles conditions, être réutilisées ? Quel est le régime juridique applicable le cas échéant aux bases de données et à leur contenu ?

Utilité des données***

Recommandations:

Public cible envisagé (grand public, communauté scientifique, acteur privé, utilisation interne à l'organisme…)

Potentiel de réutilisation*

Recommandations:

Perspectives d'application ou de développement à partir des données

Principe général de diffusion*

Recommandations:

Règles qui s'appliquent à la majorité des données produites

Publications associées aux données

Recommandations:

Publications scientifiques, Data papers
Rapports scientifiques

Documentation associée

Recommandations:

Toute information nécessaire à l'intelligibilité des données à long terme : codes, abréviations, version des logiciels de lecture, …

Données FAIR - Faciles à trouver

Standards et formats des métadonnées*

Recommandations:

Préciser les standards et formats de métadonnées (données sur les données) utilisés pour décrire vos données.
Ils sont souvent spécifiques à votre discipline.
http://www.dcc.ac.uk/resources/metadata-standards
http://rd-alliance.github.io/metadata-directory/
S'il n'existe pas de standards qui conviennent dans votre discipline, expliquez comment et quelles métadonnées ont été créées.

Identifiant unique et permanent*****

Exemple de réponse:

Ex : DOI (DOI Digital Object Identifier),... souvent attribué par l'entrepôt sur lequel vous déposez vos données

Recommandations:

L'usage d'un identifiant unique et pérenne permet de garantir la pérennité du lien vers les données.

Conventions de nommage*

Recommandations:

Conventions de nommage utilisées pour les données.

Mots clés*****

Recommandations:

Des mots clés seront-ils fournis pour optimiser les possibilités de réutilisation ?

Numéros de version*****

Recommandations:

Approche envisagée pour un versioning clair (numéros de versions) des jeux de données

Données FAIR - Données accessibles

Données ouvertes disponibles*

Recommandations:

Préciser quelles données seront ouvertes et accessibles par des tiers externes.
Préciser le niveau d'ouverture externe :

- à des membres d'un consortium sur projet, aux seules fins de la réalisation du projet de R&D,
- à des membres d'une communauté scientifique internationale, à des fins de recherche,
- au grand public.

Si des données ne peuvent être ouvertes, préciser briévement pourquoi. **Pour rappel : cet outil est hébergé à l'extérieur du CEA.**

Outil pour lire ou réutiliser des données*

Recommandations:

Préciser s’il y a besoin d’un logiciel ou d’un outil spécifique pour lire, réutiliser, retraiter les données et s'il sera inclus.

Documentation sur le logiciel si nécessaire.

Accessibilité des données*

Recommandations:

Préciser comment les données seront accessibles : ex. entrepôt de données, …

Dépôt dans un entrepôt de données*

Recommandations:

Dépôt dans une base de l’organisme, dans des bases externes **à condition que cela soit autorisé pour les données du projet** : base disciplinaire, …

Si les données sont stockées dans un entrepôt externe, attention à bien vérifier qu’il soit pérenne, que vous êtes autorisé à les y déposer et que l’entité en charge de cet entrepôt ne s’approprie pas la propriété des données.

Lien vers l’entrepôt **uniquement si cet entrepôt est accessible à tous sur internet. Pour rappel : cet outil est hébergé sur un serveur extérieur au CEA.**

Exemple de réponse:

Ex : https://zenodo.org/

Accès autorisé*

Recommandations:

Préciser comment les personnes autorisées auront accès aux données en cas de restrictions d’accès

Données FAIR - Données interopérables

Normes, vocabulaires, méthodologies pour les données et les métadonnées*

Recommandations:

Indiquer si les données et les métadonnées associées suivent des normes, vocabulaires et méthodologies permettant l’interopérabilité des données

Interopérabilité inter-disciplinaire*****

Recommandations:

Préciser si le vocabulaire utilisé pour tous les types de données de votre data set permet l’interopérabilité entre disciplines.

Sinon, préciser si une correspondance sera fournie.

Données FAIR - Réutilisation des données

Type de Licence*

Recommandations:

Spécifier les licences choisies pour préciser les conditions de partage et de réutilisation des données diffusées et les éventuelles contreparties économiques de cette réutilisation.

- Licences spécifiques déterminant les conditions, notamment financières, de l’accès aux données
- Licences de réutilisation pour les administrations : https://www.data.gouv.fr/fr/licences
- Creative Commons CC4.0 : https://creativecommons.org/choose/
- Open Data Commons : http://opendatacommons.org/

Date de mise à disposition des données*

Recommandations:

Le cas échéant, préciser la date de fin de mise à disposition.

Préciser si certaines données font l’objet d’un embargo, pourquoi et pour combien de temps (ex. publication, brevets,…).

Accès à un tiers*

Recommandations:

Préciser si les données produites et/ou utilisées dans le projet sont utilisables par un tiers, en particulier après la fin du projet

Accès restreint : exception au principe général de diffusion *

Recommandations:

Préciser succinctement la nature du risque ou de la contrainte justifiant la restriction d’accès (éthique, données personnelles, propriété intellectuelle, commerciale, données privées, sécuritaires, …)

Processus d’assurance qualité*

Recommandations:

Préciser les processus d'assurance qualité de gestion des données utilisés

Durée d'ouverture des données à la réutilisation*

Recommandations:

Durée pendant laquelle les données vont être réutilisables

Ressources

Coûts pour rendre les données FAIR et moyens pour les couvrir*

Recommandations:

Précisez les coûts associés à la mise en œuvre du plan de gestion de données.

Indiquer comment ces coûts seront pris en charge.

Responsable de la gestion des données au cours du projet de recherche*

Recommandations:

Nom, Prénom et email

Responsabilités des partenaires

Recommandations:

Décrire, s'il y a lieu, les responsabilités des autres partenaires du projet dans la gestion des données

Valeur potentielle de la conservation à long terme*

Coûts de la conservation à long terme*

Recommandations:

Attention à bien évaluer les coûts entraînés par la conservation à long terme. Ceux-ci peuvent être disproportionnés par rapport à la valeur potentielle de préservation. Il faut démontrer que le bénéfice attendu par la conservation est supérieur au coût de préservation

Archivage

Devenir des données à l'issue du projet

Recommandations:

Indiquer ce qui est prévu au terme de la phase active du projet, notamment concernant la dissémination et la valorisation des résultats.

Sélection des données*

Recommandations:

Après une analyse de risques permettant de déterminer l'intérêt de la conservation des données (valeur scientifique, stratégique, juridique, patrimoniale), indiquer, s'il y en a, celles qui devront être archivées.

Critères principaux pour sélectionner les données à archiver:

- Obligation légale ou contractuelle de conservation de certaines données pendant une durée déterminée en les rendant accessibles sur demande dans certains cas.
- Intérêt de conserver l'ensemble des données produites et de la capacité des ressources informatiques pour les accueillir. L'échantillonnage des données pour réduire le volume à conserver peut être envisagé si une sélection de données est suffisamment représentative de l'ensemble.
- Sélection de données dont les caractéristiques (format normalisé, standard répandu) faciliteront l'exploitation ultérieure sans recourir à des outils logiciels spécifiques.
- Transparence des recherches et traçabilité. Un éditeur peut demander de lier les données à la publication. Nous vous conseillons d'éviter de les déposer sur le site de l'éditeur.
- Données d'intérêt non reproductibles ou trop coûteuses à reproduire.
- Enjeux politiques et patrimoniaux : capitalisation des données.

Le type de données à archiver sera également à apprécier en fonction du projet et des objectifs de l'archivage de ces données.

Les durées de conservation seront définies conjointement par la maîtrise d'ouvrage, le service juridique et l'archiviste référent. Elles seront reportées dans un tableau de gestion.

Attention aux coûts liés à l'archivage des données !

Mettre les données dans un format recommandé par la plateforme d'archivage (ex : au CINES http://facile.cines.fr/) si la durée de conservation est supérieure à 5 ans

Volume final des données archivées

Recommandations:

Estimation

Durée de conservation préconisée*

Recommandations:

Dépend non seulement des exigences légales et réglementaires, de la valeur patrimoniale des données, mais également des préconisations émises lors de l’analyse de risques

Plateforme d’archivage*

Documentation papier associée

Recommandations:

Référence de cette documentation

Sécurité des données

Analyse de risque

Recommandations:

Préciser le type d’analyse de risques réalisée si besoin (Consulter votre ASSI si besoin)

Garantie des données*

Recommandations:

Dispositions mises en place pour la sécurité des informations (y compris récupération des données ; stockage et transfert des données sécurisés ; garantie de confidentialité, d’intégrité et de traçabilité)

Pour les données de recherche non relatives aux publications, le propriétaire des données qui les ouvrirait de manière libre et gratuite à tous, ne donne aucune garantie quant à l’utilisation des données. En particulier, en cas d’exploitation commerciale, il dégage toute responsabilité en cas d’accident, incident de toute sorte, ou perte d’exploitation commerciale pouvant résulter de l’utilisation des données.

Identification des données sensibles. Attention : cette partie n’a pas vocation à être diffusée

Recommandations:

Données protégées ou à risque, données nominatives, données stratégiques, ...
Pour rappel : l'outil que vous utilisez est hébergé sur des serveurs extérieurs au CEA.

Sécurité de la conservation à long terme*

Recommandations:

Vérifier que les données sont stockées en toute sécurité dans un entrepôt certifié pour l’organisation et la conservation à long terme

Aspects éthiques

Questions éthiques et juridiques

Recommandations:

Impact sur le partage et la réutilisation des données.
Le consentement éclairé de la personne pour le partage de ses données personnelles, leur réutilisation, et leur conservation à long terme est-il pris en compte dans les questionnaires traitant des données personnelles?

Envisager:

- l'anonymisation des données personnelles pour la préservation et/ou le partage
- la pseudonymisation
- le chiffrement des données

Indiquer la bonne prise en compte des questions éthiques

Vérifier si une revue des pratiques (par ex. un comité d'éthique) est requis pour la collecte des données

Autre

Autres procédures de gestion des données utilisées*

Recommandations:

Indiquer si d’autres procédures de gestion des données sont utilisées (national/financeur/sectorielle/…)

Informations complémentaires

Recommandations:

Toute information complémentaire pouvant être utile à la bonne gestion de données