

Université Gustave Eiffel: Université Gustave Eiffel - Modèle de PGD

1. Présentation des jeux de données collectés

Jeux de données collectés

Recommandations:

Présenter de manière synthétique les données qui seront collectées au cours du projet

2. Concevoir la collecte

Responsabilité des données

Recommandations:

Dans la grande majorité des cas, les données ne sont pas couvertes par le droit de la propriété intellectuelle. L'établissement qui produit les données ou qui est à l'initiative de la collecte est responsable de ces données. Dans un projet, certains jeux de données peuvent être par un ou plusieurs partenaires (co-responsabilité). La responsabilité des données est généralement stipulée dans l'accord de consortium.

Lorsque les jeux de données sont préexistants au projet, ouverts ou non, il convient de respecter quelques bonnes pratiques. Consultez la page "[Utiliser des données](#)" de l'intranet pour en savoir plus.

Expliquez les différents cas de responsabilité des jeux de données du projet.

Exemple de réponse:

- Université Gustave Eiffel
- Les données sont sous la coresponsabilité de l'Université Gustave Eiffel et des partenaires du projet. L'article 7 de l'accord de consortium du projet traite la question de la propriété des résultats.
- Les données produites par XXX disponibles en open data sous la licence de réutilisation Open Database Licence seront réutilisées pour le projet.
- Certaines données appartenant à XXX et partagées avec l'Université Gustave Eiffel sous les termes de la convention n°XXX seront réutilisées.

Mode de collecte

Recommandations:

Les données peuvent être collectées par :

- des capteurs : préciser quels sont les capteurs
- des enquêtes : préciser l'outil utilisé ([des plateformes d'enquêtes sont mises à disposition par la DGDIN](#))
- un logiciel (préciser le nom du logiciel et la version).

Les données peuvent également être disponibles en open data ou partagées par un partenaire. Dans ce dernier cas de figure, expliquez comment les données seront mises à votre disposition.

Préciser ici l'ensemble des moyens de collecte des données du projet.

Exemple de réponse:

Les données de l'enquête seront collectées grâce à l'outil [Limesurvey](#)

Nature des données

Recommandations:

Il peut s'agir d'images, de vidéos, de cartes, de données textuelles, chiffrées, etc.

Exemple de réponse:

Données cartographiques 2D/3D/HD utilisées pour les algorithmes de calcul de la position précise.

Règlementation et éthique

Recommandations:

Préciser dans cette partie si les données sont protégées par une réglementation (RGPD, droit d'auteur, secret défini par la loi : médical, statistique, industriel, défense, judiciaire). Pour en savoir plus sur le cadre réglementaire, nous vous invitons à consulter la page "[Maîtriser le cadre juridique](#)" sur l'intranet.

Attention ! Si vous êtes rattaché à l'Université Gustave Eiffel et que vous collectez/utilisez des données personnelles, vous devez compléter la fiche d'inscription au registre disponible sur [l'intranet](#). Cette fiche doit ensuite être transmise au service protection des données personnelles via cette adresse : protectiondesdonnees-dpo@univ-eiffel.fr

Exemple de réponse:

Les données collectées au cours du projet comportent des données à caractère personnel. Ces données sont enregistrées dans le registre des traitements de l'Université Gustave Eiffel. Pour cela, la fiche registre n°XXX a été remplie et transmise à la déléguée à la protection des données de l'université.

Format du jeu de données (si numérique)

Exemple de réponse:

- TXT (ASCII)

- CSV
- HTML
- ODT
- TIFF
- PNG
- PDF
- XML
- Le format propriétaire des récepteurs GNSS peut nécessiter l'utilisation de convertisseurs mais certains logiciels open source peuvent ouvrir certains des formats propriétaires (par exemple RTKLib, GoGPS) et permettre la conversion du format propriétaire en formats standards (RINEX, CSV).

Recommandations:

L'utilisation de formats ouverts est à privilégier afin de faciliter l'ouverture et/ou la réutilisation des données. Le [Référentiel Général d'Interopérabilité de l'état](#) et le [CINES](#) précisent les standards retenus.

De manière générale, choisir des formats standards et ouverts facilite la préservation à long terme des données.

Même si le format n'est pas standard et ouvert, le fait qu'il existe une documentation qui décrit les spécifications du format, ou que le format soit auto-documenté (par exemple, langage balisé ou usage de métadonnées...) facilite également la préservation à long terme.

Si vous utilisez différents formats, veuillez préciser les jeux de données associés aux formats.

Format des données (si non numériques)

Exemple de réponse:

Notes, fiches, photographies, bobines de film ou VHS

Langue (des données et de leur documentation)

Exemple de réponse:

- Français
- La langue à employer pour documenter les données est l'anglais

3. Documentation des données

Convention de nommage

Exemple de réponse:

Le nom des fichiers sera constitué de la façon suivante :

<Capteur>_<JJ_MM_AAAA>_<numero-version>.format . JJ_MM_AAAA correspondra à la date d'actualisation du fichier.

Recommandations:

La mise en place d'une convention de nommage garantit aux membres de l'équipe et aux utilisateurs des données de trouver facilement le fichier de données recherché. En général, il est conseillé de faire apparaître dans le nom de fichier la date, le type de données ou description brève des données, le numéro de version. Si les règles de nommage ne sont pas facilement compréhensibles, expliquer bien le raisonnement.

Métadonnées

Recommandations:

Les métadonnées (données décrivant les données et les jeux de données) sont également très utiles pour rendre les données faciles à trouver et interopérables. Il existe des descripteurs standards (comme le Dublin Core, MARC, TEI, etc.) généralistes ou spécifiques à une discipline. Pour s'informer sur les standards disciplinaires : <https://fairsharing.org/> , <https://www.dcc.ac.uk/guidance/standards/metadata> ou <http://rd-alliance.github.io/metadata-directory/> .

Certaines métadonnées peuvent être créées automatiquement par l'instrument de capture de la donnée.

Lors du dépôt des données sur un entrepôt de données (dans le cadre de l'open data), un ensemble de métadonnées devront être complétées (métadonnées de citation et parfois disciplinaires).

Exemple de réponse:

- Les enquêtes sont réalisées selon la norme DDI
- Les jeux de données seront décrit grâce à des métadonnées Dublin Core

Documentation

Exemple de réponse:

La documentation associée aux données d'enquête comprend le questionnaire et le manuel des enquêteurs.

Recommandations:

Les éléments documentaires favorisent la compréhension et in fine, la réutilisation des données. Il s'agit par exemple de décrire les raisons de la collecte (objectifs du projet de recherche) et les moyens mis en œuvre (méthodologie et instruments...).

En parallèle, un jeu de données peut être accompagné d'une documentation ou notice spécifique. Par exemple, une documentation expliquant la structure des fichiers, les unités des valeurs stockées, des éléments de description des valeurs (1 = présent, 0 = absent), ... La bonne pratique est d'utiliser un format ouvert (CSV, TXT, etc.)

4. Stocker, utiliser, traiter les données

Estimation du volume des jeux de données numériques

Exemple de réponse:

- Environ 250 Go
- 5 To

Recommandations:

Cette estimation permet d'anticiper les besoins de stockage.

Stockage du jeu de données (numérique) - Infrastructures de stockage

Exemple de réponse:

- Les données brutes sont stockées sur le disque dur des outils d'acquisition
- Les données post-traitées sont stockées sur l'espace réseau du laboratoire : lien vers l'espace réseau)

Recommandations:

Le stockage de données confidentielles, personnelles ou sensibles sur des plateformes Cloud (type Dropbox, google drive...), est interdit par la PSSI, pour des raisons de sécurité. Voir la Politique de Sécurité des Systèmes d'Information de l'État portée par la circulaire du Premier ministre n° 5725/SG du 17 juillet 2014. <https://www.legifrance.gouv.fr/circulaire/id/38641>

Le stockage permanent de données sur des ordinateurs personnels, sur des clés USB ou des disques durs externes n'est pas recommandé, également pour des raisons de sécurité et de pérennité.

Description des conditions de stockage des données non numériques

Recommandations:

Selon le support des données (papiers...) décrivez les lieux, comment sont stockées les données et les mesures de sécurité prises pour contrôler l'accès si besoin.

Sauvegarde du jeu de données (numérique)

Recommandations:

Au-delà du stockage, une solution de sauvegarde peut s'avérer utile, notamment en cas d'incidents de toute nature.

Exemple de réponse:

- Un double stockage est envisagé : 1) sous la responsabilité du responsable de WP qui s'assurera d'une sauvegarde périodique, 2) sur la plateforme omnispaces dédiée au projet.
- Disque dur de sauvegarde externe stocké dans le bureau du responsable des données (bâtiment XXX 2ème étage)

Contrôle des accès pendant le projet

Recommandations:

L'accès aux données peut être limité, notamment si les données sont protégées par une réglementation. Il convient alors d'utiliser une solution informatique permettant de limiter l'accès : mot de passe, chiffrement, espace réseau limité à certaines personnes, etc.

Consultez la ressource "[Création d'une archive chiffrée](#)" sur l'intranet pour savoir comment chiffrer des données.

Exemple de réponse:

- Les jeux de données protégés par le secret industriel seront en accès restreint. Seules les personnes habilitées auront des droits d'accès à l'espace réseau sur lequel les données sont stockées
- L'accès aux données déposées sur le cloud est limité par un mot de passe

Partage pendant le projet

Recommandations:

Le partage des données durant la vie du projet se déroule conformément aux accords de propriété évoqués dans la première partie

Le partage de données confidentielles ou sensibles sur des plateformes Cloud (type Dropbox...) est interdit par la PSSI, pour des raisons de sécurité.

Pour trouver la solution la plus adaptée (accès réseau FTP, Cloud Eiffel, Filesender, envoi de clé USB ou disque dur externe, site web spécifique...), consultez [l'intranet](#) ou contactez votre correspondant informatique ou la DGDIN.

Exemple de réponse:

Le jeu de données est accessible à toute l'équipe projet via la plateforme mise à disposition par l'IREX.

Traçabilité et assurance qualité

Recommandations:

Garantir la traçabilité et la qualité des données est un enjeu majeur en termes d'intégrité scientifique (respect des valeurs de fiabilité, de rigueur, de transparence, voire d'honnêteté)

Exemple de réponse:

La traçabilité des traitements apportés aux données est assurée par l'usage régulier d'un cahier de laboratoire. La qualité des données est garantie par différentes dispositions de prévention (suivi métrologique des instruments, connaissance des incertitudes, etc.) et de contrôle (détection de valeurs aberrantes ou manquantes, harmonisation des protocoles expérimentaux,

partage des bases des données). Au niveau des données de spectrométrie de masse, la qualité et la conformité des données seront contrôlées par les divers contrôles qualité et les calibrations des appareils réalisés régulièrement.

5. Trier et conserver les données

Durée de conservation

Recommandations:

Toutes les données n'ont pas d'intérêt à être conservées à l'issue du projet, certaines doivent être détruites si elles n'ont pas d'intérêt à long terme et certaines sont conservées pour leur valeur informationnelle, leur potentiel de réutilisation, leur valeur de preuve, leurs valeurs scientifique ou historique.

Le pôle archives de l'Université est à votre disposition pour déterminer le sort de vos données.

La durée d'utilité administrative (DUA) correspond au temps pendant lequel les données doivent être conservées, soit en vertu des prescriptions réglementaires, soit parce qu'ils restent nécessaires et utiles à la bonne marche du ou des laboratoires. Cette durée commence en général à partir de la clôture d'un dossier ou du projet. Le tableau de gestion des archives de l'Université est un bon outil pour déterminer la DUA des données. (<https://intranet.univ-eiffel.fr/luniversite/affaires-juridiques-et-institutionnelles/archivage/documents-utiles-et-fiches-pratiques>)

Exemple de réponse:

- Les données seront détruites à l'issue de la durée d'utilité administrative de 3 ans
- Le fichier de correspondance entre les noms des personnes et leur identifiant sera supprimé à l'issue du projet

Infrastructures d'archivage

Recommandations:

Préciser comment les données seront archivées. Sont-elles conservées sur des espaces de stockage du laboratoire ? Sont-elles conservées sur l'entrepôt de données de l'université en mode « non publié » ? Sont-elles archivées sur les [infrastructures du Centre Informatique National de l'Enseignement Supérieur](#) (CINES) avec l'aide de la Très Grande Infrastructure de Recherche Huma-Num (<https://www.huma-num.fr/services-et-outils/archiver>) ?

6. Publier les données (open data)

Nom(s) de(s) la personne(s) chargée(s) de l'ouverture des données

Dépôt sur un entrepôt de données

Exemple de réponse:

- Les données du projet seront déposées sur [GéOSUNA](#), l'infrastructure de données spatiales
- Les données du projet seront disponibles sur l'entrepôt institutionnel de l'Université Gustave Eiffel (<https://data.univ-gustave-eiffel.fr/>).

Recommandations:

Pour ouvrir les données de recherche, le dépôt dans un entrepôt de données est fortement recommandé.

L'Université Gustave Eiffel met à votre disposition un entrepôt institutionnel pour déposer vos données, [data.univ-gustave-eiffel](#).

Il existe un grand nombre d'entrepôts, qu'ils soient institutionnels, thématiques ou généraux. Ces entrepôts sont référencés par des répertoires :

- [Cat OPIDOR](#) (entrepôts français)
- [Re3data](#)

Attribution d'un identifiant pérenne

Recommandations:

Au moment du dépôt, il est important qu'un identifiant pérenne soit attribué aux données pour faciliter son accès et sa citation dans une publication par exemple. Vérifiez bien que l'entrepôt sur lequel sont déposées les données attribue un identifiant (DOI, Handle, par exemple).

L'Université Gustave Eiffel est en capacité d'attribuer des DOI.

L'entrepôt de l'université fournit automatiquement un DOI.

Exemple de réponse:

DOI attribué automatiquement sur l'entrepôt institutionnel de l'Université Gustave Eiffel

Date de publication

Recommandations:

L'ouverture des données peut être différée pour des raisons d'exploitation et de publication par exemple.

Exemple de réponse:

Les données seront déposées dans les six mois après la fin du projet.

Licence de réutilisation

Exemple de réponse:

Les données seront disponibles sous la licence ouverte d'Etalab

Recommandations:

Il est important d'attribuer une licence de réutilisation aux données qui seront ouvertes. Elle permettra aux réutilisateurs des données de connaître les conditions à la réutilisation des données. L'Université Gustave Eiffel invite les personnes qui publient des données à les ouvrir sous l'une des deux licences prévues par le [Décret n°2017-638 du 27 avril 2017 - art. 1](#) pouvant être utilisées par les acteurs publics : la licence ouverte ou la licence ODbL. Pour en savoir plus, consultez la page intranet « [Maîtriser le cadre juridique](#) »

Non diffusion

Recommandations:

Pour rappel, les données publiques doivent par défaut être ouvertes, vous devez donc expliquer les raisons pour lesquelles les données ne pourront être diffusées. Par essence, les données secrètes et sensibles ne pourront être diffusées. Il en est de même pour les données à caractère personnel qui ne pourront être diffusées qu'après avoir été anonymisées de manière irréversible.

Consultez la page "[Maîtriser le cadre juridique](#)" de l'intranet pour en savoir plus.

Exemple de réponse:

Les vidéos ne seront pas diffusées en open research data parce qu'elles comportent des données permettant d'identifier des personnes. Les moyens d'anonymisation automatique de vidéos ne sont actuellement pas assez performants et l'anonymisation manuelle serait démesurément trop longue à effectuer.

7. Ressources nécessaires à la gestion des données

Ressources

Exemple de réponse:

300 € / To / an

Recommandations:

La gestion des données peut engendrer des coûts de stockage, de matériel et des ressources humaines et financières par exemple pour préparer les données à l'ouverture. Préciser dans cette partie les ressources prévues pour la gestion et l'ouverture des données.

Pour vous aider, vous pouvez vous appuyer sur [cette ressource proposée par l'Inist-CNRS](#).