

IRSN - Institut de Radioprotection et de Sûreté Nucléaire: PGDR type "projet de recherche" IRSN

Informations sur le service de recherche

Nom du service (ou laboratoire, le cas échéant) de recherche

Recommandations:

Intitulé complet.

Par ex: Service des Agressions Internes et des risques Industriels

Acronyme service (ou laboratoire) et appartenance

Exemple de réponse:

PSN-RES/Sa2i ou PSN-RES/Sa2i/LEF

Structure IRSN de recherche (selon répertoire national des structures)

- URSûr 201722597T
- UREnv 201722596S
- URSan 201722594P

Recommandations:

Les 3 numéros proposés correspondent aux numéros d'identification des 3 pôles recherche de l'IRSN, tels qu'ils sont référencés dans le Répertoire NATIONAL des Structures de Recherche (RNSR).

201722597T : [Répertoire des structures. \(education.fr\)](http://education.fr)

201722596S : [Répertoire des structures. \(education.fr\)](http://education.fr)

201722594P : [Répertoire des structures. \(education.fr\)](http://education.fr)

Partenaires impliqués dans le projet de recherche

Recommandations:

Préciser les structures partenaires internes ou externes en précisant un contact pour chacune (Nom, prénom, adresse mail)

Responsable(s) de la gestion et de l'ouverture des données

Recommandations:

La personne responsable de la gestion et de l'ouverture des données d'un projet de recherche est la personne désignée pour piloter la rédaction du PGD et la vérification de la conformité de la gestion effective des données avec ce qui est décrit dans le plan. Elle est en charge, sous couvert du chef de projet, de piloter l'analyse permettant de statuer sur l'ouverture des données.

Indiquer nom, prénom, appartenance, email, numéro de téléphone et, le cas échéant, identifiants (ORCID, Id HAL). Pour en savoir plus sur ces identifiants, s'adresser au correspondant Open Science du SEARCH (audrey.legendre@irsn.fr)

Informations générales sur les données du projet

Description des données

Recommandations:

On entend par "Description des données" ce que dirait le chercheur, le plus "naturellement" possible, avec ses propres mots, pour décrire les données qu'il collecte et manipule dans le cadre du projet.

Par exemple : donnée de la cohorte..., données d'acquisition d'une campagne de mesures sismiques, données de microscopie, données de cytométrie en flux,...

Préciser le mode d'obtention des données produites

- 1- Données d'observation
- 2- Données expérimentales
- 3- Données computationnelles
- 4- Données dérivées ou compilées

Recommandations:

Préciser la modalité d'obtention des données.

Les **données d'observation** : ce sont par ex. des données d'enquête, de télédétection, d'échantillonnage, d'enregistrements *in situ* sur le terrain, des images de microscopie,...

Les **données expérimentales** : ce sont des données produites par une expérimentation par ex. des résultats d'essais, des chromatogrammes, des séquences de gènes,...

Les **données computationnelles** : elles sont issues de modèles (modèle climatique, modèle économique, modèle physique, ...) ou de simulations

Les **données dérivées ou compilées** : elles sont issues de la compilation d'un ensemble de données collectées et/ou créées, résultats de fouille de texte (text mining) ou de fouille de données (data mining)

Données utilisées par le projet et provenant de sources extérieures

Recommandations:

Le cas échéant, indiquer : Nom du/des jeu(x) de données, nature, source(s) (donner l'identifiant pérenne ou l'URL le cas échéant), volume, format(s), conditions d'acquisition (coût éventuel, conditions, licence), fréquence d'acquisition/mises à jour si nécessaire.

NB : Avant de réutiliser des données, vous devez vous assurer que vous avez le droit de les réutiliser, notamment en :

- vérifiant qu'elles ne sont pas protégées par une réglementation nationale ou internationale, un copyright ou un droit de propriété intellectuelle...

- vérifiant que les personnes sont informées de la réutilisation de leurs données personnelles, le cas échéant.

Si vous avez un doute, consultez le Service Juridique de l'IRSN.

Logiciel(s) de production ou de traitement nécessaire(s) avec sa (leur) version

Recommandations:

Préciser le cas échéant les logiciels qui ont servi à la production et/ou au traitement des données et leur version.

Ceci vise à garantir l'interopérabilité et la pérennité des données produites afin de garantir qu'on puisse toujours les ouvrir. Même si elles sont obtenues avec des logiciels dits "propriétaires", il faut s'assurer qu'elles soient, autant que possible, disponibles sous des formats libres.

"Criticité" des données : que se passerait-il si les données étaient corrompues ou perdues?

- 1- Elles sont re-générables avec peu d'effort et à faible coût
- 2- Elles sont re-générables avec un effort/coût modéré
- 3- Elles ne sont re-générables qu'au prix d'un effort/coût élevé
- 4- On ne peut pas les ré-obtenir

Recommandations:

Il s'agit de préciser dans quelle mesure les données produites/ collectées pourraient être facilement ré-obtenues, en cas de perte. Ceci oriente sur les solutions de stockage et de conservation à considérer : il faut être encore plus vigilant pour les données de niveau 3 ou 4.

Documentation et qualité des données au fil du projet de recherche

Qualification des données au fil du projet de recherche

Recommandations:

Pendant le processus de recherche, les données vont suivre un processus de "qualification" les faisant passer de l'état de données brutes ou primaires (B) aux états de données nettoyées /traitées (N) puis analysées (A) et enfin interprétées/figées (I). Ce processus est constitutif du "cycle de vie de la donnée".

Les moyens mis en œuvre pour atteindre ces différents états seront précisés dans la description du protocole expérimental (cahier de laboratoire ou autre dispositif).

Selon le projet de recherche, les données "produits" ou données d'intérêt pourront être des données B, N, A et/ou I.

Règles de nommage des ensembles de données

Recommandations:

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

Inclure si possible dans les règles la qualification des données (B, N, A, I) et une date, en évitant les caractères spéciaux, accentués et les espaces vides.

Prévoir également une gestion et un contrôle du versionning.

Etablir une liste des jeux, maintenue à jour, le cas échéant.

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'organisation de la conservation et de l'archivage

Gestion des données pendant le projet de recherche

Stockage des données tout au long du processus de recherche

Recommandations:

Il faut utiliser des espaces de stockage robustes, dotés d'une sauvegarde garantie.

Pour un stockage à l'IRSN, utiliser les espaces recommandés par D3NSI, selon les indications disponibles sous

M:\PGDRECHERCHE\Guide de remplissage du PGDR\Gestion des données pendant le processus de recherche.docx

La sécurité des données dans ces espaces est garantie par le contrôle des accès par la personne administratrice de l'espace dédié.

Le stockage des données sur des ordinateurs portables, des disques durs externes, ou des périphériques de stockage tels que des clés USB est proscrit.

Figurer l'ensemble des données à la fin du projet pour des questions d'authenticité et de traçabilité. Si lors de cette étape, un document à valeur probante est généré, il faut le conserver et indiquer son lieu de stockage dans le PGD afin qu'il soit collecté par les archivistes pour être conservé de manière sécurisée et pérenne.

Plan de sauvegarde des données

Recommandations:

Pour un stockage à l'IRSN, en utilisant les espaces préconisés par D3NSI, la sauvegarde (et la restauration le cas échéant) est garantie par les procédures d'exploitation fixées dans un référentiel (pour info: [le référentiel IRSN/S3A/REF-057](#))

Si le stockage se fait en dehors des espaces gérés par D3NSI (notamment si les données sont stockées à l'extérieur de l'institut par exemple pour des données acquises et traitées sur des plateformes externes), il faut interroger et s'assurer des procédures de stockage et de sauvegarde des données mises en œuvre.

Conservation des données au terme du projet de recherche

Définition des critères de tri et de sélection des données à conserver

Recommandations:

Les critères de tri, selon la nature des projets de recherche peuvent être très différents.

Les questions suivantes (liste non exhaustive) peuvent guider le choix des critères de tri, selon la typologie B, N, A, I :

- Y-a-t-il une obligation légale ou contractuelle de conservation de certaines données?
- Existe-t-il un standard de conservation dans le domaine scientifique ?
- Sont-ce des données d'intérêt non reproductibles ?
- Sont-ce des données liées à des publications?

Durée de conservation des jeux de données

Recommandations:

La durée de conservation des données doit être définie en accord et avec l'appui des archivistes du SEARCH et elle doit être cohérente avec le tableau de gestion des archives (TGA).

Espace de conservation des données

Recommandations:

Les recommandations IRSN sont disponibles sous *M:\PGDRECHERCHE\Guide de remplissage du PGDR\Conservation des données au terme du projet de recherche.docx*

Y-a-t-il besoin de ressources supplémentaires particulières pour la conservation des données?

Exemple de réponse:

Un espace de conservation particulier pour les cohortes en épidémiologie

Recommandations:

Précisez la nature et la gouvernance de cet espace.

Référencement des données dans le catalogue des données de l'institut

Recommandations:

Les données doivent être répertoriées dans le catalogue des données de l'institut. A faire avec la personne référente du service en charge de la mise à jour du catalogue IRSN sur PADI ou prendre contact avec la cellule de valorisation des données à D3NSI.

Exigences légales et éthiques, code de conduite

Nature des données traitées dans le projet: s'agit-il de ?

- Données à caractère personnel
- Données de santé
- Données couvertes par le secret des affaires
- Biens à double usage
- Autre sensibilité à préciser

Recommandations:

Si les données ne rentrent dans aucune des catégories énoncées, ne rien cocher

Pour toute question relative aux définitions proposées, se rapprocher du DPO et/ou du SJ

Si des données personnelles sont contenues dans les données du projet, peuvent-elles être:

- anonymisées au moment de la collecte
- anonymisées pendant la phase d'analyse
- anonymisées à la fin de l'analyse et avant publication
- pseudonymisées au moment de la collecte
- pseudonymisées pendant la phase d'analyse
- pseudonymisées à la fin de l'analyse et avant publication

Recommandations:

Il faut prendre attache du DPO pour s'assurer de traiter les données personnelles dans les constions idoines

Y -a-t-il besoin de traiter des considérations éthiques ou juridiques par rapport aux données?

Recommandations:

La question amène la réflexion sur d'éventuelles questions d'éthique et de déontologie qui devraient être prises en compte et leur impact sur le transfert et le stockage des données, leur accès et leur durée de conservation. Le cas échéant, démontrer que ces aspects sont bien pris en compte (prévoir un chapitre dédié dans la description du projet de recherche, vérifier si une revue par un comité d'éthique est requise) et sont planifiés (prévoir un/plusieurs livrable(s) dédié(s) si nécessaire).

Dans le cas positif, il faut lister les chartes et codes de conduite nationaux ou internationaux ainsi que la charte d'éthique et de déontologie et/ou le code de déontologie applicables dans le cadre du projet.

Se rapprocher de DRP/CCE

Y-a-t-il des questions relatives à la propriété intellectuelle (propriété, droits d'usage,...)?

Recommandations:

S'interroger sur les questions de propriété intellectuelle (ex.: propriété, droit d'usage) et sur la manière dont elles sont gérées? Par exemple, ces questions sont-elles réglées par un accord de consortium? Ou un contrat doit-il être mis en place? Y faire référence, le cas échéant

Indiquer notamment :

- qui est le propriétaire des données,
- quelles conditions d'accès sont applicables aux données notamment celles transmises par un tiers (liberté d'accès ou restriction);
- Existe-t-il des logiciels ou base données? Sont-ils accessibles et si oui comment (licence)?
- Evoquer éventuellement le cas du mandataire unique s'il a été désigné afin de savoir qui gère les aspects de valorisation.

S'appuyer sur les compétences de SG/SJ et DST/SP3In

Partage des données en Open data

Les données du projet de recherche peuvent-elles être publiées en Open Data?

Recommandations:

La loi pour une république Numérique de 2016 porte l'obligation d'ouverture des données publiques y compris les données de la recherche. Depuis le décret n°2021-1572 du 3 décembre 2021 *relatif aux exigences de l'intégrité scientifique par les établissements publics contribuant au service public de la recherche*, il a obligation pour les établissements publics de promouvoir la diffusion en Open Data des données produites par de la recherche financée sur fonds publics en les ayant rendues **FAIR : Faciles à trouver, Accessibles, interopérables et Réutilisables**. Le principe d'ouverture qui prévaut est celui des **"données aussi ouvertes que possible et fermées que nécessaire"**.

En pratique, pour savoir si les données peuvent être ouvertes, on peut s'appuyer, en première intention, sur le questionnaire disponible sur *M:\PGDRECHERCHE\Guide de remplissage du PGDR\Questionnaire ouvrabilité des données.docx*.

In fine, il convient de vérifier avec le Service Juridique que les données d'un projet ne sont pas contraintes par des dispositions qui empêcheraient leur diffusion en Open data.

Concernant la diffusion en Open Data, la consigne de l'IRSN est d'utiliser les entrepôts thématiques ou les entrepôts définis conjointement avec les partenaires du projet (le catalogue [re3data](#) fournit une liste d'entrepôt thématiques). A défaut, l'IRSN dépose ses jeux de données dans Recherche Data Gouv (Plateforme nationale fédérée des données de la recherche).

Constitution des jeux de données mis en Open data

Recommandations:

Lister et décrire le(s) différent(s) jeux de données qui sera (seront) mis en Open Data. Pour respecter les bonnes pratiques, chaque jeu de donnée doit avoir un titre, un descriptif, être rendu FAIR et, le cas échéant, le suivi des différentes versions doit être décrit.

Données F.A.I.R.

Choix des métadonnées pour des données Faciles à trouver

Recommandations:

Préciser les métadonnées qui seront utilisées pour faciliter la recherche et l'identification du (des) jeu(x) de données, en s'appuyant, si possible, sur des ontologies. Ces standards sont souvent spécifiques à la discipline et à l'entrepôt utilisé.

Attribution d'un identifiant unique et permanent de chacun des jeux de données

Recommandations:

L'usage d'un identifiant unique et pérenne (DOI = Digital Object Identifier) permet de garantir la pérennité du lien vers les données. Il est **indispensable** pour diffuser des données en Open data. Celui-ci est obtenu auprès de DATACITE.

La démarche à suivre est décrite sous *M:\PGDRECHERCHE\Guide de remplissage du PGDR\Attribution d'un DOI.docx*.

Qui a créé le(s) jeux de données ?

Recommandations:

Il est nécessaire de relier le jeu de données :

- à son(s) producteurs en utilisant le (leur) PID (Personal identifier, de type ORCID ou IdHAL) : pour la création de cet(ces) identifiant(s) se renseigner auprès de SEARCH
- aux établissements de recherche associés au projet avec leur n° RNSR (Répertoire National des Structures de Recherche). Pour l'IRSN, 201722597T ou 201722596S ou 201722594P.

Interopérabilité technique des données

Exemple de réponse:

csv, json, html...

Recommandations:

Il faut utiliser, autant que possible, des formats ouverts (=libre).

Voici une liste non exhaustive de formats de fichiers ouverts

Type de données	Formats ouverts les plus classiques
Données brutes	CSV, JSON, XML
Données géographiques	KML, SHP
Texte	DOCX, ODT, RTF, TXT
Image	BMP, GIF, JPG, PNG
Tableur	XLSX, CSV, ODS
Vidéo	MKV, NUT, OGM
Audio	FLAC, OGG, WAV, MP3
Présentations	ODP, PPTX
Archivage	7z (Seven Zip), TAR, ZIP
Dessin	EPS, ODG, SVG

Quelles licences pour les données réutilisables?

Recommandations:

Avant de rendre public un jeu de données, il faut lui apposer une licence de diffusion fixant les conditions de son utilisation : droits d'utilisation et de modification de la donnée, droits de réutilisation commerciale et non commerciale, obligations éventuelles comme la mention de la source des données ou le partage à l'identique.

Vous n'avez pas toujours le choix du type de licence qui va s'appliquer à vos données :

- L'entrepôt dans lequel vous souhaitez déposer votre jeu de données peut imposer une licence de diffusion particulière; vérifier ses exigences ;
- lorsque les données sont liées à un article scientifique, la licence de diffusion choisie pour les données doit répondre aux exigences de la revue. Consultez les instructions aux auteurs ;
- le type de licence préconisé est donc **un critère de sélection important pour choisir un entrepôt ou un éditeur.**

Il faut privilégier une licence largement utilisée et compatible avec les autres licences existantes. Différents types de licences sont utilisables comme les licences Creative Commons CC4.0, la licence ouverte LO créée par Etalab ,...

Pour le choix des licences, il convient de prendre l'avis de SG/SJ.

Coût des dépôts de données et coût du maintien des données dans les entrepôts de données?

Ces coûts ont-ils été évalués?

Sont-ils intégrés dans le budget du projet ou non?

Par qui seront-ils portés, le cas échéant?

Recommandations:

Les coûts d'entreposage des données dans des entrepôts devraient être traités dès la phase d'élaboration du projet.

Si cela n'a pas été fait, il est encore temps de se poser la question au démarrage du projet.

Quelques outils existent pour aider à évaluer les coûts :

- OpenAIRE : identifier et évaluer les coûts de gestion des données de la recherche [how-to-comply-to-h2020-mandates-rdm-costs](https://openaire.eu) (openaire.eu)
- Data management costing tool (UK Data service)
- Guide de l'Université d'Utrecht [Costs of data management | Utrecht University \(uu.nl\)](https://www.uu.nl/en/management-and-administration/finance-and-accounting/data-management)
- EPFL Cost Calculator for Data Management [Cost Calculator \(epfl.ch\)](https://www.epfl.ch/research/data-management/cost-calculator/)
- TU Delft Data management costs [Data management costs \(tudelft.nl\)](https://www.tudelft.nl/en/research/data-management/costs/)