



**HAL**  
open science

## Data management plan of REVERSAAL research unit

Gabrielle Favreau, Rémi Clément, Vivien Dubois, Jean-Marc Choubert

### ► To cite this version:

Gabrielle Favreau, Rémi Clément, Vivien Dubois, Jean-Marc Choubert. Data management plan of REVERSAAL research unit. 2024, <https://entrepot.recherche.data.gouv.fr/dataset.xhtml?persistentId=doi:10.57745/C6P6WW>. hal-04634231v2

**HAL Id: hal-04634231**

**<https://hal.inrae.fr/hal-04634231v2>**

Submitted on 7 Oct 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

## PLAN DE GESTION DE DONNEES DE STRUCTURE DE L'UNITE DE RECHERCHE REVERSAAL



Licence [CC-BY SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/)

Date : 2024\_10\_07

DOI PGD : 10.57745/C6P6WW

### INFORMATIONS SUR LA STRUCTURE

#### a. Nom de la structure

- UR REVERSAAL : Unité de recherche Réduire, réutiliser, valoriser les ressources des eaux résiduaires.
- Codique : 1468

#### b. Identifiant RNSN

- 201822688M

#### c. Identifiant de la structure

Localisée à Villeurbanne, l'unité de recherche (UR) REVERSAAL « Réduire Réutiliser Valoriser Les Ressources Des Eaux Résiduaires » est une unité propre de recherche du d'INRAE, rattachée au Centre Lyon-Grenoble Auvergne-Rhône-Alpes et au département scientifique TRANSFORM. Elle mène des recherches sur le traitement et la valorisation des eaux usées urbaines. Plus précisément, elle positionne son projet de recherche autour de la préfiguration de la station du futur dans une perspective de résilience des villes et de valorisation des ressources en eau, énergie et matières, en accord avec les objectifs de la bioéconomie. Elle occupe ainsi une place unique dans le périmètre

des écotecnologies pour l'environnement. Ses travaux visent à faire progresser les connaissances pour élaborer des préconisations pertinentes et innovantes en termes de conception, de dimensionnement, d'exploitation et d'optimisation des installations de valorisation et de traitement des eaux résiduaires.

L'UR REVERSAAL développe des travaux de recherche avec trois objectifs principaux qui forment trois des axes de recherche de l'unité : réduire les impacts induits par les effluents et leur traitement/valorisation, réutiliser les effluents traités et valoriser l'énergie et les matières qu'ils contiennent. Le quatrième axe, transversal, concerne l'innovation digitale. Ces activités font appel aux disciplines scientifiques du génie des procédés appliqué au traitement et à la valorisation des eaux urbaines, avec des compétences complémentaires en chimie, mécanique des fluides et géophysique. L'unité participe également à des expertises techniques, pour le compte du Ministère en charge de l'Écologie, notamment.

L'UR REVERSAAL a été créée au 1<sup>er</sup> janvier 2018 et a été rattachée au département TRANSFORM lors de la fusion d'Irstea et de l'INRA au 1<sup>er</sup> janvier 2020. L'unité compte actuellement 35 personnes dont 19 agents permanents (4 chercheurs, 15 ingénieurs et techniciens) et 9 doctorants et post-doctorants. 7 agents sont titulaires de l'habilitation à diriger les recherches (HDR) auprès de l'école doctorale Chimie-Procédés-Environnement de Lyon (ED206). Le budget annuel de l'unité est d'environ 1,5 M€ réparti sur une trentaine de projets de recherche actifs.

Les activités de l'UR REVERSAAL sont menées sur trois sites :

- l'un à Villeurbanne (350 m<sup>2</sup> de bureaux et 200 m<sup>2</sup> de locaux expérimentaux ou de stockage) ;
- le second site est le Hall expérimental distant de 2 km et situé à proximité immédiate de la station d'épuration de la Feyssine (150 m<sup>2</sup> équipés de pilotes instrumentés et automatisés alimentés en continu en eaux usées et boues et d'un laboratoire d'analyse de routine. Des équipements de mesure en conditions contrôlées (décantation, échanges gazeux, caractérisation des milieux poreux, cinétiques de transformation) y sont couramment utilisés ;
- le troisième site est la plateforme REFLET dédiée aux solutions fondées sur la nature, située à Craponne à 25 km de Villeurbanne. Elle comprend 9 pilotes de 20 m<sup>2</sup>.

Outre l'étude des procédés et des filières, l'UR REVERSAAL étudie également les processus à l'échelle laboratoire et met au point des modèles numériques de procédés et filières.

Pour plus de détails vous trouverez la vidéo de présentation des plateformes sur le lien suivant : <https://youtu.be/LubQrvsB-gc>

## ***e. Responsabilités dans la structure***

- Jean-Marc Choubert (DU)
- Pascal Molle (DUA)
- Rémi Clément (RDO)
- Chaque porteur d'un projet

## ***e. Etablissement(s) tutelle(s)***

- INRAE

## **f. Département scientifique et Centre de rattachement INRAE**

- 52 - Sciences pour l'ingénierie des aliments, des produits biosourcés et des résidus de l'activité humaine TRANSFORM
- CE45 – Lyon Grenoble Auvergne Rhône Alpes

## **g. Financeur(s) (permettant l'acquisition des jeux de données)**

- Contrats de recherche (RPC) et recettes (RPNC) permettant l'acquisition de données

## **INFORMATIONS SUR LE PLAN DE GESTION**

**DOI** (version publiée du plan de gestion) :

- Cf. numéro inscrit en page de garde

**Historique des versions :**

1. PGD\_S\_REVERSAAL\_V1\_FR, le 28/11/2023

**Rédigé par :**

- Groupe de travail composé de Gabrielle Favreau, Rémi Clément, Vivien Dubois, Jean-Marc Choubert

**Relu par :**

- Interne UR : Nicolas Forquet, Sylvie Gillot, Sophie Besnault
- Externe UR : Anne-Sophie Bage

## TYPOLOGIE GENERALE DES DONNEES

Quatre types de produits constituent les données utilisées dans les projets de recherche de l'unité.

- A - Données générées lors d'expérimentations ou de simulations
- B - Codes sources
- C - Données et codes mis à disposition par un partenaire
- D - Données d'enquêtes

Les paragraphes suivants précisent pour chaque produit :

- les types de données concernées,
- le droit de propriété intellectuelle,
- la sensibilité et la confidentialité,
- le partage, l'organisation et la documentation,
- le stockage et la sécurité
- l'archivage et la conservation.

Remarques :

- Les données collectées dans le cadre de synthèses bibliographiques ne sont concernées par aucun des 4 types de produits si elles ont déjà été publiées dans un article scientifique et qu'elles peuvent être retrouvées à l'aide de la référence
- Les données non-publiées confiées par son auteur entrent dans le champ du produit C.

## Avant d'aller plus loin

La réalisation du PGD de projet à partir de ce PGD de structure est facilité par la fiche [Mémo#5 « Faire un PGD de projet »](#) sur Data@reversaal (Nextcloud).

L'outil Nextcloud de stockage des données cité dans tout ce document fait l'objet de deux notices d'utilisation : Fiche [Mémo #3 « Utiliser Nextcloud »](#) et Fiche [Mémo #4 « Gérer les accès Nextcloud »](#) sur Data@reversaal (Nextcloud).

## A – Produit “ Données générées lors d'expérimentations ou de simulations ”

### I. Présentation générale des données

- Mode d'obtention des données

Données générées par voies expérimentales ou par calculs numériques lors du travail d'un agent de l'UR REVERSAAL d'INRAE

- Origine

Expérimentations menées sur sites, sur plateformes expérimentales et au laboratoire.  
Simulations numériques.

- Type de données

Données tabulaires (ex. sous forme de tableur)

Image

Texte

Fichier géographique

Code

Fichier binaire (données stockées dans un format compressé propriétaire ou non)

- Nature des données

Composition bio-physico-chimique (concentrations en polluants, pH, température, indice de décantation...) d'effluents, de boues, de populations bactériennes, de produits chimiques, valeurs mesurées ou calculées

Débit (de gaz, d'effluent), temps de fonctionnement d'équipements, dépenses énergétiques

Données personnelles (coordonnées téléphoniques, e-mails) d'exploitants ou de maîtres d'ouvrages

Fichiers géographiques (types de sols, coordonnées GPS d'ouvrages, points de mesures)

Mesures hydriques (précipitation, évapotranspiration) et de perméabilité

Schéma de filière de traitement, dimensions associées, liste et caractéristiques d'équipements

...

La liste ci-dessus est non exhaustive.

- Format des données

Les données collectées et produites (analysées et mises en forme) peuvent se présenter sous de multiples formats :

- Données tabulaires : txt csv, xls,xlsx, odt
- Image : png, jpeg, pdf, tiff, bmp, vtk, mpeg, avi, mov, mkv ...
- Texte : pdf, odf, doc, docx, txt, ppt, pptx
- Fichiers géographiques : shp, gpkg
- Formats génériques : rar, zip, Rdata, feather, HDF5
- Code : R, py

- Périmètre thématique des données

- Procédés et filières de traitement de valorisation des eaux urbaines, des boues
- Capteurs et instruments de mesure
- Simulateurs et codes numériques

## II. Droit de propriété intellectuelle

L'UR REVERSAAL est en accord avec les principes de l'Open Data et respecte la [Charte des Infrastructures de Recherche](#) d'INRAE. Les données collectées et celles produites à partir leur analyse sont propriétés d'INRAE. Des droits spécifiques peuvent s'appliquer en cas de partenariat avec des partenaires privés.

## III. Sensibilité et confidentialité des données

- Identification du niveau de sensibilité des jeux de données

Diffusion restreinte : données scientifiques et techniques acquises lors des expérimentations ou de calculs numériques

Confidentiel : données personnelles ayant un niveau d'impact inférieur ou égal à 2 (nom, prénom, adresse mail professionnelle)

- Garantie de la confidentialité (mesures prises et normes)

INRAE s'engage à conserver strictement ces données et à ne pas les communiquer à des tiers, même internes à l'unité (hors du projet), sans consentement écrit du coordinateur du projet (sauf si exception mentionnée dans la PGD ou la convention du projet).

- Analyse de second niveau et garantie de la confidentialité

Les données sont analysées par un (ou plusieurs) agents de l'unité. Elles peuvent faire l'objet d'une analyse de second niveau dans le cadre d'un contrat partenarial spécifique (se référer à la convention du projet).

#### IV. Partage des données

- Politique de partage (obligation, restriction, interdiction)

L'UR REVERSAAL est soumise aux règles institutionnelles formalisées par la [charte pour le libre accès aux données et aux publications](#). Les données scientifiques produites à partir de données expérimentales ou de simulations doivent être partagées à l'exception d'un contrat partenarial spécifique (se référer à la convention). L'unité recommande d'utiliser une licence libre [Creative Commons CC-by-nc-nd 4.0](#) restreignant tout usage commercial et toute modification sans l'accord des auteurs.

Le partage de données confidentielles (données personnelles, coordonnées GPS) est interdit.

- Réutilisations potentielles des données scientifiques obtenues

Au terme du projet et de l'éventuel embargo pour publication, l'UR REVERSAAL pourra réutiliser les données acquises lors des expérimentations et simulations qu'elle a menées, sauf si exception mentionnée dans le PGD ou la convention de projet.

- Modalités de lecture

Les données collectées et produites (analysées et mises en forme) peuvent se présenter sous de multiples formats et donc nécessiter de multiples logiciels pour leur lecture. La liste ci-dessous est non exhaustive :

csv, xls, xlsx, odf : Excel, LibreOffice, OpenOffice

R : R, Rstudio

py : Python, Rstudio, PyCharm, Visual Studio Code

png, jpeg, tiff : Photos, Firefox, XnView, Paint

mpeg, avi, mov, mkv : Lecteur multimédia, VLC



pdf : Adobe Reader

odf, doc, docx : Word, LibreOffice, OpenOffice

txt : Bloc-notes, Wordpad

ppt, pptx : PowerPoint, LibreOffice, OpenOffice

shp, gpkg : QGIS

rar, zip : 7-Zip, RAR

- Modalités de partage

Avec qui ?	Comment ?	Sous quelle licence ?
<b>Porteurs de projets</b>	Stockage réseau : espace NextCloud du projet	L'UR restreint aux membres du projet. Sinon se référer au PGD du projet.
<b>Participants et financeurs</b>	Diffusion par voie électronique (mail, lien de téléchargement) et/ou espace NextCloud du projet et/ou présentation, rapports	L'UR restreint aux membres du projet. Sinon se référer au PGD du projet.
<b>Tous</b>	Entrepôt <a href="http://data.gouv.fr">data.gouv.fr</a>	Sous licence <a href="https://creativecommons.org/licenses/by-nc-nd/4.0/">CC-by-nc-nd 4.0</a> . Si autre, se référer au PGD du projet.

#### V. Organisation et documentation des données

- Méthodes et outils utilisés lors du cycle de vie de la donnée (acquisition, traitement, mise à disposition, archivage, destruction)

Étape	Méthode(s)	Outil(s)
<b>Préparation</b>	Plan d'échantillonnage	Word
<b>Acquisition</b>	Expérimentations Stockage sur un support fiable	Capteur, mesures et analyses Stockage réseau : NextCloud
<b>Traitement</b>	Analyse des données Stockage sur un support fiable	Tableur Simulateur (ex. Sumo) Stockage réseau : NextCloud
<b>Synthèse</b>	Rapport / Présentation Stockage sur un support fiable	Traitement de texte et création d'un support de présentation Stockage réseau : NextCloud
<b>Partage</b>	Diffusion par voie électronique Présentations Entrepôts	Client mail (préférentiellement <a href="http://partage.inrae.fr">partage.inrae.fr</a> et <a href="http://filesender.com">filesender</a> )
<b>Archivage / Destruction (à l'issue du projet)</b>	Archivage des données brutes et produites Destruction des données personnelles	Stockage réseau : NextCloud Datacenter

- Métadonnées (nature, production et mise à jour)

Les fichiers de données brutes ou analysées (.xlsx, .shp, etc.) sont accompagnés d'un fichier de métadonnées respectant le format défini par l'UR et basé sur le DUBLIN CORE (<https://www.bnf.fr/fr/dublin-core>). Ce modèle de fichier ([téléchargeable sur ce lien](#)) est accessible à tous les agents de l'UR sur NextCloud : reversaal > donnees@reversaal > plan\_gestion\_donnees (onglet Métadonnées).

Point d'amélioration : Information large de cette pratique aux agents de l'UR REVERSAAL, choix d'une base de données plus adaptée

- Documentation complémentaire aux métadonnées

Non-concerné

- Gestion et organisation des fichiers de données

Point d'amélioration : Définir une liste harmonisée de métadonnées (ontologie), Elaborer une convention de nommage de fichier, travail à confier à un groupe de travail éphémère qui s'inspirera notamment des informations existantes au sein du réseau de RDO/RDS.

- Procédure de contrôle qualité des données

Point d'amélioration : Mise en place d'une vérification des données par les membres du projet et son coordinateur.

## VI. Stockage et sécurité des données

- Systèmes d'information

Les données sont stockées sur des serveurs situés dans un datacenter choisi par l'établissement INRAE. L'accès aux données est géré par une authentification déléguée à chaque partenaire de la fédération Edugain.

- Supports de stockage

Une fiche pratique relative au stockage numérique des données (cf. fiche [Mémo#2 « Stocker les données »](#) sur Data@reversaal (Nextcloud)) a été mise en place au sein de l'unité pour éviter la perte de donnée et adopter une démarche de science ouverte. Cette fiche permet aux agents d'identifier le(s) support(s) de stockage le plus adéquat(s) pour leurs données en fonction du stade d'acquisition et de traitement de celles-ci. Elle vise aussi à uniformiser les pratiques au sein de l'UR REVERSAAL.

Piste d'amélioration : remise de cette fiche et échange sur le sujet dès l'arrivée d'un nouvel agent.

- Sécurisation du transfert de données

Transferts physiques (clé USB, disque dur externe) : non autorisés.

Transferts réseau (FileSender, NextCloud) : mots de passe, chiffrement et utilisation de protocoles sécurisés.

- Volumétrie actuelle et prévisionnelle

L'UR REVERSAAL bénéficie actuellement d'une capacité totale de stockage de 2 To qui est ajustable chaque année.

- Politique de sécurité de l'entité stockant physiquement les données

La [Politique Sécurité des Systèmes d'Information \(PSSI\)](#) INRAE s'applique à l'entité stockant physiquement les données de l'UR REVERSAAL (hébergement interne INRAE).

- Partage avant publication et confidentialité

L'UR REVERSAAL recommande l'embargo jusqu'à publication. Si exception, se référer au PGD du projet.

- Sécurité, intégrité et traçabilité

Sécurité : protection contre les virus de tous les postes de travail, formation à la [cybersécurité](#), authentification requise pour accéder aux ressources.

Intégrité : se référer à la [charte de déontologie, d’intégrité scientifique et d’éthique d’INRAE](#).

Traçabilité : existence d’un modèle de fichier de métadonnées ([téléchargeable sur ce lien](#)) à disposition de tous les agents sur NextCloud : reversaal > donnees@reversaal > plan\_gestion\_donnees (onglet Métadonnées)

Point d’amélioration : mettre en place un suivi des modifications pour savoir ce qui est modifié, par qui et quand.

- Sensibilisation aux bonnes pratiques hygiène numérique

Les agents de l’UR REVERSAAL participent aux sensibilisations aux bonnes pratiques d’hygiène numérique effectuées au niveau du centre (fiche, séminaire).

Piste d’amélioration : planifier une intervention régulière du correspondant SSI du centre pour les nouveaux arrivants notamment.

## VII. Archivage et conservation des données

- Identification des données à conserver et de celles à détruire

Données à conserver : données scientifiques validées

Données à détruire : données personnelles

- Plateforme ou procédures pour un archivage pérenne des données

Archivage sur NEXTCLOUD.

Pont d'amélioration : Actualiser la procédure et la fiche mémo, et la diffuser.

- Durée de conservation des données

Variable selon la nature des données et des projets. Se référer au PGD du projet.

- Garantie de financements pour couvrir les coûts associés à la conservation à long terme

Fonds propres à l'unité et fonds provenant de projets de recherche.

## B – Produit “ Codes sources ”

### I. Présentation générale des données

- Mode d'obtention du code source

Code ou programme produit par un membre de l'UR REVERSAAL dans le cadre de projet de recherche.

- Origine
  - Interne à l'UR REVERSAAL
  - Modélisation/simulation/analyse de données

- Type de code source

Il est composé d'une succession d'instructions composant un programme sous une forme lisible, telles qu'elles ont été écrites dans un langage de programmation. Il peut être élaboré dans un langage propriétaire ou dans un langage libre.

- Nature des codes sources

Codes informatiques développés dans différents langages, sous différents environnements et pour différents cas d'utilisation :

- **Analyse de données produites** : Il s'agit d'un code ayant pour rôle de réaliser la mise en forme, le filtrage et la restitution graphique des données ;
- **Modélisation/Simulation** : Ce code est utilisé pour simuler le comportement d'un procédé ou d'un processus, souvent dans le but de mieux comprendre son fonctionnement ou d'effectuer des prédictions en vue de prendre des décisions éclairées (ex. optimisation) ;
- **Commande/ Contrôle** : Référence à un ensemble de fichiers et de lignes de code informatique écrits dans un langage de programmation spécifique pour contrôler le comportement d'un circuit électronique, d'un automate ou d'une centrale d'acquisition ;
- **Base de données** : Un ensemble d'instructions informatiques écrites dans un langage de programmation spécifique pour effectuer diverses opérations liées à la gestion et à l'utilisation d'une base de données. Le rôle principal d'un tel programme est de permettre aux utilisateurs ou aux applications d'interagir avec la base de données, de stocker, de récupérer, de mettre à jour et de gérer les données de manière efficace et sécurisée ;
- **Application WEB** : logiciel qui s'exécute dans le navigateur Web permettant d'accéder à une base de données ou bien à un outil de calculs (ex. Energie STEP).

- Format des codes sources

Les codes sources produits peuvent se présenter sous de multiples formats :

- **Analyse de données** : R, Python, Matlab, Excel.

- **Modélisation/Simulation** : R, Python, Matlab, GPS-X, SUMO, COMSOL, QGIS...
- **Commande/ Contrôle** : Python, Matlab
- **Application WEB et Base de données** : SQL, DDL, DBF, XML, PHP, Ruby On Rails, HTML5, CSS, frameworks, PHP, Javascript

▪ Périmètre thématique des codes sources

- Procédés et filières de traitement de valorisation des eaux urbaines, des boues
- Capteurs et instruments de mesure
- Simulateurs et codes numériques

II. Droit de propriété intellectuelle

L'UR REVERSAAL d'INRAE est en accord avec les principes de l'Open Data (FAIR) et respecte la [Charte des Infrastructures de Recherche](#) d'INRAE. Les codes sources sont propriétés d'INRAE. Des droits spécifiques peuvent s'appliquer en cas de partenariat avec des entreprises privées.

III. Sensibilité et confidentialité des codes sources

- Identification du niveau de sensibilité des code sources

Pour la majorité des codes développés, il n'y a pas de sensibilité particulière. L'unité s'engage à respecter les principes de l'Open Data (FAIR) et respecte la [Charte des Infrastructures de Recherche](#) d'INRAE.

À l'exception de certains cas :

- Diffusion restreinte : durant la durée du projet
- Confidentiel : code ayant trait au traitement de données personnelles avec un niveau d'impact inférieur ou égal à 2 (nom, prénom, adresse mail professionnelle)

- Garantie de la confidentialité (mesures prises et normes)

Non-concerné

- Analyse de second niveau et garantie de la confidentialité

Non-concerné

IV. Partage des codes sources

Les codes développés dans le cadre de projets avec des financements publics sont déposés sur la [Forge MIA INRAE](#), et rendus accessibles. L'unité recommande l'usage de l'open-access pour les codes sources, et une licence libre restreignant tout usage commercial et modification sans l'accord des auteurs [Creative Commons CC-by-nc-nd 4.0](#).

La réutilisation des codes est possible pour :

- L'usage dans leur fonction de base
- L'implémentation de nouvelles fonctionnalités
- L'installation et les évolutions fonctionnelles

Pour les projets partenariaux, ces règles sont applicables sous réserve d'accords différents avec le partenaire privé dans le cadre d'une convention spécifique.

#### V. Organisation et documentation des codes sources

Point d'amélioration : Elaborer une procédure pour l'archivage des code source (ex. fiche mémo), Former les agents (tuto)

#### VI. Stockage et sécurité des codes sources

- Systèmes d'information

Les données sont stockées sur des serveurs situés dans un datacenter choisi par l'établissement INRAE. L'accès aux données est géré par une authentification déléguée à chaque partenaire de la fédération Edugain.

- Supports de stockage

Une fiche pratique relative au stockage numérique des données (cf. fiche [Mémo#2 « Stocker les données »](#) sur Data@reversaal (Nextcloud)) a été mise en place au sein de l'unité pour éviter la perte de donnée et adopter une démarche de science ouverte. Cette fiche permet aux agents d'identifier le(s) support(s) de stockage le plus adéquat(s) pour leurs données en fonction du stade d'acquisition et de traitement de celles-ci. Elle vise aussi à uniformiser les pratiques au sein de l'unité.

Piste d'amélioration : remise de cette fiche et échange sur le sujet dès l'arrivée d'un nouvel agent.

- Sécurisation du transfert de données

Transferts physiques (clé USB, disque dur externe) : non autorisés.

Transferts réseau (FileSender, NextCloud, [Forge MIA INRAE](#)) : mots de passe, chiffrement et utilisation de protocoles sécurisés.



- Volumétrie actuelle et prévisionnelle

L'UR REVERSAAL bénéficie actuellement d'une capacité totale de stockage de 2 To qui est ajustable chaque année.

- Politique de sécurité de l'entité stockant physiquement les données

La [Politique Sécurité des Systèmes d'Information \(PSSI\)](#) INRAE s'applique à l'entité stockant physiquement les données de l'UR REVERSAAL (hébergement interne INRAE).

- Partage avant publication et confidentialité

L'UR REVERSAAL recommande l'embargo jusqu'à publication, et mis à disposition des codes sources en accès libre. Si exception, se référer au PGD du projet.

- Sécurité, intégrité et traçabilité

Sécurité : protection contre les virus de tous les postes de travail, formation à la [cybersécurité](#), authentification requise pour accéder aux ressources.

Intégrité : se référer à la [charte de déontologie, d'intégrité scientifique et d'éthique d'INRAE](#).

Traçabilité : existence d'un modèle de fichier de métadonnées à disposition de tous les agents.

Point d'amélioration : mettre en place un suivi des modifications pour savoir ce qui est modifié, par qui et quand.

- Sensibilisation aux bonnes pratiques hygiène numérique

Les agents de l'UR REVERSAAL participent aux sensibilisations aux bonnes pratiques d'hygiène numérique effectuées au niveau du centre (fiche, séminaire).

Piste d'amélioration : planifier une intervention régulière du correspondant SSI du centre pour les nouveaux arrivants notamment.

## VII. Archivage et conservation des données

- Identification des codes sources à conserver et de ceux à détruire

Données à conserver : code source de données à financement public archivé sur les plateformes de référence citées précédemment

Données à détruire : code source de données à financement privé, en respectant le PGD du projet

- Plateforme ou procédures pour un archivage pérenne des données

Archivage sur NEXTCLOUD, [Forge MIA INRAE](#), [data.gouv.fr](https://data.gouv.fr)

Pont d'amélioration : basculer vers Forge nationale INRAE début 2025.

- Durée de conservation des codes sources

Variable selon la nature des données et des projets. Se référer au PGD du projet.

- Garantie de financements pour couvrir les coûts associés à la conservation à long terme

Fonds propres à l'unité et fonds provenant de projets de recherche.

## C – Produit “ Données ou codes mis à disposition par un partenaire ”

### I. Présentation générale des données

- Mode d'obtention des données

Données ou codes produits par un tiers et mises à disposition d'INRAE

- Origine

Partenaire extérieur

- Type de données

Données tabulaires (ex. sous forme de tableur)

Image

Texte

Fichier géographique

Code

- Nature des données

Composition bio-physico-chimique (concentrations en polluants, pH, température, indice de décantation...) d'effluents, de boues, de populations bactériennes, de produits chimiques, valeurs mesurées ou calculées

Débit (de gaz, d'effluent), temps de fonctionnement d'équipements, dépenses énergétiques

Schéma de filière de traitement, dimensions associées, liste et caractéristiques d'équipements

Dimensions d'ouvrages (cahier justificatif, cahier des garanties, manuel d'autosurveillance)

Analyse de scénarios prospectifs d'évolution

Avis techniques

Données personnelles (coordonnées téléphoniques, e-mails) d'exploitants ou de maîtres d'ouvrages

Fichiers géographiques (localisation de réseau d'assainissement ou d'autres ouvrages types de sols avec coordonnées GPS, de points de mesures, type de sols)

Mesures hydriques (précipitation, évapotranspiration) et de perméabilité

Codes numériques

...

La liste ci-dessus est non exhaustive.

- Format des données

Les données collectées et produites (analysées et mises en forme) peuvent se présenter sous de multiples formats :

- Données tabulaires : csv, xls, xlsx, odf, R, py
- Image : png, jpeg, pdf, tiff, bmp, vtk, mpeg, avi, mov, mkv ...
- Texte : pdf, odf, doc, docx, txt, ppt, pptx
- Fichiers géographiques : shp, gpkg
- Formats génériques : rar, zip
- Code : R, py

▪ Périmètre thématique des données

- Procédés et filières de traitement de valorisation des eaux urbaines, des boues
- Capteurs et instruments de mesure
- Simulateurs et codes numériques

II. Droit de propriété intellectuelle

L'UR REVERSAAL d'INRAE est en accord avec les principes de l'Open Data et respecte la [Charte des Infrastructures de Recherche](#) d'INRAE. Les données et codes collectés et ceux produits à partir leur analyse sont propriétés d'INRAE. Des droits spécifiques peuvent s'appliquer en cas de partenariat avec des entreprises privées.

III. Sensibilité et confidentialité des données

▪ Identification du niveau de sensibilité des jeux de données

Diffusion restreinte : données scientifiques et techniques, codes numériques, partagés par le tiers

Confidentiel : données personnelles ayant un niveau d'impact inférieur ou égal à 2 (nom, prénom, adresse mail professionnelle)

▪ Garantie de la confidentialité (mesures prises et normes)

Les données et modèles mis à disposition dans le cadre de projets collaboratifs font l'objet d'un accord écrit (ex. accord de confidentialité) signé par le fournisseur et par INRAE, document instruit par le service Ingénierie de Projets et Partenariats (IPP) du Centre INRAE. Cet accord doit expliciter les usages possibles et la période de mise à disposition. Les données ne sont partagées qu'aux membres qui doivent en avoir connaissance. INRAE s'engage à conserver strictement ces données et à ne pas les communiquer à des tiers, même internes à l'unité (hors du projet de recherche dans lequel elles ont été confiées), sans consentement écrit de celui qui a fourni les données.

Point d'amélioration : Diffusion des bonnes pratiques en termes de RGPD au sein de l'unité, à déployer lors de la mise à disposition de données par un tiers.

- Analyse de second niveau et garantie de la confidentialité

Les données et codes sont analysés par un (ou plusieurs) agents de l'unité. Elles ne font pas l'objet d'un partage pour une analyse de second niveau, à l'exception d'un contrat partenarial spécifique (se référer à la convention).

#### IV. Partage des données

- Politique de partage (obligation, restriction, interdiction)

Les données ou codes mis à disposition dans le cadre de projet font l'objet d'un accord écrit (ex. accord de confidentialité) signé par le fournisseur et par INRAE, document instruit par le service Ingénierie de Projets et Partenariats (IPP) du Centre INRAE. Cet accord doit expliciter les usages possibles et la période de mise à disposition. Les données ne sont partagées qu'aux membres qui doivent en avoir connaissance. INRAE s'engage à conserver strictement ces données et codes et à ne pas les communiquer à des tiers, même internes, sans consentement écrit.

L'UR REVERSAAL est soumise aux règles institutionnelles formalisées par la [charte pour le libre accès aux données et aux publications](#). Les données scientifiques et codes numériques produits à partir de données mises à disposition doivent donc être partagés à l'exception d'un contrat partenarial spécifique (se référer à la convention du projet ou au PGD de projet). Pour ce faire, l'unité recommande d'utiliser une licence libre restreignant tout usage commercial et modification sans l'accord des auteurs [Creative Commons CC-by-nc-nd 4.0](#).

Le partage de données confidentielles (données personnelles, coordonnées GPS) est interdit.

- Réutilisations potentielles des données scientifiques ou codes obtenus

L'UR REVERSAAL sollicitera le partenaire au sujet d'une réutilisation des données ou codes pour de nouvelles études ou des études complémentaires. Elle demandera son consentement écrit au partenaire. La réutilisation des données pourrait être inscrite dès la convention avec un partenaire public.

- Modalités de lecture

Les données collectées et produites (analysées et mises en forme) ainsi que les codes, peuvent se présenter sous de multiples formats et donc nécessiter de multiples logiciels pour leur lecture. La liste ci-dessous est non exhaustive :

csv, xls, xlsx, odf : Excel, LibreOffice, OpenOffice  
R : R, Rstudio  
py : Python, Rstudio, PyCharm, Visual Studio Code  
png, jpeg, tiff : Photos, Firefox, XnView, Paint  
mpeg, avi, mov, mkv : Lecteur multimédia, VLC  
pdf : Adobe Reader  
odf, doc, docx : Word, LibreOffice, OpenOffice  
txt : Bloc-notes, Wordpad

ppt, pptx : PowerPoint, LibreOffice, OpenOffice  
shp, gpkg : QGis  
rar, zip : 7-Zip, RAR

- Modalités de partage

Avec qui ?	/Comment ?	Sous quelle licence ?
<b>Porteurs de projets</b>	Stockage réseau : espace NextCloud du projet	L'UR restreint aux membres du projet. Sinon se référer au PGD du projet.
<b>Participants et financeurs</b>	Diffusion par voie électronique (mail, lien de téléchargement) et/ou espace NextCloud du projet et/ou présentation, rapports	L'UR restreint aux membres du projet. Sinon se référer au PGD du projet.
<b>Tous</b>	Entrepôt <a href="http://data.gouv.fr">data.gouv.fr</a>	Sous licence <a href="https://creativecommons.org/licenses/by-nc-nd/4.0/">Creative Commons CC-by-nc-nd 4.0.</a> Si autre, se référer au PGD du projet.

#### V. Organisation et documentation des données

- Méthodes et outils utilisés lors du cycle de vie de la donnée (acquisition, traitement, mise à disposition, archivage, destruction)

Etape	Méthode(s)	Outil(s)
<b>Préparation</b>	Rédaction de l'accord de confidentialité explicitant par écrit les modalités du partage	Word
<b>Acquisition</b>	Réception par email d'un fichier de données, Réception d'un relevé papier Stockage sur un support fiable	Client mail (préférentiellement <a href="http://partage.inrae.fr">partage.inrae.fr</a> ) et filesender
<b>Traitement</b>	Analyse des données	Stockage réseau : NextCloud Tableur Simulateur (ex. Sumo)
<b>Synthèse</b>	Stockage sur un support fiable Rapport / Présentation	Stockage réseau : NextCloud Traitement de texte et création d'un support de présentation
<b>Partage</b>	Diffusion par voie électronique Présentations Entrepôts	Stockage réseau : NextCloud Client mail (préférentiellement <a href="http://partage.inrae.fr">partage.inrae.fr</a> ) et filesender Entrepôts
<b>Archivage / Destruction (à l'issue du projet)</b>	Archivage des données brutes et produites Destruction des données personnelles	Stockage réseau : NextCloud Datacenter

- Métadonnées (nature, production et mise à jour)

Les fichiers de données brutes ou analysées (.xlsx, .shp, etc.) sont accompagnés d'un fichier de métadonnées respectant le format défini par l'UR et basé sur le DUBLIN CORE (<https://www.bnf.fr/fr/dublin-core>). Ce modèle de fichier ([téléchargeable sur ce lien](#)) est accessible à tous les agents de l'UR sur NextCloud : reversaal > donnees@reversaal > plan\_gestion\_donnees (onglet Métadonnées).

Point d'amélioration : diffusion plus large de cette pratique aux agents de l'UR REVERSAAL, choix d'une base de données plus adaptée

- Documentation complémentaire aux métadonnées

Non concerné.

- Gestion et organisation des fichiers de données

Point d'amélioration : mise en place de conventions de nommage de fichier.

- Procédure de contrôle qualité des données

Point d'amélioration : mise en place d'une vérification des données par les membres du projet et son coordinateur.

## VI. Stockage et sécurité des données

- Systèmes d'information

Les données sont stockées sur des serveurs situés dans un datacenter choisi par l'établissement INRAE. L'accès aux données est géré par une authentification déléguée à chaque partenaire de la fédération Edugain.

- Supports de stockage

Une fiche pratique relative au stockage numérique des données (cf. fiche [Mémo#2 « Stocker les données »](#) sur Nextcloud Data@Reversaal) a été mise en place au sein de l'unité pour éviter la perte de donnée et adopter une démarche de science ouverte. Cette fiche permet aux agents d'identifier le(s) support(s) de stockage le plus adéquat(s) pour leurs données en fonction du stade d'acquisition et de traitement de celles-ci. Elle vise aussi à uniformiser les pratiques au sein de l'UR REVERSAAL.

Piste d'amélioration : remise de cette fiche et échange sur le sujet dès l'arrivée d'un nouvel agent.

- Sécurisation du transfert de données

Transferts physiques (clé USB, disque dur externe) : non autorisés.

Transferts réseau (FileSender, NextCloud) : mots de passe, chiffrement et utilisation de protocoles sécurisés.

- Volumétrie actuelle et prévisionnelle

L'UR REVERSAAL bénéficie actuellement d'une capacité totale de stockage de 2 To qui est ajustable chaque année.

- Politique de sécurité de l'entité stockant physiquement les données

La [Politique Sécurité des Systèmes d'Information \(PSSI\)](#) INRAE s'applique à l'entité stockant physiquement les données de l'UR REVERSAAL (hébergement interne INRAE).

- Partage avant publication et confidentialité

L'UR REVERSAAL recommande l'embargo jusqu'à publication. Si exception, se référer au PGD du projet.

- Sécurité, intégrité et traçabilité

Sécurité : protection contre les virus de tous les postes de travail, formation à la [cybersécurité](#), authentification requise pour accéder aux ressources.

Intégrité : se référer à la [charte de déontologie, d'intégrité scientifique et d'éthique d'INRAE](#).

Traçabilité : existence d'un modèle de fichier de métadonnées à disposition de tous les agents.

Point d'amélioration : mettre en place un suivi des modifications pour savoir ce qui est modifié, par qui et quand.

- Sensibilisation aux bonnes pratiques hygiène numérique

Les agents de l'UR REVERSAAL participent aux sensibilisations aux bonnes pratiques d'hygiène numérique effectuées au niveau du centre (fiche, séminaire).

Piste d'amélioration : planifier une intervention régulière du correspondant SSI du centre pour les nouveaux arrivants notamment.



VII. Archivage et conservation des données

- Identification des données à conserver et de celles à détruire

Données à conserver : données scientifiques

Données à détruire : données personnelles

- Plateforme ou procédures pour un archivage pérenne des données

Archivage sur NEXTCLOUD.

Pont d'amélioration : Actualiser la procédure et la fiche mémo, et la diffuser.

- Durée de conservation des données

Variable selon la nature des données et des projets. Se référer au PGD du projet.

- Garantie de financements pour couvrir les coûts associés à la conservation à long terme

Fonds propres à l'unité et fonds provenant de projets de recherche.

## D – Produit “Données d’enquête”

### I. Présentation générale des données

- Mode d’obtention des données

Données générées par consultation de personnes extérieures dans le cadre d’enquêtes

- Origine

Enquête

- Type de données

Données tabulaires (ex. sous forme de tableur)

Image

Texte

Fichier géographique

- Nature des données

Retours d’expérience quantitatifs et qualitatifs

Recueil de besoins

Recueil d’avis

Données personnelles (à vocation de projets scientifiques mais l’item concerne également l’organisation d’évènements, ex. EPNAC)

- Format des données

Les données collectées et produites (analysées et mises en forme) peuvent se présenter sous de multiples formats :

- Dataset : csv, xls, xlsx, odf, R, py
- Image : png, jpeg, pdf, tiff
- Texte : pdf, odf, doc, docx, txt, ppt, pptx
- Fichiers géographiques : shp, gpkg
- Formats génériques : rar, zip

- Périmètre thématique des données

- Procédés et filières de traitement de valorisation des eaux urbaines, des boues
- Capteurs et instruments de mesure
- Simulateurs et codes numériques
- Organisation d’évènements liés à ces thématiques

## II. Droit de propriété intellectuelle

L'UR REVERSAAL d'INRAE est en accord avec les principes de l'Open Data et respecte la [Charte des Infrastructures de Recherche](#) d'INRAE. Les données collectées et celles produites à partir leur analyse sont propriétés d'INRAE. Des droits spécifiques peuvent s'appliquer en cas de partenariat avec des entreprises privées.

## III. Sensibilité et confidentialité des données

- Identification du niveau de sensibilité des jeux de données

Public : données scientifiques (retours d'expérience, besoins, avis)

Confidentiel : données personnelles ayant un niveau d'impact inférieur ou égal à 2 (nom, prénom, adresse mail professionnelle)

- Garantie de la confidentialité (mesures prises et normes)

Pour réaliser leurs enquêtes, les agents de l'unité doivent avoir recours à une offre INRAE (LimeSurvey) qui assure des garanties (pas de transfert/stockage hors UE, accès sécurisé par un identifiant LDAP). Les données personnelles collectées lors des enquêtes réalisées par des agents de l'unité ont un niveau d'impact inférieur ou égal à 2 (nom, prénom et adresse mail professionnelle permettant de recontacter les participants a posteriori). Elles sont néanmoins soumises au RGPD (Règlement Général sur la Protection des Données). La [procédure interne INRAE](#) pour inscrire les projets collectant ce type de données dans un registre central est suivie par les agents de l'unité. Elle permet de garantir la bonne manipulation et la sécurité de ces données grâce à :

- la rédaction d'une fiche informative à destination des participants pour s'assurer de leur consentement éclairé (modèle Ex\_fiche\_info\_recherche\_V12 (12).docx)
- la rédaction d'un engagement de sécurité (conformité des postes de travail où transitent les données, traitement sur le réseau INRAE ou via le VPN, serveurs sécurisés, sensibilisation des agents réalisant les enquêtes, respect de la charte informatique, etc.)

Ces deux documents sont soumis à la Déléguée à la protection des données d'INRAE pour approbation avant inscription au registre.

Point d'amélioration : Diffusion des bonnes pratiques en termes de RGPD au sein de l'unité.

- Analyse de second niveau et garantie de la confidentialité

Les données sont analysées au sein de l'unité et ne font pas l'objet d'un partage pour une analyse de second niveau à l'exception d'un contrat partenarial spécifique (se référer à la convention ou PGD de projet).

## IV. Partage des données

- Politique de partage (obligation, restriction, interdiction)

L'UR est soumise aux règles institutionnelles formalisées par la [charte pour le libre accès aux données et aux publications](#). Les données scientifiques produites lors d'une enquête doivent donc être partagées à l'exception d'un contrat partenarial spécifique (se référer à la convention). L'unité recommande d'utiliser une licence libre restreignant tout usage commercial et modification sans l'accord des auteurs [Creative Commons CC-by-nc-nd 4.0](#).

Rappel : Le partage des données personnelles collectées lors d'une enquête est interdit.

- Réutilisations potentielles des données scientifiques obtenues

Réutilisation pour de nouvelles études ou études complémentaires

- Modalités de lecture

Les données collectées et produites (analysées et mises en forme) peuvent se présenter sous de multiples formats et donc nécessiter de multiples logiciels pour leur lecture. La liste ci-dessous est non exhaustive.

csv, xls, xlsx, odf : Excel, LibreOffice, OpenOffice

R : R, Rstudio

py : Python, Rstudio, PyCharm, Visual Studio Code

png, jpeg, tiff : Photos, Firefox, XnView

pdf : Adobe Reader

odf, doc, docx : Word, LibreOffice, OpenOffice

txt : Bloc-notes, Wordpad

ppt, pptx : PowerPoint, LibreOffice, OpenOffice

shp, gpkg : QGis

rar, zip : 7-Zip, RAR

- Modalités de partage (comment ? avec qui ? sous quelle licence ?)

Avec qui ?	Comment ?	Sous quelle licence ?
<b>Porteurs de projets</b>	Stockage réseau : espace NextCloud du projet	L'UR restreint aux membres du projet. Sinon se référer au PGD du projet.
<b>Participants et financeurs</b>	Diffusion par voie électronique (mail, lien de téléchargement) et/ou espace NextCloud du projet et/ou présentation, rapports	L'UR restreint aux membres du projet. Sinon se référer au PGD du projet.
<b>Tous</b>	Entrepôt <a href="https://data.gouv.fr">data.gouv.fr</a>	Sous licence <a href="#">Creative Commons CC-by-nc-nd 4.0</a> . Si autre, se référer au PGD du projet.

V. Organisation et documentation des données

- Méthodes et outils utilisés lors du cycle de vie de la donnée (acquisition, traitement, mise à disposition, archivage, destruction)

Etape	Méthode(s)	Outil(s)
<b>Préparation</b>	Rédaction du squelette de l'enquête Création de l'enquête	Tableur et traitement de texte LimeSurvey
<b>Acquisition</b>	Ouverture de l'enquête et envoi du lien aux participants Relances Clôture, export et stockage sur un support fiable	LimeSurvey Client mail (préférentiellement partage.inrae.fr) et filesender Stockage réseau : NextCloud
<b>Traitement</b>	Analyse des données Stockage sur un support fiable	Tableur QGis Stockage réseau : NextCloud
<b>Synthèse</b>	Rapport / Présentation Stockage sur un support fiable	Traitement de texte et création de slides Stockage réseau : NextCloud
<b>Partage</b>	Diffusion par voie électronique Présentations Entrepôts	Client mail (préférentiellement partage.inrae.fr) et filesender Entrepôts
<b>Archivage / Destruction (à l'issue du projet)</b>	Archivage des données brutes et produites Destruction des données personnelles	Stockage réseau : NextCloud Datacenter

- Métadonnées (nature, production et mise à jour)

Les fichiers de données brutes ou analysées (.xlsx, .shp, etc.) sont accompagnés d'un fichier de métadonnées respectant le format défini par l'UR et basé sur le DUBLIN CORE (<https://www.bnf.fr/fr/dublin-core>). Ce modèle de fichier ([téléchargeable sur ce lien](#)) est accessible à tous les agents de l'UR sur NextCloud : reversaal > donnees@reversaal > plan\_gestion\_donnees (onglet Métadonnées).

Point d'amélioration : diffusion plus large de cette pratique aux agents de l'UR REVERSAAL, choix d'une base de données plus adaptée

- Documentation complémentaire aux métadonnées

Non-concerné.

- Gestion et organisation des fichiers de données

Point d'amélioration : mise en place de conventions de nommage de fichier.

- Procédure de contrôle qualité des données

Point d'amélioration : mise en place d'une vérification des données par les membres du projet et son coordinateur.

## VI. Stockage et sécurité des données

- Systèmes d'information

Les données sont stockées sur des serveurs situés dans un datacenter choisi par l'établissement INRAE. L'accès aux données est géré par une authentification déléguée à chaque partenaire de la fédération Edugain.

- Supports de stockage

Une fiche pratique relative au stockage numérique des données a été mise en place au sein de l'unité pour éviter la perte de donnée et adopter une démarche de science ouverte. Cette fiche permet aux agents d'identifier le(s) support(s) de stockage le plus adéquat(s) pour leurs données en fonction du stade d'acquisition et de traitement de celles-ci. Elle vise aussi à uniformiser les pratiques au sein de l'UR REVERSAAL.

Piste d'amélioration : remise de cette fiche et échange sur le sujet dès l'arrivée d'un nouvel agent.

- Sécurisation du transfert de données

Transferts physiques (clé USB, disque dur externe) : non autorisés.

Transferts réseau (FileSender, NextCloud) : mots de passe, chiffrement et utilisation de protocoles sécurisés.

- Volumétrie actuelle et prévisionnelle

L'UR REVERSAAL bénéficie actuellement d'une capacité totale de stockage de 2 To qui est ajustable chaque année.

- Politique de sécurité de l'entité stockant physiquement les données

La [Politique Sécurité des Systèmes d'Information \(PSSI\)](#) INRAE s'applique à l'entité stockant physiquement les données de l'UR REVERSAAL (hébergement interne INRAE).

- Partage avant publication et confidentialité

L'UR REVERSAAL recommande l'embargo jusqu'à publication. Si exception, se référer au PGD du projet.

- Sécurité, intégrité et traçabilité

Sécurité : protection contre les virus de tous les postes de travail, formation à la [cybersécurité](#), authentification requise pour accéder aux ressources.

Intégrité : se référer à la [charte de déontologie, d'intégrité scientifique et d'éthique d'INRAE](#).

Traçabilité : existence d'un modèle de fichier de métadonnées à disposition de tous les agents.

Point d'amélioration : mettre en place un suivi des modifications pour savoir ce qui est modifié, par qui et quand.

- Sensibilisation aux bonnes pratiques hygiène numérique

Les agents de l'UR REVERSAAL participent aux sensibilisations aux bonnes pratiques d'hygiène numérique effectuées au niveau du centre (fiche, séminaire).

Piste d'amélioration : planifier une intervention régulière du correspondant SSI du centre pour les nouveaux arrivants notamment.

## VII. Archivage et conservation des données

- Identification des données à conserver et de celles à détruire

Données à conserver : données scientifiques

Données à détruire : données personnelles

- Plateforme ou procédures pour un archivage pérenne des données

Archivage sur NEXTCLOUD.

Pont d'amélioration : Actualiser la procédure et la fiche mémo, et la diffuser.

- Durée de conservation des données

Variable selon la nature des données et des projets. Se référer au PGD du projet.

- Garantie de financements pour couvrir les coûts associés à la conservation à long terme

Fonds propres à l'unité et fonds provenant de projets de recherche.