



HAL
open science

PLAN DE GESTION DE DONNEES Projet SETIER II

Rémi Clément, Stéphanie Prost-Boucle, Sylvain Moreau, Camille Arnault

► **To cite this version:**

Rémi Clément, Stéphanie Prost-Boucle, Sylvain Moreau, Camille Arnault. PLAN DE GESTION DE DONNEES Projet SETIER II. Plan de gestion de donnée du projet SETIER II, pp.1-26, 2024, 10.17180/1b85-mm70 . hal-04712649

HAL Id: hal-04712649

<https://hal.inrae.fr/hal-04712649v1>

Submitted on 27 Sep 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 Projet SETIER II



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

Date : 2024_09_22

DOI PGD : **XXXXX**

INFORMATIONS SUR LE PROJET SETIER II

a. Nom du projet

SETIER II :

b. Identifiant

2024 5624

c. Résumé du projet

La gestion durable des filières d'assainissement repose en outre sur la collecte et l'analyse des données. Cette démarche revêt une importance capitale, dès lors qu'il s'agit de piloter efficacement une station d'épuration ou un processus de traitement. L'avenir de ces installations dépendra de modèles reposant sur des données fiables, exploitant des méthodes d'intelligence artificielle ou de fusion de données. Ces modèles doivent répondre aux exigences environnementales et réglementaires de demain. D'autre part, les petites stations, jusqu'à récemment peu équipées en instrumentation, commencent à se doter de dispositifs adaptés à différents contextes (tels que la REUT et le traitement à la demande, la préservation de l'environnement, le traitement de l'azote et du phosphore), voire simplement pour une meilleure compréhension du fonctionnement de la station d'épuration. L'enquête menée par REVERSAAL dans le cadre du projet SETIER I en 2020 a souligné les difficultés rencontrées sur le terrain concernant l'instrumentation. Les répondants ont évoqué des obstacles tels que le coût des capteurs, mais surtout la programmation de l'obsolescence du matériel, ainsi que les difficultés rencontrées par les utilisateurs pour réparer les centrales (Prost-Boucle et al., 2022). Le projet SETIER II intervient dans un contexte où les pratiques évoluent vers une nouvelle mise en œuvre de l'instrumentation, nécessitant l'acquisition de données via des outils ouverts, réparables et intelligents. Dans cette optique, l'usage d'outils « low-tech » semble opportun (Spinelli et Gottesman, 2019 ; Zhu et al., 2023). Ce concept émergent vise à développer des systèmes de mesure à faible coût, simples, robustes, reproductibles et fonctionnels. Ces instruments open source et open hardware (capteurs et centrales d'acquisition) permettront notamment d'approfondir les

connaissances sur les stations de traitement des eaux usées. Le projet SETIER II ambitionne de démontrer que ces outils peuvent :

- Offrir une souplesse dans les pratiques d'instrumentation (augmentation des fréquences de mesure, diversification des capteurs, différents types de mesures), tout en renforçant les compétences des utilisateurs ;
- Être utilisés pour surveiller le bon fonctionnement des stations d'épuration et les optimiser ;
- Contribuer à la collecte de données enrichissant les bases de données existantes.

Dans le cadre du projet SETIER II, l'objectif est de consolider les avancées des centrales développées dans le projet précédent tout en intégrant de nouveaux paramètres pour répondre aux défis actuels. Cela permettrait d'apporter de nouvelles informations sur la qualité des eaux, mais également de répondre à des enjeux actuels tels que la DERU, les déversements en cas d'orage, ou encore les mesures d'émissions de gaz dans les stations d'épuration. Le projet SETIER I, financé par l'AE RMC et piloté par INRAE REVERSAAL (2019-2023), a prouvé que les centrales d'acquisition low-tech s'adaptent aux technologies classiques de débitmétrie (mesure de hauteur d'eau), de capteurs physico-chimiques (pH, redox, oxygène dissous, température) et de mesures d'énergie (puissance de fonctionnement de moteurs par pinces ampérométriques). Les financeurs et les utilisateurs (agences de l'eau, SATESE, bureaux d'études, collectivités) ont exprimé à INRAE REVERSAAL leur intérêt à poursuivre les études dans ce domaine pour répondre directement aux demandes du terrain, mais aussi pour anticiper les défis futurs liés à la révision de la DERU, à la REUT, et également pour enrichir les bases de données sur le traitement de l'eau là où cela sera nécessaire.

e. Responsabilités du projet

- Rémi Clément et Stéphanie Prost-Boucle – Coordinateurs (REVERSAAL)
- Sylvain Moreau (PROSE)

e. Établissement(s) tutelle(s)

- INRAE

g. Financeur(s)

- Agence de l'Eau Rhône-Méditerranée-Corse (Interlocuteur : Camille Arnault)

INFORMATIONS SUR LE PLAN DE GESTION

DOI (version publiée du plan de gestion) :

Dépôt dans Hal (<https://hal.inrae.fr/>) indiquer le DOI : **XXXXXXXXXXXX**

Historique des versions :

1. PGD_SETIER II_V0_FR, le 20/09/2024

Rédigé par :

Rémi Clément, Stéphanie Prost-Boucle

Relu par :

Sylvain Moreau, Camille Arnault

TYPOLOGIE GENERALE DES DONNEES

Quatre types de produits constituent les données utilisées dans le projet SETIER II.

- 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.

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 des travaux menés dans le projet SETIER II

- 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 (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

Toutes données relatives au projet

- 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, comprenant également les informations métrologique et d’assemblage
- Simulateurs et codes numériques

II. Droit de propriété intellectuelle

Le Projet SETIER II 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 de leur analyse sont la propriété d’INRAE.

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)

REVERSAAL et PROSE 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) membres du projet ayant collecté les données.

IV. Partage des données

- Politique de partage (obligation, restriction, interdiction)

Les données scientifiques produites à partir de données expérimentales ou de simulations seront partagées sous 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.**

L'ensemble des données est **sous embargo jusqu'à la fin du projet SETIER II (juin 2028)**. A l'issue du projet en juillet 2028, l'ensemble des données sera rendu accessible sous un dépôt Gitlab du projet.

- Réutilisations potentielles des données scientifiques obtenues

Au terme du projet et de l'éventuel embargo pour publication, Les membres du projet pourront réutiliser les données acquises lors des expérimentations et simulations qu'ils auront menées.

La réutilisation de données confidentielles (données personnelles, coordonnées GPS) est interdite.

- 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 ?
Membres du projets	Stockage réseau : espace NextCloud du projet	restreint aux membres du projet.
Participants et financeurs	Diffusion par voie électronique (mail, lien de téléchargement) et/ou espace NextCloud du projet et/ou présentation, rapports	restreint aux membres du projet ou financeur
Tous (Juillet 2028)	Entrepôt Gitlab INRAE	Sous licence CC-by-nc-nd 4.0 .

Les licences pourront être réévaluées au cours du projet en accord avec ces membres et le financeur

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)
Préparation	Plan d’échantillonnage	Word
Acquisition	Expérimentations Stockage sur un support fiable	Capteur, mesures et analyses Stockage réseau : NextCloud
Traitement	Analyse des données Stockage sur un support fiable	Tableur Simulateur (ex. Matlab/Comsol) Stockage réseau : NextCloud
Synthèse	Rapport / Présentation Stockage sur un support fiable	Traitement de texte et création d’un support de présentation Stockage réseau : NextCloud
Partage	Diffusion par voie électronique Présentations Entrepôts	Client mail (préférentiellement partage.inrae.fr et filesender)
Archivage / Destruction (à l’issue du projet)	Archivage des données brutes et produites Destruction des données personnelles	Stockage réseau : NextCloud Datacenter ou Gitlab

- 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 un format basé sur le DUBLIN CORE (<https://www.bnf.fr/fr/dublin-core>). .

- Documentation complémentaire aux métadonnées

Non-concerné

- Gestion et organisation des fichiers de données

A définir par les membres du projet

- Procédure de contrôle qualité des données

L’ensemble des données sera validé par les coordinateurs du projet SETIER II, avant partage.

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

Nextcloud, avec dossier projet spécifique, dont l’accès aux données est géré par une authentification

- Sécurisation du transfert de données

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

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

- Volumétrie actuelle et prévisionnelle

Le projet SETIER II bénéficie actuellement d’une capacité totale de stockage de 1 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 du projet SETIER II.

- Partage avant publication et confidentialité

L’ensemble des données du projet sera sous embargo jusqu’à la fin du projet (juillet 2028) ou leur publication par les membres du projet SETIER II.

- 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](#).

- Sensibilisation aux bonnes pratiques hygiène numérique

Les membres du projet seront sensibilisés aux bonnes pratiques d’hygiène numérique effectuées au niveau du stockage du projet SETIER, lors de la réunion de démarrage du projet.

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 ou GITLAB

- Durée de conservation des données

Les données seront conservées sans date limite de destruction, tant que les données seront jugées utiles par les membres du projet.

B – Produit “ Codes sources ”

I. Présentation générale des données

- Mode d'obtention du code source

Code ou programme produit par un membre du projet SETIER II.

- Origine

Développé par les membres du projet SETIER II

- 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, SETIER).

- 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, Arduino

- **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, comprenant également les informations métrologique et d'assemblage
- Métrologie/ instrumentation
- Simulateurs et codes numériques

II. Droit de propriété intellectuelle

Les membres du projet en accord avec les principes de l'Open Data (FAIR) respectent la [Charte des Infrastructures de Recherche](#) . Les codes sources sont la propriété des membres ayant participé à leur développement.

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) ainsi que 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 sous licence à minima [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.

V. Organisation et documentation des codes sources

L'organisation sera définie par les membres du projet.

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

Nextcloud, avec dossier projet spécifique, dont l'accès aux données est géré par une authentification

- 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

Le projet SETIER II bénéficie actuellement d'une capacité totale de stockage de 1 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 du projet SETIER II.

- Partage avant publication et confidentialité

L'ensemble des code du projet sera sous embargo jusqu'à la fin du projet (juillet 2028) ou leur publication par les membres du projet SETIER II.

- 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.

- Sensibilisation aux bonnes pratiques hygiène numérique

Les membres du projet seront sensibilisés aux bonnes pratiques d'hygiène numérique effectuées au niveau du stockage du projet SETIER, lors de la réunion de démarrage du projet.

VII. Archivage et conservation des données

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

Données à conserver : code source de données sur les plateformes de référence citées ci-dessous

- Plateforme ou procédures pour un archivage pérenne des données

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

- Durée de conservation des codes sources

Sans durée et seulement conservé sur les plateformes d'archivage.

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 des membres du projet

- 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

- 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, comprenant également les informations métrologique et d'assemblage
- Métrologie/ instrumentation

II. Droit de propriété intellectuelle

Le projet SETIER II 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.

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 les membres du projet SETIER II. 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. Les membres du projet s'engagent à conserver strictement ces données et à ne pas les communiquer à des tiers, même internes à leur propre unité de recherche (hors du projet de recherche dans lequel elles ont été confiées), sans consentement écrit de celui qui a fourni les données.

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 les coordinateurs du projet. 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. Les membres du projet s'engagent à conserver strictement ces données et codes et à ne pas les communiquer à des tiers, même internes, sans consentement écrit.

- Réutilisations potentielles des données scientifiques ou codes obtenus

Les membres du projet solliciteront 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.

- 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 ?
Porteurs de Membres du projet	Stockage réseau : espace sécurisé NextCloud du projet	NA
Participants et financeurs	Diffusion par voie électronique (mail, lien de téléchargement) et/ou espace NextCloud du projet et/ou présentation, rapports	NA
Tous (Juillet 2028)	Pas de diffusion	Pas de diffusion

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)
Préparation	Rédaction de l'accord de confidentialité explicitant par écrit les modalités du partage	Word
Acquisition	Réception par email d'un fichier de données, Réception d'un relevé papier	Client mail (préférentiellement partage.inrae.fr) et filesender
Traitement	Stockage sur un support fiable Analyse des données	Stockage réseau : NextCloud Tableur Simulateur (ex. Sumo)
Synthèse	Stockage sur un support fiable Rapport / Présentation	Stockage réseau : NextCloud Traitement de texte et création d'un support de présentation
Partage	Stockage sur un support fiable Diffusion par voie électronique Présentations Entrepôts	Stockage réseau : NextCloud Client mail (préférentiellement partage.inrae.fr) et filesender Entrepôts
Archivage / Destruction (à l'issue du projet)	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 basé sur le DUBLIN CORE (<https://www.bnf.fr/fr/dublin-core>).

- Documentation complémentaire aux métadonnées

Non concerné.

- Gestion et organisation des fichiers de données

La gestion et l'organisation des fichiers de données sera définie durant le projet

- Procédure de contrôle qualité des données

Validation avant publication par les coordinateurs du projet SETIER II

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

Nexcloud

▪ 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

Le projet SETIER bénéficie actuellement d'une capacité totale de stockage de 1 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'ensemble des données du projet sera sous embargo jusqu'à la fin du projet (juillet 2028) ou leur publication par les membres du projet SETIER II. Ces données ne seront pas partagées par les membres du projet SETIER II

▪ 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](#).

- Sensibilisation aux bonnes pratiques hygiène numérique

Les membres du projet seront sensibilisés aux bonnes pratiques d'hygiène numérique effectuées au niveau du stockage du projet SETIER, lors de la réunion de démarrage du projet .

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

Nextcloud, avec dossier projet spécifique, dont l'accès aux données est géré par une authentification

- Durée de conservation des données

Les données seront conservées sans date limite de destruction, tant que les données seront jugées utiles par les membres 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 tableau)

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, comprenant également les informations métrologique et d’assemblage
- Métrologie/ instrumentation
- Organisation d’évènements liés à ces thématiques

II. Droit de propriété intellectuelle

Les membres du projet sont 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 la propriété des membres du projet.

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.

IV. Partage des données

- Politique de partage (obligation, restriction, interdiction)

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 ?
Porteurs de projets	Stockage réseau : espace NextCloud du projet	restreint aux membres du projet. NA
Participants et financeurs	Pas de partage	NA
Tous	Pas de partage	NA

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)
Préparation	Rédaction du squelette de l'enquête Création de l'enquête	Tableur et traitement de texte LimeSurvey
Acquisition	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
Traitement	Analyse des données Stockage sur un support fiable	Tableur QGis Stockage réseau : NextCloud
Synthèse	Rapport / Présentation Stockage sur un support fiable	Traitement de texte et création de slides Stockage réseau : NextCloud
Partage	Diffusion par voie électronique	

	Présentations Entrepôts	Client mail (préférentiellement partage.inrae.fr) et filesender Entrepôts
Archivage / Destruction (à l'issue du projet)	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 basé sur le DUBLIN CORE (<https://www.bnf.fr/fr/dublin-core>).

- Documentation complémentaire aux métadonnées

Non-concerné.

- Gestion et organisation des fichiers de données

La gestion et l'organisation des fichiers de données sera définie durant le projet

- Procédure de contrôle qualité des données

Validation avant publication par les coordinateurs du projet SETIER II

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

Nextcloud, avec dossier projet spécifique, dont l'accès aux données est géré par une authentification

- 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

Le projet SETIER II bénéficie actuellement d'une capacité totale de stockage de 1 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 du projet SETIER II (hébergement interne INRAE).

- Partage avant publication et confidentialité

L'ensemble des données du projet sera sous embargo jusqu'à la fin du projet (juillet 2028) ou leur publication par les membres du projet SETIER II.

- 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.

- Sensibilisation aux bonnes pratiques hygiène numérique

Les membres du projet seront sensibilisés aux bonnes pratiques d'hygiène numérique effectuées au niveau du stockage du projet SETIER, lors de la réunion de démarrage du projet

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.

- Durée de conservation des données

Données détruites à l'issue du projet (juillet 2028).