



HAL
open science

Plan de gestion des données (PGD) de l'Observatoire de recherche en environnement sur les poissons diadromes dans les fleuves côtiers (ORE DiaPFC)

Frédéric Marchand, Nadine Renault, Clarisse Boulenger

► To cite this version:

Frédéric Marchand, Nadine Renault, Clarisse Boulenger. Plan de gestion des données (PGD) de l'Observatoire de recherche en environnement sur les poissons diadromes dans les fleuves côtiers (ORE DiaPFC). INRAE. 2024. hal-04600886v3

HAL Id: hal-04600886

<https://hal.science/hal-04600886v3>

Submitted on 29 Aug 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.



Open licence - etalab

PLAN DE GESTION DES DONNEES (PGD) DE L'OBSERVATOIRE DE RECHERCHE EN ENVIRONNEMENT POISSONS DIADROMES DANS LES FLEUVES COTIERS (ORE DIAPFC)



Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "INRA - Trame Structure" fourni par INRAE - Institut national de recherche pour l'agriculture l'alimentation et l'environnement.

Table des matières

.....		Erreur ! Signet non défini.
Plan de gestion des données (PGD) de L'Observatoire de Recherche en Environnement Poissons Diadromes dans les Fleuves Côtiers (ORE DiaPFC)		1
1	InformationS générales	4
1.1	Informations sur la structure.....	4
1.1.1	Nom de la structure	4
1.1.2	Type de structure :	4
1.1.3	Identifiant de la structure :	4
1.1.4	périmètre de la structure	4
1.1.5	Responsabilités dans la structure :	5
1.1.6	Etablissement(s) tutelle(s) :	5
1.1.7	Département de rattachement InraE :	5
1.1.8	Financeurs :	5
1.2	Information sur le document.....	5
1.2.1	Titre du plan	5
1.2.2	Domaine de recherche (selon classification OCDE).....	5
1.2.3	Langue	5
1.2.4	Identification du document	5
1.2.5	Suivi du document.....	6
1.2.6	Plans de gestion en lien avec le projet	6
1.3	Information sur les contacts / rôles	6
2	Présentation générale des données	7
2.1	Mode d'obtention des données	7
2.2	Origine des données	7
2.3	Type de données	7
2.4	Nature des données.....	7
2.4.1	Poissons :	7
2.4.2	Conditions environnementales :	7
2.5	Format des données	8
2.6	Périmètre thématique des données	8
3	Droits de propriété intellectuelle	8
4	Confidentialité	8
5	Partage des données	8
5.1	Quelles sont les réutilisations potentielles de ces données	8
5.2	Comment les données sont-elles partagées :	8

5.3	Sous quelle licence ?	9
6	Organisation et documentation des données.....	9
6.1	Quels méthodes et outils sont utilisés pour acquérir et traiter les données, depuis leur acquisition jusqu'à leur mise à disposition, leur archivage ou leur destruction ?	10
6.1.1	Données Poissons	10
6.1.2	Données environnementales.....	11
6.2	métadonnées	12
6.2.1	Données Poissons	12
6.2.2	Données environnementales.....	13
6.3	Documentation complémentaire	13
6.4	Comment les fichiers de données sont-ils gérés et organisés : contrôle des versions, conventions de nommage des fichiers, organisation des fichiers ?	13
6.4.1	Données Poissons	13
6.4.2	Données environnementales.....	14
6.5	Qualité des données	14
6.5.1	Poissons:.....	14
6.5.2	Données environnementales.....	15
7	Stockage et sécurité des données	15
7.1	Les supports de données :.....	15
7.2	Volumétrie.....	15
7.2.1	Base Posgresql de l'ORE.....	15
7.2.2	SharePoint de l'ORE	16
7.2.3	NextCloud	16
7.3	Politique de sécurité.....	16
7.4	Sécurité - Intégrité - Tracabilité :.....	16
8	Archivage et conservation des données	17
8.1	Archivage.....	17
8.2	Conservation.....	17
8.3	durée de conservation des données ?	17
8.4	Garantie de financements à la conservation à long terme ?	17

1 INFORMATIONS GENERALES

1.1 INFORMATIONS SUR LA STRUCTURE

1.1.1 NOM DE LA STRUCTURE

Observatoire de recherche sur les poissons diadromes dans les fleuves côtiers (ORE DiaPFC)

1.1.2 TYPE DE STRUCTURE :

OBSERVATOIRE DE RECHERCHE EN ENVIRONNEMENT (ORE)

1.1.3 IDENTIFIANT DE LA STRUCTURE :

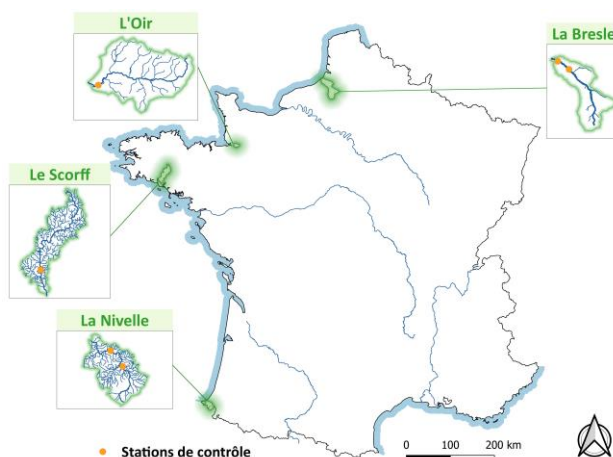
ORE DiaPFC

1.1.4 PERIMETRE DE LA STRUCTURE

La problématique de l'Observatoire de Recherche sur les Diadromes dans les Fleuves Côtiers (ORE DiaPFC) est d'étudier l'évolution des populations de poissons diadromes sous l'effet des changements environnementaux (changement climatique et changements liés à l'agriculture) qui affectent ces cours d'eau.

L'ORE DiaPFC est focalisé sur les poissons grands migrateurs amphihalins (saumon, truite, anguille, aloses, lamproies...). Ces espèces, très fragilisées par l'action de l'homme, sont une composante emblématique de la biodiversité des fleuves côtiers. Ces derniers sont aujourd'hui les principaux refuges des poissons diadromes qui ont considérablement régressés sur les hydrosystèmes fluviaux de plus grande taille.

Cette infrastructure de Recherche s'appuie sur 4 cours d'eau côtiers de la façade Manche Atlantique, L'Oir et la Bresle Normandie (Manche et Seine Maritime), le Scorff en Bretagne (Morbihan), la Nivelle dans le Pays Basque (Pyrénées Atlantiques). Ces quatre rivières, équipées de stations de de contrôle des poissons diadromes, font l'objet de suivis biologiques et physicochimiques récurrents depuis début des années 80. A ces cours d'eau, sont associées des installations expérimentales à Rennes et à Saint-Pée-sur-Nivelle. L'ensemble est complété par des simulateurs démo génétiques individu-centré pour l'expérimentation in silico sur des populations virtuelles (non traité dans ce PGD). Les données d'échantillonnage d'écaillés ou d'autres tissus de poissons alimentent le Centre de Ressources Biologique Colisa (CRB Colisa).



Les sites ateliers sont rattachés à 2 unités : l'UMR ECOBIOP (Site atelier La Nivelle) et l'U3E (Oir, Scorff et Bresle). L'OFB, via le pôle R&D sur les Migrateurs AMphihalins dans leur Environnement (Pôle MIAME) est un partenaire essentiel de cet ORE

Site web de l'ORE DiaPFC : <https://diapfc.hub.inrae.fr/>

1.1.5 RESPONSABILITES DANS LA STRUCTURE :

Nom,	Prénom	Courriel	Rôle
PREVOST,	Etienne	etienne.prevost@inrae.fr	Animateur scientifique de l'ORE DiaPFC
MARCHAND,	Frédéric	frederic.marchand@inrae.fr	Animateur opérationnel de l'ORE DiaPFC
BOULENGER,	Clarisse	clarisse.boulenger@inrae.fr	Responsable des données
AYMES,	Jean- Christophe	jean-christophe.aymes@inrae.fr	Responsable de l'IE ECP
LANGÉ,	Frédéric	frederic.lange@inrae.fr	Responsable du site de la Nivelles
JOSSET,	Quentin	quentin.josset@ofb.gouv.fr	Responsable du site de la Bresle
JEANNOT,	Nicolas	nicolas.jeannot@inrae.fr	Responsable du site du Scorff
TREMBLAY,	Julien	Julien.tremblay@inrae.fr	Responsable du site de l'Oir
RENAULT,	Nadine	nadine.renault@inrae.fr	Responsable des SI

1.1.6 ETABLISSEMENT(S) TUTELLE(S) :

INRAE - Institut national de recherche en agriculture, alimentation et environnement

OFB - Office Français de la Biodiversité

1.1.7 DEPARTEMENT DE RATTACHEMENT INRAE :

ECODIV- Écologie et biodiversité des milieux forestiers, prairiaux et aquatiques

1.1.8 FINANCEURS :

INRAE - Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement

OFB - Office Français de la Biodiversité

Europe via le FEAMPA (Data collection Framework)

1.2 INFORMATION SUR LE DOCUMENT

1.2.1 TITRE DU PLAN

Plan de gestion des données (PGD) de l'observatoire DiaPFC

1.2.2 DOMAINE DE RECHERCHE (SELON CLASSIFICATION OCDE)

Earth and related environmental sciences

1.2.3 LANGUE

Français

1.2.4 IDENTIFICATION DU DOCUMENT

Identifiant : [hal-04600886](https://hal.archives-ouvertes.fr/hal-04600886)

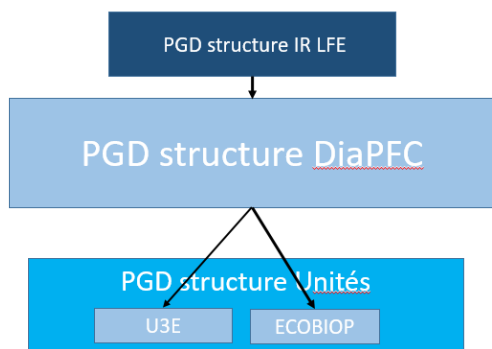
Type d'identifiant : HAL

1.2.5 SUIVI DU DOCUMENT

Date	N° de version	Auteur	Validé par	Description
2019-11-16	1	Frédéric Marchand		Création du document avec DMP Opidor
30/01/2019	2	F. Marchand		
29/05/2023	3	F. Marchand	F. Marchand	Validation DMP Opidor
06/02/2024	4	N. Renault		Version soumise au comité d'expert du département
12/03/2024	5	N. Renault	E. Prévost	Prise en compte des remarques du comité expertise département
26/04/2024	6	N. Renault	E. Prévost	Version publiée dans HAL le 10/06/2024
29/08/2024	7	N. Renault		Ajout de la référence HAL de ce document et des PGD en lien avec ce document

1.2.6 PLANS DE GESTION EN LIEN AVEC LE PROJET

- PGD du projet "IR LIFE"
- PGD de l'unité U3E ([hal-04674519](https://hal.inrae.fr/hal-04674519)) et ECOBIOP



1.3 INFORMATION SUR LES CONTACTS / ROLES

Nom	Affiliation	Rôles
Clarisse Boulenger	INRAE	- Personne contact pour les données
Frédéric Marchand	INRAE	- Coordinateur du projet

		- Responsable du plan de gestion de données
Nadine Renault	INRAE	- Responsable des systèmes d'informations

2 PRESENTATION GENERALE DES DONNEES

2.1 MODE D'OBTENTION DES DONNEES

- Données générées par la structure
- Données produites par un tiers (BD Topage pour représenter géographiquement les cours d'eau)

2.2 ORIGINE DES DONNEES

Données issues d'observations : ce sont des données de suivis à long terme. Les données de l'ORE DiaPFC portent sur les grands poissons migrateurs (saumon, truite de mer, anguille, aloses et lamproies). Ce sont des données sur la phénologie (migration ...), la biométrie (taille, poids ...) et les données concernant les conditions environnementales (qualité de l'eau, de l'habitat ...). L'ORE DiaPFC est composé de 4 sites répartis sur la façade atlantique, la Nivelle au Pays Basque, le Scorff en Bretagne et l'Oir et la Bresle en Normandie. Les poissons diadromes sont suivis depuis 1983 pour la Nivelle, la Bresle et l'Oir et 1993 pour le Scorff.

Des analyses sclérochronologiques ou génétiques peuvent être effectuées sur les échantillons afin de déterminer l'âge et/ou le sexe.

2.3 TYPE DE DONNEES

- Dataset

Données biologiques sur les migrateurs amphihalins et conditions environnementales (climatiques et physico-chimiques).

- Collection

Collecte d'écaillés et de nageoires pour le CRB Colisa

2.4 NATURE DES DONNEES

2.4.1 POISSONS :

- Événement de capture : Date, méthode, localisation (coordonnées géographiques), opérateurs (initiales), utilisation produits vétérinaires ...
- Biométrie : Espèce, Longueur, Poids, stade phénologique, sexe ...
- Prélèvements d'échantillons
- Marquage collectif ou individuel
- Données d'analyses : Âge, sexe ...

2.4.2 CONDITIONS ENVIRONNEMENTALES :

- Événement de prélèvement : Date, localisation (coordonnées géographiques), opérateurs (initiales)
- Données issues d'observations lors des campagnes de pêche (hauteur d'eau, conditions environnementales ...),
- Données enregistrées par des capteurs (T°C, turbidité, pluviométrie ...)
- Données physico_chimie : mesures de paramètres physiques sur des prélèvements d'eau dans les cours d'eau (température, pH, conductivité, turbidité, Matières en suspension (MES)) et chimiques (teneurs en ions en solution : ammonium NH4+, nitrates NO3-, nitrites NO2- et orthophosphates PO43-).

2.5 FORMAT DES DONNEES

- Texte (.txt, csv),
- Excel
- Access
- PostgreSQL

2.6 PERIMETRE THEMATIQUE DES DONNEES

- Climate
- Biodiversity and Ecology
- Water resources

3 DROITS DE PROPRIETE INTELLECTUELLE

Les droits sur les données et les autres informations créées sont la propriété conjointe de INRAE et de l'OFB.

4 CONFIDENTIALITE

Toutes nos données sont géo-référencées à 7 m de précision et pourraient permettre de localiser une parcelle et son propriétaire. La qualité des données géographiques mise en ligne (DOI) est volontairement dégradée pour les données de qualité d'eau.

Les données contiennent des informations nominatives et sont donc soumises à la RGPD.

Les bases de données contiennent des données professionnelles sur les opérateurs. Ce sont les initiales qui permettent de faire le lien entre les données et l'opérateur. Les données relatives à un opérateur sont ses Initiales, nom et prénom, statut (expert, non expert ...).

Les données opérateurs ne sont pas diffusées. En cas de nécessité, la base de données permet de les anonymiser via un identifiant.

5 PARTAGE DES DONNEES

La plupart des données de l'observatoire de Recherche en Environnement ORE DiaPFC a vocation à être partagée. Ce sont des données publiques financées par des fonds publics. Les données de marquage individuel peuvent avoir un embargo pour laisser le temps aux scientifiques de la structure de publier leur projet lorsqu'ils sont à l'initiative de ce marquage.

Nous ne diffusons pas ou peu des données concernant des espèces sensibles ou à forte valeur patrimoniale, notamment l'emplacement d'espèces rares comme les moules perlières.

5.1 QUELLES SONT LES REUTILISATIONS POTENTIELLES DE CES DONNEES

Ce sont des données à long terme (plus de 35 ans de données) rares avec un potentiel usage autour de la qualité des milieux aquatiques, la gestion des populations des espèces étudiées ou l'impact du changement climatique sur ces espèces.

5.2 COMMENT LES DONNEES SONT-ELLES PARTAGEES :

Les données non publiées dans des entrepôts de données sont accessibles sur demande auprès de l'administrateur de la base de données ou des référents des sites ateliers dans le format souhaité.

La plupart de ces données respectent le principe FAIR (Findable, Accessible, Interoperable, Reusable). Les bases de données PostgreSQL sont stockées sur le serveur de base de données de l'U3E sur le Datacenter INRAE de Toulouse. Les utilisateurs des unités constitutives de l'ORE DiaPFC y ont accès à partir du login et mot de passe du service d'annuaire LDAP INRAE et après avoir été déclaré dans la base par les administrateurs. La consultation est possible quelque soit le site atelier à partir de l'outil de gestion de bases de données de son choix (Pgadmin, Dbeaver, Access,

Base, etc.). L'insertion ou la modification des données pour un site atelier donné est limitée aux personnes autorisées (gestion à partir de rôle).

Les bases Access sont stockées sur un serveur de fichiers INRAE dont l'accès est géré par le service d'annuaire LDAP. Tous les utilisateurs des unités constituantes de l'ORE DiaPFC y ont accès, en lecture et en modification.

Quelque soit le site atelier, les données concernant le saumon subissent le même traitement (script R, GNU GPL v3.0 license) pour être agrégées. La synthèse de ces résultats est mise à disposition sur le Github de l'ORE (<https://github.com/ORE-DiaPFC>). Il est prévu de migrer ces données sur le Gitlab institutionnel.

Certains jeux de données et métadonnées sont également disponibles sur des entrepôts de données et sont identifiés par un DOI :

- Les données de biodiversité sont publiées annuellement sur le site du [GBIF, Global Biodiversity Information Facility](#).
- Les données sur les conditions environnementales sont déposées sur le portail Data Gouv (dataverse ORE) et mises à jour annuellement. (<https://entrepot.recherche.data.gouv.fr/dataverse/orediapfc>)

Les données de marquage de saumon et truite (tag) sont diffusées sur le portail [Salmo PIT Tag DB](#)

Les données phénologiques sont publiées sur le portail [TEMPO](#) (réseau national d'observatoires dédiés à la phénologie de l'ensemble du règne vivant (espèces végétales et animales, exploitées et sauvages)).

La liste des données publiées est visible sur le site web de l'ORE <https://diapfc.hub.inrae.fr/collection-donnees/jeux-de-donnees>

Quelques exemples :

- Jeannot, N., Azam, D., Guilloux, Y., and Prévost, E. 2023. Abundances indices and biological traits of juvenile salmon sampled in the Scorff river (France). Institut National de la Recherche Agronomique (INRA). <https://www.gbif.org/dataset/89064e3a-aa3c-495d-b236-092e1dae7042>
- Josset, Q., Flesselle, A., Bernardin, A., Macquet, T., Petit, L., Robin, E., Bras, R.L., Jean-Louis Fagard, Fournel, F., Euzenat, G., Beaulaton, L., Prévost, E., 2022a. Phenology and biological traits of migrating trout and sea trout (*Salmo trutta*) sampled by trapping in the survey in the Bresle river (France). <https://doi.org/10.15468/JKPYA8>
- Marchand, F., Azam, D., Delanoë, R., Destouches, J.-P., Tremblay, J., Baglinière, J.-L., Nevoux, M., *et al.* 2023. Abundances and biological traits of fish sampled by electrofishing (except specific abundance indices) on the River Oir since 1988 (France). Institut National de la Recherche Agronomique (INRA). <https://www.gbif.org/dataset/6242ca69-c65e-4723-b629-9aeb05b51c88>
- Tremblay, J., Marchand, F., Azam, D., Delanoë, R., Destouches, J.-P., Prévost, E., 2022a. Phenology and biological traits of migrating river lamprey sampled by electrofishing in the Oir river (France). <https://doi.org/10.15468/JCHNPY>
- Ollitrault, M., and Jeannot, N. 2019. Physico-chemical parameters of river water in Scorff and its tributaries (Morbihan county - France). Portail Data Inra. <https://data.inra.fr/citation?persistentId=doi:10.15454/GAXTX6>

5.3 SOUS QUELLE LICENCE ?

Les jeux de données publiés sur le GBIF sous une licence [Creative Commons Attribution Non Commercial CC BY 4.0 License](#).

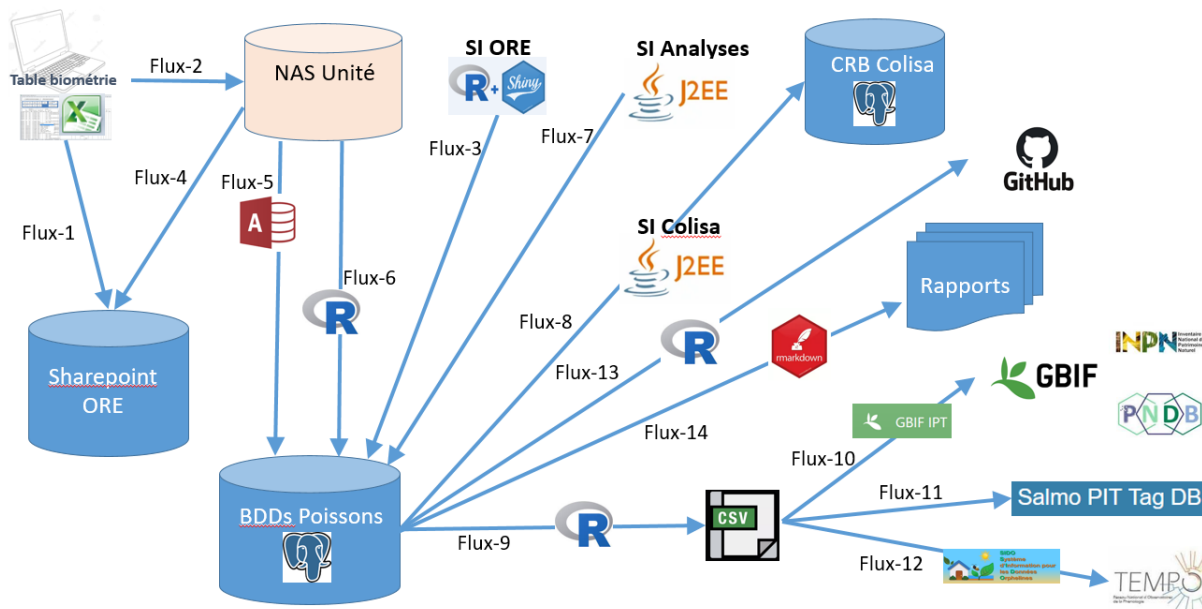
Tous les jeux de données publiés sur le portail Data Gouv sont sous [licence etalab-2.0](#).

6 ORGANISATION ET DOCUMENTATION DES DONNEES

La collecte des données est assurée par le personnel en place sur chaque site de l'ORE DiaPFC. La saisie est également de leur ressort. Aucune donnée n'a été archivée ou détruite car c'est l'intérêt de ces jeux de données d'être sur le long terme.

6.1 QUELS METHODES ET OUTILS SONT UTILISES POUR ACQUERIR ET TRAITER LES DONNEES, DEPUIS LEUR ACQUISITION JUSQU'A LEUR MISE A DISPOSITION, LEUR ARCHIVAGE OU LEUR DESTRUCTION ?

6.1.1 DONNEES POISSONS



Les données sont acquises à l'aide de méthodes standardisées et/ou documentées pour chaque site atelier. Voir les PGD ORE-U3E, PGD ORE-IECP

Les données sont saisies à l'aide d'une table de biométrie associée à une application (Excel avec macro) permettant de standardiser (utilisation de référentiel, vocabulaire commun...) et de réaliser des premiers contrôles de la qualité des données. [La table de biométrie est référencée dans les actifs numériques d'INRAE.](#)

Flux-1 : Les données ainsi acquises sont synchronisées ou déposées sur le Sharepoint de l'ORE.

Flux-2 : Les fichiers de données sont ensuite compilés

Flux-3 : Les fichiers compilés sont intégrés par les responsables de site dans la base de données PostgreSQL à l'aide de l'interface web accessible en interne et avec authentification <https://orediapfc.inrae.fr> (Shiny/R). L'authentification est basée sur l'accès à la base de données Postgres. [Ce système d'information est référencé dans les actifs numériques d'INRAE.](#) L'insertion se passe en deux phases, une phase de vérification de la cohérence de données, puis une seconde phase avec l'insertion en base. Si lors de la première phase, des anomalies sont repérées (erreur sur les données de référence, doublons suréchantillon, etc.), l'utilisateur modifie son fichier en entrée (fichier communément appelé fichier compilé/corrigé) et le resoumet à l'interface.

Flux-4 : Une fois le fichier intégré, l'utilisateur le dépose sur le Sharepoint de l'ORE.

Flux-5 : Les modifications/correction à la marge sont du ressort des unités mais les agents de l'ORE ont à disposition des formulaires Access reliés à la base Postgres. Les formulaires Access sont tous identiques mais sont reliés aux données propres à chaque site (une base Access par site atelier).

Flux-6 : Les modifications en masse sont gérées par les administrateurs de la base de données via des scripts R.

Flux-7 : La saisie des analyses sur les échantillons pour caractériser l'âge et le sexe des poissons est commune à l'ensemble des sites et est réalisée à partir d'une interface WEB accessible en interne <https://sinalyse.u3e.inrae.fr/> (J2EE). [Ce système d'information est référencé dans les actifs numériques d'INRAE.](#)

Flux-8 : Les données d'échantillons sont publiées sur le site colisa.fr à partir d'une interface d'insertion disponible après authentification sur le site (Menu échantillon/Intégration/Depuis la base de données). Les utilisateurs autorisés peuvent insérer/mettre à jour les échantillons du site atelier dont ils sont responsables. [Le SI Colisa est référencé dans les actifs numériques d'INRAE.](#)

Flux-09 : Des fichiers CSV pour les jeux de données publiées sont générés à partir de requêtes SQL spécifiques selon les modèles demandés par les portails :

- Flux-10 : Utilisation de l'interface [Integrated Publishing Toolkit \(IPT\)](#) du GBIF pour déposer les jeux de données au format CSV. (Création de deux identifiants de connexion (N. Renault et F. Marchand) demandés auprès du GBIF France (Sophie Pamerlon).
- Flux-11 : Dépôt du fichier CSV sur le portail [Salmo PIT Tag DB](#) dans l'onglet Register Deployed Tags. Aucune identification n'est nécessaire.
- Flux-12 : Utilisation de l'outil [SIDO](#) pour le dépôt sur le portail TEMPO. L'identification sur le portail via le Idap INRAE est possible.

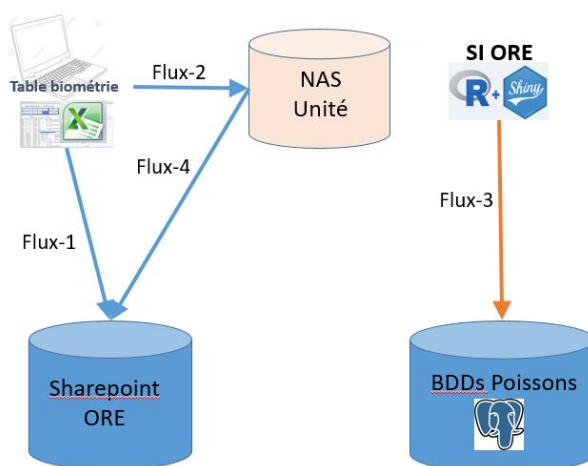
Flux-13 : Des estimations d'abondances sont réalisées dans un cadre statistique bayésien dont les résultats et les scripts sont mis à disposition sur Github (<https://github.com/ORE-DiaPFC/Abondance>).

Flux-14 : Un modèle de rapport automatique a été défini pour l'ORE. A l'heure actuelle, seul le rapportage sur le site de la Bresle est opérationnel, celui de l'Oir est en cours de développement. Ce rapportage est réalisé avec R Markdown.

Chaque site atelier est autonome sur la gestion des données générées via les antennes RFID. Voir les PGD de structure des unités.

6.1.2 DONNEES ENVIRONNEMENTALES

6.1.2.1 DONNEES ISSUES D'OBSERVATIONS LORS DES CAMPAGNES DE PECHE (HAUTEUR D'EAU, CONDITIONS ENVIRONNEMENTALES ...)



Flux-1 et 2 : Ces données sont saisies via la table de biométrie et sont stockées dans une feuille séparée des données de pêches.

Flux-3 : Ces données ne sont pas encore intégrées dans la base ORE, développement en cours sur le SI ORE.

Flux- 4 : Les données sont sur le sharepoint de l'ORE dans les fichiers compilés qui ont permis l'intégration des données poissons. La structure la base Postgres est prête à accueillir ces données.

6.1.2.2 DONNEES PHYSICO_CHIMIE

Chaque site atelier est autonome sur la manière de saisir les données de terrain (voir PGD d'unités).

Les modes opératoires pour chaque méthode d'analyse sont identiques pour tous les sites atelier sur le nexcloud INRAE.

- [MO-PRO-PHY-01 Dosage de l'ammonium](#)
- [MO-PRO-PHY-02 Dosage du phosphate](#)
- [MO-PRO-PHY-03 Dosage des Nitrates NO3](#)

Toutes les analyses des sites de l'ORE sont effectuées sur Rennes. L'ensemble des analyses est saisi dans des fichiers Excel centralisés sur le NAS de l'unité U3E avec un répertoire par site atelier.

Un schéma dédié aux données physico-chimie a été créé dans la base Postgres de l'ORE et les données ont été migrées jusqu'en 2020. L'objectif à court terme est de migrer toutes les données dans cette base Postgres et de proposer une interface de saisie commune pour le terrain avec une application web d'intégration.

6.1.2.3 DONNEES ENREGISTREES PAR DES CAPTEURS (T°C, TURBIDITE, PLUVIOMETRIE ...)

Chaque site atelier est autonome sur la gestion des données capteurs. Voir les PGD d'unité.

Toutefois, quand les capteurs seront reliés au réseau LoRaWAN, la procédure sera identique pour tous les sites ateliers. Actuellement, un seul capteur de ce type est déployé sur le site du Scorff.

Les données sont transmises automatiquement à des passerelles LoRaWAN qui les relaient vers un serveur réseau. En cas de non disponibilité du serveur réseau LoRaWAN, les passerelles sont en mesure de garder les enregistrements pendant 48H et de les retransmettre à la disponibilité du serveur LoRaWAN.

Toutes les données sont ensuite récupérées sur le serveur LoRaWAN (cool-life.u3e.inrae.fr [194.167.76.152]) via le protocole MQTT par l'application ThingsBoard, hébergée sur une autre machine virtuelle (code-life.u3e.inrae.fr [194.167.76.151]). Ces 2 machines virtuelles sont pour l'instant hébergées sur un serveur de la DSI de l'Institut Agro Rennes-Angers. A terme elles seront migrées vers le Datacenters INRAE de Toulouse.

Les données enregistrées sont accessibles sur un tableau de bord via une interface web en lecture seule uniquement et avec une connexion au VPN INRAE ([lien vers le tableau de bord des niveaux d'eau et de température pour le Scorff](#)).

6.2 METADONNEES

Les métadonnées sont communes à l'ensemble des sites.

6.2.1 DONNEES POISSONS

Métadonnées	Description - Format	Standard, associés	Vocabulaires
Date et Heure	aaaa-mm-jj hh:mm:ss	ISO 8601	
Protocole	Lien vers mode opératoire	Interne	
Secteur	Localisation du lieu de capture	Interne	
Cours d'eau	Cours d'eau du lieu de capture	SANDRE et Interne	
Coordonnées géographiques	Latitude et longitude en Degrés Décimaux	WGS84	
Taxon	Niveau taxonomique le plus précis possible	EML et TAXREF	
Longueur	Longueur en mm		
Poids	Poids en g		
Sexe	Femelle, mâle, indéterminé ou non recherché	SANDRE	
Age	Distinction âge de rivière et âge de mer pour les diadromes	Interne	
Maturité	Etat de maturité au moment de la capture	SANDRE	
Marquage individuel	PITS	ISO FDXB ou TROVAN	

6.2.2 DONNEES ENVIRONNEMENTALES

Métadonnées	Description - Format	Standard, associés	Vocabulaires
Date et Heure	aaaa-mm-jj hh:mm:ss	ISO 8601	
Protocole	Lien vers mode opératoire	Interne	
Secteur	Localisation du lieu de capture	Interne	
Cours d'eau	Cours d'eau du lieu de capture	SANDRE	
Coordonnées géographiques	Latitude et longitude en Degrés Décimaux	WGS84	
Nitrates	NO3- en mg/L	SANDRE	
Nitrites	NO2- en mg/L	SANDRE	
Ammonium	NH4+ en mg/L	SANDRE	
Orthophosphates	PO4--- en mg/L	SANDRE	
Température de mesure du pH	TemppH en °C	SANDRE	
Température de mesure de la conductivité	TempCond en °C	SANDRE	
Turbidité	Turbidité en NTU	SANDRE	
Conductivité	Cond en µS/cm	SANDRE	
Potentiel en Hydrogène	pH en pH	SANDRE	
Pluviométrie	mm de pluie par m²	SANDRE	

Les métadonnées sur les conditions environnementales sont décrites dans le dataverse INRAE-U3E (<https://entrepot.recherche.data.gouv.fr/dataverse/u3e>) et mises à jour annuellement

6.3 DOCUMENTATION COMPLEMENTAIRE

La description des tables, attributs, fonction, etc. est systématiquement réalisée dans Postgres.

Une documentation spécifique pour les données physico-chimiques est disponible sur le sharepoint de l'ORE dans le dossier [Documents/Physico-chimie](#)

6.4 COMMENT LES FICHIERS DE DONNEES SONT-ILS GERES ET ORGANISES : CONTROLE DES VERSIONS, CONVENTIONS DE NOMMAGE DES FICHIERS, ORGANISATION DES FICHIERS ?

6.4.1 DONNEES POISSONS

Les fichiers issus de la table de biométrie sont stockés :

- 1) sur l'ordinateur dédié à la prise de mesure,
- 2) une carte SD nommée "BIOMETRIE" sur l'ordinateur dédié à la prise de mesure
- 3) sauvegardé sur les NAS d'unités et éventuellement sur le site Sharepoint de l'ORE (copie ou synchronisation) sur les sous sites "Données brutes" et "Données compilées/corrigées" (sites de l'Oir et du Scorff).

Tous les fichiers issus de la table de biométrie suivent une convention de nommage : site atelier_date la plus ancienne dans le fichier au format AAAAMMJJ_date la plus récente dans le fichier au format AAAAMMJJ_listes des méthodes dans le fichier séparé par "_ " "Biometrie" _ numero de version.

Exemple : OIR_20200127_20200129_PGM_PGD_Biometrie_01.xlsx

Les fichiers générés par la table de biométrie peuvent être compilés. Les fichiers compilés suivent également une convention de nommage : site atelier + date la plus ancienne dans le fichier + date la plus récente dans le fichier + listes des méthodes dans le fichier + "Biometrie" + numero.

Exemple : OIR_20200127_20200129_PGM_PGD_compil_01.xlsx

Les différentes versions du fichier de saisie Table de biométrie sont disponibles sur le sharepoint de l'ORE dans le dossier « [Table de biométrie](#) ». Un document word retrace les modifications de chaque version. Une procédure d'installation de ce fichier et le paramétrage des sauvegardes est consultable sur le sharepoint [ORE DiaPFC/Table de biometrie](#). Une nouvelle table de biométrie est fabriquée par le bureau d'étude StreamInnov (partenariat). Les versions de fichiers Excel pour ce type de table et la documentation sont stockées dans le dossier [Streaminnov](#).

Les interfaces Access pour accéder aux données de la base sont disponibles en téléchargement sur le sharepoint de l'ORE dans le sous site « Données » et dans le dossier propre à chaque site atelier.

Les scripts de correction en masse des données sont stockés sur le NAS de l'U3E et de plus en plus sur le projet Gitlab dédié à la base de données (<https://forgemia.inra.fr/pole-migrateurs/ore-diapfc/BDPoisson>) accessible seulement aux administrateurs de la base.

Les codes des applications web de l'ORE (SI-ORE et SI-ANALYSES) sont versionnés sur le gitlab Forgemia du pôle pour la gestion des migrateurs amphihalins dans leur environnement OFB-INRAE-Institut Agro-UPPA (MIAME) :

- L'application web pour l'intégration des données : https://forgemia.inra.fr/pole-migrateurs/ore-diapfc/BDPoisson_shiny.
- L'application web pour la saisie des analyses : <https://forgemia.inra.fr/pole-migrateurs/si-analyses>.

Les images liées aux analyses des échantillons d'écaillés de chaque site atelier sont réalisées au sein des unités et les données générées à cette occasion sont stockées de manière spécifique selon les unités (voir les PGD ORE-U3E, PGD ORE-IECP).

Les requêtes sql et fichiers xml pour export vers le GBIF sont stockés dans le sharepoint [ORE DiaPFC/Documents/jeux de données](#).

Le mode opératoire d'insertion des données pour le portail Salmo PIT Tag DB, les requêtes sql et le programme R sont stockés sur le sharepoint de [l'ORE ORE DiaPFD/Documents/TAG-DB](#).

Les scripts SQL pour le portail TEMPO sont sur le NAS de l'U3E (P:\BDPOISSON\TEMPO). Accès limité.

Les résultats et les scripts concernant les estimations d'abondances sont mis à disposition sur Github (<https://github.com/ORE-DiaPFC/Abondance>). »

6.4.2 DONNEES ENVIRONNEMENTALES

Chaque site atelier est autonome sur la manière de saisir les données de terrain, aussi, il n'y a pas d'espace partagé à l'ORE pour stocker les données brutes (voir PGD d'unités).

Les fichiers de saisie des analyses sont identiques pour chaque site atelier. Ils sont stockés sur le NAS de l'U3E avec un répertoire par site (Z:\Physico_Chimie). Il existe un fichier Excel par site atelier dans lequel il y a une feuille par année. Il n'y a pas de versionnement des fichiers.

6.5 QUALITE DES DONNEES

6.5.1 POISSONS:

Un premier contrôle de la qualité des données est assuré via le fichier de table de biométrie. Des listes déroulantes reprenant les tables de références de la base de données sont utilisées. La saisie automatique des pit-tag, de la longueur et du poids limitent également les erreurs de saisie. Des contrôles de cohérence sont appliqués lors de la saisie sur certains paramètres tels que la taille et le poids en fonction des espèces, du stade. Un message demandant la confirmation de la valeur apparaît si elle est en dehors des plages définies.

Les balances sont contrôlées chaque semaine sur les stations de contrôle et à chaque déplacement lors des inventaires piscicoles (un message d'alerte s'affiche à chaque ouverture du fichier table de biométrie).

Sur un des sites ateliers de l'ORE DiaPFC (Oir), un script (R) a été développé pour vérifier la cohérence de l'ensemble des paramètres à partir de l'analyse statistique des 30 dernières années. Un fichier de sortie du script indique les incohérences observées. L'observateur peut donc corriger lorsque cela est possible à partir des données papiers. La prochaine étape sera d'appliquer un code de validité issu du script R (vérifié par le script x ou pas de vérification).

6.5.2 DONNEES ENVIRONNEMENTALES

Des scripts (R) pour la visualisation de graphiques avant la publication des données permet de vérifier à minima les points qui sortent des boxplots. L'observateur peut donc vérifier si ce sont des erreurs de saisie ou autres et corriger lorsque cela est possible à partir des données papiers (avant 2019).

Chaque unité est autonome sur le matériel à contrôler et la fréquence (voir PGD des unités) toutefois des modes opératoires sont partagés sur Nextcloud :

- [MO-MAT-PLU-03 Vérification des pluviomètres](#)
- MO-MAT-TMP-03 Méthode de contrôle des thermomètres enregistreur

7 STOCKAGE ET SECURITE DES DONNEES

Chaque unité est responsable du stockage et de la sauvegarde des données brutes sachant qu'elles sont déposées sur des serveurs de fichiers dont la sauvegarde est automatiquement assurée par les Directions des Systèmes d'Informations (INRAE ou OFB)

L'administrateur des systèmes d'information de l'ORE assure toute la partie technique des outils communs à l'ORE (Sharepoint, Postgres, Gitlab) et s'assure des sauvegardes.

Par ailleurs les jeux de données sont ou seront déposés dans des entrepôts de données (GBIF, Dataverse ...) qui en garantissent l'accès et la sauvegarde. Une sauvegarde sur nos propres serveurs est réalisée en parallèle.

7.1 LES SUPPORTS DE DONNEES :

- Papier : Données saisies sur papier
- Portable terrain : Ordinateur ou tablette de terrain
- Poste de travail : Ordinateur de bureau
- Capteurs autonomes : Capteurs enregistreurs autonomes de différents paramètres (T°, conductivité, pluviométrie ...).
- Navettes de transport : Utilitaires de terrain permettant le déchargement des capteurs autonomes puis de les transporter vers un ordinateur.
- Carte SD : Carte SD permettant de sauvegarder une copie supplémentaire avec l'utilisation de la table de mesure automatique des poissons.
- Gitlab [forgemia \(https://forgemia.inra.fr/pole-migrateurs/ore-diapfc/BDPoisson\)](https://forgemia.inra.fr/pole-migrateurs/ore-diapfc/BDPoisson). Sauvegarde et versionnement du SI ORE, des scripts SQL et du rapportage automatique.
- GitHub (<https://github.com/ORE-DiaPFC>). Sauvegarde et versionnement des données d'estimation.
- Le serveur sharepoint de l'ORE (<https://sites.inra.fr/site/ore-diapfc/donnees/SitePages/Accueil.aspx>) permet d'héberger l'ensemble des données. Des sous sites sharepoint ont été créés pour classer les données par site atelier et par type de données (« Données brutes », « Données compilées/corrigées », « Données scannées »).
- Base de données Postgresql de l'ORE (base db_u3e : schema oir, bresle, scorff, nivelle, analyses et env), (pgeodb.u3e.inrae.fr) sur le datacenter de toulouse
- Bases de données Access : interface d'accès à la base de données Postgres.

7.2 VOLUMETRIE

7.2.1 BASE POSTGRESQL DE L'ORE

- Oir : 2 Go
- Scorff : 160 Mo
- Bresle : 250 Mo
- Nivelle : 38 Mo

- Analyses : 300 Mo

7.2.2 SHAREPOINT DE L'ORE

- Données brutes/compilées/scannées : 6 Go
- Jeux de données : 1,5 Go

7.2.3 NEXTCLOUD

- Photos d'analyses d'échantillons : 100 Go

Une réflexion est engagée pour optimiser le stockage des photos, que garde t-on ? etc.

7.3 POLITIQUE DE SECURITE

Toutes les données de l'ORE sont stockées sur des infrastructures de INRAE donc sécurisées selon la « [Charte des infrastructures de recherche à INRAE](#) ».

Des personnes ressources en informatique (PRI) par unité et un administrateur de base de données assurent toute la partie technique de ces supports et flux et la sauvegarde en interne.

Les sauvegardes de la base de données Posgresql sont automatisées. Elles sont décrites sur le Git du pole MIAME (<https://forgemia.inra.fr/pole-migrateurs/ore-diapfc/BDPoisson/-/wikis/Sauvegarde-et-restauration>). Un script est exécuté tous les soirs à 23h :

- sauvegarde en local avec pg_dump de la base
- copie du fichier obtenu par le pg_dump sur [le serveur capacitif](#) de l'U3E (<https://iaas.dcidf.inra.fr/dashboard/project/containers/>, bucket savepggeodb)
- suppression des fichiers datant de plus de 15 jours sur le serveur capacitif

En parallèle l'outil pgbackrest permet d'effectuer une sauvegarde totale le dimanche à 22h, une sauvegarde différentielle le soir à 22h sauf le dimanche et une sauvegarde incrémentale toutes les heures entre 9h et 18h du lundi au vendredi. Toutes ces sauvegardes sont stockées sur le serveur capacitif de l'U3E (<https://iaas.dcidf.inra.fr/dashboard/project/containers/>, bucket savepggeodb) dans le répertoire var/lib/pgbackrest.

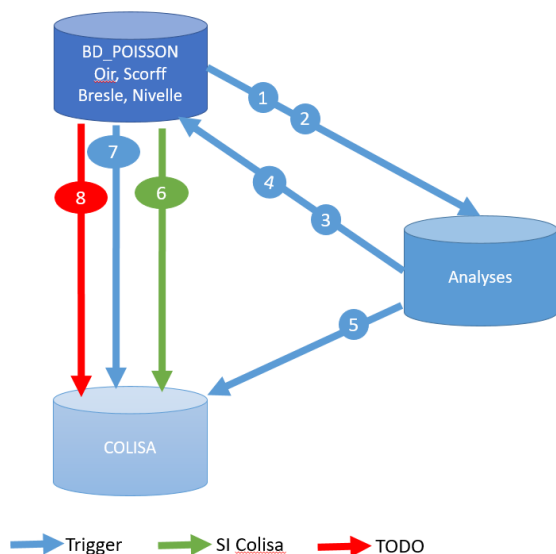
Tous les jeux de données déposés sur les portails GBIF, TEMPO et TAG-DB sont également conservés sur le sharepoint de l'ORE.

7.4 SECURITE - INTEGRITE - TRACABILITE :

Voir la [Charte des infrastructures de recherche à INRAE](#)

Concernant les données « Poissons », toutes les données brutes (fichiers Excel et feuilles de terrain scannées) sont conservées. La version de chaque fichier intégré via le SI ORE est conservée.

Des triggers sont mis en place sur la base de données Postgres afin d'assurer l'intégrité des données entre les données poissons des sites ateliers, les analyses et colisa.



1. Création d'une analyse de type observation avec en paramètre le protocole et le type de résultat. Trigger déclenché à l'insertion d'une capture.
2. Suppression des analyses liées à une capture supprimée.
3. Vérification de la validité de l'analyse après insertion d'une analyse et modification de la capture s'il s'agit d'une analyse de sexe ou d'âge.
4. Mise à jour de la table capture suite à une suppression d'une analyse.
5. Mise à jour des analyses dans Colisa suite à insertion ou suppression d'une analyse
6. Insertion des échantillons de l'ORE dans Colisa à partir de l'interface (<https://colisa.fr/>)
7. Suppression d'un échantillon dans colisa s'il est rendu indisponible (perdu) ou supprimé dans la base poisson
8. TODO Si modif d'une variable dans une capture avec un échantillon, mettre à jour COLISA

Le SI-Analyse permet de garder la traçabilité de toutes les analyses effectuées sur les échantillons de l'ORE.

Trois champs sont systématiquement intégrés dans les tables de la base de données Postgres (date d'insertion, date de mise à jour et utilisateur qui a fait la mise à jour). Une historisation des données est également mise en place sur les tables stratégiques de la base.

La publication et l'identification par DOI des jeux de données permet également d'assurer un minimum de traçabilité.

8 ARCHIVAGE ET CONSERVATION DES DONNEES

8.1 ARCHIVAGE

Les serveurs où sont hébergées les données sont maintenus par INRAE. Les données déposées sur des serveurs externes (GBIF, PNDB) peuvent être produites à nouveau à partir de nos bases de données sur les serveurs INRAE.

8.2 CONSERVATION

Toutes les données ont vocation à être conservées sur le long terme car c'est pour cela qu'elles sont collectées.

8.3 DUREE DE CONSERVATION DES DONNEES ?

Indéfinie.

8.4 GARANTIE DE FINANCEMENTS A LA CONSERVATION A LONG TERME ?

Soutiens INRAE, OFB (pôle MIAME), FEAMPA (DCF)