



HAL
open science

Plan de gestion des données (PGD) du Centre de ressource biologique (CRB) Colection of ichthyological samples (COLISA)

Nadine Renault, Frédéric Marchand

► To cite this version:

Nadine Renault, Frédéric Marchand. Plan de gestion des données (PGD) du Centre de ressource biologique (CRB) Colection of ichthyological samples (COLISA). INRAE. 2024. hal-04678572v1

HAL Id: hal-04678572

<https://hal.inrae.fr/hal-04678572v1>

Submitted on 27 Aug 2024 (v1), last revised 29 Aug 2024 (v2)

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) DU CENTRE DE RESSOURCE BIOLOGIQUE (CRB) COLECTION OF ICHTYOLOGICAL SAMPLES (COLISA)



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

Plan de gestion des données (PGD) du Centre de ressource biologique (crb) COlection of Ichthyological SAmples (colisa)	1
1 Information 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.....	4
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.....	5
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	6
2.1 Mode d'obtention des données	6
2.2 Origine des données	6
2.3 Type de données	6
2.4 Nature des données.....	7
2.5 Format des données	7
2.6 Périmètre thématique des données	7
3 Droits de propriété intellectuelle.....	7
4 Confidentialité	7
5 Partage des données.....	7
5.1 Quelles sont les réutilisations potentielles de ces données	7
5.2 Comment les données sont-elles partagées	8
5.3 Sous quelle licence ?	8
6 Organisation et documentation des données.....	8
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 ?	8

6.2	métadonnées	9
6.3	Documentation complémentaire	10
6.4	Comment les fichiers de données sont-ils gérés et organisés : contrôle des versions, conventions de nommage des fichiers, organisation des fichiers ?	10
6.5	Qualité des données	10
7	Stockage et sécurité des données	10
7.1	Les supports de données.....	11
7.2	Volumétrie.....	11
7.2.1	Base Posgresql « db_u3e », schéma « collection »	11
7.2.2	SharePoint de COLISA	11
7.3	politique de sécurité	11
7.4	Sécurité - Intégrité – Tracabilité	11
8	Archivage et conservation des données	12
8.1	Archivage	12
8.2	Conservation.....	12
8.3	durée de conservation des données ?	12
8.4	Garantie de financements à la conservation à long terme ?	12

1 INFORMATION GENERLES

1.1 INFORMATIONS SUR LA STRUCTURE

1.1.1 NOM DE LA STRUCTURE

[COLISA](#) (COLlection of Ichthyological SAmples)

1.1.2 TYPE DE STRUCTURE

Centre de ressources biologiques (CRB).

1.1.3 IDENTIFIANT DE LA STRUCTURE

<https://doi.org/10.15454/D3ODJM>

1.1.4 PERIMETRE DE LA STRUCTURE

Le CRB [Colisa](#) (COLlection of Ichthyological SAmples) est un Centre de ressources biologiques labellisé par le GIS [IBISA](#) (Infrastructures en biologie, santé et agronomie) en 2018. Le CRB Colisa est intégré dans 2 Infrastructures de Recherche, l'IR LIFE, « Living In Freshwater and Estuaries » et l'IR RARE, Ressources Agronomiques pour la Recherche (Pilier Environnement [BRC4env](#))

Depuis les années 70, différentes unités de l'INRAE ainsi que l'OFB prélèvent des échantillons sur de nombreux spécimens aquatiques. Ces échantillons, généralement des écailles ou des nageoires, servent ensuite à la communauté scientifique pour effectuer différentes analyses et recherches. Pour effectuer la surveillance des populations de saumons, de truites, mais aussi d'autres espèces moins connues, les scientifiques interviennent sur tous les bassins hydrographiques de France. Notre catalogue Colisa référence ces différents échantillons de tissus durs collectés sur les sites d'études.

Les échantillons d'écailles et otolithes proviennent :

- des programmes de recherche et de suivis à long terme menés par les unités INRAE U3E à Rennes et UMR ECOBIOP à Saint-Pée-sur-Nivelle dans le cadre de l'[ORE DiaPFC](#) ; de l'UMR CARRTEL à Thonon-les-Bains dans le cadre du [SOERE OLA](#) (Observatoire des lacs alpins) et Kerguelen.
- de la déclaration des captures de salmonidés (pêche à la ligne et professionnelle) traitées par le CNICS (Centre national d'interprétation des captures de salmonidés) dans le cadre du [Pôle Poissons migrateurs](#).

Cette collection se caractérise par son envergure historique (plus de **60 ans**), par la nature et la quantité des échantillons (plus de **480 000 échantillons** de **d'une quarantaine d'espèces** de poissons et lamproies), ainsi que par leur origine géographique (territoire national).

1.1.5 RESPONSABILITES DANS LA STRUCTURE

Nom,	Prénom	Courriel	Rôle
Marchand,	Frédéric	frederic.marchand@inrae.fr	Responsable du CRB et responsable scientifique Rennes
Starck,	Agnès	agnes.starck@inrae.fr	Responsable technique Rennes Responsable communication
Josset,	Quentin	quentin.josset@ofb.gouv.fr	Responsable scientifique Bresle
Flesselle,	Aurélie	aurelie.flesselle@ofb.gouv.fr	Responsable technique Bresle
Renault,	Nadine	nadine.renault@inrae.fr	Responsable Système d'information et gestion des données
Daverat,	Françoise	francoise.daverat@inrae.fr	Responsable scientifique Saint-Pée-sur-Nivelle
Lange,	Frédéric	frederic.lange@inrae.fr	Responsable technique Nivelle

Guéraud, François	francois.gueraud@inrae.fr	Responsable technique Kerguelen
Hamelet, Valérie	valerie.hamelet@inrae.fr	Responsable technique Thonon
Goulon, Chloé	chloe.goulon@inrae.fr	Responsable Scientifique Thonon
Martignon, Mélanie	melanie.martignon@inrae.fr	Responsable Qualité

1.1.6 ETABLISSEMENT(S) TUTELLE(S)

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

OFB - Office National pour la Biodiversité

1.1.7 DEPARTEMENT DE RATTACHEMENT INRAE

- ECODIV- Écologie et biodiversité des milieux forestiers, prairiaux et aquatiques
- AQUA - Écosystèmes aquatiques, ressources en eau et risques

1.1.8 FINANCEURS

- INRAE - Institut national de recherche en agriculture, alimentation et environnement
- OFB - Office National pour la Biodiversité

1.2 INFORMATION SUR LE DOCUMENT

1.2.1 TITRE DU PLAN

Plan de gestion des données (PGD) du CRB COLISA

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 : à venir

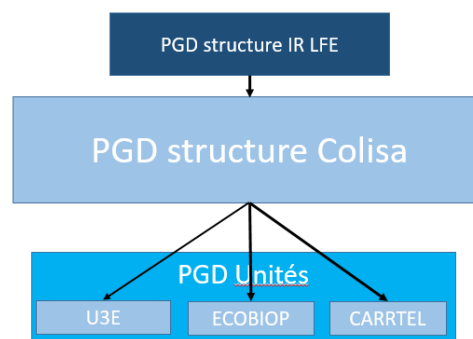
Type d'identifiant : HAL

1.2.5 SUIVI DU DOCUMENT

Date	N° de version	Auteur	Validé par	Description
30/01/2019	00	Frédéric Marchand		Création du document sur Opidor
10/08/2023	01	Nadine Renault	F Marchand	
08/02/2024	02	Nadine Renault	F Marchand	
20/06/2024	03	Nadine Renault		Prise en compte des remarques du comité expertise département (C. Pichot, A Fiocca, A Bénard)
26/08/2024	04	Nadine Renault	M Martignon	Relecture
27/08/2024	05	Nadine Renault	N Renault	Ajout des liens vers les guides scalimétrie et prise en compte remarque de Mélanie

1.2.6 PLANS DE GESTION EN LIEN AVEC LE PROJET

- PGD U3E : [hal-04674519](https://hal.archives-ouvertes.fr/hal-04674519)
- PGD CARRTEL :
- PGD ECOBIOP :
- PGD du projet "IR LIFE" :
- PGD ORE DiaPFC : [hal-04600886](https://hal.archives-ouvertes.fr/hal-04600886)



1.3 INFORMATION SUR LES CONTACTS / ROLES

Nom	Affiliation	Rôles
Frédéric Marchand	INRAE U3E	<ul style="list-style-type: none">• Coordinateur du projet• Personne contact pour les données (Colisa-Commun, Colisa-U3E)• Responsable du plan de gestion de données
Nadine Renault	INRAE U3E	<ul style="list-style-type: none">• Responsable du système d'information

2 PRESENTATION GENERALE DES DONNEES

2.1 MODE D'OBTENTION DES DONNEES

- Données générées par la structure
- Données produites par un tiers

Les données associées aux échantillons proviennent des unités INRAE U3E, Ecobiop et CARRTEL ainsi que de l'OFB (Pôle MIAME).

2.2 ORIGINE DES DONNEES

- Analyse
- Observation
- Expérimentation

Depuis 2021, le type, la date et le protocole de toutes les analyses effectuées sur un échantillon, sans donner le résultat de chacune, sont affichées dans Colisa.fr. Ces données proviennent du [SI analyses de l'ORE \(SI référencé dans les actifs numériques d'INRAE\)](#). Quand une analyse est déclarée comme référente par un expert, le résultat est directement stocké dans les métadonnées de l'échantillon (age, sexe, ecaille régénérée, etc.)

2.3 TYPE DE DONNEES

- Collection
- Dataset
- Image

Ce sont principalement des données biologiques caractérisant les échantillons du CRB Colisa. Des images sont également régulièrement acquises et permettent de nouvelles analyses. Depuis 2022, des données climatiques sont récoltées pour mesurer les conditions de conservation des échantillons.

2.4 NATURE DES DONNEES

- Données de capture : Date, lieu, technique, ...
- Données biométrique : Taille, poids, sexe, pathologie ...
- Données d'analyses : Âge, croissance ..
- Données climatiques : capteurs positionnés dans les salles de stockage pour mesurer les conditions de conservation des échantillons.

2.5 FORMAT DES DONNEES

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

2.6 PERIMETRE THEMATIQUE DES DONNEES

- Biodiversity and Ecology
- Climate

3 DROITS DE PROPRIETE INTELLECTUELLE

Ce sont des données publiques financées sur fonds publics.

4 CONFIDENTIALITE

Toutes les données sont géo référencées à 7-10 m de précision. Les données couplées aux coordonnées GPS ne révèlent pas d'un comportement ou une pratique du propriétaire de la parcelle.

Le SI Colisa enregistre des données professionnelles sur les gestionnaires des collections qui permettent de les contacter si besoin.

Lors des demandes d'emprunt ou d'expertise adressées au CRB Colisa, les données personnelles du demandeur sont enregistrées. Elles sont utilisées à des fins d'échanges entre le CRB et le demandeur, et pour l'envoi des échantillons ou données demandés.

Sur le catalogue en ligne, la qualité des données géographiques est volontairement dégradée lors de la visualisation en ligne (200m).

5 PARTAGE DES DONNEES

Ce sont des données publiques financées par des fonds publics. Elles ont vocation à être partagées. Nous ne diffusons pas ou peu des données concernant des espèces sensibles ou à forte valeur patrimoniale.

5.1 QUELLES SONT LES REUTILISATIONS POTENTIELLES DE CES DONNEES

Au plan scientifique, cette collection contribue à la caractérisation actuelle et rétrospective des individus ou des populations de poissons, au regard des traits d'histoire de vie, de l'évolution de la diversité génétique de ces populations et des modifications des conditions environnementales dans lesquelles elles évoluent.

Au plan finalisé, elle participe à la définition des opérations de protection, de gestion des populations exploitées : définition de quotas (TACs), périodes d'ouverture de pêche, etc. De même, les informations contenues et recherchées au travers des échantillons permettent de revisiter les actions de gestion des populations, dont certaines sont protégées (liste rouge de l'UICN) ou considérées comme patrimoniales.

Ces données sont sur le principe FAIR (Findable, Accessible, Interoperable, Reusable).

5.2 COMMENT LES DONNEES SONT-ELLES PARTAGEES

Les données sont stockées dans une base de données Postgresql (serveur pgeodb.u3e.inrae.fr sur le datacenter INRAE de Toulouse, dans la base db_u3e, dans le schéma « collection »). Les utilisateurs y ont accès à partir d'un login et mot de passe du service d'annuaire LDAP INRAE et après avoir été déclarés dans la base par les administrateurs. Les droits sont gérés par les administrateurs de la base.

Les données sont également accessibles au travers d'un catalogue en ligne ou de bases de données internationales :

- Catalogue [Colisa](#) en ligne avec une interface web, les données sont exportables par l'utilisateur au format csv. Les utilisateurs doivent s'inscrire via l'interface avant de pouvoir accéder aux données (les agents INRAE doivent également s'inscrire).
- Système d'information de l'IR [RARE](#). Mise à jour des données une fois par an.
- Publication de jeux de données dans le Global Biodiversity Information Facility ([GBIF](#)) et le Global Genome Biodiversity Network ([GGBN](#)) pour les espèces modèles une fois par an.
- Les données des capteurs LoraWan sont accessibles via l'[interface de visualisation des données de capteurs](#). Seules les personnes sur le réseau INRAE y ont accès.

5.3 SOUS QUELLE LICENCE ?

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

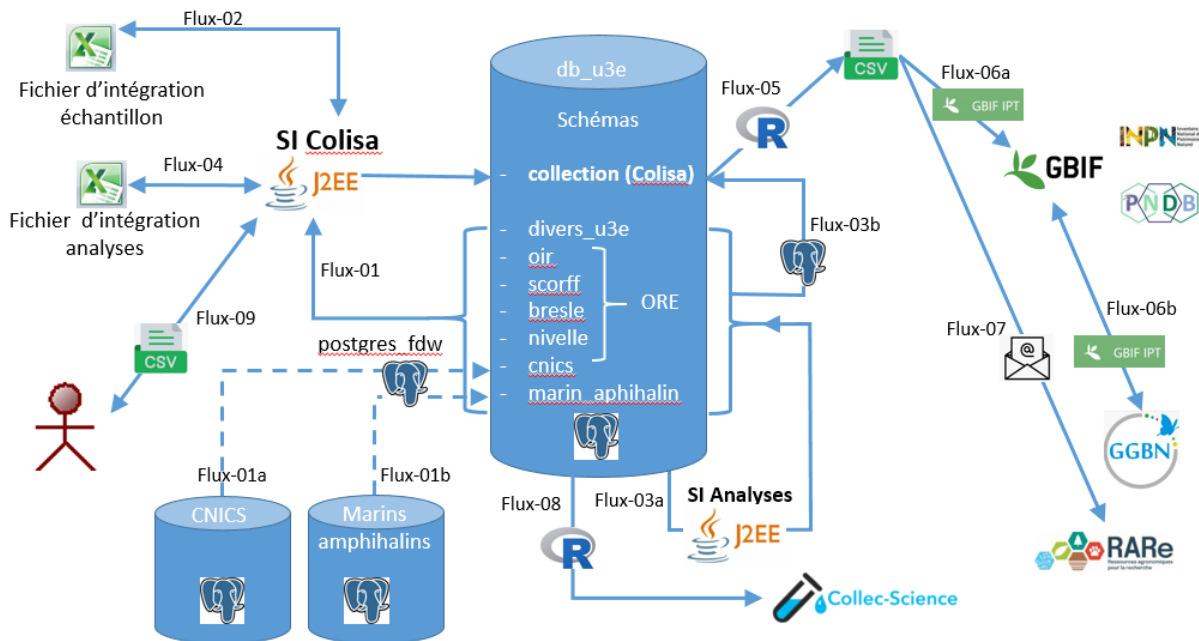
Les jeux de données publiés sur le portail RARE sont sous [licence etalab-2.0](#).

6 ORGANISATION ET DOCUMENTATION DES DONNEES

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 ?

La collecte des échantillons n'est pas assurée par le CRB Colisa.

Les données sont intégrées dans Colisa via l'interface d'intégration du SI Colisa ([colisa.fr](#)). [Le SI Colisa est référencé dans les actifs numériques d'INRAE](#). Il est nécessaire d'être connecté et avoir les droits « manager ». Toutes les données sont stockées dans le schéma « collection » de la base db_u3e.



Deux possibilités d'intégration :

- Flux-01 : A partir de la base de données de l'U3E (db_u3e) :
 - Flux-01 : échantillons de l'ORE DiaPFC : schémas oir, scorff, bresle, nivelle
 - Flux-01 : autres échantillons de l'U3e : schéma divers_u3e
 - Flux-01a : échantillons du CNICS : schéma cnics (vue des échantillons de la base db_capture construite à partir de l'extension datawrapper de Postgres)
 - Flux-01b : échantillons des marins amphihalins : schéma marins_amphihalins (vue des échantillons de la base db_captures_marins_amphihalins construite à partir de l'extension datawrapper de Postgres)
- Flux-02 : A partir d'un fichier modèle pour tous les autres échantillons (Kerguelen, CARTELL, programmes de recherche, ...). Le modèle du fichier d'import peut être téléchargé à partir de l'interface web Colisa.fr après s'être connecté.

Les données d'analyses (données d'ageage, de sexe et analyses réalisées) peuvent être intégrées de deux façons :

- Flux-03a – Flux-03b : Pour les échantillons de l'ORE et tous les échantillons détenus par l'U3E, le Système d'information SI Analyses permet de saisir les résultats des analyses effectuées sur les échantillons et d'en garder la traçabilité (<https://sinalyse.u3e.inrae.fr/>, [référéncé dans les actifs numériques d'INRAE](#)). Toutes les nuits un trigger est déclenché pour mettre à jour Colisa.
- Flux-04 : Pour tous les autres échantillons, l'intégration se fait à partir d'un fichier modèle. Ce fichier modele est disponible en ligne sur Colisa.fr après s'être connecté.

Flux-05 : Des vues ont été créées dans le schéma « collection » pour permettre l'extraction des données dans des fichiers texte selon les modèles fournis par les portails de diffusion.

- Flux-06 : GGBN (Global Genome Biodiversity Network) :
 - Flux-06a : extraction des 2 vues dédiées et dépôt des fichiers sur le portail du GBIF via leur outil d'intégration (Integrated Publishing Toolkit (IPT), <https://ipt.gbif.fr/>). Un mail est ensuite envoyé à support@biocase.org (GGBN) avec l'url colisa sur l'ipt (https://ipt.gbif.fr/resource?r=fish_sample_collection) afin de les informer de la mise à jour des données et récupérer la nouvelle archive. La procédure est disponible sur disponible sur le Sharepoint de Colisa (<https://sites.inrae.fr/site/colisa>).
 - Flux-06b : le GGBN récupère l'archive sur l'IPT du GBIF
- Flux-07 : RARe : envoi du fichier csv de la vue dédiée par mail.

Flux-08 : La gestion des échantillons (identification du lieu de stockage, mouvements, prêts, etc.) est assurée, pour la majorité des unités, par le logiciel Collec-science. Une instance de collec-science a été créée par unité sur le serveur collec-science d'INRAE. Toutes les données présentes dans la base db_u3e sont automatiquement versées dans la base de Collec-science via un script R utilisant les API Collec-Science. Voir les PGD d'unité.

Flux-09 : Le SI Colisa met à disposition un outil d'extraction par critères de recherche avec la possibilité de télécharger les données dans un fichier CSV. Il permet également d'enregistrer les demandes de prêts et d'en assurer le suivi.

6.2 METADONNEES

Les métadonnées sont mises à jour à partir de l'interface colisa.fr.

Métadonnées	Description- format	Standard, Vocabulaires associés
Date et Heure	aaaa-mm-jj hh:mm:ss	ISO 8601
Secteur	Localisation du lieu de capture	Interne
Cours d'eau	Cours d'eau du lieu de capture	SANDRE
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

6.3 DOCUMENTATION COMPLEMENTAIRE

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

Le CRB Colisa est certifié ISO 9001, dans ce cadre un site Sharepoint a été créé (<https://sites.inra.fr/site/colisa/>) avec tous les documents et liens utiles au fonctionnement du CRB.

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

Toutes les données intègrent une BDD Postgres. Il n'existe pas de versionnement de la BDD. Il n'existe pas de convention de nommage des fichiers.

Attention : les noms des 2 fichiers d'extraction pour le GGBN ne doivent pas être modifiés sinon le jeu de données sera considéré comme nouveau lors du dépôt sur l'IPT du GBIF. Ils portent le nom des vues avec l'extension txt (collection_v_ggbn_espece_modele.txt et collection_v_ggbn_espece_modele_relationship.txt)

6.5 QUALITE DES DONNEES

Des contrôles sur les données sont réalisés en amont par les structures qui fournissent les échantillons et données associées.

La qualité des données issues des analyses d'âge ou de croissance réalisées par Colisa est assurée par l'organisation de sessions d'intercalibration entre experts de l'ageage du CRB. Ces sessions ont lieu 2 à 3 fois par an pour une formation en continu et l'harmonisation des pratiques. Les images des échantillons utilisés pour cette intercalibration sont conservés dans le Sharepoint Colisa (<https://sites.inrae.fr/site/colisa/>) et servent d'échantillons de références. Par ailleurs, l'utilisation de listes pré-définies dans la base de données limite les erreurs de saisie.

Une [procédure de calibration](#) de matériel a été mise en place pour l'ensemble des sites et est disponible sur le Sharepoint Colisa.

Un indice de qualité allant de 1 à 4 a été mis en place à partir de 2021 pour la lecture d'âge des poissons du CNICS. Chaque lecteur donne une note puis indique un niveau de confiance de son interprétation.

1	Ecailles régénérées et non exploitables, aucune possibilité de donner un âge
2	Difficile de donner un âge (écaille régénérées, manque d'expérience, écailles peu lisibles) mais on suppose que...
3	Forte probabilité que l'âge donné soit le bon mais subsiste un doute : à comparer avec d'autres écailles du même type, demander un autre avis
4	Age simple à identifier

Deux guides ont été rédigés pour aider à l'interprétation de l'âge des salmonidés.

Bagliniere, J.-L., Hamelet, V., Guéraud, F., Aymes, J.-C., Goulon, C., Richard, A., Josset, Q., Marchand, F., 2020. Guide pour l'interprétation des écailles et l'estimation de l'âge chez la truite commune (*Salmo trutta*) dans les populations françaises, Office français de la biodiversité-OFB. ed, Guides et protocoles. Ce guide est téléchargeable sur le site de l'OFB (<https://professionnels.ofb.fr/fr/node/835>) et référencé dans le portail partenarial des documents sur l'eau et la biodiversité (<https://www.documentation.eauetbiodiversite.fr/notice/guide-pour-l-interpretation-des-ecailles-et-l-estimation-de-l-age-chez-la-truite-commune-salmo-trutt0>)

Bagliniere, J.-L., Jeannot, N., Nevoux, M., Lamireau, L., Lange, F., Josset, Q., Marchand, F., 2022. Guide pour l'interprétation des écailles et l'estimation de l'âge chez le saumon atlantique (*salmo salar*) dans les populations françaises. Ce guide est présenté en ligne sur le portail technique de l'OFB (<https://professionnels.ofb.fr/fr/node/1504>) et référencé dans le portail partenarial des documents sur l'eau et la biodiversité (<https://www.documentation.eauetbiodiversite.fr/notice/guide-pour-l-interpretation-des-ecailles-et-l-estimation-de-l-age-chez-la-truite-commune-salmo-trutt0>).

7 STOCKAGE ET SECURITE DES DONNEES

L'administrateur des systèmes d'information de Colisa assure toute la partie technique des outils communs à Colisa (Postgres, Collec-science) et s'assure des sauvegardes des bases Postgres.

Par ailleurs le jeu de données Colisa est déposé dans l'entrepôt de données du GBIF qui en garantit l'accès et la sauvegarde.

7.1 LES SUPPORTS DE DONNEES

- Toutes les données sont dans la base de données « db_u3e » sur le serveur postgresql de l'U3E (pggeodb.u3e.inrae.fr) sur le datacenter de toulouse.
- La base de données Postgresql pour les captures de marins amphihalins « db_captures_marins_amphihalins » sur le serveur postgresql de l'U3E (pggeodb.u3e.inrae.fr) sur le datacenter de toulouse.
- La base de données Postgresql du CNICS « db_capture » sur le serveur postgresql de l'U3E (pggeodb.u3e.inrae.fr) sur le datacenter de toulouse.
- Sharepoint COLISA (<https://sites.inrae.fr/site/colisa>)
- Gitlab forgemia (<https://forgemia.inra.fr/pole-migrateurs/colisa>). Sauvegarde et versionnement du code J2EE du SI COLISA.
- Gitlab forgemia (<https://forgemia.inra.fr/pole-migrateurs/si-analyses>). Sauvegarde et versionnement du code J2EE du SI_ANALYSE.
- GBIF : Base de données externe du GBIF
- GGBN : Base de données externe du GGBN
- Portail RARe

7.2 VOLUMETRIE

7.2.1 BASE POSGRESQL « db_u3e », SCHEMA « collection »

La volumétrie est faible et ne pose pas de problème : < 300 MB pour près de 500 000 échantillons.

7.2.2 SHAREPOINT DE COLISA

Environ 5Go

- 2,3 Go pour le stockage des documents
- 2,5 Go pour les images d'intercalibrage

7.3 POLITIQUE DE SECURITE

Une fois dans le catalogue Colisa, les données sont accessibles depuis le site Colisa.fr.

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

Les sauvegardes de la base de données Postgresql 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.

7.4 SECURITE - INTEGRITE – TRACABILITE

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

L'identification par DOI de la collection (<https://data.inra.fr/dataset.xhtml?persistentId=doi:10.15454/D3ODJM>) devrait permettre d'assurer un minimum de suivi. La mise en place d'un formulaire de demande en ligne sur le catalogue Colisa avec la [charte d'engagement du CRB Colisa](#) à signer devrait également faciliter ce suivi.

Pour l'ORE et toutes les base de données de l'U3E le SI analyse permet de garder la trace de toutes les analyses (nouvelles ou correction). Des triggers sont mis en place sur la base de données Postgres afin d'assurer l'intégrité des données entre les données de l'U3E, les analyses faites sur les échantillons de l'U3e et Colisa (voir PGD de l'unité U3E).

Le SI_COLISA assure la traçabilité des demandes de prêts.

Toutes les productions scientifiques et techniques finalisées sont intégrées et mises à dispositions des scientifiques et du grand public au travers de la collection Hal-Colisa <https://hal.inrae.fr/COLISA>.

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, GGBN) 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).