
DMP du projet "OLA-Infrastructure"

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

Renseignements sur le plan

Titre du plan	DMP du projet "OLA-Infrastructure"
Langue	fra
Date de création	2020-01-16
Date de dernière modification	2021-08-21
Identifiant	https://doi.org/10.15454/1RUGHK

Renseignements sur le projet

Titre du projet OLA-Infrastructure

Résumé

Plan de gestion des données de l'infrastructure **OLA-O**bservation au long terme et expérimentation - écosystèmes **LA**custres.

Organisée sous forme d'offre de services, l'infrastructure OLA met à la disposition des scientifiques des outils et moyens (bateaux, sondes, bassins/mésocosmes, laboratoires, drone, ...) leur permettant d'explorer les milieux lacustres et d'obtenir des données scientifiques de qualité dans l'objectif de comprendre et de modéliser l'évolution et les fonctions écologiques de ces systèmes, qui sont soumis simultanément à un changement des pressions anthropiques locales et climatiques.

La partie "Observation", opérationnelle depuis de très nombreuses années, est déjà riche d'un historique de 30 à 60 ans de données (sur les grands lacs périalpins) pour de nombreux paramètres biologiques et physicochimiques. L'imbrication étroite entre ces activités de collecte de données in situ et celles d'expérimentations permet de mieux caractériser les mécanismes de réponse de la biodiversité aux pressions.

Les données d'observation sont déposées dans le Projet « Suivi des lacs » du système d'information OLA (SI OLA) par dépôt et publication de fichiers formatés selon les types de données ; le SI OLA gère les fichiers, leurs révisions et la mise en base des données. Il permet l'accès à ces données.

Les données d'expérimentations sont déposées pour leur projet sous forme de fichiers et gérés par la fonctionnalité « Entrepôt de fichiers » ; ; le SI OLA gère les fichiers, leurs révisions et permet l'accès à ces fichiers.

Sources de financement

- INRAE (Subvention d'Etat) + financements sur projets : NA

Produits de recherche :

1. Système d'Information des Données d'observation sur le long terme des lacs et des données d'expérimentation (Jeu de données)
2. Expérimentation au sol et en lac (Service)

Contributeurs

Nom	Affiliation	Rôles
Ghislaine Monet		<ul style="list-style-type: none">• Coordinateur du projet• Personne contact pour les données (SI-OLA, Experimentation)• Responsable du plan

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

DMP du projet "OLA-Infrastructure"

Informations sur la structure

OLA - Observation au long terme et expérimentation pour les écosystèmes **L**Acustres.

- ISC (Infrastructure Scientifique Collective)
-

Nom, Prénom	Courriel	Rôle
MONET Ghislaine	Ghislaine.Monet@inrae.fr	Responsable de l'infrastructure Membre du comité de direction de OLA Responsable données
Domaizon Isabelle	Isabelle.Domaizon@inrae.fr	Directrice de l'unité de Recherche (Umr Carrtel) à laquelle est adossée l'infrastructure OLA. Membre du comité de direction de OLA
Tran-Khac Viet	Viet.Tran-Khac@inrae.fr	Responsable du laboratoire d'analyse de physico-chimie Membre du comité de direction de OLA

INRAE

- Autre (à préciser dans la zone de commentaires)

AQUA

INRAE

Informations sur le plan de gestion

<https://doi.org/10.15454/1RUGHK>

Date	n° de version	Status	Auteur	Affiliation de l'auteur (se reporter à l' annuaire Inra)	Validé par	Validé le
23/06/2021	1.0.0	en cours	G. Monet	INRAE UMR CARRTEL		
27/07/2020	1.0.1	en cours	G. Monet	INRAE UMR CARRTEL		
01/08/2021	1.1	En relecture	G. Monet	INRAE UMR CARRTEL		
20/08/2021	1.1	Relecture	V. Tran-Khac	INRAE UMR CARRTEL		
21/08/2021	1.1	Terminé	G. Monet	INRAE UMR CARRTEL	G. Monet	21/08/2021

Présentation générale des données

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

Les données d'observation sont collectées par la structure sur 4 grands lacs péri alpins (Léman, Annecy, Bourget, Aiguebelette). Ces campagnes d'observations sont commanditées par l'unité de recherche de proximité, l'UMR CARRTEL, qui gère les contrats avec les gestionnaires des Lacs : La CIPEL pour le lac Léman, le SILA pour le lac d'Annecy, le CISALB pour le lac du Bourget et le CCLA pour le lac d'Aiguebelette. Un historique de ces données de 30 à 60 ans pour ces grands lacs péri alpins est mis à disposition de la communauté scientifique sur le SI-OLA.

Ces données sont soit recueillies *in situ* (cas des sondes multiparamètres par ex.) soit résultantes de l'analyse des échantillons prélevés, analyses faites par les différents plateaux techniques de OLA (physico-chimie, analyse de la biodiversité phytoplancton et zooplancton, ...). La fréquence d'échantillonnage pour chaque lac est environ bi-mensuelle entre les mois de mars et octobre et mensuelle pour les mois d'hiver.

En ce qui concerne ces suivis à long terme, la structure OLA prend en charge les différents processus de gestions des données :

- Collecte.
- Traitement et curation des données.
- Stockage sur le poste de travail pour l'acquisition terrain.
- Transfert des fichiers ou travail direct sur un espace de stockage réseau sécurisé (*NAS de l'unité CARRTEL*).
- Mise en base de données dans le SI-OLA une fois par an sur les données de l'année écoulée.
- Mise à la disposition des données par le SI OLA après inscription sur le SI et demande d'accès selon le cas.
- Partage à la demande vers d'autres SI si justifié.
- Aide au dépôt de jeux de données sur le dataverse INRAE

Dans tous les cas, ces données sont prêtes à leur valorisation ou exploitation dans le cadre d'analyses et de modélisations.

- **Données produites par un tiers dont les fichiers sont conformes à ceux du SI OLA**

L'Observatoire des lacs d'altitude (réseau des lacs sentinelles géré par Asters, Conservatoire d'espaces naturels) se concentre sur une vingtaine de lacs situés à plus de 1800 m d'altitude pour suivre la réponse de ces écosystèmes aux changements globaux.

Les données concernent des lacs situés dans les réserves naturelles de Haute-Savoie (Anterne, Brévent, Cornu, Jovet, Pormenaz), dans le parc national de la Vanoise (Arpont, Blanc et Noir du Carro, Merlet, Mont Coua), dans le parc national des Écrins (Muzelle, Pavé, Pétaarel, Pisses, Plan-Vianney), dans le parc national du Mercantour (Bresse inférieur, Bresse supérieur, Lauzanier), des lacs gérés par le centre d'ingénierie d'EDF (Aumar, Corne, Cos, Izourt, Rabuons) et par la communauté de communes Cœur de Maurienne-Arvan (Blanc et Bramant).

Des protocoles communs réalisés chaque année en septembre permettent un relevé de données uniformes entre les lacs du réseau sur les paramètres suivants : transparence, profils de sonde multi-paramètres, analyse physico-chimique de l'eau, analyse de la chlorophylle-a, détermination du zoo et phytoplancton. Ces paramètres sont étudiés sur le point géographique ou se situe la plus grande profondeur pour le lac. Les analyses physico-chimiques de l'eau du lac sont réalisées sur deux échantillons, l'un prélevé au fond du lac et l'autre prélevé à différentes profondeurs de la zone euphotique. L'étude du phytoplancton porte également sur un échantillon prélevé dans la zone euphotique. Un trait de filet est réalisé dans toute la colonne d'eau pour étudier le zooplancton. Enfin, des capteurs sont placés à l'année dans les lacs et mesurent la température à différentes profondeurs et l'oxygène au fond du lac. Pour plus d'information sur les protocoles : www.lacs-sentinelles.org/fr/les-protocoles

Pour la plupart des lacs, l'acquisition de données a démarré en 2014. Pour certains lacs, des données plus anciennes existent (depuis

1992) avec des incertitudes notamment sur les données de chimie et des protocoles parfois légèrement différents.

Les données « brutes » produites lors des campagnes de terrain, sont stockées dans le système d'information (SI) de OLA (Observatoire des LACs), administré par l'INRAE de Thonon-les-Bains, France. Elles sont téléchargeables après création d'un compte utilisateur sur le SI.

Les points GPS du lieu de prélèvements de tous ces lacs, péri-alpins ou lacs d'altitudes sont disponible ici: [OLA-Sites-Description-gps.pdf](#)

-
- Observation
 - Expérimentation

-
- Dataset
-

Ces données couvrent différents thématiques.

Les variables du thème physico-chimie sont : azote, calcium, carbone, chlorures, conductivité, magnésium, oxygène dissous, pH, phosphore, potassium, silice réactive, sodium, sulfate, température, titre alcalimétrique complet. Données prélevées à différentes profondeurs sur toute la colonne d'eau du lac.

- *Léman* : Sans être continue la période concerne un suivi depuis les années 1957 jusqu'à nos jours.
- *Bourget* : La période concerne un suivi depuis 1999 jusqu'à nos jours.
- *Annecy* : Sans être continue la période concerne un suivi depuis les années 1965 jusqu'à nos jours (Absence de données entre 1982 et mai 1990).
- *Aiguebelette* : La période concerne un suivi depuis 1974 jusqu'à nos jours.

Les variables du thème chlorophylle sont : chlorophylle a strickland-parsons, chlorophylle c, caroténoïdes, chlorophylle a scor-unesco, phéopigments. Données prélevées à différentes profondeurs sur toute la colonne d'eau du lac.

- *Léman* : La période concerne un suivi depuis 1976 jusqu'à nos jours.
- *Bourget* : La période concerne un suivi depuis 2006 jusqu'à nos jours.
- *Annecy* : La période concerne un suivi depuis 1990 jusqu'à nos jours.
- *Aiguebelette* : La période concerne un suivi depuis 1989 jusqu'à nos jours.

Les variables du thème conditions de prélèvements sont : température de l'air, nébulosité, ensoleillement, temps, direction du vent, vitesse du vent, pression atmosphérique, aspect de l'eau, état de surface, transparence, couleur de l'eau.

- *Léman* : La période concerne un suivi depuis 1957 jusqu'à nos jours.
- *Bourget* : La période concerne un suivi depuis 2010 jusqu'à nos jours.
- *Annecy* : La période concerne un suivi depuis 1966 jusqu'à nos jours.
- *Aiguebelette* : La période concerne un suivi depuis 1989 jusqu'à nos jours.

Les variables du thème phytoplancton sont : nom du déterminateur, volume sédimenté, surface de comptage, taxon déterminé, nombre de champs comptés, nombre d'objets comptés, nombre d'objets par ml, nombre de cellules par ml, biovolume de l'espèce dans l'échantillon. Données prélevées sur une profondeur intégrée de 0 à 10m ou 0 à 20m selon les années.

- *Léman* : La période concerne un suivi depuis 1974 jusqu'à nos jours.
- *Bourget* : La période concerne un suivi depuis 1995 jusqu'à nos jours.
- *Annecy* : La période concerne un suivi depuis 1996 jusqu'à nos jours.
- *Aiguebelette* : La période concerne un suivi depuis 1998 jusqu'à nos jours.

Les variables du thème production primaire sont : durée d'incubation, production primaire par durée d'incubation réelle, production primaire par heure, production primaire par tiers médian. Données prélevées à différentes profondeurs sur la colonne d'eau du lac de 0 à 30m.

- *Léman* : La période concerne un suivi depuis 1976 jusqu'à nos jours.
- *Bourget* : La période concerne un suivi depuis 2011 jusqu'à nos jours.
- *Annecy* : La période concerne un suivi de 2011 à 2014
- *Aiguebelette* : Pas de production primaire

Les variables du thème sondes sont : profondeur, température, chlorophylle a, pH conductivité, oxygène, taux de saturation en oxygène, turbidité, transparence, par w, par a, pah, pc. Données obtenues à différentes profondeurs (tous les 10 cm environ) sur toute la colonne d'eau du lac (0 à 300m).

- *Léman* : La période concerne un suivi depuis 1991 jusqu'à nos jours.
- *Bourget* : La période concerne un suivi depuis 1991 jusqu'à nos jours. (1984 pour la température)
- *Annecy* : La période concerne un suivi depuis 1991 jusqu'à nos jours. (1966 pour la température)
- *Aiguebelette* : La période concerne un suivi depuis 1991 jusqu'à nos jours (1974 pour la température)

Les variables du thème zooplancton sont : nom du déterminateur, biovolume sédimenté, nom du taxon déterminé, stade de développement, nombre d'individus. Données prélevées sur une profondeur intégrée de 0 à 50m.

- *Léman* : La période concerne un suivi depuis 1959 jusqu'à nos jours.
- *Bourget* : La période concerne un suivi depuis 1996 jusqu'à nos jours (absence de données entre 1997 et 2003 inclus).
- *Annecy* : La période concerne un suivi depuis 1994 jusqu'à nos jours.
- *Aiguebelette* : Pas de zooplancton

-
- csv pour les fichiers déposés
 - csv pour les fichiers résultants de requêtage
 - SGBD Postgresql pour la gestion

Un format de données csv est imposé pour le dépôt des fichiers dans le SI-OLA, il est fonction de la thématique de la donnée. L'utilisateur venant chercher les données n'a pas besoin de connaître ce format de dépôt. Il peut extraire les données du SI soit dans un format csv "colonnes" (compatible avec tous les tableurs) soit dans un format csv "ligne" plus facilement compatibles avec les logiciels de traitement statistique tels que R.

-
- Biodiversity and Ecology
 - Water resources

Droits de propriété intellectuelle

Les données sont libres d'accès selon les principes de la Science Ouverte, il est cependant nécessaire de s'inscrire sur le SI (créer un compte) et éventuellement (selon les lacs) de faire une demande d'accès ce qui permet de générer des collaborations scientifiques. Une charte d'utilisation des données écrites en français et en anglais décrit les termes de réutilisations des données; elle est accessible en ligne: [Terms of Use of OLA-IS \(Observatory of LAkes - Information System\)](#). Cette charte permet d'indiquer quelle citation utilisée lors de l'utilisation des données.

Confidentialité

Aucun jeux de données du SI-OLA ne comprend de données confidentielles au sens du RGPD.

sans objet

sans objet

Partage des données

Les données acquises par OLA dans le cadre de ses suivis long terme entrent dans le cadre de la Science Ouverte.

Les données peuvent être réutilisées pour la recherche sur l'étude et le fonctionnement des lacs spécifiquement sur les lacs observés ou dans un but de méta analyses.

Les fichiers obtenus suite au requêtage de la base de données sont fournis au format CSV et conformes aux recommandations d'interopérabilités émises par l'Etat.

Ils ont les caractéristiques techniques suivantes :

- Le caractère utilisé pour séparer les colonnes est la virgule.
- Le texte utilise la norme de codage de caractères [UTF-8](#).

Accessible sur le SI OLA : <https://si-ola.inrae.fr>

et

Publication sur le portail [Data Inra](#)

-
- Tous (open acces)

Avec inscription sur le SI pour les données du SI-OLA.

-
- Licence ouverte <https://www.etalab.gouv.fr/licence-ouverte-open-licence> (compatible CC-BY)
-

Organisation et documentation des données

Les méthodes et outils pour acquérir et traiter les données des 4 grands lacs suivis par l'observatoire sont gérés en interne par l'infrastructure OLA.

Le SI OLA est aussi un service pour l'accueil de données d'autres lacs tels que les lacs d'altitude pour lesquels la gestion des données (acquisition, curation) est faite par le CEN de Hte-Savoie (Conservatoire des Espaces Naturels). Les collecteurs et fournisseurs de données assurent la responsabilité des données qu'ils mettent à disposition sur le SI OLA s'ils en ont l'autorisation.

Les informations qui suivent concernent les données obtenues sur les 4 grands lacs péri alpins (Léman, Annecy, Bourget, Aiguebelette).

- **Collecte. Les méthodes et outils de collecte dépendent du type de données :**

Les collecteurs de données suivent les protocoles définis dans le cadre de l'assurance qualité de OLA ; ces protocoles décrivent les normes AFNOR suivies (si c'est le cas) ainsi que les méthodes et outils nécessaires à la collecte de la donnée.

*Données recueillies *in situ* : *Exemple des données sondes multi paramètres :*

Les données brutes des sondes multi paramètres sont recueillies et stockées sur l'ordinateur du terrain.

Ces données sont lisibles avec le logiciel d'acquisition spécifique à la sonde qui permet la calibration de la sonde, la visualisation des données en cours d'acquisition et la création d'un fichier texte ouvert universel.

En vue d'y ajouter des métadonnées (nom du site, coordonnées GPS, et de faire un premier traitement de validation et de contrôle de cohérence des données acquises, un logiciel développé en interne est utilisé (logiciel SHL-Sonde). Il apporte une aide à l'utilisateur en lui proposant les données qui pourraient être objet d'une curation et en lui proposant un outil de visualisation des données. Le logiciel SHL-Sonde permet également d'intégrer d'autres données issues des analyses au laboratoire qui servent de faire une validation croisée avec les données brutes.

*Données recueillies sur les plateaux techniques de laboratoire:

Ces données sont obtenues par (i) l'utilisation d'appareillages spécifiques, il en est ainsi des données de physico-chimie obtenues à partir d'analyses effectuées sur des échantillons d'eau recueillis au cours de la phase de terrain ou par (ii) observation sous microscopes, c'est le cas des données de biodiversité du plancton (phyto et zoo) qui sont obtenues par observation sous microscope par les systématiciens de la thématique. Après observation et comptage ces données sont notées sur cahier et sur fichier au format excel.

Dans tous les cas, les données sont ensuite stockées sur le réseau dans un espace partagé versionné, sécurisé et régulièrement sauvegardé.

- **Traitement et curation : outils utilisés**

Le traitement des données et un premier niveau de curation est faite sur le plateau à l'origine de la création de la donnée. Les fichiers bruts et en cours de traitement sont stockés sur un espace partagé sécurisé (cft-tho-nas01-v / IP 147.99.240.215) qui est lui-même sauvegardé sur un disque cloud (IP 147.99.240.231) accessible seulement à l'administrateur des données et aux administrateurs informaticiens de l'Umr Carrtel.

Les principaux outils utilisés sont :

- Contrôle par les opérateurs (utilisation de cartes de contrôle)
- Le logiciel R
- l'intégrateur de données Talend
- logiciel SHL sonde interne

- **Mise à disposition : outils de développement**

Les données sont mises à disposition sur le SI OLA (<https://si-ola.inrae.fr>); Le SI est développé selon une architecture informatique éprouvée en langage JAVA. L'interface web utilise des standards du métier tels que le framework JSF, Spring et Hivernate complètent les outils de développement. Les sources de l'[architecture informatique de OLA](#) sont disponibles.

- **Stockage et archivage : outils**

Les données du SI sont stockés sur les datacenter INRAE ; les sauvegardes et répliquions sont gérées selon les procédures des datacenter INRAE.

L'outil de gestion des données est le SGBD Postgresql

Métadonnées	Origine, mode de production des métadonnées (ex : saisie manuelle, annotation automatique...)	Standard, Vocabulaires associés	Conditions ou fréquence de la mise à jour (si applicable) (ex : changement de l'accessibilité)
Métadonnées géospatiale	-Fournies avec le fichier des données extraites sur le SI -Saisie via interface web si dépôt manuel sur dataverse inrae -Annotation automatique en utilisant les pipelines d'annotation et d'exploitation Anaee-F sur DataVerse INRAE, API	Vocabulaire contrôlé du thésaurus AnaEE-France https://opendata.inra.fr/anaeeThes Le thésaurus AnaEE-F est aligné sur GEMET et Agrovoc ISO 19115	quand nécessaire
Citation de la donnée	-Saisie manuelle via l'interface Web pour data inrae, -Fourni automatique lors d'extraction via le SI OLA.	Fourni un DOI : 10.4081/jlimnol.2020.1944	quand nécessaire
Métadonnées sciences de la vie et environnement	-Saisie via l'interface Web pour data verse INRAE -fourni systématiquement via le SI OLA lors des extractions de données -Annotation automatique si dépôt sur dataverse inrae via les pipelines AnaEE-F d'annotation et d'exploitation.	-Vocabulaires contrôlés pour différentes métadonnées : thésaurus AnaEE-F ; http://agroportal.lirmm.fr/ontologies/ANAEETHES -Référentiel SANDRE pour les données de l'eau. - ALGAE BASE pour la taxonomie du phytoplancton - Référentiel taxonomique TAXREF pour les organismes.	quand nécessaire
Ressource Sémantique	saisie via l'interface Web, API	Ontologie AnaEE-France ; basée sur l'ontologie générique OBOE ; https://www.w3.org/2015/spatial/wiki/Alignment_to_OBOE Agroportal metadata	

Non.

Consulter éventuellement les protocoles suivis et les normes AFNOR utilisées (indiquées dans les protocoles).

Le système d'information OLA gère deux types de fichiers : les fichiers de dépôt et les fichiers d'extraction. Ces deux types de fichiers sont liés à des fichiers de données de référence (DDR) qui représente le référentiel du SI et qui sont des données partagées par tous les types de données, toutes les thématiques gérées dans le SI OLA ; ces fichiers de DDR préalablement insérés initialisent le système

de données. Il est possible d'ajouter à la demande des données de références telles que par exemple : Nouvelle variable, nouvelle unité, nouveau taxon, nouveau projet etc.

Les fichiers déposés sont ceux qui seront mis en ligne par l'administrateur de données qui peut être différents du dépositaire du fichier. Le SI propose une gestion des versions déposées et mise en ligne; tous les fichiers sont stockés jusqu'à demande de suppression par l'administrateur des données.

Convention de nommage :

La convention de nommage de tous les fichiers, quel que soit le type de données (la thématique) est la suivante :

NomDuProjet_NomDuSite_NomDuTypeDeDonnees_DateDebut_DateFin#commentaire#.csv

Exemples (plusieurs formats possible pour la date):

suivi_des_lacs_annecy_sonde_multiparametres_2013-01-16_2013-12-09.csv

suivi_des_lacs_leman_physico_chimie_01-01-1957_31-12-2012#valid GM_avec_temp#.csv

Les fichiers sont eux même construits selon un format défini qui tient compte du type de données ; Ces fichiers contiennent des champs communs :

nom du projet	nom du site	nom de la plateforme	date de prelevement	profondeur	variable 1	variable 2	variable ...
----------------------	--------------------	-----------------------------	----------------------------	------------	------------	------------	--------------

ou autre exemple :

nom du projet	nom de la plateforme	date de prélèvement	outil de mesure	outil de prélèvement	profondeur minimum	profondeur maximum	nom du déterminateur	Variable 1	Var 2
----------------------	-----------------------------	----------------------------	-----------------	----------------------	--------------------	--------------------	----------------------	------------	-------

La donnée déposée sur le système d'information et mise à disposition est obtenue après plusieurs étapes permettant d'assurer sa fiabilité.

1. Calibration/étalonnage des outils de mesures ou appareillage de laboratoire
2. Acquisition de la donnée et vérification technique par le collecteur ==>mise en fichier
3. Validation scientifique des données au fil de l'eau et en fin d'année pour la cohérence évolutive.
4. Vérification logicielle automatique lors du dépôt dans le SI à la fois sur la cohérence du fichier déposé par rapport à son contenu et sur les valeurs déposées par rapport aux valeurs attendues (encadrement de valeurs).

Les workflow des principales données du SI sont ci-dessous :

Pour la physico-chimie : [Workflow-traitement-Physico-Chimie-OLA.png](#)

Pour la détermination et le comptage du zooplancton : [Workflow-analyse-du-zooplancton.png](#)

Pour l'insertion de ces données vers le SI OLA : [Workflow-insertion-des-donnees-vers-le-SIOLA.png](#)

Stockage et sécurité des données

Les flux peuvent légèrement varier selon la donnée (terrain, physico-chimie, biodiversité...)

En règle générale :

1. Acquisition de la donnée sur matériel spécifique de terrain ou par passage d'échantillon sur appareillage ou par analyse de technicien (détermination, comptage). Inscription des résultats sur cahier pour l'analyse du zooplancton.
2. Valeur brute stockée sur ordinateur liée au matériel d'acquisition ou sur ordinateur du déterminateur, ou directement sur un espace de stockage partagé réseau.
3. Transfert du fichier sur échange de stockage partagé si non fait à l'étape 2 le jour même de l'acquisition.
4. Traitement du fichier pour sa mise au format SI-OLA sur poste individuel relié au réseau INRAE.
5. Dépôt du fichier traité mis en forme pour le SI OLA déposé sur l'espace NAS
6. Dépôt et publication des fichiers sur le SI OLA une fois par an, après validation complète des données. Les données du SI sont stockées sur les datacenters INRAE. Les données en base de données sont sauvegardées quotidiennement par dump (fonctionnalité du système de gestion de la base de données, SGBD Postgresql). Puis ce fichier dump est déposé sur une partition dont l'archivage est pris en charge par les outils de l'infrastructure du datacenter INRAE de Toulouse avec réplication sur le serveur d'Ile de France. La durée théorique de rétention des sauvegardes proposée par le datacenter INRAE de Toulouse est de 3 mois, mais la durée effective qui nous a été imposée en raison de l'important volume de données des dumps des SI gérés par l'écoinformatique a été réduite à 7 jours.

Le SI lui-même permet de faire la révision des fichiers ; tous les fichiers déposés sont stockés et archivés après dépublication si c'est le cas.

La volumétrie des données est actuellement de 2.5 Go, mais à cela vont

s'ajouter courant 2021 des données à intégrer au SI :

-100 Go par an pour les données Fishola,

-100 Go d'historique pour des données de suivis par acoustique + 10 Go par an

-10 Go de données de suivis des populations de perches

-100 Go pour des suivis phénologiques de reproduction de poissons
-2 To sur un projet initié en 2021 : LÉXFISH
D'ici quelques temps la volumétrie sera donc autour des 300 Go annuel + 2 à 3 To de stockage.

Politique locale à ce jour.

Les espaces partagés sur l'espace de stockage commun (serveur NAS) sont répliqués tous les soirs sur un disque externe cloud accessible seulement aux administrateurs informatiques.

L'espace de stockage du SI OLA est situé sur le Datacenter INRAE de Toulouse, et répliqué sur celui d'Ile de France selon la politique des datacenter INRAE.

Un espace cloud a été défini pour OLA ; il abritera une sauvegarde du NAS sur lequel sont sauves actuellement les données avant mises sur SI.

Les données du [SI OLA](#) que ce soient les données d'observations ou les données issues d'expérimentation dans le cadre d'un projet (déposées par la fonctionnalité d'entrepôt de fichiers), sont mises à disposition selon les principes FAIR.

Une [charte d'utilisation](#) est systématiquement fournie avec les données ainsi qu'un lien vers la [citation](#) à adopter.

Les utilisateurs doivent s'inscrire sur le SI (ouvrir un compte) ; les seuls éléments qui leur sont demandés sont de mettre un identifiant, un mot de passe (qui sont les seuls à connaître) et une adresse mail valide.

Certaines données sont disponibles dès inscription c'est par exemple le cas pour les lacs d'altitude, d'autres nécessitent d'en demander l'accès par une fonctionnalité du SI. Ceci permet d'avoir une traçabilité des demandes et permet une éventuelle collaboration avec les demandeurs.

L'accès est systématiquement ouvert. Les données ne sont disponibles qu'après validation technique et scientifique.

Par ailleurs, des jeux de données bien circonscrits peuvent être extraits du SI et mis à disposition sur le data verse INRAE qui fournit un DOI pour ce jeu de données déposé. C'est le cas pour les jeux de données ayant vocation à accompagner certaines publications scientifiques. Ils sont alors publiés avec des métadonnées "étendues" (métadonnées permettant d'aller au-delà de la simple réutilisation de la données).

La production des données d'observation sur les grands lacs est gérée par l'infrastructure OLA, les producteurs sont internes et le suivi fait au sein de l'infrastructure OLA par le comité de direction.

Les données produites sur les lacs d'altitude sont gérées et validées par le [Conservatoire d'espace naturel \(CEN\) de Haute-Savoie](#).

Chaque fournisseur de données valide sa production et assure la fiabilité des données qu'il fournit ; les fournisseurs de données sont validés et autorisés par le comité de direction de l'infrastructure OLA.

Archivage et conservation des données

Système d'Information des Données d'observation sur le long terme des lacs et des données d'expérimentation

Les données déposées sur le SI OLA ont vocation à être pérennes, c'est le cas depuis plus de 60 ans pour les données du Lac Léman.

Il n'y a donc pas de destruction envisagée.

Les fichiers déposés sont cependant gérés par un système de révision et peuvent être supprimés à la demande par le gestionnaire, s'ils ne sont pas publiés.

En ce qui concerne les fichiers bruts (avant validation ou fichiers intermédiaires) ils sont également conservés sur le long terme pour permettre une nouvelle exploitation si nécessaire.

Les données du SI OLA sont conservées sur le long terme et répliquées sur les 2 datacenter de INRAE.

A vie.

Les coûts de conservation de la donnée sont actuellement assurés par INRAE.

Expérimentation au sol et en lac

Question sans réponse.

Question sans réponse.

Question sans réponse.

Question sans réponse.