

---

# **PGD 1 : Suivi (fictif) de population de poissons dans le lac du Bourget**

*Plan de gestion de données créé à l'aide de DMP OPIDoR*

**Créateurs du PGD :** Marie-Claude QUIDOZ, VERONIQUE ARNAL, Anne Gorgeon

**Affiliation du créateur principal :** CNRS

**Modèle du PGD :** ANR - Modèle de PGD (français)

**Dernière modification du PGD :** 17/09/2020

**Financeur :** ANR

**Numéro de subvention :** ANR-2020-000-000

## **Résumé du projet :**

**Avertissement : il s'agit d'un PGD destiné à être utilisé à des fins pédagogiques. Le projet est fictif, toute ressemblance avec des projets existants ou ayant existé ne saurait être que fortuite**

L'objectif principal est d'observer les conséquences sur la faune marine des actions de dépollution engagées depuis le milieu des années 1970 dans le lac du Bourget qui présentait un phénomène majeur d'eutrophisation.

Le lac du Bourget n'a pour l'instant fait l'objet que de rares campagnes de prélèvements non standardisés.

Cette étude sera menée dans le cadre des missions de l'observatoire « Friture\_Du\_Lac » et respectera des protocoles éprouvés et standardisés dans d'autres lacs des Alpes pour les prélèvements d'eau et de poissons.

Les observations se feront dans le cadre de plusieurs campagnes de prélèvements couvrant l'ensemble des stations sélectionnées, avant et après la période estivale. Elles porteront essentiellement sur l'inventaire des poissons: après prélèvement de tous les poissons présents, les espèces les plus courantes seront ensuite dénombrées (nombre d'individus d'une espèce) et seuls les individus appartenant à 5 espèces, prédéfinies comme des poissons bio-indicateurs, seront pesés et mesurés, avant leur remise à l'eau.

**Chercheur Principal :** Véronique Arnal

**Contact pour les Données :** Marie-Claude QUIDOZ

## **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.

# PGD 1 : Suivi (fictif) de population de poissons dans le lac du Bourget

---

## 1. Description des données et collecte ou réutilisation de données existantes

### Données préexistantes

Aucune donnée ancienne ne sera réutilisée: quelques données dans la base de données Savoie concernent le lac du Bourget mais elles ont été obtenues sur d'autres sites géolocalisés et parfois avec des protocoles différents.

### Nouvelles données - Méthodologie

Les données recueillies dans ce projet proviendront des 2 campagnes de prélèvements annuelles avant et après la période touristique estivale (mai et septembre) et sur 3 ans entre Mai 2020 et Septembre 2022.

Les données d'observation des poissons et de l'eau du lac (comptage, mesures et dosage) seront saisies sur un carnet de terrain puis retranscrites sous tableur le soir même (un classeur par campagne) avant de les valider et de les inclure dans la base de données Savoie.

Ces données seront exploitées pour des analyses statistiques ultérieures (logiciel R), jusqu'à la fin du contrat ANR au 30 avril 2023.

En dehors des campagnes de prélèvement, les échantillons d'eau et de nageoires recueillis sur le bateau, seront amenés en laboratoire pour analyses microscopiques (données d'observation) toxicologiques et moléculaires (données expérimentales).

### Nouvelles données - Documentation

Elles seront documentées avec tout le détail du déroulement de la campagne de prélèvement, en particulier les noms de stations géolocalisées, les protocoles utilisés, les mesures réalisées, les échantillons prélevés et les photos d'identification des espèces.

L'ensemble des expérimentations en laboratoire (protocoles, suivi du déroulement de l'expérience et résultats) et les étapes d'analyses bio-informatiques et statistiques seront consignés dans un cahier de laboratoire papier.

Les données seront conservées sous formats ouverts :

- tableur (Libre Office -> \*.ods)
- base de données (SGBD PostgreSQL)

Les documents associés seront disponibles sous format Libre Office -> \*.odt ou en pdf

Les photos de chaque lot de poissons seront disponibles sous format \*.jpeg

Les séquences seront disponibles sous format \*.txt

La carte du lac du Bourget (avec le géo-référencement des stations) sera disponible en \*.shp

En termes de volume, sur la totalité du projet, environ 4000 mesures seront faites, 2300 photos seront prises, 4000 séquences seront analysées et 4000 échantillons seront conservés dans des cryoboîtes. Ce qui représente une dizaine de Go pour les mesures, les photos et les données brutes et analysées des séquences, 114 bouteilles de prélèvements d'eau et 50 cryoboîtes de 80 tubes pour les échantillons de nageoires (au total 750 cm linéaire).

## 2. Documentation et qualité des données

### Référentiels

- le référentiel taxonomique TAXREF v. 13 (INPN) sera utilisé pour référencer les espèces
- la norme internationale ISO 80000-1:2009 sera utilisée pour les unités
- les stations seront référencées en WGS84 et le référentiel de l'IGN sera utilisé pour les noms des communes
- les standards en biologie moléculaire seront utilisés, en particulier le code IUPAC pour le séquençage d'ADN

### Métadonnées

- la base de données sera décrite en utilisant le standard EML (Ecological Metadata Language). Un effort particulier

sera fait pour décrire les mesures (eml attribute)

- les informations au standard EXIF (Exchangeable image file format) des photos seront sauvegardées

#### **Organisation et codage des répertoires**

Un répertoire spécifique nommé "Terrain" recensera tous les protocoles de collecte des données, les notices d'utilisation et de calibrage des appareils de mesure et les images d'aide à l'identification des espèces de poissons. Il contiendra aussi le principe établi d'organisation et de codage des répertoires et des fichiers.

Un répertoire spécifique nommé "Laboratoire" recensera tous les protocoles, les notices des appareils, les résultats expérimentaux et d'analyses bio-informatiques et statistiques.

Un répertoire "Données Terrain" sera créé pour chaque campagne. Il sera organisé conformément aux éléments mis dans le document d'organisation et de codage des répertoires et des fichiers.

Un répertoire "Données Laboratoire" sera créé pour chaque campagne. Il sera organisé conformément aux éléments mis dans le document d'organisation et de codage des répertoires et des fichiers.

Lors de la campagne de terrain, la qualité et la conformité de la collecte des données seront contrôlées par le responsable de campagne, assisté du technicien en charge des mesures et du technicien responsable de la saisie des données. Ce contrôle portera essentiellement sur:

- \* la calibration des appareils de mesures avant le départ en bateau, leur étalonnage entre chaque mesure et la répétition des mesures (x3)
- \* le respect des protocoles de prélèvement d'eau, de poissons et de tissus
- \* le bon codage des échantillons
- \* la validation de la saisie données collectées sur le terrain

Lors des expériences en laboratoire, la qualité et la conformité des données seront contrôlées par le responsable de campagne, assisté du technicien de laboratoire. Ce contrôle portera essentiellement sur:

- \* la calibration et le bon usage des appareils de laboratoire
- \* le respect des protocoles
- \* la traçabilité des échantillons
- \* la double lecture des résultats et leur saisie dans un cahier de laboratoire papier

### **3. Stockage et sauvegarde pendant le processus de recherche**

#### **Les données seront sauvegardés sous les trois supports :**

- Carnet de terrain papier : sauvegarde dans des armoires ignifugées du CEFE
- Tableurs : sauvegarde sur l'ordinateur portable terrain puis transmission par 4G d'une copie sur le serveur de stockage du CEFE (sauvegardes automatiques quotidiennes et contrôle des sauvegardes assuré par le service informatique du CEFE)
- Base de données : sauvegarde selon le principe 3-2-1 (3 sauvegardes sur 2 supports différents (SQL/ binaire) avec 1 lieu déporté (sur le campus de l'université de Montpellier))

#### **La même attention sera accordée aux données de contexte (stations / méthode de prélèvement / campagne / ...)**

#### **Les métadonnées (au format EML) seront sauvegardées sous trois supports :**

- sur le portail metacat du PNDB (<https://openstack-192-168-100-101.genouest.org/metacatui/data>)
- sur le portail Metashark du PNDB (<https://metashark.pndb.fr/>)
- sur les serveurs de la plateforme SIE du CEFE

Avec les trois mécanismes de sauvegarde mis en place, la sécurité des données est assurée. Rien ne s'opposera à leur récupération en cas d'incident.

Les données seront accessibles aux responsables des campagnes ainsi qu'à leurs collaborateurs au cours du processus de recherche. Elles seront modifiables uniquement par le responsable présent sur le bateau.

Les données ne seront disponibles en open acces qu'à la fin du projet après exploitation, analyse et publication et éventuelle période d'embargo.

Aucune donnée sensible. Pas de protection particulière.

## 4. Exigences légales et éthiques, codes de conduite

Les données à caractère personnel des intervenants lors des campagnes de terrain (bateau, équipage et capitaine, pêcheurs bénévoles) feront l'objet d'un consentement éclairé pour leur collecte au cours du projet. Ces données ne seront jamais diffusées. Ces informations utiles à des fins d'organisation des campagnes ne seront accessibles qu'aux responsables de campagne.

Le détenteur de la propriété intellectuelle des données est l'École Pratique des Hautes Études (EPHE) dont relève le porteur de projet. Les données d'observation, traitées et achevées seront librement accessibles (licence CC-BY libre d'utilisation à la condition de l'attribuer à l'auteur en citant son nom).

Le code d'éthique institutionnel de l'établissement dont relève le porteur de projet (EPHE) est appliqué.

L'étude envisagée est conforme à l'agrément préfectoral de l'établissement pour ce qui concerne les espèces animales prélevées.

Un comité d'éthique de scientifiques du CEFE a validé la nécessité scientifique du recours aux poissons vivants qui seront tous remis en liberté, ainsi que le choix de 5 espèces bio-indicateurs pour l'étude biométrique et le prélèvement de tissus, afin de mesurer l'impact des politiques mises en place sur le lac du Bourget pour réduire la pollution aquatique.

Le comité a constaté que les principes de la bientraitance sont respectés et que les conditions d'utilisation des animaux sont optimisées (principe des 3R « Replace, reduce and refine ») compte tenu des nécessités expérimentales. Le responsable de campagne et le technicien en charge de la manipulation des poissons et du prélèvement de tissu sont habilités en expérimentation animale.

Aucune étude préliminaire n'a été exigée puisque les protocoles expérimentaux proviennent du réseau des observatoires des Lacs de Montagnes Savoyards (OLMS) et sont couramment utilisés.

## 5. Partage des données et conservation à long terme

L'ensemble des données collectées dans le cadre de cette ANR (01/05/2020 - 30/04/2023) seront disponibles en libre accès (licence CC-BY) après l'embargo de publication finale des données traitées.

Les données seront mises à disposition par l'intermédiaire de l'entrepôt Zenodo (<https://zenodo.org/>), accessibles aussi sur le portail de données du réseau des observatoires des Lacs de Montagnes Savoyards (OLMS) (<https://data.olms.fr>) et dans la banque internationale Genbank pour les données génétiques (<https://www.ncbi.nlm.nih.gov/genbank/>).

Toutes les données seront ouvertes à l'exception des données personnelles qui seront détruites.

Les données de planification et d'organisation des campagnes, d'observations et de mesures et d'analyses en laboratoire seront sauvegardées sur les serveurs du CEFE pour la durée du projet et pendant 3 années supplémentaires. A l'expiration de cette échéance, seuls les carnets de terrain papier, les cahiers de laboratoire papier et la base de données seront conservés.

La pérennisation de la base de donnée sera assurée dans le cadre des missions de la plateforme SIE du CEFE.

Les données génétiques déposées dans Genbank seront préservées dans le cadre des missions du NCBI (<https://www.ncbi.nlm.nih.gov/genbank/>).

Les données seront mises à disposition par l'intermédiaire de l'entrepôt Zenodo (<https://zenodo.org/>) et accessibles aussi sur le portail du réseau des "Observatoires des Lacs de Montagnes Savoyards (OLMS)" (<https://data.olms.fr>) et dans la base de données Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>).

Aucun outil particulier n'est nécessaire en plus des outils classiques (tableur, lecture image, éditeur de texte, ...).

Le dépôt dans l'entrepot Zenodo générera automatiquement un DOI pour les jeux de données déposés.

Un mécanisme de DOI sur requête est en cours de développement sur le portail du réseau des "Observatoires des Lacs de Montagnes Savoyards (OLMS)". Une fois finalisé, ce mécanisme nous permettra de suivre l'utilisation de nos jeux de données.

Les données génétiques déposées dans Genbank seront identifiables par un numéro d'accession unique international.

## **6. Responsabilités et ressources en matière de gestion des données**

La gestion des données sera assurée par les porteurs du projet, à savoir Véronique Arnal et Marie-Claude Quidoz

Toutes les démarches mises en place tout au long de ce projet permettent de s'assurer que les données seront FAIR. La rédaction d'un 'datapaper' contribuera à la démarche de FAIRisation des données issues de ce projet.

Au vu du peu de ressources matérielles demandées, toutes les dépenses d'équipement sont pris en charge par la plateforme SIE du CEFE.