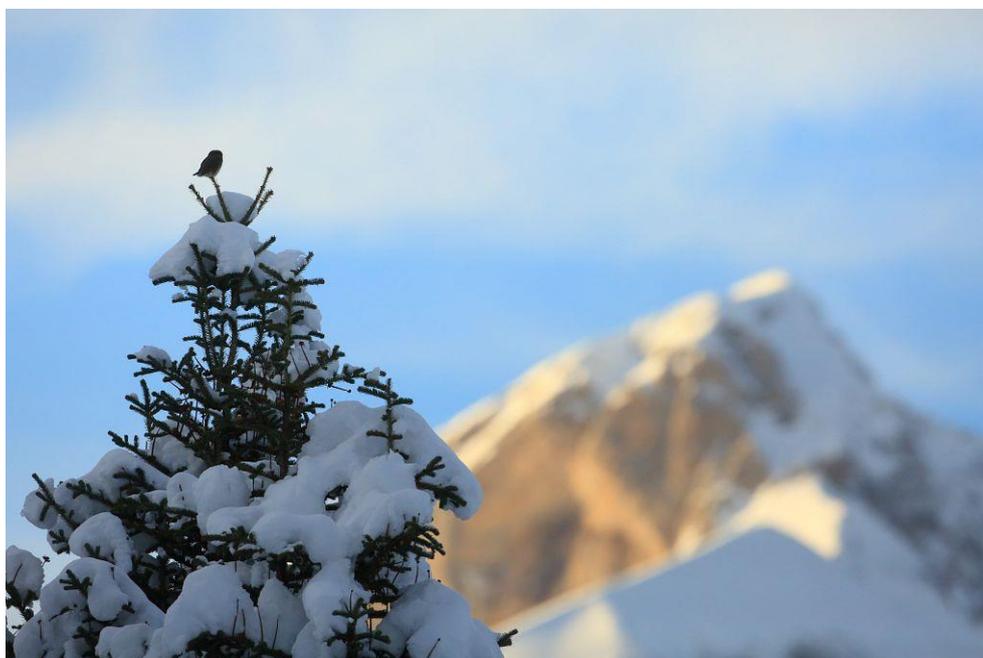




MÉMOIRE de PRÉ-SPÉCIALISATION **(2021)**

Valorisation des données de biodiversité sur le territoire transalpin (ALCOTRA)



Chevêchette d'Europe, Cabane des Pierres, Poligny, Champsaur
Copyright : Marc Corail, Parc National des Écrins

Mathilde LECLERC

Pré-spécialisation : NumAg - Numérique pour l'Agriculture
à Bordeaux Sciences Agro

Lieu du stage : Domaine de Charance - 05000 GAP

Tuteur école : Christian GERMAIN

Maitre de stage : Adrien PAJOT



Résumé

Entre la France et l'Italie se dresse l'un des plus grands massifs montagneux d'Europe : les Alpes. Recouvrant la frontière nord de l'Italie et le sud-est de la France, elles sont au cœur d'un projet transfrontalier, le PITEM Biodiv'ALP, visant à protéger et valoriser la biodiversité de ce territoire unique. Au sein de celui-ci, le projet COBIODIV SI a pour objectif d'améliorer l'interopérabilité et la valorisation des données de biodiversité. Ce document rapporte le travail effectué au cours de mon stage de 2^{ème} année à Bordeaux Sciences Agro dans le cadre de ce projet. L'objectif était de valoriser les données de biodiversité au travers d'un atlas de la biodiversité. Un atlas de biodiversité vise à faire l'inventaire des espèces de faune et de flore notamment, sur un territoire donné pour avoir une vision d'ensemble de l'état de la biodiversité observée à un instant t. Les objectifs principaux d'un atlas sont de mieux connaître la biodiversité afin de mieux la protéger, de sensibiliser et mobiliser le public, et de mieux prendre en compte la biodiversité et ses enjeux dans les actions ou projets territoriaux.

Le travail sur cet atlas a été réalisé en plusieurs étapes. La première étape, la phase d'analyse du projet composée d'une analyse des besoins, d'une analyse fonctionnelle, d'un état de l'art et d'une analyse technique réalisés seule ou en collaboration avec un autre stagiaire (Corentin Lange), a montré que le meilleur outil pour valoriser les données de biodiversité de la zone ALCOTRA est GeoNature-atlas. Cet outil open-source développé par le Parc National des Écrins s'est vu s'ajouter quelques fonctionnalités pour répondre aux besoins du projet. Le projet s'est basé sur plusieurs jeux de données reçues de différents partenaires : CBNA, CBNMed, Parc National du Mercantour, Parc National des Écrins, Flavia, Parco Naturale Alpi Marittime. Ces données (près de 3 millions) ont été intégrées, principalement par mapping PostgreSQL, dans une base GeoNature alimentant la base de l'outil GeoNature-atlas. Au terme de ce stage, couplé à celui de Corentin Lange, l'outil était fonctionnel et permet d'établir une preuve de concept de la faisabilité de ces atlas internationaux. Toutefois, il reste encore des thématiques à travailler pour le rendre complet comme celles des statuts de protection des espèces et la prise en compte des habitats dans ce genre d'atlas ainsi que dans les mesures de protection de la biodiversité.

Mots clés

Biodiversité, atlas, ALCOTRA, projet international, Alpes, Italie, organismes, base de données, SQL, intégration de données, référentiel spatial, développements web



Abstract

Between France and Italy stands one of the largest mountain ranges in Europe: the Alps. Covering northern border of Italy and south-eastern France are at the heart of a cross-border project, the PITEM Biodiv'ALP, aimed at protecting and enhancing the biodiversity of this unique territory. Within it, the COBIODIV SI project aims at the improvement of the interoperability and the enhancing of biodiversity data. This document reports on the work carried out during my Bordeaux Sciences Agro's second year internship in this project. The aim was the enhancing of biodiversity data from the ALCOTRA area by a biodiversity atlas. A biodiversity atlas aims to make an inventory of fauna and flora species, in a given territory to have an overall vision of the state of biodiversity observed at a precise moment. The main objectives of an atlas are to better understand biodiversity in order to better protect it, to raise awareness and mobilize the public, and to better take biodiversity and its issues into account in territorial actions or projects.

The work on this atlas was carried out in several stages. The first stage, the analysis phase of the project consisting of a needs analysis, a functional analysis, a state of the art, and a technical analysis carried out alone or collaboratively with another intern (Corentin Lange) has shown that the best tool for enhancing biodiversity data from the ALCOTRA area is GeoNature-atlas. This open-source tool developed by the Parc National des Écrins has been added some features to meet the needs of the project. The project was based on several datasets received from different partners: CBNA, CBNMed, Parc National du Mercantour, Parc National des Écrins, Flavia, Parco Naturale Alpi Marittime. These data (around 3 million) were integrated, mainly by PostgreSQL mapping, into a GeoNature database feeding the GeoNature-atlas tool database. At the end of this internship, coupled to the one of Corentin Lange, the tool was functional and allow to establish a proof of concept of the feasibility of these international atlases. Nevertheless, there are still issues to work on to make it complete, such as the status of species protection and the inclusion of habitats in such atlases as well as in biodiversity protection measures.



Remerciements

Je tiens, tout d'abord, à remercier mon maître de stage **Adrien Pajot** et **Camille Monchicourt** pour leur encadrement et cette superbe opportunité de stage au Parc des Écrins, sur un projet très intéressant. Je remercie également Adrien pour ses conseils et le temps qu'il a m'a consacré, ainsi que pour son partage d'expériences sur la formation AgroTIC, et l'application des nouvelles technologies dans le domaine de l'environnement.

Je souhaite également remercier **Théo Lechémia** pour son aide précieuse et le temps qu'il m'a consacré pendant mon stage sur les parties techniques.

Je veux aussi remercier l'ensemble de l'équipe du pôle SI : **Vincent Pietri**, **Elie Bouttier** et **Jean-Pascal Milcent** ainsi que les autres stagiaires du pôle : **Cendrine Hoarau**, **Baptiste Garcin** et **Corentin Lange**, pour m'avoir si bien accueillie et pour la super ambiance qu'ils ont mise.



Table des matières

- Résumé..... 1
- Mots clés..... 1
- Abstract..... 2
- Remerciements..... 3
- Table des figures 5
- Liste des tableaux..... 5
- Liste des abréviations 6
- Glossaire..... 6
- Introduction 8
- 1. Contexte 9
 - 1.1) L’organisme : Parc national des Écrins..... 9
 - 1.2) Le projet COBIODIV SI..... 9
 - 1.3) Les missions du stage : Valorisation des données de biodiversité..... 12
- 2. Phase d’analyse du projet 13
 - 2.1) Analyse des besoins 13
 - 2.2) État de l’art 14
 - 2.3) Analyse fonctionnelle 16
- 3. Étude des données de biodiversité..... 17
 - 3.1) Contacter les partenaires français et italiens..... 17
 - 3.2) Analyse des données transmises 18
 - 3.3) Conception d’une base de données de travail 19
 - 3.4) Synthèse de l’architecture choisie..... 21
- 4. Intégration des données 22
 - 4.1) Intégration des données d’observations..... 22
 - 4.2) Intégration de la géométrie de la zone d’étude..... 26
- 5. Développements techniques..... 27
 - 5.1) Intégration de la dimension « organismes » dans l’atlas 27
 - 5.2) Fonctionnalité : « Nouvelles espèces observées » 29
- 6. Bilan et perspectives..... 31
 - 6.1) Bilan de l’intégration des données..... 31
 - 6.2) Bilan du développement de l’outil..... 32
 - 6.3) Intégration d’un flux de données continu de sources agrégatives de données 33
 - 6.4) Extension du projet à l’intégralité des Alpes..... 34



6.5) Inclusion de la thématique des habitats au projet	34
Conclusion.....	35
Webographie.....	36

Table des figures

Figure 1: Carte du Parc National des Écrins et de ses secteurs	9
Figure 2: Cartographie de la zone ALCOTRA	10
Figure 3: Schéma de l'organisation du PITEM Biodiv'ALP.....	11
Figure 4 : Organigramme simplifié des structures impliquées dans les données de biodiversité et leurs flux	12
Figure 5: Schéma de répartition des missions du projet	13
Figure 6 : Modèle conceptuel de données	20
Figure 7: Schéma de l'outil de valorisation des données de biodiversité de la zone ALCOTRA.....	22
Figure 8: Script SQL de création de la source du jeu de données à importer.....	23
Figure 9: Exemple de script SQL faisant le mapping entre le jeu de données de la faune italienne et la synthèse.....	24
Figure 10: Exemple d'utilisation de conditions pour importer des données dans la synthèse (ici sur les référentiels spatiaux).....	24
Figure 11: Script sql de suppression des données hors ALCOTRA France	25
Figure 12: Géométries des territoires français et italiens.....	26
Figure 13: Visualisation des géométries de la zone ALCOTRA française et de la zone ALCOTRA italienne depuis QGIS.....	26
Figure 14: Géométrie de la zone ALCOTRA	27
Figure 15 : Exemple de fiche organisme (Parc National des Écrins).....	27
Figure 16: Script SQL des modifications de la base de données de l'atlas.....	28
Figure 17: Script de création de la classe organismes.....	29
Figure 18: Script HTML de la fonctionnalité "nouvelles espèces observées"	30
Figure 19: Front-end de la fonctionnalité "nouvelles espèces observées"	30
Figure 20 : Schéma des flux de données et des méthodes d'intégration.....	32
Figure 21: Carte de la délimitation de la convention alpine	34

Liste des tableaux

Tableau 1: Extrait du tableau récapitulatif des besoins de l'outil de valorisation des données de biodiversité	14
Tableau 2: Tableau d'analyse des fonctionnalités de l'outil GeoNature-atlas.....	15
Tableau 3: Extrait du tableau présentant les fonctionnalités de l'outil.....	16
Tableau 4: Extrait de l'évaluation fonctionnelle des outils.....	16
Tableau 5: Extrait de l'évaluation des fonctionnalités de l'outil.....	17
Tableau 6: Récapitulatifs du nombre de données par organismes.....	31



Liste des abréviations

AuRA : Auvergne-Rhône-Alpes

BDD : Base de données

CBNA : Conservatoire Botanique National Alpin

CBNMed : Conservatoire Botanique National Méditerranéen de Porquerolles

CEN : Conservatoire d'Espaces Naturels

INPN : Inventaire National du Patrimoine Naturel

PACA : Provence-Alpes-Côte d'Azur

PITEM : Plan intégré thématique

PNE : Parc National des Écrins

SI : Système d'information

SINP : Système d'information de l'inventaire du patrimoine naturel

ORCID : Open Researcher and Contributor ID

Glossaire

Biodiversité : La biodiversité est « la variabilité des êtres vivants de toute origine y compris, entre autres, les écosystèmes aquatiques et les complexes écologiques dont ils font partie : cela comprend la diversité au sein des espèces, ainsi que celle des écosystèmes » (*art.2 Convention sur la diversité biologique*)

Convention alpine : C'est une convention internationale visant à assurer la protection et le développement durable de la chaîne alpine. Elle a été signée en 1991, en Autriche, par l'Autriche, la France, l'Allemagne, l'Italie, le Liechtenstein, la Suisse et l'UE. Puis par la Slovénie en 1993. Elle est rentrée en vigueur en mars 1995.

Élus : Les élus rassemblent l'ensemble des hommes et des femmes politiques au niveau local (commune), départemental et régional.

Financeurs : Les financeurs regroupent l'ensemble des personnes et des structures finançant le projet COBIODIV.

Grand public : Le grand public définit l'ensemble des personnes sans qualifications spécifiques en biodiversité ayant donc des niveaux de connaissances et de compétences très variés dans le domaine, qui seront utilisateurs de l'outil par intérêt personnel pour la biodiversité.



Intégration de données : C'est l'importation de données de différentes sources au sein d'une même base de données.

Interopérabilité : C'est la capacité des outils informatiques/données de pouvoir communiquer et fonctionner ensemble afin de créer un réseau et faciliter les transferts de données.

Interreg : C'est un programme européen visant à promouvoir la coopération entre les régions européennes et le développement de solutions communes dans les domaines du développement urbain, rural et côtier, du développement économique et de la gestion de l'environnement.

Open-source : Fait référence à tout logiciel dont les codes sont ouverts gratuitement pour l'utilisation ou la duplication, et qui permet de favoriser le libre échange des savoirs informatiques.

Partenaires et délégataires : Ils regroupent l'ensemble des responsables, et des directeurs des différents pôles scientifiques des structures publiques de biodiversité intervenant dans la zone ALCOTRA du côté français.

Silene : Système d'Information de l'inventaire du patrimoine naturel de la région PACA.

Taxon : Un taxon est une entité regroupant l'ensemble des organismes vivants qui possèdent des caractéristiques communes définies. Un taxon peut-être une famille, un genre, une espèce, etc...

Valorisation des données : La valorisation des données est la création d'une plus-value aux données, par exemple en les traitant et en les exploitant. Notamment, à travers une interface utilisateur de visualisation de données brutes ou traitées, augmentant ainsi leur visibilité et leur intérêt.



Introduction

A travers la France, plusieurs espaces naturels sont protégés par des réglementations nationales. Au cœur des Alpes se situe le Parc National des Écrins, dont le siège est basé à Gap, dans les Hautes-Alpes. Sur ce site comme dans les autres espaces naturels protégés, l'objectif principal est la protection du patrimoine naturel et culturel du site. Les équipes accomplissent cette mission en partie en collectant des données de biodiversité.

Les données de biodiversité sont des données collectées par des agents de terrain ou par des capteurs, qui consistent en l'observation de la biodiversité que ça soit de la faune, de la flore ou bien même des habitats. La base d'une d'observation est composée de plusieurs champs principaux comme le nom du taxon observé, le ou les observateurs, la date et la localisation. D'autres critères comme le(s) sexe(s) de l'individu(s) (standard : SEXE) ou le stade phénologique (standard : STADE_VIE) complètent ces champs principaux. Ainsi, dans les espaces protégés français, des données de biodiversité sont collectées très régulièrement et à l'heure actuelle l'ensemble des données sont stockées numériquement en base de données principalement. Cette collecte d'information de biodiversité n'est pas unique à la France, et dans le cadre du projet de ce stage nous allons nous intéresser à la zone ALCOTRA, zone frontière située dans les Alpes françaises et italiennes.

L'objectif de ce stage est de travailler dans la continuité d'un projet visant à favoriser l'interopérabilité des données de biodiversité entre les acteurs en essayant de répondre à la question suivante :

Comment valoriser au mieux les données de biodiversité du domaine transalpin franco-italien ?

Ce rapport vise dans un premier temps à présenter le contexte du stage par une présentation de l'organisme d'accueil (le Parc National des Écrins), du projet COBIODIV SI, et des missions du stage. Ensuite, le travail effectué pendant le stage est décrit en quatre parties, en commençant par la phase d'analyse du projet, puis l'étude et l'intégration des données dans une base de données, et en finissant par les développements techniques. Le document se conclut par une partie faisant le bilan du stage et du développement de l'outil et de ses perspectives.

1. Contexte

1.1) L'organisme : Parc national des Écrins

Les parcs nationaux sont des organismes publics sous tutelle du ministère de l'Environnement et ont été créés dans l'objectif de protéger et préserver des espaces naturels exceptionnels face aux développements et aménagements qui menaçaient ces territoires aux paysages exceptionnels et riches en biodiversité. La France compte actuellement 11 parcs nationaux qui couvrent 9,5% du territoire français. Les parcs nationaux ne couvrent pas seulement des espaces montagnards, ils couvrent également des espaces maritimes (Parc national de Port-Cros, Guadeloupe, etc.) et forestiers (Parc national de forêts). Les parcs nationaux sont des territoires reconnus en France mais aussi internationalement pour leur patrimoine naturel, culturel et paysager exceptionnel.

Les parcs nationaux sont composés de trois pôles regroupés en un siège : un pôle aménagements et développement local, un pôle communication et accueil et un pôle scientifique. Le pôle scientifique réunit plusieurs filières : des naturalistes spécialistes de la faune, de la flore ou des écosystèmes, des scientifiques étudiant les paramètres physiques, et des spécialistes des systèmes d'informations parfois regroupés en un pôle « Système d'informations ».

Le Parc National des Écrins (PNE), créé en 1973, s'étend sur 252 000 hectares dont 91 800 ha font partie du cœur du parc : zone au sein de laquelle s'applique une charte stricte en termes d'actions et d'aménagements du territoire. Il est par exemple interdit d'y aller en voiture, de faire du camping ou de cueillir des fleurs ou des baies. Le parc est divisé en 7 secteurs, chaque secteur possède une maison de parc avec un chef de secteur et des gardes-moniteurs (des agents de terrains dont un de leur rôle est de faire respecter les lois environnementales dans la zone du parc). Le siège du PNE se trouve au domaine de Charance sur les hauteurs de Gap (Hautes-Alpes) au sein duquel j'ai effectué mon stage.

Au sein du Parc National des Écrins, se trouve un pôle SI. Celui-ci est composé d'un responsable (et aussi géomaticien) Camille Monchicourt, de deux développeurs Théo Lechémiat et Elie Bouttier, d'un administrateur système, réseaux, et téléphonie Vincent Pietri. Les missions principales du pôle sont notamment de gérer toutes les données acquises par les agents du parc, à travers des outils qu'ils ont notamment conçus : GeoNature (une base de données de biodiversité), GeoNature-atlas (un outil de valorisation de ces données) et Geotrek (un outil pour la randonnée). Ces outils, open-source, sont actuellement utilisés par de nombreux autres parcs et organismes impliqués dans la préservation de la biodiversité ainsi que par des collectivités.

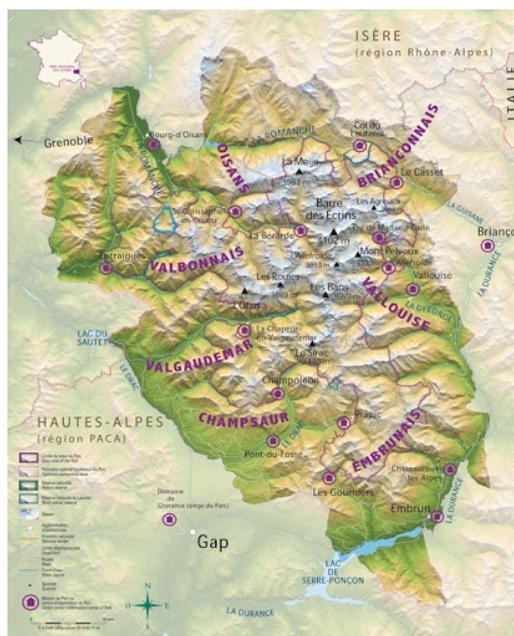


Figure 1: Carte du Parc National des Écrins et de ses secteurs

1.2) Le projet COBIODIV SI

Le stage se porte sur un axe du projet COBIODIV SI. Ce projet est l'une des cinq activités du projet COBIODIV qui fait lui-même partie des cinq grands projets du Plan Intégré Thématique (PITEM)

Biodiv'ALP. Le PITEM Biodiv'ALP est un projet transfrontalier européen entre la France et l'Italie. Il est piloté par Région Sud et se porte sur la biodiversité de la zone ALCOTRA.

Cette zone est un territoire transfrontalier entre la France et l'Italie, au cœur des Alpes. Elle s'étend en France sur deux régions (Auvergne-Rhône-Alpes et Provence-Alpes-Côte d'Azur) et particulièrement sur cinq départements (Savoie, Haute-Savoie, Hautes-Alpes, Alpes-de-Haute-Provence, Alpes-Maritimes). Du côté italien, la zone ALCOTRA chevauche trois régions (Région Autonome Vallée d'Aoste, Région Piémont, et la Région Ligurie avec la province d'Imperia). Ci-dessous une cartographie de cette zone (Figure 2).

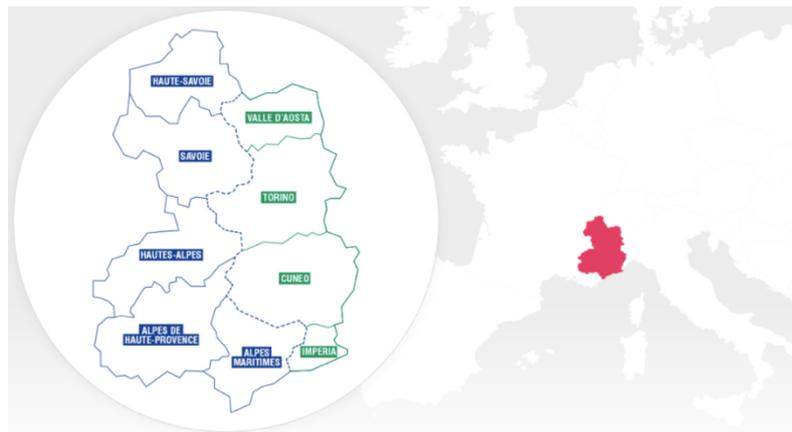


Figure 2: Cartographie de la zone ALCOTRA

Le PITEM a pour but de protéger et valoriser la biodiversité des Alpes. Il se décompose en cinq grands projets (voir Figure 3) dont le projet COBIODIV qui a pour objectif de « Connaître la biodiversité et les écosystèmes pour mieux les protéger ensemble ». L'activité 5, baptisé COBIODIV SI, dont sont chargés le Parc National des Écrins, Natural Solutions, et l'ARBE doit permettre la structuration des bases de données naturalistes dans une logique d'interopérabilité.

L'ARBE

L'agence régionale pour la biodiversité et l'environnement est une agence publique, de la région Sud, créée il y a 40 ans. Elle vise à favoriser les actions de préservation de la biodiversité et de la transition écologique. Ses missions s'articulent autour de 6 sujets majeurs : l'accompagnement des collectivités, la gestion des observatoires, l'animation de réseaux thématiques, la sensibilisation de différents publics, la contribution aux projets européens et internationaux, et la contribution à l'élaboration de stratégies régionales.

Natural Solutions

Natural Solutions est une agence web créée en 2008, qui met son expertise au service de la biodiversité. Elle détient une double compétence informatique et environnement lui permettant de conseiller et accompagner ses clients dans divers projets. Elle est notamment spécialisée dans le développement d'outils pour la collecte, la gestion, le stockage et l'exploitation des données environnementales. Elle a notamment participé aux développements des outils open source GeoNature et GeoNature-atlas, dont il est question dans ce rapport.

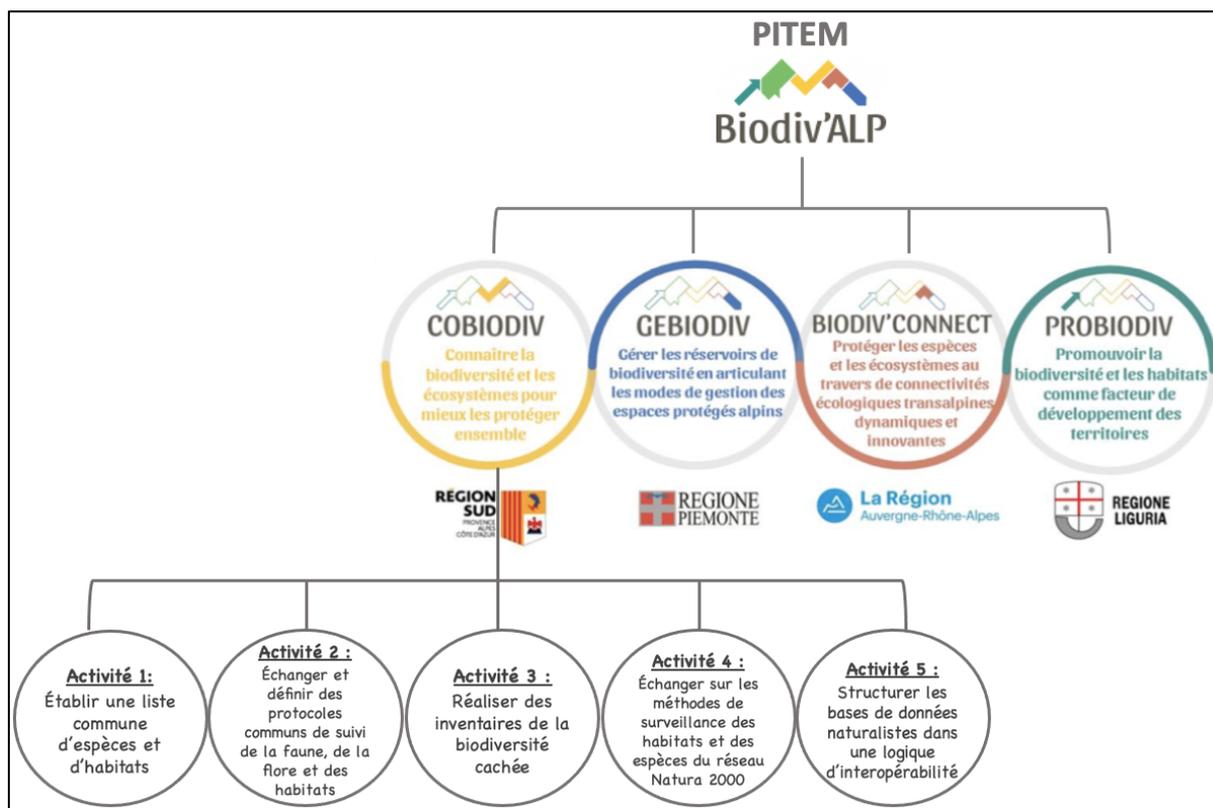


Figure 3: Schéma de l'organisation du PITEM Biodiv'ALP

Le projet COBIODIV SI se divise en deux phases. La première phase, co-pilotée par l'ARBE, Natural Solution et le PNE, consistait en un diagnostic des systèmes d'informations des différents acteurs de la biodiversité intervenant dans la zone ALCOTRA. Celle-ci a été effectuée en France par Adrien Pajot en 2020 et a abouti à une synthèse ^[1]. Du côté Italien, la situation sanitaire a fortement retardé le début de cette phase qui a donc commencé en 2021 et dont les résultats ont été connus fin juillet 2021.

La deuxième phase consiste à proposer une stratégie (et la mettre en œuvre) afin de rendre l'ensemble des données de biodiversité des Alpes interoperables à la suite du diagnostic effectué lors de la première phase. L'interopérabilité des données en France comme en Italie permettrait une meilleure valorisation de ces dernières. C'est sur cet objectif que se porte le stage.

L'objectif de cette partie du projet est de valoriser les données de biodiversité des Alpes tant du côté français que du côté italien. Plusieurs problématiques interviennent dans ce projet. En effet, d'une part, si l'on s'intéresse au côté français, les données de biodiversité des Alpes sont acquises par plusieurs organismes qui sont publics ou privés. Tous ces acteurs, agissant pour la biodiversité, ont des contextes et des problématiques propres et tous n'utilisent pas les mêmes outils de gestion de données ^[1] ce qui complexifie le travail. C'est là qu'intervient l'enjeu d'interopérabilité du projet. Le diagnostic effectué lors de la première phase ^[1] a révélé que la plupart des organismes interrogés utilisent l'outil GeoNature conçu par le Parc National des Écrins. En effet, cet outil est utilisé par 12 structures sur les 20 interrogées. Les autres outils utilisés sont SERENA, Biolovision, des bases métiers pour les CBN, SICEN, Hélix etc.

D'autre part, des problématiques se dégagent aussi au niveau de l'agrégation des données. Au niveau français, les données de biodiversité sont agrégées par le SINP national. C'est un organisme public qui a pour but de rassembler l'ensemble des données de biodiversité et d'habitats notamment,

provenant de toutes les structures qui reçoivent des financements publics (obligation légale : loi Lemaire 2016). Le SINP possède des déclinaisons régionales. De fait, dans la zone ALCOTRA il y a deux SINP impliqués : SILENE pour la région PACA et le SINP AuRA qui est décomposé en quatre pôles (PIFH, invertébré, gestion des milieux naturels, faune). Ce sont normalement les SINP régionaux qui collectent les données de toutes les structures régionales, les mettent au format unique souhaité par le SINP national pour ensuite lui envoyer. Cependant, certaines structures n'envoient pas leurs données au SINP régional puisqu'elles ne sont pas structurées pour ou préfèrent les envoyer directement au SINP national, pour diverses raisons.

Les flux de données de biodiversité, en France, et les liens entre les différentes structures impliquées dans cette thématique peuvent se schématiser (Figure 4). Ce schéma ne fait pas un recensement exhaustif de l'ensemble des structures de biodiversité de la zone ALCOTRA et encore moins de la France. Il a été conçu à partir des compte-rendu de la première phase du projet et représente ainsi uniquement les structures rencontrées lors de cette dernière. Nous nous représentons ainsi un faible échantillon des sources de données à valoriser, au travers d'outils tels que les atlas.

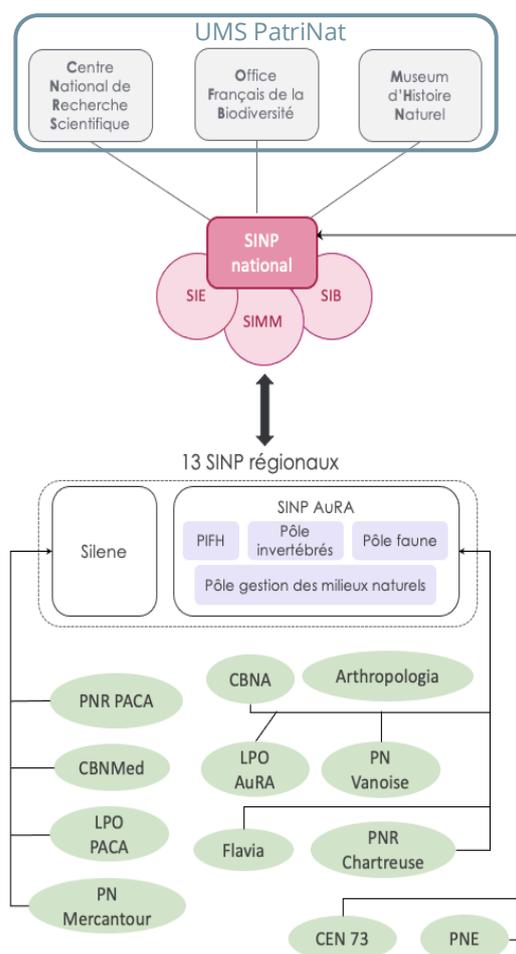


Figure 4 : Organigramme simplifié des structures impliquées dans les données de biodiversité et leurs flux

Enfin, durant la première phase du projet COBIODIV SI, la question d'un atlas de la biodiversité avait été en partie étudiée et la plupart des acteurs soulignaient l'importance d'un atlas pour valoriser les données de biodiversité des Alpes et notamment pour le partage transfrontalier des données, « puisque la biodiversité ne connaît pas de frontière ». Cette question a cependant soulevé plusieurs interrogations sur les modalités de cet atlas, chaque partenaire et délégataire a fait remonter au cours de ces entretiens leurs besoins concernant la valorisation de leurs données de biodiversité, c'est pourquoi ce stage a été conçu : comprendre les enjeux autour de la valorisation des données de biodiversité.

1.3) Les missions du stage : Valorisation des données de biodiversité

Dans le cadre du projet COBIODIV SI, une demande des financeurs et des partenaires du projet est la réalisation d'un outil de valorisation des données de biodiversité de la zone ALCOTRA en coopération avec l'Italie. Plusieurs acteurs concernés par ce projet ont été identifiés : les partenaires et délégataires du projet (c'est-à-dire des structures impliquées dans la biodiversité), les financeurs, les élus et le grand public. L'outil doit s'appuyer sur les données de biodiversité de plusieurs structures françaises présentes dans les Alpes comme les parcs nationaux, les CBN, les SINP régionaux mais aussi des associations ainsi que sur des données provenant de structures italiennes. L'outil a ainsi pour

objectif de valoriser des données de plusieurs BDD/sources différentes n'ayant pas les mêmes structures. Par conséquent, il doit répondre à deux enjeux majeurs. Premièrement, un enjeu de gestion de données et, deuxièmement, un enjeu de valorisation de données en matière de traitement et de visualisation des données.

Le stage consiste à travailler sur ces deux enjeux en travaillant sur la problématique « données de biodiversité » et en déterminant l'outil le plus adapté puis en développant dessus. Les deux objectifs s'avérant très importants, un second stagiaire Corentin Lange, a été recruté pour travailler plus sur l'aspect développement informatique.

Ayant commencé mon stage deux semaines avant l'arrivée de Corentin, je me suis occupée d'une phase d'analyse du projet comportant une analyse des besoins, un état de l'art des outils de valorisation des données de biodiversité, et une analyse fonctionnelle. Ces travaux seront développés ci-dessous. La phase d'analyse s'est achevée par une analyse technique effectuée par Corentin. À la suite de cela, nos missions respectives ont été distinctement séparées, Corentin s'occupant du côté technique avec les développements de l'outil. L'analyse et la gestion des données ainsi que leur intégration dans la base de données m'ont été confiées. Le schéma ci-contre synthétise les missions du projet qui ont été attribuées à Corentin (en bleu) et moi-même (en vert). Chacune de mes missions seront développées et illustrées dans la suite du rapport.

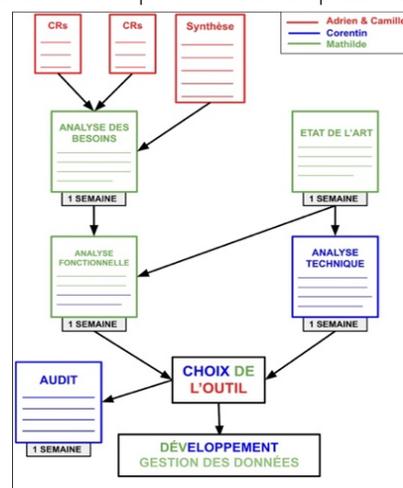


Figure 5: Schéma de répartition des missions du projet

2. Phase d'analyse du projet

Après quelques jours de recherches et de lecture des travaux déjà effectués par Adrien PAJOT pour m'approprier le contexte du projet COBIODIV SI, j'ai commencé une phase d'analyse du projet en effectuant une analyse des besoins.

2.1) Analyse des besoins

Lors de la première phase du projet, des entretiens ont été menés auprès des partenaires et délégataires du projet par Adrien PAJOT (Natural Solutions & Parc national des Écrins). Ces entretiens ont notamment permis de recueillir les besoins de ces structures et leurs attentes en ce qui concerne l'outil de valorisation et ses fonctionnalités. À l'issue de ceux-ci, des comptes rendus ont été rédigés et un premier regroupement des besoins exprimés a été fait et publié dans la synthèse^[1] de la phase 1 du projet.

La première étape de l'analyse des besoins^[2] a consisté en l'étude des comptes rendus dans le but de répertorier l'ensemble des besoins exprimés et de les préciser. Par la suite, les besoins répertoriés ont été détaillés et synthétisés dans un tableau récapitulatif reliant chaque besoin aux différents acteurs identifiés. Un système de pondération de ces acteurs permet de prioriser les différents besoins. Cinq catégories d'acteur ont été identifiées : les partenaires et délégataires, les financeurs, les scientifiques, les élus, et le grand public. Ci-dessous un extrait de ce tableau synthétique des besoins (Tableau 1).

Famille de besoin	Besoin	Détail du besoin	Membres décisionnaires des partenaires et délégataires	Financeurs	Scientifiques/Naturalistes	Élus	Grand public	Total
Portée	Atlas montrant les enjeux de biodiversité des Alpes	Éveiller des consciences écologiques	4	3	3	2	1	13
	Apporter des connaissances de biodiversité aux utilisateurs	Mise à disposition de fiches espèces, de photos, de diverses informations vulgarisées sur la biodiversité alpine	4	3	X	2	1	10
	Mise en avant des espèces endémiques et des espèces envahissantes	Visualisation des espèces endémiques pour potentiellement mettre en place des mesures spécifiques (OAD)	4	X	3	2	1	10
	Atlas avec des données métriques	Mise en avant des lacunes de prospection dans certaines zones (OAD)	4	X	3	2	1	10
	Statut des données	Possibilité de voir le statut des données (validée/invalidée, données Citizen...)	4	X	3	2	1	10
	Atlas des indicateurs de biodiversité	Évolution de la biodiversité → prioriser les mesures de conservation (OAD)	4	X	3	2	X	9
	Lier les données habitats et les données de biodiversité	La conservation des espèces ne peut se faire indépendamment de celle des habitats. Il est donc important d'avoir un outil faisant le lien entre les deux pour optimiser les mesures de conservation.	4	X	3	2	X	9
	Atlas transfrontalier	Partage de l'outil (conception de l'outil adaptable aux données italiennes : formats...)	4	3	X	2	X	9

Tableau 1: Extrait du tableau récapitulant les besoins de l'outil de valorisation des données de biodiversité

Cette analyse a permis de dégager l'ensemble des besoins auxquels l'outil doit répondre et sert de base pour la définition de l'outil et de ses fonctionnalités. Cette analyse a défini précisément les objectifs et les attentes de ce projet.

Ce tableau a notamment permis de conclure que le projet nécessite la création d'un atlas transfrontalier qui utiliserait les données italiennes et les données françaises, celles-ci provenant à terme principalement des BDD des SINP régionaux (PACA et AuRA). L'objectif n'est donc pas de créer une nouvelle base de données à partir de toutes les bases des organismes présents dans les Alpes (qui est le rôle du SINP) mais de concevoir un outil de valorisation, à partir de différentes bases de données rendues interopérables dans un format simple. Cet outil permettant de rendre accessible les données naturalistes au grand public mais aussi aux scientifiques et aux élus qui pourront s'appuyer dessus pour des décisions d'aménagements du territoire, de protection et de préservation de l'environnement.

Le document d'analyse des besoins est disponible à l'adresse suivante (<https://geonature.fr/documents/2021-06-PITEM-BiodivAlp-COBIODIV-SI-Atlas-Analyse-besoins.pdf>).

2.2) État de l'art

En parallèle de cette analyse du besoin, j'ai effectué un état de l'art^[3] des outils permettant de répondre complètement ou en partie aux besoins qui sont exprimés dans cette analyse. L'objectif étant de ne pas réinventer un outil qui existe déjà mais d'identifier les outils ou les fonctionnalités des différents outils sur lesquels le projet pourra s'appuyer.

En général, les données de biodiversité sont valorisées à travers des atlas. Un atlas de biodiversité est un portail de visualisation des données de biodiversité permettant de rendre accessible certaines informations comme la présence d'une espèce.

Ce document porte sur l'analyse de plusieurs outils de valorisation de données de biodiversité déjà existants et accessibles sur internet (6 atlas et 1 tableau de bord). Cette analyse fait une étude sur un panel d'outils qui s'appliquent à des échelles géographiques différentes (communale, régionale, nationale, internationale) et à destination de publics différents (citoyen, naturalistes, élus, etc...). Pour chaque outil étudié, un tableau récapitulatif est proposé présentant les informations accessibles, les fonctionnalités, quelques aspects techniques, les pistes d'améliorations et les « manques ». Ci-dessous, un exemple de ce tableau pour l'outil GeoNature-atlas (Tableau 2).

Points forts	
Informations accessibles : <ul style="list-style-type: none"> • Dashboard avec informations générales sur l'atlas (nombre d'observations, nombre d'espèces, nombre de communes, nombre de photographies) • Carte des observations des sept derniers jours • Listes des espèces observées pendant les sept derniers jours • Fiches espèces : <ul style="list-style-type: none"> - nom scientifique - rangs taxonomiques - nombre d'observations totales et par maille - nombre d'observateurs - nombre de communes - année de première observation - année de dernière observation - fichier audio et vidéo - information général sur l'espèce - quelques photos - Graphique des observations par altitudes - Graphique des observations par mois • Galerie • Fiches par communes et par groupe taxonomique • Suggestions/Conseils aux utilisateurs (Espèces à découvrir, A voir en ce moment) 	Fonctionnalités : <ul style="list-style-type: none"> • Carte interactive • Découpage par maille et localisation précise des observations → Précision s'adapte à l'échelle de la carte (zoom) • Fond de carte : orthophotographie, plan • Choix de la plage temporelle (année) • Lien vers des articles
Aspect technique : <ul style="list-style-type: none"> • Open-source et générique • Utilisation d'une BDD GeoNature (outil très utilisé par les structures gérant des données de biodiversité) • Performant Ergonomie : <ul style="list-style-type: none"> • Légende transparente (pour voir les informations cartographiques dessous) • Affichage de la carte en plein écran Navigation simple 	
Points à améliorer	Points faibles
<ul style="list-style-type: none"> • Dégradation des données à la maille (remarque du Parc National du Mercantour) • Pas de recherches par famille ou genre 	Les points suivants sont des informations ou fonctionnalités qui n'apparaissent pas dans cet outil mais qui peuvent être intéressantes d'ajouter. <ul style="list-style-type: none"> • Pas d'échelle sur la carte • Surface des mailles non indiquée • Informations et cartes non téléchargeables

Tableau 2: Tableau d'analyse des fonctionnalités de l'outil GeoNature-atlas

Il est également possible de trouver pour chaque outil un lien permettant de le visualiser ainsi qu'un lien vers la page GitHub de leurs codes pour les outils open-sources.

À la suite de l'étude des sept outils, une synthèse comparative des outils est présentée, sous forme de tableau. Ce tableau a pour but de synthétiser les fonctionnalités, les informations accessibles, etc.... de chaque outil afin d'identifier les similitudes et les différences de ces outils et ainsi aiguiller le choix d'un futur organisme recherchant un outil pour valoriser ses données. Ce tableau permet également de déterminer, dans le cadre du projet, l'outil qui comptabilise le plus de fonctionnalités intéressantes pour répondre aux besoins.

Cette étude a montré que l'ensemble de ces outils sont ergonomiques et pour la plupart très performants. L'ensemble des outils possèdent des caractéristiques générales que l'on peut considérer comme des fonctionnalités basiques d'un outil pour valoriser les données de biodiversité. Certains outils se démarquent des autres par leur nombre de fonctionnalités développées notamment par des fonctionnalités de paramétrage des informations visualisées mais aussi tout simplement par les informations présentées et accessibles aux utilisateurs/ visiteurs de l'outil. Le document est disponible à l'adresse suivante (<https://geonature.fr/documents/2021-06-PITEM-BiodivAlp-COBODIV-SI-Atlas-Etat-art.pdf>).

2.3) Analyse fonctionnelle

À la suite des deux documents précédents, a été effectuée l'analyse fonctionnelle d'un outil de valorisation des données de biodiversité pour la zone ALCOTRA. Cette analyse propose d'étudier les fonctionnalités que l'outil de valorisation des données de biodiversité de la zone ALCOTRA pourrait posséder afin de répondre aux différents besoins des acteurs du projet. Une fonctionnalité est un développement permettant à l'utilisateur d'effectuer une ou plusieurs actions.

Après avoir présenté le contexte et la démarche, ce document propose un tableau exprimant les besoins sous forme de « User Stories » (récit utilisateur) et la ou les fonctionnalités qui répond(ent) à chaque besoin comme montré dans l'extrait ci-dessous (Tableau 3).

Famille de besoin	ID	Besoin	Total	User story			ID fonctionnalité	Fonctionnalité associée
				En tant que...	Je veux...	Afin de...		
BP1	Atlas montrant les enjeux de biodiversité des Alpes	13	Partenaires ou financeurs du projet	montrer les enjeux de la biodiversité dans la zone ALCOTRA	éveiller des consciences écologiques	BP2F1	Affichage du statut de l'espèce (liste rouge IUCN=endémiques) sur la fiche espèce + affichage spécifique sur la carte	
						BP2F2	Chiffres clés sur les espèces à statut (protégées, réglementées, menacées, endémiques ou endémiques et menacées) : leur nombre sur le territoire, nombre d'individus d'espèces à statut observés morts / vivant	
						BP2F3	Pourcentage de surface protégée sur la zone ALCOTRA	
						BP2F4	Carte des observations avec possibilité d'afficher les observations de une ou plusieurs espèces au choix, ou d'un taxon	
BP2	Apporter des connaissances de biodiversité aux utilisateurs	10	Utilisateur de la plateforme	connaître des détails sur les différentes espèces	découvrir les caractéristiques d'une espèce	BP2F1	Fiches espèces (nom scientifique, statuts, infos générales, informations des observations (carte), lien vers les fiches observations, graphiques)	
						BP2F2	Fiches observations (type de contact, état biologique, type de prospections)	
						BP2F3	Fiches sur les organismes impliqués dans la gestion de la biodiversité (PN, PNR, ORL, etc...) avec lien vers leur site web	
						BP2F4	Galerie photographique	
						BP2F5	Galerie vidéos et/ou audio	
						BP2F6	Fiche espèce endémique / Valorisation une espèce par semaine / Carnoual des espèces endémiques (blog)	
BP3	Mise en avant des espèces endémiques et des espèces envahissantes	10	Utilisateur de la plateforme	pouvoir identifier facilement les espèces endémiques à la zone ALCOTRA	mettre en avant la faune/flore locale	BP3F1	cf BP2F1	
				pouvoir identifier les espèces envahissantes dans la zone ALCOTRA	connaître les problématiques environnementales de la biodiversité	BP3F2	Fiches espèces envahissantes, affichage de ce statut sur les fiches espèces	
				des données métriques sur l'ensemble de la zone ALCOTRA	visualiser les lacunes de prospection et essayer d'y remédier	BP3F3	Pourcentage de territoire maillé (au total, à l'espèce)	
BP4	Atlas avec des données métriques	10	Utilisateur de la plateforme	des données métriques par unité		BP4F1	Graphiques du nombre d'observations en fonction de classes d'altitudes (à l'espèce)	
				des données métriques par commune		BP4F2	Nombre d'observations, nombre d'observation par espèces, nombre d'espèces observées, nombre d'observateurs	

Tableau 3: Extrait du tableau présentant les fonctionnalités de l'outil

À la suite de ce tableau, un second tableau reposant sur l'état de l'art, évalue qualitativement la capacité fonctionnelle des outils étudiés à répondre aux besoins du projet, en utilisant un système de couleur (vert : l'outil répond au besoin, orange : l'outil répond partiellement au besoin, rouge : l'outil ne répond pas au besoin) comme présenté par l'extrait ci-dessous (Tableau 4).

Famille de besoin	ID	Besoin	Total	Biodiv'territoires	GeoNature-Atlas	GeoNature Dashboard	Atlas des oiseaux de France	Kollect	Faunapyr
Portée	BP1	Atlas montrant les enjeux de biodiversité des Alpes	13						
	BP2	Apporter des connaissances de biodiversité aux utilisateurs	10						
	BP3	Mise en avant des espèces endémiques	10						
	BP4	Atlas avec des données métriques	10						
	BP5	Statut des données	10						
	BP6	Atlas des indicateurs de biodiversité	9						
	BP7	Lier les données habitats et les données de biodiversité	9						

Tableau 4: Extrait de l'évaluation fonctionnelle des outils

Ce document se conclut par un tableau évaluant chaque fonctionnalité afin de les prioriser et de sélectionner les fonctionnalités indispensables à développer pour répondre aux attentes du projet dans le temps imparti. Un extrait de ce tableau (Tableau 5) est présenté ci-dessous.

ID fonctionnalité	Fonctionnalité associée	Notes				Valeur
		Importance	Complexité	Temps		
				Programme	Nombre d'étapes	
BP1F1	Affichage du statut de l'espèce (liste rouge UICN+endémiques) sur la fiche espèce + affichage spécifique sur la carte	3	4	5	3	28
BP1F2	Chiffres clés sur les espèces à statut (protégées, réglementées, menacées, endémiques ou endémiques et menacées) : leur nombre sur le territoire, nombre d'individus d'espèces à statut observées morts / vivant	3	6	5	3	30
BP1F3	Pourcentage de surface protégée sur la zone ALCOTRA	1	6	5	3	28
BP2F1	Carte des observations avec possibilité d'afficher les observations de une ou plusieurs espèce au choix, ou d'un taxon	3	8	5	5	31
BP2F2	Fiches espèces (nom scientifiques, statuts, infos générales, informations des observations (carte), lien vers les fiches observations, graphiques)	3	8	1	3	25
BP2F3	Fiches observations (type de contact, état biologique, type de prospections)	2	8	5	3	28
BP2F4	Fiches sur les organismes impliqués dans la gestion de la biodiversité (PN, PNR, ORB, etc...) avec lien vers leur site web	1	8	2	3	24
BP2F5	Galerie photographique	2	8	0	5	25
BP2F6	Galerie vidéos et/ou audio	1	10	5	5	31
BP3F1	Fiche espèce endémique / Valorisation une espèce par semaine / Carrousel des espèces endémiques (blog)	2	10	2	5	29
BP3F2	Cf BP1F1	Cf BP1F1	Cf BP1F1	Cf BP1F1	Cf BP1F1	Cf BP1F1
BP3F3	Fiches espèces envahissantes, affichage de ce statut sur les fiches espèces	2	4	3	3	22
BP4F1	Pourcentage de territoire maillé (au total, à l'espèce)	3	6	3	3	25
BP4F2	Graphiques du nombre d'observations en fonction de classes d'altitudes (à l'espèce)	2	4	0	3	19

Tableau 5: Extrait de l'évaluation des fonctionnalités de l'outil

Cette analyse permet de mettre en évidence trois outils parmi les six étudiés, GeoNature-atlas, Collect Obs'Nat et Biodiv/Territoires. Ces trois outils open source, ont été étudiés au cours d'une analyse technique réalisée par Corentin Lange (<https://geonature.fr/documents/2021-06-PITEM-BiodivAlp-COBIODIV-SI-Atlas-Analyse-technique.pdf>).

L'analyse fonctionnelle mise en parallèle avec l'analyse technique^[5], a permis de choisir l'outil sur lequel nous nous sommes appuyés. Ces analyses ont permis de conclure que fonctionnellement et techniquement, l'outil GeoNature-Atlas semble le plus adapté. C'est un outil open-source développé par Théo Lechémi, en 2016, dans le cadre d'un stage au Parc National des Écrins. Cet outil est depuis lors repris par de nombreuses structures/collectivités à travers la France mais aussi à l'étranger. Il a donc été retenu afin d'être déployé sur la zone ALCOTRA et des fonctionnalités supplémentaires lui ont été apportées au cours de l'été, en lien avec ces documents.

Maintenant que l'outil a été choisi, il a fallu d'une part le déployer sur la zone ALCOTRA et développer les nouvelles fonctionnalités (travail de Corentin) et d'autre part, lui intégrer les données de biodiversité (mon travail).

3. Étude des données de biodiversité

3.1) Contacter les partenaires français et italiens

La première étape de la phase d'intégration des données a été de recontacter l'ensemble des organismes collectant des données de biodiversité aussi bien français qu'italiens afin de leur demander de nous transmettre des échantillons de leurs données. Ces échantillons permettant d'étudier la structure des données et leur format. Tout ceci, dans le but de concevoir une base de données de travail la plus adaptée possible à l'ensemble des données de la zone ALCOTRA. Ces échanges avec les partenaires nous ont permis de rassembler plusieurs jeux de données aussi bien français qu'italiens pour ensuite les intégrer à la base de données et tester les performances de l'outil.



3.2) Analyse des données transmises

À la suite des demandes d'échantillons de données aux différents partenaires, nous avons reçu plusieurs jeux de données de la part de partenaires français mais aussi de partenaires italiens qu'il a fallu analyser.

- **Étude des standards et référentiels français**

En France, il existe plusieurs référentiels et standards de nomenclatures permettant de cadrer les différentes données naturalistes, cela permet d'avoir des données de biodiversité uniformes au niveau national. Les données françaises de biodiversité devraient toutes se trouver dans ce format peu importe l'observateur, l'organisme ou même le lieu des observations. L'ensemble de ces standards et référentiels sont développés par l'INPN. Nous noterons toutefois que seul l'outil GeoNature est basé sur ce standard et que les autres bases de données ont leur propre structure. En ce qui concerne les standards de nomenclatures, ils régissent principalement le format des métadonnées et complètent les observations avec des informations comme le sexe, l'état biologique, le stade phénologique de(s) individu(s), technique d'observation, etc... La liste des standards français de nomenclatures est disponible à l'adresse suivante : <http://standards-sinp.mnhn.fr/nomenclature/>

Ajouté à la problématique des données, vient la problématique des espèces. En effet, afin que chaque structure appelle un taxon par son « vrai nom », un référentiel national a été établi en France. Ce référentiel taxonomique est TAXREF. Il contient l'ensemble des taxons officiellement reconnus en France et associe à chaque taxon un identifiant, un nom scientifique, et les noms vernaculaires français entre autres informations.

Enfin, d'autres problématiques doivent être gérées au niveau national comme le statut de protection d'une espèce. Ainsi l'INPN a également conçu une base de connaissance statuts qui recense pour chaque espèce l'ensemble des statuts de protection à plusieurs niveau géographique (départemental, régional, national, européen, international, etc..).

L'étude de ces référentiels et de ces nomenclatures m'ont permis de me familiariser avec les données de biodiversité et leur structuration en France, et ainsi comprendre les jeux de données envoyés par les organismes français.

- **Visualisation et analyse des données**

Les données ont été visualisées sur Rstudio afin de comprendre et connaître la structure et le format des jeux de données. En effet, il est possible sur Rstudio d'importer des gros fichiers Excel ou csv pour y afficher le contenu comme sur un tableur. Ainsi il a été possible d'identifier pour chaque jeu de données les informations présentes et les informations manquantes pour répondre aux différentes fonctionnalités et/ou par rapport à d'autres jeux de données. Cela a également permis de faire une première correspondance entre les libellés des nomenclatures françaises et les informations italiennes recensées.

Les données ont également été visualisées sur QGIS (version 3.14) avec des couches vectorielles des limites administratives de la France et de l'Italie afin de visualiser géographiquement les données des différents partenaires en France et en Italie afin d'entre autres vérifier que les données soient bien localisées dans la zone ALCOTRA ainsi que le référentiel spatial des coordonnées des observations.

- **Les problématiques dégagées**

L'étude des différents jeux de données ont permis de dégager plusieurs problématiques à l'intégration de données de formats et de structures différents. Tout d'abord, du côté français



l'ensemble des organismes, ayant fourni des données, stockent leurs données dans une base GeoNature. Ainsi l'ensemble des jeux de données ont la même structure. Cependant, dans la base GeoNature, les nomenclatures peuvent être sous forme de libellé ou d'identifiant. Par conséquent, dans les jeux de données reçu il y avait deux formats différents. En revanche, du côté italien, le parc des alpes maritimes nous a envoyé quatre jeux de données. Chacun étant différent structurellement des autres.

D'autres problématiques ont été dégagées au niveau de la correspondance entre données françaises et italiennes.

L'une des premières et principales problématiques identifiées est la problématique du référentiel spatial de la localisation des données d'observations. En effet, en France, la norme est d'utiliser le SRID 2154 soit la projection Lambert 93 tandis que les données italiennes suivent le SRID 23032 soit la projection ED50/UTM32. Cependant, il existe un système de coordonnées international, le SRID 4326 permettant d'utiliser un même un système de coordonnées pour les données italiennes et françaises.

Une autre problématique majeure est celle du référentiel taxonomique. En France, comme expliqué précédemment, il existe un référentiel taxonomique (TAXREF) sur lequel se base tous les relevés de biodiversité permettant d'uniformiser les données et ainsi identifier chaque individu et chaque espèce par un nom scientifique commun unique pour l'ensemble des scientifiques français. Cependant, ce n'est pas le cas en Italie. Aucun référentiel taxonomique n'est utilisé et seulement quelques structures stockent leurs données en base, les autres sont stockés sous forme de fichiers Excel. De fait, les espèces ne sont pas identifiées nationalement par un code unique et les jeux de données italiens sont soumis aux erreurs d'« écriture » lors de la saisie des données dans des fichiers Excel. Étant donné que la biodiversité ne connaît pas ou peu de frontières, il est peu probable que des espèces vues en zone ALCOTRA côté italien ne soient pas recensées dans le référentiel français. Par conséquent, il a été décidé pour l'outil d'utiliser le référentiel TAXREF pour faire la correspondance entre les espèces françaises et italiennes. Une espèce seulement était présente en Italie et non référencée en France.

Enfin, les jeux de données de biodiversité contiennent des informations complémentaires à la seule information de « qui a vu quoi, quand et où ». Il est par exemple renseigné le sexe de(s) individu(s), le statut biologique, le stade phénologique, etc.... qui sont régis en France par les standards de nomenclatures, expliqué ci-dessus, mais ce n'est pas le cas en Italie. En effet, dans les quatre jeux de données ces informations ne sont pas stockées de la même manière, ni avec la même orthographe ou libellé. Ces informations étant, d'autre part, peu renseignées il a été fait le choix de ne conserver, pour les données italiennes, que les informations principales des observations (qui, quand, où, quoi). D'autant plus, que les données complémentaires sur les observations ne font pas l'objet de développements de fonctionnalités de l'outil (dans celles sélectionnées pour le projet) et ne sont donc pas essentielles à intégrer dans l'outil.

3.3) Conception d'une base de données de travail

L'étude des jeux de données, et des standards français utilisés pour les données de biodiversité a permis de faire un premier listing des informations disponibles et des informations potentiellement manquantes pour répondre aux fonctionnalités dégagées par l'analyse du projet. Cela a permis de concevoir un modèle de données permettant de répondre au maximum aux demandes

du projet décrites par les différentes fonctionnalités listées dans l'analyse fonctionnelle. Le modèle conceptuel de données est présenté ci-dessous (Figure 6).

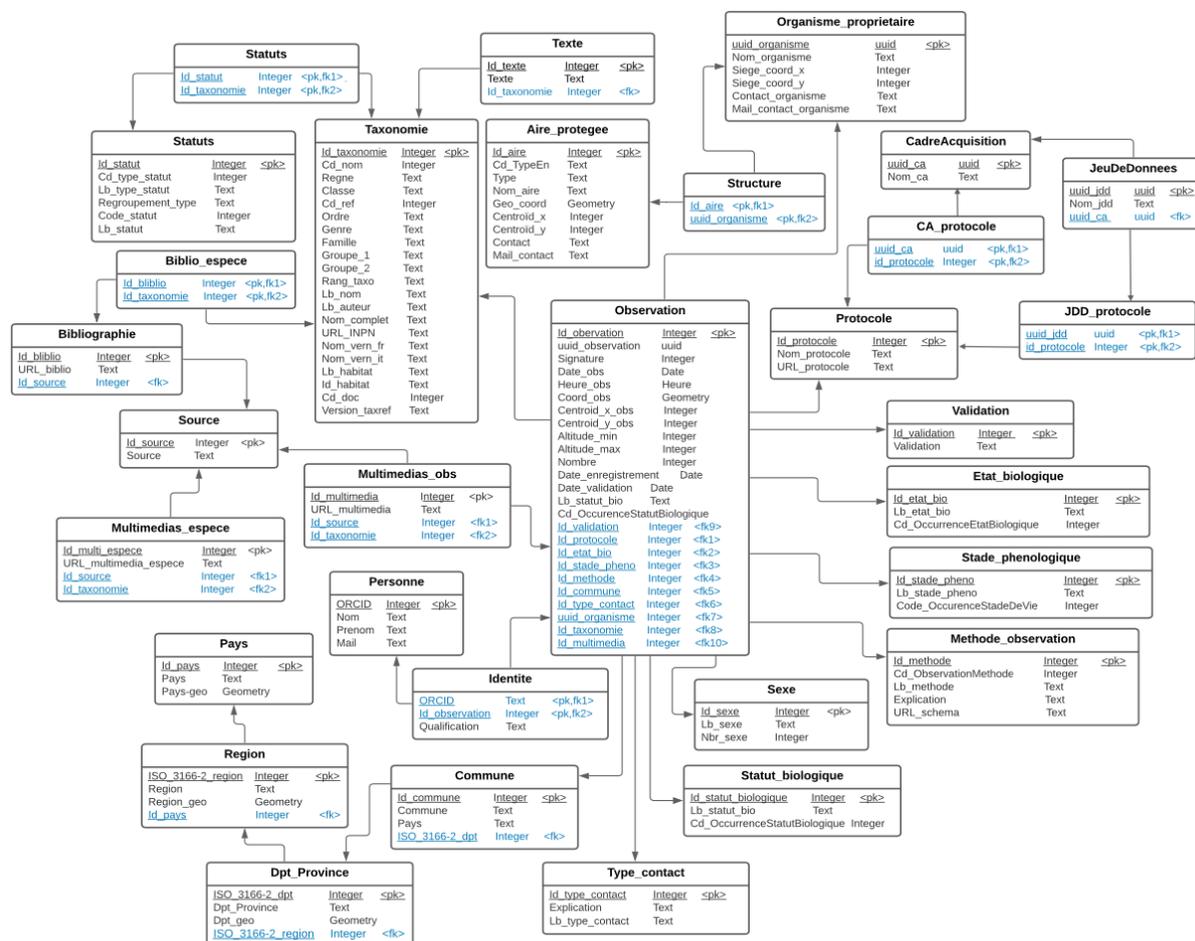


Figure 6 : Modèle conceptuel de données

À la vue de ce modèle de données et de ses nombreuses similitudes avec la base de données GeoNature, il a été décidé de ne pas concevoir une nouvelle base de données mais d'utiliser une base GeoNature (développée par le Parc National des Écrins) pour intégrer l'ensemble des données de la zone ALCOTRA. D'autant plus, que la base GeoNature a été optimisée en termes de performance pour certaines informations comme les informations spatiales par exemple. Dans ce modèle conceptuel de données, il y a une table par niveau géographique dans lesquelles les champs sont répétitifs c'est-à-dire qu'il y a une table commune, une table région, etc.. Tandis que dans la base GeoNature toutes ces informations sont regroupées dans une seule table (appelée ref_geo pour référentiel géographique) dans laquelle se trouve un champ supplémentaire: id_type qui permet d'identifier le niveau géographique (commune, région, pays...) de l'information, en récupérant le libellé du type dans une table (bib_areas_types) de correspondance de l'identifiant et du libellé entre autres informations. De plus, cette base est déjà construite et l'outil GeoNature-atlas y est facilement rattaché, ce qui représente un gain de temps considérable lors d'un stage de trois mois.

- Qu'est que la base GeoNature ?

GeoNature utilise un SGBD PostgreSQL et sa cartouche spatial PostGIS. La base GeoNature est constituée de plusieurs schémas regroupant des problématiques métiers, seul les schémas gn_imports, gn_synthese, gn_meta, taxonomie et ref_nomenclature sont utilisés dans ce projet. Le



schéma gn_synthèse est le schéma le plus important, c'est dans celui-ci qu'est la table synthèse contenant l'ensemble des informations d'observations qui sont valorisées dans l'outil. Le schéma taxonomie est le second schéma important, c'est ce schéma qui contient l'ensemble des informations de TAXREF permettant de relier les différentes observations à leur taxon et leurs informations respectives (cd_nom, nom scientifique, noms vernaculaires français, nom vernaculaire anglais...).

Le schéma gn_import a été utilisé pour importer les données des fichiers csv dans la base GeoNature avant leur transfert dans la synthèse grâce à des importations de données par mapping, dont la méthodologie est expliquée à la partie 4.2. Le schéma ref_nomenclature contient l'ensemble des standards de nomenclature de l'INPN utilisés en biodiversité et permet de faire le lien entre les codes nomenclatures des standards de l'INPN et les identifiants de la base GeoNature pour ces mêmes standards.

- **La problématique des observateurs**

Cependant, quelques différences existent entre le modèle conçu et la base GeoNature. Par exemple dans le modèle de données conçu à partir de l'étude des jeux de données, il existe une table « Personne » qui n'existe pas dans la base GeoNature. Cette table pourrait être une possible amélioration du schéma de la base GeoNature.

En effet, cette table permet d'avoir une liste des personnes liées aux données de biodiversité, que ce soient des observateurs comme des validateurs. Cette table contient pour chaque personne enregistrée dans la base leur nom, prénom, et mail. De plus, elle permet de les identifier grâce à un identifiant unique, l'ORCID qui permettrait non seulement de les identifier au sein d'un organisme mais également au sein d'autres organismes nationaux et même au niveau international puisque l'ORCID est un code alphanumérique ouvert pour les chercheurs et les contributeurs peu importe leur nationalité. Actuellement, il est impossible avec la base GeoNature de lier un observateur et/ou validateur aux observations puisqu'ils ne sont identifiés que par leur nom, prénom et ne sont pas enregistrés dans une table avec un identifiant. Ils sont seulement enregistrés au format texte dans un champs à remplir pour chaque observation dans la synthèse et ainsi soumis aux erreurs d'orthographe de la personne saisissant les données et rend très difficile les traitements des données par rapports aux observateurs.

- **Lien organismes/observations**

Une autre différence, importante étant donné les fonctionnalités ajoutées à GeoNature-atlas pour le projet, est le lien entre les observations et les organismes. En effet, les développements prévoient de promouvoir les organismes/structures fournissant les données. Pour cela, le lien doit être fait entre les observations et les organismes. Selon le modèle de données présenté ci-dessus, les organismes sont reliés directement aux observations. En revanche, dans GeoNature le lien n'est pas direct, il faut passer par un intermédiaire : le jeu de données. Cela est dû au fait que le lien organismes/observations est en réalité complexe, puisqu'une observation peut être liée à plusieurs organismes avec des rôles différents dans l'acquisition de la donnée. Par conséquent, l'établissement de ce lien nécessite la jointure de trois tables et l'ajout d'une quatrième si l'on veut les informations taxonomiques liées à l'observation contenues dans TAXREF, qui sera expliqué plus loin dans le rapport.

3.4) Synthèse de l'architecture choisie

Pour résumer, le fonctionnement est le suivant :

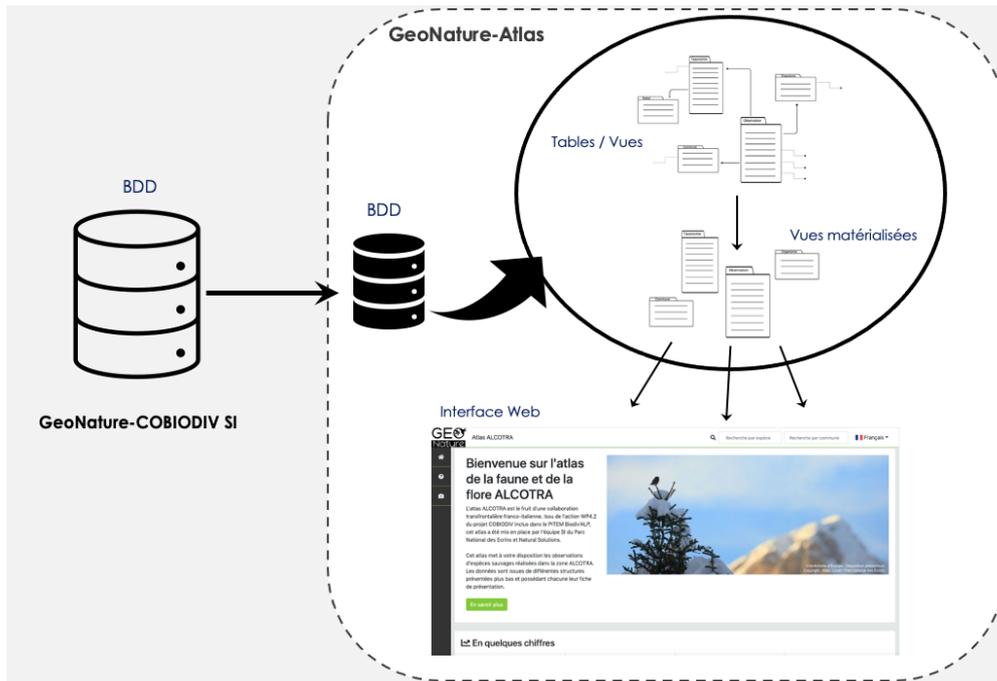


Figure 7: Schéma de l'outil de valorisation des données de biodiversité de la zone ALCOTRA

Les données de biodiversités françaises et italiennes sont intégrées dans une base GeoNature. La base GeoNature est reliée à l'outil GeoNature-atlas qui possède sa propre base de données via un foreign-data-wrapper. La base de données de l'atlas est composée de 6 schémas et d'une multitude de vues matérialisées contenant le résultat de requêtes « appliquées » sur les tables de la base de l'atlas et dont le résultat est interrogé par le backend de l'atlas grâce au toolkit SQLAlchemy.

4. Intégration des données

4.1) Intégration des données d'observations

Les données de biodiversité de la zone ALCOTRA sont acquises par plusieurs organismes, l'outil doit donc intégrer plusieurs jeux de données de structures et formats différents.

Il a été fait le choix d'utiliser une base GeoNature qui permet d'intégrer et gérer l'ensemble des données nécessaires aux différentes fonctionnalités de l'outil décrites dans l'analyse fonctionnelle. Plusieurs types d'import ont été réalisés, des imports manuels ou des imports via l'interface de GeoNature.

- **Module d'import de GeoNature**

L'outil GeoNature possède un module d'import (https://github.com/PnX-SI/gn_module_import) qui permet d'intégrer des données contenues dans un fichier csv ou GeoJSON dans une base de données GeoNature. Une fois le fichier chargé, le mapping se fait à travers l'interface, pour chaque champ de la synthèse une liste déroulante avec toutes les entêtes de colonnes est proposée. Pour importer le fichier exporté de GeoNature, il existe un mapping préenregistré qu'il suffit de sélectionner. Une fois le mapping validé, l'importation se lance et indique les données qui n'ont pas pu être importé et les erreurs rencontrées.



Le module d'import a été utilisé pour intégrer les données de la structure Flavia APE (qui étudie les lépidoptères) puisque la taille du jeu de données permettait de le faire (environ 6 000 données).

- **GN2PG (GeoNature to Postgre)**

GN2PG est un outil qui a été développé pour faciliter l'interopérabilité des données entre les différentes structures actrices de la biodiversité. Il permet d'échanger facilement grâce à une API des données d'une base GeoNature à une autre base GeoNature ou tout simplement à une base PostgreSQL. Le fonctionnement est le suivant : le GeoNature source met à disposition ces données via le module d'export. GN2PG interroge le module d'export (l'API) et récupère les données pour les intégrer dans la base GeoNature puits. Cet outil crée un schéma avec plusieurs tables et stocke dans un champ de l'une d'entre elles les données au format JSON et insère les cadres d'acquisition et les jeux de données dans le schéma des métadonnées. Ensuite les données en JSON sont fractionnées et intégrées dans les tables de GeoNature dont la synthèse. Cet outil n'est pas automatiquement installé lors de l'installation de GeoNature, il faut donc dans un premier temps l'installer puis le configurer (<https://github.com/lpoaura/GN2PG>). C'est dans le fichier de configuration que les informations concernant l'API de l'organisme « source » sont renseignées. Cet outil devait être utilisé pour intégrer les données du PNE et du Parc du Mercantour mais un problème au niveau de l'API du module d'export de GeoNature a été mis en évidence, rendant cet outil inutilisable temporairement et imposant la réalisation d'imports manuels.

- **Import manuel (Mapping)**

L'import manuel des données consiste en l'intégration des données sous format xls, csv, etc... dans la base de données GeoNature à travers différents script SQL qui importent, transforment, et font correspondre les champs de la synthèse avec les informations contenues dans les fichiers. Malgré l'aspect chronophage du Mapping, cette solution a été la plus utilisée pour importer les données du fait des besoins de transformations du format des données et de la taille de ces jeux de données, trop lourd pour le module d'import par exemple.

Démarche générale du mapping :

1. Conversion des fichiers XLSX en CSV
2. Connection à la base en superutilisateur (nécessaire pour utiliser la commande copy)
3. Utilisation d'une fonction interne de GeoNature « load_csv_file » : cette fonction crée une table vide dans le schéma d'import avec autant de champs qu'il y a de colonnes dans le fichier csv et les nomme comme les entêtes de colonnes du fichier, ensuite elle utilise la commande « copy » pour importer les données du fichier dans la table.
4. Création des métadonnées associées au jeu de données : requête pour la table t_source, utilisation de l'interface de GeoNature pour la création des cadres d'acquisition et des jeux de données (4 jeux de données italiens venant du parc national des alpes maritimes, un de Flavia APE, un du PNE, un du Mercantour, un du CBNMed, et un du CBNA, un cadre d'acquisition a été créé par organisme)

```
1  ---- CREATION SOURCE ----
2  INSERT INTO gn_synthese.t_sources(name_source, desc_source)
3  VALUES
4  ('FAUNE-ITALIE', 'Données italiennes de faune du parc des alpes maritimes ')
```

Figure 8: Script SQL de création de la source du jeu de données à importer

5. Import dans la synthèse de GeoNature : script SQL transformant et transférant les données de la table du schéma d'import dans la synthèse de la base.

```

6  ---- IMPORT DANS LA SYNTHÈSE ----
7  INSERT INTO gn_synthese.synthese(
8  unique_id_sinp,
9  id_source,
10 id_dataset,
11 id_nomenclature_valid_status,
12 id_nomenclature_observation_status,
13 id_nomenclature_source_status,
14 id_nomenclature_biogeo_status,
15 count_min,
16 count_max,
17 cd_nom,
18 cd_hab,
19 nom_cite,
20 meta_v_taxref,
21 altitude_min,
22 altitude_max,
23 the_geom_4326,
24 the_geom_point,
25 the_geom_local,
26 precision,
27 date_min,
28 date_max,
29 validator,
30 observers,
31 meta_validation_date,
32 meta_create_date,
33 meta_update_date,
34 last_action)
35 )
36 SELECT
37     uuid_generate_v4(),
38     3 AS id_source,
39     7 AS id_dataset,
40     315 AS id_nomenclature_valid_status, --Certain
41     84 AS id_nomenclature_observation_status, -- Présent
42     72 AS id_nomenclature_source_status, --Ne sait pas
43     176 AS id_nomenclature_biogeo_status, --Non renseigné
44     numero_individui_esatto::integer, --count_min
45     numero_individui_esatto::integer, --count_max
46     (select cd_nom from taxonomie.taxref t where t.nom_complet=nome_scientifico and t.cd_nom =t.cd_ref ::integer) AS cd_nom,
47     (select id_habitat from taxonomie.taxref t where t.nom_complet=nome_scientifico and t.cd_nom =t.cd_ref ::integer) AS cd_hab,
48     descrizione_specie AS nom_cite,
49     'Taxref V13.0' AS meta_v_taxref,
50     quota_da::integer, -- Altitude min
51     quota_a::integer, -- Altitude max
52     ST_Transform(ST_SetSRID(ST_MakePoint("utm_x"::numeric,"utm_y"::numeric),23032),4326) AS the_geom_4326, -- WGS84
53     ST_Centroid(ST_Transform(ST_SetSRID(ST_MakePoint("utm_x"::numeric, "utm_y"::numeric),23032),4326)) AS the_geom_point, -- WGS84
54     ST_Transform(ST_SetSRID(ST_MakePoint("utm_x"::numeric, "utm_y"::numeric),23032),2154) AS the_geom_local, -- Lambert 93
55     REPLACE(impresione, ' m', '')::integer AS precision, --Pb si km
56     CONCAT(data_osservazione, ' ',ora_osservazione_da)::timestamp AS date_min,
57     CONCAT(data_osservazione, ' ',ora_osservazione_da)::timestamp AS date_max,
58     validator,
59     altri_osservatori::text, --observateur
60     data_validazione::timestamp,
61     data_registrazione::timestamp,
62     NOW() AS meta_update_date,
63     'I' AS last_action -- code de la dernière action effectuée: Valeurs possibles 'I': insert, 'U': update
64 FROM gn_imports.importavesbis
65 ORDER BY CONCAT(data_osservazione, ' ',ora_osservazione_da)
66 ;

```

Figure 9: Exemple de script SQL faisant le mapping entre le jeu de données de la faune italienne et la synthèse

```

147     case
148     when geom<>'N' then (ST_Transform(ST_GeomFromText(substring(geom,11,length(geom)),2154),4326))
149     else (null)
150     end as the_geom_4326,

```

Figure 10: Exemple d'utilisation de conditions pour importer des données dans la synthèse (ici sur les référentiels spatiaux)

Le mapping a notamment été utilisé pour importer les données italiennes puisque qu'elles nécessitaient plusieurs modifications avant d'être intégrées dans la synthèse de GeoNature. Un premier traitement des données sur Excel a été effectué afin de rendre les données intégrables dans une base de données, c'est-à-dire remplacement des espaces dans les entêtes de colonnes par des



«_», suppressions des espaces après le nom de genre (pour le fichier BD_Faune_NO_aves) qui empêchait la correspondance avec TAXREF.

Chaque jeu de données italien ayant un format différent, un script SQL a été écrit pour chaque jeu de données qui fait le mapping entre les champs italiens et les champs de la synthèse. Ces scripts permettent également les nombreuses transformations et correspondances entre les données contenues dans les jeux de données italiens et celles demandées dans la synthèse et leur format spécifique. Un exemple de script pour un jeu de donnée italien (Données de la faune du parc des Alpes Maritimes extrait de leur outil AVES) est en figure 9 ci-dessus. En effet, les Italiens n'utilisent pas le même référentiel spatial que les Français ni le système de GPS WGS-84. L'importation des coordonnées de l'observation passe donc par la transformation de ces données dans le référentiel WGS-84 grâce à l'utilisation de différentes fonctions SQL. D'autre part, dans GeoNature, les informations temporelles sont dans un format spécifique, il est donc nécessaire de concaténer plusieurs champs italiens pour intégrer les informations dans la synthèse. La dernière action majeure de ces scripts est la correspondance des noms d'espèces ou de genre avec les espèces /genres inventoriés dans TAXREF afin de retrouver le cd_nom, identifiant de l'espèce ou du genre, sur lequel repose l'ensemble des commandes utilisant le taxon de l'observation.

Pour les données du CBNA, il existait déjà des scripts, écrits par Jean-Pascal Milcent, permettant de transférer les données d'un fichier csv, dans le format d'échange du CBN, dans la synthèse GeoNature. L'intégration de données du CBNA a donc été effectuée grâce à ces scripts. Certaines données n'ont pas pu être intégrées du fait de valeurs erronées pour l'altitude de l'observation ou bien parce que le cd_nom n'existe pas dans la version TAXREF 14 (version de la base GeoNature du projet COBIODIV SI). À la suite de l'import des données dans la synthèse de la base, les données ne concernant pas la zone ALCOTRA ont été supprimé grâce à une intersection avec la géométrie de la zone ALCOTRA côté français soit les départements 04, 05, 06, 73, 74 créée à partir d'un fichier Shape des départements français sur QGIS.

```
----- Suppression des données du cbna hors alcotra france -----  
DELETE FROM gn_synthese.synthese s  
WHERE NOT st_intersects(s.the_geom_local, (SELECT zaf.geom FROM gn_imports.zone_alcotra_france zaf))  
AND id_dataset NOT IN (4,5,6,7,8,9,10,11,187)
```

Figure 11: Script sql de suppression des données hors ALCOTRA France

Pour le CBNMed, les données ont été importées également par mapping. Le fichier csv est dans le format d'échange du CBN ainsi il contient tous les codes nomenclatures des standards INPN des différentes informations relatives aux observations. Le script SQL de mapping consiste donc essentiellement à relier les codes des nomenclatures aux identifiants de ces nomenclatures dans la base GeoNature grâce au backoffice de GeoNature. Le CBNMed n'ayant pas la même version de TAXREF que celle importée dans la base du projet COBIODIV SI, un taxon n'ayant pas un cd_nom reconnu a été supprimé des données (47 observations) afin de pouvoir intégrer les données dans la synthèse qui nécessite d'avoir un cd_nom contenu dans la table Taxref car c'est une clé étrangère. De plus, en amont de l'intégration le fichier csv a été transformé sur R pour modifier le séparateur afin d'utiliser la fonction « load_csv_file » qui ne comprend pas un fichier un csv avec un séparateur tab, et qu'il n'était pas possible d'utiliser la fonction copy car le format du fichier n'était pas non plus reconnu par la commande.

Pour le PNE, l'outil GN2PG ne pouvant être utilisé, les données ont été importées dans la base dans un premier temps, dans le schéma gn_exports par un script SQL fourni par le PNE. Les données ont ensuite été transférées dans la synthèse de la base par un mapping, en réutilisant le script utilisé pour les données du CBNMed mais en remplaçant les identifiants de nomenclatures par leur libellé,

étant donné que le format des nomenclatures des données du CBNMed étaient sous forme d'identifiants tandis que celles du PNE sous forme de libellés.

Au final ce sont 7 mappings qui ont été réalisés permettant l'import de plus de 3 millions de données concernant plus de 16000 espèces de faune et de flore.

4.2) Intégration de la géométrie de la zone d'étude

- Conception de l'emprise de la zone ALCOTRA

Dans l'outil de valorisation de la biodiversité plusieurs informations sont présentées sous forme de carte. Pour ces cartes, il est donc nécessaire d'un point de vue ergonomique et pratique d'avoir l'emprise de la zone d'étude. Cette emprise a été créée sur QGIS à partir de deux fichiers shapefile. L'un contenant les délimitations départementales de la France métropolitaine, l'autre contenant les délimitations provinciales de l'Italie. Ci-dessous, une visualisation des géométries de ces territoires sur QGIS.

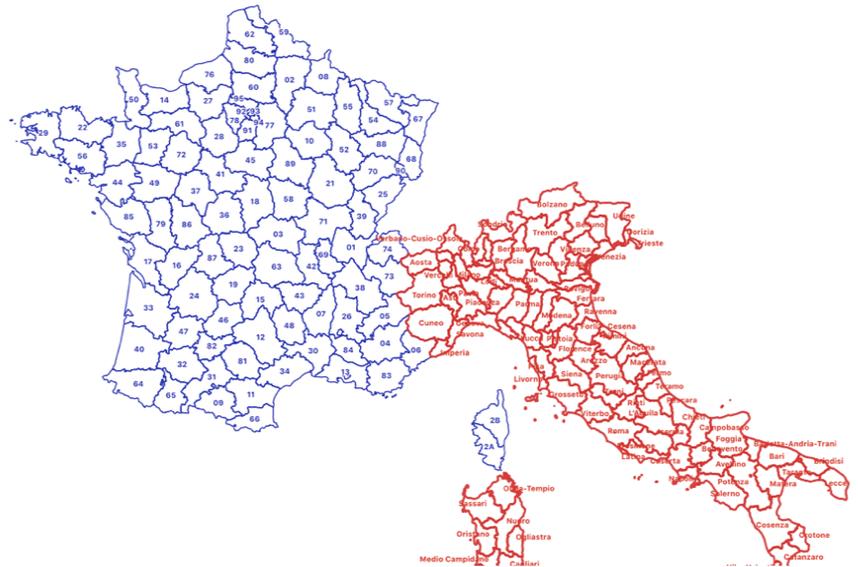


Figure 12: Géométries des territoires français et italiens

Pour chaque pays, une nouvelle couche a été créée à partir des départements ou des provinces de la zone ALCOTRA (détaillés à la page 10) sélectionnés sur les couches nationales. Sur chacune des couches, les entités ont été fusionnées afin d'obtenir une couche de la zone ALCOTRA française et une couche de la zone ALCOTRA italienne (Figure 13).

Ces deux couches ont ensuite été fusionnées et un remaniement manuel a été effectué afin de palier le problème de correspondance des frontières entre les deux pays (les deux sommets de la frontière se rejoignant, il n'y avait pas de problème de distance mais uniquement des vides à l'intérieur de l'emprise géographique), donnant la couche finale contenant l'emprise géographique de la zone ALCOTRA.

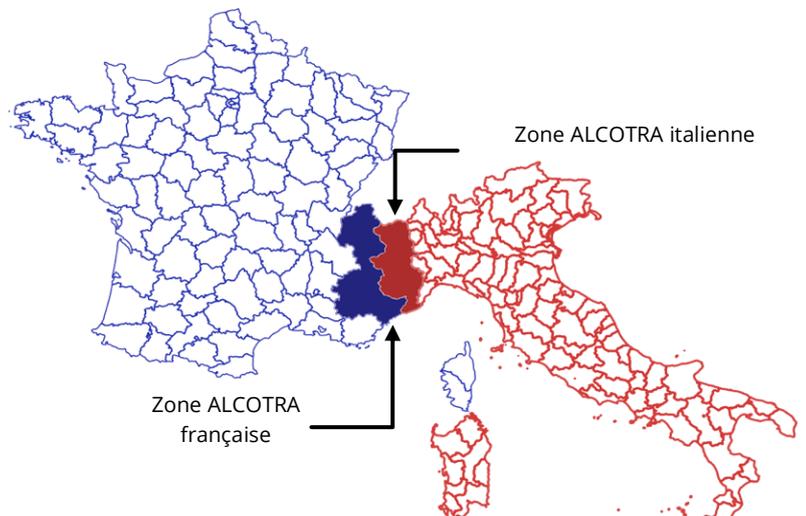


Figure 13: Visualisation des géométries de la zone ALCOTRA française et de la zone ALCOTRA italienne depuis QGIS



- Import de l'emprise dans la base de données

Une fois l'emprise géographique de la zone ALCOTRA créée, celle-ci a été importée dans la base de données depuis QGIS dans une table d'import et ensuite transférée dans la table l_area du schéma ref_geo après transformation de son système de projection. Ci-contre, la visualisation de la géométrie de la zone ALCOTRA, entrée dans la base de données, depuis DBever (outil permettant l'administration et le requêtage de base de données).

L'intégration des données a permis de soulever les problématiques d'interopérabilités d'une part mais aussi de concrètement tester l'atlas grâce à un contenu important permettant d'en faire un véritable POC (preuve de concept).

Puisqu'il me restait quelques jours avant la fin de mon stage et ayant une forte appétence pour les développements techniques je me suis mise en appuis sur les développements de fonctionnalités de l'atlas avec Corentin.

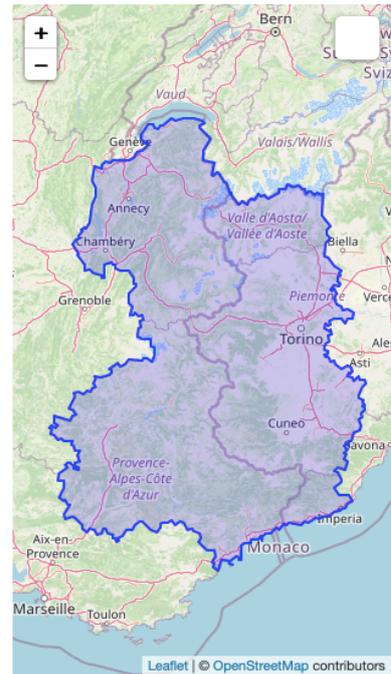


Figure 14: Géométrie de la zone ALCOTRA

5. Développements techniques

5.1) Intégration de la dimension « organismes » dans l'atlas

L'outil de valorisation développé dans le cadre de ce projet se base sur GeoNature-atlas. Afin de correspondre aux attentes du projet, plusieurs fonctionnalités ont été développées et ajoutées à cet outil. L'un des ajouts est la dimension « organismes » qui vise à valoriser le travail des organismes collectant des données de biodiversité. Cette dimension a été développée en collaboration avec Corentin Lange, qui s'est principalement occupé du front-end de la fiche organisme (Figure 15) mais aussi de plusieurs fonctions back-end, et de l'ajout de l'onglet organismes sur la fiche espèce. Une fiche organisme permet de donner les éléments clé pour une structure pourvoyeuse de donnée.

Parc National des Écrins
Région/département rattaché : Domaine de Charance 05000
Téléphone : 0492402010

Données de l'organisme

722490 Observations	23.97% Participation à l'atlas	7843 Espèces
------------------------	-----------------------------------	-----------------

Espèces les plus observées

Bouquetin des Alpes Capra ibex 30368 Observations 4.20% des observations de l'organisme	Fiche espèce
Chamois Rupicapra rupicapra 17001 Observations 2.35% des observations de l'organisme	Fiche espèce
Pinson des arbres Fringilla coelebs 12623 Observations 1.75% des observations de l'organisme	Fiche espèce

Figure 15: Exemple de fiche organisme (Parc National des Écrins)

En amont des développements backend et front-end de l'atlas, la dimension organismes a été ajoutée au niveau de la base de données de l'atlas. En effet, cette dimension n'étant pas prise en compte dans l'outil d'origine, il n'existait aucune table ou vue matérialisée dans la base de l'atlas contenant les données nécessaires. La première étape fut donc d'écrire des scripts SQL afin de créer un schéma « utilisateurs » dans la base de l'atlas, d'y importer les tables avec les données de la base GeoNature du projet COBIODIV SI contenant les données sur les organismes, et de créer une vue matérialisée associant chaque organisme (avec toutes ses informations propres : nom, localisation, les url du logo et du site de l'organisme...) à ses observations. Ce lien organismes-observations n'est pas direct dans une base de données GeoNature, cette vue matérialisée se fait donc par l'intermédiaire des données sur le jeu de données auquel est associé chaque observation, et auquel est associé un ou plusieurs organismes selon leur rôle dans l'acquisition de la donnée. Pour cela, il a été importé dans la base de données de l'atlas les tables `cor_data_actor`, reliant les jeux de données à leurs organismes, et `bib_organisme`, contenant l'ensemble des informations personnelles des organismes. Ci-dessous (Figure16) le script SQL utilisé pour ces étapes.

```

1  ----- CRÉATION DU SCHEMA -----
2  CREATE SCHEMA utilisateurs;
3
4  -- TABLE bib_organisme
5  CREATE FOREIGN TABLE utilisateurs.bib_organisms (
6      id_organisme int4 OPTIONS(column_name 'id_organisme') NOT NULL,
7      uuid_organisme uuid OPTIONS(column_name 'uuid_organisme') NULL,
8      nom_organisme varchar(500) OPTIONS(column_name 'nom_organisme') NULL,
9      adresse_organisme varchar(128) OPTIONS(column_name 'adresse_organisme') NULL,
10     cp_organisme varchar(5) OPTIONS(column_name 'cp_organisme') NULL,
11     ville_organisme varchar(50) OPTIONS(column_name 'ville_organisme') NULL,
12     tel_organisme varchar(14) OPTIONS(column_name 'tel_organisme') NULL,
13     fax_organisme varchar(14) OPTIONS(column_name 'fax_organisme') NULL,
14     email_organisme varchar(100) OPTIONS(column_name 'email_organisme') NULL,
15     url_organisme varchar(255) OPTIONS(column_name 'url_organisme') NULL,
16     url_logo varchar(255) OPTIONS(column_name 'url_logo') NULL,
17     id_parent int4 OPTIONS(column_name 'id_parent') NULL
18 )
19 SERVER geonaturedbserver
20 OPTIONS (schema_name 'utilisateurs', table_name 'bib_organismes');
21
22 -- TABLE cor_data_actor
23 CREATE FOREIGN TABLE utilisateurs.cor_dataset_actor (
24     id_cda int4 OPTIONS(column_name 'id_cda') NOT NULL,
25     id_dataset int4 OPTIONS(column_name 'id_dataset') NULL,
26     id_role int4 OPTIONS(column_name 'id_role') NULL,
27     id_organisme int4 OPTIONS(column_name 'id_organisme') NULL,
28     id_nomenclature_actor_role int4 OPTIONS(column_name 'id_nomenclature_actor_role') NULL
29 )
30 SERVER geonaturedbserver
31 OPTIONS (schema_name 'gn_meta', table_name 'cor_dataset_actor');
32
33 CREATE VIEW utilisateurs.reduced_cor_dataset_actor
34 AS SELECT DISTINCT id_dataset, id_organisme
35     FROM utilisateurs.cor_dataset_actor;
36
37
38 --CRÉATION VUE MATÉRIALISÉE
39 CREATE MATERIALIZED VIEW atlas.vm_organisms
40 AS SELECT DISTINCT cd_ref, count(*) as nb_observations, bo.id_organisme , nom_organisme , adresse_organisme , cp_organisme ,
41     ville_organisme , tel_organisme , email_organisme , url_organisme ,url_logo
42     FROM utilisateurs.bib_organisms bo
43     JOIN utilisateurs.reduced_cor_dataset_actor cda ON bo.id_organisme =cda.id_organisme
44     JOIN synthese.synthese s ON s.id_dataset =cda.id_dataset
45     JOIN taxonomie.taxref t on s.cd_nom=t.cd_nom
46     group by t.cd_ref, bo.id_organisme, bo.nom_organisme, bo.adresse_organisme, bo.cp_organisme, bo.ville_organisme, bo.tel_organisme,
47     bo.email_organisme, bo.url_organisme, bo.url_logo
48     with data;

```

Figure 16: Script SQL des modifications de la base de données de l'atlas pour intégrer la dimension "organismes"

Plusieurs développements de l'atlas ont été effectués sur cette dimension. La première étape fut de créer en backend une classe python « copiant » la vue matérialisée précédemment créée dans la base de l'atlas.

```
11 class VmOrganisms(Base):
12     __table__=Table(
13         'vm_organisms',
14         metadata,
15         Column('nb_observations', Integer),
16         Column('id_organism', Integer, primary_key=True, unique =True),
17         Column('nom_organism',String(500)),
18         Column('adresse_organism', String(128)),
19         Column('cp_organism', String(5)),
20         Column('ville_organism', String(100)),
21         Column('tel_organism',String(14)),
22         Column('email_organism', String(100)),
23         Column('url_organism', String(255)),
24         Column('url_logo', String(255)),
25         Column('cd_ref',Integer),
26         schema="atlas",
27         autoload=True,
28         autoload_with=engine
29     )
30
```

Figure 17: Script de création de la classe organismes

À la suite de cela, un nouveau fichier python a été créé, intitulé `vmOrganismesRepository`, dans lequel se trouve quatre fonctions python permettant chacune de récupérer les informations des organismes et des observations nécessaires aux fonctionnalités de cette dimension. La première fonction renvoie, à partir d'un identifiant organisme, le nombre de taxons observés et le nombre total d'observations par organisme ainsi que les informations de l'organisme (nom, courriel, logo...) à destination de la fiche organisme. La troisième fonction renvoie, pour un taxon, la liste des organismes ayant contribué à des observations avec pour chaque organisme le nombre d'observations, le nom de l'organisme et son logo, informations affichées sur la fiche taxonomique. La deuxième et la quatrième fonction, écrites par Corentin, permettent respectivement de renvoyer les trois taxons les plus observés par un organisme donné, et de renvoyer pour un organisme donné le nombre d'observation par groupe taxonomique afin de créer le graphique de répartition des observations par groupe taxonomique sur la fiche organisme.

Les différents scripts de front-end des fonctionnalités de cette dimension ont été développés par Corentin ainsi que l'ajout aux routes de l'atlas. D'autre part, le PNR du marais Poitevin a développé un bandeau partenaires (<https://github.com/PnX-SI/GeoNature-atlas/pull/245>) qui fait défiler l'ensemble des logos des organismes partenaires sur la page d'accueil. Cette fonctionnalité devait être ajoutée au prototype développé mais par manque de temps cela n'a pas été fait.

5.2) Fonctionnalité : « Nouvelles espèces observées »

La fonctionnalité « nouvelles espèces observées » est un nouveau bloc intégré sur la page d'accueil de l'atlas, où il y est présenté les dernières espèces qui ont été découvertes sur la zone ALCOTRA (bien évidemment ces espèces ont pu être découvertes précédemment hors de la zone ALCOTRA).

Dans un premier temps, une fonction python a été développée, en utilisant SQLAlchemy, permettant de récupérer les informations nécessaires à la fonctionnalité grâce à une requête SQL. Cette requête permet de ressortir le nom scientifique, le nom français, la date de première observation ainsi qu'un le chemin d'accès à une photo des cinq dernières espèces découvertes sur la zone ALCOTRA. L'ensemble de ces informations sont affichées et utilisées par le front-end.

```

1  {% block newSpecies %}
2  <!-- Nouvelles espèces-->
3  <div class="card mt-4 bg-transparent border-0">
4  <div class="card-body">
5  <div class="row">
6  <div class="col-lg-12">
7  <h3 class="card-title title-spaced">
8  <span class="fa fa-bell icon-color-title"></span> &nbsp;{{ _('new.species.observed') }}
9  <br>
10 </div>
11 </div>
12 <div class="row row-medias">
13 <div class="lastDiscoveries">
14 <center>
15 <div class="card-deck">
16   {% for species in lastDiscoveries %}
17   <div class="card border-secondary">
18     {% if species.id_media == None %}
19       {% set img_path='atlas/static/images/Photo_non_disponible.png' %}
20     {% else %}
21       {% if configuration.REDIMENSIONNEMENT_IMAGE %}
22         {% set img_path = configuration.TAXHUB_URL+'/api/tmedias/thumbnail/'+species.id_media|string+'?h=500' %}
23       {% else %}
24         {% set img_path = species.media_path %}
25       {% endif %}
26     {% endif %}
27 <a href="{{ url_for('main.ficheEspece',cd_ref=species.cd_ref) }}">
28 
29 </a>
30 <div class="card-body">
31 <p> {% if species.nom_vern != None %}
32 <b> {{ species.nom_vern }}</b> {% endif %}
33 <br/><i>{{ species.lb_nom }}</i></p>
34 <a href="{{ url_for('main.ficheEspece',cd_ref=species.cd_ref) }}"
35 class="btn btn-default rounded-pill" role="button">{{ _('species.sheet') }}</a>
36 </div>
37 <div class="card-footer border-secondary text-muted">
38   {{species.date}}
39 </div>
40 </div>
41   {% endfor %}
42 </div>
43 </center>
44 </div>
45 </div>
46 </div>
47 </div>
48 {% endblock %}

```

Figure 18: Script HTML de la fonctionnalité "nouvelles espèces observées"

Côté front-end, il a été utilisé le Framework CSS bootstrap 4. La fonctionnalité est sous forme de cartes, utilisant les card de bootstrap. Un nouveau fichier nommé « newSpecies.html » a été créé et est appelé dans le fichier main.html de la page d'accueil de l'atlas. Cette fonctionnalité forme un nouveau bloc de la page d'accueil de l'atlas. Pour finir, cette fonction a été ajoutée à la route de l'atlas de la page d'accueil. Le résultat, en front-end, est visible à la Figure 19.



Figure 19: Front-end de la fonctionnalité "nouvelles espèces observées"



L'ensemble des développements techniques effectués durant le stage par Corentin ou moi-même sont disponibles sur le GitHub de GeoNature-atlas dans la branche « dev-cobiodiv », à l'adresse suivante : <https://github.com/PnX-SI/GeoNature-atlas/tree/dev-cobiodiv> .

6. Bilan et perspectives

6.1) Bilan de l'intégration des données

La valorisation des données de biodiversité dans le cadre du projet COBIODIV SI s'est basée sur plusieurs jeux de données venant de partenaires français et italiens. Ces données ont pu être en grande majorité intégrées dans une base de données (GeoNature) dédiée au projet. En tout, 3 014 129 données de biodiversité ont pu être intégrées et valorisées dans le cadre de ce projet. Le détail des données reçues des organismes et intégrées dans la base de données du projet COBIODIV SI est présenté dans le tableau ci-dessous.

Organisme	FRANCE					ITALIE
	PNE	PNM	Flavia APE	CBNA	CBNMed	Parco alpi marittime
Données envoyées	722 368	?	6 331	≈ 5 millions	528 648	116
Données intégrées	722 368	0	6 324	1 756 728	528 611	95
Pourcentage d'intégration	100%	0%	99%	35%	99%	80%

Tableau 6: Récapitulatifs du nombre de données par organismes

Côté français, très peu de données n'ont pas été intégrées. La principale cause des données non intégrées est un problème de correspondance de taxon entre la dernière version de TAXREF (utilisé dans la base GeoNature du projet) et la version précédente encore utilisée par le CBNMed et le CBNA. Un taxon présent dans la version 13 n'est pas présent dans la version 14, ces données n'ont donc pas pu être intégrées dans la synthèse puisqu'une contrainte impose aux observations d'avoir un cd_nom (identifiant du taxon de l'observation) existant dans Taxref. D'autre part, le faible pourcentage d'intégration des données du CBNA s'explique également par le fait que toutes les données reçues se situaient hors zone ALCOTRA. Les données du Parc du Mercantour n'ont pas été intégrées, d'une part, du fait de l'impossibilité de l'utilisation de l'outil GN2PG, et d'autre part, parce que le fichier geopackage reçu ne pouvait être ouvert correctement sur QGIS.

Concernant les données italiennes, les taxons ne sont pas référencés par TAXREF dans les fichiers fournis. Ainsi pour les intégrer, il a été effectué une correspondance par égalité du nom scientifique d'espèce ou de genre avec respectivement les champs nom complet et libellé nom de TAXREF afin de retrouver l'identifiant cd_nom de l'espèce ou du genre. Par conséquent, certains noms scientifiques mal orthographié ou non connu sous ce nom en France n'ont pas trouvé de correspondance avec la base TAXREF et n'ont pas pu être intégrée à la synthèse de GeoNature, mais cela ne concerne qu'un petit nombre de données.

Ce tableau (Tableau 6) nous montre que des données structurées permettent une meilleure intégration des données au sein d'une base de données pour pouvoir être ensuite remobilisées facilement via différents outils technologiques.

Plusieurs méthodes d'intégration de données ont été utilisées. Le schéma ci-dessous récapitule les méthodes utilisées pour l'intégration des données reçues et montre les flux attendus (en pointillé) pour l'avenir du projet, c'est-à-dire en passant par les bases des SINP pour le côté français.

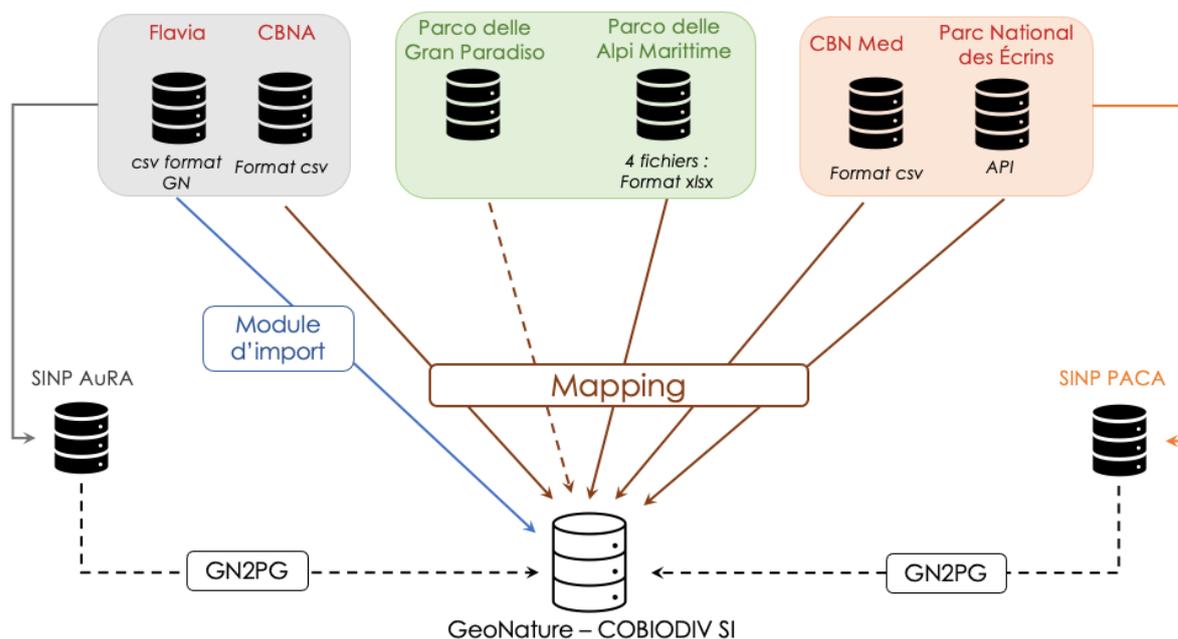


Figure 20.: Schéma des flux de données et des méthodes d'intégration

6.2) Bilan du développement de l'outil

Une fois les données intégrées dans la base GeoNature, elles ont pu être valorisées par l'outil GeoNature-atlas auquel plusieurs fonctionnalités ont été ajoutées. L'analyse fonctionnelle fait un listing de l'ensemble des fonctionnalités qui permettent de répondre aux attentes du projet. Cependant, étant donné le temps imparti pour développer l'ensemble des fonctionnalités décrites, il a été décidé de ne développer que certaines fonctionnalités, jugées les plus importantes. Les fonctionnalités sélectionnées concernent principalement le développement de la dimension « espèce », l'apport de retouches à la dimension « commune » et l'ajout de la dimension « organismes ».

Pour la dimension « espèce », j'ai travaillé sur deux fonctionnalités. La première est la fonctionnalité « nouvelles espèces observées » qui est terminée, fonctionnelle et apparaît sur la page d'accueil. La seconde fonctionnalité, très importante pour cette dimension, est la fonctionnalité d'affichage des statuts de protection des espèces à enjeux sur la fiche espèce, et la fiche commune. Cette fonctionnalité repose sur la base de connaissance statuts de l'INPN contenu dans TaxHub. Par conséquent, il a été décidé de ne pas créer de vue matérialisée dans la base de GeoNature-atlas pour les statuts mais de faire appel à l'API de TaxHub. TaxHub est une application web de gestion des taxons, utilisant le référentiel TAXREF, et permet de gérer les descriptions et les médias des espèces dans GeoNature-atlas. Dans un premier temps, nous avons, avec Adrien Pajot, réfléchi à la présentation de ses statuts sur la fiche espèce puisqu'ils sont nombreux et à différentes échelles géographiques. Après cela, nous avons ajouté, avec Théo Lechémiat, deux filtres aux routes de l'API de TaxHub. Cette fonctionnalité n'a malheureusement pas pu être développée en totalité dans le temps imparti, tous le côté front-end reste encore à développer.

En ce concerne la dimension « organismes », nous avons, en binôme avec Corentin, réussi à développer plusieurs fonctionnalités permettant de valoriser le travail des organismes collectant des



données de biodiversité. Une fiche organisme a été ajoutée à l'atlas ainsi que l'onglet organismes aux fiches espèces.

Le développement de l'outil était la mission principale de Corentin. Son rapport de stage rend compte des développements qu'il a effectués et explicite plus en détail l'outil GeoNature-atlas et son évolution dans le cadre de la mission COBIODIV SI.

6.3) Intégration d'un flux de données continu de sources agrégatives de données

Actuellement, les données intégrées dans la base de données du projet proviennent de divers organismes collectant des données de biodiversité. À terme l'objectif est d'intégrer un flux continu de données provenant de structures agrégatives des données de biodiversité afin de réduire le nombre de flux entrants. Cela permettrait, en effet, un gain de temps considérable au niveau de l'intégration des données puisque cela réduirait le nombre de structure et de formats des jeux de données à intégrer. En effet, dans l'état actuel du projet, du côté français, les données que nous avons intégrées proviennent de 5 structures différentes et quasiment autant de structures et de formats des données différents. Il serait donc plus favorable de baser l'atlas sur la valorisation des données de biodiversité provenant des SINP régionaux qui agrègent les données de tous les organismes publiques et limitant ainsi le nombre de flux français à deux. Concernant les données italiennes celles-ci proviennent d'un seul organisme et nous avons reçu 4 jeux de données de structures et de formats différents, quadruplant le travail pour un seul organisme.

Ce qui est à noter concernant la valorisation des données est que plus les données seront structurées de manière homogène, par l'utilisation de standards et de référentiels, au sein d'un pays ou bien même au sein d'un organisme, plus il sera facile de les valoriser. Tout comme le fait de posséder des structures agrégatives de données permettant de réduire le nombre de sources et peut-être aussi de rendre les données publiques plus facilement accessibles.

6.4) Extension du projet à l'intégralité des Alpes

Les alpes sont une chaîne de montagne européenne qui s'étendent sur plusieurs pays. Une perspective d'évolution de ce projet pourrait être d'intégrer l'ensemble des alpes ou du moins d'étendre le projet à la zone de la convention alpine dont les limites sont présentées sur la Figure 21.

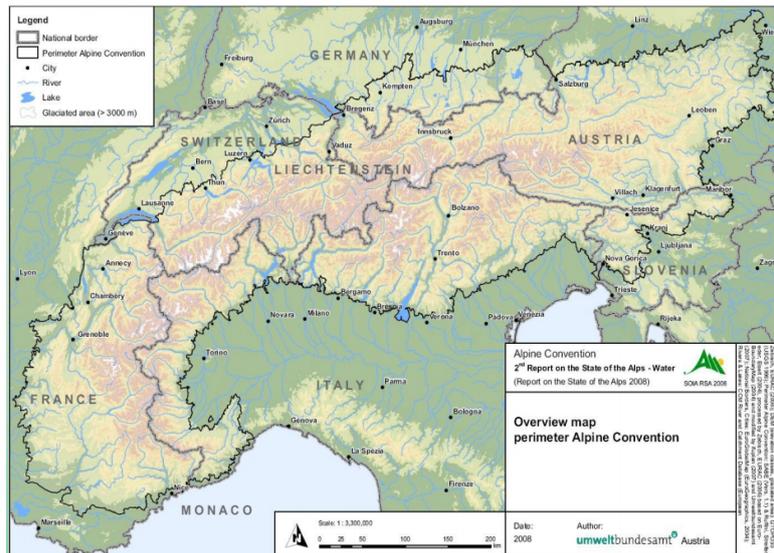


Figure 21: Carte de la délimitation de la convention alpine

Actuellement, l'atlas valorise uniquement les données de biodiversité de la zone ALCOTRA soit uniquement sur le territoire des Alpes françaises et italiennes. Il serait donc envisageable d'intégrer au projet la suisse par exemple. En effet, ce pays est frontalier aux deux pays de la zone ALCOTRA, et c'est également un pays dont les institutions scientifiques sont potentiellement capables de mener des projets internationaux de cette envergure.

En effet, pour une protection efficace des Alpes, il serait plus intéressant d'inclure dans le projet l'ensemble des pays du territoire alpin puisque la biodiversité ne s'arrête pas aux limites administratives. La biodiversité de la zone ALCOTRA est donc en continuité avec celle des autres pays du territoire alpin et dont tous sont confrontés aux mêmes problématiques de préservation et de protection de la biodiversité.

6.5) Inclusion de la thématique des habitats au projet

Depuis quelques décennies les scientifiques alertent sur les dangers qu'encourt la biodiversité. Le grand public et les dirigeants politiques et territoriaux ont pris conscience des enjeux environnementaux et prennent des mesures pour la protection de la biodiversité comme la création d'espaces protégés, la participation financière à des études et la mise en place d'organismes comme les SINP. Toutefois, les mesures visent essentiellement à la conservation des espèces alors que la protection de celles-ci ne peut se faire sans la protection de leurs habitats. Et malheureusement, contrairement aux données de biodiversité les données d'habitats se font plus rares tout comme leur valorisation et leur mise en parallèle avec les données de biodiversité.

Mais la prise de conscience de ce problème il y a quelques années par le MNHN a abouti à la création par l'UMS PatriNat, d'une base de données regroupant les typologies des habitats et des végétations utilisées en France (référentiel de typologie HABREF). Ce référentiel a pour but de répondre à plusieurs objectifs :



- « proposer un référentiel national unique réunissant les versions officielles de référence de toutes les typologies d'habitats et de végétations, terrestres et marines, concernant les territoires français de métropole et d'outre-mer.

- constituer une base de connaissance sur les unités de typologies, avec des informations sur leur statut de présence dans les territoires français, **les espèces associées (en lien avec le référentiel TAXREF)**...

- standardiser et référencer les informations, et assurer leur mise à jour régulière
- faciliter l'interopérabilité entre les programmes qui les utilisent (ZNIEFF, Natura 2000)
- permettre la mise à disposition de ces typologies »

Guide méthodologique de HABREF^[7]

Ainsi une évolution possible du projet serait d'inclure dans le projet l'acquisition et la valorisation des données d'habitats et de végétations en se basant sur le référentiel HABREF comme il est déjà fait pour les données de biodiversité avec le référentiel TAXREF. L'ajout de fonctionnalités sur la thématique des habitats permettrait aux utilisateurs d'avoir une vision d'ensemble sur l'état de l'environnement de la zone ALCOTRA et non seulement de l'état de la biodiversité. Ces fonctionnalités permettraient également aux acteurs territoriaux impliqués dans la conservation de la biodiversité de prendre en considération les habitats et leur conservation lors de leur prise de décision pour des mesures de conservation de la faune et de la flore.

Conclusion

Le projet COBIODIV SI posait la question de la valorisation des données de biodiversité des Alpes franco-italienne et plus particulièrement de la manière dont ces données pouvaient être valorisée au mieux. Après un peu moins d'un mois de stage et une phase d'analyse très complète du projet et de ses enjeux, il s'est révélé que la meilleure solution pour répondre à cette problématique serait de rendre accessible et diffusable ces données à travers un atlas numérique permettant de toucher un large public. Le projet s'est donc dirigé vers le développement d'un atlas numérique, en s'appropriant un outil open-source, GeoNature-atlas, qui a reçu plusieurs ajouts de fonctionnalités pour répondre aux attentes du projets. En parallèle de ce travail de développement de l'outil, tout un travail a été fait autour des données. Le projet utilise des données de biodiversité reçues de plusieurs organismes de deux nationalités différentes. Un premier travail d'étude de la donnée a permis de dégager les problématiques liées à la valorisation de données internationales et de données n'utilisant pas de standards ou référentiels. Outre passant ces problématiques, les données ont pu être intégrées au sein d'une base de données GeoNature (SGBD PostgreSQL avec sa cartouche spatiale PostGIS) qui est reliée par un foreign-data-wrapper à l'outil GeoNature-atlas permettant de visualiser par une application web les données de biodiversité de la zone ALCOTRA.

La mission de valorisation des données de biodiversité du projet COBIODIV SI s'est déroulée sur trois mois et a mobilisé deux stagiaires du Parc des Écrins. Ces contraintes n'ont pas permis d'explorer à leur maximum tous les aspects de la valorisation de ces données mais donne un très bon aperçu de ce qu'il est possible de faire grâce à des données mises conjointement à disposition entre deux pays.

Finalement, le projet a permis à 5 acteurs de la biodiversité de la zone ALCOTRA de valoriser leurs données ensemble au sein d'un même outil montrant ainsi la possibilité d'étudier et d'agir pour la biodiversité sur une zone transfrontalière. Reste encore à se poser la question du devenir de l'outil... Est-il voué à être diffusé ? Par qui sera-t-il pris en charge aussi bien en termes de développements que de maintenance ?



Webographie

- [1] PAJOT Adrien. (05/02/2021). Synthèse de la phase 1 du projet COBIODIV SI. [En ligne] (consulté le 17/05/2021). Disponible sur : <https://geonature.fr/documents/2021-01-PITEM-BiodivAlp-COBIODIV-SI-Synthese-Diagnostic.pdf>
- [2] LECLERC Mathilde. (18/06/2021). Analyse des besoins d'un outil de valorisation des données de biodiversité de la zone ALCOTRA. [En ligne]. Disponible sur : <https://geonature.fr/documents/2021-06-PITEM-BiodivAlp-COBIODIV-SI-Atlas-Analyse-besoins.pdf>
- [3] LECLERC Mathilde. (18/06/2021). État de l'art des outils existants de valorisation des données de biodiversité. [En ligne]. Disponible sur : <https://geonature.fr/documents/2021-06-PITEM-BiodivAlp-COBIODIV-SI-Atlas-Etat-art.pdf>
- [4] LECLERC Mathilde. (06/07/2021). Analyse fonctionnelle de l'outil de valorisation des données de biodiversité sur la zone ALCOTRA. [En ligne]. Disponible sur : <https://geonature.fr/documents/2021-06-PITEM-BiodivAlp-COBIODIV-SI-Atlas-Analyse-fonctionnelle.pdf>
- [5] LANGE Corentin. (06/07/2021). Analyse technique de l'outil de valorisation des données de biodiversité sur la zone ALCOTRA. [En ligne]. Disponible sur : <https://geonature.fr/documents/2021-06-PITEM-BiodivAlp-COBIODIV-SI-Atlas-Analyse-technique.pdf>
- [6] LECHÉMIA Théo. (30/09/2016). Rapport de stage : Création d'un atlas dynamique de la faune et de la flore au Parc National des Écrins. [En ligne]. Disponible sur : https://geonature.fr/documents/2016-09-30-rapport_stage_Theo-Lechemia.pdf
- [7] UMS PatriNat. (15/12/2020). Méthodologie de TAXREF v14.0 en PDF. [En ligne]. Disponible sur : <https://inpn.mnhn.fr/docs-web/docs/download/354850>
- [8] UMS PatriNat. (Novembre 2019). Guide méthodologique HABREF v5.0 en PDF. [En ligne]. Disponible sur : https://inpn.mnhn.fr/docs/ref_habitats/HABREF_5.0/Guide_methodologique_HABREF_V5.0.pdf